Prewhitened Long-Run Variance Estimation Robust to Nonstationarity*

Alessandro Casini†
University of Rome Tor Vergata

Pierre Perron‡
Boston University

27th September 2023

Abstract

We introduce a nonparametric nonlinear VAR prewhitened long-run variance (LRV) estimator for the construction of standard errors robust to autocorrelation and heteroskedasticity that can be used for hypothesis testing in a variety of contexts including the linear regression model. Existing methods either are theoretically valid only under stationarity and have poor finite-sample properties under nonstationarity (i.e., fixed-b methods), or are theoretically valid under the null hypothesis but lead to tests that are not consistent under nonstationary alternative hypothesis (i.e., both fixed-b and traditional HAC estimators). The proposed estimator accounts explicitly for nonstationarity, unlike previous prewhitened procedures which are known to be unreliable, and leads to tests with accurate null rejection rates and good monotonic power. We also establish MSE bounds for LRV estimation that are sharper than previously established and use them to determine the data-dependent bandwidths.

JEL Classification: C12, C13, C18, C22, C32, C51
Keywords: Asymptotic Minimax MSE, Data-dependent bandwidths, HAC, HAR, Long-run variance, Nonstationarity, Prewhitening, Spectral density.

*This paper previously circulated with the title "Minimax MSE Bounds and Nonlinear VAR Prewhitening for Long-Run Variance Estimation Under Nonstationarity". We thank two anonymous referees for helpful suggestions. We thank Oliver Linton, Whitney Newey and Tim Vogelsang for helpful discussions. We also thank seminar participants at University College London, University of Cambridge and University of Connecticut for comments.

†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

Heteroskedasticity and autocorrelation robust (HAR) inference requires estimation of the relevant asymptotic variance or simply the long-run variance (LRV). A large literature has considered this problem. In econometrics, Andrews (1991) and Newey and West (1987; 1994) extended the scope of kernel-based autocorrelation and heteroskedastic consistent (HAC) estimators of the LRV [see also de Jong and Davidson (2000) and Hansen (1992)]. 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. 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 HAC estimators is that the LRV estimator is not consistent under fixed-b asymptotics and inference is nonstandard. Test statistics under the null hypotheses asymptotically 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 yields test statistics with more accurate null rejection rates when there is strong dependence.[1]

There is widespread evidence that the processes governing economic data are nonstationary. By nonstationary we mean non-constant moments. As in the literature, we consider processes whose sum of absolute autocovariances is finite. That is, we rule out processes with unbounded second moments (e.g., unit root). The latter can be handled by taking first-differences or applying some de-trending technique. Nonstationarity can occur for several reasons: changes in the moments induced by changes in the model parameters that govern the data (e.g., the Great Moderation with the decline in variance for many macroeconomic variables or the effects of the COVID-19 pandemic); smooth changes in the distributions of the processes that arise from transitory dynamics; and so on. HAR inference requires the estimation of the LRV of some relevant process, V_t say.[2] We first analyze the case with $\mathbb{E}(V_t) = 0$ for all t, since it is the leading case that applies under the null hypothesis. This will allow us to derive useful properties to construct bandwidths (and so on) to have tests with the correct null rejection rates. Thus, under the null hypothesis, nonstationary occurs through time-varying autocovariances $\mathbb{E}(V_t V_{t-k})$. We recognize that in some cases, the null hypothesis may involve a non-constant mean (e.g., when the model is misspecified).

[1] See Jansson (2004) and Sun, Phillips and Jin (2008) for theoretical results based on asymptotic expansions.

[2] For example, in the linear regression model $V_t = x_t e_t$ where x_t is a vector of regressors and e_t is a disturbance.

As in the literature, we do not address this case since the results can only be obtained on a case by case basis. Under the alternative hypothesis, $\mathbb{E}(V_t) \neq 0$, and $\mathbb{E}(V_t)$ as well as $\mathbb{E}(V_t V_{t-k})$ can be time-varying. In most HAR inference problems the leading case is with a non-zero mean. The literature has so far not properly addressed this leading case. Our aim is to devise a method for this leading case that delivers useful estimates such that the tests have good power. Hence, we shall also consider the properties of our estimator when the mean of V_t is non-zero and show that it leads to test having good monotonic power, unlike what is available in the literature.

The objective of this paper is to propose an estimator of the LRV that has the following properties: (i) it can be used for any hypothesis testing problem both within and outside the linear regression model and is valid under both stationarity and nonstationarity; (ii) it can be used without the need to develop further asymptotic analyses to determine the null limiting distribution of the test statistics; (iii) it leads to tests that have accurate null rejection rates even with strong dependence; such tests are consistent in any hypothesis testing problem, and in particular, in testing problems characterized by a nonstationary alternative hypothesis.[3] None of the existing procedures satisfies all three properties. Fixed-b methods rely on nonstandard limit theory and require one to derive the null limiting distribution on a case-by-case basis.[4] Casini (2023a) showed that the original fixed-b methods are not theoretically valid under nonstationarity since the null limiting distribution of the test statistics is then not pivotal. More recently, a variant of the fixed-b approach [see, e.g., Sun (2014) and Lazarus, Lewis, Stock and Watson (2018)] considered the use of small-b asymptotics (i.e., small-bandwidths) in conjunction with fixed-b critical values. In general, the latter methods do not satisfy (i)-(ii) since they use fixed-b critical values, and we show below that they may not lead to consistent tests under nonstationary alternative hypothesis. Traditional HAC estimators satisfy (i)-(ii) since they are consistent for the LRV so that a test statistic studentized by an HAC estimator follows asymptotically a standard distribution. A long-lasting problem with HAC estimators is that they lead to HAR tests that can be oversized when there is strong dependence. To address this issue, Andrews and Monahan (1992) proposed the prewhitened HAC estimators which substantially reduce the oversized problem under stationarity with HAR tests having null rejection rates similar to those of recent methods based on fixed-b [e.g., the EWP and EWC methods of Lazarus et al. (2021) and Lazarus et al. (2018), respectively]. However, we show theoretically that existing prewhitened and non-prewhitened LRV estimators

[3]By nonstationary alternative hypothesis we mean alternative hypothesis such that $\mathbb{E}(V_t)$ is time-varying.

[4]Lazarus, Lewis and Stock (2021) pointed out the usefulness for empirical work of having test statistics that follow asymptotically standard distributions rather than nonstandard distributions whose critical value has to be obtained by simulations.

lead to HAR tests that are not consistent in contexts characterized by nonstationary alternative hypotheses. This has been a recurrent problem in the time series econometrics literature.[5] It occurs, for instance, when using tests involving structural breaks based on estimating the model under the null hypothesis; e.g., tests for forecast evaluation [e.g., Diebold and Mariano (1995), Giacomini and White (2006) and West (1996)], tests for forecast instability [cf. Casini (2018), Giacomini and Rossi (2009, 2010) and Perron and Yamamoto (2021)], CUSUM tests for structural change [see, e.g., Brown, Durbin and Evans (1975) and Ploberger and Krämer (1992)] tests and inference in time-varying parameters models [e.g., Cai (2007) and Chen and Hong (2012)], tests and inference for regime switching models [e.g., Hamilton (1989) and Qu and Zhuo (2020)].

To improve the power properties of HAR tests based on HAC estimators, Casini (2023b) proposed to modify the HAC estimators by adding a second kernel which applies smoothing over time. Such double kernel HAC estimators (DK-HAC) satisfy properties (i)-(iii) except that they can be oversized when there is high serial correlation. We introduce a novel nonparametric nonlinear VAR prewhitening procedure to apply prior to constructing the DK-HAC estimators. The key property is that our prewhitening procedure is applied locally in time through nonparametric time smoothing. This allows us to account flexibly for the time-varying second-order properties of the data and to reduce the asymptotic bias arising from nonparametric estimation. Our prewhitening is robust to nonstationarity unlike previous prewhitened procedures [e.g., Andrews and Monahan (1992), Preinerstorfer (2017), Rho and Shao (2013) and Xiao and Linton (2002)]. 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 $\widehat{J}_{\text{HAC,pw}} = \widehat{J}_{\text{HAC},V^*}/(1-\widehat{a}_1)^2$ 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 $\widehat{J}_{\text{HAC},V^*}$ 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 nonstationarity increases. Hence, $\widehat{J}_{\text{HAC,pw}}$ is inflated and test statistics lose power.

The consistency, rate of convergence and MSE of the new prewhitening procedure are established under nonstationarity using the segmented locally stationary framework. We then establish the consistency, rate of convergence and minimax MSE bounds for the DK-HAC estimator under general nonstationarity (i.e., unconditionally heteroskedastic processes) and discuss how these

[5] Simulation evidence of serious (e.g., non-monotonic) power problems was documented by Altissimo and Corradi (2003), Casini (2018), Casini and Perron (2019, 2021a, 2020), Chan (2022b), Chang and Perron (2018), Crainiceanu and Vogelsang (2007), Demetrescu and Salish (2020), Deng and Perron (2006), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Otto and Breitung (2021), Perron and Yamamoto (2021), Shao and Zhang (2010), Vogelsang (1999) and Zhang and Lavitas (2018) among others.

results can be used to show that the prewhitened DK-HAC estimators are valid under general nonstationarity. The new minimax MSE bounds generalizes the MSE bounds in Andrews (1991) as follows. Andrews (1991) expressed the bounds in terms of the distributions of two different second-order stationary processes. The two distributions provide upper and lower bounds, respectively, to the autocovariances of the nonstationary processes in some class. We show that this class can be enlarged substantially if the two distributions are taken to be those of some nonstationary processes that satisfy segmented locally stationarity. This allows for more variability of $\mathbb{E}(V_t V_{t-k})$ and serial dependence of $\{V_t\}$. Thus, our bounds apply to a richer class of processes. The new bounds also provide information on how nonstationarity influences the estimation bias.

The prewhitened DK-HAC estimators lead to HAR tests with null rejection rates close to the nominal even with strong dependence. Furthermore, we show theoretically that the prewhitened DK-HAC estimators lead to HAR tests that are consistent even under nonstationary alternative hypotheses whereas existing HAC-based and fixed-b-based HAR tests are not consistent with their power converging to zero as nonstationarity increases. The simulations demonstrate that these theoretical results provide accurate predictions about the finite-sample behavior of the tests.

The paper is organized as follows. Section 2 introduces the nonlinear VAR prewhitening procedure and its asymptotic results are established in Section 3. Section 4 establishes the theoretical validity of the DK-HAC estimators under general nonstationarity and presents new minimax MSE bounds. Section 5 presents some theoretical results about the power of HAR tests under nonstationary alternative hypotheses. Section 6 presents the simulation results. Section 7 concludes. The supplemental materials [cf. Casini and Perron (2021b)] contain all mathematical proofs.

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 $V_t(\beta)$ a random p-vector for each $\beta \in \Theta \subset \mathbb{R}^{p_\beta}$ and $\mathbb{E}(V_t(\beta_0)) = 0$ for all t under the null hypothesis provided that the underlying model is correctly specified. We allow for $\mathbb{E}(V_t) \neq 0$ in Section 5 when we analyze the theoretical properties of the power of the tests. 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,

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

where B_T is a nonrandom $p_\beta \times p$ matrix. Often it is easy to construct estimators $\widehat{B}_T$ such that $\widehat{B}_T - B_T \xrightarrow{\mathbb{P}} 0$. Thus, one needs a consistent estimator of $J = \lim_{T\to\infty} J_T$ to construct a consistent estimator of $\lim_{T\to\infty} B_T J_T B_T'$. Our goal is to consider the estimation of J under nonstationarity.

Under nonstationarity the autocovariance of V_t depends on the calendar time at which it is computed in addition to the lag. That is, $\Gamma_u(k) \triangleq \mathbb{E}(V_{Tu}V'_{Tu-k})$ where $u = t/T$ for some lag $k \in \mathbb{Z}$. The rescaled time index $u \in [0,\, 1]$ is introduced because under nonstationarity we use the infill asymptotics. We now define the local spectral density of V_t at time u and frequency ω, $f(u,\, \omega)$. It is an important quantity because it summarizes the second-order properties of V_t. It is defined as the squared modulus of the transfer function $A(u,\, \omega)$ where the latter appears in the spectral representation of V_t [see eq. (S.1) in the supplement]. That is, $f(u,\, \omega) = |A(u,\, \omega)|^2$. The local spectral density can also be defined implicitly from the definition of $c(u,\, k)$ which is the approximation to the local autocovariance $\Gamma_u(k)$ where

$$c(u,\, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} f(u,\, \omega)\, d\omega, \tag{2.1}$$

and $i = \sqrt{-1}$. In fact, Lemma S.A.1 in Casini (2023b) showed that, under the assumptions we introduce below, $\Gamma_u(k) = c(u,\, k) + O(T^{-1})$ where $O(T^{-1})$ is the error due to the infill asymptotic approximation. Eq. (2.1) relates the local autocovariance of $\{V_t\}$ at rescaled time u and lag k to its local spectral density at u. Thus, the nonstationary properties of $\{V_t\}$, which are reflected in the time-varying behavior of the autocovariance function $\Gamma_u(k)$, depend on the smoothness properties of $f(u,\, \omega)$ in u. For example, if $\{V_t\}$ is stationary, then $\Gamma_u(k) = \Gamma(k)$ for all u, $c(u,\, k)$ is constant in u, $f(u,\, \omega) = f(\omega)$ and (2.1) reduces to $\Gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k} f(\omega)\, d\omega$. These coincide with textbook definitions used under stationarity [see, e.g., Brillinger (1975)]. If $f(u,\, \omega)$ is continuous in u then $\{V_t\}$ is locally stationary [cf. Dahlhaus (1997)].[6] For example, consider a time-varying AR(1) $V_t = a(t/T)\, V_{t-1} + u_t$ where u_t is a zero-mean i.i.d. process with unit variance and $a(\cdot)$ is continuous with $a(t/T) \in (-1,\, 1)$ for all t. Then V_t is a locally stationary AR(1) with a local spectral density $f(u,\, \omega)$ that is continuous in u. We impose restrictions on the smoothness of $f(u,\, \omega)$ in u which allow for considerable forms of nonstationarity in $\{V_t\}$ including most of the

[6] In econometrics, locally stationary processes are often referred to as time-varying parameter processes.

nonstationary models used in econometrics.[7]

Assumption 2.1. *(i)* $\{V_t\}$ *is zero-mean with local spectral density* $f(u, \omega)$ *that is piecewise Lipschitz continuous with* m_0 *discontinuity points; (ii)* $f(u, \omega)$ *is twice continuously differentiable in* u *at all continuity points with bounded derivatives* $(\partial/\partial u) f(u, \cdot)$ *and* $(\partial^2/\partial u^2) f(u, \cdot)$*, and Lipschitz continuous in the second component; (iii)* $(\partial^2/\partial u^2) f(u, \cdot)$ *is Lipschitz continuous at all continuity points; (iv)* $f(u, \omega)$ *is twice left-differentiable at all discontinuity points with bounded derivatives* $(\partial/\partial_- u) f(u, \cdot)$ *and* $(\partial^2/\partial_- u^2) f(u, \cdot)$ *and has piecewise Lipschitz continuous derivative* $(\partial^2/\partial_- u^2) f(u, \cdot)$*.*

Assumption 2.1 implies that $\{V_t\}$ is segmented locally stationary (SLS) (see Definition S.A.1 in the supplement). It is similar to Assumption 3.1 in Casini (2023b) where the latter imposes smoothness conditions on the transfer function $A(u, \omega)$ whereas here we directly make assumptions on the local spectral density $f(u, \omega)$. The class of SLS processes allows for relevant features such as structural change, regime switching-type and threshold model and includes general time-varying parameter processes, locally stationary processes and stationary processes.[8] Assumption 2.1 requires $f(u, \cdot)$ to be twice differentiable at the continuity points and left-differentiable at the discontinuity points. The zero-mean assumption holds under the null hypothesis. To focus on the main intuition, we first consider the case of SLS processes and then extend the results to general nonstationarity processes in Section 4.[9] The latter require more technical notations and assumptions. In Section 2.1 we present the prewhitening DK-HAC estimator while in Section 2.2 we discuss its data-dependent bandwidths.

2.1 Prewhitening DK-HAC Estimator

Under Assumption 2.1, the argument at the beginning of Section 2.1 in Casini (2023b) suggests that $J = 2\pi \int_0^1 f(u, 0)\, du$. The right-hand side can be seen as a function, say $\widetilde{f}(\omega)$, evaluated at the zero frequency $\omega = 0$. The intuition behind prewhitening is simple, though the mechanics under nonstationarity are quite different. Suppose one is estimating $\widetilde{f}(0)$ nonparametrically by averaging asymptotically unbiased estimators of $\widetilde{f}(\omega)$ at a number of points ω in a neighborhood of 0. The

[7]A function $g(\cdot) : [0, 1] \mapsto \mathbb{R}$ is said to be piecewise (Lipschitz) continuous if there exists a finite subdivision $\{x_0, x_1, \ldots, x_n\}$ of $[0, 1]$ where $x_0 = 0$ and $x_n = 1$, such that for all $i \in \{1, 2, \ldots, n\}$ g is (Lipschitz) continuous on (x_{i-1}, x_i).

[8]For general time-varying parameter processes we mean linear and nonlinear processes whose parameters can change smoothly as well as abruptly. See Example 2.1 in Casini (2023b) for some examples.

[9]For general nonstationarity we mean a process with a time-varying spectral density that does not satisfy piecewise Lipschitz continuity.

flatter is the function $\widetilde{f}(\omega)$ around 0, the smaller the estimation bias. The idea is to transform the data such that the function of the transformed data, say $\widetilde{f}^*(\omega)$, is flatter in the neighborhood of $\omega = 0$. Then, using the transformed data one can estimate $\widetilde{f}^*(0)$ by averaging asymptotically unbiased estimators of $\widetilde{f}^*(\omega)$ at points ω in the neighborhood of 0. The resulting bias should be less than that incurred by estimating $\widetilde{f}(0)$ since $\widetilde{f}^*(\omega)$ is flatter than $\widetilde{f}(\omega)$. Finally, one can apply the inverse of the transformation from $\widetilde{f}(\omega)$ to $\widetilde{f}^*(\omega)$ to obtain an estimator of $\widetilde{f}(\omega)$ from the estimator of $\widetilde{f}^*(\omega)$. This is how it works under stationarity. However, under nonstationarity one applies both the transformation and the inverse transformation locally in time, otherwise the prewhitening procedure may be unreliable as nonstationarity induces an additional source of bias in both the transformation and its inverse.

The proposed prewhitening procedure is based on the following three steps:

Step 1 (whitening step): Divide the sample in $\lfloor T/n_T \rfloor$ blocks, each with n_T observations. Let $\widehat{V}_t = V_t(\widehat{\beta})$, where $\widehat{\beta}$ is a $\sqrt{T}$-consistent estimator of β_0. For each block $r = 0, \ldots, \lfloor T/n_T \rfloor$, run the following VAR($p_A$),

$$\widehat{V}_t = \sum_{j=1}^{p_A} \widehat{A}_{r,j} \widehat{V}_{t-j} + \widehat{V}_t^* \quad \text{for} \quad t = rn_T + 1, \ldots, (r+1)\, n_T, \tag{2.2}$$

where $\widehat{A}_{r,j}$ for $j = 1, \ldots, p_A$ are $p \times p$ least-squares estimators and $\widehat{V}_t^* = V_t^*(\widehat{\beta})$ are the prewhitened residuals. The VAR in (2.2) is used to "soak up" some of the serial dependence in $\widehat{V}_t^*$ and to leave one with residuals $\{\widehat{V}_t^*\}$ that are closer to white noise.[10] That is why it is called "whitening step".

Step 2 (recoloring step): Take the prewhitened residuals $\widehat{V}_t^*$, transform them by applying an inverse transformation $\widehat{V}_t^* \mapsto \widehat{V}_{D,t}^* = \widehat{D}_t \widehat{V}_t^*$ where $\widehat{D}_t = (I_p - \sum_{j=1}^{p_A} \widehat{A}_{D,t,j})^{-1}$ with $\widehat{A}_{D,t,j} = \widehat{A}_{r,j}$ for $t = rn_T + 1, \ldots, (r+1)\, n_T$. This implies that the transformed residuals $\widehat{V}_{D,t}^*$ have been "recolored" (i.e., the dependence has been added back). Note that the matrix $\widehat{D}_t$ is the same for all t in a given block. In this way the appropriate amount of dependence is added back, i.e., no contamination from possibly different strengths of dependence occurring in other blocks.

Step 3 (prewhitened DK-HAC estimation): Construct the prewhitened DK-HAC estimator

[10] Since the residuals $\{\widehat{V}_t^*\}$ are closer to a white noise process, they have a flatter spectral density at $\omega = 0$ than $\{\widehat{V}_t\}$ because a white noise process has a flat spectral density.

$\widehat{J}_{T,\text{pw}}$ using $\widehat{V}^*_{D,t}$:

$$\widehat{J}_{\text{pw},T}\left(\widehat{b}^*_{1,T},\,\widehat{b}^*_{2,T}\right)=\frac{T}{T-p}\sum_{k=-T+1}^{T-1}K_1\left(\widehat{b}^*_{1,T}k\right)\widehat{\Gamma}^*_D\left(k\right), \tag{2.3}$$
$$\text{where}\quad \widehat{\Gamma}^*_D\left(k\right)\triangleq\frac{n_T}{T-n_T}\sum_{r=0}^{\lfloor (T-n_T)/n_T\rfloor}\widehat{c}^*_{T,D}\left(rn_T/T,\,k\right),$$

with $K_1(\cdot)$ a real-valued kernel in the class $\boldsymbol{K}_3$ defined below, $\widehat{b}^*_{1,T}$ is a data-dependent bandwidth sequence to be discussed below, $n_T\to\infty$, and

$$\widehat{c}^*_{D,T}\left(rn_T/T,\,k\right)\triangleq\begin{cases}\left(T\widehat{b}^*_{2,T}\right)^{-1}\sum_{s=k+1}^{T}K^*_2\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}^*_{2,T}}\right)\widehat{V}^*_{D,s}\widehat{V}^{*\prime}_{D,s-k}, & k\geq 0\\ \left(T\widehat{b}^*_{2,T}\right)^{-1}\sum_{s=-k+1}^{T}K^*_2\left(\frac{((r+1)n_T-(s+k/2))/T}{\widehat{b}^*_{2,T}}\right)\widehat{V}^*_{D,s+k}\widehat{V}^{*\prime}_{D,s}, & k<0\end{cases},$$

K^*_2 being a kernel, $\widehat{b}^*_{2,T}$ a data-dependent bandwidth sequence to be defined below.

In order to guarantee positive semi-definiteness, one needs to use a data taper or, e.g., for $k\geq 0$,

$$K^*_2\left(\frac{(r+1)\,n_T-(s-k/2)}{T\widehat{b}^*_{2,T}}\right)=\left(K_2\left(\frac{(r+1)\,n_T-s}{T\widehat{b}^*_{2,T}}\right)K_2\left(\frac{(r+1)\,n_T-(s-k)}{T\widehat{b}^*_{2,T}}\right)\right)^{1/2},$$

see Casini (2023b).

In Step 1 the the last block is $t=\lfloor T/n_T\rfloor n_T+1,\ldots,T$. 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 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 $\{\widehat{V}_t\}$ is introduced in Casini and Perron (2023). The latter could be employed here to efficiently determine the sample-splitting. This would result in the sample being split in blocks with the property that within each block $\{V_t(\widehat{\beta})\}$ is locally stationary. However, this is not required for the theoretical validity. The least-squares estimation within blocks yields consistent estimators $\widehat{A}_{r,j}$ for some $A_{r,j}$ even when the fitted VAR is not the true model. The fitted VAR is used only to yield residuals $\{\widehat{V}^*_t\}$ that are closer to white noise so that their spectral density at zero is flatter, implying less asymptotic bias when estimating it nonparametrically.

Below we assume that $\widehat{A}_{r,j}\overset{\mathbb{P}}{\to}A_{r,j}$ for some $A_{r,j}\in\mathbb{R}^{p\times p}$ for all r and j which follow from standard arguments. For K_1 we suggest using the Quadratic Spectral (QS) kernel

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

and for K_2 a quadratic-type kernel [cf. Epanechnikov (1969)] given by $K_2(x) = 6x(1-x)$, $0 \leq x \leq 1$. These kernels are optimal under an MSE criterion [see Casini (2023b)].

There has been some recent works on LRV estimation in statistics that relate to ours. Kawka (2020) studied the asymptotic properties of classical spectral estimators for a linear time-varying AR process where the AR coefficients can have a finite number of discontinuities. Since classical spectral estimators do not involve any local smoothing over time, and since he focused on linear processes and did not consider data-dependent bandwidths, his framework required simpler assumptions. He also considered an estimate of the spectrum profile which is defined similarly to the variance profile of Cavaliere and Taylor (2007). That is based on a recursive estimate of the spectral density which is, however, different from applying local smoothing. The local smoothing is important to better account for nonstationarity as shown in Casini, Deng and Perron (2023). Potiron and Mykland (2020) showed that in the context of estimation of higher powers of volatility for high frequency data the local smoothing can lead to substantial efficiency gains. Although our setting is complicated by serial dependence and the fact that the class of estimators has a slower rate of convergence than the parametric $\sqrt{T}$-rate, the theoretical results on the power of the HAR tests below suggest that the local smoothing yields more powerful tests. In addition, Casini et al. (2023) showed that under nonstationarity the sample autocovariance can be upward biased asymptotically relative to the integrated local sample autocovariance, both for fixed lag k and for $k \to \infty$. An alternative way to deal with a time-varying mean has been considered by Chan (2022a, 2022b) who proposed a LRV estimator which uses difference-based statistics that combine local smoothing and lagged differences of the series. His results confirmed that the local smoothing is important to enhance efficiency. However, he required covariance stationarity and did not study the theoretical properties of HAR tests normalized by the proposed LRV estimator.

2.2 Data-Dependent Bandwidths

For data-dependent bandwidths, we use plug-in estimates of the optimal value that minimizes some MSE criterion, see Section 4 and Casini (2023b). Let $\Gamma_{D,u}(k) = \mathrm{Cov}(V_{D,Tu}^*, V_{D,Tu-k}^*)$ 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 some $p^2 \times p^2$ weight matrices. Let $F(K_2) \triangleq \int_0^1 K_2^2(x)\,dx$, $H(K_2) \triangleq (\int_0^1 x^2 K_2(x)\,dx)^2$,

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

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

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

The optimal $b_{2,T}$ is given by [see Casini (2023b)]

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

Let

$$K_{1,q} \triangleq \lim_{x \downarrow 0} (1 - K_1(x)) / |x|^q \qquad \text{for} \qquad q \in [0,\, \infty); \tag{2.4}$$

$K_{1,q} < \infty$ 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

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

The optimal $b_{1,T}$ given the optimal value $b_{2,T}^{\text{opt},*}$ is given by [see Casini (2023b)],

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

with $\overline{b}_{2,T}^{\text{opt},*} = \int_0^1 b_{2,T}^{\text{opt},*}(u)\, du$. 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 $H(K_2^{\text{opt}}) = 0.09$ and $F(K_2^{\text{opt}}) = 1.2$.

The bandwidths $(b_{1,T}^{\text{opt},*},\, \overline{b}_{2,T}^{\text{opt},*})$ are optimal under a sequential MSE criterion that determines the optimal b_1 as a function of the optimal $b_2(u)$. Thus, the latter influences the former but not vice-versa. However, this has the advantage that the optimal $b_2(u)$ is allowed to change over time. Belotti et al. (2023) proposed an alternative criterion that determines the optimal b_1 and b_2 that jointly minimize the global MSE. The latter yields an optimal b_2 that does not depend on u and so it does not perform as well as the sequential method when the data is far from stationary.

In order to construct a data-dependent bandwidth for $b_{2,T}(u)$, we need consistent estimators of $D_{1,D}(u)$ and $D_{2,D}(u)$. We set $\widetilde{W}^{(r,r)} = p^{-1}$ for all r which corresponds to the normalization used below for W. The estimator of $D_{1,D}(u)$ is then,

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

where $[S_\omega]$ is the cardinality of S_ω and $\omega_{s+1} > \omega_s$ with $\omega_1 = -\pi$, $\omega_{[S_\omega]} = \pi$. We set $S_\omega = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\}$. The estimator of $D_{2,D}(u)$ is given by

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

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

$$\widehat{\overline{b}}_{2,T}^{*} = (n_T/T)\sum_{r=1}^{\lfloor T/n_T\rfloor-1}\widehat{b}_{2,T}^{*}(u_r), \tag{2.5}$$

$$\text{where}\quad \widehat{b}_{2,T}^{*}(u_r) = 1.7781(\widehat{D}_{1,D}(u_r))^{-1/5}(\widehat{D}_{2,D}(u_r))^{1/5}T^{-1/5}, \quad u_r = rn_T/T. \tag{2.6}$$

The data-dependent bandwidth parameter $\widehat{b}_{1,T}^{*}$ is then defined as follows. First, one specifies p univariate approximating parametric models given by $\{V_{D,t}^{*(r)}\}$ 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

$$\widehat{b}_{1,T}^{*} = \left[2qK_{1,q}^2\widehat{\phi}_D(q)\,T\widehat{\overline{b}}_{2,T}^{*}/\left(\int K_1^2(x)\,dx\int K_2^2(x)\,dx\right)\right]^{-1/(2q+1)}. \tag{2.7}$$

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(1)) models given by $V_{D,t}^{*(r)} = a_1^{(r)}(t/T)\, V_{D,t-1}^{*(r)} + u_t^{(r)}$, $r = 1, \ldots, p$. Let $\widehat{a}_1^{(r)}(u)$ and $(\widehat{\sigma}^{(r)}(u))^2$ be the least-squares estimators of the autoregressive and innovation variance parameters computed using data close to $u = t/T$:

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

where $n_{2,T} \to \infty$.[11] These are simply least-squares estimators based on rolling windows. Then, for $q = 2$, we have

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

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 in which case it is zero.

3 Large-Sample Results When $\mathbb{E}(V_t) = 0$

In this section, we analyze the asymptotic properties of $\widehat{J}_{\mathrm{pw},T}$ for the case with $\mathbb{E}(V_t) = 0$ for all t, which is relevant under the null hypothesis provided that the model is correctly specified. Let

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

[11] See, for example, Dahlhaus and Giraitis (1998) for a discussion about nonparametric local parameter estimates in the context of locally stationary time series.

$\boldsymbol{K}_3$ 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, $q = 1$.[12] The condition $q < 34/4$ in part (iv) is a technical condition needed to control the deviation $|\widehat{b}_{1,T}^{*} - b_{\theta_1,T}|$, where $b_{\theta_1,T}$ is defined similarly to $\widehat{b}_{1,T}^{*}$ [cf. (3.1) below] but with $\widehat{\phi}_D(q)$ replaced by some ϕ_{θ^*} such that $\phi_{\theta^*} \xrightarrow{\mathbb{P}} \widehat{\phi}_D(q)$ (see Assumption 3.5 below).

We consider the same class of kernels $\boldsymbol{K}_2$ as considered by Casini (2023b):

$$\begin{aligned}\boldsymbol{K}_2 = \Bigg\{ & K_2(\cdot) : \mathbb{R} \to [0, \infty] : K_2(x) = K_2(1-x), \int K_2(x)\, dx = 1, \\ & \int_0^1 K_2(x)\, dx < \infty,\ K_2(x) = 0, \text{ for } x \notin [0, 1] \\ & |K_2(x) - K_2(y)| \le C_4 |x-y| \text{ for all } x, y \in \mathbb{R} \text{ and some constant } \mathrm{C}_4 < \infty \Bigg\}.\end{aligned}$$

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 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 3.1. *(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 $a_1, a_2, a_3, a_4 \le p$. (ii) For all $a_1, a_2, a_3, a_4 \le p$ there exists a function $\widetilde{\kappa}_{a_1,a_2,a_3,a_4} : [0, 1] \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ that is piecewise continuous in the first argument such that $\sup_{u\in[0,1]} |\kappa_{V,\lfloor Tu \rfloor}^{(a_1,a_2,a_3,a_4)}(k, s, l) - \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, k, s, l)| \le KT^{-1}$ for some $K < \infty$; $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, k, s, l)$ is twice differentiable in u at all continuity points with 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 all discontinuity points with 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*

[12] Recall that $K_{1,q}$ is defined in (2.4).

piecewise Lipschitz continuous derivative $(\partial^2/\partial_- u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$.

If $\{V_{t,T}\}$ is stationary then the cumulant condition of Assumption 3.1-(i) reduces to the standard one used in the time series literature [see, e.g., Assumption A in Andrews (1991)]. We do not require fourth-order stationarity but only that the time-$t = Tu$ fourth order cumulant is locally constant in a neighborhood of a continuity point u. As explained in Casini (2023b), using an argument similar to that used in Lemma 1 in Andrews (1991), one can show that α-mixing and moment conditions imply that the cumulant condition of Assumption 3.1-(i) holds. Part (ii) essentially requires that the approximating cumulant function $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\,k,\,s,\,l)$ satisfies similar smoothness restrictions as $f(u,\,\cdot)$ (i.e., twice differentiability at the continuity points and twice left-differentiable at the discontinuity points).

Assumption 3.2. *(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)* $\sup_{u\in[0,\,1]}\mathbb{E}\sup_{\beta\in\Theta}||\,(\partial/\partial\beta')\,V_{\lfloor Tu\rfloor}(\beta)\,||^2<\infty$.

Assumption 3.2-(i,iii) is an extension of Assumption B in Andrews (1991) to a nonstationary setting. Part (i) follows from asymptotic normality of $\sqrt{T}(\widehat{\beta}-\beta_0)$. Part (ii)-(iii) are common conditions used to obtain the asymptotic normality of $\sqrt{T}(\widehat{\beta}-\beta_0)$ under nonstationarity. In order to obtain rate of convergence results we shall replace Assumption 3.1 with the following assumption.

Assumption 3.3. *(i) Assumption 3.1-(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) $\sup_{u\in[0,\,1]}\mathbb{E}(\sup_{\beta\in\Theta}||\,(\partial^2/\partial\beta\partial\beta')\,V^{(r)}_{\lfloor Tu\rfloor}(\beta)\,||)^2<\infty$ *for all* $r=1,\ldots,p$.

Assumption 3.4. *Let W_T denote a $p^2\times p^2$ weight matrix such that $W_T\overset{\mathbb{P}}{\to}W$.*

Assumption 3.5. *(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* $n_{2,T}/T+n_{3,T}/T\to 0$, $n_{2,T}^{5/4}/T\to[c_2,\,\infty)$, $n_{3,T}^{10/6}/T\to[c_3,\,\infty)$ *with* $0<c_2,\,c_3<\infty$*; (iii)* $\sup_{u\in[0,1]}\lambda_{\max}(\Gamma^*_{D,u}(k))\le C_3k^{-l}$ *for all* $k\ge 0$ *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)* $\omega_{s+1}-\omega_s\to 0$, $[S_\omega]^{-1}\to\infty$ *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* $u\in[0,\,1]$.

Assumption 3.3 is needed to show that the effect of using $\widehat{\beta}$ rather than β_0 when constructing $\widehat{J}_{\mathrm{pw},T}$ is at most $o_{\mathbb{P}}(1)$; it is an extension of Assumption C in Andrews (1991). Parts (i)-(ii) of Assumption 3.5 are the nonparametric analogue to Assumption E-F in Andrews (1991). Part (iii) is satisfied if $\{V_t\}$ is strong mixing with mixing numbers that are less stringent than those sufficient for the cumulant condition in Assumption 3.1-(i). Part (iv) and (vi) extend (i)-(ii) to $\widehat{D}_1$ and $\widehat{D}_2$. Part (v) is needed to apply the convergence of Riemann sums. Under Assumption 3.5 the effect of using the bandwidths $\widehat{b}_{1,T}^*$ and $\widehat{b}_{2,T}^*$ rather than $b_{\theta_1,T}$ and $\bar{b}_{\theta_2,T}$ (defined below in (3.1)) when constructing $\widehat{J}_{\mathrm{pw},T}$ is at most $o_{\mathbb{P}}(1)$.

Assumption 3.6. *$\sqrt{n_T}(\widehat{A}_{r,j}-A_{r,j})=O_{\mathbb{P}}(1)$ for some $A_{r,j}\in\mathbb{R}^{p\times p}$ for all $j=1,\ldots,p_A$ and all $r=0,\ldots,\lfloor T/n_T\rfloor$.*

Given the restrictions below on n_T, Assumption 3.6 is satisfied by standard nonparametric estimators. For the consistency of $\widehat{J}_{T,\mathrm{pw}}$, Assumption 2.1, 3.1-3.2, 3.5-(i,iv) and 3.6 are sufficient. For the rate of convergence and asymptotic MSE results additional conditions are needed. Let

$$b_{\theta_1,T}=\left(2qK_{1,q}^2\phi_{\theta^*}T\bar{b}_{\theta_2,T}/\left(\int K_1^2(y)\,dy\int_0^1 K_2^2(x)\,dx\right)\right)^{-1/(2q+1)}, \tag{3.1}$$

where $\bar{b}_{\theta_2,T}\triangleq\int_0^1[H(K_2)\,D_1(u)]^{-1/5}\left(F(K_2)\,D_2(u)\right)^{1/5}T^{-1/5}du$. Recall that the bandwidths $\widehat{\bar{b}}_{2,T}^*$, $\widehat{b}_{2,T}^*$ and $\widehat{b}_{1,T}^*$ are defined by (2.5), (2.6) and (2.7), respectively.

Theorem 3.1. *Suppose $K_1(\cdot)\in\boldsymbol{K}_3$, q is as in $\boldsymbol{K}_3$, $K_2(\cdot)\in\boldsymbol{K}_2$, $\|\int_0^1 f_D^{*(q)}(u,0)\|<\infty$. Then, we have:*

(i) If Assumption 2.1, 3.1-3.2, 3.5-(i,iv) and 3.6 hold, $\sqrt{n_T}\widehat{b}_{1,T}^*\to\infty$, and $q>1/2$, then $\widehat{J}_{T,\mathrm{pw}}(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*)-J_T\xrightarrow{\mathbb{P}}0$.

(ii) If Assumption 2.1, 3.1-(ii), 3.2-3.3, 3.5-(ii,iii,v,vi) and 3.6 hold, and $n_T/(T\widehat{b}_{1,T}^*)\to0$, $n_T/(T(\widehat{b}_{1,T}^*)^q)\to0$, $T\widehat{\bar{b}}_{2,T}^*/(n_T^2\widehat{b}_{1,T}^*)\to0$, $T\widehat{\bar{b}}_{2,T}^*\widehat{b}_{1,T}^*/n_T\to0$, 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 $\gamma_{K,q}=2qK_{1,q}^2\phi_{\theta^*}/(\int K_1^2(y)\,dy\int_0^1 K_2^2(x)\,dx)$. If Assumption 3.2-3.4 and 3.5-(ii,iii,v,vi)

hold, then

$$\lim_{T\rightarrow\infty}\mathrm{MSE}(T^{8q/5(2q+1)},\, \widehat{J}_{\mathrm{pw},T}(\widehat{b}_{1,T}^{*},\, \widehat{\overline{b}}_{2,T}^{*}),\, J_{T},\, W_{T})$$
$$= 4\pi^{2}\left[\gamma_{K,q}K_{1,q}^{2}\mathrm{vec}\left(\int_{0}^{1}f_{D}^{*(q)}\left(u,\,0\right)du\right)'W\mathrm{vec}\left(\int_{0}^{1}f_{D}^{*(q)}\left(u,\,0\right)du\right)\right]$$
$$+\int K_{1}^{2}\left(y\right)dy\int K_{2}^{2}\left(x\right)dx\,\mathrm{tr}\left[W\left(I_{p^{2}}-C_{pp}\right)\left(\int_{0}^{1}f_{D}^{*}\left(u,\,0\right)du\right)\otimes\left(\int_{0}^{1}f_{D}^{*}\left(u,\,0\right)du\right)\right].$$

A result corresponding to Theorem 3.1 for non-prewhitened DK-HAC estimators is established in Theorem 5.1 in Casini (2023b) under the same assumptions with the exception of Assumption 3.6. Note that for u a continuity point, $f_{D}^{*}\left(u,\,\omega\right)=D\left(u,\,\omega\right)f^{*}\left(u,\,\omega\right)D\left(u,\,\omega\right)'$, where $D\left(u,\,\omega\right)=(I_{p}-\sum_{j=1}^{p_{A}}A_{D,j}\left(u\right)e^{-ij\omega})^{-1}$ with $A_{D,j}\left(u\right)=A_{D,Tu,j}+O\left(T^{-1}\right)$ and $f^{*}\left(u,\,\omega\right)$ is the local spectral density of $\{V_{t}^{*}\}$. Since $D\left(u-k/T,\,\omega\right)=D\left(u,\,\omega\right)+O\left(T^{-1}\right)$ by local stationarity, we have

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

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 $\widehat{J}_{\mathrm{pw},T}$ and $\widehat{J}_{T}$ are used we find that $\widehat{J}_{T,\mathrm{pw}}$ has smaller asymptotic MSE than $\widehat{J}_{T}$ if and only if (assuming $p=1$, i.e., the scalar case, with $w_{1,1}=1$)

$$\underbrace{\int_{0}^{1}f_{D}^{*(q)}\left(u,\,0\right)du}_{\text{squared bias}}\underbrace{\left(\int_{0}^{1}f_{D}^{*}\left(u,\,0\right)du\right)^{2q}}_{\text{variance}}<\underbrace{\int_{0}^{1}f^{(q)}\left(u,\,0\right)du}_{\text{squared bias}}\underbrace{\left(\int_{0}^{1}f\left(u,\,0\right)du\right)^{2q}}_{\text{variance}}.\qquad(3.2)$$

A numerical comparison would be tedious since the condition depends on the true data-generating process of $\{V_{t}\}$ and the VAR approximation for $\widehat{V}_{t}=V_{t}(\widehat{\beta})$. 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 (3.2) 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

$$|f^{*(q)}\left(0\right)|D^{2}<|f^{q}\left(0\right)|,\qquad(3.3)$$

where the quantities $f^{q}\left(0\right)$ and $f^{*(q)}\left(0\right)$ do not depend on u by stationarity. The main difference between the two conditions (3.2)-(3.3) is that the part involving the asymptotic variance is missing in (3.3). 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 (3.2) 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.

Both the smoothing over lagged autocovariances and over time influence the bias of $\widehat{J}_{\mathrm{pw},T}$. The contribution to the bias due to smoothing over lagged autocovariances is $O(b_{1,T}^q)$ while the contribution due to smoothing over time is $O(b_{2,T}^2)$. Note that the continuity points and the discontinuity points here induce a bias of the same order $b_{2,T}^2$. For the continuity points, $O(b_{2,T}^2)$ follows from the usual argument. In the neighborhood of a discontinuity point $[\lambda_j^0 - b_{2,T},\, \lambda_j^0 + b_{2,T}]$, the bias of the local smoothing is $O(b_{2,T})$. However, when averaging over blocks or equivalently integrating over $u \in [0,\, 1]$, this bias becomes $O(b_{2,T}^2)$ since there are only a finite number of discontinuity points and so each discontinuity point contributes $O(b_{2,T}^2)$ to the integrated bias. For $\widehat{J}_{\mathrm{pw},T}$ we have $(\widehat{\overline{b}}_{2,T}^{*})^2/(\widehat{b}_{1,T}^{*})^q \to 0$ since $q = 2$. Thus, the bias due to smoothing over lagged autocovariances dominates the bias due to smoothing over time.

4 Extension to General Nonstationary Random Variables

In this section we discuss the case where $\{V_t\}$ is unconditionally heteroskedastic 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 $\widehat{V}_{D,t}^*$. We use the new MSE bounds to show that the data-dependent bandwidths for the DK-HAC estimator are minimax MSE-optimal also under general nonstationarity. Corresponding results for the prewhitened estimator $\widehat{J}_{\mathrm{pw},T}$ can be obtained by using the results of Section 3, 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

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

where

$$\Gamma_{\mathscr{P},T}(k) = \begin{cases} T^{-1} \sum_{t=k+1}^{T} \mathbb{E}_{\mathscr{P}}(V_t V'_{t-k}), & \text{for } k \geq 0 \\ T^{-1} \sum_{t=-k+1}^{T} \mathbb{E}_{\mathscr{P}}(V_{t+k} V'_t), & \text{for } k < 0 \end{cases}.$$

The estimand is then given by

$$J_{\mathscr{P},T} \triangleq \sum_{k=-T+1}^{T-1} \Gamma_{\mathscr{P},T}(k). \tag{4.1}$$

The theoretical bounds are derived in terms of two distributions $\mathscr{P}_w$, $w = L, U$, under which $\{V_t\}$ is zero-mean SLS with m_0+1 regimes and satisfies Assumption 2.1 and 3.1 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

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

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

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

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

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. This approach was also used by Andrews (1991) who, however, relied on stationarity. $\boldsymbol{P}_U$ 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 $\mathscr{P}_L$. 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

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

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 2.1 and 3.1:

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

Let $\text{MSE}_{\mathscr{P}}(\cdot)$ denote the MSE of $\cdot$ under $\mathscr{P}$ and let $\boldsymbol{K}_{1,+} = \{K_1(\cdot) \in \boldsymbol{K}_1 :\ K_1(x) \geq 0\,\forall x\}$. $\boldsymbol{K}_{1,+}$ 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

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

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

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* $K_{1,q}$*,* $|\int_0^1 f_{\mathscr{P}_w,a}^{(q)}(u, 0)\, du| \in [0, \infty)$*,* $w = L, U$*,* $a \in \mathbb{R}^p$*, then we have:*

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

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

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

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

The theoretical bounds in Theorem 4.1 are sharper than the ones in Andrews (1988; 1991) which are based on stationarity (i.e., the autocovariances that dominate the autocovariances of any $\mathscr{P} \in \boldsymbol{P}_U$ are assumed by Andrews (1991) to be those of a stationary process).[13] Given that stationarity is a special case of SLS, our bounds apply to a wider class of processes. Furthermore, they are more informative because they change with the specific type of nonstationarity unlike Andrews' (1991) bounds that depend on the spectral density of a stationary process.

The theorem is derived under $b_{2,T}^2/b_{1,T}^q \to 0$ (i.e., the bias due to smoothing over time is of smaller order than that due to smoothing over lagged autocovariances). 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

$$\left(\pi\nu \int_0^1 x^2 K_2(x)\, dx \int_{\widetilde{\mathbf{C}}_{\mathscr{P}_\mathbf{U}}} \left(\partial^2/\partial u^2\right) f_{\mathscr{P}_U,a}(u, 0)\, du + 2\pi\nu\Delta_{f_{\mathscr{P}_U,a}}(0)\right)^2 + \Xi,$$

[13]There are a couple of technical issues in Section 8 in Andrews (1991). In particular, the MSE bound is not correct. See Casini (2022) for details.

where $\widetilde{\mathbf{C}}_{\mathscr{P}_U}$ is the set of continuity points under $\mathscr{P}_U$,

$$\Delta_{f_{\mathscr{P}_U,a}}(\omega)=\sum_{j=1}^{m_0}\int_0^1\left(\frac{\partial}{\partial u_-}f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)\int_0^{1-s}xK_2(x)\,dx+\frac{\partial}{\partial u_+}f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)\int_{1-s}^1 xK_2(x)\,dx\right)ds,$$

with $\left\{\lambda_j^0\right\}_{j=1}^{m_0}$ being the discontinuity points, m_0 being a finite integer,

$$\frac{\partial}{\partial u_-}f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)=\lim_{h\uparrow 0}\frac{f_{\mathscr{P}_U,a}\left(\lambda_j^0+h,\,\omega\right)-f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)}{h},$$
$$\frac{\partial}{\partial u_+}f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)=\lim_{h\downarrow 0}\frac{f_{\mathscr{P}_U,a}\left(\lambda_j^0+h,\,\omega\right)-f_{\mathscr{P}_U,a}\left(\lambda_j^0,\,\omega\right)}{h},$$

and Ξ depends on the cross-products of the bias terms due to smoothing over time and lagged autocovariances. 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 et al. (2021). 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 $V_t(\widehat{\beta})$. The following assumptions extend Assumption 3.2-3.3 to the distribution $\mathscr{P}$.

Assumption 4.2. *Assumption 3.2 holds with $\mathbb{E}$ replaced by $\mathbb{E}_{\mathscr{P}}$.*

Assumption 4.3. *(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) $\sup_{u\in[0,\,1]}\mathbb{E}_{\mathscr{P}}(\sup_{\beta\in\Theta}||(\partial^2/\partial\beta\partial\beta')V_{\lfloor Tu\rfloor}^{(a_r)}(\beta)\,||^2)<\infty$ $(r=1,\ldots,p)$.*

To show the asymptotic equivalence of the MSE of $a'\widehat{J}_Ta$ to that of $a'\widetilde{J}_Ta$ we need an additional assumption which was also used by Andrews (1991). Define

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

Let $H_{1,T}^{(r)}$, $\widehat{\beta}^{(r)}$ and $\beta_0^{(r)}$ denote the r-th elements of $H_{1,T}$, $\widehat{\beta}$ and β_0, respectively, for $r = 1, \ldots, p$.

Assumption 4.4. *For all* $r = 1, \ldots, p$, $\limsup_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathbb{E}_{\mathscr{P}}(H_{1,T}^{(r)}\sqrt{T}(\widehat{\beta}^{(r)} - \beta_0^{(r)}))^2 < \infty$ *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 4.1-4.2 hold, $\sqrt{T}b_{1,T} \to \infty$, *then* $\widehat{J}_T - J_{\mathscr{P},T} \xrightarrow{\mathbb{P}} 0$ *and* $\widehat{J}_T - \widetilde{J}_T \xrightarrow{\mathbb{P}} 0$ *where* $J_{\mathscr{P},T}$ *is defined in* (4.1).

(ii) If Assumption 4.1-4.3 hold, $n_T/(Tb_{1,T}) \to 0$, $n_T/(Tb_{1,T}^q) \to 0$ *and* $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0, \infty)$ *for some* $q \in [0, \infty)$ *for which* $K_{1,q}$, $|\int_0^1 f_{\mathscr{P}_w,a}^{(q)}(u, 0)\, du| \in [0, \infty)$, $w = U, L$, $a \in \mathbb{R}^p$, *then* $\sqrt{Tb_{1,T}b_{2,T}}(\widehat{J}_T - J_{\mathscr{P},T}) = O_{\mathscr{P}}(1)$ *and* $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathscr{P}}(1)$.

(iii) Under the assumptions of part (ii) and Assumption 4.4,

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

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

Theorem 4.2 extends the consistency, rate of convergence, MSE results of Theorem 3.2 in Casini (2023b). The asymptotic equivalence of the MSE implies that the bounds in Theorem 4.1 apply to $\widehat{J}_T$ as well. 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 depend 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 (2023b). The results for the global MSE criterion can easily be 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 $b_{2,T}^{\text{opt}}(u)$. 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 $b_{2,T}^{\text{opt}}(u)$ 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 $K_1^{\mathrm{opt}}(\cdot)$ and $K_2^{\mathrm{opt}}(\cdot)$ from Section 3.

Let $\widetilde{\mathbf{C}}_{\mathscr{P}_U}$ denote the set of continuity points $u \in (0, 1)$ under $\mathscr{P}_U$. For any $a \in \mathbb{R}^p$ and $u_0 \in \widetilde{\mathbf{C}}_{\mathscr{P}_U}$ consider the following inequality,

$$\left| a' \left(\frac{\partial^2}{\partial^2 u} c_{\mathscr{P}}(u_0, k) \right) a \right| \leq \left| a' \left(\frac{\partial^2}{\partial^2 u} c_{\mathscr{P}_U}(u_0, k) \right) a \right|, \tag{4.2}$$

which essentially requires that the distribution $\mathscr{P}_U$ has locally a larger degree of nonstationarity than that of the distribution $\mathscr{P}$. We consider the following class of distributions,

$$\boldsymbol{P}_{U,2} \triangleq \{\mathscr{P} : \mathscr{P} \in \boldsymbol{P}_U,\, m_0 = 0, \text{ and } (4.2) \text{ holds } \forall k \in \mathbb{R} \text{ and } \forall u_0 \in (0, 1)\}.$$

Let

$$D_{1,U,a}(u_0) \triangleq \left(a' \left(\frac{\partial^2 c_{\mathscr{P}_U}(u_0, k)}{\partial u^2} \right) a \right)^2,$$

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

Proposition 4.1. *Suppose Assumption 4.1-4.4 hold and $u_0 \in \widetilde{\mathbf{C}}_{\mathscr{P}_U}$. For any sequence of bandwidth parameters $\{b_{2,T}\}$ such that $b_{2,T} \to 0$, we have*

$$\begin{aligned}
\sup_{\mathscr{P} \in \boldsymbol{P}_{U,2}} \mathrm{MSE}_{\mathscr{P}}\left(a' \widehat{c}_T(u_0, k)\, a\right) &= \sup_{\mathscr{P} \in \boldsymbol{P}_{U,2}} \mathbb{E}_{\mathscr{P}} \left(a' \widehat{c}_T(u_0, k)\, a - a' c_{\mathscr{P}}(u_0, k)\, a\right)^2 \qquad (4.3)\\
&\leq \frac{1}{4} b_{2,T}^4 \left(\int_0^1 x K_2(x)\, dx \right)^2 \left(\frac{\partial^2}{\partial^2 u} a' c_{\mathscr{P}_U}(u_0, k)\, a \right)^2 \\
&\quad + \frac{1}{T b_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} a' \left(c_{\mathscr{P}_U}(u_0, l) \left[c_{\mathscr{P}_U}(u_0, l) + c_{\mathscr{P}_U}(u_0, l+2k) \right]' \right) a \\
&\quad + \frac{1}{T b_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \kappa_{\mathscr{P}_U, aV, Tu_0}(h_1, 0, h_2) + o\left(b_{2,T}^4\right) + O\left(1/(b_{2,T} T)\right),
\end{aligned}$$

which is minimized for

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

where

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

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

We now obtain the optimal $K_1(\cdot)$ and $b_{1,T}$ as a function of $\overline{b}_{2,T}^{\mathrm{opt}}=\int_0^1 b_{2,T}^{\mathrm{opt}}(u)\,du$ and $K_2^{\mathrm{opt}}(\cdot)$. For some results below, we consider a subset of $\boldsymbol{K}_1$ defined by $\widetilde{\boldsymbol{K}}_1=\{K_1(\cdot)\in\boldsymbol{K}_1|\,\widetilde{K}(\omega)\geq 0\,\forall\,\omega\in\mathbb{R}\}$ where $\widetilde{K}(\omega)=(2\pi)^{-1}\int_{-\infty}^{\infty}K_1(x)\,e^{-ix\omega}dx$. The function $\widetilde{K}(\omega)$ is referred to as the spectral window generator corresponding to the kernel $K_1(\cdot)$. 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 $K_{2,0}(\cdot)\in\boldsymbol{K}_2$, $b_{1,T}$ and $b_{2,T}=\overline{b}_{2,T}^{\mathrm{opt}}+o(T^{-1/5})$ where $\overline{b}_{2,T}^{\mathrm{opt}}=\int_0^1 b_{2,T}^{\mathrm{opt}}(u)\,du$. Let $\widehat{J}_T^{\mathrm{QS}}(b_{1,T})$ denote the estimator based on the QS kernel $K_1^{\mathrm{QS}}(\cdot)$. We then compare two kernels K_1 using comparable bandwidths $b_{1,T}$ which are defined as follows. Given $K_1(\cdot)\in\widetilde{\boldsymbol{K}}_1$, the QS kernel $K_1^{\mathrm{QS}}(\cdot)$, and a bandwidth $\{b_{1,T}\}$ to be used with the QS kernel, define a comparable bandwidth $\{b_{1,T,K_1}\}$ for use with $K_1(\cdot)$ 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

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

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

Theorem 4.3. *Suppose Assumption 4.1-4.4 hold,* $\int_0^1|f_{U,a}^{(2)}(u,\,0)|du<\infty$, *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* $K_1(\cdot)\in\widetilde{\boldsymbol{K}}_1$ *used to construct* $\widehat{J}_T$, *the QS kernel is preferred to* $K_1(\cdot)$ *in the sense that for all* $a\in\mathbb{R}^p$,

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

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

We now consider the asymptotically optimal choice of $b_{1,T}$ for a given kernel $K_1(\cdot)$ for which $K_{1,q} \in (0, \infty)$ for some q, and given K_2^{opt} and $\overline{b}_{2,T}^{\mathrm{opt}}$. 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 $\phi(q)$ that is a function of $\mathscr{P}_U$ and 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

$$\phi(q) = \sum_{r=1}^{p} w_r \left(\int_0^1 f_{U,a^{(r)}}^{(q)}(u,\, 0)\, du \right)^2 / \sum_{r=1}^{p} w_r \left(\int_0^1 f_{U,a^{(r)}}(u,\, 0)\, du \right)^2, \tag{4.4}$$

where $a^{(r)}$ is a p-vector with the r-th element one and all other elements zero. For any given $\phi(q) \in (0, \infty)$, let $\boldsymbol{P}_U(\phi)$ denote some set $\boldsymbol{P}_U$ whose dominating distribution $\mathscr{P}_U$ satisfies (4.4).

Theorem 4.4. *Suppose Assumption 4.1-4.4 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*

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

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

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

We consider the data-dependent bandwidths $\widehat{b}_{1,T}$ and $\widehat{\overline{b}}_{2,T}$ from Casini (2023b) which are defined as $\widehat{b}_{1,T}^{*}$ and $\widehat{\overline{b}}_{2,T}^{*}$, repetitively, with $\widehat{V}_t$ in place of $\widehat{V}_{D,t}^{*}$. We choose a parametric model for $\{a^{(r)\prime}V_t\}$, $r=1,\ldots,p$, where $a^{(r)}$ is a p-vector with the r-th element one and all other elements zero. We use the same locally stationary AR(1) models as in Section 3, i.e.,

$$V_t^{(r)} = a_1^{(r)}(t/T)\,V_{t-1}^{(r)} + u_t^{(r)},$$

with estimated parameters $\widehat{a}_1^{(r)}(\cdot)$ and $\widehat{\sigma}^{(r)}(\cdot)$. Let

$$\widehat{\theta} = \left(\int_0^1 \widehat{a}_1^{(1)}(u)\,du,\ \int_0^1 \left(\widehat{\sigma}^{(1)}(u)\right)^2 du,\ldots,\ \int_0^1 \widehat{a}_1^{(p)}(u)\,du,\ \int_0^1 \left(\widehat{\sigma}^{(p)}(u)\right)^2 du\right)',$$

and $\theta_{\mathscr{P}}^{*}$ denote the probability limit of $\widehat{\theta}$. We only consider distributions $\mathscr{P}$ for which $\theta_{\mathscr{P}}^{*}$ exists. We construct $\widehat{\phi}(q)=\widehat{\phi}_D(q)$ as in Section 2 but using the estimate $\widehat{\theta}$. 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 (4.4) 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

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

where

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

with $b>1+1/q$. The class of distributions $\boldsymbol{P}_{U,3}$ corresponds to the class $P_{1,1}$ used by Andrews (1988). The lower bound $0<\underline{\phi}\leq\phi_{\mathscr{P}}(q)$ in part (i) 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. Part (ii) imposes a condition on the temporal dependence of the distribution $\mathscr{P}_U$ and is similar to Assumption 3.5-(iii). Part (iii) is satisfied by a wide class of SLS processes as shown by Casini (2023b). Part (iv) was also used by Andrews (1988), though the interval $\boldsymbol{S}(q,\,b,\,l)$ is tighter as it takes into account of the time smoothing.

Let

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

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

$$\widehat{D}_{2,a}(u)\triangleq2\sum_{l=-\lfloor T^{4/25}\rfloor}^{\lfloor T^{4/25}\rfloor}a'\widehat{c}_T(u_0,\,l)\,\widehat{c}_T(u_0,\,l)'\,a,$$

where $\widehat{c}_T$ is defined as $\widehat{c}^*_{D,T}$ with $\widehat{V}_t$ and $\widehat{b}_{2,T}$ in place of $\widehat{V}^*_{D,t}$ and $\widehat{b}^*_{2,T}$, respectively,

$$\widehat{c}_T(rn_T/T,\,k)\triangleq\begin{cases}\left(T\widehat{b}_{2,T}\right)^{-1}\sum_{s=k+1}^TK_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}_{2,T}}\right)\widehat{V}_s\widehat{V}'_{s-k}, & k\geq0\\ \left(T\widehat{b}_{2,T}\right)^{-1}\sum_{s=-k+1}^TK_2^*\left(\frac{((r+1)n_T-(s+k/2))/T}{\widehat{b}_{2,T}}\right)\widehat{V}_{s+k}\widehat{V}'_s, & k<0\end{cases}.$$

Assumption 4.5. *(i) We have* $\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left[\frac{\inf\{T/n_{3,T},\,\sqrt{n_{2,T}}\}\left(\widehat{\phi}(q)^{1/(2q+1)}-\phi_{\theta^*_{\mathscr{P}}}^{1/(2q+1)}\right)}{\widehat{\phi}(q)^{1/(2q+1)}}\right]^4=O(1)$ *as* $T\to\infty$*, where q is as defined in* $\boldsymbol{K}_3$*,* $\widehat{\phi}(q)\leq\overline{\phi}<\infty$*, and* $n_{2,T}/T+n_{3,T}/T\to0$*,* $n_{2,T}^{10/6}/T\to[c_2,\,\infty)$*,* $n_{3,T}^{10/6}/T\to[c_3,\,\infty)$ *with* $0<c_2,\,c_3<\infty$*; (ii)* $\sqrt{Tb_{2,T}(u)}(\widehat{D}_{2,a}(u)-D_{2,U,a}(u))=O_{\mathscr{P}}(1)$ *for all* $u\in[0,\,1]$*; (iii) Assumption 3.5-(v) hold.*

Any estimator $\widehat{\phi}$ based on kernel nonparametric estimators of $\widehat{a}_1^{(r)}(\cdot)$ and $\widehat{\sigma}^{(r)}(\cdot)$ satisfies

Assumption 4.5-(i). Assumption 4.5-(ii) extends Assumption 3.5-(vi) to the distribution $\mathscr{P}$ and is are useful to show that the effect of using $\widehat{b}_{1,T}$ and $\widehat{\overline{b}}_{2,T}$ rather than $b_{1,\theta_{\mathscr{P}},T}$ and $\overline{b}_{2,T}^{\text{opt}}$ when constructing $\widehat{J}_T$ is at most $o_{\mathbb{P}}(1)$. 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 the estimators with fixed bandwidth parameters have been determined in Section 4.2, from this result 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 4.1-4.5 hold. Then, for all $a \in \mathbb{R}^p$,*

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

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 $((T/n_{3,T})^2 + n_{2,T})/T^{4/5} \to c \in [0, \infty)$ in Assumption 4.5-(i) we have $\widehat{b}_{1,T} = O_{\mathscr{P}}(1)$. Thus, $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}) - J_{\mathscr{P},T} \xrightarrow{\mathscr{P}} 0$ 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 Theoretical Results About the Power of HAR Tests Under General $\mathbb{E}(V_t)$

A long-lasting problem in time series econometrics is the low/non-monotonic power of HAR inference tests under nonstationary alternative hypotheses. The problem involves HAR tests outside the regression model that can be characterized by an alternative hypothesis involving $\mathbb{E}(V_t) = \mu_t$ with $\mu_t \neq 0$ for at least one t. The process μ_t can be any piecewise continuous function of t. For example, tests for structural breaks, tests for regime switching and tests for time-varying parameters can be framed in this way. To see this, consider a linear regression model,

$$y_t = x_t' \beta_t + e_t, \qquad t = 1, \ldots, T. \tag{5.1}$$

The null hypothesis of no break in the regression coefficient of x_t is written as $H_{\beta,0}:\ \beta_t = \beta_0$ for all t for some $\beta_0 \in \mathbb{R}^p$ [see, e.g., Andrews (1993)]. The alternative hypothesis may be of several forms. Let $H_{\beta,1}:\ \beta_t = \beta(t/T)$ $(t = 1, \ldots, T)$ for some piecewise continuous function $\beta(\cdot)$. Estimating (5.1) by least-squares yields $y_t = x_t'\widehat{\beta} + \widehat{e}_t$ for all t where $\widehat{\beta}$ is the least-squares estimate and $\{\widehat{e}_t\}$ are the least-squares residuals. Letting $V_t = x_t\widehat{e}_t$, the null hypothesis $H_{\beta,0}$ can be rewritten as $H_0 : \mathbb{E}(V_t) = 0$ for all t while the alternative hypothesis $H_{\beta,1}$ can be rewritten as $H_1 : \mathbb{E}(V_t) = \mu_t$ where $\mu_t \neq 0$ for at least one t. Structural break tests are based on an estimate of the LRV of $V_t = x_t\widehat{e}_t$. While under H_0 V_t is zero-mean, under H_1 the mean of V_t is time-varying. Under the alternative hypothesis, it is required for consistency of the test that the LRV estimator converges to some positive semidefinite matrix since the numerator of the test statistics diverges to infinity. However, time-variation in the mean of V_t severely biases upward traditional LRV estimators which then lead to tests with non-monotonic power. Casini et al. (2023) established analytical results for this phenomenon, which they referred to as low frequency contamination. We show that the proposed nonlinear prewhitened DK-HAC estimator accounts for nonstationarity also under the alternative hypothesis and leads to consistent tests with good monotonic power. Although the main theoretical result of this section is presented for a particular HAR test and a particular form of H_1, this result is general enough to provide guidance for most cases discussed in the literature.

We present theoretical results about the power of a popular forecast evaluation test, namely the Diebold-Mariano test [cf. Diebold and Mariano (1995)], which can be also framed as above. We focus on the Diebold-Mariano test for ease of the exposition. Similar results hold for the other HAR inference tests that can be framed as above, though the proofs change slightly depending on the specific test statistic. Suppose the goal is to forecast some variable y_t. Two forecast models are used: $y_t = \beta^{(1)} + \beta^{(2)} x_{t-1}^{(i)} + e_t$ where $x_{t-1}^{(i)}$ is some predictor and $i = 1, 2$. That is, each forecast model uses an intercept and a predictor. The parameters $\beta^{(1)}$ and $\beta^{(2)}$ are estimated using least-squares in the in-sample $t = 1, \ldots, T_m$ with a fixed forecasting scheme. Each forecast model generates a sequence of $\tau\,(=1)$-step ahead out-of-sample losses $L_t^{(i)}$ $(i = 1, 2)$ for $t = T_m + 1, \ldots, T - \tau$. Then $d_t \triangleq L_t^{(2)} - L_t^{(1)}$ denotes the loss differential at time t. Let $\overline{d}_L$ denote the average of the loss differentials. 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 $\widehat{J}_{d_L,T}$ 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. The true model is $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1}^{(0)} + e_t$ where $x_{t-1}^{(0)}$ is a predictor and e_t is a zero-mean error. We assume that the conditions for consistency and asymptotic normality of the least-squares estimates of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ are satisfied.

In this setting, $\widehat{V}_t = d_t$. The hypothesis testing problem is given by

$$H_0 : \mathbb{E}\left(\widehat{V}_t\right) = 0, \quad \text{for all } t, \qquad (5.2)$$
$$H_1 : \mathbb{E}\left(\widehat{V}_t\right) = \mu_t, \quad \text{with } \mu_t \neq 0 \text{ for at least one } t.$$

H_0 corresponds to equal predictive ability between the two forecast models while H_1 corresponds to the two forecast models performing differently.

Since we want to study the power of t_{DM}, we need to work under the alternative hypothesis. The two competing forecast models are as follows: the first model uses the actual true predictor (i.e., $x_{t-1}^{(1)} = x_{t-1}^{(0)}$ for all t) 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 $t \leq T_b$ and $x_{t-1}^{(2)} = \delta + x_{t-1}^{(0)} + u_{X_2,t}$ for $t > T_b$ with $T_b > T_m$, and $u_{X_2,t}$ is a zero-mean error term. Evidently, the null hypotheses of equal predictive ability should be rejected whenever $\delta > 0$. We consider t_{DM} normalized by different LRV estimators. The HAC estimator is defined as

$$\widehat{J}_{d_L,\mathrm{HAC},T} \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_T k\right) \widehat{\Gamma}\left(k\right), \qquad \widehat{\Gamma}\left(k\right) = T^{-1} \sum_{t=|k|+1}^{T} \widehat{V}_t \widehat{V}_{t-|k|},$$

where $K_1(\cdot)$ is a kernel (e.g., the Bartlett and QS) and b_T a bandwidth. Kiefer et al. (2000) proposed to use a LRV estimator that keeps b_T at a fixed fraction of T, i.e., $\widehat{J}_{\mathrm{KVB},T} \triangleq T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(1 - |t-s|/T\right) \widehat{V}_t \widehat{V}_s$ which is equivalent to the Newey-West estimator with $b_T = T^{-1}$.

We present theoretical results about the power of t_{DM}. Let $t_{\mathrm{DM},i} = T_n^{1/2} \overline{d}_L / \sqrt{\widehat{J}_{d_L,i,T}}$ denote the DM test statistic where i = DK, pwDK, KVB, EWC, A91, pwA91, NW87 and pwNW87. $\widehat{J}_{d_L,\mathrm{A91},T}$ and $\widehat{J}_{d_L,\mathrm{NW87},T}$ are $\widehat{J}_{d_L,\mathrm{HAC},T}$ where $K_1(\cdot)$ is the Bartlett and QS kernel, respectively.[14] $\widehat{J}_{d_L,\mathrm{pwA91},T}$ and $\widehat{J}_{d_L,\mathrm{pwNW87},T}$ are the prewhitened HAC estimators using the QS and Bartlett kernel, respectively, and the prewhitening procedure of Andrews and Monahan (1992). "DK" refers to the DK-HAC estimator from Casini (2023b) with the MSE-optimal kernels and bandwidths whereas "pwDK" refers to the prewhitened DK-HAC estimator $\widehat{J}_{\mathrm{pw},T}$ in (2.3). Define the power of $t_{\mathrm{DM},i}$ as $\mathbb{P}_\delta(|t_{\mathrm{DM},i}| > z_\alpha)$ where z_α is the two-sided standard normal critical value and $\alpha \in (0, 1)$ is the significance level. To avoid repetitions we present the results only for i = DK, pwDK, KVB, NW87 and pwNW87. The results concerning the EWC estimator are the same as those for the KVB's fixed-b estimator. The results pertaining to Andrews' (1991) HAC estimator (with and without prewhitening) are the same as those corresponding to Newey and West's (1987) estimator (with and without prewhitening, respectively). For the HAC and DK-HAC estimators we report the results

[14] Since $\{\widehat{V}_t\}$ is only observed in the out-of-sample, the LRV estimators use a sample of T_n observations.

for the MSE-optimal bandwidth [see Andrews (1991), Casini (2023b) and Whilelm (2015)].[15] We set $n_T = n_{2,T} = n_{3,T} = T^{2/3}$ which satisfy the growth rate bounds [see Casini (2023b) for details]. Let $n_\delta = T - T_b - 2$ denote the length of the regime in which $x_t^{(2)}$ exhibits a shift δ in the mean. The alternative hypothesis depends on the shift magnitude δ and on how long the shift lasts for. Here the latter is n_δ. More generally, this is the set of time points such that $\mathbb{E}(\widehat{V}_t) = \mu_t \neq 0$ holds.

Theorem 5.1. *Let $\{d_t - \mathbb{E}(d_t)\}$ be a SLS process satisfying Assumption 2.1 and 3.1, and $n_\delta = O(T_n^{1/2+\zeta})$ where $\zeta \in (0, 1/2)$ such that $T_n^\zeta b_T^{1/2} \to 0$ and $T_n^\zeta (\widehat{b}_{1,T}^*)^{1/2} \to 0$. Then, we have:*

(i) If $b_T \to 0$, then $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}| > z_\alpha) \to 0$. If $b_T = O(T^{-1/3})$, then $|t_{\mathrm{DM,NW87}}| = O_{\mathbb{P}}(T_n^{\zeta-1/6})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}| > z_\alpha) \to 0$.

(ii) If $b_T \to 0$, then $\mathbb{P}_\delta(|t_{\mathrm{DM,pwNW87}}| > z_\alpha) \to 0$. If $b_T = O(T^{-1/3})$, then $|t_{\mathrm{DM,pwNW87}}| = O_{\mathbb{P}}(T_n^{\zeta-1/6})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,pwNW87}}| > z_\alpha) \to 0$.

(iii) If $b_T = T^{-1}$, then $|t_{\mathrm{DM,KVB}}| = O_{\mathbb{P}}(T_n^{\zeta-1/2})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,KVB}}| > z_\alpha) \to 0$.

(iv) Under Assumption 3.2-(i-iii), $|t_{\mathrm{DM,DK}}| = \delta^2 O_{\mathbb{P}}(T_n^\zeta)$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,DK}}| > z_\alpha) \to 1$.

(v) Under Assumption 3.2-(i-iii), 3.5-(i,iv) and 3.6, $|t_{\mathrm{DM,pwDK}}| = \delta^2 O_{\mathbb{P}}(T_n^\zeta)$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,pwDK}}| > z_\alpha) \to 1$.

Note that $b_T = O(T^{-1/3})$ in parts (i)-(ii) refers to the MSE-optimal bandwidth for the Newey and West's (1987) estimator. The conditions $T_n^\zeta b_T^{1/2} \to 0$ and $T_n^\zeta (\widehat{b}_{1,T}^*)^{1/2} \to 0$ mean that the length of the regime in which $x_t^{(2)}$ exhibits a shift δ in the mean increases to infinity at a slower rate than T. Theorem 5.1 implies that when the (prewhitened or non-prewhitened) HAC estimators or the fixed-b LRV estimators are used, the DM test is not consistent and its power converges to zero. The theorem suggests that prewhitened and non-prewhitened HAC estimators suffer from this problem in a similar way. The theorem also implies that the power functions corresponding to tests based on HAC estimators lie above the power functions corresponding to those based on fixed-b/EWC LRV estimators. An additional feature is that $|t_{\mathrm{DM,NW87}}|$, $|t_{\mathrm{DM,pwNW87}}|$ and $|t_{\mathrm{DM,KVB}}|$ do not increase in magnitude with δ because δ appears in both the numerator and denominator. The results concerning the DK-HAC estimator and the prewhitened DK-HAC estimator $\widehat{J}_{\mathrm{pw},T}$ show that these issues do not occur when these estimators are used. In fact, the test is consistent and its power increases with δ and with the sample size. We provide finite-sample evidence in support of these theoretical results in Section 6.

[15] For the HAC estimators we also report the result for any bandwidth choice $b_T \to 0$ such that $Tb_T \to \infty$, which is sufficient for the consistency of the estimator.

6 Small-Sample Evaluations

We now show that the prewhitened DK-HAC estimators lead to HAR inference tests that have accurate null rejection rates when there is strong dependence and have superior power properties relative to those based on 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 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 estimator of the limit of $\mathrm{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ 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,

$$y_t = \beta_0^{(1)} + \delta + \beta_0^{(2)} x_t + e_t, \qquad t = 1, \ldots, T, \tag{6.1}$$

for the t-test on the intercept and

$$y_t = \beta_0^{(1)} + (\beta_0^{(2)} + \delta) x_t + e_t, \qquad t = 1, \ldots, T, \tag{6.2}$$

for the t-test on $\beta_0^{(2)}$ where $\delta = 0$ under the null hypotheses. In model M1 we set $\beta_0^{(1)} = 0$, $\beta_0^{(2)} = 1$, $x_t \sim$ i.i.d. $\mathscr{N}(1, 1)$ and $e_t = \rho e_{t-1} + u_t$, $\rho = 0.4, 0.9$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.7)$. Model M2 involves segmented locally stationary errors: $\beta_0^{(1)} = \beta_0^{(2)} = 0$, $x_t = 0.6 + 0.8 x_{t-1} + u_{x,t}$, $u_{x,t} \sim$ i.i.d. $\mathscr{N}(0, 1)$ 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 $X = [X_1, \ldots, X_T]'$ and $X_t = [1, x_t]'$.

Next, we move to the forecast evaluation tests. The Diebold-Mariano test statistic is defined as in Section 5, $t_{\mathrm{DM}} \triangleq T_n^{1/2} \overline{d}_L / \widehat{J}_{d_L,T}^{1/2}$. In model M3 we consider an out-of-sample forecasting exercise with a fixed 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, we specify the following data-generating process and the two forecasting models that have equal predictive ability. The true model for y_t 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 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 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 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. $\mathscr{N}(0,\,1)$. 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 the t-test for forecast breakdown of Giacomini and Rossi (2009). Suppose we want to forecast a variable y_t which follows $y_t=\beta_0^{(1)}+\beta_0^{(2)}x_{t-1}+\delta x_{t-1}\mathbf{1}\{t>T_1^0\}+e_t$ where $x_t\sim$ i.i.d. $\mathscr{N}(1.5,\,1.5)$ and $e_t=0.3e_{t-1}+u_t$ 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 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, $SL_{t+\tau}=L_{t+\tau}-\overline{L}_{t+\tau}$. Here 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 a LRV 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 (2.2)] (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 (2.2)]. 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' (1991) HAC estimator with automatic bandwidth; Andrews' (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 et al. (2000); the Empirical Weighted Cosine (EWC) of Lazarus et al. (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 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 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 are 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 oversized tests with $T = 200$.

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 tests. 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 evidence and in part suggest new facts. 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 null rejection rates 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 leading to tests that control the null rejection rates 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 null rejection rates. Tests based on the Newey-West with fixed-b (KVB) sacrifices power more than using the DK-HAC estimators 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 previously 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 and 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 estimators. The latter have excellent power. The reason for the severe power problems for 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. The theoretical results about the power in Theorem 5.1 suggest that 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 estimators allows to avoid this problem because it flexibly accounts for nonstationarity. The key idea is not to mix observations belonging to different regimes. Simulation results for additional data-generating processes involving ARMA, ARCH and heteroskedastic errors are not discussed here because the results are qualitatively equivalent.

7 Conclusions

We introduce a nonparametric nonlinear VAR prewhitened long-run variance (LRV) estimator for the construction of standard errors robust to autocorrelation and heteroskedasticity that can be used for hypothesis testing both within and outside the linear regression model. HAR tests normalized by the proposed estimator exhibit accurate null rejection rates even when there is strong dependence. We show theoretically that existing estimators lead to HAR tests that have low/non-monotonic power under nonstationary alternative hypotheses while the proposed estimator has good monotonic power thereby addressing a long-lasting problem in time series econometrics. The proposed method is theoretically valid under general nonstationary random variables. We also establish mean-squared error bounds for LRV estimation that are sharper than previously established and use them to determine the data-dependent bandwidths.

Supplemental Materials

The supplement for online publication [cf. Casini and Perron (2021b)] presents the proofs of the results in the paper.

References

Altissimo, F., Corradi, V., 2003. Strong rules for detecting the number of breaks in a time series. Journal of Econometrics 117, 207–244.

Andrews, D.W.K., 1988. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Cowles Foundation Discussion Paper No. 877, Yale University .

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

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change-point. Econometrica 61, 821–56.

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

Belotti, F., Casini, A., Catania, L., Grassi, S., Perron, P., 2023. Simultaneous bandwidths determination for double-kernel HAC estimators and long-run variance estimation in nonparametric settings. Econometric Reviews 42, 281–306.

Brillinger, D., 1975. Time Series Data Analysis and Theory. New York: Holt, Rinehart and Winston.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society. Series B (Methodological) 37, 149–192.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Casini, A., 2018. Tests for forecast instability and forecast failure under a continuous record asymptotic framework. arXiv preprint arXiv:1803.10883.

Casini, A., 2022. Comment on Andrews (1991) "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". Econometrica 90, 1–2.

Casini, A., 2023a. The Fixed-b limiting distribution and the ERP of HAR tests under nonstationarity. arXiv preprint arXiv 2111.14590.

Casini, A., 2023b. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Journal of Econometrics, 235, 372–392.

Casini, A., Deng, T., Perron, P., 2023. Theory of low frequency contamination from nonstationarity and misspecification: consequences for HAR inference. arXiv preprint arXiv:2103.01604.

Casini, A., Perron, P., 2019. Structural breaks in time series. Oxford Research Encyclopedia of

Economics and Finance, Oxford University Press.

Casini, A., Perron, P., 2020. Generalized Laplace inference in multiple change-points models. Econometric Theory 38, 35–65.

Casini, A., Perron, P., 2021a. Continuous record Laplace-based inference about the break date in structural change models. Juornal of Econometrics 224, 3–21.

Casini, A., Perron, P., 2021b. Supplement to "prewhitened long-run variance estimation robust to nonstationarity". arXiv preprint arXiv:2103.02235.

Casini, A., Perron, P., 2023. Change-point analysis of time series with evolutionary spectra. arXiv preprint arXiv 2106.02031.

Cavaliere, G., Taylor, A.M.R., 2007. Testing for unit roots in time series models with non-stationary volatility. Journal of Econometrics 140, 919–947.

Chan, K.W., 2022a. Optimal difference-based variance estimators in time series: a general framework. Annals of Statistics 50, 1376–1400.

Chan, K.W., 2022b. Mean-structure and autocorrelation consistent covariance matrix estimation. Journal of Business and Economic Statistics, 40, 201–215.

Chang, S.Y., Perron, P., 2018. A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Econometric Reviews 37, 577–601.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80, 1157–1183.

Crainiceanu, C.M., Vogelsang, T.J., 2007. Nonmonotonic power for tests of a mean shift in a time series. Journal of Statistical Computation and Simulation 77, 457–476.

Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Annals of Statistics 25, 1–37.

Dahlhaus, R., Giraitis, L., 1998. On the optimal segment length for parameter estimates for locally stationary time series. Journal of Time Series Analysis 19, 629–655.

Demetrescu, M., Salish, N., 2020. (Structural) VAR models with ignored changes in mean and volatility. Unpublished Manuscript, SSRN https://ssrn.com/abstract=3544676.

Deng, A., Perron, P., 2006. A comparison of alternative asymptotic frameworks to analyse a structural change in a linear time trend. Econometrics Journal 9, 423–447.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–63.

Epanechnikov, V., 1969. Non-parametric estimation of a multivariate probability density. Theory of Probability and its Applications 14, 153–158.

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. Review of Economic Studies 76, 669–705.

Giacomini, R., Rossi, B., 2010. Forecast comparisons in unstable environments. Journal of Applied Econometrics 25, 595–620.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. New York: Wiley.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hansen, B., 1992. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica 60, 967–972.

Jansson, M., 2004. The error in rejection probability of simple autocorrelation robust tests. Econometrica 72, 937–946.

de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. Econometrica 68, 407–423.

Juhl, T., Xiao, Z., 2009. Testing for changing mean with monotonic power. Journal of Econometrics 148, 14–24.

Kawka, R., 2020. Convergence of spectral density estimators in the locally stationary framework. Econometrics and Statistics, forthcoming.

Kiefer, N.M., Vogelsang, T.J., 2002. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica 70, 2093–2095.

Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.

Kiefer, N.M., Vogelsang, T.J., Bunzel, H., 2000. Simple robust testing of regression hypotheses. Econometrica 69, 695–714.

Kim, D., Perron, P., 2009. Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope. Journal of Econometrics 149, 26–51.

Lazarus, E., Lewis, D.J., Stock, J.H., 2021. The size-power tradeoff in HAR inference. Econometrica 89, 2497–2516.

Lazarus, E., Lewis, D.J., Stock, J.H., Watson, M.W., 2018. HAR inference: recommendations for practice. Journal of Business and Economic Statistics 36, 541–559.

Martins, L., Perron, P., 2016. Improved tests for forecast comparisons in the presence of instabilities. Journal of Time Series Analysis 37, 650–659.

Newey, W.K., West, K.D., 1987. A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

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

Otto, S., Breitung, J., 2021. Backward CUSUM for testing and monitoring structural change. Econometric Theory, forthcoming.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.

Perron, P., Yamamoto, Y., 2021. Testing for changes in forecast performance. Journal of Business and Economic Statistics 39, 148–165.

Ploberger, W., Krämer, W., 1992. The CUSUM test with OLS residuals. Econometrica 60, 271–285.

Potiron, Y., Mykland, P., 2020. Local parametric estimation in high frequency data. Journal of Business and Economic Statistics 38, 679–692.

Preinerstorfer, D., 2017. Finite sample properties of tests based on prewhitened nonparametric covariance estimators. Electronic Journal of Statistics 11, 2097–2167.

Qu, Z., Zhuo, F., 2020. Likelihood ratio based tests for Markov regime switching. Review of Economic Studies 88, 937–968.

Rho, Y., Shao, X., 2013. Improving the bandwidth-free inference methods by prewhitening. Journal of Statistical Planning and Inference 143, 1912–1922.

Shao, X., Zhang, X., 2010. Testing for change points in time series. Journal of the American Statistical Association 105, 122–1240.

Sun, Y., 2014. Let's fix it: fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. Journal of Econometrics 178, 659–677.

Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. Econometrica 76, 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, 283–299.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

Whilelm, D., 2015. Optimal bandwidth selection for robust generalized methods of moments estimation. Econometric Theory 31, 1054–1077.

Xiao, Z., Linton, O., 2002. A nonparametric prewhitened covariance estimator. Journal of Time Series Analysis 23, 215–250.

Zhang, T., Lavitas, L., 2018. Unsupervised self-normalized change-point testing for time series. Journal of the American Statistical Association 113, 637–648.

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

Prewhitened Long-Run Variance Estimation Robust to Nonstationarity

Alessandro Casini — University of Rome Tor Vergata
Pierre Perron — Boston University

27th September 2023

Abstract

This supplemental material is for online publication and is structured as follows. Section S.A presents some preliminary notions. Section S.B, S.C and S.D present the proofs of the results of Section 3, 4 and 5, respectively.

S.A Preliminaries

In this section we present a formal definition of SLS processes which is implied by Assumption 2.1 on $f(u, \omega)$. Let $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_{m_0} < \lambda_{m_0+1} = 1$ where m_0 may be fixed or grow to infinity. A function $G(u, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C}$ is said to be piecewise (Lipschitz) continuous in u with $m_0 + 1$ segments if for each segment $j = 1, \ldots, m_0 + 1$ 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 S.A.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,

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

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 $A_j : (\lambda^0_{j-1}, \lambda^0_j] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$, $\lambda^0_j \triangleq T^0_j / T$ and for all T,

$$A(u, \omega) = A_j(u, \omega) \text{ for } \lambda^0_{j-1} < u \leq \lambda^0_j, \tag{S.2}$$

$$\sup_{1 \leq j \leq m_0+1} \sup_{T^0_{j-1} < t \leq T^0_j, \omega} \left| A^0_{j,t,T}(\omega) - A_j(t/T, \omega) \right| \leq K T^{-1}. \tag{S.3}$$

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

In the context of HAR inference $\mathbb{E}(V_t) = 0$ and so $\mu(t/T) = 0$ for all t in Definition S.A.1. In view of Definition S.A.1, Assumption 2.1 also holds with $f(u, \omega)$ replaced by $A(u, \omega)$ and this property is used in some parts of the proofs. In Assumption 3.1-(ii), the continuity points are those $u \in [0, 1]$ such that $u \neq \lambda^0_j$ $(j = 1, \ldots, m_0 + 1)$ whereas the discontinuity points are those $u \in [0, 1]$ such that $u = \lambda^0_j$ $(j = 1, \ldots, m_0 + 1)$.

S.B Proofs of the Results in Section 3

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

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

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.B.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 3.1, $\|\int_0^1 f^{*(0)}(u, 0)\| < \infty$, 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 (2023b) [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 2.1 and 3.1 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),

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

where we have used the boundedness of the kernel K_2 (and thus of K_2^*), Assumption 3.2-(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 3.2-(ii,iii) by a mean-value expansion and,

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

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

Lemma S.B.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,

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

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,

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

We first show that $A_{1,T}\xrightarrow{\mathbb{P}}0$. 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 $K_1(\cdot)\in\boldsymbol{K}_3$ to yield,

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

for some $C<\infty$. By Assumption 3.5-(i),

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

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

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

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

$$\begin{aligned} B_{1,T} &\leq \left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)} n_T^{-1/2} \sum_{k=1}^{S_T} k \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}} \right) \sqrt{n_T} \left(\widehat{\beta} - \beta_0\right) \right| \\ &\leq C \left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)} \left(Tb_{\theta_{2,T}}\right)^{2r/(2q+1)} n_T^{-1/2} \sup_{k \geq 1} \left\| \frac{\partial}{\partial \beta} \widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}} \right\| \sqrt{n_T} \left\| \widehat{\beta} - \beta_0 \right\|, \end{aligned} \tag{S.12}$$

since $r < (10q+17)/24$, and $\sup_{k\geq 1} ||(\partial/\partial\beta)\widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}}|| = O_{\mathbb{P}}(1)$ using (S.5) and Assumption 3.2-(ii,iii) (the latter continues to hold for $\{\overline{V}_t^*\}$). In addition,

$$\begin{aligned} \mathbb{E}\left(B_{2,T}^2\right) &\leq \mathbb{E}\left(\left(T\widehat{b}_{2,T}^*\right)^{-2/(2q+1)} \sum_{k=1}^{S_T}\sum_{j=1}^{S_T} kj \left|\widetilde{\Gamma}^*(k) - \Gamma_T^*(k)\right| \left|\widetilde{\Gamma}^*(j) - \Gamma_T^*(j)\right| \right) \\ &\leq \left(T\widehat{b}_{2,T}^*\right)^{-2/(2q+1)-1} S_T^4 \sup_{k\geq 1} T\widehat{b}_{2,T}^* \mathrm{Var}\left(\widetilde{\Gamma}^*(k)\right) \\ &\leq \left(T\widehat{b}_{2,T}^*\right)^{-2/(2q+1)-1} \left(T\overline{b}_{\theta_{2,T}}\right)^{4r/(2q+1)} \sup_{k\geq 1} T\widehat{b}_{2,T}^* \mathrm{Var}\left(\widetilde{\Gamma}^*(k)\right) \\ &\leq \left(\widehat{b}_{2,T}^*\right)^{-2/(2q+1)-1} T^{-1-2/(2q+1)} T^{16r/5(2q+1)} \sup_{k\geq 1} T\widehat{b}_{2,T}^* \mathrm{Var}\left(\widetilde{\Gamma}^*(k)\right) \to 0, \end{aligned} \tag{S.13}$$

given that $r < (3+2q)/4$ and $\sup_{k\geq 1} T\widehat{b}_{2,T}^* \mathrm{Var}(\widetilde{\Gamma}^*(k)) = O(1)$ by Lemma S.A.5 in Casini (2023b) that also holds with $\widetilde{\Gamma}^*(k)$ in place of $\widetilde{\Gamma}(k)$. Next,

$$\begin{aligned} B_{3,T} &\leq \left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)} S_T \sum_{k=1}^{\infty} \left|\Gamma_T^*(k)\right| \\ &\leq \left(T\widehat{b}_{2,T}^*\right)^{(r-1)/(2q+1)} O_{\mathbb{P}}(1) \to 0, \end{aligned} \tag{S.14}$$

using Assumption 3.1-(i) since $r < 1$. This gives $A_{1,T} \overset{\mathbb{P}}{\to} 0$. Next, we show that $A_{2,T} \overset{\mathbb{P}}{\to} 0$. Let $A_{2,1,T} = L_{1,T} + L_{2,T} + L_{3,T}$, where

$$\begin{aligned} L_{1,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}^* k\right)\left(\widehat{\Gamma}^*(k) - \widetilde{\Gamma}^*(k)\right), \\ L_{2,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}^* k\right)\left(\widetilde{\Gamma}^*(k) - \Gamma_T^*(k)\right), \quad \text{and} \\ L_{3,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}^* k\right)\Gamma_T^*(k). \end{aligned} \tag{S.15}$$

We apply a mean-value expansion and use $\sqrt{n_T}(\widehat{\beta} - \beta_0) = O_{\mathbb{P}}(1)$ as well as (S.5) to obtain

$$|L_{1,T}| = n_T^{-1/2} \sum_{k=S_T+1}^{T-1} C_1 \left(\widehat{b}_{1,T}^* k\right)^{-b} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}^*(k)\right)|_{\beta=\overline{\beta}} \sqrt{n_T}\left(\widehat{\beta} - \beta_0\right) \right| \tag{S.16}$$

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

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

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

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 3.5-(i). Let us turn to $L_{3,T}$. We have,

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

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 $(\widehat{J}_T(b_{\theta_1,T}, \widehat{b}^*_{2,T}) - \widehat{J}_T(b_{\theta_1,T}, \overline{b}_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. Its proof is the same as in Theorem 5.1-(i) in Casini (2023b) 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.B.3. *Under the assumptions of Theorem 3.1-(ii), we have*

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

Proof. Write

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

Applying Theorem 3.1-(ii) in Casini (2023b) 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

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

Using Assumption 3.3-(ii),

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

and thus,

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

since $Tb_{\theta_2,T}/(n_T^2b_{\theta_1,T})\rightarrow0$. Next, we want to show that $G_T=o_{\mathbb{P}}(1)$. Following Andrews (1991) (cf. the last paragraph of p. 852), we apply the results of Theorem 3.1-(i,ii) in Casini (2023b) to $\widetilde{J}_T^*$ where the latter is constructed using $(V_t^{*\prime},\,\partial V_t^*/\partial\beta'-\mathbb{E}(\partial V_t^*/\partial\beta'))'$ 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 (2023b) each expression above is $O_{\mathbb{P}}\left(1\right)$. Given,

$$G_{T}\leq\frac{\sqrt{Tb_{\theta_{2},T}}}{\sqrt{n_{T}}}\sqrt{b_{\theta_{1},T}}\left(A_{1}+A_{2}\right)+\frac{\sqrt{Tb_{\theta_{2},T}}}{\sqrt{n_{T}}}\sqrt{b_{\theta_{1},T}}\sum_{k=-T+1}^{T-1}K_{1}\left(b_{\theta_{1},T}k\right)\frac{n_{T}}{T}\sum_{r=0}^{T/n_{T}}\frac{1}{Tb_{\theta_{2},T}}$$

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

$$\triangleq\frac{\sqrt{Tb_{\theta_{2},T}}}{\sqrt{n_{T}}}\sqrt{b_{\theta_{1},T}}\left(A_{1}+A_{2}\right)+A_{3}\sup_{s}\left|\mathbb{E}\left(\frac{\partial}{\partial\beta}V_{s}^{*}\right)\right|,$$

and the fact that $Tb_{\theta_{2},T}b_{\theta_{1},T}/n_{T}\rightarrow0$ it remains to show that A_{3} is $o_{\mathbb{P}}\left(1\right)$. Note that

$$\mathbb{E}\left(A_{3}^{2}\right)\leq\frac{Tb_{\theta_{2},T}}{n_{T}}b_{1,T}\sum_{k=-T+1}^{T-1}\sum_{j=-T+1}^{T-1}\left|K_{1}\left(b_{\theta_{1},T}k\right)K_{1}\left(b_{\theta_{1},T}j\right)\right|4\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}$$

$$\times\frac{1}{Tb_{\theta_{2},T}}\frac{1}{Tb_{\theta_{2},T}}\sum_{s=1}^{T}\sum_{l=1}^{T}K_{2}^{*}\left(\frac{\left(\left(r+1\right)n_{T}-\left(s+k/2\right)\right)/T}{b_{\theta_{2},T}}\right)$$

$$\times K_{2}^{*}\left(\frac{\left(\left(b+1\right)n_{T}-\left(l+j/2\right)\right)/T}{b_{\theta_{2},T}}\right)\left|\mathbb{E}\left(V_{s}^{*}V_{l}^{*}\right)\right|,$$

and that $\mathbb{E}\left(V_{s}^{*}V_{l}^{*}\right)=c^{*}\left(u,\,h\right)+O\left(T^{-1}\right)$ uniformly in $h=s-l$ and $u=s/T$ by Lemma S.A.1 in Casini (2023b). Since $\sum_{h=-\infty}^{\infty}\sup_{u\in\left[0,\,1\right]}\left|c^{*}\left(u,\,h\right)\right|<\infty$,

$$\mathbb{E}\left(A_{3}^{2}\right)\leq\frac{1}{n_{T}b_{\theta_{1},T}}\left(b_{\theta_{1},T}\sum_{k=-T+1}^{T-1}\left|K_{1}\left(b_{\theta_{1},T}k\right)\right|\right)^{2}\int_{0}^{1}K_{2}^{2}\left(x\right)dx\int_{0}^{1}\sum_{h=-\infty}^{\infty}\left|c^{*}\left(u,\,h\right)\right|du=o\left(1\right).$$

This implies $G_{T}=o_{\mathbb{P}}\left(1\right)$ which concludes the proof. □

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

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

Proof. Let

$$r\in(\max\{\{\{\left(-10+4q+24b\right)/24\left(b-1\right)\},\,\{\left(b-1/2\right)/\left(b-1\right)\}\,\text{for}\,b>\max\left(1+1/q,\,4\right)\},$$

$$\{\{(8b-4)\,/\,(b-1)\,(10q+5)\},\ \{(b-2/3-q/3)\,/\,(b-1)\}\ \text{for}\ b>1+1/q\},\ q/\,(l-1)\},$$
$$\min\{16q/48+44/48,\ 46/48+20q/48,\ 2/3+q/3\}),$$

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

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

Let

$$N_1\triangleq\{-S_T,\,-S_T+1,\ldots,\,-1,\,1,\ldots,\,S_T-1,\,S_T\}$$
$$N_2\triangleq\{-T+1,\ldots,\,-S_T-1,\,S_T+1,\ldots,\,T-1\}\,.$$

Let us consider the first term above,

$$\begin{aligned}T^{8q/10(2q+1)}&\left(\widehat{J}_T^*\left(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*\right)-\widehat{J}_T^*\left(b_{\theta_1,T},\,\widehat{b}_{2,T}^*\right)\right)\\&=T^{8q/10(2q+1)}\sum_{k\in N_1}\left(K_1\left(\widehat{b}_{1,T}^*k\right)-K_1\left(b_{\theta_1,T}k\right)\right)\widehat{\Gamma}^*(k)\\&\quad+T^{8q/10(2q+1)}\sum_{k\in N_2}K_1\left(\widehat{b}_{1,T}^*k\right)\widehat{\Gamma}^*(k)\\&\quad-T^{8q/10(2q+1)}\sum_{k\in N_2}K_1\left(b_{\theta_1,T}k\right)\widehat{\Gamma}^*(k)\\&\triangleq A_{1,T}+A_{2,T}-A_{3,T}.\end{aligned}\tag{S.21}$$

We first show that $A_{1,T}\xrightarrow{\mathbb{P}}0$. 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,

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

for some $C<\infty$. By Assumption 3.5-(ii), ($\widetilde{n}_T|\widehat{\phi}_D(q)-\phi_{\theta^*}|=O_{\mathbb{P}}(1)$) and using the delta method, it suffices to show that $B_{1,T}+B_{2,T}+B_{3,T}\xrightarrow{\mathbb{P}}0$, where

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

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

$$B_{3,T}=\left(\widehat{b}_{2,T}^{*}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_{T}^{-1}\sum_{k=1}^{S_{T}}k\left|\Gamma_{T}^{*}\left(k\right)\right|.$$

By a mean-value expansion, we have

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

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 $\sup_{k\geq1}||\left(\partial/\partial\beta\right)\widehat{\Gamma}^{*}\left(k\right)|_{\beta=\overline{\beta}}||=O_{\mathbb{P}}(1)$ using (S.5) and Assumption 3.2-(ii,iii). In addition,

$$\begin{aligned}\mathbb{E}\left(B_{2,T}^{2}\right)&\leq\mathbb{E}\left(\left(\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)}T^{(8q-10)/5(2q+1)}\widetilde{n}_{T}^{-2}\sum_{k=1}^{S_{T}}\sum_{j=1}^{S_{T}}kj\left|\widetilde{\Gamma}^{*}\left(k\right)-\Gamma_{T}^{*}\left(k\right)\right|\left|\widetilde{\Gamma}^{*}\left(j\right)-\Gamma_{T}^{*}\left(j\right)\right|\right) \qquad (S.25)\\&\leq\left(\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1}T^{(8q-10)/5(2q+1)-2/3-1}S_{T}^{4}\sup_{k\geq1}Tb_{2,T}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\\&\leq\left(\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1}T^{(8q-10)/5(2q+1)-2/3-1}\left(Tb_{2,T}\right)^{4r/(2q+1)}\sup_{k\geq1}Tb_{2,T}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\\&\leq T^{1/5}T^{2/5(2q+1)}T^{(8q-10)/5(2q+1)-2/3-1}T^{4r/(2q+1)}T^{-4r/5(2q+1)}\sup_{k\geq1}Tb_{2,T}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\to0,\end{aligned}$$

given that $\sup_{k\geq1}Tb_{2,T}\mathrm{Var}(\widetilde{\Gamma}^{*}(k))=O(1)$ using Lemma S.A.5 in Casini (2023b) and $r<46/48+20q/48$. Assumption 3.5-(iii) and $\sum_{k=1}^{\infty}k^{1-l}<\infty$ for $l>2$ yield

$$\begin{aligned}B_{3,T}&\leq\widehat{b}_{2,T}^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_{T}^{-1}C_{3}\sum_{k=1}^{\infty}k^{1-l} \qquad (S.26)\\&\leq T^{(-21-14q)/10(2q+1)}C_{3}\sum_{k=1}^{\infty}k^{1-l}\to0,\end{aligned}$$

where we have used the fact that $\widetilde{n}_{T}/T^{1/3}\to\infty$. Combining (S.22)-(S.26) we deduce that $A_{1,1,T}\overset{\mathbb{P}}{\to}0$. The same argument applied to $A_{1,T}$ where the summation now also extends over negative integers k gives $A_{1,T}\overset{\mathbb{P}}{\to}0$. Next, we show that $A_{2,T}\overset{\mathbb{P}}{\to}0$. 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

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

$$L_{2,T}=L_{2,T}^{A}+L_{2,T}^{B}=T^{8q/10(2q+1)}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/2}\rfloor}+\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}\right)K_1\left(\widehat{b}_{1,T}^{*}k\right)\left(\widetilde{\Gamma}^{*}\left(k\right)-\Gamma_{T}^{*}\left(k\right)\right),$$

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

$$\begin{aligned}\left|L_{1,T}^{A}\right|&=T^{8q/10(2q+1)-1/3}\sum_{k=S_T+1}^{\lfloor D_T T^{1/2}\rfloor}C_1\left(\widehat{b}_{1,T}^{*}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)}\sum_{k=S_T+1}^{\lfloor D_T T^{1/2}\rfloor}C_1k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)}O_{\mathbb{P}}\left(1\right)O_{\mathbb{P}}\left(1\right),\end{aligned}\tag{S.28}$$

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

$$\begin{aligned}\left|L_{1,T}^{B}\right|&=T^{8q/10(2q+1)-1/3}\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}C_1\left(\widehat{b}_{1,T}^{*}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)}\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}C_1k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)+(1-b)/2}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &=T^{8q/10(2q+1)-1/3+4b/5(2q+1)+(1-b)/2}O_{\mathbb{P}}\left(1\right)\overset{\mathbb{P}}{\to}0,\end{aligned}$$

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

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

Note that

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

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,

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

Note that

$$\begin{aligned}
&\mathbb{E}\left(T^{8q/10(2q+1)+b/(2q+1)-1/2}\left(\widehat{b}_{2,T}^{*}\right)^{b/(2q+1)-1/2}\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}k^{-b}\sqrt{T\widehat{b}_{2,T}^{*}}\left|\widetilde{\Gamma}^{*}(k)-\Gamma_T^{*}(k)\right|\right)^2 \qquad \text{(S.32)}\\
&\leq T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}\left(\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}k^{-b}\sqrt{T\widehat{b}_{2,T}^{*}}\left(\mathrm{Var}\left(\widetilde{\Gamma}^{*}(k)\right)\right)^{1/2}\right)^2\\
&=T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}\left(\sum_{k=\lfloor D_T T^{1/2}\rfloor+1}^{T-1}k^{-b}\right)^2 O(1)\\
&=T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}D_T^{2(1-b)}T^{(1-b)}O(1)\to 0,
\end{aligned}$$

since $r>(8b-4)/((b-1)(10q+5))$ and $\sqrt{T\widehat{b}_{2,T}^{*}}\mathrm{Var}\left(\widetilde{\Gamma}^{*}(k)\right)=O(1)$ as above. Combining (S.29)-(S.30) yields $L_{2,T} \xrightarrow{\mathbb{P}} 0$. Let us turn to $L_{3,T}$. By Assumption 3.5-(iii) and $|K_1(\cdot)| \leq 1$, we have,

$$\begin{aligned} |L_{3,T}| &\leq T^{8q/10(2q+1)} \sum_{k=S_T}^{T-1} C_3 k^{-l} \leq T^{8q/10(2q+1)} C_3 S_T^{1-l} \\ &\leq C_3 T^{8q/10(2q+1)} T^{-4r(l-1)/5(2q+1)} \to 0, \end{aligned} \tag{S.33}$$

since $r > q/(l-1)$. In view of (S.27)-(S.33) we deduce that $A_{2,1,T} \xrightarrow{\mathbb{P}} 0$. Applying the same argument to $A_{2,2,T}$, we have $A_{2,T} \xrightarrow{\mathbb{P}} 0$. Using similar arguments, one has $A_{3,T} \xrightarrow{\mathbb{P}} 0$. It remains to show that $T^{8q/10(2q+1)}(\widehat{J}_T^*(b_{\theta_1,T}, \widehat{b}_{2,T}^*) - \widehat{J}_T^*(b_{\theta_1,T}, b_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. The proof of the latter result follows from the proof of the corresponding result in Theorem 5.1-(ii) in Casini (2023b) with $r < 2/3 + q/3$ and $r > (b - 2/3 - q/3)/(b-1)$. □

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

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

$$\times\left(V_t^*\left(I_p-\sum_{j=1}^{p_A}A_{D,t,j}\right)^{-1}\right)'.$$

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

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

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

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

It follows that

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

Altogether, this implies $J_{T,D}^* \xrightarrow{\mathbb{P}} J_T$. Using Assumption 3.6 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

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

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

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

Equation (S.38) follows from

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

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

since the presence of $\widehat{D}_s$ in $\widehat{V}^*_{D,s}$ is irrelevant. Thus, Lemma S.B.3-S.B.4 yield (S.38). Given that $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}/n_T}\to 0$, Assumption 3.6 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 (2023b) and Assumption 3.6, we deduce that

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

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 (2023b). The same argument in Casini (2023b) now with reference to Theorem 3.1-(i,ii) yields

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

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

S.C Proofs of the Results in Section 4

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

S.C.1 Proof of Theorem 4.1

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

Lemma S.C.1. *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.C.1. 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 $r = 0,\ldots,\lfloor T/n_T \rfloor$,

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

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

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

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

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

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

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

By a similar argument,

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

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

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

$$
\begin{aligned}
&+2\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\sum_{k=-T+1}^{T-1}\sum_{j=-T+1}^{T-1}\left|K_{1}\left(b_{1,T}k\right)K_{1}\left(b_{1,T}j\right)\right|\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}\left(\frac{1}{Tb_{2,T}}\right)^{2}\\
&\left(\frac{1}{Tb_{2,T}}\right)^{2}\sum_{s=k+1}^{T}\sum_{l=j+1}^{T}\left|K_{2}^{*}\left(\frac{\left(\left(r+1\right)n_{T}-\left(s-k/2\right)\right)/T}{b_{2,T}}\right)K_{2}^{*}\left(\frac{\left(\left(b+1\right)n_{T}-\left(l-j/2\right)\right)/T}{b_{2,T}}\right)\right|\\
&\times\kappa_{s}^{*}\left(-k,\,l-s,\,l-j-s\right)\\
=&\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_{U}}\left(a'\widetilde{J}_{T,K}a\right), \qquad \text{(S.3)}
\end{aligned}
$$

where the last inequality uses (S.1). For $K_1(\cdot)\in\boldsymbol{K}_{1,+}$, we can rely on an argument analogous to that of (S.3) using (S.2) in place of (S.1) to yield,

$$
\begin{aligned}
\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_{L}}\left(a'\widetilde{J}_{T}a\right)&\geq\lim_{T\rightarrow\infty}\inf_{\mathscr{P}\in\boldsymbol{P}_{L}}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{J}_{T}a\right)\\
&\geq\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_{L}}\left(a'\widetilde{J}_{T,K}a\right). \qquad \text{(S.4)}
\end{aligned}
$$

By Theorem 3.1 in Casini (2023b),

$$
\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_{w}}\left(a'\widetilde{J}_{T}a\right)=8\pi^{2}\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx\left(\int_{0}^{1}f_{\mathscr{P}_{w},a}\left(u,\,0\right)du\right)^{2},\qquad\text{and}\qquad \text{(S.5)}
$$

$$
\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_{w}}\left(a'\widetilde{J}_{K,T}a\right)=8\pi^{2}\int\left|K_{1}\left(y\right)\right|^{2}dy\int_{0}^{1}\left|K_{2}\left(x\right)\right|^{2}dx\left(\int_{0}^{1}f_{\mathscr{P}_{w},a}\left(u,\,0\right)du\right)^{2}, \qquad \text{(S.6)}
$$

for $w=L,\,U$. Equations (S.3), (S.5) and (S.6) combine to establish part (i) of the lemma:

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

By a similar reasoning, equations (S.4) and (S.5) 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.C.2. *Let Assumption 4.1 hold,* $K_1(\cdot)\in\boldsymbol{K}_1$, $K_2(\cdot)\in\boldsymbol{K}_2$, $b_{1,T},\,b_{2,T}\rightarrow0$, $n_T\rightarrow\infty$, $n_T/T\rightarrow0$, $1/Tb_{1,T}b_{2,T}\rightarrow0$, $1/Tb_{1,T}^{q}b_{2,T}\rightarrow0$, $n_T/Tb_{1,T}^{q}\rightarrow0$ *and* $b_{2,T}^{2}/b_{1,T}^{q}\rightarrow0$ *for some* $q\in[0,\,\infty)$ *for which* $K_{1,q}$, $|\int_0^1 f_{\mathscr{P}_w,a}^{(q)}(u,\,0)\,du|\in[0,\,\infty)$, $w=U,\,L$. *We have for all* $a\in\mathbb{R}^{p_\beta}$*:*

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

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

Proof of Lemma S.C.2. We begin with part (i). We have,

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

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

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

Thus,

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

By Lemma S.A.1 in Casini (2023b), $\mathbb{E}_{\mathscr{P}_U}(V_sV_{s-k}')=c\left(s/T,\,k\right)+O\left(T^{-1}\right)$ uniformly in s and k. By the proof of Lemma S.A.8 in Casini (2023b),

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

It then follows that $\lim_{T\rightarrow\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q}|G_{1,\mathscr{P},T}|=0$ given the conditions $n_T/Tb_{1,T}^q\rightarrow 0$ and $b_{2,T}^2/b_{1,T}^q\rightarrow 0$. Next, given that $1-K_1(b_{1,T}k)\geq 0$,

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

Write the right-hand side above as,

$$\begin{aligned}&\lim_{T\rightarrow\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1\left(b_{1,T}k\right)\right)a'\left(\mathbb{E}_{\mathscr{P}_U}\left(\widetilde{\Gamma}\left(k\right)\right)-\int_0^1 c_{\mathscr{P}_U}\left(u,\,k\right)du\right)a\\ &\quad+\lim_{T\rightarrow\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1\left(b_{1,T}k\right)\right)a'\left(\int_0^1 c_{\mathscr{P}_U}\left(u,\,k\right)du\right)a.\end{aligned}\qquad\text{(S.7)}$$

By Lemma S.A.1 in Casini (2023b), the first term above is less than,

$$\lim_{T\rightarrow\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1\left(b_{1,T}k\right)\right)O\left(T^{-1}\right)=0.\qquad\text{(S.8)}$$

Thus, it remains to consider the second term of (S.7). 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: $w(x)\rightarrow K_{1,q}$ as $x\rightarrow 0$; $w(\cdot)$ 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 $|\int_0^1 f_{\mathscr{P}_w,a}^{(q)}(u,\,0)\,du|\in[0,\,\infty)$, $w=U,\,L$. 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.8), we have

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

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 $f_{\mathscr{P}_U,a}^{(q)}$ are replaced by $\inf_{\mathscr{P}\in\boldsymbol{P}_L}$, $\Gamma_{\mathscr{P}_L,u}$ and $f_{\mathscr{P}_L,a}^{(q)}$. $\square$

Proof of Theorem 4.1. Parts (i) and (ii) of the theorem follow from Lemma S.C.1-(i) and Lemma S.C.2-(i), and Lemma S.C.1-(ii) and Lemma S.C.2-(ii), respectively. □

S.C.2 Proof of Theorem 4.2

Lemma S.C.1-S.C.2 [with $q=0$ in part (ii)] implies $\widetilde{J}_T - J_{\mathscr{P},T} = o_{\mathscr{P}}(1)$. 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 3.2. 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,

$$\sqrt{T} b_{1,T}\left(\widehat{J}_T - \widetilde{J}_T\right) = b_{1,T} \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k) \frac{\partial}{\partial \beta'} \widehat{\Gamma}(k) |_{\beta=\bar{\beta}} \sqrt{T}\left(\widehat{\beta} - \beta_0\right), \tag{S.9}$$

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.9) is

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

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

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

We can use the same argument as in (S.5) but now using Assumption 4.3-(ii), sot that

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

and thus,

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

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

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

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

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

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

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

$$\times K_2^*\left(\frac{((b+1)\,n_T-(l+j/2))\,/T}{b_{2,T}}\right)|\mathbb{E}_{\mathscr{P}}\left(V_sV_l\right)|\,.$$

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

$$\mathbb{E}_{\mathscr{P}}\left(A_3^2\right)\leq\frac{1}{Tb_{1,T}b_{2,T}}\left(b_{1,T}\sum_{k=-T+1}^{T-1}|K_1\left(b_{1,T}k\right)|\right)^2\int_0^1K_2^2\left(x\right)dx\int_0^1\sum_{h=-\infty}^{\infty}|c_{\mathscr{P}_U}\left(u,\,h\right)|\,du=o\left(1\right), \tag{S.10}$$

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

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

$$\begin{aligned}
&\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\left|\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{J}_Ta\right)-\mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_Ta\right)\right| \qquad \text{(S.11)}\\
&=\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\left|2\mathbb{E}_{\mathscr{P}}\left(a'\widetilde{J}_Ta-a'J_{\mathscr{P},T}a\right)\overline{G}_T+\mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right)\right|\\
&\leq2\lim_{T\to\infty}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_Ta\right)\right)^{1/2}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right)\right)^{1/2}\\
&\quad+\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right).
\end{aligned}$$

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) $\limsup_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}(a'\widetilde{J}_Ta)<\infty$. Result (b) follows by Lemma S.C.1-(i). A second-order expansion yields,

$$\overline{G}_T=\left[\frac{\partial}{\partial\beta}a'\widetilde{J}_T\left(\beta_0\right)a\right]\left(\widehat{\beta}-\beta_0\right)+\frac{1}{2}\left(\widehat{\beta}-\beta_0\right)'\left[\frac{\partial^2}{\partial\beta\partial\beta'}a'\widetilde{J}_T\left(\overline{\beta}\right)a\right]\left(\widehat{\beta}-\beta_0\right)=\overline{G}_{1,T}+\overline{G}_{2,T}, \tag{S.12}$$

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 4.4,

$$\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_{1,T}^2\right) \qquad \text{(S.13)}\\
&\leq Tb_{1,T}b_{2,T}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta^{(r)}}a'\widetilde{J}_T\left(\beta_0\right)a\left(\widehat{\beta}^{(r)}-\widehat{\beta}_0^{(r)}\right)\right)^2\\
&\leq\frac{1}{Tb_{1,T}}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(H_{1,T}^{(r)}\sqrt{T}\left(\widehat{\beta}^{(r)}-\widehat{\beta}_0^{(r)}\right)\right)^2\\
&\to0,
\end{aligned}$$

and

$$\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_{2,T}^2\right) \tag{S.14}$$

$$\leq \frac{1}{4}Tb_{1,T}b_{2,T}p^2 \max_{r\leq p} \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathbb{E}_{\mathscr{P}} \left(\left|\widehat{\beta}^{(r)} - \beta_0^{(r)}\right| \frac{\partial^2}{\partial\beta^{(r)}\partial\beta^{(r)\prime}} a' \widetilde{J}_T\left(\overline{\beta}\right) a \left|\widehat{\beta}^{(r)} - \beta_0^{(r)}\right|\right)^2$$
$$\leq \frac{b_{2,T}}{Tb_{1,T}}p^2 \max_{r\leq p} \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathbb{E}_{\mathscr{P}} \left(\sqrt{T}\left|\widehat{\beta}^{(r)} - \beta_0^{(r)}\right| H_{2,T}^{(r)}\sqrt{T}\left|\widehat{\beta}^{(r)} - \beta_0^{(r)}\right|\right)^2$$
$$\to 0.$$

Equations (S.12) to (S.14) and the Cauchy-Schwartz inequality yield result (a) and thus the desired result of the theorem. □

S.C.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.1),

$$\begin{aligned}
&\operatorname{Var}_{\mathscr{P}_U}\left(a'\widetilde{c}_T(u_0, k)\, a\right) \\
&\quad\leq \sup_{\mathscr{P}\in\boldsymbol{P}_U} \operatorname{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\leq (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\leq \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= \operatorname{Var}_{\mathscr{P}_U}\left(a'\overline{c}_T(u_0, k)\, a\right),
\end{aligned} \tag{S.15}$$

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 $K_2(\cdot) \geq 0$ by definition, Proposition 3.1 in Casini (2023b) implies

$$\begin{aligned}
&\operatorname{Var}_{\mathscr{P}_U}\left(a'\widetilde{c}_T(u_0, k)\, a\right) \\
&\qquad= \frac{1}{Tb_{2,T}}\int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} a'\left(c_{\mathscr{P}_U}(u_0, l)\left[c_{\mathscr{P}_U}(u_0, l) + c_{\mathscr{P}_U}(u_0, l+2k)\right]'\right) a
\end{aligned}$$

$$\begin{aligned}
&+ \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \kappa_{\mathscr{P}_U, aV, Tu_0}(h_1,\, 0,\, h_2) \\
&+ o\left(b_{2,T}^4\right) + O\left(1/\left(b_{2,T}T\right)\right) \\
&= \mathrm{Var}_{\mathscr{P}_U}\left(a'\overline{c}_T(u_0,\, k)\, a\right). \qquad \text{(S.16)}
\end{aligned}$$

Next, we discuss the bias. We have,

$$\begin{aligned}
&\sup_{\mathscr{P}\in \boldsymbol{P}_{U,2}} \left|\mathbb{E}_{\mathscr{P}}\left(a'\widetilde{c}_T(u_0,\, k)\, a - a'c_{\mathscr{P}}(u_0,\, k)\, a\right)\right| \\
&\quad = \lim_{T\to\infty} \sup_{\mathscr{P}\in \boldsymbol{P}_{U,2}} \left| (Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2\left(\frac{\left((r+1)\, n_T - (s+k/2)\right)/T}{b_{2,T}}\right) a'\mathbb{E}_{\mathscr{P}}\left(V_s V_{s-k}'\right) a - a'c_{\mathscr{P}}(u_0,\, k)\, a \right| \\
&\quad \leq \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2(x)\, dx \int_0^1 \left| a' \frac{\partial^2}{\partial^2 u} c_{\mathscr{P}_U}(u_0,\, k)\, a \right| du + o\left(b_{2,T}^2\right) + O\left(\frac{1}{Tb_{2,T}}\right), \qquad \text{(S.17)}
\end{aligned}$$

where the inequality above follows from (4.2). Combining (S.16)-(S.17), 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 (4.3). 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 (4.3). The form for the optimal $b_{2,T}(\cdot)$ and $K_2(\cdot)$ follow from the same argument as in Proposition 4.1 in Casini (2023b). □

S.C.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 $K_{1,q}$, $|\int_0^1 f_{U,a}^{(q)}(u,\, 0)\, du| \in [0,\, \infty)$, then by Lemma S.C.1-(i) and Lemma S.C.2-(i),

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

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

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

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

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

The optimality of K_1^{QS} then follows from the same argument as in the proof of Theorem 4.1 in Casini (2023b). □

S.C.5 Proof of Theorem 4.4

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

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

By Theorem 4.1-(i),

$$\begin{aligned}
&\liminf_{T\to\infty} \left(Tb_{2,T}\right)^{2q/(2q+1)} \sup_{\mathscr{P}\in \boldsymbol{P}_U(\phi(q))} \mathbb{E}_{\mathscr{P}} \mathrm{L}\left(\widetilde{J}_T(b_{1,T}),\, J_{\mathscr{P},T}\right) \qquad \text{(S.18)}\\
&= \liminf_{T\to\infty} \left(\gamma^{-1/(2q+1)} + o(1)\right) Tb_{1,T}b_{2,T} \sup_{\mathscr{P}\in \boldsymbol{P}_U(\phi)} \sum_{r=1}^{p} w_r \mathrm{MSE}_{\mathscr{P}}\left(a^{(r)\prime}\widetilde{J}_T(b_{1,T})\, a^{(r)}\right)\\
&= \gamma^{-1/(2q+1)} 4\pi^2 \Bigg[\sum_{r=1}^{p} w_r \Bigg(\gamma K_{1,q}^2 \left(\int_0^1 f^{(q)}_{\mathscr{P}_U,a^{(r)}}(u,\,0)\, du\right)^2\\
&\quad + 2\int K_1^2(x)\, dx \int_0^1 K_2^2(y)\, dy \left(\int_0^1 f_{\mathscr{P}_U,a^{(r)}}(u,\,0)\, du\right)^2\Bigg)\Bigg].
\end{aligned}$$

The right-hand side above is minimized at $\gamma^{\text{opt}} = (2qK_{1,q}^2\phi(q))^{-1}(\int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx)$. Note that $\gamma^{\text{opt}} > 0$ provided that $f_{\mathscr{P}_U,a^{(r)}}(u,\,0) > 0$ and $f^{(q)}_{\mathscr{P}_U,a^{(r)}}(u,\,0) > 0$ for some $u \in [0,\,1]$ and some r for which $w_r > 0$. Hence, $\{b_{1,T}\}$ is optimal in the sense that $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma^{\text{opt}}$ if and only if $b_{1,T} = b_{1,T}^{\text{opt}} + o((Tb_{2,T})^{-1/(2q+1)})$. In virtue of Theorem 4.2-(iii), eq. (S.18) 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.C.6 Proof of Theorem 4.5

The proof of the theorem uses the following lemmas.

Lemma S.C.3. *Let $K_1(\cdot)$, $K_2(\cdot)$, $\{b_{1,\theta_{\mathscr{P}},T}\}$, $\{S_{\mathscr{P},T}\}$, $\widehat{\phi}(\cdot)$ and q be as in Theorem 4.5. Then, for all $a \in \mathbb{R}^p$, (i)*

$$T^{8q/5(2q+1)} \sup_{\mathscr{P}\in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1} K_1\left(\widehat{b}_{1,T}k\right) a'\widehat{\Gamma}(k)\, a\right)^2 \to 0;$$

(ii)

$$T^{8q/5(2q+1)} \sup_{\mathscr{P}\in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(K_1\left(\widehat{b}_{1,T}k\right) - K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)\right) a'\widehat{\Gamma}(k)\, a\right)^2 \to 0.$$

Proof of Lemma S.C.3. First we prove part (i). We have,

$$\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)a'\widehat{\Gamma}\left(k\right)a\right)^2\right)^{1/2} \tag{S.19}$$
$$\leq\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)\left(a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}a\right)\right)^2\right)^{1/2}$$
$$+\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)a'\Gamma_{\mathscr{P},T}a\right)^2\right)^{1/2}$$
$$\triangleq B_{1,T}+B_{2,T}.$$

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

$$B_{2,T}\leq\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}\left|K_1\left(\widehat{b}_{1,T}k\right)\right|a'\left(\int_0^1\Gamma_{\mathscr{P}_U,u}\left(k\right)du\right)a\right)^2\right)^{1/2} \tag{S.20}$$
$$\leq T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\sum_{k=S_{\mathscr{P},T}+1}^{T-1}\sup_{u\in[0,\,1]}a'\left(\int_0^1\Gamma_{\mathscr{P}_U,u}\left(k\right)du\right)a$$
$$\leq T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\sum_{k=S_{\mathscr{P},T}+1}^{T-1}C_3k^{-l}$$
$$\leq C_{3,1}T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\int_{S_{\mathscr{P},T}}^{\infty}k^{-l}dk$$
$$\leq C_{3,1}T^{8q/10(2q+1)}S_{T,\mathscr{P}}^{1-l}$$
$$=T^{8q/10(2q+1)+4r(1-l)/5(2q+1)}\to 0,$$

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

$$B_{1,1,T}^{2}\leq T^{8q/5(2q+1)-4/5}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\left\lfloor D_{T}T^{1/2}\right\rfloor}C_{1}\left(\widehat{b}_{1,T}k\right)^{-b}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left|a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}\left(k\right)a\right|\right)^{2} \tag{S.21}$$

$$\leq T^{8q/5(2q+1)-4/5+8b/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\left\lfloor D_{T}T^{1/2}\right\rfloor}C_{1}k^{-b}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left|a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}\left(k\right)a\right|\right)^{2}$$

$$\times\left(2qK_{1,q}^{2}\widehat{\phi}\left(q\right)/(\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx)\right)^{2b/(2q+1)}$$

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

$$\times\sup_{\mathscr{P}\in\mathscr{P}_{U,3}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\left\lfloor D_{T}T^{1/2}\right\rfloor}\sum_{j=S_{\mathscr{P},T}+1}^{\left\lfloor D_{T}T^{1/2}\right\rfloor}k^{-b}j^{-b}T\overline{b}_{\theta_{2},T}\left(\mathrm{Var}_{\mathscr{P}}\left(a'\widehat{\Gamma}\left(k\right)a\right)\mathrm{Var}_{\mathscr{P}}\left(a'\widehat{\Gamma}\left(j\right)a\right)\right)^{1/2}\right)$$

$$\leq C_{1,2}T^{8q/5(2q+1)-4/5+8b/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left(\left(\sum_{k=S_{\mathscr{P},T}+1}^{\left\lfloor D_{T}T^{1/2}\right\rfloor}k^{-b}\right)^{2}T\overline{b}_{2,T}^{\mathrm{opt}}\left(\sup_{k\geq1}\mathrm{Var}_{\mathscr{P}_{U}}\left(a'\widehat{\Gamma}\left(k\right)a\right)\right)\right)$$

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

$$\leq C_{1,3}T^{8q/5(2q+1)-4/5+8b/5(2q+1)-8(b-1)r/5(2q+1)}\rightarrow0,$$

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

$$B_{1,2,T}^{2}\leq T^{8q/5(2q+1)-4/5}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=\left\lfloor D_{T}T^{1/2}\right\rfloor+1}^{T}C_{1}\left(\widehat{b}_{1,T}k\right)^{-b}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left|a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}\left(k\right)a\right|\right)^{2} \tag{S.22}$$

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

$$\leq C_{1,3}T^{8q/5(2q+1)-4/5+8b/5(2q+1)-(b-1)}\rightarrow0,$$

for some constants $0<C_{1,2},\,C_{1,3}<\infty$ and with q satisfying $8/q-20q<6$. Equations (S.19)-(S.22) combine to establish part (i). We now prove part (ii). Using the Lipschitz condition on $K_{1}\left(\cdot\right)$, we get

$$A_{1,T}=T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=1}^{S_{\mathscr{P},T}}\left(K_{1}\left(\widehat{b}_{1,T}k\right)-K_{1}\left(b_{1,\theta_{\mathscr{P}},T}k\right)\right)a'\widehat{\Gamma}\left(k\right)a\right)^{2} \tag{S.23}$$

$$
\begin{aligned}
&\leq T^{8q/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} C_2 \left(\widehat{b}_{1,T} - b_{1,\theta_{\mathscr{P}},T} \right) k a' \widehat{\Gamma}(k) a \right)^2 \\
&\leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)} \widetilde{n}_T^{-1} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right) k a' \widehat{\Gamma}(k) a \right)^2 \\
&\leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right) k a' \widehat{\Gamma}(k) a \right)^2
\end{aligned}
$$

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,

$$
\begin{aligned}
A_{1,T}^{1/2} \leq &\left(C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2 \right. \\
&\left. \times \left(\sum_{k=1}^{S_{\mathscr{P},T}} k \left(a' \widehat{\Gamma}(k) a - a' \Gamma_{\mathscr{P},T}(k) a \right) \right)^2 \right)^{1/2} \\
&+ \left(C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2 \right. \\
&\left. \times \left(\sum_{k=1}^{S_{\mathscr{P},T}} k a' \Gamma_{\mathscr{P},T}(k) a \right)^2 \right)^{1/2} \\
= &\, A_{1,1,T} + A_{1,2,T}.
\end{aligned} \tag{S.24}
$$

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,

$$
\begin{aligned}
A_{1,1,T}^2 \leq &\, C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5} S_{\mathscr{P},T}^4 \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2 \\
&\times \left(\frac{1}{S_{\mathscr{P},T}} \sum_{k=1}^{S_{\mathscr{P},T}} \frac{k}{S_{\mathscr{P},T}} \left(a' \widehat{\Gamma}(k) a - a' \Gamma_{\mathscr{P},T}(k) a \right) \right)^2 \\
\leq &\, C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5+16r/5(2q+1)-4/5} \\
&\times \left(\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q)\, \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^4 \right)^{1/2}
\end{aligned} \tag{S.25}
$$

$$\times\left(\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\frac{1}{S_{\mathscr{P},T}}\sum_{k=1}^{S_{\mathscr{P},T}}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left(a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}\left(k\right)a\right)\right)^{4}\right)^{1/2}$$
$$\times\left(2qK_{1,q}^{2}\phi_{\theta_{\mathscr{P}}^{*}}\left(q\right)/\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx\right)^{4r/(2q+1)}\rightarrow0,$$

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

$$\begin{aligned}A_{1,2,T}^{2}&\leq C_{2,1}T^{8q/5(2q+1)-8/5(2q+1)-3/5}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\frac{\sqrt{\widetilde{n}_{T}}\left(\widehat{\phi}\left(q\right)^{1/(2q+1)}-\phi_{\theta_{\mathscr{P}}^{*}}\left(q\right)^{1/(2q+1)}\right)}{\left(\widehat{\phi}\left(q\right)\phi_{\theta_{\mathscr{P}}^{*}}\left(q\right)\right)^{1/(2q+1)}}\right)^{2}\\&\quad\times\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left(\sum_{k=1}^{S_{\mathscr{P},T}}ka'\Gamma_{\mathscr{P},T}\left(k\right)a\right)^{2}\\&\leq C_{2,1}T^{8q/5(2q+1)-8/5(2q+1)-3/5}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\frac{\sqrt{\widetilde{n}_{T}}\left(\widehat{\phi}\left(q\right)^{1/(2q+1)}-\phi_{\theta_{\mathscr{P}}^{*}}\left(q\right)^{1/(2q+1)}\right)}{\left(\widehat{\phi}\left(q\right)\phi_{\theta_{\mathscr{P}}^{*}}\left(q\right)\right)^{1/(2q+1)}}\right)^{2}\\&\quad\times\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left(\sum_{k=1}^{S_{\mathscr{P},T}}k^{1-l}\right)^{2}\rightarrow0,\end{aligned}\tag{S.26}$$

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.24)-(S.26) combine to establish part (ii) of the lemma. □

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

$$\left\Vert \widehat{J}_{1}-\widehat{J}_{2}\right\Vert _{\mathscr{P}}\geq\left|\left\Vert \widehat{J}_{1}-J\right\Vert _{\mathscr{P}}-\left\Vert J-\widehat{J}_{2}\right\Vert _{\mathscr{P}}\right|.\tag{S.27}$$

Hence, it suffices to show that

$$T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left\Vert a'\widehat{J}_{T}\left(\widehat{b}_{1,T},\,\widehat{\overline{b}}_{2,T}\right)a-a'\widehat{J}_{T}\left(b_{1,\theta_{\mathscr{P}},T},\,\overline{b}_{2,T}^{\mathrm{opt}}\right)a\right\Vert _{\mathscr{P}}^{2}\rightarrow0.\tag{S.28}$$

The latter follows from

$$\begin{aligned}&T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left\Vert a'\widehat{J}_{T}\left(\widehat{b}_{1,T},\,\widehat{\overline{b}}_{2,T}\right)a-a'\widehat{J}_{T}\left(b_{1,\theta_{\mathscr{P}},T},\,\widehat{\overline{b}}_{2,T}\right)a\right\Vert _{\mathscr{P}}^{2}\\&\quad+T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\left\Vert a'\widehat{J}_{T}\left(b_{1,\theta_{\mathscr{P}},T},\,\widehat{\overline{b}}_{2,T}\right)a-a'\widehat{J}_{T}\left(b_{1,\theta_{\mathscr{P}},T},\,\overline{b}_{2,T}^{\mathrm{opt}}\right)a\right\Vert _{\mathscr{P}}^{2}\rightarrow0.\end{aligned}\tag{S.29}$$

Note that

$$a'\widehat{J}_{T}\left(\widehat{b}_{1,T},\,\widehat{\overline{b}}_{2,T}\right)a-a'\widehat{J}_{T}\left(b_{1,\theta_{\mathscr{P}},T},\,\widehat{\overline{b}}_{2,T}\right)a\tag{S.30}$$

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

We can apply Lemma S.C.3-(ii) to the first term of (S.30) and Lemma S.C.3-(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.29) 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 $k\geq 0$,

$$
\begin{aligned}
&\widehat{c}_T(rn_T/T,\,k)-\widehat{c}_{\theta_2,T}(rn_T/T,\,k)\\
&\quad=\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\sum_{s=k+1}^{T}\left(K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))\,/T}{\widehat{b}_{2,T}((r+1)\,n_T/T)}\right)-K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))\,/T}{b_{2,T}^{\mathrm{opt}}((r+1)\,n_T/T)}\right)\right)\widehat{V}_s\widehat{V}_{s-k}\\
&\qquad+O_{\mathbb{P}}\left(1/T\overline{b}_{2,T}^{\mathrm{opt}}\right).
\end{aligned}
\tag{S.31}
$$

Given Assumption 3.5-(v) 4.5-(ii,iii) and using the delta method, we have for $s\in\{Tu-\lfloor T\overline{b}_{2,T}^{\mathrm{opt}}\rfloor,\ldots,\,Tu+\lfloor T\overline{b}_{2,T}^{\mathrm{opt}}\rfloor\}$:

$$
\begin{aligned}
&K_2\left(\frac{(Tu-(s-k/2))\,/T}{\widehat{b}_{2,T}(u)}\right)-K_2\left(\frac{(Tu-(s-k/2))\,/T}{b_{2,T}^{\mathrm{opt}}(u)}\right)\\
&\quad\leq C_4\left|\frac{Tu-(s-k/2)}{T\widehat{b}_{2,T}(u)}-\frac{Tu-(s-k/2)}{Tb_{2,T}^{\mathrm{opt}}(u)}\right|\\
&\quad\leq CT^{-4/5-2/5}T^{2/5}\left|\left(\frac{\widehat{D}_2(u)}{\widehat{D}_1(u)}\right)^{-1/5}-\left(\frac{D_2(u)}{D_{1,\theta}(u)}\right)^{-1/5}\right|\left|Tu-(s-k/2)\right|\\
&\quad\leq CT^{-4/5-2/5}O_{\mathbb{P}}(1)\left|Tu-(s-k/2)\right|.
\end{aligned}
\tag{S.32}
$$

Therefore,

$$
\begin{aligned}
&T^{8q/10(2q+1)}\left(a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\,\widehat{\overline{b}}_{2,T}\right)a\,-\,a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\,\overline{b}_{2,T}^{\mathrm{opt}}\right)a\right)\\
&\quad=T^{8q/10(2q+1)}\sum_{k=-T+1}^{T-1}K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(a'\widehat{c}\,(rn_T/T,\,k)\,a-a'\widehat{c}_{\theta_2,T}(rn_T/T,\,k)\,a\right)\\
&\quad\leq T^{8q/10(2q+1)}C\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)\right|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\frac{1}{T\overline{b}_{2,T}^{\mathrm{opt}}}\\
&\qquad\times\sum_{s=k+1}^{T}\left|K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))\,/T}{\widehat{b}_{2,T}((r+1)\,n_T/T)}\right)-K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))\,/T}{b_{2,T}^{\mathrm{opt}}((r+1)\,n_T/T)}\right)\right|\\
&\qquad\times\left|\left(a'\widehat{V}_s\widehat{V}_{s-k}'a-\mathbb{E}_{\mathscr{P}}\left(a'V_sV_{s-k}'a\right)\right)+\mathbb{E}_{\mathscr{P}}\left(a'V_sV_{s-k}'a\right)\right|
\end{aligned}
\tag{S.33}
$$

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

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 $\boldsymbol{P}_{U,3}$,

$$\begin{aligned}\mathbb{E}\left(H_{1,1,T}^{2}\right)\leq{}& T^{8q/5(2q+1)}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T}K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)K_1\left(b_{1,\theta_{\mathscr{P}},T}j\right)\left(\frac{n_T}{T}\right)^{2}\sum_{r_1=0}^{\lfloor T/n_T\rfloor}\sum_{r_2=0}^{\lfloor T/n_T\rfloor}\frac{1}{\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{2}} \qquad \text{(S.34)}\\ &\times\sum_{s=k+1}^{T}\sum_{t=j+1}^{T}\left(K_2^{*}\left(\frac{\left(\left(r_1+1\right)n_T-\left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}\left(\left(r_1+1\right)n_T/T\right)}\right)-K_2^{*}\left(\frac{\left(\left(r_1+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{2,T}^{\mathrm{opt}}\left(\left(r_1+1\right)n_T/T\right)}\right)\right)\\ &\times\left(K_2^{*}\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{\widehat{b}_{2,T}\left(\left(r_2+1\right)n_T/T\right)}\right)-K_2^{*}\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{b_{2,T}^{\mathrm{opt}}\left(\left(r_2+1\right)n_T/T\right)}\right)\right)\\ &\times\mathbb{E}_{\mathscr{P}}\left(a'\widehat{V}_s\widehat{V}_{s-k}'a-\mathbb{E}_{\mathscr{P}}\left(V_sV_{s-k}\right)\right)\left(a'\widehat{V}_t\widehat{V}_{t-j}'a-\mathbb{E}_{\mathscr{P}}\left(V_tV_{t-j}\right)\right)\\ \leq{}& CT^{8q/5(2q+1)}S_T^{2}T^{-2/5}\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\sup_{k\geq1}T\overline{b}_{2,T}^{\mathrm{opt}}\mathrm{Var}_{\mathscr{P}}\left(\widehat{\Gamma}\left(k\right)\right)O_{\mathbb{P}}\left(1\right)\\ \leq{}& CT^{8q/5(2q+1)}S_T^{2}T^{-2/5}\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\sup_{k\geq1}T\overline{b}_{2,T}^{\mathrm{opt}}\mathrm{Var}_{\mathscr{P}_U}\left(\widehat{\Gamma}\left(k\right)\right)O_{\mathbb{P}}\left(1\right)\\ \leq{}& CT^{(8q+8r)/5(2q+1)-2/5-1}O_{\mathbb{P}}\left(\left(\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\right)\to0,\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\left(H_{1,2,T}^{2}\right)\leq{}& T^{8q/5(2q+1)-2/5}\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}b_{1,\theta_{\mathscr{P}},T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left(\mathrm{Var}_{\mathscr{P}}\left(\widehat{\Gamma}\left(k\right)\right)\right)^{1/2}O\left(1\right)\right)^{2} \qquad \text{(S.35)}\\ \leq{}& T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}b_{1,\theta_{\mathscr{P}},T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{T\overline{b}_{2,T}^{\mathrm{opt}}}\left(\mathrm{Var}_{\mathscr{P}_U}\left(\widehat{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^{2}\\ \leq{}& T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}b_{1,\theta_{\mathscr{P}},T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}O\left(1\right)\right)^{2}\\ \leq{}& T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}b_{1,\theta_{\mathscr{P}},T}^{-2b}S_T^{2(1-b)}\to0,\end{aligned}$$

since $r>\left(b-3/4-q/2\right)/\left(b-1\right)$. Eq. (S.34) and (S.35) yield $H_{1,T}\overset{\mathbb{P}}{\to}0$. Given $\left|K_1\left(\cdot\right)\right|\leq1$ and (S.32), we have

$$\left|H_{2,1,T}\right|\leq CT^{8q/10(2q+1)}T^{-2/5}\sum_{k=1}^{\infty}k^{-l}\to0,$$

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

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

which completes the proof. □

S.D Proof of the Results of Section 5

S.D.1 Proof of Theorem 5.1

Consider first the numerator of $t_{\mathrm{DM},i}$. We have

$$\begin{aligned}T_n^{1/2} \overline{d}_L &= \delta^2 O_{\mathbb{P}} \left(T_n^{1/2} T_n^{-1} n_\delta\right) + O_{\mathbb{P}} \left(T_n^{1/2} T_n^{-1} (T_n - n_\delta)^{1/2}\right) \mathscr{N}(0,\, J_{\mathrm{DM}}) \\ &= \delta^2 O_{\mathbb{P}} \left(T_n^{-1/2} n_\delta\right) + O_{\mathbb{P}}(1),\end{aligned}$$

for some $J_{\mathrm{DM}} \in (0,\, \infty)$ where the factor δ^2 follows from the quadratic loss.

Next, we focus on the expansion of the denominator of $t_{\mathrm{DM},i}$ which hinges on which LRV estimator is used. We begin with part (i). Under $b_T \to 0$ as $T \to \infty$, Theorem 3.1 in Casini et al. (2023) yields

$$\begin{aligned}\widehat{J}_{d_L,\mathrm{NW87},T} &= \sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor} (1 - |b_T k|)\, \widehat{\Gamma}(k) \\ &= \sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor} (1 - |b_T k|) \int_0^1 c(u,\, k)\, du \\ &\quad + \sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor} (1 - |b_T k|) \left(2^{-1} \left(\frac{T_b - T_m - 1}{T_n}\right) \left(\frac{T_n - T_b - 2}{T_n}\right) \delta^4 + o_{\mathbb{P}}(1)\right) \\ &= CJ_{\mathrm{DM}} + \sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor} (1 - |b_T k|) \left(2^{-1} \left(\frac{T_b - T_m - 1}{T_n}\right) \left(\frac{T_n - T_b - 2}{T_n}\right) \delta^4 + o_{\mathbb{P}}(1)\right),\end{aligned}$$

for some $C > 0$ such that $C < \infty$. By Exercise 1.7.12 in Brillinger (1975),

$$\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor} (1 - |b_T k|) \exp(-i\omega k) = b_T \left(\frac{\sin \frac{\lfloor b_T^{-1}\rfloor \omega}{2}}{\sin \frac{\omega}{2}}\right)^2.$$

Evaluating the expression above at $\omega=0$ and applying L'Hôpital's rule we yield,

$$\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}(1-|b_Tk|)=b_T\left(\frac{\frac{\lfloor b_T^{-1}\rfloor}{2}}{\frac{1}{2}}\right)^2=\left\lfloor b_T^{-1}\right\rfloor.$$

Therefore, $\widehat{J}_{d_L,\mathrm{NW87},T}=CJ_{\mathrm{DM}}+\delta^4O_{\mathbb{P}}(b_T^{-1})$ and

$$\begin{aligned}|t_{\mathrm{DM,NW87}}|&\leq\frac{\delta^2O_{\mathbb{P}}\left(T_n^{-1/2}n_\delta\right)+O_{\mathbb{P}}(1)}{\left(\delta^4O\left(b_T^{-1}\right)\right)^{1/2}}\\&=\frac{\delta^2O\left(T_n^{\zeta}\right)}{\delta^2O\left(b_T^{-1/2}\right)}=O\left(T_n^{\zeta}b_T^{1/2}\right),\end{aligned}\tag{S.1}$$

which implies $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}|>z_\alpha)\to0$ since $T_n^{\zeta}b_T^{1/2}\to0$.

If $b_T=O(T^{-1/3})$, similar derivations yield $|t_{\mathrm{DM,NW87}}|=O(T_n^{\zeta-1/6})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}|>z_\alpha)\to0$.

We now consider part (ii). We have $b_T\to0$ as $T\to\infty$. The whitening step for $\widehat{J}_{d_L,\mathrm{pwNW87},T}$ involves the following fitted least-squares regression,

$$\widehat{V}_t=\widehat{A}_1\widehat{V}_{t-1}+\widehat{V}_t^*\quad\text{for}\quad t=2,\ldots,T_n-1,$$

where $\widehat{A}_1$ is the least-squares estimate and $\{\widehat{V}_t^*\}$ is the corresponding least-squares residual. Under H_1, $\widehat{V}_t^*$ exhibits a break in the mean of magnitude δ^2 because $\widehat{V}_t$ has a break in the mean of the same magnitude after T_b. From Casini et al. (2023) it follows that $\widehat{A}_1\overset{\mathbb{P}}{\to}A_{1,B}$ with $|1-A_{1,B}|<|1-A_1|$ where A_1 is such that $\widehat{A}_1\overset{\mathbb{P}}{\to}A_1$ under the null hypothesis (i.e., under $\delta=0$). Let

$$\widehat{J}_T^*=\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}K_1(b_Tk)\,\widehat{\Gamma}^*(k),\qquad\widehat{\Gamma}^*(k)=(T_n-2)^{-1}\sum_{t=|k|+2}^{T_n-1}\widehat{V}_t^*\widehat{V}_{t-|k|}^*,$$

and $c^*(u,k)=\mathrm{Cov}(\widehat{V}_{\lfloor Tu\rfloor}^*,\widehat{V}_{\lfloor Tu\rfloor-k}^*)$. Using Theorem 3.1 in Casini et al. (2023),

$$\begin{aligned}\widehat{J}_{d_L,\mathrm{pwNW87},T}&=\left(1-\widehat{A}_1\right)^{-2}\widehat{J}_T^*\\&=\left(1-\widehat{A}_1\right)^{-2}\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}(1-|b_Tk|)\,\widehat{\Gamma}^*(k)\\&=\left(1-\widehat{A}_1\right)^{-2}\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}(1-|b_Tk|)\int_0^1c^*(u,k)\,du\end{aligned}$$

$$
\begin{aligned}
&+\left(1-\widehat{A}_1\right)^{-2}\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}\left(1-|b_Tk|\right)\left(2^{-1}\left(\frac{T_b-T_m}{T_n}\right)\left(\frac{T_n-T_b}{T_n}\right)\delta^4+o_{\mathbb{P}}(1)\right)\\
&=C\left(1-A_{1,B}\right)^{-2}J_{\mathrm{DM}}\\
&+\left(1-A_{1,B}\right)^{-2}\sum_{k=-\lfloor b_T^{-1}\rfloor}^{\lfloor b_T^{-1}\rfloor}\left(1-|b_Tk|\right)\left(2^{-1}\left(\frac{T_b-T_m}{T_n}\right)\left(\frac{T_n-T_b}{T_n}\right)\delta^4+o_{\mathbb{P}}(1)\right),
\end{aligned}
$$

for some finite $C>0$. Thus, $\widehat{J}_{d_L,\mathrm{pwNW87},T}=C\left(1-A_{1,B}\right)^{-2}J_{\mathrm{DM}}+\left(1-A_{1,B}\right)^{-2}\delta^4O_{\mathbb{P}}(b_T^{-1})$ and

$$
\begin{aligned}
|t_{\mathrm{DM,pwNW87}}|&\leq\frac{\delta^2O_{\mathbb{P}}\left(T_n^{-1/2}n_\delta\right)+O_{\mathbb{P}}(1)}{\left(\left(1-A_{1,B}\right)^{-2}\delta^4O\left(b_T^{-1}\right)\right)^{1/2}}\\
&=\frac{\left(1-A_{1,B}\right)\delta^2O\left(T_n^{\zeta}\right)}{\delta^2O\left(b_T^{-1/2}\right)}=O\left(T_n^{\zeta}b_T^{1/2}\right),
\end{aligned}
\tag{S.2}
$$

which implies $\mathbb{P}_\delta(|t_{\mathrm{pwDM,NW87}}|>z_\alpha)\to0$ since $T_n^{\zeta}b_{1,T}^{1/2}\to0$.

In part (iii), $b_T=T^{-1}$. Proceeding as in (S.1) we have $|t_{\mathrm{DM,KVB}}|=O(T_n^{\zeta-1})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,KVB}}|>z_\alpha)\to0$ since $T_n^{\zeta-1}\to0$.

We consider part (iv). Using Theorem 3.3 in Casini et al. (2023), we have

$$
\begin{aligned}
\widehat{J}_{d_L,\mathrm{DK},T}&=\sum_{k=-T_n+2}^{T_n-2}K_1\left(\widehat{b}_{1,T}k\right)\frac{n_{T_n}}{T_n}\sum_{r=1}^{\lfloor T_n/n_{T_n}\rfloor}\widehat{c}_T\left(rn_{T_n}/T_n,\,k\right)\\
&=\sum_{k=-T_n+2}^{T_n-2}K_1\left(\widehat{b}_{1,T}k\right)\frac{n_{T_n}}{T_n}\sum_{r=1}^{\lfloor T_n/n_{T_n}\rfloor}\Big(c\left(rn_{T_n}/T_n,\,k\right)\\
&+\delta^2\mathbf{1}\left\{\left(|rn_{T_n}+k/2+T_n\widehat{b}_{2,T}/2+1)-T_b|/(T_n\widehat{b}_{2,T})\right)\in(0,\,1)\right\}\Big)+o_{\mathbb{P}}(1)\\
&=J_{\mathrm{DM}}+\delta^2O_{\mathbb{P}}\left(\left(\widehat{b}_{1,T}\right)^{-1}\widehat{b}_{2,T}\right)+o_{\mathbb{P}}(1).
\end{aligned}
$$

Using $\widehat{b}_{2,T}/\widehat{b}_{1,T}\to0$ it follows that

$$
\begin{aligned}
|t_{\mathrm{DM,DK}}|&=\frac{\delta^2O_{\mathbb{P}}\left(T_n^{-1/2}n_\delta\right)+O_{\mathbb{P}}(1)}{\left(J_{\mathrm{DM}}+\delta^2O_{\mathbb{P}}\left(\left(\widehat{b}_{1,T}\right)^{-1}\widehat{b}_{2,T}\right)\right)^{1/2}}\\
&=\delta^2O\left(T_n^{\zeta}\right).
\end{aligned}
$$

Since $T_n^{\zeta}\to\infty$, we have $\mathbb{P}_\delta(|t_{\mathrm{DM,DK}}|>z_\alpha)\to1$.

Finally, we consider part (v). The whitening step for $\widehat{J}_{d_L,\mathrm{pwDK},T}$ involves the following fitted least-squares regression,

$$\widehat{V}_t = \widehat{A}_{r,1}\widehat{V}_{t-1} + \widehat{V}_t^* \quad \text{for} \quad t = rn_{T_n} + 1, \ldots, (r+1)\, n_{T_n},$$

(for the last block, $t = \lfloor T_n/n_{T_n} \rfloor n_{T_n} + 1, \ldots, T_n$) where $\widehat{A}_{r,1}$ is the least-squares estimate and $\{\widehat{V}_t^*\}$ is the corresponding least-squares residual. Under H_1, $\widehat{V}_t^*$ exhibits a break in the mean of magnitude δ^2 in the r^*th block such that $T_b \in \{r^* n_{T_n} + 1, \ldots, (r^*+1)\, n_{T_n}\}$. This follows because $\widehat{V}_t$ has a break in the mean of the same magnitude after T_b. Note that over the blocks $r \neq r^*$, $\widehat{V}_t$ does not have a break in the mean. Using Theorem 3.3 in Casini et al. (2023), we have

$$\begin{aligned}
\widehat{J}_{d_L,\mathrm{pwDK},T} &= \sum_{k=-T_n+1}^{T_n-1} K_1\left(\widehat{b}_{1,T}^* k\right) \frac{n_{T_n}}{T_n} \sum_{r=1}^{\lfloor T_n/n_{T_n} \rfloor} \widehat{c}_{T,D}^*\left(rn_{T_n}/T_n,\, k\right) \\
&= \sum_{k=-T_n+1}^{T_n-1} K_1\left(\widehat{b}_{1,T}^* k\right) \frac{n_{T_n}}{T_n} \sum_{r=1}^{\lfloor T_n/n_{T_n} \rfloor} \Bigg(c_D^*\left(rn_{T_n}/T_n,\, k\right) \\
&\quad + \delta^2 \mathbf{1}\left\{ \left(|r^* n_{T_n} + k/2 + T_n \widehat{b}_{2,T}^*/2 + 1) - T_b| / (T_n \widehat{b}_{2,T}^*) \right) \in (0,\, 1) \right\} \Bigg) + o_{\mathbb{P}}(1) \\
&= J_{\mathrm{DM}} + \delta^2 O_{\mathbb{P}}\left(\left(\widehat{b}_{1,T}^*\right)^{-1} \widehat{b}_{2,T}^* \right) + o_{\mathbb{P}}(1).
\end{aligned}$$

It follows that

$$\begin{aligned}
|t_{\mathrm{DM,pwDK}}| &= \frac{\delta^2 O_{\mathbb{P}}\left(T_n^{-1/2} n_\delta\right) + O_{\mathbb{P}}(1)}{\left(J_{\mathrm{DM}} + \delta^2 O_{\mathbb{P}}\left(\left(\widehat{b}_{1,T}^*\right)^{-1} \widehat{b}_{2,T}^* \right) \right)^{1/2}} \\
&= \delta^2 O\left(T_n^{\zeta}\right).
\end{aligned}$$

Since $T_n^{\zeta} \to \infty$ we have $\mathbb{P}_\delta(|t_{\mathrm{DM,pwDK}}| > z_\alpha) \to 1$. □

References

Andrews, D.W.K., 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica 59, 817–858.

Brillinger, D.R., 1975. Time Series Data Analysis and Theory. New York: Holt, Rinehart and Winston.

Casini, A., 2022. Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models. Journal of Econometrics, forthcoming.

Casini, A., Deng, T., Perron, P., 2022. Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference. arXiv preprint arXiv:2103.01604.