Simultaneous Bandwidths Determination for DK-HAC Estimators and Long-Run Variance Estimation in Nonparametric Settings*

Federico Belotti[†]
University of Rome Tor Vergata

Alessandro Casini[‡]
University of Rome Tor Vergata

Leopoldo Catania[§]
Aarhus University

Stefano Grassi[¶]
University of Rome Tor Vergata

Pierre Perron[‖]
Boston University

4th March 2021

Abstract

We consider the derivation of data-dependent simultaneous bandwidths for double kernel heteroskedasticity and autocorrelation consistent (DK-HAC) estimators. In addition to the usual smoothing over lagged autocovariances for classical HAC estimators, the DK-HAC estimator also applies smoothing over the time direction. We obtain the optimal bandwidths that jointly minimize the global asymptotic MSE criterion and discuss the trade-off between bias and variance with respect to smoothing over lagged autocovariances and over time. Unlike the MSE results of Andrews (1991), we establish how nonstationarity affects the bias-variance trade-off. We use the plug-in approach to construct data-dependent bandwidths for the DK-HAC estimators and compare them with the DK-HAC estimators from Casini (2021) that use data-dependent bandwidths obtained from a sequential MSE criterion. The former performs better in terms of size control, especially with stationary and close to stationary data. Finally, we consider long-run variance estimation under the assumption that the series is a function of a nonparametric estimator rather than of a semiparametric estimator that enjoys the usual $\sqrt{T}$ rate of convergence. Thus, we also establish the validity of consistent long-run variance estimation in nonparametric parameter estimation settings.

JEL Classification: C12, C13, C18, C22, C32, C51

Keywords: Fixed-b, HAC standard errors, HAR, Long-run variance, Nonstationarity, Misspecification, Outliers, Segmented locally stationary.

---

*We thank Zhongjun Qu for useful comments.

[†]Dep. of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, IT. Email: federico.belotti@uniroma2.it.

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

[§]Dep. of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, DE. Email: leopoldo.catania@econ.au.dk.

[¶]Dep. of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, IT. Email: stefano.grassi@uniroma2.it.

[‖]Dep. of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, US. Email: perron@bu.edu.

1 Introduction

Long-run variance (LRV) estimation has a long history in econometrics and statistics since it plays a key role for heteroskedasticity and autocorrelation robust (HAR) inference. The classical approach in HAR inference relies on consistent estimation of the LRV. Newey and West (1987) and Andrews (1991) proposed kernel heteroskedasticity and autocorrelation consistent (HAC) estimators and showed their consistency. However, recent work by Casini (2021) showed that, both in the linear regression model and other contexts, their results do not provide accurate approximations in that test statistics normalized by classical HAC estimators may exhibit size distortions and substantial power losses. Issues with the power have been shown for a variety of HAR testing problems outside the regression model [e.g., Altissimo and Corradi (2003), Casini (2018), Casini and Perron (2019, 2020a, 2020b), Chan (2020), Chang and Perron (2018), Crainiceanu and Vogelsang (2007), Deng and Perron (2006), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Perron and Yamamoto (2021) and Vogelsang (1999)]. Casini, Deng, and Perron (2021) showed theoretically that such power issues are generated by low frequency contamination induced by nonstationarity. More specifically, nonstationarity biases upward each sample autocovariance. Thus, LRV estimators are inflated and HAR test statistics lose power. These issues can also be provoked by misspecification, nonstationary alternative hypotheses and outliers. They also showed that LRV estimators that rely on fixed-b or versions thereof suffer more from these problems than classical HAC estimators since the former use a larger number of sample autocovariances.[1]

In order to flexibly account for nonstationarity, Casini (2021) introduced a double kernel HAC (DK-HAC) estimator that applies kernel smoothing over two directions. In addition to the usual smoothing over lagged autocovariances used in classical HAC estimators, the DK-HAC estimator uses a second kernel that applies smoothing over time. The latter accounts for time variation in the covariance structure of time series which is a relevant feature in economics and finance. Since the DK-HAC uses two kernels and bandwidths, one cannot rely on the theory of Andrews (1991) or Newey and West (1994) for selecting the bandwidths. Casini (2021) considered a sequential MSE criterion that determines the optimal bandwidth controlling the number of lags as a function of the optimal bandwidth controlling the smoothing over time. Thus, the latter influences the former but not viceversa. However, each smoothing affects the bias-variance trade-off so that the two bandwidths should affect each others optimal value. Consequently, it is useful to

[1] The fixed-b literature is extensive. Pioneering contribution of Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002; 2005) introduced the fixed-b LRV estimators. Additional contributions can be found in Dou (2019), Lazarus, Lewis, and Stock (2020), Lazarus, Lewis, Stock, and Watson (2018), Gonçalves and Vogelsang (2011), de Jong and Davidson (2000), Ibragimov and Müller (2010), Jansson (2004), Müller (2007, 2014), Phillips (2005), Politis (2011), Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018; 2019), Robinson (1998), Sun (2013; 2014a; 2014b), Velasco and Robinson (2001) and Zhang and Shao (2013).

consider an alternative criterion to select the bandwidths. In this paper, we consider simultaneous bandwidths determination obtained by jointly minimizing the asymptotic MSE of the DK-HAC estimator. We obtain the asymptotic optimal formula for the two bandwidths and use the plug-in approach to replace unknown quantities by consistent estimates. Our results are established under the nonstationary framework characterized by segmented locally stationary processes [cf. Casini (2021)]. The latter extends the locally stationary framework of Dahlhaus (1997) to allow for discontinuities in the spectrum. Thus, the class of segmented locally stationary processes includes structural break models [see e.g., Bai and Perron (1998) and Casini and Perron (2021b)], time-varying parameter models [see e.g., Cai (2007)] and regime switching [cf. Hamilton (1989)].

We establish the consistency, rate of convergence and asymptotic MSE results for the DK-HAC estimators with data-dependent simultaneous bandwidths. The optimal bandwidths have the same order $O(T^{-1/6})$ whereas under the sequential criterion the optimal bandwidths smoothing over time has an order $O(T^{-1/5})$ and the optimal bandwidth smoothing the lagged autocovariances has an order $O(T^{-4/25})$. Thus, asymptotically, the joint MSE criterion implies the use of (marginally) more lagged autocovariances and a longer segment length for the smoothing over time relative to the sequential criterion. Hence, the former should control more accurately the variance due to nonstationarity while the latter should control better the bias. If the degree of nonstationarity is high then the theory suggests that one should expect the sequential criterion to perform marginally better. The difference in the smoothing over lags is very minor between the order of the corresponding bandwidths implied by the two criteria. Our simulation analysis supports this view as we show that the joint MSE criterion performs better especially when the degree of nonstationarity is not too high.

Overall, we find that HAR tests normalized by DK-HAC estimators strike the best balance between size and power among the existing LRV estimators and we also find that using the bandwidths selected from the joint MSE criterion yields tests that perform better than the sequential criterion in terms of size control. The optimal rate $O(T^{-1/6})$ is also found by Neumann and von Sachs (1997) and Dahlhaus (2012) in the context of local spectral density estimates under local stationarity. Under both sequential and joint MSE criterion the optimal kernels are found to be the same, i.e., the quadratic spectral kernel for smoothing over autocovariance lags [similar to Andrews (1991)] and a parabolic kernel [cf. Epanechnikov (1969)] for smoothing over time.

Another contribution of the paper is develop asymptotic results for consistent LRV estimation in nonparametric parameter estimation settings. Newey and West (1987) and Andrews (1991) established the consistency of HAC estimators for the long-run variance of some series $\{V_t(\widehat{\beta})\}$ where $\widehat{\beta}$ is a semiparametric estimator of β_0 having the usual parametric rate of convergence $\sqrt{T}$ [i.e., they assumed that $\sqrt{T}(\widehat{\beta}-\beta_0) = O_{\mathbb{P}}(1)$]. For example, in the linear regression model

estimated by least-squares, $V_t(\widehat{\beta}) = \widehat{e}_t x_t$ where $\{\widehat{e}_t\}$ are the least-squares residuals and $\{x_t\}$ is a vector of regressors. Unfortunately, the condition $\sqrt{T}(\widehat{\beta} - \beta_0) = O_{\mathbb{P}}(1)$ does not hold for nonparametric estimators $\widehat{\beta}_{\text{np}}$ since they satisfy $T^{\vartheta}(\widehat{\beta}_{\text{np}} - \beta_0) = O_{\mathbb{P}}(1)$ for some $\vartheta \in (0, 1/2)$. For example, for tests for forecast evaluation often forecasters use nonparametric kernel methods to obtain the forecasts [i.e., $\{V_t(\widehat{\beta})\} = L(e_t(\widehat{\beta}_{\text{np}}))$ where $L(\cdot)$ is a forecast loss, $e_t(\cdot)$ is a forecast error and $\widehat{\beta}_{\text{np}}$ is, e.g., a rolling window estimate of a parameter that is used to construct the forecasts]. Given the widespread use of nonparametric methods in applied work, it is useful to extend the theoretical results of HAC and DK-HAC estimators for these settings. We establish the validity of HAC and DK-HAC estimators including the validity of the corresponding estimators based on data-dependent bandwidths.

The remainder of the paper is organized as follows. Section 2 introduces the statistical setting and the joint MSE criterion. Section 3 presents consistency, rates of convergence, asymptotic MSE results, and optimal kernels and bandwidths for the DK-HAC estimators using the joint MSE criterion. Section 4 develops a data-dependent method for simultaneous bandwidth parameters selection and its asymptotic properties are then discussed. Section 5 presents theoretical results for LRV estimation in nonparametric parameter estimation. Section 6 presents Monte Carlo results about the small-sample size and power of HAR tests based on the DK-HAC estimators using the proposed automatic simultaneous bandwidths. We also provide comparisons with a variety of other approaches. Section 7 concludes the paper. The supplemental material [Belotti, Casini, Catania, Grassi, and Perron (2021)] contains the mathematical proofs. The code to implement the proposed methods is available online in `Matlab`, `R` and `Stata` languages.

2 The Statistical Environment

We consider the estimation of the LRV $J \triangleq \lim_{T\to\infty} J_T$ where $J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(V_s(\beta_0)V_t(\beta_0)')$ with $V_t(\beta)$ being a random p-vector for each $\beta \in \Theta$. For example, for the linear model $V_t(\beta) = (y_t - x_t'\beta)x_t$. The classical approach for inference in the context of serially correlated data is based on consistent estimation of J. Newey and West (1987) and Andrews (1991) considered the class of kernel HAC estimators, where the subscript Cla stands for classical,

$$\widehat{J}_{\text{Cla},T} = \widehat{J}_{\text{Cla},T}(b_{1,T}) \triangleq \frac{T}{T-p}\sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\,\widehat{\Gamma}_{\text{Cla}}(k),$$

$$\text{with} \qquad \widehat{\Gamma}_{\text{Cla}}(k) \triangleq \begin{cases} T^{-1}\sum_{t=k+1}^{T}\widehat{V}_t\widehat{V}_{t-k}', & k \geq 0 \\ T^{-1}\sum_{t=-k+1}^{T}\widehat{V}_{t+k}\widehat{V}_t', & k < 0, \end{cases}$$

$\widehat{V}_t = V_t(\widehat{\beta})$, $K_1(\cdot)$ is a real-valued kernel in the class $\boldsymbol{K}_1$ defined below and $b_{1,T}$ is a bandwidth sequence. The factor $T/(T-p)$ is an optional small-sample degrees of freedom adjustment. For the Newey-West estimator K_1 corresponds to the Bartlett kernel while for Andrews' (1991) K_1 corresponds to the quadratic spectral (QS) kernel. Data-dependent methods for the selection of $b_{1,T}$ were proposed by Newey and West (1994) and Andrews (1991), respectively. Under appropriate conditions on $b_{1,T} \to 0$ they showed that $\widehat{J}_{\text{Cla},T} \xrightarrow{\mathbb{P}} J$. When $\{V_t\}$ is second-order stationary, $J = 2\pi f(0)$ where $f(0)$ is the spectral density of $\{V_t\}$ at frequency zero. Most of the LRV estimation literature has focused on the stationarity assumption for $\{V_t\}$ [e.g., Kiefer et al. (2000), Müller (2007) and Lazarus et al. (2020)]. Unlike the HAC estimators, fixed-b (and versions thereof) LRV estimators require stationarity of $\{V_t\}$. The latter assumption is restrictive for economic and financial time series. The properties of J under nonstationarity were studied recently by Casini (2021) who showed that if $\{V_t\}$ is either locally stationary or segmented locally stationary (SLS), then $J = 2\pi \int_0^1 f(u, 0)\, du$ where $f(u, 0)$ is the time-varying spectral density at rescaled time $u = t/T$ and frequency zero. For locally stationary processes, $f(u, 0)$ is smooth in u while for SLS processes $f(u, 0)$ can in addition contain a finite-number of discontinuities. The number of discontinuities can actually grow to infinity with unchanged results though at the expense of slightly more complex derivations. Since the assumption of a finite number of discontinuities capture well the idea that a finite number of regimes or structural breaks is enough to account for structural changes (or big events) in economic time series we maintain this assumption here. The latter is relaxed by Casini and Perron (2021c).

Under nonstationarity Casini (2021) argued that an extension of the classical HAC estimators can actually account flexibly for the time-varying properties of the data. He proposed the class of double kernel HAC (DK-HAC) estimators,

$$\widehat{J}_T = \widehat{J}_T(b_{1,T}, b_{2,T}) \triangleq \frac{T}{T-p} \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\, \widehat{\Gamma}(k), \quad \text{with}$$
$$\widehat{\Gamma}(k) \triangleq \frac{n_T}{T-n_T} \sum_{r=0}^{\lfloor (T-n_T)/n_T \rfloor} \widehat{c}_T(rn_T/T, k),$$

where $n_T \to \infty$ satisfies the conditions given below, and

$$\widehat{c}_T(rn_T/T, k) \triangleq \begin{cases} (Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2^* \left(\frac{((r+1)n_T - (s-k/2))/T}{b_{2,T}} \right) \widehat{V}_s \widehat{V}'_{s-k}, & k \geq 0 \\ (Tb_{2,T})^{-1} \sum_{s=-k+1}^{T} K_2^* \left(\frac{((r+1)n_T - (s+k/2))/T}{b_{2,T}} \right) \widehat{V}_{s+k} \widehat{V}'_s, & k < 0 \end{cases}, \tag{2.1}$$

with K_2^* being a real-valued kernel and $b_{2,T}$ is a bandwidth sequence. $\widehat{c}_T(u, k)$ is an estimate of the local autocovariance $c(u, k) = \mathbb{E}(V_{\lfloor Tu \rfloor}, V'_{\lfloor Tu \rfloor - k}) + O(T^{-1})$ [under regularity conditions; see

Casini (2021)] at lag k and time $u = rn_T/T$. $\widehat{\Gamma}(k)$ estimates the local autocovariance across blocks of length n_T and then takes an average over the blocks. The estimator $\widehat{J}_T$ involves two kernels: K_1 smooths the lagged autocovariances—akin to the classical HAC estimators—while K_2 applies smoothing over time. The smoothing over time better account for nonstationarity and makes $\widehat{J}_{\mathrm{DK},T}$ robust to low frequency contamination. See Casini et al. (2021) who showed theoretically that existing LRV estimators are contaminated by nonstationarity so that they become inflated with consequent large power losses when the estimators are used to normalize HAR test statistics.

Casini (2021) considered adaptive estimators $\widehat{J}_{\mathrm{DK},T}$ for which $b_{1,T}$ and $b_{2,T}$ are data-dependent. Observe that the optimal $b_{2,T}$ actually depends on the properties of $\{V_{t,T}\}$ in any given block [i.e., $b_{2,T} = b_{2,T}(t/T)$]. Let

$$\begin{aligned}&\mathrm{MSE}\left(b_{2,T}^{-4},\, \widehat{c}_T(u_0,\, k,\,),\, \widetilde{W}_T\right)\\ &\quad\triangleq b_{2,T}^{-4}\mathbb{E}\left[\mathrm{vec}\left(\widehat{c}_T(u_0,\, k) - c(u_0,\, k)\right)\right]' \widetilde{W}_T\left[\mathrm{vec}\left(\widehat{c}_T(u_0,\, k) - c(u_0,\, k)\right)\right],\end{aligned}$$

where $\widetilde{W}_T$ is some $p \times p$ positive semidefinite matrix. He considered a sequential MSE criterion to determine the optimal kernels and bandwidths. For K_1, the result states that the QS kernel minimizes the asymptotic MSE for any $K_2(\cdot)$. The optimal $b_{1,T}^{\mathrm{opt}}$ and $b_{2,T}^{\mathrm{opt}}$ satisfy the following,

$$\begin{aligned}&\mathrm{MSE}\left(Tb_{1,T}^{\mathrm{opt}}\overline{b}_{2,T}^{\mathrm{opt}},\, \widehat{J}_T\left(b_{1,T}^{\mathrm{opt}},\, \overline{b}_{2,T}^{\mathrm{opt}}\right),\, W_T\right) \leq \mathrm{MSE}\left(Tb_{1,T}^{\mathrm{opt}}\overline{b}_{2,T}^{\mathrm{opt}},\, \widehat{J}_T\left(b_{1,T},\, \overline{b}_{2,T}^{\mathrm{opt}}\right),\, W_T\right) \qquad (2.2)\\ &\text{where } \overline{b}_{2,T}^{\mathrm{opt}} = \int_0^1 b_{2,T}^{\mathrm{opt}}(u)\, du\\ &\text{and } b_{2,T}^{\mathrm{opt}}(u) = \underset{b_{2,T}}{\mathrm{argmin}}\mathrm{MSE}\left(b_{2,T}^{-4},\, \widehat{c}_T(u_0,\, k) - c(u_0,\, k),\, \widetilde{W}_T\right).\end{aligned}$$

$\widehat{J}_T(b_{1,T},\, \overline{b}_{2,T}^{\mathrm{opt}})$ indicates the estimator $\widehat{J}_T$ that uses $b_{1,T}$ and $\overline{b}_{2,T}^{\mathrm{opt}}$. Eq. (2.2) holds as $T \to \infty$. The above criterion determines the globally optimal $b_{1,T}^{\mathrm{opt}}$ given the integrated locally optimal $b_{2,T}^{\mathrm{opt}}(u)$. Under (2.2), only $b_{2,T}$ affects $b_{1,T}$ but not vice-versa. Intuitively, this is a limitation because it is likely that in order to minimize the global MSE the bandwidths $b_{1,T}$ and $b_{2,T}$ affect each other.

In this paper, we consider a more theoretically appealing criterion to determine the optimal bandwidths. That is, we consider bandwidths $(\widetilde{b}_{1,T}^{\mathrm{opt}},\, \widetilde{b}_{2,T}^{\mathrm{opt}})$ that jointly minimize the global asymptotic relative MSE, denoted by ReMSE,

$$\begin{aligned}&\lim_{T\to\infty}\mathrm{ReMSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T(b_{1,T},\, b_{2,T})\, J^{-1},\, W_T\right), \qquad (2.3)\\ &\qquad = \lim_{T\to\infty} Tb_{1,T}b_{2,T}\mathbb{E}\left(\mathrm{vec}\left(\widehat{J}_T J^{-1} - I_p\right)' W_T \mathrm{vec}\left(\widehat{J}_T J^{-1} - I_p\right)\right),\end{aligned}$$

where W_T is $p^2 \times p^2$ weight matrix. Under (2.3), $\widetilde{b}_{1,T}^{\mathrm{opt}}$ and $\widetilde{b}_{2,T}^{\mathrm{opt}}$ affect each other simultaneously. This

is a more reasonable property. In Section 3 we solve for the sequences $(\widetilde{b}_{1,T}^{\text{opt}}, \widetilde{b}_{2,T}^{\text{opt}})$ that minimize (2.3). We propose a data-dependent method for $(\widetilde{b}_{1,T}^{\text{opt}}, \widetilde{b}_{2,T}^{\text{opt}})$ in Section 4.

The literature on LRV estimation has routinely focused on the case where $\widehat{V}_t$ is a function of a parameter estimate $\widehat{\beta}$ that enjoys a standard $\sqrt{T}$ parametric rate of convergence. While this is an important case, the recent increasing use of nonparametric methods suggests that the case where $\widehat{\beta}$ enjoys a nonparametric rate of convergence slower than $\sqrt{T}$ is of potential interest. Hence, in Section 5 we consider consistent LRV estimation under the latter framework and develop corresponding results for the classical HAC as well as the DK-HAC estimators.

We consider the following standard classes of kernels [cf. Andrews (1991)],

$$\boldsymbol{K}_1 = \{K_1(\cdot) : \mathbb{R} \to [-1, 1] : K_1(0) = 1, K_1(x) = K_1(-x), \forall x \in \mathbb{R} \tag{2.4}$$
$$\int_{-\infty}^{\infty} K_1^2(x)\, dx < \infty, K_1(\cdot) \text{ is continuous at 0 and at all but finite numbers of points}\Big\}.$$

$$\boldsymbol{K}_2 = \Big\{K_2(\cdot) : \mathbb{R} \to [0, \infty] : K_2(x) = K_2(1-x), \int K_2(x)\, dx = 1, \tag{2.5}$$
$$K_2(x) = 0, \text{ for } x \notin [0, 1], K_2(\cdot) \text{ is continuous}\Big\}.$$

The class $\boldsymbol{K}_1$ was also considered by Andrews (1991). Examples of kernels in $\boldsymbol{K}_1$ include the Truncated, Bartlett, Parzen, Quadratic Spectral (QS) and Tukey-Hanning kernels. The QS kernel was shown to be optimal for $\widehat{J}_{\text{Cla},T}$ under the MSE criterion by Andrews (1991) and for $\widehat{J}_T$ under a sequential MSE criterion by Casini (2021),

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

The class $\boldsymbol{K}_2$ was also considered by, for example, Dahlhaus and Giraitis (1998).

Throughout we adopt the following notational conventions. The jth element of a vector x is indicated by $x^{(j)}$ while the (j, l)th element of a matrix X is indicated as $X^{(j,l)}$. $\text{tr}(\cdot)$ denotes the trace function and $\otimes$ denotes the tensor (or Kronecker) product operator. The $p^2 \times p^2$ matrix C_{pp} is a commutation matrix that transforms $\text{vec}(A)$ into $\text{vec}(A')$, i.e., $C_{pp} = \sum_{j=1}^{p}\sum_{l=1}^{p} \iota_j \iota_l' \otimes \iota_l \iota_j'$, where ι_j is the jth elementary p-vector. $\lambda_{\max}(A)$ denotes the largest eigenvalue of the matrix A. W and $\widetilde{W}$ are used for $p^2 \times p^2$ weight matrices. $\mathbb{C}$ is used for the set of complex numbers and $\overline{A}$ for the complex conjugate of $A \in \mathbb{C}$. Let $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m < \lambda_{m+1} = 1$. A function $G(\cdot, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C}$ is said to be piecewise (Lipschitz) continuous with $m+1$ segments if it is (Lipschitz) continuous within each segment. For example, it is piecewise Lipschitz continuous if for each segment $j = 1, \ldots, m+1$ it satisfies $\sup_{u \neq v} |G(u, \omega) - G(v, \omega)| \leq K|u - v|$ for any $\omega \in \mathbb{R}$ with $\lambda_{j-1} < u, v \leq \lambda_j$ for some

$K < \infty$. We define $G_j(u, \omega) = G(u, \omega)$ for $\lambda_{j-1} < u \leq \lambda_j$, so $G_j(u, \omega)$ is Lipschitz continuous for each j. If we say piecewise Lipschitz continuous with index $\vartheta > 0$, then the above inequality is replaced by $\sup_{u \neq v} |G(u, \omega) - G(v, \omega)| \leq K |u - v|^{\vartheta}$. 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 $\omega \in \mathbb{R}$. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function. The symbol "$\triangleq$" is for definitional equivalence.

3 Simultaneous Bandwidths Determination for DK-HAC Estimators

In Section 3.1 we present the consistency, rate of convergence and asymptotic MSE properties of predetermined bandwidths for the DK-HAC estimators. We use the MSE results to determine the optimal bandwidths and kernels in Section 3.2. We use the framework for nonstationarity introduced in Casini (2021). That is, we assume that $\{V_{t,T}\}$ is segmented locally stationary (SLS). Suppose $\{V_t\}_{t=1}^T$ is defined on an abstract probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space, $\mathscr{F}$ is the σ-algebra and $\mathbb{P}$ is a probability measure. We use an infill asymptotic setting and rescale the original discrete time horizon $[1, T]$ by dividing each t by T. Letting $u = t/T$ and $T \to \infty$, this defines a new time scale $u \in [0, 1]$. Let $i \triangleq \sqrt{-1}$.

Definition 3.1. A sequence of stochastic processes $\{V_{t,T}\}_{t=1}^T$ is called Segmented Locally Stationarity (SLS) with $m_0 + 1$ regimes, transfer function A^0 and trend $\mu_{\cdot}$ if there exists a representation

$$V_{t,T} = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t) A_{j,t,T}^0(\omega) \, d\xi(\omega), \qquad \left(t = T_{j-1}^0 + 1, \ldots, T_j^0\right), \tag{3.1}$$

for $j = 1, \ldots, m_0 + 1$, where by convention $T_0^0 = 0$ and $T_{m_0+1}^0 = T$ and the following holds:

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

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

where $\operatorname{cum}\{\cdot\}$ is the cumulant of rth order, $g_1 = 0$, $g_2(\omega) = 1$, $|g_r(\omega_1, \ldots, \omega_{r-1})| \leq M_r < \infty$ 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 $K > 0$ and a piecewise continuous function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0 + 1$, there exists a 2π-periodic function $A_j : (\lambda_{j-1}, \lambda_j] \times \mathbb{R} \to \mathbb{C}$

with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$, $\lambda_j^0 \triangleq T_j^0/T$ and for all T,

$$A(u, \omega) = A_j(u, \omega) \text{ for } \lambda_{j-1}^0 < u \leq \lambda_j^0, \tag{3.2}$$

$$\sup_{1\leq j\leq m_0+1} \sup_{T_{j-1}^0 < t \leq T_j^0, \omega} \left| A_{j,t,T}^0(\omega) - A_j(t/T, \omega) \right| \leq KT^{-1}. \tag{3.3}$$

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

Observe that this representation is similar to the spectral representation of stationary processes [see Anderson (1971), Brillinger (1975), Hannan (1970) and Priestley (1981) for introductory concepts]. The main difference is that $A(t/T, \omega)$ and $\mu(t/T)$ are not constant in t. Dahlhaus (1997) used the time-varying spectral representation to define the so-called locally stationary processes which are characterized, broadly speaking, by smoothness conditions on $\mu(\cdot)$ and $A(\cdot, \cdot)$. Locally stationary processes are often referred to as time-varying parameter processes [see e.g., Cai (2007) and Chen and Hong (2012)]. However, the smoothness restrictions exclude many prominent models that account for time variation in the parameters. For example, structural change and regime switching-type models do not belong to this class because parameter changes occur suddenly at a particular time. Thus, the class of SLS processes is more general and likely to be more useful. Stationarity and local stationarity are recovered as special cases of the SLS definition.

Let $\mathcal{T} \triangleq \{T_1^0, \ldots, T_{m_0}^0\}$. The spectrum of $V_{t,T}$ is defined (for fixed T) as

$$f_{j,T}(u, \omega) \triangleq \begin{cases} (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \operatorname{Cov}\left(V_{\lfloor uT-3|s|/2\rfloor,T}, V_{\lfloor uT-|s|/2\rfloor,T}\right) \exp(-i\omega s), & Tu \in \mathcal{T},\ u = T_j^0/T \\ (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \operatorname{Cov}\left(V_{\lfloor uT-s/2\rfloor,T}, V_{\lfloor uT+s/2\rfloor,T}\right) \exp(-i\omega s), & Tu \notin \mathcal{T},\ T_{j-1}^0/T < u < T_j^0/T \end{cases},$$

with $A_{1,t,T}^0(\omega) = A_1(0, \omega)$ for $t < 1$ and $A_{m+1,t,T}^0(\omega) = A_{m+1}(1, \omega)$ for $t > T$. Casini (2021) showed that $f_{j,T}(u, \omega)$ tends in mean-squared to $f_j(u, \omega) \triangleq |A_j(u, \omega)|^2$ for $T_{j-1}^0/T < u = t/T \leq T_j^0/T$, which is the spectrum that corresponds to the spectral representation. Therefore, we call $f_j(u, \omega)$ the time-varying spectral density matrix of the process. Given $f(u, \omega)$, we can define the local covariance of $V_{t,T}$ at rescaled time u with $Tu \notin \mathcal{T}$ and lag $k \in \mathbb{Z}$ as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} f(u, \omega)\, d\omega$. The same definition is also used when $Tu \in \mathcal{T}$ and $k \geq 0$. For $Tu \in \mathcal{T}$ and $k < 0$ it is defined as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} A(u, \omega) A(u - k/T, -\omega)\, d\omega$.

3.1 Asymptotic MSE Properties of DK-HAC estimators

Let $\widetilde{J}_T$ denote the pseudo-estimator identical to $\widehat{J}_T$ but based on $\{V_{t,T}\} = \{V_{t,T}(\beta_0)\}$ rather than on $\{\widehat{V}_{t,T}\} = \{V_{t,T}(\widehat{\beta})\}$.

Assumption 3.1. *(i)* $\{V_{t,T}\}$ *is a mean-zero SLS process with* m_0+1 *regimes; (ii)* $A(u,\omega)$ *is twice continuously differentiable in* u *at all* $u \neq \lambda_j^0$ $(j=1,\ldots,m_0+1)$ *with uniformly bounded derivatives* $(\partial/\partial u)\,A(u,\cdot)$ *and* $(\partial^2/\partial u^2)\,A(u,\cdot)$*, and Lipschitz continuous in the second component with index* $\vartheta=1$*; (iii)* $(\partial^2/\partial u^2)\,A(u,\cdot)$ *is Lipschitz continuous at all* $u \neq \lambda_j^0$ $(j=1,\ldots,m_0+1)$*; (iv)* $A(u,\omega)$ *is twice left-differentiable in* u *at* $u=\lambda_j^0$ $(j=1,\ldots,m_0+1)$ *with uniformly bounded derivatives* $(\partial/\partial_- u)\,A(u,\cdot)$ *and* $(\partial^2/\partial_- u^2)\,A(u,\cdot)$ *and has piecewise Lipschitz continuous derivative* $(\partial^2/\partial_- u^2)\,A(u,\cdot)$*.*

We also need to impose conditions on the temporal dependence of $V_t = V_{t,T}$. Let

$$\begin{aligned}\kappa_{V,t}^{(a,b,c,d)}(u,\,v,\,w) &\triangleq \kappa^{(a,b,c,d)}(t,\,t+u,\,t+v,\,t+w) - \kappa_{\mathscr{N}}^{(a,b,c,d)}(t,\,t+u,\,t+v,\,t+w)\\ &\triangleq \mathbb{E}\left(V_t^{(a)}V_{t+u}^{(b)}V_{t+v}^{(c)}V_{t+w}^{(d)}\right) - \mathbb{E}\left(V_{\mathscr{N},t}^{(a)}V_{\mathscr{N},t+u}^{(b)}V_{\mathscr{N},t+v}^{(c)}V_{\mathscr{N},t+w}^{(d)}\right),\end{aligned}$$

where $\{V_{\mathscr{N},t}\}$ is a Gaussian sequence with the same mean and covariance structure as $\{V_t\}$. $\kappa_{V,t}^{(a,b,c,d)}(u,\,v,\,w)$ is the time-t fourth-order cumulant of $(V_t^{(a)},\,V_{t+u}^{(b)},\,V_{t+v}^{(c)},\,V_{t+w}^{(d)})$ while $\kappa_{\mathscr{N}}^{(a,b,c,d)}(t,\,t+u,\,t+v,\,t+w)$ is the time-t centered fourth moment of V_t if V_t were Gaussian.

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

We do not require fourth-order stationarity but only that the time-$t=Tu$ fourth order cumulant is locally constant in a neighborhood of u.

Following Parzen (1957), we define $K_{1,q}\triangleq\lim_{x\downarrow 0}(1-K_1(x))/|x|^q$ for $q\in[0,\,\infty)$; q increases with the smoothness of $K_1(\cdot)$ with the largest value being such that $K_{1,q}<\infty$. When q is an even integer, $K_{1,q}=-\left(d^qK_1(x)/dx^q\right)|_{x=0}/q!$ and $K_{1,q}<\infty$ if and only if $K_1(x)$ is q times differentiable at zero. We define the index of smoothness of $f(u,\,\omega)$ at $\omega=0$ by $f^{(q)}(u,\,0)\triangleq(2\pi)^{-1}\sum_{k=-\infty}^{\infty}|k|^q\,c(u,\,k)$, for $q\in[0,\,\infty)$. If q is even, then $f^{(q)}(u,\,0)=(-1)^{q/2}\left(d^qf(u,\,\omega)/d\omega^q\right)|_{\omega=0}$. Further, $||f^{(q)}(u,\,0)||<\infty$ if and only if $f(u,\,\omega)$ is q times differentiable at $\omega=0$. We define

$$\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widetilde{J}_T,\,W\right)=Tb_{1,T}b_{2,T}\mathbb{E}\left[\mathrm{vec}\left(\widetilde{J}_T-J_T\right)'W\,\mathrm{vec}\left(\widetilde{J}_T-J_T\right)\right]. \tag{3.4}$$

Theorem 3.1. *Suppose* $K_1(\cdot) \in \boldsymbol{K}_1$, $K_2(\cdot) \in \boldsymbol{K}_2$, *Assumption 3.1-3.2 hold,* $b_{1,T}, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ *and* $1/Tb_{1,T}b_{2,T} \to 0$. *We have: (i)*

$$\lim_{T\to\infty} Tb_{1,T}b_{2,T}\mathrm{Var}\left[\mathrm{vec}\left(\widetilde{J}_T\right)\right] = 4\pi^2 \int K_1^2(y)\,dy \int_0^1 K_2^2(x)\,dx\,(I + C_{pp})\left(\int_0^1 f(u,0)\,du\right)\otimes\left(\int_0^1 f(v,0)\,dv\right).$$

(ii) If $1/Tb_{1,T}^q b_{2,T} \to 0$, $n_T/Tb_{1,T}^q \to 0$ *and* $b_{2,T}^2/b_{1,T}^q \to \nu \in (0,\infty)$ *for some* $q \in [0,\infty)$ *for which* $K_{1,q}$, $||\int_0^1 f^{(q)}(u,0)\,du|| \in [0,\infty)$ *then* $\lim_{T\to\infty} b_{1,T}^{-q}\mathbb{E}(\widetilde{J}_T - J_T) = \mathsf{B}_1 + \mathsf{B}_2$ *where* $\mathsf{B}_1 = -2\pi K_{1,q}\int_0^1 f^{(q)}(u,0)\,du$ *and* $\mathsf{B}_2 = 2^{-1}\nu\int_0^1 x^2 K_2(x)\sum_{k=-\infty}^{\infty}\int_0^1 (\partial^2/\partial u^2)\,c(u,k)\,du$.

(iii) If $n_T/Tb_{1,T}^q \to 0$, $b_{2,T}^2/b_{1,T}^q \to \nu$ *and* $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0,\infty)$ *for some* $q \in [0,\infty)$ *for which* $K_{1,q}$, $||\int_0^1 f^{(q)}(u,0)\,du|| \in [0,\infty)$ *, then*

$$\lim_{T\to\infty}\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widetilde{J}_T,\,W\right) = 4\pi^2\left[\gamma\left(4\pi^2\right)^{-1}\mathrm{vec}\left(\mathsf{B}_1+\mathsf{B}_2\right)'W\mathrm{vec}\left(\mathsf{B}_1+\mathsf{B}_2\right) + \int K_1^2(y)\,dy\int K_2^2(x)\,dx\,\mathrm{tr}W\left(I_{p^2}+C_{pp}\right)\left(\int_0^1 f(u,0)\,du\right)\otimes\left(\int_0^1 f(v,0)\,dv\right)\right].$$

The bias expression in part (ii) of Theorem 3.1 is different from the corresponding one in Casini (2021) because $b_{2,T}^2/b_{1,T}^q \to \nu \in (0,\infty)$ replaces $b_{2,T}^2/b_{1,T}^q \to 0$ there. The extra term is B_2. This means that both $b_{1,T}$ and $b_{2,T}$ affect the bias as well as the variance. It is therefore possible to consider a joint minimization of the asymptotic MSE with respect to $b_{1,T}$ and $b_{2,T}$. Note that $\mathsf{B}_2 = 0$ when $\int_0^1 (\partial^2/\partial^2 u)\,c(u,k)\,du = 0$. The latter occurs when the process is stationary. We now move to the results concerning $\widehat{J}_T$.

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

Assumption 3.3(i)-(iii) is the same as Assumption B in Andrews (1991). Part (i) is satisfied by standard (semi)parametric estimators. In Section 5 we relax this assumption and consider nonparametric estimators that satisfy $T^{\vartheta}(\widehat{\beta}-\beta_0) = O_{\mathbb{P}}(1)$ where $\vartheta \in (0,1/2)$. In order to obtain rate of convergence results we replace Assumption 3.2 with the following assumptions.

Assumption 3.4. *(i) Assumption 3.2 holds with* $V_{t,T}$ *replaced by*

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

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

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

Theorem 3.2. *Suppose $K_1(\cdot) \in \boldsymbol{K}_1$, $K_2(\cdot) \in \boldsymbol{K}_2$, $b_{1,T}$, $b_{2,T} \to 0$, $n_T \to \infty$, $n_T/Tb_{1,T} \to 0$, and $1/Tb_{1,T}b_{2,T} \to 0$. We have:*

(i) If Assumption 3.1-3.3 hold, $\sqrt{T}b_{1,T} \to \infty$, $b_{2,T}/b_{1,T} \to \nu \in [0, \infty)$ then $\widehat{J}_T - J_T \xrightarrow{\mathbb{P}} 0$ and $\widehat{J}_T - \widetilde{J}_T \xrightarrow{\mathbb{P}} 0$.

(ii) If Assumption 3.1, 3.3-3.4 hold, $n_T/Tb_{1,T}^q \to 0$, $1/Tb_{1,T}^q b_{2,T} \to 0$, $b_{2,T}^2/b_{1,T}^q \to \nu \in [0, \infty)$ and $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0, \infty)$ for some $q \in [0, \infty)$ for which $K_{1,q}$, $\|\int_0^1 f^{(q)}(u, 0)\, du\| \in [0, \infty)$, then $\sqrt{Tb_{1,T}b_{2,T}}(\widehat{J}_T - J_T) = O_{\mathbb{P}}(1)$ and $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$.

(iii) Under the conditions of part (ii) with $\nu \in (0, \infty)$ and Assumption 3.5,

$$\lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T,\, W_T\right) = \lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widetilde{J}_T,\, W\right).$$

Part (ii) yields the consistency of $\widehat{J}_T$ with $b_{1,T}$ only required to be $o(Tb_{2,T})$. This rate is slower than the corresponding rate $o(T)$ of the classical kernel HAC estimators as shown by Andrews (1991) in his Theorem 1-(b). However, this property is of little practical import because optimal growth rates typically are less than $T^{1/2}$—for the QS kernel the optimal growth rate is $T^{1/5}$ while it is $T^{1/3}$ for the Barteltt. Part (ii) of the theorem presents the rate of convergence of $\widehat{J}_T$ which is $\sqrt{Tb_{2,T}b_{1,T}}$, the same rate shown by Casini (2021) when $b_{2,T}^2/b_{1,T}^q \to 0$. Thus, the presence of the bias term B_2 does not alter the rate of convergence. In Section 3.2, we compare the rate of convergence of $\widehat{J}_T$ with optimal bandwidths $(\widetilde{b}_{1,T}^{\mathrm{opt}}, \widetilde{b}_{2,T}^{\mathrm{opt}})$ from the joint MSE criterion (2.3) with that using $(b_{1,T}^{\mathrm{opt}}, b_{2,T}^{\mathrm{opt}})$ from the sequential MSE criterion (2.2), and with that of the classical HAC estimators when the corresponding optimal bandwidths are used.

3.2 Optimal Bandwidths and Kernels

We consider the optimal bandwidths $\left(\widetilde{b}_{1,T}^{\mathrm{opt}}, \widetilde{b}_{2,T}^{\mathrm{opt}}\right)$ and kernels $\widetilde{K}_1^{\mathrm{opt}}$ and $\widetilde{K}_2^{\mathrm{opt}}$ that minimize the global asymptotic relative MSE (2.3) given by

$$\begin{aligned}\lim_{T\to\infty} &\mathrm{ReMSE}\left(\widehat{J}_T\left(b_{1,T},\, b_{2,T}\right) J^{-1},\, W_T\right), \\ &= \mathbb{E}\left(\mathrm{vec}\left(\widehat{J}_T J^{-1} - I_p\right)' W_T \mathrm{vec}\left(\widehat{J}_T J^{-1} - I_p\right)\right).\end{aligned}$$

Let $\Xi_{1,1} = -K_{1,q}$, $\Xi_{1,2} = (4\pi)^{-1}\int_0^1 x^2 K_2(x)\, dx$, $\Xi_2 = \int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx$,

$$\Delta_{1,1,0} \triangleq \int_0^1 f^{(q)}(u,\,0)\,du \left(\int_0^1 f(u,\,0)\,du\right)^{-1} \qquad \text{and}$$
$$\Delta_{1,2} \triangleq \sum_{k=-\infty}^{\infty} \int_0^1 \left(\partial^2/\partial u^2\right) c(u,\,k)\,du \left(\int_0^1 f(u,\,0)\,du\right)^{-1}.$$

Theorem 3.3. *Suppose Assumption 3.1, 3.3-3.5 hold,* $\int_0^1 ||f^{(2)}(u,\,0)||du < \infty$, $\mathrm{vec}(\Delta_{1,1,0})' W \mathrm{vec}(\Delta_{1,1,0}) > 0$, $\mathrm{vec}(\Delta_{1,2})' W \mathrm{vec}(\Delta_{1,2}) > 0$ *and* W *is positive definite. Then,* $\lim_{T\to\infty} \mathrm{ReMSE}\left(\widehat{J}_T(b_{1,T},\,b_{2,T})\,J^{-1},\,W_T\right)$ *is jointly minimized by*

$$\widetilde{b}_{1,T}^{\mathrm{opt}} = 0.46 \left(\frac{\mathrm{vec}(\Delta_{1,2})' W \mathrm{vec}(\Delta_{1,2})}{\left(\mathrm{vec}(\Delta_{1,1,0})' W \mathrm{vec}(\Delta_{1,1,0})\right)^5}\right)^{1/24} T^{-1/6},$$
$$\widetilde{b}_{2,T}^{\mathrm{opt}} = 3.56 \left(\frac{\mathrm{vec}(\Delta_{1,1,0})' W \mathrm{vec}(\Delta_{1,1,0})}{\left(\mathrm{vec}(\Delta_{1,2})' W \mathrm{vec}(\Delta_{1,2})\right)^5}\right)^{1/24} T^{-1/6}.$$

Furthermore, the optimal kernels are given by $K_1^{\mathrm{opt}} = K_1^{\mathrm{QS}}$ *and* $K_2^{\mathrm{opt}}(x) = 6x(1-x)$ *for* $x \in [0,\,1]$.

The requirement $\int_0^1 ||f^{(2)}(u,\,0)||du < \infty$ is not stringent and reduces to the one used by Andrews (1991) when $\{V_{t,T}\}$ is stationary. Note that $\Delta_{1,1,0}$ accounts for the relative variation of $\int_0^1 f(u,\,\omega)$ around $\omega = 0$ whereas $\Delta_{1,2}$ accounts for the relative time variation (i.e., nonstationarity). The theorem states that as $\Delta_{1,1,0}$ increases $\widetilde{b}_{1,T}^{\mathrm{opt}}$ becomes smaller while $\widetilde{b}_{2,T}^{\mathrm{opt}}$ becomes larger. This is intuitive. With more variation around the zero frequency, more smoothing is required over the frequency direction and less over the time direction. Conversely, the more nonstationary is the data the more smoothing is required over the time direction (i.e., $\widetilde{b}_{2,T}^{\mathrm{opt}}$ is smaller and the optimal block length $T\widetilde{b}_{2,T}^{\mathrm{opt}}$ smaller) relative to the frequency direction. Both optimal bandwidths $(\widetilde{b}_{1,T}^{\mathrm{opt}},\,\widetilde{b}_{2,T}^{\mathrm{opt}})$ have the same order $O(T^{-1/6})$. We can compare it with $b_{1,T}^{\mathrm{opt}} = O(T^{-4/25})$ and $\overline{b}_{2,T}^{\mathrm{opt}} = O(T^{-1/5})$ resulting from the sequential MSE criterion in Casini (2021). The latter leads to a slightly smaller block length relative to the global criterion (2.3) [i.e., $O(T\overline{b}_{2,T}^{\mathrm{opt}}) < O(T\widetilde{b}_{2,T}^{\mathrm{opt}})$]. Since K_2 applies overlapping smoothing, a smaller block length is beneficial if there is substantial nonstationarity. On the same note, a smaller block length is less exposed to low frequency contamination since it allows to better account for nonstationarity. The rate of convergence when the optimal bandwidths are used is $O(T^{1/3})$ which is sightly faster than the corresponding rate of convergence when $(b_{1,T}^{\mathrm{opt}},\,b_{2,T}^{\mathrm{opt}})$. The latter is $O(T^{0.32})$, so the difference is small.

4 Data-Dependent Bandwidths

In this section we consider estimators $\widehat{J}_T$ that use bandwidths $b_{1,T}$ and $b_{2,T}$ whose values are determined via data-dependent methods. We use the "plug-in" method which is characterized by plugging-in estimates of unknown quantities into a formula for an optimal bandwidth parameter (i.e., the expressions for $\widetilde{b}_{1,T}^{\text{opt}}$ and $\widetilde{b}_{2,T}^{\text{opt}}$). Section 4.1 discusses the implementation of the automatic bandwidths, while Section 4.2 presents the corresponding theoretical results.

4.1 Implementation

The first step for the construction of data-dependent bandwidth parameters is to specify p univariate parametric models for $V_t = (V_t^{(1)}, \ldots, V_t^{(p)})$. The second step involves the estimation of the parameters of the parametric models. Here standard estimation methods are local least-squares (LS) (i.e., LS method applied to rolling windows) and nonparametric kernel methods. Let

$$\phi_1 \triangleq \frac{\text{vec}\left(\Delta_{1,2}\right)' W \text{vec}\left(\Delta_{1,2}\right)}{\left(\text{vec}\left(\Delta_{1,1,0}\right)' W \text{vec}\left(\Delta_{1,1,0}\right)\right)^5}, \qquad \phi_2 \triangleq \frac{\text{vec}\left(\Delta_{1,1,0}\right)' W \text{vec}\left(\Delta_{1,1,0}\right)}{\left(\text{vec}\left(\Delta_{1,2}\right)' W \text{vec}\left(\Delta_{1,2}\right)\right)^5}.$$

In a third step, we replace the unknown parameters in ϕ_1 and ϕ_2 with corresponding estimates. Such estimates $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are then substituted into the expression for $\widetilde{b}_{1,T}^{\text{opt}}$ and $\widetilde{b}_{2,T}^{\text{opt}}$ to yield

$$\widehat{b}_{1,T} = 0.46\phi_1^{1/24} T^{-1/6}, \qquad \widehat{b}_{2,T} = 3.56\phi_2^{1/24} T^{-1/6}. \tag{4.1}$$

In practice, a reasonable candidate to be used as an approximating parametric model is the first order autoregressive [AR(l)] model for $\{V_t^{(r)}\}$, $r = 1, \ldots, p$ (with different parameters for each r) or a first order vector autoregressive [VAR(l)] model for $\{V_t\}$ [see Andrews (1991)]. However, in our context it is reasonable to allow the parameters to be time-varying. For parsimony, we consider time-varying AR(1) models with no break points in the spectrum (i.e., $V_t^{(r)} = a_1\left(t/T\right) V_{t-1}^{(r)} + u_t^{(r)}$).

The use of p univariate parametric models requires W to be a diagonal matrix. This leads to $\phi_1 = \phi_{1,1}/\phi_{1,2}^5$ and $\phi_2 = \phi_{1,2}/\phi_{1,1}^5$ where

$$\phi_{1,1} = \sum_{r=1}^{p} W^{(r,r)} \left(\sum_{k=-\infty}^{\infty} \int_0^1 \frac{\partial^2}{\partial u^2} c^{(r,r)}\left(u,\, k\right) du \right)^2 \Big/ \left(\int_0^1 f^{(r,r)}\left(u,\, 0\right) du \right)^2.$$

$$\phi_{1,2} = \sum_{r=1}^{p} W^{(r,r)} \left(\int_0^1 f^{(q)(r,r)}\left(u,\, 0\right) du \right)^2 \Big/ \left(\int_0^1 f^{(r,r)}\left(u,\, 0\right) du \right)^2.$$

The usual choice is $W^{(r,r)} = 1$ for all r. An estimate of $f^{(r,r)}\left(u,\, 0\right)$ $(r = 1, \ldots, p)$ is $\widehat{f}^{(r,r)}\left(u,\, 0\right) =$

$(2\pi)^{-1}(\widehat{\sigma}^{(r)}(u))^2(1-\widehat{a}_1^{(r)}(u))^{-2}$ while $f^{(2)(r,r)}(u,0)$ can be estimated by $\widehat{f}^{(2)(r,r)}(u,0) = 3\pi^{-1}((\widehat{\sigma}^{(r)}(u))^2\widehat{a}_1^{(r)}(u))(1-\widehat{a}_1^{(r)}(u))^{-4}$ where $\widehat{a}_1^{(r)}(u)$ and $\widehat{\sigma}^{(r)}(u)$ are the LS estimates computed using local data to the left of $u = t/T$:

$$\widehat{a}_1^{(r)}(u) = \frac{\sum_{j=t-n_{2,T}+1}^{t}\widehat{V}_j^{(r)}\widehat{V}_{j-1}^{(r)}}{\sum_{j=t-n_{2,T}+1}^{t}\left(\widehat{V}_{j-1}^{(r)}\right)^2}, \qquad \widehat{\sigma}^{(r)}(u) = \left(\sum_{j=t-n_{2,T}+1}^{t}\left(\widehat{V}_j^{(r)} - \widehat{a}_1^{(r)}(u)\,\widehat{V}_{j-1}^{(r)}\right)^2\right)^{1/2}, \tag{4.2}$$

where $n_{2,T} \to \infty$. More complex is the estimation of $\overline{\Delta}_{1,2,1} \triangleq \sum_{k=-\infty}^{\infty}\int_0^1 (\partial^2/\partial u^2)\, c(u,k)$ because it involves the second partial derivative of $c(u,k)$. We need a further parametric assumption. We assume that the parameters of the approximating time-varying AR(1) models change slowly such that the smoothness of $f(\cdot,\omega)$ and thus of $c(\cdot,\cdot)$ is the same to the one that would arise if $a_1(u) = 0.8(\cos 1.5 + \cos 4\pi u)$ and $\sigma(u) = \sigma = 1$ for all $u \in [0,1]$ [cf. Dahlhaus (2012)]. Then, $\Delta_{1,2,1}^{(r,r)}(u,k) \triangleq (\partial^2/\partial u^2)\, c^{(r,r)}(u,k)$ can be computed analytically:

$$\begin{aligned}\Delta_{1,2,1}^{(r,r)}(u,k) = \int_{-\pi}^{\pi} e^{ik\omega}\Big[&\frac{3}{\pi}\left(1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega)\right)^{-4}\left(0.8\left(-4\pi\sin(4\pi u)\right)\right)\exp(-i\omega)\\ &-\frac{1}{\pi}\left|1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega)\right|^{-3}\left(0.8\left(-16\pi^2\cos(4\pi u)\right)\right)\exp(-i\omega)\Big]\,d\omega).\end{aligned}$$

An estimate of $\Delta_{1,2,1}^{(r,r)}(u,k)$ is given by

$$\begin{aligned}&\widehat{\Delta}_{1,2,1}^{(r,r)}(u,k) \triangleq\\ &\quad[S_\omega]^{-1}\sum_{s\in S_\omega} e^{ik\omega_s}\Big[\frac{3}{\pi}\left(1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega_s)\right)^{-4}\left(0.8\left(-4\pi\sin(4\pi u)\right)\right)\exp(-i\omega_s)\\ &\quad-\frac{1}{\pi}\left|1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega_s)\right|^{-3}\left(0.8\left(-16\pi^2\cos(4\pi u)\right)\right)\exp(-i\omega_s)\Big],\end{aligned}$$

where $[S_\omega]$ is the cardinality of S_ω and $\omega_{s+1} > \omega_s$ with $\omega_1 = -\pi$, $\omega_{[S_\omega]} = \pi$. In our simulations, we use $S_\omega = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\}$. We can average over u and sum over k to obtain an estimate of $\overline{\Delta}_{1,2,1}^{(r,r)}$: $\widehat{\overline{\Delta}}_{1,2,1}^{(r,r)} = \sum_{k=-\lfloor T^{1/6}\rfloor}^{\lfloor T^{1/6}\rfloor}\frac{n_{3,T}}{T}\sum_{j=0}^{\lfloor T/n_{3,T}\rfloor}\widehat{\Delta}_{1,2,1}^{(r,r)}(jn_T/T,k)$ where the number of summands over k grows at the same rate as $1/\widetilde{b}_{1,T}^{\text{opt}}$; a different choice is allowed as long as it grows at a slower rate than $T^{2/5}$ but our sensitivity analysis does not indicate significant changes.

Then, $\widehat{b}_{1,T} = 0.46\widehat{\phi}_1^{1/24}T^{-1/6}$ and $\widehat{b}_{2,T} = 3.56\widehat{\phi}_2^{1/24}T^{-1/6}$ where $\widehat{\phi}_1 = \widehat{\phi}_{1,1}/\widehat{\phi}_{1,2}^5$, $\widehat{\phi}_2 = \widehat{\phi}_{1,2}/\widehat{\phi}_{1,1}^5$,

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

$$\widehat{\phi}_{1,2} = 36 \sum_{r=1}^{p} W^{(r,r)} \left(\frac{\frac{n_{3,T}}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} ((\widehat{\sigma}^{(r)}(jn_{3,T}+1))^2 \widehat{a}_1^{(r)}(jn_{3,T}+1))(1-\widehat{a}_1^{(r)}(jn_{3,T}+1))^{-4}}{\frac{n_{3,T}}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} (\widehat{\sigma}^{(r)}(jn_{3,T}+1))^2 (1-\widehat{a}_1^{(r)}(jn_{3,T}+1))^{-2}} \right)^2.$$

For most of the results below we can take $n_{3,T} = n_{2,T} = n_T$.

4.2 Theoretical Results

We establish results corresponding to Theorem 3.2 for the estimator $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$ that uses $\widehat{b}_{1,T}$ and $\widehat{b}_{2,T}$. We restrict the class of admissible kernels to the following,

$$\begin{aligned}\boldsymbol{K}_3 = \Big\{ & K_3(\cdot) \in \boldsymbol{K}_1 : (i)\ |K_1(x)| \le C_1 |x|^{-b} \text{ with } b > \max(1+1/q, 3) \text{ for } |x| \in [\overline{x}_L, D_T h_T \overline{x}_U], \\ & b_{1,T}^2 h_T \to \infty,\ D_T > 0,\ \overline{x}_L, \overline{x}_U \in \mathbb{R},\ 1 \le \overline{x}_L < \overline{x}_U, \text{ and} \\ & \text{with } b > 1 + 1/q \text{ for } |x| \notin [\overline{x}_L, D_T h_T \overline{x}_U], \text{ and some } C_1 < \infty, \\ & \text{where } q \in (0, \infty) \text{ is such that } K_{1,q} \in (0, \infty),\ (ii)\ |K_1(x) - K_1(y)| \le C_2 |x-y|\ \forall x, \\ & y \in \mathbb{R} \text{ for some costant } C_2 < \infty \Big\}.\end{aligned}$$

Let $\widehat{\theta}$ denote the estimator of the parameter of the approximate (time-varying) parametric model(s) introduced above [i.e., $\widehat{\theta} = (\int_0^1 \widehat{a}_1(u)\,du, \int_0^1 \widehat{\sigma}_1^2(u)\,du, \ldots, \int_0^1 \widehat{a}_p^2(u)\,du, \int_0^1 \widehat{\sigma}_p^2(u)\,du)'$]. Let θ^* denote the probability limit of $\widehat{\theta}$. $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are the values of ϕ_1 and ϕ_2, receptively, with $\widehat{\theta}$ instead of θ. The probability limits of $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are denoted by ϕ_{1,θ^*} and ϕ_{2,θ^*}, respectively.

Assumption 4.1. *(i)* $\widehat{\phi}_1 = O_{\mathbb{P}}(1)$, $1/\widehat{\phi}_1 = O_{\mathbb{P}}(1)$, $\widehat{\phi}_2 = O_{\mathbb{P}}(1)$, *and* $1/\widehat{\phi}_2 = O_{\mathbb{P}}(1)$*; (ii)* $\inf\left\{T/n_{3,T}, \sqrt{n_{2,T}}\right\}((\widehat{\phi}_1 - \phi_{1,\theta^*}), (\widehat{\phi}_2 - \phi_{2,\theta^*}))' = O_{\mathbb{P}}(1)$ *for some* $\phi_{1,\theta^*}, \phi_{2,\theta^*} \in (0, \infty)$ *where* $n_{2,T}/T + n_{3,T}/T \to 0$, $n_{2,T}^{10/6}/T \to [c_2, \infty)$, $n_{3,T}^{10/6}/T \to [c_3, \infty)$ *with* $0 < c_2, c_3 < \infty$*; (iii)* $\sup_{u \in [0,1]} \lambda_{\max}(\Gamma_u(k)) \le C_3 k^{-l}$ *for all* $k \ge 0$ *for some* $C_3 < \infty$ *and some* $l > 3$*, where* q *is as in* $\boldsymbol{K}_3$*; (iv)* $|\omega_{s+1} - \omega_s| = O(T^{-1})$ *and* $[S_\omega] = O(T)$*; (v)* $\boldsymbol{K}_2$ *includes kernels that satisfy* $|K_2(x) - K_2(y)| \le C_4 |x-y|$ *for all* $x, y \in \mathbb{R}$ *and some constant* $C_4 < \infty$*.*

Parts (i) and (v) are sufficient for the consistency of $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$. Parts (ii)-(iii) and (iv)-(v) are required for the rate of convergence and MSE results. Note that ϕ_{1,θ^*} and ϕ_{2,θ^*} coincide with the optimal values ϕ_1 and ϕ_2, respectively, only when the approximate parametric model indexed by θ^* corresponds to the true data-generating mechanism.

Let $b_{\theta_1,T} = 0.46\phi_{1,\theta^*}^{1/24}T^{-1/6}$ and $b_{\theta_2,T} = 3.56\phi_{2,\theta^*}^{1/24}T^{-1/6}$. The asymptotic properties of $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T})$ are shown to be equivalent to those of $\widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})$ where the theoretical properties of the latter follow from Theorem 3.2.

Theorem 4.1. *Suppose $K_1(\cdot) \in \boldsymbol{K}_3$, q is as in $\boldsymbol{K}_3$, $K_2(\cdot) \in \boldsymbol{K}_2$, $n_T \to \infty$, $n_T/Tb_{\theta_1,T} \to 0$, and $||\int_0^1 f^{(q)}(u,\, 0)\, du|| < \infty$. Then, we have:*

(i) If Assumption 3.1-3.3 and 4.1-(i,v) hold and $n_{3,T} = n_{2,T} = n_T$, then $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T}) - J_T \xrightarrow{\mathbb{P}} 0$.

(ii) If Assumption 3.1, 3.3-3.4 and 4.1-(ii,iii,iv,v) hold $q \leq 2$ and $n_T/Tb_{\theta_1,T}^2 \to 0$, then $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\,(\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T}) - J_T) = O_{\mathbb{P}}(1)$ and $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})) = o_{\mathbb{P}}(1)$.

(iii) If Assumption 3.1, 3.3-3.5 and 4.1-(ii,iii,iv,v) hold, then

$$\lim_{T\to\infty} \mathrm{MSE}\left(T^{2/3},\, \widehat{J}_T\left(\widehat{b}_{1,T},\, \widehat{b}_{2,T}\right),\, W_T\right)$$
$$= \lim_{T\to\infty} \mathrm{MSE}\left(Tb_{\theta_1,T}b_{\theta_2,T},\, \widehat{J}_T\left(b_{\theta_1,T},\, b_{\theta_2,T}\right),\, W_T\right).$$

When the chosen parametric model indexed by θ is correct, it follows that $\phi_{1,\theta^*} = \phi_1$, $\phi_{2,\theta^*} = \phi_2$, $\widehat{\phi}_1 \xrightarrow{\mathbb{P}} \phi_1$ and $\widehat{\phi}_2 \xrightarrow{\mathbb{P}} \phi_2$. The theorem then implies that $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T})$ exhibits the same optimality properties presented in Theorem 3.3.

5 Consistent LRV in the Context of Nonparametric Parameter Estimates

We relax the assumption that $\{V_t(\widehat{\beta})\}$ is a function of a semiparametric estimator $\widehat{\beta}$ satisfying $\sqrt{T}(\widehat{\beta} - \beta_0) = O_{\mathbb{P}}(1)$. This holds, for example, in the linear regression model estimated by least-squares where $V_t(\widehat{\beta}) = \widehat{e}_t x_t$ with $\{\widehat{e}_t\}$ being the fitted residuals and $\{x_t\}$ being a vector of regressors. However, there are many HAR inference contexts where one needs an estimate of the LRV based on a sequence of observations $\{V_t(\widehat{\beta}_{\mathrm{np}})\}$ where $\widehat{\beta}_{\mathrm{np}}$ is a nonparametric estimator that satisfies $T^{\vartheta}(\widehat{\beta}_{\mathrm{np}} - \beta_0) = O_{\mathbb{P}}(1)$ for some $\vartheta \in (0,\, 1/2)$. For example, in forecasting one needs an estimate of the LRV to obtain a pivotal asymptotic distribution for forecast evaluation tests while one has access to a sequence $\{V_t(\widehat{\beta}_{\mathrm{np}})\}$ obtained from nonparametric estimation using some in-sample. Given that nonparametric methods have received a great deal of attention in applied work lately, it is useful to extend the theory of HAC and DK-HAC estimators to these settings. We consider the HAC estimators in Section 5.1 and the DK-HAC estimators in Section 5.2.

5.1 Classical HAC Estimators

We show that the classical HAC estimators that use the data-dependent bandwidths suggested in Andrews (1991) remain valid when $\sqrt{T}(\widehat{\beta}-\beta_0)=O_{\mathbb{P}}(1)$ is replaced by $T^{\vartheta}(\widehat{\beta}_{\text{np}}-\beta_0)=O_{\mathbb{P}}(1)$ for some $\vartheta\in(0,\,1/2)$. We work under the same assumptions as in Andrews (1991). Under stationarity we have $\Gamma_u(k)=\Gamma(k)$ and $\kappa_{V,\lfloor Tu\rfloor}^{(a,b,c,d)}(k,\,s,\,l)=\kappa_{V,0}^{(a,b,c,d)}(k,\,s,\,l)$ for any $u\in[0,\,1]$.

Assumption 5.1. *$\{V_t\}$ is a mean-zero, fourth-order stationary sequence with $\sum_{k=-\infty}^{\infty}\|\Gamma(k)\|<\infty$ and $\sum_{k=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa_{V,0}^{(a,b,c,d)}(k,\,s,\,l)|<\infty\ \forall a,\,b,\,c,\,d\leq p$.*

Assumption 5.2. *(i) $T^{\vartheta}(\widehat{\beta}_{\text{np}}-\beta_0)=O_{\mathbb{P}}(1)$ for some $\vartheta\in(0,\,1/2)$; (ii) $\sup_{t\geq1}\mathbb{E}\|V_t\|^2<\infty$; (iii) $\sup_{t\geq1}\mathbb{E}\sup_{\beta\in\Theta}\|(\partial/\partial\beta)V_t(\beta)\|^2<\infty$; (iv) $\int|K_1(y)|\,dy<\infty$.*

Assumption 5.3. *(i) Assumption 5.1 holds with V_t replaced by*

$$\left(V_t',\,\text{vec}\left(\left(\frac{\partial}{\partial\beta'}V_t(\beta_0)\right)-\mathbb{E}\left(\frac{\partial}{\partial\beta'}V_t(\beta_0)\right)\right)'\right)'.$$

(ii) $\sup_{t\geq1}\mathbb{E}(\sup_{\beta\in\Theta}||(\partial^2/\partial\beta\partial\beta')V_t^{(a)}(\beta)||^2)<\infty$ for all $a=1,\ldots,p$.

Let $\widehat{b}_{\text{Cla},1,T}=\left(qK_{1,q}^2\widehat{\alpha}(q)T/\int K_1^2(x)\,dx\right)^{-1/(2q+1)}$. The form of $\widehat{\alpha}(q)$ depends on the approximating parametric model for $\{V_t^{(r)}\}$. Andrews (1991) considered stationarity AR(1) models for $\{V_t^{(r)}\}$, which result in

$$\widehat{\alpha}(2)=\sum_{r=1}^{p}W^{(r,r)}\frac{4\left(\widehat{a}_1^{(r)}\right)^2(\widehat{\sigma}^r)^4}{\left(1-\widehat{a}_1^{(r)}\right)^8}\Big/\sum_{r=1}^{p}W^{(r,r)}\frac{(\widehat{\sigma}^r)^4}{\left(1-\widehat{a}_1^{(r)}\right)^4},\quad\text{and}\tag{5.1}$$

$$\widehat{\alpha}(1)=\sum_{r=1}^{p}W^{(r,r)}\frac{4\left(\widehat{a}_1^{(r)}\right)^2(\widehat{\sigma}^r)^4}{\left(1-\widehat{a}_1^{(r)}\right)^6\left(1+\widehat{a}_1^{(r)}\right)^2}\Big/\sum_{r=1}^{p}W^{(r,r)}\frac{(\widehat{\sigma}^r)^4}{\left(1-\widehat{a}_1^{(r)}\right)^4}.$$

Let

$$\begin{aligned}\boldsymbol{K}_{\text{Cla},3}=\Big\{&K_1(\cdot)\in\boldsymbol{K}_1:\ (i)\ |K_1(y)|\leq C_1|y|^{-b}\ \text{for some}\ b>1+1/q\ \text{and some}\ C_1<\infty,\\&\text{where}\ q\in(0,\,\infty)\ \text{is such that}\ K_{1,q}\in(0,\,\infty),\ \text{and}\ (ii)\ |K_1(x)-K_1(y)|\leq C_2|x-y|\ \forall x,\\&y\in\mathbb{R}\ \text{for some costant}\ C_2<\infty\Big\}.\end{aligned}\tag{5.2}$$

Assumption 5.4. *$\widehat{\alpha}(q)=O_{\mathbb{P}}(1)$ and $1/\widehat{\alpha}(q)=O_{\mathbb{P}}(1)$.*

Theorem 5.1. *Suppose $K_1(\cdot)\in\boldsymbol{K}_{3,\text{Cla}}$, $||f^{(q)}||<\infty$, $q>(1/\vartheta-1)/2$, and Assumption 5.1-5.4 hold, then $\widehat{J}_{\text{Cla},T}(\widehat{b}_{\text{Cla},1,T})-J_T\xrightarrow{\mathbb{P}}0$.*

5.2 DK-HAC Estimators

We extend the consistency result in Theorem 4.1-(i) assuming that $\widehat{V}_t = V_t(\widehat{\beta}_{\text{np}})$. Thus, we replace Assumption 3.3 by the following.

Assumption 5.5. *(i) $T^{\vartheta}(\widehat{\beta}_{\text{np}} - \beta_0) = O_{\mathbb{P}}(1)$ for some $\vartheta \in (0, 1/2)$; (ii)-(iv) from Assumption 3.3 continue to hold.*

Theorem 5.2. *Suppose $K_1(\cdot) \in \boldsymbol{K}_3$, $K_2(\cdot) \in \boldsymbol{K}_2$, $T^{\vartheta} b_{\theta_1,T} b_{\theta_2,T} \to \infty$, $n_T \to \infty$, $n_T/T b_{\theta_1,T} \to 0$, and $\| \int_0^1 f^{(q)}(u, 0)\, du \| < \infty$. If Assumption 3.1-3.2, 4.1-(i,v), 5.5 hold and $n_{3,T} = n_{2,T} = n_T$, then $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T}) - J_T \xrightarrow{\mathbb{P}} 0$.*

Theorem 5.1-5.2 require different conditions on the parameter ϑ that controls the rate of convergence of the nonparametric estimator. In Theorem 5.1 this conditions depends on q while in Theorem 5.2 it depends on q through $b_{\theta_1,T}$ and also on the smoothing over time through $b_{\theta_2,T}$. For both HAC and DK-HAC estimators, the condition allows for standard nonparametric estimators with optimal nonparametric convergence rate.

6 Small-Sample Evaluations

In this section, we conduct a Monte Carlo analysis to evaluate the performance of the DK-HAC estimator based on the data-dependent bandwidths determined via the joint MSE criterion (2.3). We consider HAR tests in the linear regression model as well as HAR tests for forecast breakdown, i.e., the test of Giacomini and Rossi (2009). The linear regression models have an intercept and a stochastic regressor. We focus on the t-statistics $t_r = \sqrt{T}(\widehat{\beta}^{(r)} - \beta_0^{(r)})/\sqrt{\widehat{J}_T^{(r,r)}}$ where $\widehat{J}_T$ is an estimate of the limit of $\text{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ and $r = 1, 2$. t_1 is the t-statistic for the parameter associated to the intercept while t_2 is associated to the stochastic regressor x_t. We omit the discussion of the results concerning to the F-test since they are qualitatively similar. Three basic regression models are considered. We run a t-test on the intercept in model M1 and a t-test on the coefficient of the stochastic regressor in model M2 and M3. The models are based on,

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

for the t-test on the intercept (i.e., t_1) and

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

for the t-test on $\beta_0^{(2)}$ (i.e., t_2) where $\delta = 0$ under the null. We consider the following models:

- M1: $e_t = 0.4e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0,\, 0.5)$, $x_t \sim$ i.i.d. $\mathscr{N}(1,\, 1)$, $\beta_0^{(1)} = 0$ and $\beta_0^{(2)} = 1$.
- M2: $e_t = 0.4e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0,\, 1)$, $x_t \sim$ i.i.d. $\mathscr{N}(1,\, 1)$, and $\beta_0^{(1)} = \beta_0^{(2)} = 0$.
- M3: segmented locally stationary errors $e_t = \rho_t e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0,\, 1)$, $\rho_t = \max\{0,\, -1(\cos(1.5 - \cos(5t/T)))\}$[2] for $t \notin (4T/5+1,\, 4T/5+h)$ and $e_t = 0.99e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0,\, 1)$ for $t \in (4T/5+1,\, 4T/5+h)$ where $h = 10$ for $T = 200$ and $h = 30$ for $T = 400$, and $x_t = 1 + 0.6x_{t-1} + u_{X,t}$, $u_{X,t} \sim$ i.i.d. $\mathscr{N}(0,\, 1)$.

Finally, we consider model M4 which we use to investigate the performance of Giacomini and Rossi's (2009) test for forecast breakdown. Suppose we want to forecast a variable y_t generated by $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1} + e_t$ where $x_t \sim$ i.i.d. $\mathscr{N}(1,\, 1.2)$ and $e_t = 0.3e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathscr{N}(0,\, 1)$. For a given forecast model and forecasting scheme, the test of Giacomini and Rossi (2009) detects a forecast breakdown when the average of the out-of-sample losses differs significantly from the average of the in-sample losses. The in-sample is used to obtain estimates of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ which are in turn used to construct out-of-sample forecasts $\widehat{y}_t = \widehat{\beta}_0^{(1)} + \widehat{\beta}_0^{(2)} x_{t-1}$. We set $\beta_0^{(1)} = \beta_0^{(2)} = 1$. We consider a fixed forecasting scheme. GR's (2009) test statistic is defined as $t^{\mathrm{GR}} \triangleq \sqrt{T_n \overline{SL}}/\sqrt{\widehat{J}_{SL}}$ where $\overline{SL} \triangleq T_n^{-1} \sum_{t=T_m}^{T-\tau} SL_{t+\tau}$, $SL_{t+\tau}$ is the surprise loss at time $t+\tau$ (i.e., the difference between the time $t+\tau$ out-of-sample loss and in-sample loss, $SL_{t+\tau} = L_{t+\tau} - \overline{L}_{t+\tau}$), T_n is the sample size in the out-of-sample, T_m is the sample size in the in-sample and $\widehat{J}_{SL}$ is a LRV estimator. We restrict attention to one-step ahead forecasts (i.e., $\tau = 1$). Under $H_1 : \mathbb{E}(\overline{SL}) \neq 0$, we have $y_t = 1 + x_{t-1} + \delta x_{t-1} \mathbf{1}\{t > T_1^0\} + e_t$, where $T_1^0 = T\lambda_1^0$ with $\lambda_1^0 = 0.7$. Under this specification there is a break in the coefficient associated with x_{t-1}. Thus, there is a forecast instability or failure and the test of Giacomini and Rossi (2009) should reject H_0. We set $T_m = 0.4T$ and $T_n = 0.6T$.

Throughout our study we consider the following LRV estimators: $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T})$ with K_1^{opt}, K_2^{opt} and automatic bandwidths; the same $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T})$ with in addition the prewhitening of Casini and Perron (2021c); Casini's (2021) DK-HAC with K_1^{opt}, K_2^{opt} and automatic bandwidths from the sequential method; DK-HAC with prewhitening; Newey and West's (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994); the same with the prewhitening procedure of Andrews and Monahan (1992); Newey-West estimator with the fixed-b method of Kiefer et al. (2000); the Empirical Weighted Cosine (EWC) of Lazarus et al. (2020).[3] Casini and Perron (2021c) proposed three methods related to prewhitening: (1) $\widehat{J}_{T,\mathrm{pw},1}$ uses a stationary model to whiten the data; (2) $\widehat{J}_{T,\mathrm{pw,SLS}}$ uses a nonstationary model to whiten the data; (3) $\widehat{J}_{T,\mathrm{pw,SLS},\mu}$ is the same as $\widehat{J}_{T,\mathrm{pw,SLS}}$ but it adds a time-varying intercept in the VAR to whiten the data.

[2] That is, ρ_t varies smoothly between 0 and 0.8071.

[3] We have excluded Andrews's (1991) HAC estimator since its performance is similar to that of the Newey-West estimator.

We set $n_T = T^{0.66}$ as explained in Casini (2021) and $n_{2,T} = n_{3,T} = n_T$. Simulation results for models involving ARMA, ARCH and heteroskedastic errors are not discussed here because the results are qualitatively equivalent. The significance level is $\alpha = 0.05$ throughout.

6.1 Empirical Sizes of HAR Inference Tests

Table 1-2 report the rejection rates for model M1-M4. As a general pattern, we confirm previous evidence that Newey-West's (1987) HAC estimator leads to t-tests that are oversized when the data are stationary and there is substantial dependence [cf. model M1-M2]. This is a long-discussed issue in the literature. Newey-West with prewhitening is often effective in reducing the oversize problem under stationarity. However, the simulation results below and in the literature show that the prewhitened Newey-West-based tests can be oversized when there is high serial dependence. Among the existing methods, the rejection rates of the Newey-West-based tests with fixed-b are accurate in model M1-M2. Overall, the results in the literature along with those in Casini (2021) and Casini and Perron (2021c) showed that under stationarity the original fixed-b method of KVB is the method which is in general the least oversized across different degrees of dependence among all existing methods. EWC performs similarly to KVB's fixed-b. Among the recently introduced DK-HAC estimators, Table 1 reports evidence that the non-prewhitened DK-HAC from Casini (2021) leads to HAR tests that are a bit oversized whereas the tests based on the new DK-HAC with simultaneous data-dependent bandwidths, $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, are more accurate. The results also show that the tests based on the prewhitened DK-HAC estimators are competitive with those based on KVB's fixed-b in controlling the size. In particular, tests based on $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$ with prewhitening are more accurate than those using the prewhitened DK-HAC with sequential data-dependent bandwidths. Since $\widehat{J}_{T,\text{pw},1}$ uses a stationarity VAR model to whiten the data, it works as well as $\widehat{J}_{T,\text{pw,SLS}}$ and $\widehat{J}_{T,\text{pw,SLS},\mu}$ when stationarity actually holds, as documented in Table 1.

Turning to nonstationary data and to the GR test, Table 2 casts concerns about the finite-sample performance of existing methods in this context. For both model M3 and M4, existing long-run variance estimators lead to HAR tests that have either size equal or close to zero. The methods that use long bandwidths (i.e., many lagged autocovariances) such as KVB's fixed-b and EWC suffer most from this problem relative to using the Newey-West estimator. This is demonstrated in Casini et al. (2021) who showed theoretically that nonstationarity induces positive bias for each sample autocovariance. That bias is constant across lag orders. Since existing LRV estimators are weighted sum of sample autocovariances (or weighted sum of periodogram ordinates), the more lags are included the larger is the positive bias. Thus, LRV estimators are inflated and HAR tests have lower rejection rates than the significance level. As we show below, this mechanism has consequences for power as well. In model M3-M4, tests based on the non-prewhitened DK-HAC

$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$ performs well although tests based on the prewhitened DK-HAC are more accurate. $\widehat{J}_{T,\text{pw},1}$ leads to tests that are slightly less accurate because it uses stationarity and when the latter is violated its performance is affected. In model M4, KVB's fixed-b and prewhitened DK-HAC are associated to rejection rates relatively close to the significance level.

In summary, the prewhitened DK-HAC estimators yield t-test in regression models with rejection rates that are relatively close to the exact size. The DK-HAC with simultaneous bandwidths developed in Section 4 performs better (i.e., the associated null rejection rates are closer to the significance level and approach it from below) than the corresponding DK-HAC estimators with sequential bandwidths when the data are stationary. This is in accordance with our theoretical results. Also for nonstationary data the simultaneous bandwidths perform in general better than the sequential bandwidths, though the margin is smaller. The non-prewhitened DK-HAC can lead to oversized t-test on the intercept if there is high dependence. Our results confirm the oversize problem induced by the use of the Newey-West estimators documented in the literature under stationarity. Fixed-b HAR tests control the size well when the data are stationary but can be severely undersized under nonstationarity, a problem that also affects tests based on the Newey-West. Thus, prewhitened DK-HAC estimators are competitive to fixed-b methods under stationarity and they perform well also when the data are nonstationary.

6.2 Empirical Power of HAR Inference Tests

For model M1-M4 we report the power results in Table 3-6. The sample size is $T = 200$. For model M1, tests based on the Newey and West's (1987) HAC and on the non-prewhitened DK-HAC estimators have the highest power but they were more oversized than the tests based on other methods. KVB's fixed-b leads to t-tests that sacrifices some power relative to using the prewhitened DK-HAC estimators while EWC-based tests have lower power locally to $\delta = 0$ (i.e., $\delta = 0.1$ and 0.2). In model M2, a similar pattern holds. HAR tests normalized by either classical HAC or DK-HAC estimators have similarly good power while HAR tests based on KVB's fixed-b have relatively less power. In model M3, the best power is achieved with Newey-West's (1987) HAC estimator followed by $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$ and EWC. Using KVB's fixed-b leads to large power losses. In model M4, it appears that all versions of the classical HAC estimators of Newey and West (1987), the KVB's fixed-b and EWC lead to t-tests that have, essentially, zero power for all δ. In contrast, the t-test standardized by the DK-HAC estimators have good power. Among the latter DK-HAC estimators, the ones that use the sequential bandwidths have slightly higher power but they margin is very small. This follows from the usual size-power trade-off since the simultaneous bandwidths led to tests that have more accurate size control.

The severe power problems of tests based on classical HAC estimators, KVB's fixed-b and

EWC can be simply reconciled with the fact that under the alternative hypotheses the spectrum of V_t is not constant. Existing estimators estimate an average of a time-varying spectrum. Because of this instability in the spectrum, they overestimate the dependence in V_t. Casini et al. (2021) showed that nonstationarity/misspecification alters the low frequency components of a time series making the latter appear as more persistent. Since classical HAC estimators are a weighted sum of an infinite number of low frequency periodogram ordinates, these estimates tend to be inflated. Similarly, LRV estimators using long bandwidths are weighted sum of a large number of sample autocovariances. Each sample autocovariance is biased upward so that the latter estimates are even more inflated than the classical HAC estimators. This explains why KVB's fixed-b and EWC HAR tests have large power problems, even though classical HAC estimators are also affected.

Casini et al. (2021) showed that the introduction of the smoothing over time in the DK-HAC estimators avoids such low frequency contamination. This follows because observations belonging to different regimes do not overlap when computing sample autocovariances. This guarantees excellent power properties also under nonstationarity/misspecification or under nonstationary alternative hypotheses (e.g., GR test discussed above). Simulation evidence suggests that tests based on the DK-HAC with simultaneous bandwidths are robust to low frequency contamination and overall performs better than tests based on the DK-HAC with sequential bandwidths especially with respect to size control.

7 Conclusions

We considered the derivation of data-dependent simultaneous bandwidths for double kernel heteroskedasticity and autocorrelation consistent (DK-HAC) estimators. We obtained the optimal bandwidths that jointly minimize the global asymptotic MSE criterion and discussed the trade-off between bias and variance with respect to smoothing over lagged autocovariances and over time. We highlighted how the derived MSE bounds are influenced by nonstationarity unlike the MSE bounds in Andrews (1991). We compared the DK-HAC estimators with simultaneous bandwidths to the DK-HAC estimators with bandwidths from the sequential MSE criterion. The new method leads to HAR tests that performs better in terms of size control, especially with stationary and close to stationary data. Finally, we considered long-run variance estimation where the relevant observations are a function of a nonparametric estimator and established the validity of the HAC and DK-HAC estimators in this setting. Hence, we also extended the consistency results in Andrews (1991) and Newey and West (1987) to nonparametric estimation settings.

References

Altissimo, F., Corradi, V., 2003. Strong rules for detecting the number of breaks in a time series. Journal of Econometrics 117 (2), 207–244.

Anderson, T.W., 1971. The Stastical Analysis of Time Series. New York: Wiley.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60 (4), 953–966.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

Belotti, F., Casini, A., Catania, L., Grassi, S., Perron, P., 2021. Supplement to "Simultaneous bandwidths determination for double-kernel HAC estimators and long-run variance estimation in nonparametric settings". arXiv preprint arXiv 2103.00060.

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

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136 (1), 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., 2019. Improved methods for statistical inference in the context of various types of parameter variation. Ph.D Dissertation, Boston University.

Casini, A., 2021. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor Vergata.

Casini, A., Deng, T., Perron, P., 2021. 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., 2020a. Continuous record Laplace-based inference about the break date in structural change models. Juornal of Econometrics forthcoming.

Casini, A., Perron, P., 2020b. Generalized Laplace inference in multiple change-points models. Econometric Theory forthcoming.

Casini, A., Perron, P., 2021a. Change-point analysis of time series with evolutionary spectra. Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor

Vergata.

Casini, A., Perron, P., 2021b. Continuous record asymptotics for change-point models. arXiv preprint arXiv:1803.10881.

Casini, A., Perron, P., 2021c. Minimax MSE bounds and nonlinear VAR prewhitening for long-run variance estimation under nonstattionarity. arXiv preprint arXiv 2103.02235.

Chan, K.W., 2020. Mean-structure and autocorrelation consistent covariance matrix estimation. Journal of Business and Economic Statistics, forthcoming.

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 (6), 577–601.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80 (3), 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 (6), 457–476.

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

Dahlhaus, R., 2012. Locally stationary processes. Vol. 30. Elsevier, pp. 351–413.

Dahlhaus, R., Giraitis, L., 1998. On the optimal segment length for parameter estimates for locally stationary time series. Journal of Time Series Analysis 19 (6), 629–655.

de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. Econometrica 68 (2), 407–423.

Deng, A., Perron, P., 2006. A comparison of alternative asymptotic frameworks to analyse a structural change in a linear time trend. Econometrics Journal 9 (3), 423–447.

Dou, L., 2019. Optimal HAR inference. Unpublished Manuscript, Department of Economics, Princeton University.

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

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. Review of Economic Studies 76 (2), 669–705.

Gonçalves, S., Vogelsang, T.J., 2011. Block bootstrap HAC robust tests: the sophistication of the naïve bootstrap. Econometric Theory 27 (4), 745–791.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Hannan, E.J., 1970. Multiple Time Series. New York: Wiley.

Ibragimov, R., Müller, U.K., 2010. t-statistic based correlation and heterogeneity robust inference.

Journal of Business and Economic Statistics 28 (4), 453–468.

Jansson, M., 2004. The error in rejection probability of simple autocorrelation robust tests. Econometrica 72 (3), 937–946.

Juhl, T., Xiao, Z., 2009. Testing for changing mean with monotonic power. Journal of Econometrics 148 (1), 14–24.

Kiefer, N., Vogelsang, T.J., 2002. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica 70 (5), 2093–2095.

Kiefer, N., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21 (6), 1130–1164.

Kiefer, N., Vogelsang, T.J., Bunzel, H., 2000. Simple robust testing of regression hypotheses. Econometrica 69 (3), 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 (1), 26–51.

Lazarus, E., Lewis, D.J., Stock, J.H., 2020. The size-power tradeoff in HAR inference. Econometrica, forthcoming.

Lazarus, E., Lewis, D.J., Stock, J.H., Watson, M.W., 2018. HAR inference: recommendations for practice. Journal of Business and Economic Statistics 36 (4), 541–559.

Martins, L., Perron, P., 2016. Improved tests for forecast comparisons in the presence of instabilities. Journal of Time Series Analysis 37 (5), 650–659.

Müller, U.K., 2007. A theory of robust long-run variance estimation. Journal of Econometrics 141 (2), 1331–1352.

Müller, U.K., 2014. HAC corrections for strongly autocorrelated time series. Journal of Business and Economic Statistics 32 (3), 311–322.

Neumann, M.H., von Sachs, R., 1997. Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Annals of Statistics 25 (1), 38–76.

Newey, W.K., West, K.D., 1987. A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61 (4), 631–653.

Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. Annals of Mathematical Statistics 28 (2), 329–348.

Perron, P., Yamamoto, Y., 2021. Testing for changes in forecast performance. Journal of Business and Economic Statistics 39 (1), 148–165.

Phillips, P.C.B., 2005. HAC estimation by automated regression. Econometric Theory 21 (1), 116–

142.

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

Pötscher, B.M., Preinerstorfer, D., 2018. Controlling the size of autocorrelation robust tests. Journal of Econometrics 207 (2), 406–431.

Pötscher, B.M., Preinerstorfer, D., 2019. Further results on size and power of heteroskedasticity and autocorrelation robust tests, with an application to trend testing. Electronic Journal of Statistics 13 (2), 3893–3942.

Preinerstorfer, D., Pötscher, B. M., 2016. On size and power of heteroskedasticity and autocorrelation robust tests. Econometric Theory 32 (2), 261–358.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Vol. I and II. New York: Academic Press.

Robinson, P.M., 1998. Inference-without smoothing in the presence of nonparametric autocorrelation. Econometrica 66 (5), 1163–1182.

Sun, Y., 2013. Heteroscedasticity and autocorrelation robust F test using orthonormal series variance estimator. Econometrics Journal 16 (1), 1–26.

Sun, Y., 2014a. Fixed-smoothing asymptotics in a two-step GMM framework. Econometrica 82 (6), 2327–2370.

Sun, Y., 2014b. Let's dix it: fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. Journal of Econometrics 178 (3), 659–677.

Velasco, C., Robinson, P.M., 2001. Edgeworth expansions for spectral density estimates and studentized sample mean. Econometric Theory 17 (3), 497–539.

Vogelsang, T.J., 1999. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. Journal of Econometrics 88 (2), 283–299.

Zhang, X., Shao, X., 2013. Fixed-smoothing asymptotics for time series. Annals of Statistics 41 (3), 1329–1349.

A Appendix

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

	Model M1, t_1		Model M2, t_2	
$\alpha = 0.05$	$T = 200$	$T = 400$	$T = 200$	$T = 400$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.079	0.059	0.074	0.073
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.042	0.049	0.046	0.063
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.055	0.044	0.057	0.055
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.057	0.052	0.057	0.056
$\widehat{J}_T$, Casini (2020)	0.117	0.102	0.083	0.082
$\widehat{J}_T$, prewhite, CP	0.062	0.055	0.063	0.061
$\widehat{J}_T$, prewhite, SLS, CP	0.060	0.059	0.068	0.065
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.061	0.060	0.074	0.054
Newey-West (1987)	0.101	0.086	0.086	0.085
Newey-West (1987), prewhite	0.076	0.060	0.078	0.080
Newey-West (1987), fixed-b (KVB)	0.058	0.055	0.061	0.059
EWC	0.062	0.047	0.059	0.069

CP stands for Casini and Perron (2021c).

Table 2: Empirical small-sample size of t_2-tests for model M3 and of model M4

	Model M3, t_2		Model M4, GR test	
$\alpha = 0.05$	$T = 200$	$T = 400$	$T = 200$	$T = 800$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.064	0.068	0.079	0.061
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.029	0.062	0.054	0.047
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.032	0.043	0.062	0.054
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.033	0.043	0.073	0.064
$\widehat{J}_T$, Casini (2020)	0.024	0.027	0.055	0.060
$\widehat{J}_T$, prewhite, CP	0.015	0.000	0.057	0.042
$\widehat{J}_T$, prewhite, SLS, CP	0.063	0.062	0.061	0.055
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.073	0.069	0.061	0.056
Newey-West (1987)	0.074	0.069	0.000	0.000
Newey-West (1987), prewhite	0.041	0.000	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.021	0.005	0.061	0.044
EWC	0.031	0.010	0.017	0.017

CP stands for Casini and Perron (2021c).

Table 3: Empirical small-sample rejection rates of the t_1-test for model M1

	Model M1, t_1			
$\alpha = 0.05$, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.218	0.589	0.980	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.132	0.465	0.960	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.172	0.553	0.958	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.174	0.544	0.958	1.000
$\widehat{J}_T$, Casini (2020)	0.291	0.620	0.980	1.000
$\widehat{J}_T$, prewhite, CP	0.191	0.518	0.949	1.000
$\widehat{J}_T$, prewhite, SLS, CP	0.161	0.509	0.969	1.000
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.165	0.508	0.970	1.000
Newey-West (1987)	0.248	0.629	0.987	1.000
Newey-West (1987), prewhite	0.197	0.576	0.979	1.000
Newey-West (1987), fixed-b (KVB)	0.141	0.373	0.844	0.998
EWC	0.150	0.493	0.963	1.000

CP stands for Casini and Perron (2021c).

Table 4: Empirical small-sample rejection rates of the t_2-tests for model M2

$\alpha = 0.05$, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.263	0.642	0.988	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.191	0.532	0.968	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.221	0.592	0.982	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.221	0.597	0.983	1.000
$\widehat{J}_T$, Casini (2020)	0.276	0.653	0.988	1.000
$\widehat{J}_T$, prewhite, CP	0.237	0.611	0.986	1.000
$\widehat{J}_T$, prewhite, SLS, CP	0.225	0.598	0.982	1.000
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.165	0.598	0.988	1.000
Newey-West (1987)	0.268	0.332	0.992	1.000
Newey-West (1987), prewhite	0.258	0.374	0.990	1.000
Newey-West (1987), fixed-b (KVB)	0.199	0.463	0.914	1.000
EWC	0.193	0.571	0.978	1.000

CP stands for Casini and Perron (2021c).

Table 5: Empirical small-sample rejection rates of the t_2-test for model M3

$\alpha = 0.05$, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.167	0.429	0.794	0.962
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.112	0.230	0.691	0.921
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.104	0.325	0.687	0.912
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.104	0.328	0.688	0.912
$\widehat{J}_T$, Casini (2020)	0.091	0.293	0.687	0.940
$\widehat{J}_T$, prewhite, CP	0.046	0.228	0.537	0.836
$\widehat{J}_T$, prewhite, SLS, CP	0.152	0.381	0.728	0.946
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.164	0.395	0.741	0.954
Newey-West (1987)	0.212	0.487	0.839	0.973
Newey-West (1987), prewhite	0.154	0.416	0.779	0.950
Newey-West (1987), fixed-b (KVB)	0.101	0.298	0.661	0.906
EWC	0.151	0.409	0.793	0.960

CP stands for Casini and Perron (2021c).

Table 6: Empirical small-sample rejection rates of the GR tests

$\alpha = 0.05$, $T = 800$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$	0.127	0.719	0.982	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite	0.108	0.681	0.982	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS	0.113	0.730	0.991	1.000
$\widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T})$, prewhite, SLS, μ	0.121	0.719	0.982	1.000
$\widehat{J}_T$, Casini (2020)	0.139	0.622	0.812	0.915
$\widehat{J}_T$, prewhite, CP	0.114	0.699	0.965	0.994
$\widehat{J}_T$, prewhite, SLS, CP	0.134	0.779	0.989	0.999
$\widehat{J}_T$, prewhite, SLS, μ, CP	0.152	0.793	0.989	0.999
Newey-West (1987)	0.000	0.000	0.000	0.000
Newey-West (1987), prewhite	0.000	0.000	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.062	0.042	0.000	0.000
EWC	0.050	0.044	0.004	0.000

CP stands for Casini and Perron (2021c).

Supplemental Material to

Simultaneous Bandwidths Determination for DK-HAC Estimators and Long-Run Variance Estimation in Nonparametric Settings

Federico Belotti — University of Rome Tor Vergata

Alessandro Casini — University of Rome Tor Vergata

Leopoldo Catania — Aarhus University

Stefano Grassi — University of Rome Tor Vergata

Pierre Perron — Boston University

4th March 2021

Abstract

This supplemental material includes the proofs of the results in the paper.

S.A Mathematical Appendix

In some of the proofs below $\overline{\beta}$ is understood to be on the line segment joining $\widehat{\beta}$ and β_0. We discard the degrees of freedom adjustment $T/(T-p)$ from the derivations since asymptotically it does not play any role. Similarly, we use T/n_T in place of $(T-n_T)/n_T$ in the expression for $\widehat{\Gamma}(k)$. Let $\widetilde{c}_T(u,\,k)$ denote the estimator that uses $\{V_{t,T}\}$.

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

S.A.1.1 Proof of Theorem 3.1

Part (i) follows from Theorem 3.1 in Casini (2021). For part (ii), let $J_{c,T} = \int_0^1 c(u,\,0) + 2\sum_{k=1}^{T-1}\int_0^1 c(u,\,k)\,du$ and $\mathcal{T}_C \triangleq \{0,\,n_T,\,\ldots,\,T-n_T,\,T\}/\mathcal{T}$. We begin with the following relationship,

$$\mathbb{E}\left(\widetilde{J}_T - J_T\right) = \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\,\mathbb{E}\left(\widetilde{\Gamma}(k)\right) - J_{c,T} + \left(J_{c,T} - J_T\right).$$

Using Lemma S.B.4 in Casini (2021), we have for any $0 \le k \le T-1$,

$$\begin{aligned}
&\mathbb{E}\left(\frac{n_T}{T}\sum_{r=0}^{T/n_T}\widetilde{c}_T(rn_T/T,\,k) - \int_0^1 c(u,\,k)\,du\right)\\
&= \frac{n_T}{T}\sum_{r\in\mathcal{T}_C}\left(c(rn_T/T,\,k) + \frac{1}{2}b_{2,T}^2\int_0^1 x^2K_2(x)\,dx\frac{\partial^2}{\partial^2 u}c(u,\,k) + o\left(b_{2,T}^2\right) + O\left(\frac{1}{b_{2,T}T}\right)\right)\\
&\quad + \frac{n_T}{T}\sum_{r\notin\mathcal{T}_C}\left(c(rn_T/T,\,k) + \frac{1}{2}b_{2,T}^2\int_0^1 x^2K_2(x)\,dx\right.\\
&\quad \times\int_{-\pi}^{\pi}\exp(i\omega k)\left(C_1(rn_T/T,\,\omega) + C_2(rn_T/T,\,\omega) + C_3(rn_T/T,\,\omega)\right)d\omega\\
&\quad \left. + o\left(b_{2,T}^2\right) + O\left(\frac{1}{b_{2,T}T}\right)\right) - \int_0^1 c(u,\,k)\,du,
\end{aligned}$$

where

$$C_1(u_0,\,\omega) = 2\frac{\partial A_j(u_0,\,-\omega)}{\partial_- u}\frac{\partial A_{j+1}(v_0,\,\omega)}{\partial_+ v} \qquad C_2(u_0,\,\omega) = \frac{\partial^2 A_{j+1}(v_0,\,\omega)}{\partial_+ v^2}A_j(u_0,\,-\omega),$$
$$C_3(u_0,\,\omega) = \frac{\partial^2 A_j(u_0,\,\omega)}{\partial_- u^2}A_{j+1}(v_0,\,\omega),$$

with $u_0 = rn_T/T$ and $v_0 = u_0 - k/2T$. The right-hand side above is equal to

$$\begin{aligned}
&\frac{n_T}{T}\sum_{r=0}^{T/n_T}c(rn_T/T,\,k) - \int_0^1 c(u,\,k)\,du\\
&\quad + \frac{1}{2}b_{2,T}^2\int_0^1 x^2K_2(x)\,dx\int_0^1\frac{\partial^2}{\partial^2 u}c(u,\,k)\,du + o\left(b_{2,T}^2\right) + O\left(\frac{1}{Tb_{2,T}}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\\
&\times\int_{0}^{1}\left(\int_{-\pi}^{\pi}\exp\left(i\omega k\right)\left(C_{1}\left(u,\,\omega\right)+C_{2}\left(u,\,\omega\right)+C_{3}\left(u,\,\omega\right)\right)d\omega\mathbf{1}\left\{Tu\in\mathcal{T}\right\}\right)du\\
&+o\left(b_{2,T}^{2}\right)+O\left(\frac{1}{Tb_{2,T}}\right)\\
=\;&O\left(\frac{n_{T}}{T}\right)+\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du+o\left(b_{2,T}^{2}\right)+O\left(\frac{1}{Tb_{2,T}}\right),
\end{aligned}
$$

where the last equality follows from the convergence of approximations to Riemann sums and from the fact that $\mathbf{1}\left\{Tu\in\mathcal{T}\right\}$ has zero Lebesgue measure. This leads to,

$$
\begin{aligned}
&b_{1,T}^{-q}\mathbb{E}\left(\widetilde{J}_{T}-J_{c,T}\right)\\
&=-b_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_{1}\left(b_{1,T}k\right)\right)\int_{0}^{1}c\left(u,\,k\right)du\\
&\quad+\frac{1}{2}\frac{b_{2,T}^{2}}{b_{1,T}^{q}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{k=-T+1}^{T}K_{1}\left(b_{1,T}k\right)\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du+O\left(\frac{1}{Tb_{1,T}^{q}b_{2,T}}\right)+O\left(\frac{n_{T}}{Tb_{1,T}^{q}}\right)\\
&=-b_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_{1}\left(b_{1,T}k\right)\right)\int_{0}^{1}c\left(u,\,k\right)du\\
&\quad-\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}\left(x\right)dxb_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_{1}\left(b_{1,T}k\right)\right)\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du\\
&\quad+\frac{1}{2}\frac{b_{2,T}^{2}}{b_{1,T}^{q}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{k=-T+1}^{T}\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du+O\left(\frac{1}{Tb_{1,T}^{q}b_{2,T}}\right)+O\left(\frac{n_{T}}{Tb_{1,T}^{q}}\right)\\
&=-b_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_{1}\left(b_{1,T}k\right)\right)\int_{0}^{1}c\left(u,\,k\right)du\\
&\quad-\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}\left(x\right)dxO\left(1\right)+\frac{1}{2}\frac{b_{2,T}^{2}}{b_{1,T}^{q}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{k=-T+1}^{T}\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du\\
&\quad+O\left(\frac{1}{Tb_{1,T}^{q}b_{2,T}}\right)+O\left(\frac{n_{T}}{Tb_{1,T}^{q}}\right),
\end{aligned}
$$

since $\left|\sum_{k=-\infty}^{\infty}\left|k\right|^{q}\int_{0}^{1}\left(\partial^{2}/\partial^{2}u\right)c\left(u,\,k\right)du\right|<\infty$ by Assumption 3.2-(i). Since $J_{c,T}-J_{T}=O\left(T^{-1}\right)$ we conclude that,

$$
\begin{aligned}
\lim_{T\to\infty}b_{1,T}^{-q}\mathbb{E}\left(\widetilde{J}_{T}-J_{T}\right)=&-2\pi K_{1,q}\int_{0}^{1}f^{(q)}\left(u,\,0\right)du\\
&+\frac{1}{2}\frac{b_{2,T}^{2}}{b_{1,T}^{q}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{k=-T+1}^{T}\int_{0}^{1}\frac{\partial^{2}}{\partial^{2}u}c\left(u,\,k\right)du,
\end{aligned}
$$

using $b_{2,T}^2/b_{1,T}^q \to \nu$. Part (iii) follows from part (i)-(ii) and the commutation-tensor product formula follows from the same argument as in Theorem 3.1-(iii) in Casini (2021). □

S.A.1.2 Proof of Theorem 3.2

The proof follows the same steps as in Theorem 3.2 in Casini (2021) with references to Theorem 3.1 there replaced by references to Theorem 3.1 here and in addition it uses the fact that $\widetilde{J}_T$ is still asymptotically unbiased because

$$\begin{aligned}\mathbb{E}\left(\widetilde{J}_T - J_T\right) &= -2\pi K_{1,q} b_{1,T}^q \int_0^1 f^{(q)}(u,\,0)\,du \\ &\quad + \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2(x)\,dx \sum_{k=-T+1}^{T} \int_0^1 \frac{\partial^2}{\partial^2 u} c(u,\,k)\,du, \\ &\to 0.\end{aligned}$$

Note that the above result continues to hold even when $q = 0$ since then $K_{1,q} = 0$. □

S.A.1.3 Proof of Theorem 3.3

Let $\Delta_2 \triangleq \mathrm{tr} W \left(I_{p^2} + C_{pp}\right) I_p \otimes I_p$. We focus on the scalar case. The derivations for the multivariate case are straightforward but tedious and so we omit them. Note that

$$\begin{aligned}\mathrm{ReMSE} &= \mathbb{E}\left(\left(\frac{\widehat{J}_T(b_{1,T},\,b_{2,T})}{J} - 1\right)^2\right) \\ &= \mathbb{E}\left(\left(\frac{\widehat{J}_T(b_{1,T},\,b_{2,T})}{J} - \mathbb{E}\left(\frac{\widehat{J}_T(b_{1,T},\,b_{2,T})}{J}\right)\right)^2\right) + \left(\mathbb{E}\left(\frac{\widehat{J}_T(b_{1,T},\,b_{2,T})}{J}\right) - 1\right)^2 \\ &= (T b_{2,T} b_{1,T})^{-1} \varpi_3 + \left(b_{1,T}^q \varpi_1 + b_{2,T}^2 \varpi_2\right)^2,\end{aligned}$$

where $\varpi_1 = \left(\int_0^1 f(u,\,0)\,du\right)^{-1} \Xi_{1,1}\Delta_{1,1,0}$, $\varpi_2 = \left(\int_0^1 f(u,\,0)\,du\right)^{-1} \Xi_{1,2}\Delta_{1,2}$ and $\varpi_3 = \Xi_2$. Minimizing the right-hand side above with respect to $b_{1,T}$ and $b_{2,T}$ yields

$$b_{1,T} = T^{-\frac{2}{2+5q}} \left(2q^{-1}\frac{\varpi_2}{\varpi_1}\right)^{1/q} \left(\frac{\left(\left(2q^{-1}\frac{\varpi_2}{\varpi_1}\right)^{1/q}\right)^{-1} \varpi_3}{(2q^{-1}+1)\,4\varpi_2^2}\right)^{2/(2+5q)}, \quad b_{2,T} = \left(\frac{\left(T\left(2q^{-1}\frac{\varpi_2}{\varpi_1}\right)^{1/q}\right)^{-1} \varpi_3}{(2q^{-1}+1)\,4\varpi_2^2}\right)^{q/(2+5q)}.$$

At this point to solve for $\widetilde{b}_{1,T}^{\mathrm{opt}}$ and $\widetilde{b}_{2,T}^{\mathrm{opt}}$ we guess and verify that the optimal K_1 is such that $q = 2$. Note that $q = 2$ holds for the Parzen, Tukey-Hanning and the QS kernels. Thus, substituting out $q = 2$, we have

$$\widetilde{b}_{1,T}^{\mathrm{opt}} = T^{-1/6}\left(\frac{\varpi_2}{\varpi_1^5}\right)^{1/12}\left(\frac{\varpi_3}{8}\right)^{1/6}, \qquad \widetilde{b}_{2,T}^{\mathrm{opt}} = T^{-1/6}\left(\frac{\varpi_2^5}{\varpi_1}\right)^{-1/12}\left(\frac{\varpi_3}{8}\right)^{1/6}.$$

The optimal relative MSE is then given by

$$\begin{aligned}\text{ReMSE}\left(\widetilde{b}_{1,T}^{\text{opt}},\,\widetilde{b}_{2,T}^{\text{opt}}\right) &= \left(Tb_{2,T}b_{1,T}\right)^{-1}\varpi_3 + \left(b_{1,T}^2\varpi_1 + b_{2,T}^2\varpi_2\right)^2 \\ &= 3\left(4\pi\right)^{-1/3}\left(2\pi\right)^{1/3}\left(-K_{1,q}\right)^{1/3}\Delta_{1,1,0}^{1/3}\Delta_{1,2}^{1/3}\Delta_2^{2/3}\left(\int_0^1 x^2K_2\left(x\right)dx\right)^{1/3} \\ &\quad\times\left(\int K_1^2\left(y\right)dy\int_0^1 K_2^2\left(x\right)dx\right)^{2/3}.\end{aligned}$$

Recall that $K_{1,2} = -\left(d^2K_1\left(x\right)/dx^2\right)|_{x=0}/2$. The spectral window generator of K_1 is defined as $K_{1,\text{SWG}}\left(\omega\right) = \left(2\pi\right)^{-1}\int_{-\infty}^{\infty}K_1\left(x\right)e^{-ix\omega}dx$. It follows from Priestley (1981, Ch. 6) that for K_1^{QS} we have

$$K_{1,\text{SWG}}^{\text{QS}}\left(\omega\right) = \begin{cases}\frac{3}{4\pi}\left(1-\left(\omega/\pi\right)\right), & |\omega|\leq\pi \\ 0 & |\omega|>\pi\end{cases}.$$

We also have the following properties: $K_{1,2} = \int_{-\infty}^{\infty}\omega^2K_{1,\text{SWG}}\left(\omega\right)d\omega$, $K_1\left(0\right) = \int_{-\infty}^{\infty}K_{1,\text{SWG}}\left(\omega\right)d\omega$, and $\int_{-\infty}^{\infty}K_1^2\left(x\right)dx = \int_{-\infty}^{\infty}K_{1,\text{SWG}}^2\left(\omega\right)d\omega$. We now minimize $\text{ReMSE}\left(\widetilde{b}_{1,T}^{\text{opt}},\,\widetilde{b}_{2,T}^{\text{opt}}\right)$ with respect to K_1 and K_2 under the restrictions that $\int_0^1 K_2\left(x\right)dx = 1$ and that (a) $\int_{-\infty}^{\infty}K_{1,\text{SWG}}\left(\omega\right)d\omega = 1$, (b) $K_{1,\text{SWG}}\left(\omega\right)\geq 0$, $\forall\,\omega\in\mathbb{R}$, and (c) $K_{1,\text{SWG}}\left(\omega\right) = K_{1,\text{SWG}}\left(-\omega\right)$, $\forall\,\omega\in\mathbb{R}$. This is equivalent to minimizing (1) $\left(\int_0^1 K_2^2\left(x\right)dx\right)^2\left(\int_0^1 x^2K_2\left(x\right)dx\right)$ subject to $\int_0^1 K_2\left(x\right)dx = 1$, and (2) $\int_{-\infty}^{\infty}\omega^2\widetilde{K}_1\left(\omega\right)d\omega\left(\int_{-\infty}^{\infty}\widetilde{K}_1^2\left(\omega\right)d\omega\right)^2$ subject to (a)-(c). Using a calculus of variations, Priestley (1981, Ch. 7) showed that $K_2^{\text{opt}} = 6x\left(1-x\right)$ for $x\in[0,\,1]$ solves (1) and that K_1^{QS} solves (2). Since the equivalence between the optimization problem (1)-(2) and our problem is independent of q, this verifies that our guess was correct because $q = 2$ for K_1^{QS}. Therefore,

$$\widetilde{b}_{1,T}^{\text{opt}} = 0.46T^{-1/6}\left(\Delta_{1,1,0}/\int_0^1 f\left(u,\,0\right)du\right)^{-5/12}\left(\Delta_{1,2}/\int_0^1 f\left(u,\,0\right)du\right)^{1/12}, \qquad \text{and}$$

$$\widetilde{b}_{2,T}^{\text{opt}} = 3.56T^{-1/6}\left(\Delta_{1,1,0}/\int_0^1 f\left(u,\,0\right)du\right)^{1/12}\left(\Delta_{1,2}/\int_0^1 f\left(u,\,0\right)du\right)^{-5/12}.$$

The result for $\widetilde{b}_{1,T}^{\text{opt}}$ and $\widetilde{b}_{2,T}^{\text{opt}}$ for the multivariate case follows from the matrix-form of the above expressions. □

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

S.A.2.1 Proof of Theorem 4.1

Without loss of generality, we assume that V_t is a scalar. The constant $C<\infty$ may vary from line to line. We begin with the proof of part (ii) because it becomes then simpler to prove part (i). By Theorem 3.2-(ii), $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_T(b_{\theta_1,T},\,b_{\theta_2,T})-J_T) = O_{\mathbb{P}}\left(1\right)$. It remains to establish the second result of Theorem 4.1-(ii). Let $S_T = \left\lfloor b_{\theta_1,T}^{-r}\right\rfloor$ where

$$r\in\left(\max\left\{1,\,\left(\bar{b}-1/2\right)/\left(\bar{b}-1\right),\,2/\left(l-1\right),\,\left(b-2\right)/\left(b-1\right)\right\},\,5/4\right),$$

with $b > 1 + 1/q$ and $\bar{b} > 3$. We will use the following decomposition

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

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,

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

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 Lipschitz condition on $K_1(\cdot) \in \boldsymbol{K}_3$ to yield,

$$\begin{aligned} |A_{1,1,T}| &\leq T^{1/3} \sum_{k=1}^{S_T} C_2 \left|\widehat{b}_{1,T} - b_{\theta_1,T}\right| k \left|\widehat{\Gamma}(k)\right| \\ &\leq C\widetilde{n}_T \left|\widehat{\phi}_1^{1/24} - \phi_{1,\theta^*}^{1/24}\right| \left(\widehat{\phi}_1 \phi_{1,\theta^*}\right)^{-1/24} T^{1/3-1/6} \widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k)\right|, \end{aligned} \tag{S.3}$$

for some $C < \infty$. By Assumption 4.1-(ii) ($\widetilde{n}_T|\widehat{\phi}_1 - \phi_{1,\theta^*}| = O_{\mathbb{P}}(1)$) and using the delta method it suffices to show that $B_{1,T} + B_{2,T} + B_{3,T} \xrightarrow{\mathbb{P}} 0$ where

$$\begin{aligned} B_{1,T} &= T^{1/6}\widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right|, \\ B_{2,T} &= T^{1/6}\widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|, \qquad \text{and} \\ B_{3,T} &= T^{1/6}\widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\Gamma_T(k)\right|, \end{aligned} \tag{S.4}$$

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

$$\begin{aligned} B_{1,T} &\leq T^{1/6}\widetilde{n}_T^{-1} T^{-1/2} \sum_{k=1}^{S_T} k \left|\left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}\right) \sqrt{T}\left(\widehat{\beta} - \beta_0\right)\right| \\ &\leq CT^{1/6-1/2} T^{2r/6} \widetilde{n}_T^{-1} \sup_{k \geq 1} \left\|\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}\right\| \sqrt{T} \left\|\widehat{\beta} - \beta_0\right\| \end{aligned} \tag{S.5}$$

$$\leq CT^{1/6-1/2+r/3}\widetilde{n}_T^{-1}\sup_{k\geq 1}\left\|\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right\|\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\|\xrightarrow{\mathbb{P}}0,$$

since $\widetilde{n}_T/T^{1/3}\to\infty$, $r<2$, $\sqrt{T}||\widehat{\beta}-\beta_0||=O_{\mathbb{P}}(1)$, and $\sup_{k\geq 1}||(\partial/\partial\beta')\widehat{\Gamma}(k)|_{\beta=\overline{\beta}}||=O_{\mathbb{P}}(1)$ using (S.28) in Casini (2021) and Assumption 3.3-(ii,iii). In addition,

$$\begin{aligned}\mathbb{E}\left(B_{2,T}^2\right)&\leq\mathbb{E}\left(T^{2/3}\widetilde{n}_T^{-2}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T}kj\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\left|\widetilde{\Gamma}(j)-\Gamma_T(j)\right|\right)\\&\leq b_{\theta_2,T}^{-1}T^{2/3-1-2/3}S_T^4\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\\&\leq b_{\theta_2,T}^{-1}T^{-1}T^{4r/6}\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\to 0,\end{aligned}\tag{S.6}$$

given that $\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}(\widetilde{\Gamma}(k))=O(1)$ using Lemma S.B.5 in Casini (2021) and $r<5/4$. Assumption 4.1-(iii) and $\sum_{k=1}^{\infty}k^{1-l}<\infty$ for $l>2$ yield,

$$B_{3,T}\leq T^{1/6}\widetilde{n}_T^{-1}C_3\sum_{k=1}^{\infty}k^{1-l}\to 0,\tag{S.7}$$

where we have used the fact that $\widetilde{n}_T/T^{3/10}\to\infty$. Combining (S.3)-(S.7) we deduce that $A_{1,1,T}\xrightarrow{\mathbb{P}}0$. The same argument applied to $A_{1,T}$ where the summation now extends over negative integers k gives $A_{1,T}\xrightarrow{\mathbb{P}}0$. Next, we show that $A_{2,T}\xrightarrow{\mathbb{P}}0$. Again, we use the notation $A_{2,1,T}$ (resp., $A_{2,2,T}$) to denote $A_{2,T}$ with the summation over positive (resp., negative) integers. Let $A_{2,1,T}=L_{1,T}+L_{2,T}+L_{3,T}$, where

$$\begin{aligned}L_{1,T}&=L_{1,T}^A+L_{1,T}^B=T^{1/3}\left(\sum_{k=S_T+1}^{\lfloor D_TT^{1/3}\rfloor}+\sum_{k=\lfloor D_TT^{1/3}\rfloor+1}^{T-1}\right)K_1\left(\widehat{b}_{1,T}k\right)\left(\widehat{\Gamma}(k)-\widetilde{\Gamma}(k)\right),\\L_{2,T}&=L_{2,T}^A+L_{2,T}^B=T^{1/3}\left(\sum_{k=S_T+1}^{\lfloor D_TT^{1/3}\rfloor}+\sum_{k=\lfloor D_TT^{1/3}\rfloor+1}^{T-1}\right)K_1\left(\widehat{b}_{1,T}k\right)\left(\widetilde{\Gamma}(k)-\Gamma_T(k)\right),\\\text{and}\qquad L_{3,T}&=T^{1/3}\sum_{k=S_T+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)\Gamma_T(k).\end{aligned}\tag{S.8}$$

We apply a mean-value expansion and use $\sqrt{T}(\widehat{\beta}-\beta_0)=O_{\mathbb{P}}(1)$ to obtain

$$\begin{aligned}\left|L_{1,T}^A\right|&=T^{1/3-1/2}\sum_{k=S_T+1}^{\lfloor D_TT^{1/3}\rfloor}C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right|\\&=T^{1/3-1/2+b/6}\sum_{k=S_T+1}^{\lfloor D_TT^{1/3}\rfloor}C_1k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right|\\&=T^{1/3-1/2+b/6+r(1-b)/6}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right|\\&=T^{1/3-1/2+b/6+r(1-b)/6}O_{\mathbb{P}}(1)\,O_{\mathbb{P}}(1),\end{aligned}\tag{S.9}$$

which goes to zero since $r > 1$ and

$$\begin{aligned}\left|L_{1,T}^{B}\right| &= T^{1/3-1/2}\sum_{k=\lfloor D_T T^{1/3}\rfloor+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &= CT^{1/3-1/2+b/6}\sum_{k=\lfloor D_T T^{1/3}\rfloor+1}^{T-1} C_1 k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right| \\ &= CD_T^{1-b}T^{1/3-1/2+b/6+(1-b)/3}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right| \xrightarrow{\mathbb{P}} 0,\end{aligned}$$

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

$$\begin{aligned}\left|L_{2,T}^{A}\right| &= T^{1/3}\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right| \\ &= C_1\left(\widehat{\phi}_1\right)^{-b/24}T^{1/3+b/6-1/2}b_{\theta_2,T}^{-1/2}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor}k^{-b}\right) \\ &\quad\times\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|.\end{aligned} \tag{S.10}$$

Note that

$$\begin{aligned}&\mathbb{E}\left(T^{1/3+b/6-1/2}b_{\theta_2,T}^{-1/2}\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor}k^{-b}\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\right)^2 \\ &\le CT^{2/3+b/3-1}b_{\theta_2,T}^{-1}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2}\right)^2 \\ &= T^{2/3+b/3-1}b_{\theta_2,T}^{-1}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor}k^{-b}\right)^2 O(1) \\ &= T^{2/3+b/3-1}b_{\theta_2,T}^{-1}S_T^{2(1-b)}O(1)\to 0,\end{aligned} \tag{S.11}$$

since $r > (b-1/2)/(b-1)$, $b>3$, and $Tb_{\theta_2,T}\mathrm{Var}(\widetilde{\Gamma}(k)) = O(1)$ as above. Next,

$$\begin{aligned}\left|L_{2,T}^{B}\right| &= T^{1/3}\sum_{k=\lfloor D_T T^{1/3}\rfloor+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right| \\ &= C_1\left(\widehat{\phi}_1\right)^{-b/24}T^{1/3+b/6-1/2}b_{\theta_2,T}^{-1/2}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\right) \\ &\quad\times\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|,\end{aligned}$$

and

$$\mathbb{E}\left(T^{1/3+b/6-1/2}b_{\theta_2,T}^{-1/2}\sum_{k=\lfloor D_T T^{1/3}\rfloor+1}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\right)^2 \tag{S.12}$$
$$= T^{2/3+b/3-1}b_{\theta_2,T}^{-1}\left(\sum_{k=\lfloor D_T T^{1/3}\rfloor+1}^{T-1}k^{-b}\right)^2 O(1)$$
$$= T^{2/3+b/3-1}b_{\theta_2,T}^{-1}T^{2(1-b)/3}D_T^2 O(1)\to 0,$$

since $2b>3$ given $b>1+1/q$ and $q\leq 2$. The cross-product term involving

$$\left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/3}\rfloor}\sum_{j=\lfloor D_T T^{1/3}\rfloor+1}^{T-1}\right)K_1\left(\widehat{b}_{1,T}k\right)K_1\left(\widehat{b}_{1,T}j\right)\left(\widetilde{\Gamma}(j)-\Gamma_T(j)\right)\left(\widetilde{\Gamma}(j)-\Gamma_T(j)\right),$$

can be treated in a similar fashion. Combining (S.10)-(S.12) yields $L_{2,T}\xrightarrow{\mathbb{P}}0$. Let us turn to $L_{3,T}$. By Assumption 4.1-(iii) and $|K_1(\cdot)|\leq 1$, we have,

$$|L_{3,T}|\leq T^{1/3}\sum_{k=S_T}^{T-1}C_3k^{-l}\leq T^{1/3}C_3S_T^{1-l}\leq C_3T^{1/3}T^{-r(l-1)/6}\to 0, \tag{S.13}$$

since $r>2/(l-1)$. In view of (S.8)-(S.13) 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 $T^{1/3}(\widehat{J}_T(b_{\theta_1,T},\,\widehat{b}_{2,T})-\widehat{J}_T(b_{\theta_1,T},\,b_{\theta_2,T}))\xrightarrow{\mathbb{P}}0$. Let $\widehat{c}_{\theta_2,T}(rn_T/T,\,k)$ denote the estimator that uses $b_{\theta_2,T}$ in place of $\widehat{b}_{2,T}$. We have for $k\geq 0$,

$$\begin{aligned}&\widehat{c}_T(rn_T/T,\,k)-\widehat{c}_{\theta_2,T}(rn_T/T,\,k)\\&\quad=(Tb_{\theta_2,T})^{-1}\sum_{s=k+1}^{T}\left(K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{\theta_2,T}}\right)\right)\widehat{V}_s\widehat{V}_{s-k}\\&\qquad+O_{\mathbb{P}}\left(1/Tb_{2,b_{\theta_2,T}}\right).\end{aligned} \tag{S.14}$$

Given Assumption 4.1-(ii,v) and using the delta method, we have for $s\in\{Tu-\lfloor Tb_{\theta_2,T}\rfloor,\ldots,Tu+\lfloor Tb_{\theta_2,T}\rfloor\}$

$$\begin{aligned}&K_2\left(\frac{(Tu-(s-k/2))/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{(Tu-(s-k/2))/T}{b_{\theta_2,T}}\right)\\&\leq C_4\left|\frac{Tu-(s-k/2)}{T\widehat{b}_{2,T}}-\frac{Tu-(s-k/2)}{Tb_{\theta_2,T}}\right|\\&\leq C_4T^{-5/6}\widetilde{n}_T^{-1}\widetilde{n}_T\left|\widehat{\phi}_2^{-1/24}-\phi_{2,\theta^*}^{-1/24}\right||Tu-(s-k/2)|\\&\leq CT^{-5/6}\widetilde{n}_T^{-1}O_{\mathbb{P}}(1)\,|Tu-(s-k/2)|.\end{aligned} \tag{S.15}$$

Therefore,

$$
\begin{aligned}
&T^{1/3}\left(\widehat{J}_T\left(b_{\theta_1,T},\,\widehat{b}_{2,T}\right)-\widehat{J}_T\left(b_{\theta_1,T},\,b_{\theta_2,T}\right)\right)\\
&\quad=T^{1/3}\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(\widehat{c}\left(rn_T/T,\,k\right)-\widehat{c}_{\theta_2,T}\left(rn_T/T,\,k\right)\right)\\
&\quad\le T^{1/3}C\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{\theta_1,T}k\right)\right|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\frac{1}{Tb_{\theta_2,T}}\\
&\qquad\times\sum_{s=k+1}^{T}\left|K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{\theta_2,T}}\right)\right|\\
&\qquad\times\left|\left(\widehat{V}_s\widehat{V}_{s-k}-V_sV_{s-k}\right)+\left(V_sV_{s-k}-\mathbb{E}\left(V_sV_{s-k}\right)\right)+\mathbb{E}\left(V_sV_{s-k}\right)\right|\\
&\quad\triangleq H_{1,T}+H_{2,T}+H_{3,T}.
\end{aligned}
$$

We have to show that $H_{1,T}+H_{2,T}+H_{3,T}\xrightarrow{\mathbb{P}}0$. Let $H_{1,1,T}$ (resp. $H_{1,2,T}$) be defined as $H_{1,T}$ but with the sum over k be restricted to $k=1,\ldots,S_T$ (resp. $k=S_T+1,\ldots,T$). By a mean-value expansion, using (S.15),

$$
\begin{aligned}
\left|H_{1,1,T}\right|&\le CT^{1/3-1/2}\sum_{k=1}^{S_T}\left|K_1\left(b_{\theta_1,T}k\right)\right|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\frac{1}{Tb_{\theta_2,T}}\\
&\qquad\times\sum_{s=k+1}^{T}\left|K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{\theta_2,T}}\right)\right|\\
&\qquad\times\left\|V_s\left(\overline{\beta}\right)\frac{\partial}{\partial\beta}V_{s-k}\left(\overline{\beta}\right)+V_{s-k}\left(\overline{\beta}\right)\frac{\partial}{\partial\beta}V_s\left(\overline{\beta}\right)\right\|\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\|\\
&\le Cb_{\theta_2,T}^{-1}T^{1/3-1/2-1/3}S_T\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(CO_{\mathbb{P}}\left(1\right)\right)\\
&\qquad\times\left(\left(T^{-1}\sum_{s=1}^{T}\sup_{\beta\in\Theta}V_s^2\left(\beta\right)\right)^2\left(T^{-1}\sum_{s=1}^{T}\sup_{\beta\in\Theta}\left\|\frac{\partial}{\partial\beta}V_s\left(\beta\right)\right\|^2\right)^{1/2}\right)\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\|,
\end{aligned}
$$

where we have used the fact that $\widetilde{n}_T/T^{1/3}\to\infty$. Using Assumption 3.3 the right-hand side above is

$$
CT^{-1/2}b_{\theta_2,T}^{-1}S_T\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}O_{\mathbb{P}}\left(1\right)\xrightarrow{\mathbb{P}}0,
$$

since $r<2$. Next,

$$
\begin{aligned}
\left|H_{1,2,T}\right|&\le CT^{1/3-1/2}\sum_{k=S_T+1}^{T-1}\left(b_{\theta_1,T}k\right)^{-b}\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\frac{1}{Tb_{\theta_2,T}}\\
&\qquad\times\sum_{s=k+1}^{T}\left|K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{\theta_2,T}}\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\times \left\| V_s\left(\overline{\beta}\right)\frac{\partial}{\partial\beta}V_{s-k}\left(\overline{\beta}\right)+V_{s-k}\left(\overline{\beta}\right)\frac{\partial}{\partial\beta}V_s\left(\overline{\beta}\right)\right\|\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\| \\
&\leq Cb_{\theta_2,T}^{-1}T^{1/3-1/2-1/3}b_{\theta_1,T}^{-b}\sum_{k=S_T+1}^{T-1}k^{-b}\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}O_{\mathbb{P}}(1) \\
&\quad\times\left(\left(T^{-1}\sum_{s=1}^{T}\sup_{\beta\in\Theta}V_s^2(\beta)\right)^2\left(T^{-1}\sum_{s=1}^{T}\sup_{\beta\in\Theta}\left\|\frac{\partial}{\partial\beta}V_s(\beta)\right\|^2\right)^{1/2}\right)\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\| \\
&\leq Cb_{\theta_2,T}^{-1}T^{-1/2}b_{\theta_1,T}^{-b}\sum_{k=S_T+1}^{T-1}k^{-b}O_{\mathbb{P}}(1) \\
&\leq Cb_{\theta_2,T}^{-1}T^{-1/2}b_{\theta_1,T}^{-b}S_T^{1-b}O_{\mathbb{P}}(1) \\
&\leq Cb_{\theta_2,T}^{-1}T^{-1/2}b_{\theta_1,T}^{-b}b_{\theta_1,T}^{-r(1-b)}O_{\mathbb{P}}(1) \\
&\leq Cb_{\theta_2,T}^{-1}T^{-1/2}b_{\theta_1,T}^{-b}T^{r(1-b)/6}O_{\mathbb{P}}(1)\to 0,
\end{aligned}
$$

since $r>(b-2)/(b-1)$. This shows $H_{1,T}\xrightarrow{\mathbb{P}}0$. 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 $|K_1(\cdot)|\leq 1$ we have,

$$
\begin{aligned}
\mathbb{E}\left(H_{2,1,T}^2\right) &\leq CT^{2/3}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T}K_1\left(b_{\theta_1,T}k\right)K_1\left(b_{\theta_1,T}j\right)\left(\frac{n_T}{T}\right)^2\sum_{r_1=0}^{\lfloor T/n_T\rfloor}\sum_{r_2=0}^{\lfloor T/n_T\rfloor}\frac{1}{\left(Tb_{\theta_2,T}\right)^2} \qquad \text{(S.16)}\\
&\quad\times\sum_{s=k+1}^{T}\sum_{t=j+1}^{T}\left|K_2\left(\frac{\left(\left(r_1+1\right)n_T-\left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r_1+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{\theta_2,T}}\right)\right| \\
&\quad\times\left|K_2\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{b_{\theta_2,T}}\right)\right| \\
&\quad\times\left|\mathbb{E}\left(V_sV_{s-k}-\mathbb{E}\left(V_sV_{s-k}\right)\right)\left(V_tV_{t-k}-\mathbb{E}\left(V_tV_{t-k}\right)\right)\right| \\
&\leq CT^{2/3}S_T^2\widetilde{n}_T^{-2}\left(Tb_{\theta_2,T}\right)^{-1}\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)O_{\mathbb{P}}(1) \\
&\leq CT^{2/3-2/3-1+2r/6}O_{\mathbb{P}}\left(b_{\theta_2,T}^{-1}\right)\to 0,
\end{aligned}
$$

where we have used Lemma S.B.5 in Casini (2021) and $r<3$. Turning to $H_{2,2,T}$,

$$
\begin{aligned}
\mathbb{E}\left(H_{2,2,T}^2\right) &\leq CT^{2/3-2/3}\left(Tb_{\theta_2,T}\right)^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2}O(1)\right)^2 \qquad \text{(S.17)}\\
&\leq CT^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2}\right)^2 \\
&\leq CT^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}O(1)\right)^2 \\
&\leq CT^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}S_T^{2(1-b)}\to 0,
\end{aligned}
$$

since $r > (b-5/2)/(b-1)$. Eq. (S.16) and (S.17) yield $H_{2,T} \xrightarrow{\mathbb{P}} 0$. Let $H_{3,1,T}$ (resp. $H_{3,2,T}$) be defined as $H_{3,T}$ but with the sum over k be restricted to $k = 1, \ldots, S_T$ (resp. $k = S_T+1, \ldots, T$). Given $|K_1(\cdot)| \leq 1$ and (S.15), we have

$$|H_{3,1,T}| \leq CT^{1/3}\widetilde{n}_T^{-1} \sum_{k=1}^{S_T} |\Gamma_T(k)| \leq CT^{1/3}\widetilde{n}_T^{-1} \sum_{k=1}^{\infty} k^{-l} \to 0,$$

since $\sum_{k=1}^{\infty} k^{-l} < \infty$ for $l > 1$ and $\widetilde{n}_T/T^{1/3} \to \infty$. Finally,

$$\begin{aligned} |H_{3,2,T}| &\leq CT^{1/3}\widetilde{n}_T^{-1} \sum_{k=S_T+1}^{T-1} |\Gamma_T(k)| \leq CT^{1/3}\widetilde{n}_T^{-1} \sum_{k=S_T+1}^{T-1} k^{-l} \\ &\leq CT^{1/3}\widetilde{n}_T^{-1} S_T^{1-l} \leq CT^{1/3}\widetilde{n}_T^{-1} T^{r(1-l)/6} \to 0, \end{aligned}$$

since $l > 1$ and $\widetilde{n}_T/T^{1/3} \to \infty$. This completes the proof of part (ii).

We now move to part (i). For some $\phi_{1,\theta^*}, \phi_{2,\theta^*} \in (0, \infty)$, $\widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T}) - J_T = o_{\mathbb{P}}(1)$ by Theorem 3.2-(i) since $O(b_{\theta_1,T}) = O(b_{\theta_2,T})$ and $\sqrt{T}b_{1,T} \to \infty$ hold. Hence, it remains to show $\widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T}) - \widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T}) = o_{\mathbb{P}}(1)$. Note that this result differs from the result of part (ii) only because the scale factor $T^{1/3}$ does not appear, Assumption 4.1-(ii) is replaced by part (i) of the same assumption and Assumption 4.1-(iii) is not imposed. Let

$$r \in (\max\{(2b-5)/2(b-1)\}, 1),$$

with $b > 1 + 1/q$ and let S_T be defined as in part (ii). We will use the decomposition in (S.1), and N_1 and N_2 as defined after (S.1). Let $A_{1,T}$, $A_{2,T}$ and $A_{3,T}$ be as in (S.2) without the scale factor $T^{1/3}$. Proceeding as in (S.3),

$$\begin{aligned} |A_{1,1,T}| &\leq \sum_{k=1}^{S_T} C_2 \left|\widehat{b}_{1,T} - b_{\theta_1,T}\right| k \left|\widehat{\Gamma}(k)\right| \\ &\leq C \left|\widehat{\phi}_1{}^{1/24} - \phi_{1,\theta^*}^{1/24}\right| \left(\widehat{\phi}_1 \phi_{1,\theta^*}\right)^{-1/24} T^{-1/6} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k)\right|, \end{aligned} \tag{S.18}$$

for some $C < \infty$. By Assumption 4.1-(i),

$$\left|\widehat{\phi}_1{}^{1/24} - \phi_{1,\theta^*}^{1/24}\right| \left(\widehat{\phi}_1 \phi_{1,\theta^*}\right)^{-1/24} = O_{\mathbb{P}}(1).$$

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

$$\begin{aligned} B_{1,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right| \\ B_{2,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right| \\ B_{3,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\Gamma_T(k)\right|. \end{aligned} \tag{S.19}$$

By a mean-value expansion, we have

$$B_{1,T} \leq T^{-1/6}T^{-1/2}\sum_{k=1}^{S_T} k\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right)\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right| \tag{S.20}$$
$$\leq CT^{-1/6}T^{2r/6}T^{-1/2}\sup_{k\geq 1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right\|\sqrt{T}\left\|\widehat{\beta}-\beta_0\right\|,$$

since $r<2$, and $\sup_{k\geq 1}\|(\partial/\partial\beta)\widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\| = O_{\mathbb{P}}(1)$ using (S.28) in Casini (2021) and Assumption 3.3-(ii,iii). In addition,

$$\mathbb{E}\left(B_{2,T}^2\right) \leq \mathbb{E}\left(T^{-1/3}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T} kj\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\left|\widetilde{\Gamma}(j)-\Gamma_T(j)\right|\right) \tag{S.21}$$
$$\leq \mathbb{E}\left(T^{-1/3}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T} kj\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\left|\widetilde{\Gamma}(j)-\Gamma_T(j)\right|\right)$$
$$\leq T^{-1/3-5/6}S_T^4\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)$$
$$\leq T^{-1/3-5/6}T^{4r/6}\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\to 0,$$

given that $\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}(\widetilde{\Gamma}(k)) = O(1)$ by Lemma S.B.5 in Casini (2021) and $r<7/4$. The bound in equation (S.7) is replaced by,

$$B_{3,T} \leq T^{-1/6}S_T\sum_{k=1}^{\infty}|\Gamma_T(k)| \leq T^{(r-1)/6}O_{\mathbb{P}}(1)\to 0, \tag{S.22}$$

using Assumption 3.2-(i) since $r<1$. This gives $A_{1,T}\xrightarrow{\mathbb{P}}0$. Next, we show that $A_{2,T}\xrightarrow{\mathbb{P}}0$. As above, let $A_{2,1,T} = L_{1,T}+L_{2,T}+L_{3,T}$ where each summand is defined as in (S.8) without the factor $T^{1/3}$. We have

$$|L_{1,T}| = T^{-1/2}\sum_{k=S_T+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right| \tag{S.23}$$
$$= T^{-1/2+b/6}\sum_{k=S_T+1}^{T-1} C_1k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right|$$
$$= T^{-1/2+b/6+r(1-b)/6}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\right|$$
$$= T^{-1/2+b/6+r(1-b)/6}O(1)\,O_{\mathbb{P}}(1),$$

which converges to zero since $r>(b-3)/(b-1)$. The bound for $L_{2,T}$ is given by

$$|L_{2,T}| = \sum_{k=S_T+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right| \tag{S.24}$$
$$= C_1\widehat{\phi}_1^{-b/24}T^{b/6-1/2}b_{\theta_2,T}^{-1/2}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\right)\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|,$$

and the bound in (S.11) is replaced by,

$$\begin{aligned}
&\mathbb{E}\left(T^{b/6-1/2}b_{\theta_2,T}^{-1/2}\sum_{k=S_T}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left|\widetilde{\Gamma}(k)-\Gamma_T(k)\right|\right)^2 \\
&\leq T^{b/3-1}b_{\theta_2,T}^{-1}\left(\sum_{k=S_T}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2}\right)^2 \\
&= T^{b/3-1}b_{\theta_2,T}^{-1}\left(\sum_{k=S_T}^{T-1}k^{-b}\right)^2 O(1) \\
&= T^{b/3-1+1/6}S_T^{2(1-b)}O(1)\to 0,
\end{aligned} \tag{S.25}$$

since $r>(2b-5)/2(b-1)$ and $Tb_{2,T}\mathrm{Var}(\widetilde{\Gamma}(k))=O(1)$ as above. Equations (S.24)-(S.25) combine to yield $L_{2,T}\xrightarrow{\mathbb{P}}0$ since $\widehat{\phi}_1=O_{\mathbb{P}}(1)$. The bound for $L_{3,T}$ is given by

$$\begin{aligned}
\left|\sum_{k=S_T+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)\Gamma_T(k)\right| &\leq \sum_{k=S_T+1}^{T-1}\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left|c\left(rn_T/T,\,k\right)\right| \\
&\leq \sum_{k=S_T+1}^{T-1}\sup_{u\in[0,\,1]}\left|c(u,\,k)\right|\to 0.
\end{aligned} \tag{S.26}$$

Equations (S.23)-(S.26) imply $A_{2,1,T}\xrightarrow{\mathbb{P}}0$. Thus, as in the proof of part (ii), we have $A_{2,T}\xrightarrow{\mathbb{P}}0$ and $A_{3,T}\xrightarrow{\mathbb{P}}0$. It remains to show that $(\widehat{J}_T(b_{\theta_1,T},\,\widehat{b}_{2,T})-\widehat{J}_T(b_{\theta_1,T},\,b_{\theta_2,T}))\xrightarrow{\mathbb{P}}0$. Let $\widehat{c}_{\theta_2,T}(rn_T/T,\,k)$ be defined as in part (ii). We have (S.14) and (S.15) is replaced by

$$\begin{aligned}
&K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{\theta_2,T}}\right) \\
&\leq C_4\left|\frac{Tu-(s-k/2)}{T\widehat{b}_{2,T}}-\frac{Tu-(s-k/2)}{Tb_{\theta_2,T}}\right| \\
&\leq C_4T^{-1}\left|\frac{Tu-(s-k/2)\left(\widehat{b}_{2,T}-b_{\theta_2,T}\right)}{\widehat{b}_{2,T}b_{\theta_2,T}}\right| \\
&\leq C_4T^{-5/6}\left(\frac{1}{\widehat{\phi}_2\phi_{2,\theta^*}}\right)^{1/24}\left(\widehat{\phi}_2^{1/24}-\phi_{2,\theta^*}^{1/24}\right)(Tu-(s-k/2)),
\end{aligned} \tag{S.27}$$

for $s\in\{Tu-\lfloor Tb_{\theta_2,T}\rfloor,\ldots,Tu+\lfloor Tb_{\theta_2,T}\rfloor\}$. Therefore,

$$\begin{aligned}
&\widehat{J}_T\left(b_{\theta_1,T},\,\widehat{b}_{2,T}\right)-\widehat{J}_T\left(b_{\theta_1,T},\,b_{\theta_2,T}\right) \\
&=\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(\widehat{c}\left(rn_T/T,\,k\right)-\widehat{c}_{\theta_2,T}\left(rn_T/T,\,k\right)\right) \\
&\leq C\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)
\end{aligned} \tag{S.28}$$

$$
\begin{aligned}
&\times \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T b_{\theta_2,T}} \sum_{s=k+1}^{T} \\
&\left| K_2\left(\frac{\left((r+1)\, n_T - (s-k/2)\right)/T}{\widehat{b}_{2,T}}\right) - K_2\left(\frac{\left((r+1)\, n_T - (s-k/2)\right)/T}{b_{\theta_2,T}}\right)\right| \\
&\times \left(\left|\widehat{V}_s \widehat{V}_{s-k} - V_s V_{s-k}\right| + \left|V_s V_{s-k} - \mathbb{E}\left(V_s V_{s-k}\right)\right| + \left|\mathbb{E}\left(V_s V_{s-k}\right)\right|\right) \\
&\triangleq H_{1,T} + H_{2,T} + H_{3,T}.
\end{aligned}
$$

We have to show that $H_{1,T} + H_{2,T} + H_{3,T} \xrightarrow{\mathbb{P}} 0$. By a mean-value expansion, using (S.27),

$$
\begin{aligned}
\left|H_{1,T}\right| &\leq C T^{-1/2} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{\theta_1,T} k\right)\right| \\
&\quad \times \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T b_{\theta_2,T}} \sum_{s=k+1}^{T} \\
&\quad \left| K_2\left(\frac{\left((r+1)\, n_T - (s-k/2)\right)/T}{\widehat{b}_{2,T}}\right) - K_2\left(\frac{\left((r+1)\, n_T - (s-k/2)\right)/T}{b_{\theta_2,T}}\right)\right| \\
&\quad \times \left\| V_s\left(\overline{\beta}\right) \frac{\partial}{\partial \beta} V_{s-k}\left(\overline{\beta}\right) + V_{s-k}\left(\overline{\beta}\right) \frac{\partial}{\partial \beta} V_s\left(\overline{\beta}\right) \right\| \sqrt{T} \left\|\widehat{\beta} - \beta_0\right\| \\
&\leq C b_{\theta_2,T}^{-1} T^{-1/2} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{\theta_1,T} k\right)\right| \\
&\quad \times \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} C O_{\mathbb{P}}(1) \\
&\quad \times \left(\left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} V_s^2(\beta)\right)^2 \left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} V_s(\beta) \right\|^2\right)^{1/2}\right) \sqrt{T} \left\|\widehat{\beta} - \beta_0\right\|.
\end{aligned}
$$

Using Assumption 3.3 the right-hand side above is

$$
C T^{-1/2} b_{\theta_2,T}^{-1} b_{\theta_1,T}^{-1} b_{\theta_1,T} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{\theta_1,T} k\right)\right| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,
$$

since $T^{-1/2} b_{\theta_1,T}^{-1} b_{\theta_2,T}^{-1} \to 0$. This shows $H_{1,T} \xrightarrow{\mathbb{P}} 0$. 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. $k = S_T + 1, \ldots, T$). We have

$$
\begin{aligned}
\mathbb{E}\left(H_{2,1,T}^2\right) &\leq \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} K_1\left(b_{\theta_1,T} k\right) K_1\left(b_{\theta_1,T} j\right) \\
&\quad \times \left(\frac{n_T}{T}\right)^2 \sum_{r_1=0}^{\lfloor T/n_T \rfloor} \sum_{r_2=0}^{\lfloor T/n_T \rfloor} \frac{1}{\left(T b_{\theta_2,T}\right)^2} \sum_{s=k+1}^{T} \sum_{t=j+1}^{T} \\
&\quad \times \left| K_2\left(\frac{\left((r_1+1)\, n_T - (s-k/2)\right)/T}{\widehat{b}_{2,T}}\right) - K_2\left(\frac{\left((r_1+1)\, n_T - (s-k/2)\right)/T}{b_{\theta_2,T}}\right)\right|
\end{aligned}
\tag{S.29}
$$

$$
\begin{aligned}
&\times \left| K_2\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{\widehat{b}_{2,T}}\right)-K_2\left(\frac{\left(\left(r_2+1\right)n_T-\left(t-j/2\right)\right)/T}{b_{\theta_2,T}}\right)\right| \\
&\times \left|\left(V_sV_{s-k}-\mathbb{E}\left(V_sV_{s-k}\right)\right)\left(V_tV_{t-k}-\mathbb{E}\left(V_tV_{t-k}\right)\right)\right| \\
&\leq CS_T^2\left(Tb_{\theta_2,T}\right)^{-1}\sup_{k\geq 1}Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)O_{\mathbb{P}}\left(1\right) \\
&\leq CT^{r/3}O_{\mathbb{P}}\left(T^{-1}b_{\theta_2,T}^{-1}\right)\to 0,
\end{aligned}
$$

where we have used Lemma S.B.5 in Casini (2021), (S.27) and $r<5/2$. Turning to $H_{2,2,T}$,

$$
\begin{aligned}
\mathbb{E}\left(H_{2,2,T}^2\right) &\leq \left(Tb_{\theta_2,T}\right)^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2}O\left(1\right)\right)^2 \\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^2 \\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1}k^{-b}O\left(1\right)\right)^2 \\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}S_T^{2(1-b)}\to 0,
\end{aligned}
\tag{S.30}
$$

since $r>\left(2b-5\right)/2\left(b-1\right)$. Eq. (S.29) and (S.30) yield $H_{2,T}\xrightarrow{\mathbb{P}}0$. Given $\left|K_1\left(\cdot\right)\right|\leq 1$ and (S.27), we have

$$
\left|H_{3,T}\right|\leq C\sum_{k=-\infty}^{\infty}\left|\Gamma_T\left(k\right)\right|o_{\mathbb{P}}\left(1\right)\to 0.
$$

This concludes the proof of part (i).

The result of part (iii) follows from the same argument as in Theorem 3.2-(iii) with references to Theorem 3.2-(i,ii) changed to Theorem 4.1-(i,ii). □

S.A.3 Proof of the Results in Section 5

S.A.3.1 Proof of Theorem 5.1

We begin with the following lemma which extends Theorem 1 in Andrews (1991) to the present setting. Let $\widetilde{J}_{\mathrm{Cla},T}$ denote the estimator that uses $\{V_t(\beta_0)\}$ in place of $\{\widehat{V}_t\}$ and let $\widetilde{J}_{\mathrm{Cla},T}\left(\beta\right)$ denote the estimator calculated using $\{V_t(\beta)\}$. Let $b_{\mathrm{Cla},\theta_1,T}=\left(qK_{1,q}^2\alpha_\theta T/\int K_1^2\left(y\right)dy\right)^{-1/(2q+1)}$ where $\alpha_\theta\in(0,\,\infty)$.

Lemma S.A.1. *Suppose $K_1\left(\cdot\right)\in\boldsymbol{K}_1$ and $b_{1,T}\to 0$.*

(i) If Assumption 5.1-5.2 hold and $T^{\vartheta}b_{1,T}\to\infty$, then $\widehat{J}_{\mathrm{Cla},T}-J_T\xrightarrow{\mathbb{P}}0$ and $\widehat{J}_{\mathrm{Cla},T}-\widetilde{J}_{\mathrm{Cla},T}\xrightarrow{\mathbb{P}}0$.

(ii) If Assumption 5.2-5.3 hold, $T^{1/2-2\vartheta}b_{1,T}^{-1/2}\to 0$, $T^{-2\vartheta}b_{1,T}^{-1}\to 0$ and $Tb_{1,T}^{2q+1}\to\gamma\in(0,\,\infty)$ for some $q\in(0,\,\infty)$ for which $K_{1,q}$, $||f^{(q)}||<\infty$, then $\sqrt{Tb_{1,T}}(\widehat{J}_{\mathrm{Cla},T}-J_T)=O_{\mathbb{P}}\left(1\right)$ and $\sqrt{Tb_{1,T}}(\widehat{J}_{\mathrm{Cla},T}-\widetilde{J}_{\mathrm{Cla},T})\xrightarrow{\mathbb{P}}0$.

(iii) Under the conditions of part (b),

$$
\lim_{T\to\infty}\mathrm{MSE}\left(Tb_{1,T},\,\widehat{J}_{\mathrm{Cla},T},\,W\right)=\lim_{T\to\infty}\mathrm{MSE}\left(Tb_{1,T},\,\widetilde{J}_{\mathrm{Cla},T},\,W\right)
$$

$$= 4\pi^2 \left(\gamma K_{1,q}^2 \left(\mathrm{vec} f^{(q)} \right)' W \mathrm{vec} f^{(q)} + \int K_1^2 (y)\, dy\, \mathrm{tr} W \left(I + C_{pp} \right) f \otimes f \right).$$

Proof of Lemma S.A.1. Using the same arguments as in Theorem 1 of Andrews (1991), we have $\widetilde{J}_{\mathrm{Cla},T} - J_T = o_{\mathbb{P}}(1)$ and $\sqrt{Tb_{1,T}}(\widetilde{J}_{\mathrm{Cla},T} - J_T) = O_{\mathbb{P}}(1)$. Thus, Lemma S.A.1-(i,ii) hold if the second result stated in each of these parts holds. The latter hold if and only if they hold with $\widehat{J}_{\mathrm{Cla},T} - \widetilde{J}_{\mathrm{Cla},T}$ replaced by $a'\widehat{J}_{\mathrm{Cla},T}a - a'\widetilde{J}_{\mathrm{Cla},T}a$ for arbitrary $a \in \mathbb{R}^p$. Thus, it is sufficient to consider the scalar case. We now show that $T^{\vartheta} b_{1,T}(\widehat{J}_{\mathrm{Cla},T} - \widetilde{J}_{\mathrm{Cla},T}) = O_{\mathbb{P}}(1)$ provided $b_{1,T} \to 0$ and Assumption 5.2 holds. This yields the second result of Lemma S.A.1-(i). A mean-value expansion of $\widetilde{J}_{\mathrm{Cla},T}(\widehat{\beta}_{\mathrm{np}})(= \widehat{J}_{\mathrm{Cla},T})$ about β_0 yields

$$\begin{aligned} T^{\vartheta} b_{1,T} \left(\widehat{J}_{\mathrm{Cla},T} - \widetilde{J}_{\mathrm{Cla},T} \right) &= b_{1,T} \frac{\partial}{\partial \beta'} \widetilde{J}_{\mathrm{Cla},T} \left(\bar{\beta} \right) T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right) \\ &= b_{1,T} \sum_{k=-T+1}^{T-1} K_1 \left(b_{1,T} k \right) \frac{\partial}{\partial \beta'} \widehat{\Gamma}_{\mathrm{Cla}}(k) |_{\beta = \bar{\beta}} T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right). \end{aligned} \tag{S.31}$$

Andrews (1991) showed that

$$\sup_{k \geq 1} \left\| \frac{\partial}{\partial \beta'} \widehat{\Gamma}_{\mathrm{Cla}}(k) \right\| |_{\beta = \bar{\beta}} = O_{\mathbb{P}}(1). \tag{S.32}$$

This result, Assumption 5.2-(i), and the fact that $b_{1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{1,T}k)| \to \int_{-\infty}^{\infty} |K_1(y)|\, dy < \infty$ imply that the right-hand side of (S.31) is $O_{\mathbb{P}}(1)$ and Lemma S.A.1-(i) follows because $T^{\vartheta} b_{1,T} \to \infty$.

Next we show that $\sqrt{Tb_{1,T}}(\widehat{J}_{\mathrm{Cla},T} - \widetilde{J}_{\mathrm{Cla},T}) = o_{\mathbb{P}}(1)$ under the assumptions of Lemma S.A.1-(ii). A second-order Taylor expansion gives

$$\begin{aligned} &\sqrt{Tb_{1,T}} \left(\widehat{J}_{\mathrm{Cla},T} - \widetilde{J}_{\mathrm{Cla},T} \right) \\ &\quad = \left[\sqrt{b_{1,T}} \frac{\partial}{\partial \beta'} \widetilde{J}_{\mathrm{Cla},T}(\beta_0) \right] T^{1/2-\vartheta} T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right) \\ &\qquad + \frac{1}{2} \sqrt{T} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right)' \left[\sqrt{b_{1,T}} \frac{\partial^2}{\partial \theta \partial \theta'} \widetilde{J}_T \left(\overline{\beta} \right) / \sqrt{T} \right] \sqrt{T} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right) \\ &\quad = G_T T^{1/2-\vartheta} T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right) + \frac{1}{2} T^{1/2-\vartheta} T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right)' H_T T^{1/2-\vartheta} T^{\vartheta} \left(\widehat{\beta}_{\mathrm{np}} - \beta_0 \right), \end{aligned}$$

where G_T $(\in \mathbb{R}^p)$ and H_T $(\in \mathbb{R}^p)$ are defined implicitly. Assumption 5.2-(ii,iii), 5.3-(ii) and simple manipulations yield

$$\begin{aligned} T^{1-2\vartheta} \|H_T\| &= T^{(1-2\vartheta)} \left(\frac{b_{1,T}}{T} \right)^{1/2} \sum_{k=-T+1}^{T-1} |K_1(b_{1,T}k)| \frac{1}{T} \sum_{t=|k|+1}^{T} \sup_{\beta \in \Theta} \left\| \frac{\partial^2}{\partial \beta \partial \beta'} V_t(\beta) V_{t-|k|}(\beta) \right\| \\ &= T^{(1-2\vartheta)} \left(\frac{1}{Tb_{1,T}} \right)^{1/2} \left(b_{1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{1,T}k)| \right) O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1), \end{aligned} \tag{S.33}$$

since $T^{1/2-2\vartheta} b_{1,T}^{-1/2} \to 0$. Andrews (1991) showed that $G_T = o_{\mathbb{P}}(1)$. Thus, we have to show that

$T^{1/2-\vartheta}G_T = o_{\mathbb{P}}(1)$. Let

$$D_T = \sqrt{b_{1,T}} \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k) \frac{1}{T} \sum_{t=|k|+1}^{T} \left(V_t + V_{t-|k|}\right).$$

Using eq. (A.13) in Andrews (1991) we have

$$\begin{aligned} T^{1-2\vartheta}\mathbb{E}\left(D_T^2\right) &\leq T^{1-2\vartheta} b_{1,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} \left|K_1(b_{1,T}k) K_1(b_{1,T}j)\right| \frac{4}{T^2} \sum_{s=1}^{T}\sum_{t=1}^{T} \left|\mathbb{E}(V_s V_t)\right| \\ &\leq T^{1-2\vartheta} \frac{1}{Tb_{1,T}} \left(b_{1,T} \sum_{k=-T+1}^{T-1} \left|K_1(b_{1,T}k)\right| \right)^2 \sum_{u=-T+1}^{T-1} \left|\Gamma(u)\right| \\ &= T^{1-2\vartheta} \frac{1}{Tb_{1,T}} O_{\mathbb{P}}(1), \end{aligned}$$

since $T^{-2\vartheta} b_{1,T}^{-1} \to 0$. This concludes the proof of part (ii). The proof of part (iii) of the lemma follows the same argument as in the corresponding proof in Andrews (1991). □

Proof of Theorem 5.1. By Lemma S.A.1-(i) $\widehat{J}_{\mathrm{Cla},T}(b_{\mathrm{Cla},\theta_1,T}) - J_T = o_{\mathbb{P}}(1)$, since $q > (1/\vartheta - 1)/2$ implies $T^{\vartheta} b_{1,T} \to \infty$. Hence, it suffices to show that $\widehat{J}_{\mathrm{Cla},T}(\widehat{b}_{\mathrm{Cla},1,T}) - \widehat{J}_{\mathrm{Cla},T}(b_{\mathrm{Cla},\theta_1,T}) = o_{\mathbb{P}}(1)$. Let $S_T = \lfloor (b_{\mathrm{Cla},\theta_1,T})^{-r} \rfloor$ with

$$r \in \left(\max\left((b-q-1/2)/(b-1),\, 1-(2q-1)/(2b-2)\right),\, \min\left(1,\, 3/4+q/2\right)\right).$$

We have

$$\begin{aligned} \widehat{J}_{\mathrm{Cla},T}(\widehat{b}_{\mathrm{Cla},1,T}) - \widehat{J}_{\mathrm{Cla},T}(b_{\mathrm{Cla},\theta_1,T}) &= 2\sum_{k=1}^{S_T} \left(K_1\left(\widehat{b}_{\mathrm{Cla},1,T}k\right) - K_1\left(b_{\mathrm{Cla},\theta_1,T}k\right)\right) \widehat{\Gamma}_{\mathrm{Cla}}(k) \\ &\quad + 2\sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{\mathrm{Cla},1,T}k\right) \widehat{\Gamma}_{\mathrm{Cla}}(k) \\ &\quad - 2\sum_{k=S_T+1}^{T-1} K_1\left(b_{\mathrm{Cla},\theta_1,T}k\right) \widehat{\Gamma}_{\mathrm{Cla}}(k) \\ &= 2A_{1,T} + 2A_{2,T} - 2A_{3,T}. \end{aligned} \tag{S.34}$$

We show $A_{1,T} \xrightarrow{\mathbb{P}} 0$ as follows. Using the Lipschitz condition on $K_1(\cdot)$,

$$\begin{aligned} |A_{1,T}| &\leq \sum_{k=1}^{S_T} C_2 \left|\widehat{b}_{\mathrm{Cla},1,T} - b_{\mathrm{Cla},\theta_1,T}\right| k \left|\widehat{\Gamma}_{\mathrm{Cla}}(k)\right| \\ &\leq C \left|\widehat{\alpha}(q)^{1/(2q+1)} - \alpha_\theta^{1/(2q+1)}\right| \left(\widehat{\alpha}(q)\alpha_\theta\right)^{-1/(2q+1)} T^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}_{\mathrm{Cla}}(k)\right|, \end{aligned} \tag{S.35}$$

for some constant $C < \infty$. By Assumption 5.4,

$$\left|\widehat{\alpha}(q)^{1/(2q+1)} - \alpha_\theta^{1/(2q+1)}\right| \left(\widehat{\alpha}(q)\alpha_\theta\right)^{-1/(2q+1)} = O_{\mathbb{P}}(1),$$

and so it suffices to show that $B_{1,T}+B_{2,T}+B_{3,T}\overset{\mathbb{P}}{\to}0$, where

$$B_{1,T}=T^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\widehat{\Gamma}_{\text{Cla}}(k)-\widetilde{\Gamma}_{\text{Cla}}(k)\right| \tag{S.36}$$
$$B_{2,T}=T^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\widetilde{\Gamma}_{\text{Cla}}(k)-\Gamma_T(k)\right|$$
$$B_{3,T}=T^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\Gamma_T(k)\right|.$$

By a mean-value expansion, we have

$$\begin{aligned}B_{1,T}&\leq T^{-1/(2q+1)-\vartheta}S_T\sum_{k=1}^{S_T}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}_{\text{Cla}}(k)\,|_{\beta=\overline{\beta}}\right)T^{\vartheta}\left(\widehat{\beta}_{\text{np}}-\beta_0\right)\right| \\ &\leq CT^{-1/(2q+1)-\vartheta+2r/(2q+1)}\sup_{k\geq1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}_{\text{Cla}}(k)\,|_{\beta=\overline{\beta}}\right\|T^{\vartheta}\left\|\widehat{\beta}_{\text{np}}-\beta_0\right\|\overset{\mathbb{P}}{\to}0,\end{aligned} \tag{S.37}$$

since $r<3/4+q/2$, $T^{\vartheta}||\widehat{\beta}_{\text{np}}-\beta_0||=O_{\mathbb{P}}(1)$ by Assumption 5.2, and $\sup_{k\geq1}||(\partial/\partial\beta)\widehat{\Gamma}_{\text{Cla}}(k)\,|_{\beta=\overline{\beta}}||=O_{\mathbb{P}}(1)$ by (S.32) and Assumption 5.2-(ii,iii). Andrews (1991) showed that $\mathbb{E}\left(B_{2,T}^2\right)\to0$ if $r<3/4+q/2$ and $B_{3,T}\to0$ if $r<1$. Altogether, this yields $A_{1,T}\overset{\mathbb{P}}{\to}0$. Let $A_{2,T}=L_{1,T}+L_{2,T}+L_{3,T}$, where

$$L_{1,T}=\sum_{k=S_T+1}^{T-1}K_1\left(\widehat{b}_{\text{Cla},1,T}k\right)\left(\widehat{\Gamma}_{\text{Cla}}(k)-\widetilde{\Gamma}_{\text{Cla}}(k)\right)$$
$$L_{2,T}=\sum_{k=S_T+1}^{T-1}K_1\left(\widehat{b}_{\text{Cla},1,T}k\right)\left(\widetilde{\Gamma}_{\text{Cla}}(k)-\Gamma_T(k)\right)$$
$$L_{3,T}=\sum_{k=S_T+1}^{T-1}K_1\left(\widehat{b}_{\text{Cla},1,T}k\right)\Gamma_T(k).$$

We now show $A_{2,T}\overset{\mathbb{P}}{\to}0$. By a mean-value expansion and the definition of $\boldsymbol{K}_{3,\text{Cla}}$,

$$\begin{aligned}|L_{1,T}|&=T^{-\vartheta}\sum_{k=S_T+1}^{T-1}C_1\left(\widehat{b}_{\text{Cla},1,T}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}_{\text{Cla}}(k)\right)|_{\beta=\overline{\beta}}T^{\vartheta}\left(\widehat{\beta}_{\text{np}}-\beta_0\right)\right| \\ &=T^{-\vartheta+b/(2q+1)}\left(\sum_{k=S_T+1}^{\infty}k^{-b}\right)O_{\mathbb{P}}(1) \\ &=T^{-\vartheta+b/(2q+1)-(b-1)r/(2q+1)}O_{\mathbb{P}}(1)\to0,\end{aligned} \tag{S.38}$$

where the second equality uses (S.32) and Assumption 5.2, and the convergence to zero follows from $r>(b-q-1/2)/(b-1)$. Further, Andrews (1991) showed that $L_{2,T}=o_{\mathbb{P}}(1)$ if $r>1-(2q-1)/(2b-2)$. Using $|K_1(\cdot)|\leq1$ we have $L_{3,T}\leq\sum_{k=S_T+1}^{T-1}|\Gamma(k)|\to0$. Thus, $A_{2,T}\overset{\mathbb{P}}{\to}0$. and an analogous argument yields $A_{3,T}\overset{\mathbb{P}}{\to}0$. Combined with $A_{1,T}\overset{\mathbb{P}}{\to}0$, the proof of Theorem 5.1 is completed. □

S.A.3.2 Proof of Theorem 5.2

We begin with the following lemma which extends Theorem 3.2 to the present setting.

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

(i) If Assumption 3.1-3.2 and 5.5 hold, $T^{\vartheta}b_{1,T} \to \infty$, $b_{2,T}/b_{1,T} \to \nu \in [0, \infty)$ *then* $\widehat{J}_T - J_T \xrightarrow{\mathbb{P}} 0$ *and* $\widehat{J}_T - \widetilde{J}_T \xrightarrow{\mathbb{P}} 0$.

(ii) If Assumption 3.1, 3.3-3.4 hold, $T^{1/2-2\vartheta}b_{1,T}^{-1/2} \to 0$, $T^{-2\vartheta}(b_{1,T}b_{2,T})^{-1} \to 0$, $n_T/Tb_{1,T}^q \to 0$, $1/Tb_{1,T}^q b_{2,T} \to 0$, $b_{2,T}^2/b_{1,T}^q \to \nu \in [0, \infty)$ *and* $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0, \infty)$ *for some* $q \in [0, \infty)$ *for which* $K_{1,q}, \|\int_0^1 f^{(q)}(u, 0)\, du\| \in [0, \infty)$, *then* $\sqrt{Tb_{1,T}b_{2,T}}(\widehat{J}_T - J_T) = O_{\mathbb{P}}(1)$ *and* $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$.

(iii) Under the conditions of part (ii) with $\nu \in (0, \infty)$,

$$\lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T}, \widehat{J}_T, W_T\right) = \lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T}, \widetilde{J}_T, W\right).$$

Proof of Lemma S.A.2. As in Theorem 3.2 $\widetilde{J}_T - J_T = o_{\mathbb{P}}(1)$. Proceeding as in Theorem 3.2-(ii), we first show that $T^{\vartheta}b_{1,T}(\widehat{J}_T - \widetilde{J}_T) = O_{\mathbb{P}}(1)$ under Assumption 5.5. A mean-value expansion of $\widehat{J}_T$ about β_0 yields

$$\begin{aligned} T^{\vartheta}b_{1,T}(\widehat{J}_T - \widetilde{J}_T) &= b_{1,T}\frac{\partial}{\partial\beta'}\widetilde{J}_T(\bar{\beta})T^{\vartheta}(\widehat{\beta}_{\mathrm{np}} - \beta_0) \\ &= b_{1,T}\sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)|_{\beta=\bar{\beta}}T^{\vartheta}(\widehat{\beta}_{\mathrm{np}} - \beta_0). \end{aligned} \tag{S.39}$$

Using (S.28) in Casini (2021) we have

$$\begin{aligned} & b_{1,T}\sum_{k=T+1}^{T-1} K_1(b_{1,T}k)\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)|_{\beta=\bar{\beta}}T^{\vartheta}\left(\widehat{\beta}_{\mathrm{np}} - \beta_0\right) \\ &\le b_{1,T}\sum_{k=-T+1}^{T-1} |K_1(b_{1,T}k)|\, O_{\mathbb{P}}(1) \\ &= O_{\mathbb{P}}(1), \end{aligned}$$

where the last equality uses $b_{1,T}\sum_{k=-T+1}^{T-1}|K_1(b_{1,T}k)| \to \int |K_1(y)|\, dy < \infty$. This concludes the proof of part (i) of Lemma S.A.2 because $T^{\vartheta}b_{1,T} \to \infty$. The next step is to show that $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$ under the assumptions of Lemma S.A.2-(ii). A second-order Taylor expansion gives

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

Using Assumption 3.4-(ii), Casini (2021) showed that

$$\left\| \frac{\partial^2}{\partial\beta\partial\beta'} \widehat{c}\left(rn_T/T,\, k\right) \right\|\Bigg|_{\beta=\bar{\beta}} = O_{\mathbb{P}}\left(1\right),$$

and thus,

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

since $T^{1/2-2\vartheta}b_{1,T}^{-1/2} \to 0$. Next, we want to show that $T^{1/2-\vartheta}G_T = o_{\mathbb{P}}\left(1\right)$. Following Casini (2021), it is sufficient to prove $\mathbb{E}\left(A_3^2\right) \to 0$ where

$$\begin{aligned} A_3 = T^{1/2-\vartheta}\sqrt{b_{1,T}} &\sum_{k=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)\right| \frac{n_T}{T} \sum_{r=0}^{T/n_T} \frac{1}{Tb_{2,T}} \\ &\times \sum_{s=k+1}^{T} \left| K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s-k/2\right)\right)/T}{b_{2,T}} \right) \right| \left|\left(V_s + V_{s-k}\right)\right|. \end{aligned}$$

Using the same steps as in Casini (2021),

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

since $T^{-2\vartheta}\left(b_{1,T}b_{2,T}\right)^{-1} \to 0$. This implies $G_T = o_{\mathbb{P}}\left(1\right)$. It follows that $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}\left(1\right)$ which concludes the proof of part (ii) because $\sqrt{Tb_{1,T}b_{2,T}}(\widetilde{J}_T - J_T) = O_{\mathbb{P}}\left(1\right)$ by Theorem 3.1-(iii). Part (iii) follows from the same argument used in the proof of Theorem 3.2-(iii) in Casini (2021). □

Proof of Theorem 5.2. Without loss of generality, we assume that V_t is a scalar. The constant $C < \infty$ may vary from line to line. By Lemma S.A.2-(i), $\widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T}) - J_T = o_{\mathbb{P}}\left(1\right)$. It remains to establish $\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T}) = o_{\mathbb{P}}\left(1\right)$. Let

$$r \in \left(\max\left\{\left(2b-5\right)/2\left(b-1\right),\, \left(b-6\vartheta\right)/\left(b-1\right)\right\},\, \min\left(\left(3+6\vartheta\right)/2,\, 7/4\right)\right),$$

and $S_T = \lfloor b_{\theta_1,T}^{-r} \rfloor$. We will use the decomposition (S.1) and N_1 and N_2 as defined after (S.1). Let us consider the first term on the right-hand side of (S.1),

$$\begin{aligned} &\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{b}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, \widehat{b}_{2,T}) \\ &\quad = \sum_{k\in N_1} \left(K_1(\widehat{b}_{1,T}k) - K_1(b_{\theta_1,T}k)\right)\widehat{\Gamma}\left(k\right) + \sum_{k\in N_2} K_1(\widehat{b}_{1,T}k)\widehat{\Gamma}\left(k\right) \end{aligned} \tag{S.40}$$

$$
\begin{aligned}
&- \sum_{k \in N_2} K_1(b_{\theta_1,T} k) \widehat{\Gamma}(k) \\
&\triangleq A_{1,T} + A_{2,T} - A_{3,T}.
\end{aligned}
$$

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 Lipschitz condition on $K_1(\cdot) \in \boldsymbol{K}_3$ to yield,

$$
\begin{aligned}
|A_{1,1,T}| &\leq \sum_{k=1}^{S_T} C_1 \left|\widehat{b}_{1,T} - b_{\theta_1,T}\right| k \left|\widehat{\Gamma}(k)\right| \\
&\leq C \left|\widehat{\phi}_1^{1/24} - \phi_{1,\theta^*}^{1/24}\right| \left(\widehat{\phi}_1 \phi_{1,\theta^*}\right)^{-1/24} T^{-1/6} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k)\right|,
\end{aligned} \tag{S.41}
$$

for some $C < \infty$. By Assumption 4.1-(i) ($|\widehat{\phi}_1^{1/24} - \phi_{1,\theta^*}^{1/24}| \left(\widehat{\phi}_1 \phi_{1,\theta^*}\right)^{-1/24} = O_{\mathbb{P}}(1)$) and so it suffices to show that $B_{1,T} + B_{2,T} + B_{3,T} \xrightarrow{\mathbb{P}} 0$ where

$$
\begin{aligned}
B_{1,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right|, \\
B_{2,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|, \qquad \text{and} \\
B_{3,T} &= T^{-1/6} \sum_{k=1}^{S_T} k \left|\Gamma_T(k)\right|.
\end{aligned} \tag{S.42}
$$

By a mean-value expansion, we have

$$
\begin{aligned}
B_{1,T} &\leq T^{-\vartheta-1/6} \sum_{k=1}^{S_T} k \left|\left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right) T^{\vartheta} \left(\widehat{\beta}_{\text{np}} - \beta_0\right)\right| \\
&\leq C T^{-\vartheta-1/6} T^{2r/6} \sup_{k \geq 1} \left\|\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right\| T^{\vartheta} \left\|\widehat{\beta}_{\text{np}} - \beta_0\right\| \\
&\leq C T^{-\vartheta-1/6+r/3} \sup_{k \geq 1} \left\|\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k)|_{\beta=\overline{\beta}}\right\| O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,
\end{aligned} \tag{S.43}
$$

since $r < (3 + 6\vartheta)/2$, $\sqrt{T}||\widehat{\beta}_{\text{np}} - \beta_0|| = O_{\mathbb{P}}(1)$, and $\sup_{k \geq 1} || (\partial/\partial \beta') \widehat{\Gamma}(k)|_{\beta=\overline{\beta}}|| = O_{\mathbb{P}}(1)$ using (S.28) in Casini (2021) and Assumption 5.5-(i-iii). In addition,

$$
\begin{aligned}
\mathbb{E}\left(B_{2,T}^2\right) &\leq \mathbb{E}\left(T^{-1/3} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} kj \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right| \left|\widetilde{\Gamma}(j) - \Gamma_T(j)\right|\right) \\
&\leq b_{\theta_2,T}^{-1} T^{-1/3-1} S_T^4 \sup_{k \geq 1} T b_{\theta_2,T} \text{Var}\left(\widetilde{\Gamma}(k)\right) \\
&\leq b_{\theta_2,T}^{-1} T^{-4/3} T^{4r/6} \sup_{k \geq 1} T b_{\theta_2,T} \text{Var}\left(\widetilde{\Gamma}(k)\right) \to 0,
\end{aligned} \tag{S.44}
$$

given that $\sup_{k \geq 1} T b_{\theta_2,T} \text{Var}(\widetilde{\Gamma}(k)) = O(1)$ using Lemma S.B.5 in Casini (2021) and $r < 7/4$. Assumption

4.1-(iii) and $\sum_{k=1}^{\infty} k^{1-l} < \infty$ for $l > 2$ yield,

$$B_{3,T} \leq T^{-1/6} C_3 \sum_{k=1}^{\infty} k^{1-l} \to 0. \tag{S.45}$$

Combining (S.41)-(S.45) we deduce that $A_{1,1,T} \xrightarrow{\mathbb{P}} 0$. The same argument applied to $A_{1,T}$ where the summation now extends over negative integers k gives $A_{1,T} \xrightarrow{\mathbb{P}} 0$. Next, we show that $A_{2,T} \xrightarrow{\mathbb{P}} 0$. Again, we use the notation $A_{2,1,T}$ (resp., $A_{2,2,T}$) to denote $A_{2,T}$ with the summation over positive (resp., negative) integers. Let $A_{2,1,T} = L_{1,T} + L_{2,T} + L_{3,T}$, where

$$\begin{aligned} L_{1,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T} k\right)\left(\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right), \\ L_{2,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T} k\right)\left(\widetilde{\Gamma}(k) - \Gamma_T(k)\right), \quad \text{and} \\ L_{3,T} &= \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T} k\right)\Gamma_T(k). \end{aligned} \tag{S.46}$$

We have

$$\begin{aligned} |L_{1,T}| &= T^{-\vartheta} \sum_{k=S_T+1}^{T-1} C_1\left(\widehat{b}_{1,T} k\right)^{-b} \left|\left(\frac{\partial}{\partial \beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}} T^{\vartheta}\left(\widehat{\beta}_{\text{np}} - \beta_0\right)\right| \\ &= T^{-\vartheta+b/6} \sum_{k=S_T+1}^{T-1} C_1 k^{-b} \left|\left(\frac{\partial}{\partial \beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}} T^{\vartheta}\left(\widehat{\beta}_{\text{np}} - \beta_0\right)\right| \\ &= T^{-\vartheta+b/6+r(1-b)/6} \left|\left(\frac{\partial}{\partial \beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}} T^{\vartheta}\left(\widehat{\beta}_{\text{np}} - \beta_0\right)\right| \\ &= T^{-\vartheta+b/6+r(1-b)/6} O(1)\, O_{\mathbb{P}}(1), \end{aligned} \tag{S.47}$$

which converges to zero since $r > (b - 6\vartheta)/(b-1)$. The bound for $L_{2,T}$ remains the same as in the proof of Theorem 4.1-(i) since $r > (2b-5)/2(b-1)$. Similarly, $L_{3,T} \to 0$ as in the proof of the aforementioned theorem. Altogether, we have $A_{2,T} \xrightarrow{\mathbb{P}} 0$ and using the same steps $A_{3,T} \xrightarrow{\mathbb{P}} 0$. It remains to show that $(\widehat{J}_T(b_{\theta_1,T}, \widehat{b}_{2,T}) - \widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. The proof is different from the proof of the same result in Theorem (4.1)-(i) because Assumption 5.5 replaces Assumption 3.3. We have (S.27) and we have to show $H_{1,T} + H_{2,T} + H_{3,T} \xrightarrow{\mathbb{P}} 0$, where $H_{i,T}$ $(i = 1, 2, 3)$ is defined in (S.28). We have

$$\begin{aligned} |H_{1,T}| \leq{}& CT^{-\vartheta} \sum_{k=-T+1}^{T-1} |K_1(b_{\theta_1,T} k)| \\ &\times \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T b_{\theta_2,T}} \sum_{s=k+1}^{T} \left| K_2\left(\frac{((r+1)n_T - (s-k/2))/T}{\widehat{b}_{2,T}}\right) - K_2\left(\frac{((r+1)n_T - (s-k/2))/T}{b_{\theta_2,T}}\right)\right| \\ &\times \left\| V_s\left(\overline{\beta}\right)\frac{\partial}{\partial \beta} V_{s-k}\left(\overline{\beta}\right) + V_{s-k}\left(\overline{\beta}\right)\frac{\partial}{\partial \beta} V_s\left(\overline{\beta}\right)\right\| T^{\vartheta} \left\|\widehat{\beta}_{\text{np}} - \beta_0\right\| \end{aligned}$$

$$\leq C b_{\theta_2,T}^{-1} T^{-\vartheta} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{\theta_1,T} k\right)\right|$$
$$\times \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \left(C O_{\mathbb{P}}(1)\right) \left(\left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} V_s^2(\beta) \right)^2 \left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} V_s(\beta) \right\|^2 \right)^{1/2} \right) T^{\vartheta} \left\| \widehat{\beta}_{\mathrm{np}} - \beta_0 \right\|.$$

Using Assumption 3.3 the right-hand side above is

$$C T^{-\vartheta} b_{\theta_2,T}^{-1} b_{\theta_1,T}^{-1} b_{\theta_1,T} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{\theta_1,T} k\right)\right| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,$$

since $T^{-\vartheta} b_{\theta_1,T}^{-1} b_{\theta_2,T}^{-1} \to 0$. This shows $H_{1,T} \xrightarrow{\mathbb{P}} 0$. The proof of $H_{2,T} + H_{3,T} \xrightarrow{\mathbb{P}} 0$ remains the same as that of Theorem 4.1-(i) because it does not depend on $\widehat{\beta}_{\mathrm{np}}$. □