Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models*

Alessandro Casini† / University of Rome Tor Vergata

26th July 2021

First Version: November 2018

Abstract

The literature on heteroskedasticity and autocorrelation robust (HAR) inference is extensive but its usefulness relies on stationarity of the relevant process, say V_t, usually a function of the data and estimated model residuals. Yet, a large body of work shows widespread evidence of various forms of nonstationarity in the latter. Also, many testing problems are such that V_t is stationary under the null hypothesis but nonstationary under the alternative. In either case, the consequences are possible size distortions and, especially, a reduction in power which can be substantial (e.g., non-monotonic power), since all such estimates are based on weighted sums of the sample autocovariances of V_t, which are inflated. We propose HAR inference methods valid under a broad class of nonstationary processes, labelled Segmented Local Stationary, which possess a spectrum that varies both over frequencies and time. It is allowed to change either slowly and continuously and/or abruptly at some time points, thereby encompassing most nonstationary models used in applied work. We introduce a double kernel estimator (DK-HAC) that applies a smoothing over both lagged autocovariances and time. The optimal kernels and bandwidth sequences are derived under a mean-squared error criterion. The data-dependent bandwidths rely on the "plug-in" approach using approximating parametric models having time-varying parameters estimated with standard methods applied to local data. Our method yields tests with good size and power under both stationary and nonstationary, thereby encompassing previous methods. In particular, the power gains are achieved without notable size distortions, the exact size being as good as those delivered by the best fixed-b approach, when the latter works well.

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

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

---

*This paper is based on the first chapter of my doctoral dissertation at Boston University. I am grateful to Pierre Perron for his continuous support and advice. I thank Whitney Newey and Tim Vogelsang for helpful discussions. I also thank Marine Carrasco, Giuseppe Cavaliere, Bin Chen, Jean-Jacques Forneron, Massimo Franchi, Raffaella Giacomini, Liudas Giraitis, Jesus Gonzalo, Chris Hansen, Hiroaki Kaido, Ivana Komunjer, Adam McCloskey, Zhongjun Qu, Myung Seo, Rasmus Varneskov, Dacheng Xiu, Yohei Yamamoto, and seminar participants at the BC-BU Econometrics Workshop, Chicago Booth, IWEEE 2020, KU Leuven and Georgetown for comments.

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

1 Introduction

The literature on heteroskedasticity and autocorrelation robust (HAR) inference is extensive and quite mature by now. For concreteness, consider the linear model where x_t is a vector of regressors and e_t is an unobservable disturbance, which can be serially correlated. It is now common practice to use OLS and correct the standard errors. This entails the estimation of the covariance matrix (referred to as the long-run variance, LRV) of $V_t = x_t e_t$ or (2π times) the spectral density of V_t at frequency zero when the latter is stationary (of course, in general, the relevant process V_t can be generated from a more complex model; e.g., a moment condition in a GMM context). Early important contributions in econometrics are Newey and West (1987; 1994) and Andrews (1991) who proposed heteroskedasticity and autocorrelation consistent (HAC) estimators with some optimal properties. This approach aims at devising good estimate of the LRV of V_t. An alternative method foregoes that aim and concentrates on having a test with a pivotal non-normal limit distribution that is obtained through an inconsistent estimate of the LRV of V_t that keeps the bandwidth at a fixed fraction of the sample size. This is the so-called fixed-b HAR inference initiated by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002; 2005). The drawback of this approach is that the limit distribution changes depending on the context and critical values are to be obtained numerically on a case by case basis. The literature since then has focused on providing various refinements, mostly to have tests having exact sizes closer to the nominal level.[1]

Most of this literature relies on stationarity with exception of the consistency results in Newey and West (1987) and of some results in Andrews (1991) which, however, do not provide accurate approximations. Yet, another strand of the literature has argued convincingly that the processes governing economic data $\{x_t\}$ and the errors in the relevant regressions $\{e_t\}$ are nonstationary.[2] This can occur for several reasons: changes in the moments of x_t induced by changes in the model parameters that govern the data [cf. Perron (1989), Stock and Watson (1996) and the surveys of Ng and Wright (2013) and Giacomini and Rossi (2015)]; changes in the moments of e_t (think about the Great Moderation with the decline in variance for many macroeconomic variables or the effects of the COVID-19 pandemic); smooth changes in the distributions governing either processes that arise from transitory dynamics; and so on. All these induces nonstationarity in $\{V_t\}$, which then

[1] The fixed-b or post-HAC literature is vast; see Dou (2019), Lazarus, Lewis, and Stock (2020), Lazarus et al. (2018), 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), Rho and Vogelsang (2020), Robinson (1998), Sun (2013; 2014a; 2014b), Sun, Phillips, and Jin (2008) and Zhang and Shao (2013).

[2] By nonstationary we mean non-constant moments. As in the literature, we consider processes whose sum of absolute autocovariances is finite. That is, we rule out processes with unbounded second moments (e.g., unit root).

makes $\mathbb{E}(V_t V_{t-k}')$ depend on both k and t. Furthermore, even if the data and primitive shocks $\{e_t\}$ are stationary, many HAR testing problems are such that the relevant process $\{V_t\}$ is stationary under the null hypothesis but is affected by changes in means (or other forms of nonstationarity) under the alternative. This occurs, for instance, when using tests involving structural breaks based on estimating the model under the null hypothesis; e.g., popular tests for forecast evaluation [e.g., Diebold and Mariano (1995)], tests for forecast instability [cf. Casini (2018), Giacomini and Rossi (2009) and Perron and Yamamoto (2021)], tests for structural change [cf. Casini and Perron (2019) and Perron (2006)]. When standardized by classical HAC estimators such tests may suffer from issues such as non-monotonic power, i.e., power that goes to zero as the alternative gets farther away from the null value. Various forms of misspecication and/or nonstationarity generate low frequency contamination and make the series or residuals appear much more persistent. As a consequence, HAC standard errors are too large and when used as normalizing factors of test statistics, the tests lose power [see Casini, Deng, and Perron (2021) for formal details].[3] This applies even more forcefully to the fixed-b type methods and to the recent refinements by Lazarus, Lewis, and Stock (2020) and Lazarus et al. (2018), since they involve more lagged autocovariances (or long bandwidths) and, hence, larger contaminations.

This points to the importance of extending the methods for HAR inference so that they have the correct size and good power even under nonstationarity. This is the aim of the paper. We first develop a theoretical framework under which to analyze the statistical properties of our suggested estimate. We introduce a class of nonstationary processes which possess a spectrum that varies both over frequencies and time, thereby encompassing the nonstationary models used in applied work. We work in an infill asymptotic setting akin to the one used in nonparametric regression [cf. Robinson (1989)]. For a process V_t, its spectrum at frequency ω and time $u = t/T$, denoted by $f(u, \omega)$, is allowed to change slowly yet continuously as well as to change abruptly in u at a finite number of time points; the latter allows for structural breaks in the spectrum of V_t. We label this class as Segmented Locally Stationary (SLS). It is related to the locally stationary processes introduced by Dahlhaus (1997) that have the characterizing property of behaving as a stationary process in a small neighborhood of u. This is achieved via smoothness of $f(u, \omega)$ in u. By allowing discontinuities across some segments, we can deal with relevant features such as structural change, regime switching-type and threshold models [cf. Bai and Perron (1998), Casini and Perron (2019,

[3] A partial list of works that present evidence of power issues with HAR inference is Altissimo and Corradi (2003), Casini and Perron (2019, 2020a, 2020b), Chan (2020), Crainiceanu and Vogelsang (2007), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Perron and Yamamoto (2021), Shao and Zhang (2010), Vogelsang (1999), Xu (2013), Zhang and Lavitas (2018).

2020a, 2020b and 2021b), Hamilton (1989) and Hansen (2000)]. The SLS class extends some of the analysis of Dahlhaus (1997) to processes having a more general time-varying spectrum.[4] Our framework is of independent interest and can be useful in many contexts in econometrics if one is interested in deriving the properties of estimators or inference under nonstationarity.

Under this framework, we introduce a double kernel HAC (DK-HAC) estimator in order to flexibly account for nonstationarity and we show that it is robust to low frequency contamination and other misspecifications. This entails an extension of the classical HAC estimators since in addition to the usual smoothing procedure over lagged autocovariances, it applies a second smoothing over time for each lagged autocovariance, involving a second kernel and bandwidth. If $\{V_t\}$ is Segmented Locally Stationary, $\mathbb{E}(V_t V'_{t-k})$ changes smoothly in t, as long as t is away from the change-points in the spectrum $f(t/T, \omega)$. Thus, the smoothing over time yields good estimates for the time path of $\mathbb{E}(V_t V'_{t-k})$ for all k. We determine the optimal kernels and optimal values for both bandwidth sequences under a mean-squared error (MSE) criterion. We establish new MSE bounds that show how nonstationarity affects the bias-variance trade-off and are more informative than previously established MSE bounds. We use them to construct data-dependent bandwidths relying on the "plug-in" approach. Unlike Andrews (1991), our candidate parametric models have time-varying parameters which can be estimated by applying standard methods to local data, akin to using rolling regressions. The procedure depends on three elements: the bandwidths for the smoothing over autocovariances and over time, and a block size to separate the regimes. In this paper, we consider a sequential bandwidth selection procedure by first deriving the optimal bandwidth for smoothing over time, then conditioning on this to obtain the optimal bandwidth for smoothing over autocovariances.

The DK-HAC estimators can result in HAR tests that are oversized when there is high temporal dependence in the data, a well-known problem for all methods, though for ours these distortions are relatively minor compared to, e.g., the methods of Newey and West (1987) and Andrews (1991). Still, in order to improve the size control of HAR tests, Casini and Perron (2021c), using the theory of this paper, propose a nonparametric nonlinear VAR prewhitened DK-HAC estimators. This form of prewhitening differs from those discussed previously [e.g., Andrews and Monahan (1992)

[4] A few authors used a notion of local stationarity that allows for breaks [see, e.g., Dahlhaus (2009) and Last and Shumway (2008)]. However, none of these works was concerned with HAR inference. Dahlhaus (2009) presented some results for local spectral density estimation and required smoothness (see Example 4.2 there). Last and Shumway (2008) considered testing for change-points in a locally stationary series which under the alternative hypothesis results in a piecewise locally stationary series. Furthermore, our notion of SLS processes and related framework are more general from a theoretical standpoint since we provide precise definitions and establish theoretical results about the identification of the local spectral density when there are discontinuities.

and Rho and Shao (2013)] in that it accounts explicitly for nonstationarity. HAR tests based on prewhitened DK-HAC estimators have size control competitive to fixed-b HAR tests when the latter work well (i.e., under stationarity). Notably, non-prewhitened and prewhitened DK-HAC have excellent power properties even when existing HAR tests have serious issues with power.

Comparison to Existing Literature

There are two main approaches to HAR inference differing on whether the LRV estimator is consistent or not. The classical approach relies on consistency, which results in HAC estimators [cf. Newey and West (1987; 1994), Andrews (1991) and Hansen (1992)], and on bandwidths chosen via MSE criterion. Inference is standard because HAR tests follow asymptotically standard distributions. The researcher then uses corrected standard errors and asymptotic critical values. It was shown early that classical HAC standard errors can result in oversized tests when there is substantial temporal dependence. This stimulated a second approach based on inconsistent LRV estimators that keep the bandwidth at some fixed fraction of the sample size [cf. Kiefer, Vogelsang, and Bunzel (2000)]. Because of the inconsistency, inference is nonstandard and HAR tests do not asymptotically follow standard distributions. Critical values are to be obtained numerically. Long bandwidths/fixed-b methods require stationarity and reduce the oversize problem of HAR tests. The bandwidth choice is often based on testing-oriented criteria [e.g., Sun, Phillips, and Jin (2008)]. Our approach falls in the first category; we propose HAC estimators and standard HAR inference.

We now compare in detail our approach to the existing literature. We believe that a fair comparison has to consider the following four criteria: (1) applicability to general HAR inference tests; (2) size of HAR tests; (3) power of HAR tests; (4) theoretical validity under stationarity/non-stationarity. In terms of (1), it is clear that the HAC and DK-HAC estimators are generally and immediately applicable to any HAR inference test and that they are simple to use in practice. This also explains why the classical HAC estimators have become the standard practice in econometrics and statistics. Methods that rely on long bandwidths/fixed-b do not share the same property. They are not generally applicable to HAR inference tests because a researcher would first need to derive a new asymptotic non-standard fixed-b distribution. This can be challenging/unfeasible in non-standard testing problems [e.g., tests for parameter instability, etc.]. Turning to (2), all existing HAR inference tests are known to be oversized when there is strong serial dependence. However, fixed-b HAR tests (or versions thereof) are less oversized than other tests based on the classical HAC estimators. The stronger is the temporal dependence the larger is the difference in size between the two approaches. Our prewhitened DK-HAC estimators are competitive with fixed-b HAR tests in controlling the size. Moving to (3), prewhitened and non-prewhitened DK-HAC estimators have excellent power under either stationarity or nonstationarity whereas existing methods have serious problems with power under nonstationarity or under nonstationary alternative hypotheses. These problems result in non-monotonic power and little or no power in relevant circumstances especially in HAR tests outside the stable linear regression model. Fixed-b or long bandwidths methods suffer most from these problems. Finally, turning to (4), our method like the classical HAC approach is valid under nonstationarity whereas methods using long bandwidths/fixed-b are only valid under stationarity. It should be mentioned that the fixed-b approach is shown to achieve (pointwise) higher-order refinements under stationarity while the MSE-based optimality of the HAC or DK-HAC estimators pertains only to the first-order but it holds under nonstationarity.[5]

Recently, Lazarus, Lewis, and Stock (2020) and Lazarus et al. (2018) made some progress to generalize the applicability of fixed-b methods. They showed that the t-test in the linear model using a LRV estimator based on equally-weighted cosine (EWC) under fixed-b asymptotics can achieve a t-distribution with the degrees of freedom depending on the bandwidth choice. However, Casini and Perron (2021c) showed that EWC is oversized relative to the original fixed-b of Kiefer, Vogelsang, and Bunzel (2000) and to the prewhitened DK-HAC when there is strong dependence. Lazarus, Lewis, and Stock (2020) relied on the Neyman-Pearson Lemma or simply "apple-to-apple comparison" to compare HAR tests. This is certainly a reasonable criterion. The lemma suggests to compare the power of tests that have an empirical size no greater than the significance level. However, the Neyman-Pearson Lemma alone does not suffice to find the "best" test in this context because all tests are oversized when there is strong dependence. Indeed, it does not even apply in this context. We face a trade-off between size and power. Our proposed method is competitive with the existing methods which is least oversized [i.e., original fixed-b of Kiefer, Vogelsang, and Bunzel (2000)] and has excellent power even when existing HAR tests do not have any. We believe that our method strikes a good balance with respect to criteria (1)-(4).

Our approach is different from methods based on subsampling of t-statistics [see, e.g., Ibragimov and Müller (2010)]. The latter rely on splitting the sample in subsamples and estimating the model within each subsample. Under the assumption that the estimates from the subsamples are asymptotically independent, the test statistic based on an average of estimates across subsamples follows asymptotically a t-distribution. In terms of point (1) above, this approach is not general enough compared to the HAC/DK-HAC approach because this changes the test statistic and its

[5]Pointwise means that higher-order refinements of fixed-b hold only for a given data-generating process. These results have been established for t- and F-test in a baseline linear model with stationary Gaussian errors. Preinerstorfer and Pötscher (2016) pointed out some limitations of this approach [e.g., no uniform (over DGPs) refinements].

asymptotic distribution. Also, subsampling test statistics which are not t-tests can be challenging/unfeasible and would require at best extra work in general HAR inference contexts to derive the new distribution. Finally, our simulation experience (not reported) suggests that this method suffers from the same finite-sample issues about size and power as the fixed-b methods.

Related Work

This paper is part of a set of papers on HAR inference by the author and collegues. The current paper provides the core theoretical and empirical elements that are used in all other papers, which can be viewed as providing extensions or refinements. Casini and Perron (2021c) used our theoretical framework to derive minimax MSE bounds for LRV estimation that are sharper than previously established and extended some of our theoretical results to general nonstationarity. As a finite-sample refinement, they also developed a new prewhitening procedure robust to nonstationarity for DK-HAC estimators. Even though the latter procedure is included in our simulations, we established the corresponding theoretical results in a separate paper because of the extent of the work needed in the analysis. Casini, Deng, and Perron (2021) showed analytically that the poor finite-sample performance of existing LRV-based HAR tests under nonstationarity and misspecification is induced by low frequency contamination. Belotti et al. (2021) used our theoretical framework to propose alternative data-dependent bandwidths for DK-HAC estimators that are optimal under a global MSE criterion. Casini and Perron (2021a) considered change-point detection in time series with evolutionary spectra. Initially, it was intended to be used in the method suggested in this paper to improve the finite-sample size and power properties, given that we work with segmented locally stationary processes. However, it turns out that the current method to select the blocks (see Section 4.4) is able to handle even abrupt structural change.

The remainder of the paper is organized as follows. Section 2 introduces the statistical setting and the new HAC estimator. Section 3 presents consistency, rates of convergence and the asymptotic MSE results for the DK-HAC estimators. Asymptotically optimal kernels and bandwidths are derived in Section 4. A data-dependent method for choosing the bandwidths and its asymptotic properties are discussed in Section 5. Section 6 presents a Monte Carlo study. Section 7 concludes the paper. The supplemental materials [cf. Casini (2021) and an additional supplement not for publication] contain some implementation details and all mathematical proofs. The code to implement our methods is provided in Matlab, R and Stata languages through a Github repository.

2 The Statistical Environment

To motivate our approach, consider the linear regression model estimated by least-squares (LS): $y_t = x_t'\beta_0 + e_t$ $(t = 1, \ldots, T)$, where $\beta_0 \in \Theta \subset \mathbb{R}^p$, y_t is an observation on the dependent variable, x_t is a p-vector of regressors and e_t is an unobserved disturbance. The LS estimator is given by $\widehat{\beta} = (X'X)^{-1}X'Y$, where $Y = (y_1, \ldots, y_T)'$ and $X = (x_1, \ldots, x_T)'$. Classical inference about β_0 requires estimation of $\mathrm{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ where

$$\mathrm{Var}(\sqrt{T}(\widehat{\beta} - \beta_0)) \triangleq \mathbb{E}\left[\left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} e_s x_s (e_t x_t)' \left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\right],$$

where "$\triangleq$" is used for definitional equivalence. Consistent estimation of $\mathrm{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ relies on consistent estimation of $\lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathbb{E}(e_s x_s (e_t x_t)')$. More generally, one needs a consistent estimate of $J \triangleq \lim_{T\to\infty} J_T$ 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 the linear regression model, $V_t(\beta) = (y_t - x_t'\beta)x_t$. For HAR tests outside the regression model $V_t(\beta)$ takes different forms.[6] Hence, our problem is to estimate J when $\{V_t\}$ is a Segmented Locally Stationary process, as defined in Section 2.1.[7] Such estimate can then be used to conduct HAR inference in the usual way using the theory developed below. By a change of variables, J_T can be rewritten as

$$J_T = \sum_{k=-T+1}^{T-1} \Gamma_{T,k}, \qquad \text{where} \qquad \Gamma_{T,k} = \begin{cases} T^{-1}\sum_{t=k+1}^{T} \mathbb{E}(V_t V_{t-k}') & \text{for } k \geq 0 \\ T^{-1}\sum_{t=-k+1}^{T} \mathbb{E}(V_{t+k} V_t') & \text{for } k < 0 \end{cases},$$

and $V_t = V_t(\beta_0)$. The rest of this section is structured as follows. In Section 2.1 we introduce a new class of nonstationary time series that we use as the underlying framework for our theoretical analysis. Section 2.2 presents the DK-HAC estimator. 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 by $X^{(j,l)}$. $\mathrm{tr}(\cdot)$ denotes the trace and $\otimes$ denotes the tensor product. The $p^2 \times p^2$ matrix C_{pp} is a commutation matrix that transforms $\mathrm{vec}(A)$ into $\mathrm{vec}(A')$, 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 A. W and $\widetilde{W}$ are used for $p^2 \times p^2$ weight matrices. $\mathbb{C}$ is used for the set of complex

[6] If one suspects that β_0 may not be constant, one can use appropriate tests for parameter instability. However, our discussions and methods still apply because these tests are HAR inference tests and one needs a LRV estimate of J based on the appropriate V_t.

[7] Casini and Perron (2021c) extended the results to the case where $\{V_t\}$ is generally nonstationary (i.e., $\{V_t\}$ is a sequence of unconditionally heteroskedastic random variables).

numbers. $\overline{A}$ is used for the complex conjugate of $A \in \mathbb{C}$. Let $0 = \lambda_0 < \ldots < \lambda_{m+1} = 1$. 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 (e.g., it is piecewise Lipschitz continuous if for each $j = 1, \ldots, m+1$ 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 $\lambda_{j-1} < u \leq \lambda_j$. 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 $\forall \omega \in \mathbb{R}$.

2.1 Segmented Locally Stationary Processes

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. In order to introduce a framework to analyze time series models with a time-varying spectrum it is necessary to introduce an infill asymptotic setting whereby we 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 $u \in [0, 1]$ which we interpret as saying that as $T \to \infty$ we observe more and more realizations of V_t close to time t, i.e., we observe the rescaled process V_{Tu} on the interval $[u - \varepsilon, u + \varepsilon]$, where $\varepsilon > 0$ is a small number.

In order to define a general class of nonstationary processes, we shall start from processes that have a time-varying spectral representation specified by:

$$V_{t,T} = \mu(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t) A(t/T, \omega) d\xi(\omega), \tag{2.1}$$

where $i \triangleq \sqrt{-1}$, $\mu(t/T)$ is the trend function, $A(t/T, \omega)$ is the transfer function and $\xi(\omega)$ is some stochastic process whose properties are specified below. Observe that this representation is similar to the spectral representation of stationary processes [see Anderson (1971), Brillinger (1975), Hannan (1970) and Priestley (1981)]. We shall see that the main difference is that $A(t/T, \omega)$ and $\mu(t/T)$ are not constant in t.[8] 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 have been used widely in both statistics and economics, though in the latter field they are best known as time-varying

[8] In HAR inference, a minimal assumption on V_t under the null hypothesis is that it has zero mean (i.e., $\mu(t/T) = 0$ for all t). However, in this subsection we allow for arbitrary $\mu(t/T)$ so as to introduce a general framework, also applicable under various alternative hypotheses in both within and outside the regression model.

parameter processes [see, e.g., Cai (2007) and Chen and Hong (2012)]. 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 point in time. We propose a class of nonstationarity processes which allow both continuous and discontinuous changes in the parameters. Stationarity and local stationarity are recovered as special cases.

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

$$V_{t,T}=\mu_j\left(t/T\right)+\int_{-\pi}^{\pi}\exp\left(i\omega t\right)A_{j,t,T}^{0}\left(\omega\right)d\xi\left(\omega\right),\qquad\left(t=T_{j-1}^{0}+1,\ldots,\,T_{j}^{0}\right),\tag{2.2}$$

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

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

where $\mathrm{cum}\left\{\cdot\right\}$ is the cumulant of rth order, $g_1=0$, $g_2\left(\omega\right)=1$, $\left|g_r\left(\omega_1,\ldots,\,\omega_{r-1}\right)\right|\leq M_r<\infty$ and $\varphi\left(\omega\right)=\sum_{j=-\infty}^{\infty}\delta\left(\omega+2\pi j\right)$ is the period 2π extension of the Dirac delta function $\delta\left(\cdot\right)$.

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

$$A\left(u,\,\omega\right)=A_j\left(u,\,\omega\right)\text{ for }\lambda_{j-1}^0<u\leq\lambda_j^0,\tag{2.3}$$

$$\sup_{1\leq j\leq m_0+1}\ \sup_{T_{j-1}^0<t\leq T_j^0,\,\omega}\left|A_{j,t,T}^0\left(\omega\right)-A_j\left(t/T,\,\omega\right)\right|\leq KT^{-1}.\tag{2.4}$$

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

The smoothness properties of A in u guarantees that $V_{t,T}$ has a piecewise locally stationary behavior. Later we will require additional smoothness properties for A.

Example 2.1. (i) Suppose X_t is a stationary process with spectral representation $X_t=\int_{-\pi}^{\pi}\exp\left(i\omega t\right)A\left(\omega\right)d\xi\left(\omega\right)$, and $\mu,\,\sigma:\,[0,\,1]\rightarrow\mathbb{R}$ are piecewise continuous. Then, $V_{t,T}=\mu_j\left(t/T\right)+\sigma_j\left(t/T\right)X_t$, with $T_{j-1}^0<t\leq T_j^0\,(1\leq j\leq m_0+1)$ is a SLS process with m_0+1 regimes where $A_{j,t,T}^0\left(\omega\right)=A_j\left(t/T,\,\omega\right)=\sigma_j\left(t/T\right)A\left(\omega\right)$. Within each segment, $V_{t,T}$ is locally stationary. When $t=Tu$ is away

from the change-points, as $T \to \infty$ more and more realizations of $V_{Tu,T}$ with $u \in [u-\varepsilon,\, u+\varepsilon]$ are observed, that is, realizations with amplitude close to $\sigma_j(u)$ for the appropriate j.

(ii) Suppose e_t is an *i.i.d.* sequence and $V_{t,T} = \sum_{k=0}^{\infty} a_{j,k}(t/T)\, e_{t-k}$, $T_{j-1}^0 < t \leq T_j^0$ $(1 \leq j \leq m_0+1)$. Then, $V_{t,T}$ is SLS with $A_{j,t,T}^0(\omega) = A_j(t/T,\, \omega) = \sum_{k=0}^{\infty} a_{j,k}(t/T) \exp(-i\omega k)$.

(iii) Autoregressive processes with time-varying coefficients, known as TVAR, augmented with structural breaks are SLS. In this case, we do not have the exact relationship $A_{j,t,T}^0(\omega) = A_j(t/T,\, \omega)$ but only the approximate relationship (2.4).

If there is only a single regime (i.e., $m_0 = 0$) then $V_{t,T}$ is locally stationary [cf. Dahlhaus (1997)]. If μ and A^0 do not depend on t, then $V_{t,T}$ is stationary and the spectral representation of stationary processes applies. However, $m_0 = 0$ rules out structural change and regime switching models. With $m_0 > 0$, we propose a framework where parameter variation can occur either smoothly or abruptly, both being relevant for economic data.[9]

Let $\lfloor \cdot \rfloor$ denote the largest smaller integer function and let $\mathcal{T} \triangleq \{T_1^0,\, \ldots,\, T_{m_0}^0\}$. We define the spectrum of $V_{t,T}$ in (2.1) (for fixed T) as

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

with $A_{1,t,T}^0(\omega) = A_1(0,\, \omega)$ for $t < 1$ and $A_{m_0+1,t,T}^0(\omega) = A_{m_0+1}(1,\, \omega)$ for $t > T$. Our definition coincides with the Wigner-Ville spectrum [cf. Martin and Flandrin (1985)] when there are no change-points (i.e., $m_0 = 0$). Below we show 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.

Assumption 2.1. *$A(u,\, \omega)$ is piecewise Lipschitz continuous in the first component and uniformly Lipschitz continuous in the second component, with index $\vartheta > 1/2$ for both.*

Theorem 2.1. *Assume $V_{t,T}$ is Segmented Locally Stationary with m_0+1 regimes and Assumption 2.1 holds. Then, for all $u \in (0,\, 1)$, $\int_{-\pi}^{\pi} \sum_{j=1}^{m_0+1} |f_{j,T}(u,\, \omega) - f_j(u,\, \omega)|^2 d\omega = o(1)$.*

[9] Some authors have used alternative notions of local stationarity that allow for discontinuities (i.e., piecewise locally stationary) and have established some results in other contexts which are not related to HAR inference [see, e.g., Dahlhaus (2009), Last and Shumway (2008) and Zhou (2013)]. In particular, our framework is more general because we also define (and work with) the covariance between observations belonging to different regimes whereas previous works considered only the covariance between observations belonging to the same regime thereby using smoothness which restricts the framework substantially.

Let $f(u,\,\omega)=f_j(u,\,\omega)$ if $Tu\in(T^0_{j-1},\,T^0_j]$ so as to suppress the subscript j from f. It is well-known that, even when $m_0=0$, the spectral representation (2.2) is not unique [cf. Priestley (1981), Chapter 11.1]. A consequence of Theorem 2.1 is that $\{f_j(u,\,\omega)=|A_j(u,\,\omega)|^2,\,j=1,\ldots,\,m_0+1\}$ is uniquely determined from the whole triangular array $\{V_{t,T}\}$.

For $Tu\notin\mathcal{T}$ with $T^0_{j-1}/T<u=t/T<T^0_j/T$, only the realizations of $V_{t,T}$ in the time interval $u\in[u-n/T,\,u+n/T]$ with $n\to\infty$ contribute to $f_j(u,\,\omega)$. Since this interval is fully contained in a segment j where $A_j(u,\,\omega)$ is smooth, and given that the length of this interval tends to zero, $V_{t,T}$ becomes "asymptotically stationary" on this interval. The length of the interval in which $V_{t,T}$ can be considered stationary is given by $n\ln n/T^{\vartheta}\to 0$. For $Tu\in\mathcal{T}$, the arguments are different. Suppose $Tu=T^0_j$. The spectrum $f_{j,T}(u,\,\omega)$ is defined in such a way that only observations prior to T^0_j are used in order to construct an approximation to $f_j(u,\,\omega)$. Since the length of this interval tends to zero and $A_j(u,\,\omega)$ is left-Lipschitz continuous, then those observations become "asymptotically stationary" and thus provide the same information about $f_j(u,\,\omega)$.

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 $c(u,\,k)\triangleq\int_{-\pi}^{\pi}e^{i\omega k}A(u,\,\omega)\,A(u-k/T,\,-\omega)\,d\omega$.

2.2 DK-HAC Estimation

In model (2.2), if $m_0=0$ and A^0 is constant in its first argument, then $\{V_{t,T}\}$ is second-order stationary. Its spectral density matrix is then equal to $f(\omega)\triangleq(2\pi)^{-1}\sum_{k=-\infty}^{\infty}\Gamma(k)\,e^{-i\omega k}$ where $\Gamma(k)\triangleq\mathbb{E}(V_{t,T}V'_{t-k,T})$. When evaluated at frequency $\omega=0$ it plays a prominent role because $\lim_{T\to\infty}J_T=2\pi f(0)$. Nonstationarity implies that the spectral density is time-varying since $\mathbb{E}(V_tV'_{t-k})$ now depends on k as well as on t. The SLS processes introduced above accommodate this property because they have a time-varying spectrum $f(u,\,\omega)$. Accordingly, we introduce the notation $\Gamma_u(k)\triangleq\mathbb{E}(V_{Tu,T}V'_{Tu-k,T})$ where $u=t/T$. We show below that $\Gamma_u(k)=c(u,\,k)+O(T^{-1})$ uniformly in $1\leq j\leq m+1$, $Tu\leq T^0_j$ and $k\in\mathbb{Z}$. Under the rescaling $u=t/T$, $u\in[0,\,1]$, the limit of J_T for SLS processes is given by,

$$J\triangleq\lim_{T\to\infty}J_T=\int_0^1 c(u,\,0)\,du+\sum_{k=1}^{\infty}\int_0^1\left(c(u,\,k)+c(u,\,k)'\right)du.$$

Using the definition of $f(u,\,\omega)$ it can be shown that $J = 2\pi\int_0^1 f(u,\,0)\,du$. Dahlhaus (2009) discussed how to estimate $f(u,\,\omega)$ for the scalar case under smoothness in both arguments using the smoothed local periodogram. Our goals are to estimate J using a time-domain method and to relax the smoothness assumption in u. This is different from Dahlhaus' work that considered local problems (i.e., estimation of $f(u,\,\omega)$ under smoothness) and not full-sample problems (i.e., estimation of J). The class of estimators of J relies on double kernel smoothing over lags and time,

$$\widehat{J}_T = \widehat{J}_T\left(b_{1,T},\,b_{2,T}\right) \triangleq \frac{T}{T-p}\sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right)\widehat{\Gamma}\left(k\right),\ \text{with}$$

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

where $K_1(\cdot)$ is a real-valued kernel in the class $\boldsymbol{K}_1$ defined below, $b_{1,T}$ is a bandwidth sequence discussed below, $n_T \to \infty$ satisfying the conditions given below, and

$$\widehat{c}_T\left(rn_T/T,\,k\right) \triangleq \begin{cases} \left(Tb_{2,T}\right)^{-1}\sum_{s=k+1}^{T} K_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{2,T}}\right)\widehat{V}_s\widehat{V}'_{s-k}, & k \geq 0 \\ \left(Tb_{2,T}\right)^{-1}\sum_{s=-k+1}^{T} K_2^*\left(\frac{((r+1)n_T-(s+k/2))/T}{b_{2,T}}\right)\widehat{V}_{s+k}\widehat{V}'_{s}, & k < 0 \end{cases}, \tag{2.5}$$

with K_2^* being a real-valued kernel and $b_{2,T}$ is a bandwidth sequence discussed below. $\widehat{c}_T(u,\,k)$ is an estimate of the local autocovariance $c(u,\,k)$ of lag k at time $u = rn_T/T$. Estimation of $c(u,\,k)$ for locally stationary processes was considered by Dahlhaus (2012). For positive semi-definiteness, it is necessary that K_2^* takes the following form:

$$K_2^*\left(\frac{(r+1)\,n_T-(s-k/2)}{Tb_{2,T}}\right) = \left(K_2\left(\frac{(r+1)\,n_T-s}{Tb_{2,T}}\right)K_2\left(\frac{(r+1)\,n_T-(s-k)}{Tb_{2,T}}\right)\right)^{1/2} \quad \text{for } k \geq 0,$$

$$K_2^*\left(\frac{(r+1)\,n_T-(s+k/2)}{Tb_{2,T}}\right) = \left(K_2\left(\frac{(r+1)\,n_T-s}{Tb_{2,T}}\right)K_2\left(\frac{(r+1)\,n_T-(s+k)}{Tb_{2,T}}\right)\right)^{1/2} \quad \text{for } k < 0.$$

Setting $K_2(x) = (\int_0^1 h(x)^2\,dx)^{-1}h(x+1/2)^2$ and $N_T = Tb_{2,T}$, we see that positive semi-definiteness requires the use of a data taper $h(\cdot)$ with length N_T. This follows because we need each $\widehat{V}_t$ $(t = 1,\ldots,T)$ to be assigned the same weight across different k for any given r. Then, letting $\widehat{V}_t^{\circ} = \left(K_2\left(\left((r+1)\,n_T-t\right)/Tb_{2,T}\right)\right)^{1/2}\widehat{V}_t$ we can use the same arguments as in Andrews (1991) applied now to $\widehat{V}_t^{\circ}$ to show that J_T is positive semi-definite for the appropriate choice of K_1.

The estimator $\widehat{J}_T$ involves two kernels: K_1 smooths the lagged sample autocovariances, akin to the classical HAC estimators, while K_2 applies smoothing over time. The factor $T/\left(T-p\right)$ is an optional small-sample degrees of freedom adjustment. In Section 3-4, we consider estimators $\widehat{J}_T$ for which $b_{1,T}$ and $b_{2,T}$ are given sequences. In Section 5, we consider adaptive estimators $\widehat{J}_T$ 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. Since the order of $b_{2,T}(\cdot)$ is the same across blocks, we omit this notation for the developments of the asymptotic results. However, when we determine the data-dependent estimate of $b_{2,T}(\cdot)$, we will estimate $b_{2,T}(rn_T/T)$ for each r. We consider the following class of kernels [cf. Andrews (1991)],

$$\boldsymbol{K}_1 = \{K_1(\cdot) : \mathbb{R} \rightarrow [-1, 1] : K_1(0) = 1,\ K_1(x) = K_1(-x),\ \forall x \in \mathbb{R} \tag{2.6}$$
$$\int_{-\infty}^{\infty} K_1^2(x)\,dx < \infty,\ K_1(\cdot) \text{ is continuous at 0 and at all but finite numbers of points}\}.$$

Examples of kernels in $\boldsymbol{K}_1$ include the Truncated, Bartlett, Parzen, Quadratic Spectral (QS) and Tukey-Hanning kernel. We shall show below that the QS kernel has certain optimality properties:

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

3 HAC Estimation with Predetermined Bandwidths

In Section 3.1 we present some asymptotic properties of $\widehat{c}(\cdot, \cdot)$. We use them in Section 3.2 in order to establish consistency, rate of convergence and MSE properties of predetermined bandwidths 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 $\{\widehat{V}_{t,T}\} = \{V_{t,T}(\widehat{\beta})\}$. We first require some smoothness of $A(u, \cdot)$ in u.

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

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

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* $\widetilde{\kappa}_{a,b,c,d}:[0,\,1]\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\rightarrow\mathbb{R}$ *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* $(\partial^2/\partial\_u^2)\,\widetilde{\kappa}_{a,b,c,d}(u,\cdot,\cdot,\cdot)$*.*

If $\{V_{t,T}\}$ is stationary then the cumulant condition of Assumption 3.2-(i) reduces to the standard one used in the time series literature [see also Assumption A in Andrews (1991)]. We do not require fourth-order stationarity but only that the time-$t=Tu$ fourth order cumulant is locally constant in a neighborhood of u. One can show that α-mixing and moment conditions imply that the cumulant condition of Assumption 3.2 holds.

3.1 Estimation of the Local Covariance

Let $\widetilde{c}_T(u,\,k)$ denote the estimator that uses $\{V_{t,T}\}$. We consider the following class of kernels:

$$\boldsymbol{K}_2=\{K_2(\cdot):\mathbb{R}\rightarrow[0,\,\infty]:\,K_2(x)=K_2(1-x)\,,\,\int K_2(x)\,dx=1,\\ K_2(x)=0,\text{ for }x\notin[0,\,1]\,,\,K_2(\cdot)\text{ is continuous}\}. \tag{3.1}$$

Lemma 3.1. *Suppose that Assumption 3.1-3.2 hold. If* $b_{2,T}\rightarrow 0$ *and* $Tb_{2,T}^5\rightarrow\eta\in(0,\,\infty)$*, then* $\widetilde{c}_T(u_0,\,k)-c(u_0,\,k)=O_{\mathbb{P}}(\sqrt{Tb_{2,T}})$ *for all* $u_0\in(0,\,1)$*.*

3.2 Results on DK-HAC Estimation with Predetermined Bandwidths

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, $K_{1,q}=-(d^qK_1(x)/dx^q)|_{x=0}/q!$ 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 $f^{(q)}(u, 0) = (-1)^{q/2} (d^q f(u, \omega)/d\omega^q)|_{\omega=0}$. Further, $||f^{(q)}(u, 0)|| < \infty$ if and only if $f(u, \omega)$ is q times differentiable at $\omega = 0$. We define

$$\text{MSE}\left(Tb_{1,T}b_{2,T},\, \widetilde{J}_T,\, W\right) = Tb_{1,T}b_{2,T}\mathbb{E}\left[\text{vec}\left(\widetilde{J}_T - J_T\right)' W \text{vec}\left(\widetilde{J}_T - J_T\right)\right]. \tag{3.2}$$

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)*

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

(ii) If $1/Tb_{1,T}^q b_{2,T} \to 0$, $n_T/Tb_{1,T}^q \to 0$ and $b_{2,T}^2/b_{1,T}^q \to 0$ for some $q \in [0, \infty)$ for which $K_{1,q}$, $||\int_0^1 f^{(q)}(u, 0)\,du|| \in [0, \infty)$, then $\lim_{T\to\infty} b_{1,T}^{-q}\mathbb{E}(\widetilde{J}_T - J_T) = -2\pi K_{1,q}\int_0^1 f^{(q)}(u, 0)\,du$.

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

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

If $b_{2,T}^2/b_{1,T}^q \to \nu < \infty$ replaces $b_{2,T}^2/b_{1,T}^q \to 0$ in part (ii), then the asymptotic bias for the case of locally stationary processes becomes

$$\lim_{T\to\infty} b_{1,T}^{-q}\mathbb{E}(\widetilde{J}_T - J_T) = -2\pi K_{1,q}\int_0^1 f^{(q)}(u, 0)\,du + \frac{\nu}{2}\int_0^1 x^2 K_2(x) \sum_{k=-\infty}^{\infty}\int_0^1 \frac{\partial^2}{\partial u^2} c(u, k)\,du. \tag{3.3}$$

For the general case of SLS processes the term involving $(\partial^2/\partial^2 u)\, c(u, k)$ is different. The second summand on the right-hand side of (3.3) cancels when $\int_0^1 (\partial^2/\partial^2 u)\, c(u, k)\,du = 0$. The latter occurs when the process is stationary. Dahlhaus (2012) presented MSE results for a pointwise estimate of $f(u, \omega)$ under continuity in both components by applying smoothing over u and ω. His results depends on the local behavior of $f(u, \omega)$ at time u and frequency ω whereas in our problem the MSE results depend on properties of the full time path of $f(u, 0)$. The theorem suggests that the optimal choice of $b_{1,T}$ hinges on the degree of nonstationary in the data, a feature that does not appear from the corresponding results in the literature. The results are derived as $n_T \to \infty$. It is possible and indeed easier to keep n_T fixed, in which case the results are unchanged. However, the case with n_T fixed can have some disadvantages when the spectrum is discontinuous because then the estimator would be often dealing with observations from different regimes, which as explained above might lead to low frequency contamination. 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)* $\sup_{u\in[0,\,1]}\mathbb{E}\sup_{\beta\in\Theta}||\left(\partial/\partial\beta'\right)V_{\lfloor Tu\rfloor}\left(\beta\right)||^2<\infty$; *(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). As remarked above, we interpret β_0 as the pseudo-true parameter β^* when the model is misspecified. Part (iv) of the assumption is satisfied by most commonly used kernels. 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*

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

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

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

Theorem 3.2. *Suppose $K_1(\cdot)\in\boldsymbol{K}_1$, $K_2(\cdot)\in\boldsymbol{K}_2$, $b_{1,T},\,b_{2,T}\to0$, $n_T\to\infty$, $n_T/Tb_{1,T}\to0$, and $1/Tb_{1,T}b_{2,T}\to0$. We have:*

(i) If Assumption 3.1-3.3 hold, $\sqrt{T}b_{1,T}\to\infty$, $b_{2,T}/b_{1,T}\to0$ then $\widehat{J}_T-J_T\overset{\mathbb{P}}{\to}0$ and $\widehat{J}_T-\widetilde{J}_T\overset{\mathbb{P}}{\to}0$.

(ii) If Assumption 3.1, 3.3-3.4 hold, $n_T/Tb_{1,T}^q\to0$, $1/Tb_{1,T}^qb_{2,T}\to0$, $b_{2,T}^2/b_{1,T}^q\to0$ and $Tb_{1,T}^{2q+1}b_{2,T}\to\gamma\in(0,\,\infty)$ for some $q\in[0,\,\infty)$ for which $K_{1,q}$, $||\int_0^1f^{(q)}(u,\,0)\,du||\in[0,\,\infty)$, 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) and Assumption 3.5,

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

The consistency result of $\widehat{J}_T$ in part (i) applies to kernels $K_1(\cdot)$ with unbounded support and to bandwidths $b_{1,T}$ and $b_{2,T}$ such that $1/b_{1,T}b_{2,T}$ grows at rate $o(\sqrt{T/b_{2,T}})$. Part (ii) yields the consistency of $\widehat{J}_T$ with $b_{1,T}$ only required to be $o\left(Tb_{2,T}\right)$. This rate is slower than the corresponding rate $o\left(T\right)$ 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 HAC estimator the optimal growth rate is $T^{1/5}$ while it is $T^{1/3}$ for the Newey-West HAC estimator. Part (ii) of the theorem presents the rate of convergence of $\widehat{J}_T$ which is $\sqrt{Tb_{2,T}b_{1,T}}$. In Section 4, we compare the rate of convergence of $\widehat{J}_T$ with that of the classical HAC estimators when the respective optimal bandwidths are used.

4 Optimal Kernels, Bandwidths and Choice of n_T

In this section, we show the optimality of quadratic-type kernels under MSE criterion.[10] For K_1, the result states that the QS kernel minimizes the asymptotic MSE for any $K_2(\cdot)$. Let

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

where $\widetilde{W}_T$ is some $p\times p$ positive semidefinite matrix. The optimal bandwidths $b_{1,T}^{\mathrm{opt}}$ and $b_{2,T}^{\mathrm{opt}}$ satisfy the following sequential MSE criterion:

$$\begin{aligned}&\mathrm{MSE}\left(Tb_{1,T}^{\mathrm{opt}}\overline{b}_{2,T}^{\mathrm{opt}},\, \widehat{J}_T\left(b_{1,T}^{\mathrm{opt}},\, \overline{b}_{2,T}^{\mathrm{opt}}\right),\, W_T\right)\leq \mathrm{MSE}\left(Tb_{1,T}^{\mathrm{opt}}\overline{b}_{2,T}^{\mathrm{opt}},\, \widehat{J}_T\left(b_{1,T},\, \overline{b}_{2,T}^{\mathrm{opt}}\right),\, W_T\right) \qquad (4.1)\\ &\quad\text{where}\quad \overline{b}_{2,T}^{\mathrm{opt}}=\int_0^1 b_{2,T}^{\mathrm{opt}}(u)\,du\\ &\quad\text{and}\qquad b_{2,T}^{\mathrm{opt}}(u)=\underset{b_{2,T}}{\operatorname{argmin}}\ \mathrm{MSE}\left(b_{2,T}^{-4},\, \widehat{c}_T(u_0,\, k)\,,\, \widetilde{W}_T\right).\end{aligned}$$

The first inequality above has to hold as $T\rightarrow\infty$. The above criterion determines the globally optimal $b_{1,T}^{\mathrm{opt}}$ given the integrated locally optimal $b_{2,T}^{\mathrm{opt}}(u)$. Thus, $b_{1,T}^{\mathrm{opt}}$ and $\overline{b}_{2,T}^{\mathrm{opt}}$ need not be the same as the bandwidths $(\widetilde{b}_{1,T}^{\mathrm{opt}},\, \widetilde{b}_{2,T}^{\mathrm{opt}})$ that jointly minimize the global asymptotic MSE,

$$\lim_{T\rightarrow\infty}\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T(b_{1,T},\, b_{2,T})\,,\, W_T\right). \qquad (4.2)$$

Theorem 3.1-(ii) states that, under the condition $b_{2,T}^2/b_{1,T}^q\rightarrow 0$, the bias only depends on the smoothing over lagged autocovariances but not on $b_{2,T}$. Then, the global solution $\widetilde{b}_{2,T}^{\mathrm{opt}}$ would be trivial: $b_{2,T}$ affects the MSE only through the variance term and optimality requires to set the bandwidth as large as possible. In contrast, the MSE criterion (4.1) based on the MSE given

[10]Besides Andrews (1991) and Newey and West (1987) in the context of LRV estimation, the MSE-optimality criterion was also used more recently by Whilelm (2015) in a GMM context to determine the optimal bandwidth of the nonparametric estimator of the optimal weighting matrix.

in Theorem 3.1-(iii) leads to a unique solution which can be obtained analytically. Under the condition $b_{2,T}^2/b_{1,T}^q \to \nu < \infty$, Belotti et al. (2021) determined the bandwidths $(\widetilde{b}_{1,T}^{\text{opt}}, \widetilde{b}_{2,T}^{\text{opt}})$ that jointly minimize (4.2). They showed that $\widetilde{b}_{1,T}^{\text{opt}}, \widetilde{b}_{2,T}^{\text{opt}} = O(T^{-1/6})$ while the optimal bandwidths $(b_{1,T}^{\text{opt}}, \overline{b}_{2,T}^{\text{opt}})$ from (4.1) satisfy $b_{1,T}^{\text{opt}} = O(T^{-4/25})$ and $\overline{b}_{2,T}^{\text{opt}} = O(T^{-1/5})$. Thus, the criterion (4.1) leads to a slightly shorter block length relative to the global criterion (4.2) (i.e., $T\overline{b}_{2,T}^{\text{opt}} < T\widetilde{b}_{2,T}^{\text{opt}}$). A shorter bock length is beneficial if there is substantial nonstationarity and implies less sensitivity to low frequency contamination from not properly accounting for nonstationarity [cf. Casini et al. (2021)]. For a throughout comparison between the two criteria see Belotti et al. (2021).

4.1 Optimal $K_2(\cdot)$ and $b_{2,T}$

Let $F(K_2) \triangleq \int_0^1 K_2^2(x)\,dx$, $H(K_2) = (\int_0^1 x^2 K_2(x)\,dx)^2$, and for any $k \in \mathbb{Z}$,

$$D_1(u_0) \triangleq \operatorname{vec}\left(\partial^2 c(u_0, k)/\partial u^2\right)' \widetilde{W} \operatorname{vec}\left(\partial^2 c(u_0, k)/\partial u^2\right),$$
$$D_2(u_0) \triangleq \operatorname{tr}\widetilde{W}(I_{p^2} + C_{pp}) \sum_{l=-\infty}^{\infty} c(u_0, l) \otimes [c(u_0, l) + c(u_0, l+2k)].$$

Proposition 4.1. *Suppose Assumption 3.1, 3.3-3.4 hold and $\widetilde{W}_T \xrightarrow{\mathbb{P}} \widetilde{W}$. We have for all $a, b \le p$,*

$$\begin{aligned}
&\operatorname{MSE}\left(1, \widehat{c}_T^{(a,b)}(u_0, k), 1\right) \\
&\quad = \frac{1}{4} b_{2,T}^4 \left(\int_0^1 x K_2(x)\,dx\right)^2 \left(\frac{\partial^2}{\partial^2 u} c^{(a,b)}(u_0, k)\right)^2 \\
&\qquad + \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\,dx \sum_{l=-\infty}^{\infty} c^{(a,b)}(u_0, l)\left[c^{(a,b)}(u_0, l) + c^{(a,b)}(u_0, l+2k)\right] \\
&\qquad + \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\,dx \sum_{h_1=-\infty}^{\infty} \kappa_{V,\lfloor Tu_0 \rfloor}^{(a,b,a,b)}(-k, h_1, h_1 - k) + o\left(b_{2,T}^4\right) + O\left(1/(b_{2,T}T)^2\right).
\end{aligned}$$

$\operatorname{MSE}(b_{2,T}^{-4}, \widehat{c}_T(u_0, k) - c(u_0, k), \widetilde{W}_T)$ *is minimized with*

$$b_{2,T}^{\text{opt}}(u_0) = [H\left(K_2^{\text{opt}}\right) D_1(u_0)]^{-1/5} \left(F\left(K_2^{\text{opt}}\right)(D_2(u_0) + D_3(u_0))\right)^{1/5} T^{-1/5},$$

where $D_3(u_0)$ depends on $\widetilde{\kappa}$ (for $p = 1$, $D_3(u_0) = \sum_{h_1=-\infty}^{\infty} \kappa_{V,\lfloor Tu_0 \rfloor}(-k, h_1, h_1 - k)$), and $K_2^{\text{opt}}(x) = 6x(1-x)$, $0 \le x \le 1$. In addition if V_t is Gaussian, then $D_3(u_0) = 0$, for $u_0 \in (0, 1)$.

The optimal kernel $K_2^{\text{opt}}(x)$ is a transformation of the Epanechnikov kernel. Optimality of quadratic kernels under a MSE criterion has been shown in many contexts [cf. Epanechnikov

(1969) and Priestley (1981)]. The optimal bandwidth sequence decreases at rate $T^{-1/5}$ which is the same optimal rate derived in the context of parameter estimation of locally stationary processes [see e.g., Dahlhaus and Giraitis (1998)]. The term $D_1(u_0)$ is due to nonstationary, while the term $D_2(u_0)$ measures the variability of $\widehat{c}_T(u_0, k)$. The bandwidth $b_{2,T}^{\mathrm{opt}}$ converges to zero at a slower rate as the process becomes closer to stationary (i.e., as the square root of $D_1(u_0)$ decreases).

4.2 Optimal $K_1(\cdot)$

We next determine the optimal kernel K_1 and the optimal bandwidth sequence $b_{1,T}$ given any K_2 and any $b_{2,T}$ of order $O(T^{-1/5})$, i.e., the same order of $b_{2,T}^{\mathrm{opt}}(u)$ for any $u \in [0, 1]$. Let $\widehat{J}_T^{\mathrm{QS}}$ denote $\widehat{J}_T$ when the latter is based on the QS kernel. For some results below, we consider a subset of $\boldsymbol{K}_1$. Let $\widetilde{\boldsymbol{K}}_1 = \{K_1(\cdot) \in \boldsymbol{K}_1 | \, \widetilde{K}(\omega) \geq 0 \, \forall \, \omega \in \mathbb{R}\big\}$ where $\widetilde{K}(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} K_1(x) \, e^{-ix\omega} dx$. The function $\widetilde{K}(\omega)$ is referred to as the spectral window generator. The set $\widetilde{\boldsymbol{K}}_1$ contains all kernels K_1 that necessarily generate positive semidefinite estimators in finite samples.

We adopt the notation $\widehat{J}_T(b_{1,T}) = \widehat{J}_T(b_{1,T}, \, b_{2,T}, \, K_2)$ to denote the estimator $\widehat{J}_T$ that uses $b_{1,T}, \, b_{2,T} = \overline{b}_{2,T}^{\mathrm{opt}} + o\left(T^{-1/5}\right)$ and $K_2(\cdot)$. We then compare two kernels K_1 using comparable bandwidths $b_{1,T}$ which are defined as follows. Given $K_1(\cdot) \in \widetilde{\boldsymbol{K}}_1$, the QS kernel $K_1^{\mathrm{QS}}(\cdot)$, and a bandwidth sequence $\{b_{1,T}\}$ to be used with the QS kernel, define a comparable bandwidth sequence $\{b_{1,T,K_1}\}$ for use with $K_1(\cdot)$ such that both kernel/bandwidth combinations have the same asymptotic variance when scaled by the same factor $Tb_{1,T}b_{2,T}$. This means that

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{MSE}(Tb_{1,T}b_{2,T}, \, \widehat{J}_T^{\mathrm{QS}}(b_{1,T}) - \mathbb{E}(\widetilde{J}_T^{\mathrm{QS}}(b_{1,T})) + J_T, \, W_T) \\
= &\lim_{T\to\infty} \mathrm{MSE}(Tb_{1,T}b_{2,T}, \, \widehat{J}_T(b_{1,T,K_1}) - \mathbb{E}(\widetilde{J}_T(b_{1,T,K_1})) + J_T, \, W_T).
\end{aligned}$$

This definition yields $b_{1,T,K_1} = b_{1,T}/(\int K_1^2(x) \, dx)$ and $b_{1,T,\mathrm{QS}} = b_{1,T}$ since $\int \left(K_1^{\mathrm{QS}}\right)^2(x) \, dx = 1$.

Theorem 4.1. *Suppose Assumption 3.1, 3.3-3.5 hold,* $\int_0^1 ||f^{(2)}(u, 0)|| du < \infty$, $b_{2,T} \to 0$, $b_{2,T}^5 T \to \eta \in (0, \infty)$, $(\mathrm{vec}(\int_0^1 f^{(q)}(u, 0) \, du))' W \mathrm{vec}(\int_0^1 f^{(q)}(u, 0) \, du) > 0$ *and* W *is positive semidefinite. For any bandwidth sequence* $\{b_{1,T}\}$ *such that* $b_{2,T}/b_{1,T} \to 0$, $n_T/Tb_{1,T}^2 \to 0$ *and* $Tb_{1,T}^5 b_{2,T} \to \gamma \in (0, \infty)$, *and for any kernel* $K_1(\cdot) \in \widetilde{\boldsymbol{K}}_1$ *used to construct* $\widehat{J}_T$, *the QS kernel is preferred to* $K_1(\cdot)$ *in the* *sense that*

$$\begin{aligned}&\lim_{T\to\infty}\left(\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widehat{J}_T\left(b_{1,T,K_1}\right),\,W_T\right)-\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widehat{J}_T^{\mathrm{QS}}\left(b_{1,T}\right),\,W_T\right)\right)\\&\quad=4\gamma\pi^2\left(\mathrm{vec}\left(\int_0^1 f^{(2)}\left(u,\,0\right)du\right)\right)'W\mathrm{vec}\left(\int_0^1 f^{(2)}\left(u,\,0\right)du\right)\int_0^1\left(K_2^{\mathrm{opt}}\left(x\right)\right)^2dx\\&\quad\times\left[K_{1,2}^2\left(\int K_1^2\left(y\right)dy\right)^4-\left(K_{1,2}^{\mathrm{QS}}\right)^2\right]\geq 0.\end{aligned}$$

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

The requirement $\int_0^1||f^{(2)}(u,\,0)||du<\infty$ is not stringent and it reduces to the one used by Andrews (1991) when $\{V_{t,T}\}$ is stationary. If $\int_0^1||f^{(q)}(u,\,0)||du<\infty$ only for some $1\leq q<2$, one can show that any kernel with $K_{1,q}=0$ has smaller asymptotic MSE than a kernel with $K_{1,q}>0$. The QS, Parzen, and Tukey-Hanning kernels have $K_{1,q}=0$ for $1\leq q<2$, whereas the Bartlett has $K_{1,q}>0$ for $1\leq q<2$. Thus, the asymptotic superiority of the former kernels over the Bartlett kernel holds even if $\int_0^1||f^{(q)}(u,\,0)||du<\infty$ only for $1\leq q<2$.

4.3 Optimal Predetermined Bandwidth Sequence $b_{1,T}$

We now present the predetermined bandwidth sequence that minimizes the asymptotic MSE given $b_{2,T}=O(b_{2,T}^{\mathrm{opt}})$ and $K_2=K_2^{\mathrm{opt}}$. This optimality result applies to each kernel $K_1(\cdot)\in\boldsymbol{K}_1$ for which $K_{1,q}\in(0,\,\infty)$ for some $q\in(0,\,\infty)$. Thus, most commonly used kernels are allowed with the exception of the truncated kernel. Let

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

The optimal bandwidth is $b_{1,T}^{\mathrm{opt}}=(2qK_{1,q}^2\phi(q)\,Tb_{2,T}^{\mathrm{opt}}/(\int K_1^2(y)\,dy\int_0^1 K_2^2(x)\,dx))^{-1/(2q+1)}$, where $\phi(q)$ is a function of the unknown spectral density $f(\cdot,\,\cdot)$. Hence, the optimal bandwidth $b_{1,T}^{\mathrm{opt}}$ is unknown in practice, and we consider data-dependent estimates of $\phi(q)$ in Section 5.

Condition 1. $b_{1,T},\,b_{2,T}\to 0$ with $b_{2,T}/b_{1,T}\to 0$, and $Tb_{1,T}^{2q+1}b_{2,T}\to\gamma\in(0,\,\infty)$ for some $q\in[0,\,\infty)$ for which $K_{1,q},\,||\int_0^1 f^{(q)}(u,\,0)\,du||\in[0,\,\infty)$, where $b_{2,T}=O(T^{-1/5})$.

Corollary 4.1. *Suppose Assumption 3.1, 3.3-3.5 hold, $||\int_0^1 f^{(q)}(u,\,\omega)\,du||<\infty$, $\phi(q)\in(0,\,\infty)$, and W is positive definite. Consider $K_1(\cdot)\in\boldsymbol{K}_1$ for which $K_{1,q}\in(0,\,\infty)$ for some $q\in(0,\,\infty)$.*

Then, $\{b_{1,T}^{\mathrm{opt}}\}$ is optimal among the sequences $\{b_{1,T}\}$ that satisfy Condition 1 in the sense that,

$$\lim_{T\to\infty}\Bigg(\mathrm{MSE}\left(\left(Tb_{2,T}\right)^{2q/(2q+1)},\ \widehat{J}_T\left(b_{1,T},\, b_{2,T}\right),\, W_T\right) \\ - \mathrm{MSE}\left(\left(Tb_{2,T}\right)^{2q/(2q+1)},\ \widehat{J}_T\left(b_{1,T}^{\mathrm{opt}},\, b_{2,T}\right),\, W_T\right)\Bigg) \geq 0.$$

The inequality is strict unless $b_{1,T} = b_{1,T}^{\mathrm{opt}} + o((Tb_{2,T})^{-1/(2q+1)})$.

In Corollary 4.1, $q = 2$ for the QS kernel and so $b_{1,T}^{\mathrm{opt}} = 0.6584(\phi\,(2)\,Tb_{2,T}^{\mathrm{opt}})^{-1/5}(\int_0^1 K_2^2\,(y)\,dy)^{1/5}$. For $K_2\,(y) = K_2^{\mathrm{opt}}\,(y)$, the latter reduces to,

$$b_{1,T}^{\mathrm{opt}} = 0.6828(\phi\,(2)\,Tb_{2,T}^{\mathrm{opt}})^{-1/5}. \tag{4.3}$$

The optimal bandwidth is of order $T^{-4/25}$. Thus, the optimal bandwidth sequence decreases to zero at a slower rate than the optimal bandwidth sequence for the QS kernel-based HAC estimator of Andrews (1991), for which the rate is of order $T^{-1/5}$. The slower rate is due to the fact that our estimator smooths the spectrum over time through $K_2\,(\cdot)$ and this restricts the smoothing of $K_1\,(\cdot)$. In particular, the optimal choice of $b_{1,T}$ hinges on the degree of nonstationary through $b_{2,T}^{\mathrm{opt}}$. The more nonstationary are the data, the smaller is $b_{2,T}^{\mathrm{opt}}$ and the large is $b_{1,T}^{\mathrm{opt}}$ which means that less weight is given to $\widehat{\Gamma}\,(k)$ for $k \neq 0$. In contrast, the optimal choice of $b_{1,T}$ for the methods proposed in the literature is independent of the degree of nonstationarity. When $b_{1,T}$ and $b_{2,T}$ are chosen optimally, the convergence rate from Theorem 3.2 reduces to $T^{8/25}$. Thus, the rate is slower than the corresponding one for the QS kernel HAC estimator considered in Andrews (1991). However, it is misleading to compare our DK-HAC estimator with the classical HAC estimators only on the basis of the rate of convergence. In fact, the DK-HAC estimators account flexibly for nonstationarity and are robust to low frequency contamination induced by nonstationarity/misspecification whereas the classical HAC estimators are not in general [cf. Casini et al. (2021)].

4.4 Choice of n_T

Our MSE analysis does not indicate an optimal value for n_T. It only suggests growth rate bounds. When K_1^{QS} is used, n_T cannot grow faster than $T^{2/3}$. We set $n_T = T^{0.66}$ for the QS kernel. That is, we choose n_T to be the largest possible value allowed by the condition. Our sensitivity analysis (not reported) suggests that choosing a smaller n_T might result in excessive overlapping of regimes when the process is SLS (i.e., $m_0 > 0$). See Belotti et al. (2021) for more details.

5 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 an asymptotic formula for an optimal bandwidth parameter (i.e., the expressions for $b_{1,T}^{\mathrm{opt}}$ and $b_{2,T}^{\mathrm{opt}}$ from Section 4). Section 5.1 explains how to construct the automatic bandwidths while Section 5.2 presents the corresponding theoretical results.

5.1 Implementation

Let us begin with $b_{1,T}^{\mathrm{opt}}$ and then move to $b_{2,T}^{\mathrm{opt}}$. The first step for the construction of data-dependent bandwidth parameters is to specify p univariate parametric models for the elements of $V_t = (V_t^{(1)}, \ldots, V_t^{(p)})'$. The second step involves the estimation of the parameters. In our context, the logical estimation methods to use are local (weighted) least-squares (LS) (i.e., LS method applied to rolling windows) and nonparametric kernel methods. In a third step, we replace the unknown parameters in $\phi(q)$ with corresponding estimates. Such estimate $\widehat{\phi}(q)$ is then substituted into the expression for $b_{1,T}^{\mathrm{opt}}$ to yield the data-dependent bandwidth $\widehat{b}_{1,T}$:

$$\widehat{b}_{1,T} = \left(2qK_{1,q}^2 \widehat{\phi}(q)\, T \widehat{\overline{b}}_{2,T} / \left(\int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx\right)\right)^{-1/(2q+1)}, \tag{5.1}$$

where $\widehat{\overline{b}}_{2,T} = (n_T/T) \sum_{r=1}^{\lfloor T/n_T \rfloor - 1} \widehat{b}_{2,T}(rn_T/T)$. $\widehat{\overline{b}}_{2,T}$ is an average of the estimates $\widehat{b}_{2,T}(\cdot)$. Since $b_{2,T}$ depends on u, it is more efficient to estimate it for each block as its optimal value can change over time. In practice, a reasonable candidate for an approximating parametric model is the class of first order autoregressive [AR(1)] models for $\{V_t^{(r)}\}$, $r = 1, \ldots, p$ (with different parameters for each r) or a first order vector autoregressive [VAR(l)] model for $\{V_t\}$. These classes were also used by Andrews (1991). However, in our context it is reasonable to allow the parameters of the AR(1) model to be time-varying. For parsimony, we consider a time-varying AR(1) with no breaks in $f(u, \omega)$, i.e., $V_t^{(r)} = a_1(t/T)\, V_{t-1}^{(r)} + u_t^{(r)}$, where the $u_t^{(r)}$ need not be independent across r.

The use of p univariate parametric models requires a simple form for the weight matrix W.

In particular, W has to be a diagonal matrix which in turn implies that $\phi(q)$ reduces to

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

The usual choice is $W^{(r,r)} = 1$ for all r except that which corresponds to an intercept for which it is set to zero. An estimate of $f^{(r,r)}(u,\, 0)$ is $\widehat{f}^{(r,r)}(u,\, 0) = (2\pi)^{-1} (\widehat{\sigma}^{(r)}(u))^2 (1 - \widehat{a}_1^{(r)}(u))^{-2}$ 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 $u = t/T$:

$$\widehat{a}_1^{(r)}(u) = \frac{\sum_{j=\lfloor Tu \rfloor - n_{2,T}+1}^{\lfloor Tu \rfloor} \widehat{V}_j^{(r)} \widehat{V}_{j-1}^{(r)}}{\sum_{j=\lfloor Tu \rfloor - n_{2,T}+1}^{\lfloor Tu \rfloor} \left(\widehat{V}_{j-1}^{(r)} \right)^2}, \tag{5.2}$$

$$\widehat{\sigma}^{(r)}(u) = \left(\sum_{j=\lfloor Tu \rfloor - n_{2,T}+1}^{\lfloor Tu \rfloor} \left(\widehat{V}_j^{(r)} - \widehat{a}_1^{(r)}(u)\, \widehat{V}_{j-1}^{(r)} \right)^2 \right)^{1/2},$$

where $n_{2,T} \to \infty$. Then, for the QS kernel K_1,

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

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

For most of the results below we can take $n_{3,T} = n_{2,T} = n_T$. After plugging-in $\widehat{\phi}(2)$ into the formula (4.3), we have $\widehat{b}_{1,T} = 0.6828 (\widehat{\phi}(2)\, T \widehat{\overline{b}}_{2,T})^{-1/5}$.

We now propose a data-dependent procedure for $b_{2,T}(u_r)$, where $u_r = rn_T/T$ for $r = 1, \ldots, \lfloor (T - n_T)/n_T \rfloor$. 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 $c(\cdot,\, k)$ is the same as 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)]. The reason for imposing this condition is that it is otherwise difficult to estimate $(\partial^2/\partial u^2)\, c(u,\, k)$, which enters $D_1(u)$, from the data. Under the above specification, the exact expression of $D_1(u)$ can be computed analytically:

$$D_1(u) \triangleq (\int_{-\pi}^{\pi} \Big[\frac{3}{\pi}(1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega))^{-4}(0.8(-4\pi\sin(4\pi u)))\exp(-i\omega)$$
$$-\frac{1}{\pi}|1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega)|^{-3}\left(0.8\left(-16\pi^2\cos(4\pi u)\right)\right)\exp(-i\omega)\Big]d\omega)^2.$$

An estimate of $D_1(u)$ is given by

$$\widehat{D}_1(u) \triangleq ([S_\omega]^{-1}\sum_{s\in S_\omega}\Big[\frac{3}{\pi}(1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega_s))^{-4}(0.8(-4\pi\sin(4\pi u)))\exp(-i\omega_s)$$
$$-\frac{1}{\pi}|1+0.8(\cos 1.5+\cos 4\pi u)\exp(-i\omega_s)|^{-3}\left(0.8\left(-16\pi^2\cos(4\pi u)\right)\right)\exp(-i\omega_s)\Big])^2,$$

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 $S_\omega = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\}$. Note that we have computed $\widehat{D}_1(u)$ for $k = 0$ because it makes the computation simpler. Further, this is consistent with our sequential MSE criterion because $k = 0$ is the only lag for which $K_1(0) = 1$ for all K_1 so that the choice of K_1 does not influence $b_2^{\mathrm{opt}}(\cdot)$. It remains to derive an estimate of $D_2(u)$ since $F(K_2)$ and $H(K_2)$ can be computed for a given $K_2(\cdot)$. We assume that the innovations of the approximating time-varying AR(1) model satisfy $\mathbb{E}(u_t^{(r)}) = 0$, $\mathbb{E}((u_t^{(r)})^2) = \sigma^2$ and $\mathbb{E}((u_t^{(r)})^4) = 3\sigma^4$ so that $D_3(u) = 0$ for all $u \in (0, 1)$. That is, the term involving the cumulant drops from $b_2^{\mathrm{opt}}(u)$. In practice this is convenient because it is complex to deal with consistent estimation of cumulant terms. Note also that $D_3(u) = 0$ if u_t is Gaussian. Since $c(u, k)$ can be consistently estimated by $\widehat{c}_T(u, k)$, an estimate of $D_2(u)$ is given by

$$\widehat{D}_2(u_0) \triangleq p^{-1}\sum_{r=1}^{p}\sum_{l=-\lfloor T^{4/25}\rfloor}^{\lfloor T^{4/25}\rfloor}\widehat{c}_T^{(r,r)}(u_0, l)\left[2\widehat{c}_T^{(r,r)}(u_0, l)\right],$$

where the number of summands grows at the same rate as $(b_{1,T}^{\mathrm{opt}})^{-1}$; a different choice is allowed as long as it grows at a slower rate than $T^{2/5}$. Hence, the estimate of the optimal $b_{2,T}$ is given by

$$\widehat{b}_{2,T}(u_r) = 1.6786\left(\widehat{D}_1(u_r)\right)^{-1/5}(\widehat{D}_2(u_r))^{1/5}T^{-1/5}, \qquad \text{where} \qquad u_r = rn_T/T.$$

5.2 Theoretical Results

Next, we establish consistency, rate of convergence and asymptotic MSE results for the estimator $\widehat{J}_T(\widehat{b}_{1,T},\, \overline{\widehat{b}}_{2,T})$ that uses the data-dependent bandwidths $\widehat{b}_{1,T}$ and $\overline{\widehat{b}}_{2,T}$. As in Andrews (1991), we need to restrict the class of admissible kernels to the following class:

$$\begin{aligned}\boldsymbol{K}_3 = \{&K_3(\cdot) \in \boldsymbol{K}_1 :\, (i)\ |K_1(x)| \leq C_1 |x|^{-b} \text{ with } b > \max(1+1/q,\, 3) \\ &\text{for } |x| \in [\overline{x}_L,\, D_T h_T \overline{x}_U],\, b_{1,T}^2 h_T \to \infty,\, D_T > 0,\, \overline{x}_L,\, \overline{x}_U \in \mathbb{R},\, 1 \leq \overline{x}_L < \overline{x}_U, \text{ and} \\ &\text{with } b > 1+1/q \text{ for } |x| \notin [\overline{x}_L,\, D_T h_T \overline{x}_U],\, \text{and some } C_1 < \infty, \text{ where } q \in (0,\, \infty) \\ &\text{is such that } K_{1,q} \in (0,\, \infty),\, (ii)\ |K_1(x) - K_1(y)| \leq C_2 |x-y|\ \forall x,\, y \in \mathbb{R} \text{ for some} \\ &\text{costant } C_2 < \infty, \text{ and } (iii)\ q < 34/4\}.\end{aligned} \tag{5.3}$$

Let $\widehat{\theta}$ denote the estimator of the parameter of the approximate (time-varying) parametric model(s) introduced above. For example, with univariate AR(1) approximating parametric models, $\widehat{\theta} = (\int_0^1 \widehat{a}_1^{(1)}(u)\, du,\, \int_0^1 (\widehat{\sigma}^{(1)}(u))^2 du, \ldots,\, \int_0^1 \widehat{a}_1^{(p)}(u)\, du,\, \int_0^1 (\widehat{\sigma}^{(p)}(u))^2 du)'$. Let θ^* denote the probability limit of $\widehat{\theta}$. $\widehat{\phi}(q)$ is the value of $\phi(q)$ with $\widehat{\theta}$ instead of θ. Its probability limit is denoted by ϕ_{θ^*}.

Assumption 5.1. *(i) $\widehat{\phi}(q) = O_{\mathbb{P}}(1)$ and $1/\widehat{\phi}(q) = O_{\mathbb{P}}(1)$; (ii) $\inf\{T/n_{3,T},\, \sqrt{n_{2,T}}\}(\widehat{\phi}(q) - \phi_{\theta^*}) = O_{\mathbb{P}}(1)$ for some $\phi_{\theta^*} \in (0,\, \infty)$ where $n_{2,T}/T + n_{3,T}/T \to 0$, $n_{2,T}^{10/6}/T \to [c_2,\, \infty)$, $n_{3,T}^{10/6}/T \to [c_3,\, \infty)$ with $0 < c_2,\, c_3 < \infty$; (iii) $\sup_{u\in[0,\,1]} \lambda_{\max}(\Gamma_u(k)) \leq C_3 k^{-l}$ for all $k \geq 0$ for some $C_3 < \infty$ and some $l > \max\{2,\, 1 + 48q/(46+20q),\, 1 + q/(3/4 + q/2)\}$, where q is as in $\boldsymbol{K}_3$; (iv) uniformly in $u \in [0,\, 1]$, $\widehat{D}_1(u),\, \widehat{D}_2(u) = O_{\mathbb{P}}(1)$ and $1/\widehat{D}_1(u),\, 1/\widehat{D}_2(u) = O_{\mathbb{P}}(1)$; (v) $|\omega_{s+1} - \omega_s| = O(T^{-1})$ and $[S_\omega] = O(T)$; (vi) $\sqrt{T b_{2,T}(u)}(\widehat{D}_2(u) - D_2(u)) = O_{\mathbb{P}}(1)$ for all $u \in [0,\, 1]$; (vii) $\boldsymbol{K}_2$ includes kernels that satisfy $|K_2(x) - K_2(y)| \leq C_4 |x-y|$ for all $x,\, y \in \mathbb{R}$ and some constant $C_4 < \infty$.*

Parts (i)-(ii) of Assumption 5.1 are the nonparametric analogue to Assumption E and F, respectively, in Andrews (1991). Part (iii) is satisfied if $\{V_t\}$ is strong mixing with mixing numbers that are less stringent than those sufficient for the cumulant condition in Assumption 3.2-(i). Part (iv) and (vi) extend (i)-(ii) to $\widehat{D}_1$ and $\widehat{D}_2$. Part (v) is needed to apply the convergence of Riemann sums. Part (vi) follows from the asymptotic results about $\widehat{c}_T(u,\, k)$. Part (vii) requires K_2 to satisfy Lipschitz continuity. Note that ϕ_{θ^*} coincides with the optimal value $\phi(q)$ only when the approximate parametric model indexed by θ^* corresponds to the true data-generating mechanism.

Let $b_{\theta_1,T} = (2qK_{1,q}^2 \phi_{\theta^*} T \overline{b}_{\theta_2,T} / \int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx)^{-1/(2q+1)}$, where $\overline{b}_{\theta_2,T} \triangleq \int_0^1 b_{2,T}^{\mathrm{opt}}(u)\, du$. The asymptotic properties of $\widehat{J}_T(\widehat{b}_{1,T},\, \overline{\widehat{b}}_{2,T})$ are shown to be equivalent to those of $\widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})$.

Theorem 5.1. *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,*

(i) If Assumption 3.1-3.3 and 5.1-(i,iv,vii) hold, $n_{3,T} = n_{2,T} = n_T$, and $q > 1/2$, then $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T}) - J_T \xrightarrow{\mathbb{P}} 0$.

(ii) If Assumption 3.1, 3.3-3.4 and 5.1-(ii,iii,v,vi,vii) hold and $n_T/Tb^2_{\theta_1,T} \to 0$, then $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}$ $(\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T}) - J_T) = O_{\mathbb{P}}(1)$ and $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T})) = o_{\mathbb{P}}(1)$.

(iii) Let $\gamma_\theta = 2qK^2_{1,q}\phi_\theta/(\int K_1^2(y)\,dy \int_0^1 K_2^2(x)\,dx)$. If Assumption 3.1, 3.3-3.5 and 5.1-(ii,iii,v,vi,vii) hold, then

$$
\begin{aligned}
&\lim_{T\to\infty} \mathrm{MSE}\left(T^{4q/10(2q+1)}, \widehat{J}_T\left(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T}\right), W_T\right) \\
&\quad = \lim_{T\to\infty} \mathrm{MSE}\left(Tb_{\theta_1,T}b_{\theta_2,T}, \widehat{J}_T\left(b_{\theta_1,T}, b_{\theta_2,T}\right), W_T\right) \\
&\quad = 4\pi^2\left[\gamma_\theta K^2_{1,q}\mathrm{vec}\left(\int_0^1 f^{(q)}(u,0)\,du\right)' W\mathrm{vec}\left(\int_0^1 f^{(q)}(u,0)\,du\right)\right] \\
&\qquad + \int K_1^2(y)\,dy \int K_2^2(x)\,dx\,\mathrm{tr}\left(W\left(I_{p^2} - C_{pp}\right)\left(\int_0^1 f(u,0)\,du\right) \otimes \left(\int_0^1 f(v,0)\,dv\right)\right).
\end{aligned}
$$

When the chosen parametric model indexed by θ is correct, it follows that $\phi_{\theta^*} = \phi(q)$ and $\widehat{\phi}(q) \xrightarrow{\mathbb{P}} \phi(q)$. The theorem then implies that $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T})$ exhibits the same optimality properties presented in Theorem 4.1 and Corollary 4.1. We omit the details.

6 Small-Sample Evaluations

We conduct a Monte Carlo analysis to evaluate the properties of HAR inference based on the HAC estimator $\widehat{J}_T$. We consider HAR tests in the linear regression model as well as HAR tests used in the forecast evaluation literature, namely the Diebold and Mariano's (1995) test and the forecast breakdown test of Giacomini and Rossi (2009). The linear regression models have an intercept and a stochastic regressor. We focus on the t-statistics $t_r = \sqrt{T}(\widehat{\beta}^{(r)} - \beta_0^{(r)})/\sqrt{\widehat{J}^{(r,r)}_{X,T}}$ where

$$
\widehat{J}_{X,T} = \left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \widehat{J}_T \left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1},
$$

is a consistent estimate of the limit of $\mathrm{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ and $r = 1, 2$. t_1 is the t-statistic for the parameter associated with the intercept while t_2 is associated with the stochastic regressor x_t. Results for the F-test are qualitatively similar [see Casini (2019)]. Six basic regression models are considered. We run a t-test on the intercept in model M1 and M5 whereas a t-test on the coefficient of x_t is run in model M2-M4 and M6. 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 hypotheses. In Model M1 $e_t = 0.5e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.5)$, $x_t \sim$ i.i.d. $\mathscr{N}(1, 1)$, $\beta_0^{(1)} = 0$ and $\beta_0^{(2)} = 1$.[11] Model M2 involves $e_t = 0.8e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, $x_t \sim$ i.i.d. $\mathscr{N}(1, 1)$, and $\beta_0^{(1)} = \beta_0^{(2)} = 0$. In Model M3 we have segmented locally stationary errors $e_t = \rho_t e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, $\rho_t = \max\{0, -1(\cos(1.5 - \cos(5t/T)))\}$ for $t < 4T/5$ and $e_t = 0.9e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $t \geq 4T/5$, and $x_t = 0.4x_{t-1} + u_{X,t}$, $u_{X,t} \sim$ i.i.d. $\mathscr{N}(0, 1)$. Note that ρ_t varies smoothly between 0 and 0.8071. Model M4 involves some misspecification that induces nonstationarity in the errors,

$$y_t = \beta_0^{(1)} + \left(\beta_0^{(2)} + \delta\right) x_t + w_t \mathbf{1}\{t \geq 4T/5\} + e_t, \qquad t = 1, \ldots, T,$$

where $e_t = \rho_t e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, ρ_t as in M3, $x_t \sim$ i.i.d. $\mathscr{N}(1, 1)$, and $w_t \sim$ i.i.d. $\mathscr{N}(2, 1)$ independent from x_t. Model M5 involves misspecification under H_1 via a smooth change in the coefficient $\beta_0^{(2)}$ toward the end of the sample. This situation is very common in practice and it is motivated by the model for the variable "cay" from Bianchi, Lettau, and Ludvigson (2018) (cf. Figure 3 in their paper). The model is given by

$$y_t = \beta_0^{(1)} + \delta + \left(\beta_0^{(2)} + d_t \mathbf{1}\{t \geq 4.5T/5\}\right) x_t + e_t, \qquad t = 1, \ldots, T,$$

where $d_t = 1.5\delta(t - 4.5T/5)/T$, $e_t = \rho_t e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, $\rho_t = 0.8(\cos(1.5 - \cos(t/2T)))$ for $t \in \{1, \ldots, T/2 - 1\} \cup \{T/2 + T/4 + 1, \ldots, T\}$ and $e_t = 0.2e_{t-1} + 2u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $T/2 \leq t \leq T/2 + T/4$, and $x_t = 2 + 0.5x_{t-1} + u_{X,t}$, $u_{X,t} \sim$ i.i.d. $\mathscr{N}(0, 1)$. That is, ρ_t varies smoothly between 0 and 0.7021. Model M6 is given by (6.2) where $e_t = \rho_t e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, $\rho_t = \max\{0, 0.3(\cos(1.5 - \cos(t/5T)))\}$ for $t \in \{1, \ldots, T/2 - 1\} \cup \{T/2 + 4, \ldots, T - 16\}$ and $e_t = 0.99e_{t-1} + 2u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $T/2 \leq t \leq T/2 + 3$ and $e_t = 0.9e_{t-1} + 2u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $T - 15 \leq t \leq T$, and $x_t \sim$ i.i.d. $\mathscr{N}(1, 1)$. Note that $\rho_t \in [0, 0.2633]$.

[11] For the results with AR coefficient 0.9 see Table 1 in Casini and Perron (2021c) and footnote 12 below.

Next, we move to the forecast evaluation tests. The Diebold-Mariano test statistic is defined as $t_{\mathrm{DM}} \triangleq \sqrt{T_n}\,\overline{d}_L/\sqrt{\widehat{J}_{d_L,T}}$, where $\overline{d}_L$ is the average of the loss differentials between two competing forecast models, $\widehat{J}_{d_L,T}$ is an estimate of the asymptotic variance of the the loss differential series and T_n is the number of observations in the out-of-sample. Throughout we use the quadratic loss. In model M7, we consider an out-of-sample forecasting exercise with a fixed scheme where, given a sample of T observations, $0.5T$ observations are used for the in-sample and the remaining half is used for prediction. The true model for the target variable is given by $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1}^{(0)} + e_t$ where $x_{t-1}^{(0)} \sim$ i.i.d. $\mathscr{N}(1,\,1)$, $e_t = 0.3e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathscr{N}(0,\,1)$ and we set $\beta_0^{(1)} = \beta_0^{(2)} = 1$. The two competing models both involve an intercept but differ on the predictor used in place of $x_t^{(0)}$. The first forecast model uses $x_t^{(1)}$ while the second uses $x_t^{(2)}$ where $x_t^{(1)}$ and $x_t^{(2)}$ are independent i.i.d. $\mathscr{N}(1,\,1)$ sequences, both independent from $x_t^{(0)}$. Each forecast model generates a sequence of $\tau\,(=1)$-step ahead out-of-sample losses $L_t^{(i)}$ $(i = 1,\,2)$ for $t = T/2+1, \ldots, T-\tau$. Then $d_t \triangleq L_t^{(2)} - L_t^{(1)}$ denotes the loss differential at time t. The Diebold-Mariano test rejects the null of equal predictive ability when (after normalization) $\overline{d}$ is sufficiently far from zero.

Finally, we consider model M8 which we use to investigate the performance of a t-test for forecast breakdown [cf. Giacomini and Rossi (2009)]. Suppose we want to forecast a variable y_t following the equation $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1} + e_t$ where $x_t \sim$ i.i.d. $\mathscr{N}(1,\,1.5)$ 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) (GR) 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 and one-step ahead forecasts. The 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 an HAC estimator. We restrict attention to $\tau = 1$.

Throughout our study we consider the following LRV estimators: $\widehat{J}_T$ with automatic bandwidths; $\widehat{J}_T$ with automatic bandwidths and the prewhitening of Casini and Perron (2021c); Andrews's (1991) HAC estimator with automatic bandwidth; Andrews's (1991) HAC estimator with automatic bandwidth and the prewhitening procedure of Andrews and Monahan (1992); Newey and West's (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994); Newey and West's (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994) and the prewhitening procedure; Newey-West with the fixed-b method of Kiefer, Vogelsang, and Bunzel (2000).[12] Casini and Perron (2021c) proposed three forms of prewhitening: (1) $\widehat{J}_{T,\text{pw},1}$ uses a stationary model to whiten the data; (2) $\widehat{J}_{T,\text{pw,SLS}}$ uses a nonstationary model to whiten the data; (3) $\widehat{J}_{T,\text{pw,SLS},\mu}$ is the same as $\widehat{J}_{T,\text{pw,SLS}}$ but it adds a time-varying intercept in the VAR to whiten the data. For model M7 we also report results using $\widehat{J}_{T}$ and $\widehat{J}_{T,\text{pw,SLS}}$ with the pre-test for breaks in the spectrum as developed in Casini and Perron (2021a). We do not report the results for the pre-test for model M1-M6 and M8 because they are equivalent to those without the pre-test.

For all versions of $\widehat{J}_{T}$ we use K_1^{opt} and K_2^{opt}. We set $n_T = T^{0.66}$ as explained in Section 4.4 and $n_{2,T} = n_{3,T} = n_T$. We consider the following sample sizes for M1-M6: $T = 200$, 400. Simulation results for additional data-generating processes involving ARMA, ARCH and heteroskedastic errors are not discussed here because the results are qualitatively equivalent [see, e.g., Casini (2019) and Casini and Perron (2021c)]. The significance level is $\alpha = 0.05$ throughout the study.

6.1 Empirical Sizes of HAR Inference Tests

Table 1-4 report the rejection rates for model M1-M8. We begin with the t-test in the linear regression models. As a general pattern, we confirm previous evidence that the Newey-West's (1987) and Andrews' (1991) HAC estimators lead to t-tests that are oversized when the data are stationary [cf. model M1-M2]. The same problem occurs for the Newey-West (1987) HAC estimator using the usual "rule" to determine the number of lags (not reported). For extreme temporal dependence, simulations in Casini and Perron (2021c) showed that the size distortions can be even larger especially for the t-test on the intercept. Prewhitening is often effective in helping the HAC estimators to better control the size under stationarity. However, the simulation results in Casini and Perron (2021c) and in the literature show that the prewhitened HAC estimators can lead to oversized tests when there is high serial dependence. The rejection rates of tests normalized by the Newey-West estimator with fixed-b are the most accurate in model M1-M2 for $T = 200$. Overall, the results in the literature along with those in Casini (2019) and Casini and Perron (2021c) showed that under stationarity the original fixed-b method of KVB is in general the least oversized across different degrees of dependence among all existing methods. Table 1 shows that

[12] To save space, we do not report results for the Empirical Weighted Periodogram (EWP) or Empirical Weighted Cosine (EWC) of Lazarus et al. (2020) and Lazarus et al. (2018), respectively. Their performance is similar to the method of Kiefer, Vogelsang, and Bunzel (2000). Indeed, the LRV estimator of Kiefer, Vogelsang, and Bunzel (2000) leads to HAR tests that have better size control. Casini and Perron (2021c) showed that EWC leads to oversized tests when there is strong dependence in the data relative to the fixed-b method of Kiefer, Vogelsang, and Bunzel (2000) and to the prewhitened DK-HAC. The power properties of tests normalized by the EWP and EWC are similar to those using the method of Kiefer, Vogelsang, and Bunzel (2000).

for the t-test on the intercept the non-prewhitened DK-HAC leads to HAR tests that are oversized while they are accurate for the t-test on the coefficient on the stochastic regressor. The table also shows that the prewhitened DK-HAC estimators are competitive with the KVB's fixed-b in controlling the size. $\widehat{J}_{T,\mathrm{pw},1}$ is the most accurate among the DK-HAC estimators. Since $\widehat{J}_{T,\mathrm{pw},1}$ uses a stationarity VAR model to whiten the data, it works better than $\widehat{J}_{T,\mathrm{pw,SLS}}$ and $\widehat{J}_{T,\mathrm{pw,SLS},\mu}$ when stationarity actually holds which is consistent with the results of Table 1.

Turning to nonstationarity, Table 2 casts concerns about the performance of existing methods in this context. For both model M3 and M4, existing LRV estimators lead to HAR tests that have either size equal or close to zero. The methods that use long bandwidths (i.e., many lags) such as KVB's fixed-b suffer most from this problem relative to the classical HAC estimators. This is consistent with the argument in Casini, Deng, and Perron (2021) who showed analytically that nonstationarity induces a positive bias for each sample autocovariance. That bias is constant across different lags. Since existing LRV estimators are weighted sum of sample autocovariances, the larger the bandwidth (i.e., the more lagged autocovariances are included) the larger the positive bias. Thus, LRV estimators are inflated and HAR tests have rejection rates lower than the significance level. This mechanism has consequences for power as well, as we show below that traditional HAR tests have low power. In model M3-M4 the non-prewhitened DK-HAC and the prewhitened DK-HAC (except $\widehat{J}_{T,\mathrm{pw},1}$) perform well. $\widehat{J}_{T,\mathrm{pw},1}$ suffers from the same problem as the existing estimators because it uses stationarity and when this is violated its performance is affected. In model M5, the classical HAC estimators yield HAR tests that are oversized. Also the non-prewhitened DK-HAC is oversized. In contrast, the KVB's fixed-b and the prewhitened DK-HAC have rejection rates close to the significance level. In model M6, the KVB's fixed-b HAR tests tend to be undersized whereas the HAC and DK-HAC estimators lead to tests that control the size more accurately.

Turning to the HAR tests for forecast evaluations, Table 4 shows that for model M7 the KVB's fixed-b HAR test has size essentially equal to zero while the classical HAC estimators yield HAR tests that are somewhat oversized. In contrast, the tests normalized by the prewhitened DK-HAC estimators have most accurate rejection rates. In model M8, the KVB's fixed-b HAR tests are well-sized whereas the classical HAC estimators lead to tests that are severely undersized. The DK-HAC estimators control the size reasonably well.

In summary, the prewhitened DK-HAC estimators yield t-tests in regression models with rejection rates that are relatively close to the nominal size. The non-prewhitened DK-HAC can lead to oversized tests for the t-tests on the intercept if there is high dependence. Our results confirm the oversize problem of the HAR tests normalized by the classical HAC estimators documented in the literature under stationarity. The Fixed-b HAR tests control the size well when the data are stationary but can show severe undersized issues under nonstationarity, a problem also affecting the tests normalized by the classical HAC estimators. Thus, with regards to size control, the prewhitened DK-HAC estimators are competitive with fixed-b methods under stationarity and they also perform well when the data are nonstationary.

6.2 Empirical Power of HAR Inference Tests

For model M1-M6 we report the values of the power in Table 5-10. The sample size is $T = 200$. Power functions for the Diebold-Maraino and for the forecast breakdown test are presented next. For model M1, the non-prewhitened HAC and DK-HAC lead to tests that have the highest power but they were more oversized than the other methods. The KVB's fixed-b LRV leads to t-tests that sacrify some power relative to the prewhitened HAC and DK-HAC estimators. In model M2, a similar picture arises. 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 prewhitening HAC estimators and $\widehat{J}_{T,\mathrm{pw},1}$ (which uses a stationary model to whiten the data) have low power. The best power is achieved by tests normalized by Andrews' (1991) HAC estimator and $\widehat{J}_{T}$, followed by $\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$ and $\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$. The KVB's fixed-$b$ leads to relatively less power than the latter methods. The Newey-West's (1987) estimator leads to tests that have good power but they were shown to be oversized. Similar comments apply to model M4. Here Andrews' (1991) HAC estimator leads to tests that have better power for small to medium breaks while tests based on $\widehat{J}_{T}$ have better power for large breaks. In model M5, prewhitening HAC estimators and KVB's fixed-b lead to HAR tests that have non-monotonic power and reach zero as δ increases. This does not occur for the classical HAC estimators which, however, were oversized. HAR tests based on $\widehat{J}_{T,\mathrm{QS}}$, $\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$ and $\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$ perform best for this model. $\widehat{J}_{T,\mathrm{pw},1}$ results in HAR tests that have lower power relative to the tests based on the other DK-HAC because stationarity is violated. In model M6, all HAR tests enjoy monotonic power with small differences.

Next, let us move to the evaluation of the power properties of the t-tests used in the forecasting literature. We begin with the Diebold-Mariano test. For this test, the separation between the null and alternative hypotheses does not depend on the value of a single parameter. Thus, the data-generating mechanism is different from the one under the null. The two competing forecast models are as follows: the first model uses the actual true data-generating process while the second model differs in that in place of $x_{t-1}^{(0)}$ it uses $x_{t-1}^{(2)} = x_{t-1}^{(0)} + u_{X_2,t}$ for $t \leq 3T/4$ and $x_{t-1}^{(2)} = \delta + x_{t-1}^{(0)} + u_{X_2,t}$ for $t > 3T/4$ with $u_{X_2,t} \sim$ i.i.d. $\mathscr{N}(0,\, 1)$. Evidently, the null hypothesis of equal predictive ability should be rejected by the Diebold-Mariano test whenever $\delta > 0$. Table 11 reports the power for several values of δ. The HAR tests based on existing estimators have lower power relative to the $\widehat{J}_T$ DK-HAC estimators for small values of δ. When we raise δ the tests based on the HAC estimators of Andrews (1991) and Newey and West (1987), and KVB's fixed-b method display non-monotonic power gradually converging to zero. In contrast, the DK-HAC estimators lead to tests that have monotonic power that reach and maintain unit power. The only exception is the test based on $\widehat{J}_{T,\mathrm{pw},1}$ that has lower power because stationarity is violated. The table also reports $\widehat{J}_T$ and $\widehat{J}_{T,\mathrm{pw,SLS}}$ with the pre-test for breaks in the spectrum [cf. Casini and Perron (2021a)] that is used for choosing more efficiently how to split the sample in blocks to compute $\widehat{\Gamma}(k)$. The pre-test yields HAR tests with higher power while having the same size as the corresponding HAR tests with no pre-test. We have not reported the results with the pre-test for model M1-M6 and M8 because they are the same as with no pre-test.

Finally, we move to the t-test of Giacomini and Rossi (2009). The data-generating process under $H_1 : \mathbb{E}\left(\overline{SL}\right) \neq 0$ is given by $y_t = 1 + x_{t-1} + \delta x_{t-1}\mathbf{1}\{t > T_1^0\} + e_t$, where $x_{t-1} \sim$ i.i.d. $\mathscr{N}(1.5,\, 1)$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim$ i.i.d.$\mathscr{N}(0,\, 1)$ and $T_1^0 = T\lambda_1^0$ with $\lambda_1^0 = 0.8$. Under this specification there is a break in the coefficient associated to the predictor x_{t-1}. Thus, there is a forecast failure and the test of Giacomini and Rossi (2009) should reject H_0. From Table 12 it appears that all versions of the classical HAC estimators of Andrews (1991) and Newey and West (1987), and KVB's fixed-b lead to t-tests that have, essentially, zero power for all δ. The only exception is Andrews' (1991) HAC estimator with prewhitening that shows some power but it is not monotonic. In contrast, the t-test based on the DK-HAC estimators have good power. The failure of existing LRV estimators cannot be attributed to the sample size because as we raise the sample size to 400 or 800, the tests still display no power [see Casini (2019)].

The failure of the HAR tests based on the existing LRV estimators occurring in some of the data-generating mechanisms reported here can be simply reconciled with the fact that in such models the spectrum of V_t is not constant. In other words, the autocovariance of V_t depends not only on the lag order but also on t. Existing LRV estimators estimate an average of a time-varying spectrum. Because of this instability in the spectrum, they overestimate the extent of the dependence or variation in V_t. This is explained analytically in Casini et al. (2021) who showed in a general setting that nonstationarity/misspecification alters the low frequency components of a time series making the latter appear as more persistent. Since traditional LRV estimators are a weighted sum of a large number of low frequency periodogram ordinates, these estimates turn to be inflated. Similarly, LRV estimators using long bandwidths (i.e., fixed-b) 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 that use a smaller number of sample autocovariances. This explains why KVB's fixed-b HAR tests are subject to more power problems, even though the classical HAC estimators are also largely affected.

The introduction of the smoothing over time in the DK-HAC estimators avoids the low frequency contamination because observations belonging to different regimes are not mixed up when computing sample autocovariances. This guarantees good power properties also under nonstationarity/misspecification or under nonstationary alternative hypotheses (e.g., HAR tests for forecast evaluation discussed above). Casini et al. (2021) reconciled this issue with some results in the unit root and long memory literature. Tests for a unit root are known to struggle to reject the unit root hypotheses if a process is second-order stationary (i.e., no unit root) but it is contaminated by breaks in the mean or trend [cf. Perron (1989, 1990)]. Similarly, a short memory sequence contaminated by structural breaks can approximate a long memory series in that the autocorrelation function has the same properties as that of a long memory series [cf. Diebold and Inoue (2001), Hillebrand (2005), McCloskey and Hill (2017) and Mikosch and Stărică (2004)].

7 Conclusions

Economic time series are highly nonstationary. Methods constructed under the assumption of stationarity might then have undesirable properties. This paper developed a theoretical framework for inference in settings where the data may be nonstationary. A new class of double kernel heteroskedasticity and autocorrelation consistent (DK-HAC) estimators was presented. In addition to the usual smoothing procedure over lagged autocovariances, the estimator applies smoothing over time. This is important in order to account flexibly for the variation over time of the structural properties of the economic time series. Optimality results under MSE criterion concerning bandwidths and kernels have been established. A data-dependent method based on the "plug-in" approach has been proposed. There are empirical relevant circumstances where HAR tests, either in linear regression models or other contexts, standardized by existing LRV estimators perform poorly. These may result in size distortions as well as significant power losses, even when the sample size is large. In contrast, when the proposed DK-HAC estimator is used the same HAR tests do not suffer from those issues. DK-HAC estimators lead to HAR tests that have competitive size control relative to fixed-b HAR tests, when the latter work well, and have good power, irrespective of whether there is weak or strong dependence in the data.

References

Altissimo, F., and V. Corradi (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., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

Bai, J., and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66(1), 47–78.

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

Bianchi, F., M. Lettau, and S. C. Ludvigson (2018): "Monetary Policy and Asset Valuation," *Unpublished manuscript, Department of Economics, New York University*.

Brillinger, D. R. (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*.

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

——— (2021): "Supplement to "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models"," *arXiv preprint arXiv:2103.02981*.

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

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

——— (2020a): "Continuous Record Laplace-based Inference About the Break Date in Structural Change Models," *Journal of Econometrics*, 224(1), 3–21.

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

——— (2021a): "Change-Point Analysis of Time Series with Evolutionary Spectra," *arXiv preprint arXiv 2106.02031*.

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

——— (2021c): "Minimax MSE Bounds and Nonlinear VAR Prewhitening for Long-Run Variance Estimation Under Nonstationarity," *arXiv preprint arXiv:2103.02235*.

Chan, K. W. (2020): "Mean-Structure and Autocorrelation Consistent Covariance Matrix Estimation," *Journal of Business and Economic Statistics, forthcoming*.

Chen, B., and Y. Hong (2012): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica*, 80(3), 1157–1183.

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

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

——— (2009): "Local Inference for Locally Stationary Time Series Based on the Empirical Spectral Measure," *Journal of Econometrics*, 151(2), 101–112.

——— (2012): "Locally Stationary Processes," in *Handbook of Statistics*, ed. by T. Subba Rao, S. Subba Rao, and C. Rao, vol. 30, pp. 351–413. Elsevier.

Dahlhaus, R., and L. Giraitis (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., and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68(2), 407–423.

Diebold, F. X., and A. Inoue (2001): "Long Memory and Regime Switching," *Journal of Econometrics*, 105(1), 131–159.

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

Dou, L. (2019): "Optimal HAR Inference," *Unpublished Manuscript, Department of Economics, Princeton University*.

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

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

Giacomini, R., and B. Rossi (2015): "Forecasting in Nonstationary Environments: What Works and What Doesn't in Reduced-Form and Structural Models," *Annual Review of Economics*, 7, 207–229.

Gonçalves, S., and T. J. Vogelsang (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.

Hansen, B. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60(4), 967–972.

——— (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68(3), 575–603.

Hillebrand, E. (2005): "Neglecting Parameter Changes in GARCH models," *Journal of Econometrics*, 129(1), 121–138.

Ibragimov, R., and U. K. Müller (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., and Z. Xiao (2009): "Testing for Changing Mean with Monotonic Power," *Journal of Econometrics*, 148(1), 14–24.

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

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

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

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

Last, M., and R. Shumway (2008): "Detecting Abrupt Changes in a Piecewise Locally Stationary Time Series," *Journal of Multivariate Analysis*, 99(2), 191–214.

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

Inference," *Econometrica, forthcoming.*

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

Martin, W., and P. Flandrin (1985): "Wigner-Ville Spectral Analysis of Nonstationary Processes," *IEEE Transactions on Acoustics Speech and Signal Processing*, 33, 1465–1470.

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

McCloskey, A., and J. B. Hill (2017): "Parameter Estimation Robust to Low Frequency Contamination," *Journal of Business and Economic Statistics*, 35(4), 598–610.

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

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

——— (2014): "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32(3), 311–322.

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

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

Ng, S., and J. H. Wright (2013): "Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling," *Journal of Economic Literature*, 51(4), 1120–54.

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

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

Perron, P. (1990): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, 8(2), 153–162.

Perron, P. (2006): "Dealing with Structural Breaks," in *Palgrave Handbook of Econometrics*, ed. by K. Patterson, and T. Mills, vol. 1: Econometric Theory, pp. 278–352. Palgrave Macmillan.

Perron, P., and Y. Yamamoto (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., and D. Preinerstorfer (2018): "Controlling the Size of Autocorrelation Robust Tests," *Journal of Econometrics*, 207(2), 406–431.

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

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

Rho, S., and T. J. Vogelsang (2020): "Inference in Time Series Models using Smoothed-Clustered Standard Errors," *Journal of Econometrics*, forthcoming.

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

Robinson, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by H. P. Springer, Berlin, Heidelberg.

——— (1998): "Inference-Without Smoothing in the Presence of Nonparametric Autocorrelation," *Econometrica*, 66(5), 1163–1182.

Shao, X., and X. Zhang (2010): "Testing for Change Points in Time Series," *Journal of the American Statistical Association*, 105(491), 122–1240.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Stability in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 14(1), 11–30.

Sun, Y. (2013): "Heteroscedasticity and Autocorrelation Robust F Test using Orthonormal Series Variance Estimator," *Econometrics Journal*, 16(1), 1–26.

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

——— (2014b): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference," *Journal of Econometrics*, 178(3), 659–677.

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

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

Whilelm, D. (2015): "Optimal Bandwidth Selection for Robust Generalized Methods of Moments Estimation," *Econometric Theory*, 31(5), 1054–1077.

Xu, K.-L. (2013): "Powerful Tests for Structural Changes in Volatility," *Journal of Econometrics*, 173(1), 126–142.

Zhang, T., and L. Lavitas (2018): "Unsupervised Self-Normalized Change-Point Testing for Time Series," *Journal of the American Statistical Association*, 113(522), 637–648.

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

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

A Appendix

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

	Model M1, t_1		Model M2, t_2	
5% nominal size	$T = 200$	$T = 400$	$T = 200$	$T = 400$
$\widehat{J}_T$	0.086	0.067	0.054	0.038
$\widehat{J}_{T,\mathrm{pw},1}$	0.052	0.047	0.060	0.043
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$	0.053	0.041	0.069	0.037
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$	0.048	0.044	0.065	0.039
Andrews (1991)	0.081	0.060	0.082	0.072
Andrews (1991), prewhite	0.059	0.048	0.062	0.047
Newey-West (1987)	0.091	0.068	0.058	0.052
Newey-West (1987), prewhite	0.073	0.054	0.071	0.064
Newey-West (1987), fixed-b (KVB)	0.057	0.054	0.059	0.059

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

	Model M3, t_2		Model M4, t_2	
5% nominal size	$T = 200$	$T = 400$	$T = 200$	$T = 400$
$\widehat{J}_T$	0.063	0.056	0.064	0.054
$\widehat{J}_{T,\mathrm{pw},1}$	0.013	0.011	0.008	0.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$	0.062	0.061	0.063	0.043
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$	0.056	0.054	0.034	0.042
Andrews (1991)	0.047	0.025	0.043	0.016
Andrews (1991), prewhite	0.013	0.019	0.000	0.000
Newey-West (1987)	0.072	0.064	0.023	0.032
Newey-West (1987), prewhite	0.014	0.016	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.003	0.001	0.000	0.000

Table 3: Empirical small-sample size for model M5-M6

	M5, t_1		M6, t_2	
5% nominal size	$T=200$	$T=400$	$T=200$	$T=400$
$\widehat{J}_T$	0.095	0.093	0.065	0.060
$\widehat{J}_{T,\mathrm{pw},1}$	0.057	0.056	0.044	0.049
$\widehat{J}_{T,\mathrm{pw,SLS}}$	0.064	0.059	0.056	0.052
$\widehat{J}_{T,\mathrm{pw,SLS},\mu}$	0.067	0.065	0.059	0.056
Andrews (1991)	0.081	0.058	0.049	0.048
Andrews (1991), prewhite	0.069	0.051	0.040	0.044
Newey-West (1987)	0.111	0.084	0.052	0.048
Newey-West (1987), prewhite	0.078	0.057	0.048	0.045
Newey-West (1987), fixed-b (KVB)	0.063	0.054	0.034	0.035

Table 4: Empirical small-sample size for model M7-M8

	DM test		GR test	
5% nominal size	$T_n=200$	$T_n=400$	$T_n=240$	$T_n=380$
$\widehat{J}_T$	0.035	0.063	0.029	0.043
$\widehat{J}_{T,\mathrm{pw},1}$	0.026	0.031	0.028	0.033
$\widehat{J}_{T,\mathrm{pw,SLS}}$	0.045	0.042	0.036	0.039
$\widehat{J}_{T,\mathrm{pw,SLS},\mu}$	0.043	0.046	0.047	0.045
Andrews (1991)	0.083	0.085	0.000	0.000
Andrews (1991), prewhite	0.082	0.085	0.000	0.003
Newey-West (1987)	0.080	0.083	0.000	0.000
Newey-West (1987), prewhite	0.079	0.083	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.002	0.002	0.068	0.049

Table 5: Empirical small-sample power for model M1

5% nominal size, $T=200$	$\delta=0.2$	$\delta=0.4$	$\delta=0.8$	$\delta=1.6$
$\widehat{J}_T$	0.481	0.924	1.000	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.394	0.887	1.000	1.000
$\widehat{J}_{T,\mathrm{pw,SLS}}$	0.381	0.907	1.000	1.000
$\widehat{J}_{T,\mathrm{pw,SLS},\mu}$	0.370	0.907	1.000	1.000
Andrews (1991)	0.479	0.943	1.000	1.000
Andrews (1991), prewhite	0.436	0.899	1.000	1.000
Newey-West (1987)	0.549	0.961	1.000	1.000
Newey-West (1987), prewhite	0.454	0.934	1.000	1.000
Newey-West (1987), fixed-b (KVB)	0.323	0.769	0.998	1.000

Table 6: Empirical small-sample power for model M2

5% nominal size, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.6$	$\delta = 0.8$
$\widehat{J}_T$	0.153	0.403	0.906	0.996	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.150	0.366	0.858	0.987	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$	0.177	0.390	0.878	0.992	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$	0.175	0.388	0.876	0.990	1.000
Andrews (1991)	0.203	0.503	0.930	0.997	0.999
Andrews (1991), prewhite	0.148	0.416	0.914	0.997	1.000
Newey-West (1987)	0.163	0.448	0.925	0.998	1.000
Newey-West (1987), prewhite	0.178	0.463	0.924	0.997	1.000
Newey-West (1987), fixed-b (KVB)	0.133	0.332	0.781	0.957	0.995

Table 7: Empirical small-sample power for model M3

5% nominal size, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.6$	$\delta = 2.5$
$\widehat{J}_T$	0.165	0.230	0.488	0.811	0.975	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.020	0.047	0.189	0.545	0.913	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$	0.080	0.131	0.303	0.661	0.954	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$	0.068	0.105	0.275	0.651	0.931	1.000
Andrews (1991)	0.097	0.242	0.570	0.836	0.967	1.000
Andrews (1991), prewhite	0.026	0.074	0.254	0.599	0.874	1.000
Newey-West (1987)	0.108	0.195	0.448	0.793	0.976	1.000
Newey-West (1987), prewhite	0.035	0.094	0.298	0.627	0.874	1.000
Newey-West (1987), fixed-b (KVB)	0.012	0.061	0.254	0.605	0.882	0.996

Table 8: Empirical small-sample power for model M4

5% nominal size, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.6$	$\delta = 3$
$\widehat{J}_T$	0.112	0.135	0.310	0.645	0.969	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.010	0.021	0.073	0.339	0.856	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS}}$	0.064	0.089	0.166	0.431	0.874	1.000
$\widehat{J}_{T,\mathrm{pw},\mathrm{SLS},\mu}$	0.039	0.051	0.104	0.332	0.832	1.000
Andrews (1991)	0.108	0.218	0.484	0.749	0.915	0.995
Andrews (1991), prewhite	0.000	0.000	0.007	0.186	0.708	0.956
Newey-West (1987)	0.031	0.071	0.200	0.538	0.931	1.000
Newey-West (1987), prewhite	0.000	0.000	0.033	0.280	0.740	0.965
Newey-West (1987), fixed-b (KVB)	0.000	0.009	0.096	0.398	0.753	0.952

Table 9: Empirical small-sample power for model M5

5% nominal size, $T = 200$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.6$	$\delta = 2.5$
$\widehat{J}_T$	0.365	0.705	0.935	0.977	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.213	0.446	0.717	0.795	0.890
$\widehat{J}_{T,\mathrm{pw,SLS}}$	0.232	0.511	0.792	0.908	1.000
$\widehat{J}_{T,\mathrm{pw,SLS},\mu}$	0.242	0.542	0.804	0.902	1.000
Andrews (1991)	0.249	0.427	0.532	0.816	0.718
Andrews (1991), prewhite	0.214	0.320	0.122	0.035	0.340
Newey-West (1987)	0.319	0.737	0.849	0.918	0.937
Newey-West (1987), prewhite	0.212	0.563	0.146	0.062	0.403
Newey-West (1987), fixed-b (KVB)	0.095	0.108	0.127	0.132	0.143

Table 10: Empirical small-sample power for model M6

			M6, t_2		
5% nominal size, $T = 200$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.6$
$\widehat{J}_T$	0.186	0.504	0.945	1.000	1.000
$\widehat{J}_{T,\mathrm{pw},1}$	0.112	0.360	0.888	0.995	1.000
$\widehat{J}_{T,\mathrm{pw,SLS}}$	0.103	0.339	0.884	0.996	1.000
$\widehat{J}_{T,\mathrm{pw,SLS},\mu}$	0.102	0.334	0.888	0.996	1.000
Andrews (1991)	0.201	0.564	0.936	0.998	1.000
Andrews (1991), prewhite	0.169	0.499	0.916	0.996	1.000
Newey-West (1987)	0.223	0.543	0.935	0.990	1.000
Newey-West (1987), prewhite	0.215	0.530	0.925	0.996	1.000
Newey-West (1987), fixed-b (KVB)	0.131	0.368	0.776	0.974	1.000

Table 11: Empirical small-sample power of the DM (1995) test

	Model M7				
5% nominal size, $T = 400$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_T$	0.323	0.451	0.925	0.970	1.000
$\widehat{J}_{T,\text{pw},1}$	0.245	0.365	0.914	0.964	0.972
$\widehat{J}_{T,\text{pw,SLS}}$	0.351	0.505	0.922	0.962	1.000
$\widehat{J}_{T,\text{pw,SLS},\mu}$	0.341	0.499	0.934	1.000	1.000
$\widehat{J}_T$, auto, pretest	0.329	0.457	0.932	1.000	1.000
$\widehat{J}_{T,\text{pw,SLS}}$, pretest	0.372	0.516	0.942	1.000	1.000
Andrews (1991)	0.300	0.350	0.151	0.000	0.000
Andrews (1991), prewhite	0.293	0.345	0.371	0.080	0.000
Newey-West (1987)	0.297	0.350	0.598	0.817	0.782
Newey-West (1987), prewhite	0.288	0.314	0.191	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.231	0.201	0.000	0.000	0.000

Table 12: Empirical small-sample power of the GR (2009) test

	Model M8				
5% nominal size, $T = 800$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.6$	$\delta = 2.5$
$\widehat{J}_T$	0.066	0.496	0.999	1.000	1.000
$\widehat{J}_{T,\text{pw},1}$	0.059	0.491	0.997	1.000	1.000
$\widehat{J}_{T,\text{pw,SLS}}$	0.082	0.406	0.995	1.000	1.000
$\widehat{J}_{T,\text{pw,SLS},\mu}$	0.104	0.560	0.996	1.000	1.000
Andrews (1991)	0.000	0.350	0.000	0.000	0.000
Andrews (1991), prewhite	0.000	0.345	0.133	0.591	0.742
Newey-West (1987)	0.000	0.350	0.598	0.000	0.000
Newey-West (1987), prewhite	0.000	0.314	0.191	0.000	0.000
Newey-West (1987), fixed-b (KVB)	0.026	0.201	0.000	0.000	0.000

B Supplemental Materials

The supplement for online publication [cf. Casini (2021)] reviews how to apply the proposed DK-HAC estimator in GMM and IV contexts and contains the proofs of the results of Section 3. An additional supplement, not for publication, includes the proofs of the results of Section 2 and 4-5.

Supplement to "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models"

Alessandro Casini

Department of Economics and Finance
University of Rome Tor Vergata

26th July 2021

Abstract

This supplemental material is for online publication only. It contains the proofs of the results of Section 3 in the paper.

S.A Appendix: Proofs of the Results of Section 3

In the proofs below, 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 $\widehat{\Gamma}(k)$. In some of the proofs below we first consider the locally stationary case under Assumption S.A.1-S.A.2 and then extend the results to the SLS case. Note that Assumption S.A.1-S.A.2 are implied by Assumption 3.1-3.2 since the former are weaker because local stationarity does not allow for break points in the spectrum. A function $G(\cdot,\cdot):[0,1]\times\mathbb{R}\rightarrow\mathbb{C}$ is said to be right-differentiable at u_0 if $\partial G(u_0,\omega)/\partial_+ u \triangleq \lim_{u\rightarrow u_0^+}(G(u_0,\omega)-G(u,\omega))/(u-u_0)$ exists for any $\omega\in\mathbb{R}$. We sometimes use $\sum_t$ omitting the limits of the summation for the sum in $\widetilde{c}_T(u,k)$.

Assumption S.A.1. *$\{V_{t,T}\}$ is a mean-zero locally stationary process, $A(u,\omega)$ is twice differentiable in u with uniformly bounded and Lipschitz continuous 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 $\vartheta=1$.*

Assumption S.A.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 $\widetilde{\kappa}_{a,b,c,d}:[0,1]\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\rightarrow\mathbb{R}$ 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 with uniformly bounded and Lipschitz continuous derivative $(\partial^2/\partial u^2)\widetilde{\kappa}_{a,b,c,d}(u,\cdot,\cdot,\cdot)$.*

S.A.1 Preliminary Lemmas

Lemma S.A.1. *Under Assumption 3.1-3.2,*

$$\sup_{u\in\{\{(0,1)\}/\{\lambda_j^0,\,j=1,\ldots,m_0\}\}}\sup_{k\in\mathbb{Z}}\left\|\mathrm{Cov}\left(V_{\lfloor Tu\rfloor,T},\,V_{\lfloor Tu\rfloor-k,T}\right)-c(u,k)\right\|=O(T^{-1}),$$
$$\sup_{u\in(0,1)}\sup_{k\geq0}\left\|\mathrm{Cov}\left(V_{\lfloor Tu\rfloor,T},\,V_{\lfloor Tu\rfloor-k,T}\right)-c(u,k)\right\|=O(T^{-1}),$$
$$\max_{u=\lambda_j^0,\,j=1,\ldots,m_0}\sup_{k<0}\left\|\mathrm{Cov}\left(V_{\lfloor Tu\rfloor,T},\,V_{\lfloor Tu\rfloor-k,T}\right)-c(u,-k)\right\|=O(T^{-1}).$$

Proof of Lemma S.A.1. It is sufficient to consider the scalar case $p=1$. Consider first $Tu\notin\mathcal{T}$, $\lambda_{j-1}^0<u<\lambda_j^0$. Using the spectral representation (2.2), (2.4) and Assumption 2.1 leads to

$$\begin{aligned}\mathrm{Cov}\left(V_{\lfloor Tu\rfloor,T},\,V_{\lfloor Tu\rfloor-k,T}\right)&=\int_{-\pi}^{\pi}\exp(i\omega k)\,A_{j,\lfloor Tu\rfloor,T}^0(\omega)\,A_{j,\lfloor Tu\rfloor-k,T}^0(-\omega)\,d\omega\\&=\int_{-\pi}^{\pi}\exp(i\omega k)\,A_j(u,\omega)\,A_j(u-k/T,\,-\omega)\,d\omega+O(T^{-1})\\&=c(u,k)+O(T^{-1}),\end{aligned}\tag{S.1}$$

where the $O(T^{-1})$ term is uniform in $u\in\{\{(0,1)\}/\{\lambda_j^0,\,j=1,\ldots,m_0\}\}$ and k. Now consider the case $Tu\in\mathcal{T}$, $u=T_j^0/T$ and $k\geq0$. Using (2.2) and (2.4) yields

$$\mathrm{Cov}\left(V_{\lfloor Tu\rfloor,T},\,V_{\lfloor Tu\rfloor-k,T}\right)=\int_{-\pi}^{\pi}\exp(i\omega k)\,A_{j,\lfloor Tu\rfloor,T}^0(\omega)\,A_{j+1,\lfloor Tu\rfloor-k,T}^0(-\omega)\,d\omega$$

$$
\begin{aligned}
&= \int_{-\pi}^{\pi} \exp(i\omega k) A_j(u,\,\omega) A_{j+1}(u-k/T,\,-\omega)\, d\omega + O(T^{-1}) \\
&= c(u,\,-k) + O(T^{-1}),
\end{aligned} \tag{S.2}
$$

where the $O(T^{-1})$ term is uniform in u and $k \geq 0$. The argument for the case $Tu \in \mathcal{T}$ and $k<0$ is the same as for the case $Tu \notin \mathcal{T}$. □

Lemma S.A.2. *Under Assumption S.A.1-S.A.2,* $\sup_{u\in(0,\,1)} \sup_{v,\,k\in\mathbb{Z}} ||\Gamma_u(v) - \Gamma_{u+k/T}(v)|| = O(T^{-1})$.

Proof of Lemma S.A.2. We know that $\Gamma_u(v) = c(u,\,v) + O(T^{-1})$ uniformly in u and v by Lemma S.A.1 where $c(u,\,v) = \int_{-\pi}^{\pi} e^{i\omega v} f(u,\,\omega)\, d\omega$. Using Assumption S.A.1,

$$
\begin{aligned}
c(u,\,v) &= \int_{-\pi}^{\pi} e^{i\omega v} f(u+k/T,\,\omega)\, d\omega + O(k/T) \\
&= c(u+k/T,\,v) + O(k/T) \\
&= \Gamma_{u+k/T}(v) + O(k/T) + O(T^{-1}),
\end{aligned}
$$

uniformly in $u \in (0,\,1)$ and $v,\,k \in \mathbb{Z}$. □

Let

$$
\mathrm{MSE}(\widetilde{c}_T(u_0,\,k)) = Tb_{2,T}\mathbb{E}\left[\mathrm{vec}(\widetilde{c}_T(u_0,\,k) - c(u_0,\,k))' W \mathrm{vec}(\widetilde{c}_T(u_0,\,k) - c(u_0,\,k))\right],
$$

where W is some $p^2 \times p^2$ weight matrix.

Lemma S.A.3. *Suppose Assumption S.A.1-S.A.2 hold and $b_{2,T} \to 0$ as $T \to \infty$. Then, for all $u_0 \in (0,\,1)$,*

$$
\mathbb{E}[\widetilde{c}_T(u_0,\,k)] = c(u_0,\,k) + \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2(x)\, dx \left[\frac{\partial^2}{\partial^2 u} c(u_0,\,k)\right] + o\left(b_{2,T}^2\right) + O(1/(Tb_{2,T})), \tag{S.3}
$$

and for all $j,\,l,\,r,\,w \leq p$,

$$
\begin{aligned}
&\mathrm{Cov}\left[\widetilde{c}_T^{(j,l)}(u_0,\,k),\,\widetilde{c}_T^{(r,w)}(u_0,\,k)\right] \qquad \text{(S.4)} \\
&= \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} \left[c^{(j,r)}(u_0,\,l)\, c^{(l,w)}(u_0,\,l) + c^{(j,w)}(u_0,\,l)\, c^{(l,r)}(u_0,\,l+2k)\right] \\
&\quad + \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{h_1=-\infty}^{\infty} \widetilde{\kappa}_{j,l,r,w}(u_0,\,-k,\,h_1,\,h_1-k) + o\left(\frac{1}{Tb_{2,T}}\right).
\end{aligned}
$$

If $Tb_{2,T}^5 \to \eta \in (0,\,\infty)$, then, for all $u_0 \in (0,\,1)$, $\widetilde{c}_T(u_0,\,k) - c(u_0,\,k) = O_{\mathbb{P}}\left(\sqrt{Tb_{2,T}}\right)$.

If in addition $V_{t,T}$ is Gaussian, then for all $u_0 \in (0,\,1)$,

$$
\begin{aligned}
&\mathrm{Cov}\left[\widetilde{c}_T^{(j,l)}(u_0,\,k),\,\widetilde{c}_T^{(r,w)}(u_0,\,k)\right] \\
&= \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} \left[c^{(j,r)}(u_0,\,l)\, c^{(l,w)}(u_0,\,l) + c^{(j,w)}(u_0,\,l)\, c^{(l,r)}(u_0,\,l+2k)\right] \qquad \text{(S.5)} \\
&\quad + o(1/(Tb_{2,T})),
\end{aligned}
$$

and if $Tb_{2,T}^5 \to \eta \in (0,\, \infty)$*, then*

$$\lim_{T\to\infty} \mathrm{MSE}\left(\widetilde{c}_T\left(u_0,\, k\right)\right) = \frac{\eta}{4}\left(\int_0^1 x^2 K_2\left(x\right) dx\right)^2 \left[\frac{\partial^2}{\partial^2 u}\mathrm{vec}\left(c\left(u_0,\, k\right)\right)\right]' W \left[\frac{\partial^2}{\partial^2 u}\mathrm{vec}\left(c\left(u_0,\, k\right)\right)\right]$$
$$+ \int_0^1 K_2^2\left(x\right) dx\, \mathrm{tr}W \sum_{l=-\infty}^{\infty} \mathrm{vec}\left(c\left(u_0,\, l\right)\right)\left[\mathrm{vec}\left(c\left(u_0,\, l\right)\right)' + \mathrm{vec}\left(c\left(u_0,\, l+2k\right)\right)'\right].$$

Proof of Lemma S.A.3. The bias expression follows from Dahlhaus (1997). For the second moment and MSE of $\widetilde{c}_T\left(u_0,\, k\right)$, we first present the proof for the case where $V_{t,T}$ is Gaussian and $p=1$. Evaluating the expectation, we have for $k<0$,

$$\begin{aligned}
&\mathrm{Var}\left[\widetilde{c}_T\left(u_0,\, k\right)\right] \\
&\quad= \frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\mathbb{E}\left(V_{t,T}V_{s,T}\right)\mathbb{E}\left(V_{t+k,T}V_{s+k,T}\right) \\
&\quad+ \frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\mathbb{E}\left(V_{t,T}V_{t+k,T}\right)\mathbb{E}\left(V_{s,T}V_{s+k,T}\right) \\
&\quad+ \frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\mathbb{E}\left(V_{t,T}V_{s+k,T}\right)\mathbb{E}\left(V_{s,T}V_{t+k,T}\right) \\
&\quad- \left[\mathbb{E}\left(\widetilde{c}_T\left(u_0,\, k\right)\right)\right]^2.
\end{aligned}$$

Using the continuity of K_2, $\left(s-t\right)/T \to 0$ for fixed s and t, the smoothness of $\Gamma_u\left(\cdot\right)$ and Lemma S.A.1, implies that the first term on the right-hand side is equal to

$$\frac{1}{Tb_{2,T}}\int_0^1 x^2 K_2\left(x\right) dx \sum_{l=-\infty}^{\infty} c\left(u_0,\, l\right)^2.$$

For the second and third terms we use a similar argument with in addition Lemma 6.2.1 in Fuller (1995) so that

$$\begin{aligned}
&\mathrm{Var}\left[\widetilde{c}_T\left(u_0,\, k\right)\right] \\
&\quad= \frac{1}{Tb_{2,T}}\int_0^1 x^2 K_2\left(x\right) dx \sum_{l=-\infty}^{\infty} c\left(u_0,\, l\right)^2 + \frac{1}{Tb_{2,T}}\int_0^1 x^2 K_2\left(x\right) dx \sum_{l=-\infty}^{\infty} c\left(u_0,\, l\right) c\left(u_0,\, l+2k\right) + o\left(\frac{1}{Tb_{2,T}}\right).
\end{aligned} \tag{S.6}$$

Next, (S.5) follows similarly. We have

$$\begin{aligned}
&\mathrm{Cov}\left[\widetilde{c}_T^{(j,l)}\left(u_0,\, k\right),\, \widetilde{c}_T^{(r,w)}\left(u_0,\, k\right)\right] \\
&\quad= \frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\mathbb{E}\left(V_{t,T}^{(j)}V_{s,T}^{(r)}\right)\mathbb{E}\left(V_{t+k,T}^{(l)}V_{s+k,T}^{(w)}\right) \\
&\quad+ \frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\mathbb{E}\left(V_{t,T}^{(j)}V_{s+k,T}^{(w)}\right)\mathbb{E}\left(V_{t+k,T}^{(l)}V_{s,T}^{(r)}\right)
\end{aligned}$$

$$= \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} \left[c^{(j,r)}(u_0,\, l)\, c^{(l,w)}(u_0,\, l) + c^{(j,w)}(u_0,\, l)\, c^{(l,r)}(u_0,\, l+2k) \right] + o\left(1/\left(Tb_{2,T}\right)\right).$$

Using a standard bias-variance argument, we have $\widetilde{c}_T(u_0,\, k) - c(u_0,\, k) = o_{\mathbb{P}}(1)$. If $Tb_{2,T}^5 \to \eta \in (0,\, \infty)$, the asymptotic MSE of $\widetilde{c}_T(u_0,\, k)$ is given by

$$\lim_{T\to\infty} \mathrm{MSE}\left(\widetilde{c}_T(u_0,\, k)\right) = \frac{\eta}{4}\left(\int_0^1 x^2 K_2(x)\, dx\right)^2 \left[\frac{\partial^2}{\partial^2 u} c(u_0,\, k)\right]^2 + \int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} c(u_0,\, l)\left[c(u_0,\, l) + c(u_0,\, l+2k)\right]. \tag{S.7}$$

The latter suggests that if $Tb_{2,T}^5 \to \eta \in (0,\, \infty)$, then $\widetilde{c}_T(u_0,\, k) - c(u_0,\, k) = O_{\mathbb{P}}\left(\sqrt{Tb_{2,T}}\right)$ for all $u_0 \in (0,\, 1)$. The MSE expression for the multivariate case follows from (S.7).

Consider now the second moment of $\widetilde{c}_T(u_0,\, k)$ for the general case. When $V_{t,T}$ is non-Gaussian, there is an extra term in $\mathrm{Cov}[\widetilde{c}_T^{(j,l)}(u_0,\, k),\, \widetilde{c}_T^{(r,w)}(u_0,\, k)]$, namely

$$\frac{1}{\left(Tb_{2,T}\right)^2} \sum_t \sum_s K_2^*\left(\frac{u_0 - (t+k/2)/T}{b_T}\right) K_2^*\left(\frac{u_0 - (s+k/2)/T}{b_T}\right) \kappa_{V,t}^{(j,l,r,w)}(-k,\, s-t,\, s-t-k).$$

By Assumption S.A.2 with $u = t/T$,

$$\sup_{u\in(0,\,1)} \left| \kappa_{V,Tu}^{(j,l,r,w)}(-k,\, s-Tu,\, s-Tu-k) - \widetilde{\kappa}_{j,l,r,w}(u,\, -k,\, s-Tu,\, s-Tu-k) \right| = O(T^{-1}).$$

Taking a second-order Taylor's expansion of $\kappa_{V,Tu}^{(j,l,r,w)}$ around u_0 we have

$$\begin{aligned}
&\frac{1}{\left(Tb_{2,T}\right)^2} \sum_t \sum_s K_2^*\left(\frac{u_0 - (t+k/2)/T}{b_T}\right) K_2^*\left(\frac{u_0 - (s+k/2)/T}{b_T}\right) \kappa_{V,Tu}^{(j,l,r,w)}(-k,\, s-Tu,\, s-Tu-k) \\
&\quad = \frac{1}{\left(Tb_{2,T}\right)^2} \sum_t \sum_s K_2^*\left(\frac{u_0 - (t+k/2)/T}{b_T}\right) K_2^*\left(\frac{u_0 - (s+k/2)/T}{b_T}\right) \\
&\qquad \times \widetilde{\kappa}_{j,l,r,w}(u_0,\, -k,\, s-Tu_0,\, s-Tu_0-k) \\
&\qquad + \frac{1}{\left(Tb_{2,T}\right)^2} \sum_t \sum_s K_2^*\left(\frac{u_0 - (t+k/2)/T}{b_T}\right) K_2^*\left(\frac{u_0 - (s+k/2)/T}{b_T}\right) \\
&\qquad \times \frac{\partial \widetilde{\kappa}_{j,l,r,w}}{\partial u}(u_0,\, -k,\, s-Tu_0,\, s-Tu_0-k)(u_0-u) \\
&\qquad + \frac{1}{\left(Tb_{2,T}\right)^2} \sum_t \sum_s K_2^*\left(\frac{u_0 - (t+k/2)/T}{b_T}\right) K_2^*\left(\frac{u_0 - (s+k/2)/T}{b_T}\right) \\
&\qquad \times \frac{\partial^2 \widetilde{\kappa}_{j,l,r,w}}{\partial u^2}(u_0,\, -k,\, s-Tu_0,\, s-Tu_0-k)(u_0-u)^2 \\
&\quad = \frac{1}{Tb_{2,T}} \int_0^1 K_2^2(x)\, dx \sum_{h_1=-\infty}^{\infty} \widetilde{\kappa}_{j,l,r,w}(u_0,\, -k,\, h_1,\, h_1-k) + o\left(\frac{1}{Tb_{2,T}}\right). \square
\end{aligned}$$

Lemma S.A.4. *Suppose Assumption 3.1-3.2 hold and $b_{2,T} \to 0$ as $T \to \infty$. For each $T\lambda_j^0 = Tu_0 \in \mathcal{T}$ ($j = 1, \ldots, m_0$) and $|k|/Tb_{2,T} \to \eta_2 \in (0, \lambda_{j+1}^0 - \lambda_j^0)$,*

$$\mathbb{E}[\widetilde{c}_T(u_0, k)] = c(u_0, k) + \frac{1}{2}b_{2,T}^2 \int_0^1 x^2 K_2(x)\, dx$$
$$\times \int_{-\pi}^{\pi} \exp(i\omega k)\left(C_1(u_0, \omega) + C_2(u_0, \omega) + C_3(u_0, \omega)\right) d\omega + O\left(\frac{1}{Tb_{2,T}}\right) + o\left(b_{2,T}^2\right),$$

where

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

and $v_0 = u_0 - k/2T$. For $Tu_0 \notin \mathcal{T}$ or for $Tu_0 \in \mathcal{T}$ and $|k|/Tb_{2,T} \to 0$, (S.3) and (S.4) hold. For all $u_0 \in (0, 1)$, $\lim_{T\to\infty} b_{2,T}^{-2}\, \mathbb{E}[\widetilde{c}_T(u_0, k) - c(u_0, k)] < \infty$, and if further it holds that $Tb_{2,T}^5 \to \eta \in (0, \infty)$, then $\lim_{T\to\infty} Tb_{2,T}\mathrm{Var}[\widetilde{c}_T(u_0, k)] < \infty$. Furthermore, we have $\widehat{c}_T(u_0, k) - c(u_0, k) = O_{\mathbb{P}}(\sqrt{Tb_{2,T}})$ for all $u_0 \in (0, 1)$.

Proof of Lemma S.A.4. If $Tu_0 \notin \mathcal{T}$ then the result follows from Lemma S.A.3. Suppose $Tu_0 \in \mathcal{T}$ and $k/Tb_{2,T} \to 0$ (the case $k < 0$ is similar and omitted). We omit the subscript j from $A_{j,s-k,T}^0(\omega)$ and from $A_j((s-k)/T, \omega)$ since the value j is determined by s and $s-k$, respectively, and can thus be omitted. Using (2.2) we have,

$$\mathbb{E}[\widetilde{c}_T(u_0, k)] = \frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T} K_2^*\left(\frac{u_0 - (s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi} \exp(i\omega k)\, A_{s-k,T}^0(\omega)\, A_{s,T}^0(-\omega)\, d\omega.$$

Since $K_2(x) = 0$ for $x < 0$, the above sum runs up to $s = Tu_0 + k/2T$. Hence, the behavior of $A_{s,T}^0(\omega)$ only matters on a left neighborhood of u_0. Using (2.4) we have,

$$\mathbb{E}[\widetilde{c}_T(u_0, k)] = \frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T} K_2^*\left(\frac{u_0 - (s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi} \exp(i\omega k)\, A\left(\frac{s-k}{T}, \omega\right) A\left(\frac{s}{T}, -\omega\right) d\omega + O\left(T^{-1}\right).$$

By the definition of $f(\cdot, \cdot)$, it follows that,

$$\mathbb{E}[\widetilde{c}_T(u_0, k)] = \frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T} K_2^*\left(\frac{u_0 - (s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi} \exp(i\omega k)\, f\left(\frac{s-k/2}{T}, \omega\right) d\omega + O\left(T^{-1}\right).$$

Let $u_{\epsilon,T} = u_0 - \epsilon_T$, where $\epsilon_T > 0$. Since $f(u, \omega)$ is twice differentiable in u at $u \neq \lambda_j^0$ (cf. Assumption 3.1), by taking a second-order Taylor's expansion of f around $u_{\epsilon,T}$ we have

$$\mathbb{E}[\widetilde{c}_T(u_0, k)] = \frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T} K_2^*\left(\frac{u_0 - (s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi} \exp(i\omega k)\, f(u_{\epsilon,T}, \omega)\, d\omega$$

$$+\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_{2}^{*}\left(\frac{u_{0}-(s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)\frac{\partial f(u_{\epsilon,T},\,\omega)}{\partial u}\left(\frac{s-k/2}{T}-u_{\epsilon,T}\right)d\omega$$
$$+\frac{1}{2}\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_{2}^{*}\left(\frac{u_{0}-(s-k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)\frac{\partial^{2} f(u_{\epsilon,T},\,\omega)}{\partial u^{2}}\left(\frac{s-k/2}{T}-u_{\epsilon,T}\right)^{2}d\omega$$
$$+o\left(b_{2,T}^{2}\right)+O\left(\frac{1}{Tb_{2,T}}\right).$$

Choose $\epsilon_T = o_{\mathbb{P}}(\max\{b_{2,T}^2,\, 1/(Tb_{2,T})\})$. Using $\int_0^1 K_2(x)\,dx = 1$, $K_2(x) = K_2(1-x)$ and the definition of $c(u_{\epsilon,T},\, k)$, the right-hand side above is equal to

$$c(u_{\epsilon_T},\, k)+\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}(x)\,dx\int_{-\pi}^{\pi}\exp(i\omega k)\frac{\partial^{2} f(u_{\epsilon_T},\,\omega)}{\partial u^{2}}d\omega+O\left(\frac{1}{Tb_{2,T}}\right)+o\left(b_{2,T}^{2}\right).$$

Since $c(u_0,\, k)$ and $\partial^2 f(u_0,\, \omega)/\partial u^2$ are left-Lipschitz continuous by Assumption 3.1-(iii),

$$c(u_{\epsilon_T},\, k)-c(u_0,\, k)=O_{\mathbb{P}}(\epsilon_T),\qquad \frac{\partial^{2} f(u_{\epsilon_T},\,\omega)}{\partial u^{2}}-\frac{\partial^{2} f(u_0,\,\omega)}{\partial_{-} u^{2}}=O_{\mathbb{P}}(\epsilon_T).$$

Then,

$$\mathbb{E}\left[\widetilde{c}_{T}(u_0,\, k)-c(u_0,\, k)\right]=\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}(x)\,dx\int_{-\pi}^{\pi}\exp(i\omega k)\frac{\partial^{2} f(u_0,\,\omega)}{\partial_{-} u^{2}}d\omega+O\left(\frac{1}{Tb_{2,T}}\right)+o\left(b_{2,T}^{2}\right).$$

It remains to consider the case $Tu_0 = T\lambda_j^0 \in \mathcal{T}$ and $|k|/T \to \eta_2 \in (0,\, \lambda_{j+1}^0 - \lambda_j^0)$. Suppose $k < 0$ (the case $k > 0$ is similar and omitted). The derivations for the bias expression are different. Again, using (2.2) we have,

$$\mathbb{E}\left[\widetilde{c}_{T}(u_0,\, k)\right]=\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_{2}^{*}\left(\frac{u_{0}-(s+k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)A_{s+k,T}^{0}(\omega)A_{s,T}^{0}(-\omega)\,d\omega.$$

Using (2.4), we have

$$\mathbb{E}\left[\widetilde{c}_{T}(u_0,\, k)\right]=\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_{2}^{*}\left(\frac{u_{0}-(s+k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)A\left(\frac{s+k}{T},\,\omega\right)A\left(\frac{s}{T},\,-\omega\right)d\omega+O\left(T^{-1}\right).$$

We cannot use the property $f_j(u,\, \omega) = |A_j(u,\, \omega)|^2$ for $T_{j-1}^0/T < u = t/T \le T_j^0/T$ because now $u_0 = s + k/2$ implies $s = Tu_0 - k/2 > Tu_0$. That is, $A_j((s+k)/T,\, \omega)\, A_{j+1}(s/T,\, -\omega)$ cannot be approximated by $f_j(s - k/2,\, \omega)$ for those s such that $s > T_j^0$. However, by taking a second-order Taylor's expansion of A_j about $u_0 - \epsilon_{1,T}$ and of A_{j+1} about $v_0 + \epsilon_{2,T}$ where $v_0 = u_0 - k/2T$ and $\epsilon_{1,T},\, \epsilon_{2,T} > 0$, we have

$$\mathbb{E}\left[\widetilde{c}_{T}(u_0,\, k)\right]$$
$$=\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_{2}^{*}\left(\frac{u_{0}-(s+k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)A_{j+1}(v_0+\epsilon_{2,T},\,\omega)A_{j}(u_0-\epsilon_{1,T},\,-\omega)\,d\omega$$

$$
\begin{aligned}
&+\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_2^*\left(\frac{u_0-(s+k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)\left[\frac{\partial A_{j+1}(v_0+\epsilon_{2,T},\,\omega)}{\partial v}\right.\\
&\times A_j(u_0-\epsilon_{1,T},\,-\omega)\left(\frac{s}{T}-v_0-\epsilon_{2,T}\right)\\
&\left.+\frac{\partial A_j(u_0-\epsilon_{1,T},\,-\omega)}{\partial u}A_{j+1}(v_0+\epsilon_{2,T},\,\omega)\left(\frac{s+k/2}{T}-u_0+\epsilon_{1,T}\right)\right]d\omega\\
&+\frac{1}{2}\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_2^*\left(\frac{u_0-(s+k/2)/T}{b_{2,T}}\right)\int_{-\pi}^{\pi}\exp(i\omega k)\left[\frac{\partial^2 A_{j+1}(v_0+\epsilon_{2,T},\,\omega)}{\partial v^2}\right.\\
&\times A_j(u_0-\epsilon_{1,T},\,-\omega)\left(\frac{s}{T}-v_0-\epsilon_{2,T}\right)^2\\
&\left.+\frac{\partial^2 A_j(u_0-\epsilon_{1,T},\,-\omega)}{\partial u^2}A_{j+1}(v_0+\epsilon_{2,T},\,\omega)\left(\frac{s+k/2}{T}-u_0+\epsilon_{1,T}\right)^2\right]d\omega\\
&+\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}K_2^*\left(\frac{u_0-(s+k/2)/T}{b_{2,T}}\right)\\
&\times\int_{-\pi}^{\pi}\exp(i\omega k)\left[\frac{\partial A_{j+1}(v_0+\epsilon_{2,T},\,\omega)}{\partial v}\frac{\partial A_j(u_0-\epsilon_{1,T},\,-\omega)}{\partial u}\left(\frac{s}{T}-v_0-\epsilon_{2,T}\right)\left(\frac{s+k/2}{T}-u_0+\epsilon_{1,T}\right)\right]d\omega \qquad \text{(S.8)}\\
&+o\left(b_{2,T}^2\right).
\end{aligned}
$$

By Assumption 3.1, $A_j(u,\,-\omega)$, $\partial A_j(u,\,-\omega)/\partial u$ and $\partial^2 A_j(u,\,-\omega)/\partial u^2$ are left-continuous at $u=u_0$, and $A_{j+1}(u,\,\omega)$, $\partial A_{j+1}(u,\,\omega)/\partial u$ and $\partial^2 A_{j+1}(u,\,\omega)/\partial u^2$ are right-continuous at $u=v_0$, thus we have,

$$
\begin{aligned}
&A_j(u_0-\epsilon_{1,T},\,-\omega)-A_j(u_0,\,-\omega)=O_{\mathbb{P}}(\epsilon_{1,T}), \qquad \frac{\partial A_j(u_0-\epsilon_{1,T},\,-\omega)}{\partial u}-\frac{\partial A_j(u_0,\,-\omega)}{\partial_- u}=O_{\mathbb{P}}(\epsilon_{1,T}),\\
&\frac{\partial^2 A_j(u_0-\epsilon_{1,T},\,-\omega)}{\partial u^2}-\frac{\partial^2 A_j(u_0,\,-\omega)}{\partial_- u^2}=O_{\mathbb{P}}(\epsilon_{1,T})\\
&A_{j+1}(v_0+\epsilon_{2,T},\,\omega)-A_{j+1}(v_0,\,\omega)=O_{\mathbb{P}}(\epsilon_{2,T}), \qquad \frac{\partial A_{j+1}(v_0+\epsilon_{2,T},\,\omega)}{\partial v}-\frac{\partial A_{j+1}(v_0,\,\omega)}{\partial_+ v}=O_{\mathbb{P}}(\epsilon_{2,T}),\\
&\frac{\partial^2 A_{j+1}(v_0+\epsilon_{2,T},\,\omega)}{\partial v^2}-\frac{\partial^2 A_{j+1}(v_0,\,\omega)}{\partial_+ v^2}=O_{\mathbb{P}}(\epsilon_{2,T}).
\end{aligned}
$$

Choose $\epsilon_{1,T}=o_{\mathbb{P}}(\max\{b_{2,T}^2,\,1/(Tb_{2,T})\})$ and $\epsilon_{2,T}=o_{\mathbb{P}}(\max\{b_{2,T}^2,\,1/(Tb_{2,T})\})$. Using the definition of $c(u_0,\,k)$ for $k<0$, (S.8) is equal to,

$$
\begin{aligned}
&c(u_0,\,k)+b_{2,T}^2\int_0^1 x^2K_2(x)\,dx\int_{-\pi}^{\pi}\exp(i\omega k)\left(C_1(u_0,\,\omega)+C_2(u_0,\,\omega)+C_3(u_0,\,\omega)\right)d\omega\\
&\quad+O\left(\frac{1}{Tb_{2,T}}\right)+o\left(b_{2,T}^2\right).
\end{aligned}
$$

For $Tu_0\notin\mathcal{T}$, (S.3) and (S.4) follow by a similar proof as for Lemma S.A.3. Next, let us consider $\mathrm{Var}[\widetilde{c}_T(u_0,\,k)]$ for $p=1$ and $V_{t,T}$ Gaussian. Assume $u_0=\lambda_j^0$ and $|k|/Tb_{2,T}\to\eta_2\in(0,\,\lambda_{j+1}^0-\lambda_j^0)$, we have for $k<0$,

$$\begin{aligned}
&\mathrm{Var}\left[\widetilde{c}_T\left(u_0,\,k\right)\right]\\
&\quad=\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)\mathbb{E}\left(V_{t,T}V_{s,T}\right)\mathbb{E}\left(V_{t+k,T}V_{s+k,T}\right)\\
&\qquad+\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)\mathbb{E}\left(V_{t,T}V_{s+k,T}\right)\mathbb{E}\left(V_{s,T}V_{t+k,T}\right).
\end{aligned}$$

By (2.4), $A^0_{s+k,T}\left(\omega\right)A^0_{t,T}\left(-\omega\right)=A_j\left(\left(s+k\right)/T,\,\omega\right)A_{j+1}\left(t/T,\,-\omega\right)+O(T^{-1})$ and $A^0_{t,T}\left(\omega\right)A^0_{s,T}\left(-\omega\right)=A_{j+1}\left(t/T,\,\omega\right)A_{j+1}\left(s/T,\,-\omega\right)+O(T^{-1})$ for $s,\,t=Tu_0-k/2$. Now take a second order Taylor's expansion of A_{j+1} around $v_0=u_0-k/2T+\epsilon_{2,T}$ and of A_j around $u_{\epsilon,T}=u_0-\epsilon_T$, where $\epsilon_{2,T},\,\epsilon_T>0$. Applying the manipulations in (S.8) involving A_j and A_{j+1} combined with the same derivations that led to (S.6) we obtain,

$$\begin{aligned}
\mathrm{Var}\left[\widetilde{c}_T\left(u_0,\,k\right)\right]=\int_0^1 K_2\left(x\right)^2dx\Bigg\{&\sum_{l=-\infty}^{\infty}\left[c\left(v_0,\,l\right)c\left(u_0,\,l+2k\right)\right]\\
&+\sum_{l=-\infty}^{0}\left[c\left(u_0,\,l\right)c\left(u_0,\,l\right)\right]+\sum_{l=1}^{\infty}\left[c\left(v_0,\,l\right)c\left(v_0,\,l\right)\right]\Bigg\},
\end{aligned}\tag{S.9}$$

where $c\left(u_0,\,\cdot\right)$ in the second line above takes the form [cf. the definition of $c\left(u_0,\,l\right)$ for $l<0$ at the end of Section 2.1],

$$c\left(u_0,\,l\right)=\int_{-\pi}^{\pi}\exp\left(i\omega l\right)A_j\left(u_0,\,\omega\right)A_{j+1}\left(u_0-l/T,\,\omega\right)d\omega.$$

When $V_{t,T}$ is non-Gaussian, there is an extra term in $\mathrm{Cov}[\widetilde{c}_T^{(j,l)}\left(u_0,\,k\right),\,\widetilde{c}_T^{(r,w)}\left(u_0,\,k\right)]$, namely

$$\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)\kappa_{V,t}^{(j,l,r,w)}\left(-k,\,s-t,\,s-t-k\right).\tag{S.10}$$

By Assumption 3.2 with $u=t/T$,

$$\sup_{1\leq j\leq m_0+1}\sup_{\lambda^0_{j-1}<u\leq\lambda^0_j}\left|\kappa_{V,Tu}^{(j,l,r,w)}\left(-k,\,s-Tu,\,s-Tu-k\right)-\widetilde{\kappa}_{j,l,r,w}\left(u,\,-k,\,s-Tu,\,s-Tu-k\right)\right|=O\left(T^{-1}\right).$$

Taking a second-order Taylor's expansion of $\kappa_{V,Tu}^{(j,l,r,w)}$ with respect to the first argument around $v_0=u_0-k/2T+\epsilon_{2,T}$ with $\epsilon_{2,T}>0$, we have

$$\begin{aligned}
&\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)\kappa_{V,t}^{(j,l,r,w)}\left(-k,\,s-t,\,s-t-k\right)\\
&\qquad=\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)
\end{aligned}\tag{S.11}$$

$$\times \widetilde{\kappa}_{j,l,r,w}\left(v_0,\,-k,\,s-Tv_0,\,s-Tv_0-k\right)$$
$$+\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)$$
$$\times \frac{\partial\widetilde{\kappa}_{j,l,r,w}}{\partial v}\left(v_0,\,-k,\,s-Tv_0,\,s-Tv_0-k\right)\left(v_0-t/T\right)$$
$$+\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_{2,T}}\right)$$
$$\times \frac{\partial^2\widetilde{\kappa}_{j,l,r,w}}{\partial v^2}\left(v_0,\,-k,\,s-Tv_0,\,s-Tv_0-k\right)\left(v_0-t/T\right)^2+O\left(T^{-1}\right).$$

Let $\epsilon_{2,T}=o_{\mathbb{P}}(\max\{b_{2,T}^2,\,1/\left(Tb_{2,T}\right)\})$. Since $\widetilde{\kappa}_{j,l,r,w}\left(v_0,\,\cdot,\,\cdot,\,\cdot\right)$ is uniformly piecewise Lipschitz continuous by Assumption 3.2-(ii), the first term on the right-hand side above is equal to

$$\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{u_0-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{u_0-\left(s+k/2\right)/T}{b_T}\right)$$
$$\times\left(\widetilde{\kappa}_{j,l,r,w}\left(v_0,\,-k,\,s-Tv_0,\,s-Tv_0-k\right)+O\left(T^{-1}\right)\right).$$

The second and third term of (S.11) are of smaller order $o\left(1/Tb_{2,T}\right)$. Thus, (S.10) is equal to

$$\frac{1}{Tb_{2,T}}\int_0^1 K_2^2\left(x\right)dx\sum_{h_1=-\infty}^{\infty}\widetilde{\kappa}_{j,l,r,w}\left(v_0,\,-k,\,h_1,\,h_1-k\right)+o\left(\frac{1}{Tb_{2,T}}\right).$$

It remains to derive the expressions for $\mathrm{Var}\left[\widetilde{c}_T\left(u_0,\,k\right)\right]$ and $\mathrm{Cov}[\widetilde{c}_T^{(j,l)}\left(u_0,\,k\right),\,\widetilde{c}_T^{(r,w)}\left(u_0,\,k\right)]$ for the case $\left|k\right|/Tb_{2,T}\rightarrow 0$. As seen when studying the bias, the behavior of $A^0_{\cdot,T}\left(\cdot\right)$ only matters on a left neighborhood of u_0 and thus the result remains the same as in the locally stationary case. The argument involves using first a Taylor's expansion around $u_0-\epsilon_{1,T}$ with $\epsilon_{1,T}>0$ and then exploiting left-Lipschitz continuity. As in the proof of Lemma S.A.3, basic manipulations lead to the bound for the MSE. Then, consistency and the rate of convergence follow from the same arguments used there. □

Lemma S.A.5. *Cosnider $p=1$. Under Assumption 3.1-3.2,* $\sup_{k\geq1}Tb_{2,T}\mathrm{Var}(\widetilde{\Gamma}(k))=O\left(1\right).$

Proof of Lemma S.A.5. We have for $k\geq0$,

$$\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)=\left(\frac{n_T}{T}\right)^2\sum_{r=0}^{\lfloor T/n_T\rfloor}\sum_{w=0}^{\lfloor T/n_T\rfloor}\mathrm{Cov}\left(\widetilde{c}_T\left(rn_T/T,\,k\right),\,\widetilde{c}_T\left(wn_T/T,\,k\right)\right),$$

with

$$\mathrm{Cov}\left(\widetilde{c}_T\left(rn_T/T,\,k\right),\,\widetilde{c}_T\left(wn_T/T,\,k\right)\right)$$
$$=\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{rn_T/T-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{wn_T/T-\left(s+k/2\right)/T}{b_T}\right)$$
$$\times\mathbb{E}\left(V_{t,T}V_{t+k,T}V_{s,T}V_{s+k,T}\right)$$
$$+\frac{1}{\left(Tb_{2,T}\right)^2}\sum_t\sum_s K_2^*\left(\frac{rn_T/T-\left(t+k/2\right)/T}{b_T}\right)K_2^*\left(\frac{wn_T/T-\left(s+k/2\right)/T}{b_T}\right)$$

$$
\begin{aligned}
&\times \mathbb{E}\left(V_{t,T}V_{s,T}\right)\mathbb{E}\left(V_{t+k,T}V_{s+k,T}\right)\\
&+\frac{1}{\left(Tb_{2,T}\right)^{2}}\sum_{t}\sum_{s}K_{2}^{*}\left(\frac{rn_{T}/T-\left(t+k/2\right)/T}{b_{T}}\right)K_{2}^{*}\left(\frac{wn_{T}/T-\left(s+k/2\right)/T}{b_{T}}\right)\\
&\times \mathbb{E}\left(V_{t,T}V_{t+k,T}\right)\mathbb{E}\left(V_{s,T}V_{s+k,T}\right)\\
&+\frac{1}{\left(Tb_{2,T}\right)^{2}}\sum_{t}\sum_{s}K_{2}^{*}\left(\frac{rn_{T}/T-\left(t+k/2\right)/T}{b_{T}}\right)K_{2}^{*}\left(\frac{wn_{T}/T-\left(s+k/2\right)/T}{b_{T}}\right)\\
&\times \mathbb{E}\left(V_{t,T}V_{s+k,T}\right)\mathbb{E}\left(V_{s,T}V_{t+k,T}\right)-\mathbb{E}\left(\widetilde{c}_{T}\left(rn_{T}/T,\,k\right)\right)\mathbb{E}\left(\widetilde{c}_{T}\left(wn_{T}/T,\,k\right)\right).
\end{aligned}
$$

Proceeding as in the proof of Lemma S.A.4, we have

$$
\begin{aligned}
&\mathrm{Cov}\left(\widetilde{c}_{T}\left(rn_{T}/T,\,k\right),\,\widetilde{c}_{T}\left(wn_{T}/T,\,k\right)\right)\\
&=\frac{1}{Tb_{2,T}}\int_{0}^{1}K_{2}\left(x\right)dx\sum_{h_{1}=-\infty}^{\infty}\widetilde{\kappa}\left(rn_{T}/T,-k,\,h_{1},\,h_{1}-k\right)\\
&\quad+\frac{1}{Tb_{2,T}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{l=-\infty}^{\infty}c\left(rn_{T}/T,\,l\right)c\left(wn_{T}/T,\,l\right)\\
&\quad+\frac{1}{Tb_{2,T}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dx\sum_{l=-\infty}^{\infty}c\left(rn_{T}/T,\,l\right)c\left(wn_{T}/T,\,l+2k\right)+o\left(\frac{1}{Tb_{2,T}}\right),
\end{aligned}
$$

where $\widetilde{\kappa}=\widetilde{\kappa}_{1,1,1,1}$ is the cumulant for the univariate case. Note that

$$
\begin{aligned}
\sum_{l=-\infty}^{\infty}c\left(rn_{T}/T,\,l\right)c\left(wn_{T}/T,\,l+2k\right)&\leq\sum_{l=-\infty}^{\infty}\left|c\left(rn_{T}/T,\,l\right)\right|\sum_{s=-\infty}^{\infty}\left|c\left(wn_{T}/T,\,s+2k\right)\right|\\
&\leq\sum_{l=-\infty}^{\infty}\left|c\left(rn_{T}/T,\,l\right)\right|\sum_{s=-\infty}^{\infty}\left|c\left(wn_{T}/T,\,s\right)\right|.
\end{aligned}
$$

The desired result then follows by Assumption 3.2-(i) and the convergence of approximation to Riemann sums. □

S.A.2 Proofs of the Results of Section 3

S.A.2.1 Proof of Lemma 3.1

It follows by Lemma S.A.4. □

S.A.2.2 Proof of Theorem 3.1

We first prove the result for the locally stationary case (i.e., $m=0$) and then extend it to the general case $m>0$. We begin with the result for the scalar case ($p=1$) and then extend it to the vector case.

Lemma S.A.6. *Suppose $p=1$, $K_{1}(\cdot)\in\boldsymbol{K}_{1}$, $K_{2}(\cdot)\in\boldsymbol{K}_{2}$, Assumption S.A.1-S.A.2 hold, $b_{1,T}$, $b_{2,T}\rightarrow0$, $n_{T}\rightarrow\infty$, $n_{T}/T\rightarrow0$ and $1/Tb_{1,T}b_{2,T}\rightarrow0$. We have:*

(i) $\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}(\widetilde{J}_{T})=4\pi^{2}\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx(\int_{0}^{1}f\left(u,\,0\right)du)^{2}$.

(ii) If $1/Tb_{1,T}^{q}b_{2,T} \to 0$, $n_T/Tb_{1,T}^{q} \to 0$ and $b_{2,T}^{2}/b_{1,T}^{q} \to 0$ for some $q \in [0,\,\infty)$ for which $K_{1,q}$, $|\int_0^1 f^{(q)}(u,\,0)\,du| \in [0,\,\infty)$, then $\lim_{T\to\infty} b_{1,T}^{-q}[\mathbb{E}(\widetilde{J}_T - J_T)] = -2\pi K_{1,q}\int_0^1 f^{(q)}(u,\,0)\,du$.

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

$$
\begin{aligned}
&\lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widehat{J}_T,\,1\right)\\
&\qquad = 4\pi^2\left[\gamma K_{1,q}^2\left(\int_0^1 f^{(q)}(u,\,0)\,du\right)^2 + \int K_1^2(y)\,dy\int K_2^2(x)\,dx\left(\int_0^1 f(u,\,0)\,du\right)^2\right].
\end{aligned}
$$

Proof of Lemma S.A.6. We begin with part (i). Note that for any fixed non-negative $\tau_1,\,\tau_2 \in \mathbb{R}$,

$$
\begin{aligned}
&\mathrm{Cov}\left(V_sV_{s-\tau_1},\,V_lV_{l-\tau_2}\right)\\
&\quad = \mathbb{E}\left[\left(V_sV_{s-\tau_1} - \mathbb{E}\left(V_sV_{s-\tau_1}\right)\right)\left(V_lV_{l-\tau_2} - \mathbb{E}\left(V_lV_{l-\tau_2}\right)\right)\right]\\
&\quad = \mathbb{E}\left(V_sV_{s-\tau_1}V_lV_{l-\tau_2}\right) - \Gamma_{s/T}(\tau_1)\Gamma_{l/T}(\tau_2) - \Gamma_{s/T}(\tau_1)\Gamma_{l/T}(\tau_2) - \Gamma_{l/T}(l-s)\Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)\\
&\qquad - \Gamma_{(l-\tau_2)/T}(l-s-\tau_2)\Gamma_{l/T}(l-s+\tau_1) + \Gamma_{s/T}(\tau_1)\Gamma_{l/T}(\tau_2) + \Gamma_{l/T}(l-s)\Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)\\
&\qquad + \Gamma_{(l-\tau_2)/T}(l-s-\tau_2)\Gamma_{l/T}(l-s+\tau_1)\\
&\quad = \kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2) + \Gamma_{s/T}(s-l)\Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)\\
&\qquad + \Gamma_{s/T}(s-l+\tau_2)\Gamma_{l/T}(l-s+\tau_1)\\
&\quad = \kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2) + \Gamma_{l/T}(l-s)\Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)\\
&\qquad + \Gamma_{(l-\tau_2)/T}(l-s-\tau_2)\Gamma_{l/T}(l-s+\tau_1).
\end{aligned}
$$

For large T, we have by Lemma S.A.2: $\Gamma_{(l-\tau_2)/T}(k) - \Gamma_{l/T}(k) = O(\tau_2/T)$, and $\Gamma_{(s-\tau_1)/T}(k) = \Gamma_{s/T}(k) + O(\tau_1/T)$ uniformly in k, l and s. Apply the changes in variables $w = s-l$ and $v = l$, then

$$
\begin{aligned}
&\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}\mathrm{Cov}\left(V_sV_{s-\tau_1},\,V_lV_{l-\tau_2}\right)\\
&\qquad = \sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}\kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2)\\
&\qquad\quad + \sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}\Big[\Gamma_{v/T}(-w)\Gamma_{v/T}(-w+\tau_2-\tau_1) + \Gamma_{v/T}(-w-\tau_2)\Gamma_{v/T}(-w+\tau_1)\Big]\\
&\qquad\quad + \sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}\Big[\Gamma_{v/T}(-w)\,O(\tau_2/T) + O(\tau_2/T)\,\Gamma_{v/T}(-w+\tau_1)\Big].
\end{aligned}
\tag{S.12}
$$

A bound for the term involving $\Gamma_{v/T}(-w)\,O(\tau_2/T)$ in (S.12) is

$$
\begin{aligned}
\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}\left|\Gamma_{v/T}(-w)\right|O(\tau_2/T) &\le O(\tau_2/T)\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}\sup_{(v/T)\in[0,\,1]}\left|\Gamma_{v/T}(w)\right|\\
&\le O\left(T^{-1}\right),
\end{aligned}
\tag{S.13}
$$

where we have used Assumption S.A.2-(i). The argument for the term involving $O(\tau_2/T)\,\Gamma_{v/T}(-w+\tau_1)$ is analogous. We next evaluate the covariance of $\widetilde{c}_T(t/T,\,k)$. For any $1\leq t_1,\,t_2\leq T$ and (without loss of generality) non-negative integers $\tau_1,\,\tau_2\in\mathbb{R}$, apply the following changes in variables $w=s-l$ and $v=l$, so that

$$
\begin{aligned}
&Tb_{2,T}\mathrm{Cov}\left[\widetilde{c}_T(t_1/T,\,\tau_1),\,\widetilde{c}_T(t_2/T,\,\tau_2)\right]\\
&\quad=Tb_{2,T}\left(\frac{1}{Tb_{2,T}}\right)^2\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}\\
&\qquad\times K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(l-\tau_2/2))/T}{b_{2,T}}\right)\mathrm{Cov}\left(V_sV_{s-\tau_1},\,V_lV_{l-\tau_2}\right)\\
&\quad=\frac{1}{Tb_{2,T}}\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}K_2^*\left(\frac{(t_1-(v+w-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\\
&\qquad\times\left\{\left[\Gamma_{v/T}(-w)\,\Gamma_{v/T}(-w+\tau_2-\tau_1)+\Gamma_{v/T}(-w-\tau_2)\,\Gamma_{v/T}(-w+\tau_1)\right]\right\}\\
&\qquad+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)\\
&\qquad\times K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2)+A_T,
\end{aligned}
$$

where

$$
\begin{aligned}
A_T\triangleq\;&\frac{1}{Tb_{2,T}}\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}K_2^*\left(\frac{(t_1-(v+w-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\\
&\times\left\{\left[\Gamma_{v/T}(w)\,O(\tau_2/T)+O(\tau_2/T)\,\Gamma_{v/T}(w+\tau_1)\right]\right\}.
\end{aligned}
$$

Using (S.13), we have $A_T=o\left(T^{-1}\right)$. Then, using the change of variable $z=v/Tb_{2,T}$,

$$
\begin{aligned}
&Tb_{2,T}\mathrm{Cov}\left[\widetilde{c}_T(t_1/T,\,\tau_1),\,\widetilde{c}_T(t_2/T,\,\tau_2)\right]\\
&\quad=\frac{1}{Tb_{2,T}}\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}K_2^*\left(\frac{(t_1-v-w+\tau_1/2+v-v)/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-v+\tau_2/2)/T}{b_{2,T}}\right)\\
&\qquad\times\left\{\left[\Gamma_v(-w)\,\Gamma_v(-w+\tau_2-\tau_1)+\Gamma_v(-w-\tau_2)\,\Gamma_v(-w+\tau_1)\right]\right\}\\
&\qquad+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\\
&\qquad\times\kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2)+A_T\\
&\quad=\frac{1}{Tb_{2,T}}\sum_{z=(\tau_2+1)/Tb_{2,T}}^{1/b_{2,T}}\sum_{w=\tau_1+1-Tb_{2,T}z}^{T-Tb_{2,T}z}K_2^*\left(\frac{(t_1+w+\tau_1/2)/T}{b_{2,T}}-z\right)K_2^*\left(\frac{(t_2+\tau_2/2)/T}{b_{2,T}}-z\right)\qquad\text{(S.14)}\\
&\qquad\times\left\{\left[\Gamma_{zTb_{2,T}}(-w)\,\Gamma_{zTb_{2,T}}(-w+\tau_2-\tau_1)+\Gamma_{zTb_{2,T}}(-w-\tau_2)\,\Gamma_{zTb_{2,T}}(-w+\tau_1)\right]\right\}
\end{aligned}
$$

$$+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s+\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v+\tau_2/2))/T}{b_{2,T}}\right)$$
$$\times\,\kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2)+A_T.$$

Thus, with $u=t_1/T$ and $v=t_2/T$, the limit of the first term of (S.14) is equal to

$$\int_0^1 K_2^2(x)\,dx\left\{\sum_{w=-\infty}^{\infty}\left[\Gamma_u(w)\,\Gamma_v(-w+\tau_2-\tau_1)+\Gamma_u(w+\tau_2)\,\Gamma_v(-w+\tau_1)\right]\right\}.$$

When $\tau_1=\tau_2=k$ and $t=t_1=t_2$, we have

$$\begin{aligned}Tb_{2,T}\mathrm{Var}(\widetilde{c}_T(t/T,\,k))&=\int_0^1 K_2(x)^2\,dx\left\{\sum_{w=-\infty}^{\infty}\left[\Gamma_u(w)\,\Gamma_u(w)+\Gamma_u(w+k)\,\Gamma_u(w-k)\right]\right\}\\&=\int_0^1 K_2(x)^2\,dx\left\{\sum_{h=-\infty}^{\infty}\left[\Gamma_u(h)\,\Gamma_u(h)+\Gamma_u(h+2k)\,\Gamma_u(h)\right]\right\},\end{aligned}$$

where $u=t/T$ and we have used the change in variable $h=w-k$. Next, we consider $\mathrm{Cov}[\widetilde{\Gamma}(\tau_1),\,\widetilde{\Gamma}(\tau_2)]$. Note that,

$$\begin{aligned}&Tb_{2,T}\mathrm{Cov}\left[\widetilde{\Gamma}(\tau_1),\,\widetilde{\Gamma}(\tau_2)\right]\\&\quad\to\int_0^1 K_2^2(x)\,dx\int_0^1\int_0^1\left\{\sum_{h=-\infty}^{\infty}\left[\Gamma_u(h)\,\Gamma_u(h-\tau_2+\tau_1)+\Gamma_v(-h-\tau_2)\,\Gamma_v(-h-\tau_1)\right]\right\}dvdu.\end{aligned}$$

The latter can be used to evaluate $\mathrm{Var}[\sum_{k=-T+1}^{T-1}K_1(b_{1,T}k)\,\widetilde{\Gamma}(k)]$ as follows,

$$\begin{aligned}&Tb_{1,T}b_{2,T}\mathrm{Var}\left[\sum_{k=-T+1}^{T-1}K_1(b_{1,T}k)\,\widetilde{\Gamma}(k)\right]\\&\quad=2b_{1,T}\sum_{k=-T+1}^{T-1}\sum_{j=0}^{T-1}K_1(b_{1,T}k)\,K_1(b_{1,T}j)\qquad\qquad(\text{S.15})\\&\qquad\times\left(\frac{n_T}{T}\right)^2\sum_{r=0}^{T/n_T}\sum_{b=0}^{T/n_T}\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}\sum_{l=j+1}^{T}K_2^*\left(\frac{(rn_T+1)-(s+k/2)}{Tb_{2,T}}\right)K_2^*\left(\frac{(bn_T+1)-(l+j/2)}{Tb_{2,T}}\right)\\&\qquad\times\left(\left[\Gamma_{l/T}(l-s)\,\Gamma_{l/T}(l-s-j+k)+\Gamma_{l/T}(-s+l-\tau_2)\,\Gamma_{l/T}(-s+l+k)\right]\right.\\&\qquad\left.+\,\kappa_{V,s}(-k,\,l-s,\,l-s-j)\right)+o(1),\end{aligned}$$

where the $o(1)$ term follows from $A_T=o(b_{1,T}/T)$. The term involving $\kappa_{V,s}(-k,\,l-s,\,l-s-j)$ is dominated by

$$Cb_{1,T}\left|\sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty}\sum_{w=-\infty}^{\infty}\sup_{s}\kappa_{V,s}\left(-k,\,-w,\,-w-j\right)\right|=O\left(b_{1,T}\right),$$

where $C<\infty$ and we have used Assumption S.A.2-(i). Now let $w=s-l$ and $v=l$ and rewrite (S.15) as

$$\begin{aligned}
&Tb_{1,T}b_{2,T}\mathrm{Var}\left[\sum_{k=-T+1}^{T-1}K_{1}\left(b_{1,T}k\right)\widetilde{\Gamma}\left(k\right)\right]\\
&\quad=2b_{1,T}\sum_{k=-T+1}^{T-1}\sum_{j=0}^{T-1}K_{1}\left(b_{1,T}k\right)K_{1}\left(b_{1,T}j\right)\\
&\quad\times\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}\frac{1}{Tb_{2,T}}\sum_{v=j+1}^{T}\sum_{w=k+1-v}^{T-v}\\
&\quad\times K_{2}^{*}\left(\frac{\left(rn_{T}+1\right)-\left(w+v-k/2\right)}{Tb_{2,T}}\right)K_{2}^{*}\left(\frac{\left(bn_{T}+1\right)-\left(v-j/2\right)}{Tb_{2,T}}\right)\\
&\quad\times\left[\Gamma_{v/T}\left(-w\right)\Gamma_{v/T}\left(-w+j-k\right)+\Gamma_{v/T}\left(-w-j\right)\Gamma_{v/T}\left(-w+k\right)\right]+o\left(1\right)+O\left(b_{1,T}\right).
\end{aligned}$$

We next show that the term involving $\Gamma_{v/T}\left(-w-j\right)\Gamma_{v/T}\left(-w+k\right)$ vanishes in the limit. Using a change in variables $z_1=j+k$ and $z=w+j$, the latter is bounded by

$$\begin{aligned}
&4b_{1,T}\sum_{j=0}^{T-1}\sum_{z_{1}=j}^{T-1+j}K_{1}\left(b_{1,T}\left(z_{1}-j\right)\right)K_{1}\left(b_{1,T}j\right)\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}\\
&\quad\times\frac{1}{Tb_{2,T}}\sum_{v=j+1}^{T}\sum_{z=\left(z_{1}-j\right)+1-v+j}^{T-v+j}K_{2}^{*}\left(\frac{\left(rn_{T}+1\right)-\left(z-j+v-\left(z_{1}-j\right)/2\right)}{Tb_{2,T}}\right)\\
&\quad\times K_{2}^{*}\left(\frac{\left(\left(bn_{T}+1\right)-\left(v-j/2\right)\right)/T}{b_{2,T}}\right)\left[\Gamma_{v/T}\left(-z\right)\Gamma_{v/T}\left(-z+z_{1}\right)\right].
\end{aligned}\tag{S.16}$$

Making the change in variable $z_2=jb_{1,T}$, (S.16) can be expressed as,

$$\begin{aligned}
&4b_{1,T}\sum_{z_{2}=0}^{(T-1)/b_{1,T}}\sum_{z_{1}=z_{2}/b_{1,T}}^{T-1+z_{2}/b_{1,T}}K_{1}\left(b_{1,T}\left(z_{1}-z_{2}/b_{1,T}\right)\right)K_{1}\left(z_{2}\right)\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}\\
&\quad\times\frac{1}{Tb_{2,T}}\sum_{v=z_{2}/b_{1,T}+1}^{T}\sum_{z=z_{1}+1-v}^{T-v+z_{2}/b_{1,T}}K_{2}^{*}\left(\frac{\left(rn_{T}+1\right)-\left(z-z_{2}/b_{1,T}+v-\left(z_{1}-z_{2}/b_{1,T}\right)/2\right)}{Tb_{2,T}}\right)\\
&\quad\times K_{2}^{*}\left(\frac{\left(\left(bn_{T}+1\right)-\left(v-z_{2}/2b_{1,T}\right)\right)/T}{b_{2,T}}\right)\left[\Gamma_{v/T}\left(-z\right)\Gamma_{v/T}\left(-z+z_{1}\right)\right],
\end{aligned}$$

which converges to zero because the range of summation over z_1 tends to infinity.

Next, let us consider the term of (S.15) involving $\Gamma_{v/T}\left(-w\right)\Gamma_{v/T}\left(-w+j-k\right)$. With the changes in variables $u_1 = k - j$ and $u_2 = j$, this term becomes

$$b_{1,T} \sum_{u_2=-T+1}^{T-1} \sum_{u_1=-u_2-T+1}^{T-1-u_2} K_1\left(b_{1,T}\left(u_2+u_1\right)\right) K_1\left(b_{1,T} u_2\right) \left(\frac{n_T}{T}\right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \frac{1}{T b_{2,T}} \sum_{v=u_2+1}^{T} \sum_{w=u_2+u_1+1-v}^{T-v} \tag{S.17}$$
$$\times K_2^*\left(\frac{\left(r n_T+1\right)-\left(w+v-\left(u_1+u_2\right)/2\right)}{T b_{2,T}}\right) K_2^*\left(\frac{\left(b n_T+1\right)-\left(v-u_2/2\right)}{T b_{2,T}}\right)$$
$$\times \left[\Gamma_{v/T}\left(w\right) \Gamma_{v/T}\left(-w-u_1\right)\right].$$

Apply the change in variable $z = b_{1,T} u_2$ and consider the lattice points $z_n = n b_{1,T}$, where $n = -T, \ldots, T$. As $T \to \infty$, the distance between the lattice points $z_n = n b_{1,T}$ converges to zero and the highest lattice point converges to infinity. Hence, (S.17) can be expressed as,

$$\sum_{z_n=-(T-1)b_{1,T}}^{(T-1)b_{1,T}} \sum_{u_1=-z_n/b_{1,T}-T+1}^{T-1-z_n/b_{1,T}} K_1\left(b_{1,T} u_1+z_n\right) K_1\left(z_n\right) \tag{S.18}$$
$$\times \left(\frac{n_T}{T}\right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \frac{1}{T b_{2,T}}$$
$$\times \sum_{v=z_n/b_{1,T}+1}^{T} \sum_{w=z_n/b_{1,T}+u_1+1-v}^{T-v} K_2\left(\frac{\left(\left(r n_T+1\right)-\left(w+v-\left(z_n/b_{1,T}+u_1\right)/2\right)\right)/T}{b_{2,T}}\right)$$
$$\times K_2\left(\frac{\left(\left(b n_T+1\right)-\left(v-z/2b_{1,T}\right)\right)/T}{b_{2,T}}\right) \left[\Gamma_{v/T}\left(w\right) \Gamma_{v/T}\left(-w-u_1\right)\right].$$

By Lemma S.A.1, $\Gamma_{v/T}\left(w\right) \Gamma_{v/T}\left(-w-u_1\right) = c\left(v/T,\, -w\right) c\left(v/T,\, w+u_1\right) + O\left(T^{-1}\right)$. By taking a second order Taylor's expansion of $c\left(v/T,\, -w\right)$ around $r n_T/T$ and of $c\left(\left(v-u_1/1\right)/T,\, w+u_1/2\right)$ around $b n_T/T$, we have

$$\sum_{v=z_n/b_{1,T}+1}^{T} \sum_{w=z_n/b_{1,T}+u_1+1-v}^{T-v} K_2\left(\frac{\left(\left(r n_T+1\right)-\left(w+v-\left(z_n/b_{1,T}+u_1\right)/2\right)\right)/T}{b_{2,T}}\right)$$
$$\times K_2\left(\frac{\left(\left(b n_T+1\right)-\left(v-z/2b_{1,T}\right)\right)/T}{b_{2,T}}\right) \left[c\left(v/T,\, -w\right) c\left(v/T,\, w+u_1\right)\right]$$
$$= \int_0^1 K_2\left(x\right)^2 dx\, c\left(r n_T/T,\, -w\right) c\left(b n_T/T,\, w+u_1\right)$$
$$+ b_{2,T}^2 \int_0^1 x^2 K_2\left(x\right)^2 dx \frac{\partial}{\partial v} c\left(v,\, -w\right)|_{v=r n_T/T} \frac{\partial}{\partial v} c\left(v,\, w+u_1\right)|_{v=b n_T/T}$$
$$+ 2^{-1} b_{2,T}^2 \int_0^1 x^2 K_2\left(x\right)^2 dx \frac{\partial^2}{\partial v^2} c\left(v,\, -w\right)|_{v=r n_T/T} c\left(b n_T/T,\, w+u_1\right)$$
$$+ 2^{-1} b_{2,T}^2 \int_0^1 x^2 K_2\left(x\right)^2 dx\, c\left(r n_T/T,\, -w\right) \frac{\partial^2}{\partial v^2} c\left(v,\, w+u_1\right)|_{v=b n_T/T}$$

$$+ o\left(b_{2,T}^{2}\right) + O\left(\frac{1}{Tb_{2,T}}\right).$$

We can now use Lemma S.A.1 backward to show that the limit of (S.18) is equal to

$$\begin{aligned}&\int K_1(y)^2 dy \int_0^1 K_2(x)^2 dx \int_0^1 \int_0^1 \sum_{u_1=\infty}^{\infty} \sum_{w=-\infty}^{\infty} \left[\Gamma_u(w)\Gamma_a(w+u_1)\right] duda \\ &\qquad = 4\pi^2 \int K_1(y)^2 dy \int_0^1 K_2(x)^2 dx \left(\int_0^1 f(u,0)\,du\right)\left(\int_0^1 f(a,0)\,da\right).\end{aligned}$$

This proves the result of part (i). We now move to part (ii). Let

$$J_{c,T} \triangleq \int_0^1 c(u,0) + 2\sum_{k=1}^{T-1}\int_0^1 c(u,k)\,du.$$

We begin with the following relationship,

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

Using Lemma S.A.3, we have for any $-T+1 \le k \le T-1$,

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

where the last equality follows from the convergence of approximations to Riemann sums. This leads to,

$$\begin{aligned}&b_{1,T}^{-q}\mathbb{E}\left(\widetilde{J}_T - J_{c,T}\right) \\ &\quad = -b_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_1(b_{1,T}k)\right)\int_0^1 c(u,k)\,du \\ &\qquad + \frac{1}{2}\frac{b_{2,T}^2}{b_{1,T}^q}\int_0^1 x^2K_2(x)\,dx\sum_{k=-T+1}^{T}K_1(b_{1,T}k)\int_0^1\frac{\partial^2}{\partial^2 u}c(u,k)\,du + O\left(\frac{1}{Tb_{1,T}^q b_{2,T}}\right) + O\left(\frac{n_T}{Tb_{1,T}^q}\right) \\ &\quad = -b_{1,T}^{-q}\sum_{k=-T+1}^{T}\left(1-K_1(b_{1,T}k)\right)\int_0^1 c(u,k)\,du\end{aligned}$$

$$-\frac{1}{2}b_{2,T}^{2}\int_{0}^{1}x^{2}K_{2}\left(x\right)dxO\left(1\right)+\frac{1}{2}\frac{b_{2,T}^{2}}{b_{1,T}^{q}}\int_{0}^{1}x^{2}K_{2}\left(x\right)dxO\left(1\right)+O\left(\frac{1}{Tb_{1,T}^{q}b_{2,T}}\right)+O\left(\frac{n_{T}}{Tb_{1,T}^{q}}\right),$$

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

$$\lim_{T\rightarrow\infty}b_{1,T}^{-q}\mathbb{E}\left(\widetilde{J}_{T}-J_{T}\right)=-2\pi K_{1,q}\int_{0}^{1}f^{(q)}\left(u,\,0\right)du,$$

because $b_{2,T}^{2}/b_{1,T}^{q}\rightarrow0$. It remains to show part (iii). Note that $Tb_{1,T}b_{2,T}=Tb_{1,T}b_{2,T}b_{1,T}^{2q}/b_{1,T}^{2q}=b_{1,T}^{-2q}/(1/Tb_{1,T}^{2q+1}b_{2,T})=b_{1,T}^{-2q}/\left(1/\left(\gamma+o\left(1\right)\right)\right)$. Hence, using part (i)-(ii), we deduce the desired result, namely,

$$\begin{aligned}&\lim_{T\rightarrow\infty}\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widetilde{J}_{T},\,1\right)\\&\quad=\lim_{T\rightarrow\infty}b_{1,T}^{-2q}\mathbb{E}\left[\left(\widetilde{J}_{T}-J_{T}\right)^{2}\right]\left(\gamma+o\left(1\right)\right)+\lim_{T\rightarrow\infty}Tb_{1,T}b_{2,T}\mathrm{Var}\left(\widetilde{J}_{T}\right)\\&\quad=4\pi^{2}\left[\gamma K_{1,q}^{2}\left(\int_{0}^{1}f^{(q)}\left(u,\,0\right)du\right)^{2}+\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx\left(\int_{0}^{1}f\left(u,\,0\right)du\right)^{2}\right].\qquad\square\end{aligned}$$

Lemma S.A.7. *Suppose $K_{1}\left(\cdot\right)\in\boldsymbol{K}_{1}$, $K_{2}\left(\cdot\right)\in\boldsymbol{K}_{2}$, Assumption S.A.1-S.A.2 hold, $b_{1,T},\,b_{2,T}\rightarrow0$, $n_{T}\rightarrow\infty$, $n_{T}/T\rightarrow0$ and $1/Tb_{1,T}b_{2,T}\rightarrow0$. Then, part (i)-(iii) of Theorem 3.1 hold.*

Proof of Lemma S.A.7. We begin with part (i). We provide the expression for the asymptotic covariance between the $(a,\,l)$ and $(m,\,n)$ elements of $\widetilde{J}_{T}$:

$$\begin{aligned}&Tb_{1,T}b_{2,T}\mathrm{Cov}\left[\sum_{k=-T+1}^{T-1}K_{1}\left(b_{1,T}k\right)\widetilde{\Gamma}^{(a,l)}\left(k\right),\,\sum_{j=-T+1}^{T-1}K_{1}\left(b_{1,T}j\right)\widetilde{\Gamma}^{(m,n)}\left(j\right)\right]\\&\quad=4b_{1,T}\sum_{k=0}^{T-1}\sum_{j=0}^{T-1}K_{1}\left(b_{1,T}k\right)K_{1}\left(b_{1,T}j\right)\left(\frac{n_{T}}{T}\right)^{2}\sum_{r=0}^{T/n_{T}}\sum_{b=0}^{T/n_{T}}\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}\sum_{h=j+1}^{T}\\&\qquad\times K_{2}^{*}\left(\frac{\left(\left(rn_{T}+1\right)-\left(s-k/2\right)\right)/T}{b_{2,T}}\right)K_{2}^{*}\left(\frac{\left(\left(bn_{T}+1\right)-\left(h-j/2\right)\right)/T}{b_{2,T}}\right)\\&\qquad\times\left\{\kappa_{V,s}^{(a,l,m,n)}\left(-k,\,h-s,\,h-s-j\right)\right.\\&\qquad\left.+\left[\Gamma_{h/T}^{(a,m)}\left(h-s\right)\Gamma_{h/T}^{(l,n)}\left(h-s-j+k\right)+\Gamma_{h/T}^{(a,n)}\left(h-s-j\right)\Gamma_{h/T}^{(l,m)}\left(h-s+k\right)\right]\right\}+o\left(1\right),\end{aligned}\tag{S.19}$$

where the $o\left(1\right)$ term follows from using (S.13). The term involving $\kappa_{V,s}^{(a,l,m,n)}\left(-k,\,h-s,\,h-s-j\right)$ is negligible as for the scalar case. The limit of the term involving $\Gamma_{h/T}^{(a,m)}\left(h-s\right)\Gamma_{h/T}^{(l,n)}\left(h-s+j-k\right)$ is, according to the derivations to prove part (i) of Lemma S.A.6,

$$4\pi^{2}\int K_{1}\left(y\right)^{2}dy\int_{0}^{1}K_{2}\left(x\right)^{2}dx\left(\int_{0}^{1}f^{(a,m)}\left(u,\,0\right)du\right)\left(\int_{0}^{1}f^{(l,n)}\left(v,\,0\right)dv\right).\tag{S.20}$$

Similarly, the limit of the term involving $\Gamma_{h/T}^{(a,n)}(s-h-j)\,\Gamma_{h/T}^{(l,m)}(s-h+k)$ is the same as (S.20) but with m and n interchanged. The commutation-tensor product formula arises from the fact that the asymptotic covariances between $\widetilde{J}_T^{(a,l)}$ and $\widetilde{J}_T^{(m,n)}$ for $a,\,l,\,m,\,n \leq p$ are of the same form as the covariances between $X_a X_l$ and $X_m X_n$, where $X = (X_1, \ldots, X_p)' \sim \mathscr{N}(0,\,\Sigma)$. The formula then follows from $\mathrm{Var}(\mathrm{vec}(XX')) = \mathrm{Var}(X \otimes X) = (I + C_{pp})\,\Sigma \otimes \Sigma$. The proof of part (ii) of the lemma follows that of the scalar case with minor changes. Since part (iii) simply uses part (i)-(ii), it follows that

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\,\widetilde{J}_T,\,W\right)\\
&\qquad= \lim_{T\to\infty} \gamma b_{1,T}^{-2q}\mathbb{E}\left(\widetilde{J}_T - J_T\right)' W \mathbb{E}\left(\widetilde{J}_T - J_T\right) + \lim_{T\to\infty} Tb_{1,T}b_{2,T}\,\mathrm{tr}W\,\mathrm{Var}\left(\mathrm{vec}\left(\widetilde{J}_T\right)\right),
\end{aligned}$$

converges to the desired limit. □

Lemma S.A.8. *Suppose $p = 1$, $K_1(\cdot) \in \boldsymbol{K}_1$, $K_2(\cdot) \in \boldsymbol{K}_2$, Assumption 3.1-3.2 hold, $b_{1,T}, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ and $1/Tb_{1,T}b_{2,T} \to 0$. Then, (i)-(iii) of Lemma S.A.6 continue to hold.*

Proof of Lemma S.A.8. We assume without loss of generality that $m_0 = 1$ and provide the proof only for the single break case. Hence, the break date is T_2^0 (i.e., $T_1^0 = 0$ and $T_3^0 = T$). Note that by standard properties of approximations to Riemann sums, $\overline{\Gamma}(k) \to \int_0^1 (c(u,\,k))\,du$ even when $c(\cdot,\,k)$ has a finite number of discontinuities in u, where

$$\overline{\Gamma}(k) \triangleq \frac{n_T}{T-n_T} \sum_{r=0}^{\lfloor (T-n_T)/n_T \rfloor} c(rn_T/T,\,k).$$

Since the results in Lemma S.A.4 about the order of the bias and variance of $\widetilde{c}_T(u_0,\,k)$ are the same to their counterpart results in Lemma S.A.3, the proof of Lemma S.A.6 can be repeated with the following changes. We begin with part (i). For any fixed non-negative $\tau_1,\,\tau_2 \in \mathbb{R}$,

$$\begin{aligned}
&\mathrm{Cov}\left(V_s V_{s-\tau_1},\,V_l V_{l-\tau_2}\right)\\
&\quad= \kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2) + \Gamma_{l/T}(l-s)\,\Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)\\
&\qquad+ \Gamma_{(l-\tau_2)/T}(l-s-\tau_2)\,\Gamma_{l/T}(l-s+\tau_1).
\end{aligned}$$

When $l = T_2^0$ and $\tau_2 < 0$, Lemma S.A.2 cannot be applied because of the discontinuity in the spectrum of $\{V_{t,T}\}$ at time $t = T_2^0$. Thus, the relation $\Gamma_{(l-\tau_2)/T}(k) - \Gamma_{l/T}(k) = (\tau_2/T)$ for $l = T_2^0$ and $\tau_2 < 0$ does not hold. One has to carry $\Gamma_{(l-\tau_2)/T}(k)$ through the proof. Applying the changes in variables $w = s - l$ and $v = l$, we have

$$\begin{aligned}
&\sum_{s=\tau_1+1}^{T} \sum_{l=\tau_2+1}^{T} \mathrm{Cov}\left(V_{s/T}V_{(s-\tau_1)/T},\,V_{l/T}V_{(l-\tau_2)/T}\right)\\
&\quad= \sum_{s=\tau_1+1}^{T} \sum_{l=\tau_2+1}^{T} \kappa_{V,s}(-\tau_1,\,l-s,\,l-s-\tau_2)\\
&\qquad+ \sum_{v=\tau_2+1}^{T} \sum_{w=\tau_1+1-v}^{T-\tau_2-v} \left[\Gamma_{v/T}(-w)\,\Gamma_{(v-\tau_2)/T}(-w+\tau_2-\tau_1) + \Gamma_{(v-\tau_2)/T}(-w-\tau_2)\,\Gamma_{v/T}(-w+\tau_1)\right].
\end{aligned} \tag{S.21}$$

We next evaluate the covariance of $\widetilde{c}_T(t/T,\,k)$. For any $1 \leq t_1,\, t_2 \leq T$ and (without loss of generality) non-negative integers $\tau_1,\, \tau_2 \in \mathbb{R}$,

$$
\begin{aligned}
&Tb_{2,T}\mathrm{Cov}\left[\widetilde{c}_T(t_1/T,\,\tau_1),\, \widetilde{c}_T(t_2/T,\,\tau_2)\right]\\
&\quad= Tb_{2,T}\left(\frac{1}{Tb_{2,T}}\right)^2 \sum_{s=\tau_1+1}^{T}\sum_{v=\tau_2+1}^{T}\\
&\quad\times K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\mathrm{Cov}\left(V_sV_{s-\tau_1},\, V_lV_{l-\tau_2}\right)\\
&\quad= \frac{1}{Tb_{2,T}}\sum_{v=\tau_2+1}^{T}\sum_{w=\tau_1+1-v}^{T-v}K_2^*\left(\frac{(t_1-(v+w-\tau_1/2))/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\\
&\quad\times\left\{\left[\Gamma_{v/T}(-w)\,\Gamma_{(v-\tau_2)/T}(-w+\tau_2-\tau_1)+\Gamma_{(v-\tau_2)/T}(-w-\tau_2)\,\Gamma_{v/T}(-w+\tau_1)\right]\right\}\\
&\quad+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)\\
&\quad\times K_2^*\left(\frac{(t_2-(v-\tau_2/2))/T}{b_{2,T}}\right)\kappa_{V,s}\left(-\tau_1,\, l-s,\, l-s-\tau_2\right).
\end{aligned}
$$

Then, using the change of variable $z = v/Tb_{2,T}$,

$$
\begin{aligned}
&Tb_{2,T}\mathrm{Cov}\left[\widetilde{c}_T(t_1/T,\,\tau_1),\, \widetilde{c}_T(t_2/T,\,\tau_2)\right]\\
&\quad= \frac{1}{Tb_{2,T}}\sum_{v=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-v-w-\tau_1/2+v-v)/T}{b_{2,T}}\right)K_2^*\left(\frac{(t_2-zTb_{2,T}-\tau_2/2)/T}{b_{2,T}}\right)\\
&\quad\times\left\{\left[\Gamma_{zb_{2,T}}(-w)\,\Gamma_{zb_{2,T}-\tau_2/T}(-w+\tau_2-\tau_1)+\Gamma_{zb_{2,T}-\tau_2/T}(-w-\tau_2)\,\Gamma_{zb_{2,T}}(-w+\tau_1)\right]\right\}\\
&\quad+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)\\
&\quad\times K_2^*\left(\frac{(t_2-(v+\tau_2/2))/T}{b_{2,T}}\right)\kappa_{V,s}\left(-\tau_1,\, l-s,\, l-s-\tau_2\right)\\
&\quad= \frac{1}{Tb_{2,T}}\sum_{z=(\tau_2+1)/Tb_{2,T}}^{1/b_{2,T}}\sum_{w=\tau_1+1-zTb_{2,T}}^{T-z/Tb_{2,T}}K_2^*\left(\frac{(t_1+w-\tau_1/2)/T}{b_{2,T}}-z\right)K_2^*\left(\frac{(t_2-\tau_2/2)/T}{b_{2,T}}-z\right)\\
&\quad\times\left\{\left[\Gamma_{zb_{2,T}}(-w)\,\Gamma_{zb_{2,T}-\tau_2/T}(-w+\tau_2-\tau_1)+\Gamma_{zb_{2,T}-\tau_2/T}(-w-\tau_2)\,\Gamma_{zb_{2,T}}(-w+\tau_1)\right]\right\}\\
&\quad+\frac{1}{Tb_{2,T}}\sum_{s=\tau_1+1}^{T}\sum_{l=\tau_2+1}^{T}K_2^*\left(\frac{(t_1-(s-\tau_1/2))/T}{b_{2,T}}\right)\\
&\quad\times K_2^*\left(\frac{(t_2-(v+\tau_2/2))/T}{b_{2,T}}\right)\kappa_{V,s}\left(-\tau_1,\, l-s,\, l-s-\tau_2\right).
\end{aligned}
\tag{S.22}
$$

By Lemma S.A.1, $\Gamma_{zb_{2,T}}(-w)\,\Gamma_{zb_{2,T}-\tau_2/T}(-w+\tau_2-\tau_1) = c\left(zb_{2,T}/T,\, w\right)c\left(zb_{2,T}-\tau_2/T,\, w-\tau_2+\tau_1\right)+$

$O\left(T^{-1}\right)$. We need to distinguish two cases. The first case involves both t_1 and t_2 being continuity points (i.e., $t_1,\, t_2 \neq T_2^0$). The second case involves either t_1 or t_2 (or both) being discontinuity points (i.e., $t_1 = T_2^0$ or $t_2 = T_2^0$, or $t_1 = t_2 = T_2^0$). The first case is the one considered in Lemma S.A.6 and thus we omit the details. For the second case, we cannot apply the same argument as in Lemma S.A.6. Suppose $t_1 = T_2^0$ whereas $t_2 \neq T_2^0$. Let $u_{1,\epsilon,T} = t_1/T - \epsilon_{1,T}$, $\epsilon_{1,T} > 0$. We proceed as in (S.8) by taking a second order Taylor's expansion of $c\left(zb_{2,T}/T,\, w\right)$ around $u_{1,\epsilon,T}$ and then use the left-Lipschitz continuity at t_1/T. Repeat this argument for $c\left(zb_{2,T} - \tau_2/T,\, w + \tau_2\right)$. For $c\left(zb_{2,T} - \tau_2/T,\, w - \tau_2 + \tau_1\right)$ and $c\left(zb_{2,T}/T,\, w - \tau_1\right)$, take a Taylor's expansion around t_2/T. Finally, use Lemma S.A.1 backward to obtain

$$c\left(t_1/T,\, w\right) c\left(t_2,\, w - \tau_2 + \tau_1\right) = \Gamma_{t_1/T}\left(-w\right) \Gamma_{t_2/T}\left(-w + \tau_2 - \tau_1\right) + O\left(T^{-1}\right).$$

Thus, with $u = t_1/T$ and $v = t_2/T$, the limit of the first term of (S.22) is equal to

$$\int_0^1 K_2^2\left(x\right) dx \left\{ \sum_{w=-\infty}^{\infty} \left[\Gamma_u\left(w\right) \Gamma_v\left(-w + \tau_2 - \tau_1\right) + \Gamma_u\left(w + \tau_2\right) \Gamma_v\left(-w + \tau_1\right)\right] \right\}. \tag{S.23}$$

For the sub-case where only t_2 is a discontinuity point, use a Taylor's expansion of $c\left(zb_{2,T}/T,\, w\right)$ and $c\left(zb_{2,T} - \tau_2/T,\, w + \tau_2\right)$ around t_1/T, and proceed as in (S.8) by taking a second order Taylor's expansion of $c\left(zb_{2,T} - \tau_2/T,\, w - \tau_2 + \tau_1\right)$ and $c\left(zb_{2,T}/T,\, w - \tau_1\right)$ around $u_{2,\epsilon,T} = t_2/T - \epsilon_{2,T}$, $\epsilon_{2,T} > 0$ and then use the left-Lipschitz continuity at t_2/T. Again using Lemma S.A.1 backward leads to (S.23). For the final case where $t_1 = t_2 = T_2^0$ we need to proceed as in the previous two sub-cases with $t_1 = T_2^0$ and $t_2 = T_2^0$ being discontinuity points. This would lead to (S.23). We can use (S.23) to obtain,

$$\begin{aligned} &Tb_{2,T} \mathrm{Cov}\left[\widetilde{\Gamma}\left(\tau_1\right), \widetilde{\Gamma}\left(\tau_2\right)\right] \\ &\quad \to \int_0^1 K_2^2\left(x\right) dx \int_0^1 \int_0^1 \left\{ \sum_{h=-\infty}^{\infty} \left[\Gamma_u\left(h\right) \Gamma_u\left(h - \tau_2 + \tau_1\right) + \Gamma_v\left(-h - \tau_2\right) \Gamma_v\left(-h - \tau_1\right)\right] \right\} dv du. \end{aligned}$$

In (S.22) the term involving $\kappa_{V,s}\left(-\tau_1,\, l - s,\, l - s - \tau_2\right)$ is negligible as in Lemma S.A.6 while the term involving $\Gamma_{(l-\tau_2)/T}\left(-w - j\right) \Gamma_{l/T}\left(-w + k\right)$ vanishes in the limit using the same argument as in the proof of Lemma S.A.6. This proves the result of part (i).

We move to part (ii). Let

$$J_{c,T} = \int_0^1 c\left(u,\, 0\right) du + 2 \sum_{k=1}^{T-1} \int_0^1 c\left(u,\, k\right) du,$$

and $\mathcal{T}_C \triangleq \left\{\left\{0,\, n_T, \ldots,\, T - n_T,\, T\right\} / \mathcal{T}\right\}$. We begin with the following relationship,

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

Using Lemma S.A.4, we have for any $-T+1 \leq k \leq T-1$,

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

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. Thus, $b_{1,T}^{-q}\mathbb{E}(\widetilde{J}_T-J_{c,T})$ has the same form as in the locally stationary case. The relation $J_{c,T}-J_T=O(T^{-1})$ continues to hold for SLS processes in virtue of Lemma S.A.1. Hence, $\lim_{T\to\infty}b_{1,T}^{-q}\mathbb{E}(\widetilde{J}_T-J_T)=-2\pi K_{1,q}\int_0^1 f^{(q)}\left(u,\,0\right)du$. Part (iii) follows from part (i)-(ii). $\square$

Proof of Theorem 3.1. We can now complete the proof of Theorem 3.1. We begin with part (i). We provide the expression for the asymptotic covariance between the $(a,\,l)$ and $(m,\,n)$ elements of $\widetilde{J}_T$:

$$
\begin{aligned}
&Tb_{1,T}b_{2,T}\mathrm{Cov}\left[\sum_{k=-T+1}^{T-1}K_1\left(b_{1,T}k\right)\widetilde{\Gamma}^{(a,l)}\left(k\right),\,\sum_{j=-T+1}^{T-1}K_1\left(b_{1,T}j\right)\widetilde{\Gamma}^{(m,n)}\left(j\right)\right]\\
&=b_{1,T}\sum_{k=-T+1}^{T-1}\sum_{j=-T+1}^{T-1}K_1\left(b_{1,T}k\right)K_1\left(b_{1,T}j\right)\left(\frac{n_T}{T}\right)^2\sum_{r=0}^{T/n_T}\sum_{b=0}^{T/n_T}\frac{1}{Tb_{2,T}}\sum_{s=k+1}^{T}\sum_{h=j+1}^{T} \qquad \text{(S.24)}\\
&\quad\times K_2^*\left(\frac{\left(\left(rn_T+1\right)-\left(s-k/2\right)\right)/T}{b_{2,T}}\right)K_2^*\left(\frac{\left(\left(bn_T+1\right)-\left(h-j/2\right)\right)/T}{b_{2,T}}\right)\\
&\quad\times\left\{\kappa_{V,s}^{(a,l,m,n)}\left(-k,\,h-s,\,h-s-j\right)\right.\\
&\quad\left.+\left[\Gamma_{h/T}^{(a,m)}\left(h-s\right)\Gamma_{(h-j)/T}^{(l,n)}\left(h-s-j+k\right)+\Gamma_{(h-j)/T}^{(a,n)}\left(h-s-j\right)\Gamma_{h/T}^{(l,m)}\left(h-s+k\right)\right]\right\}.
\end{aligned}
$$

As for the scalar case, the term involving $\kappa_{V,s}^{(a,l,m,n)}\left(-k,\,h-s,\,h-s-j\right)$ is negligible. The limit of the term involving $\Gamma_{h/T}^{(a,m)}\left(h-s\right)\Gamma_{(h-j)/T}^{(l,n)}\left(h-s-j+k\right)$ is, according to the derivations for the proof of part

(i) of Lemma S.A.8,

$$4\pi^2 \int K_1 (y)^2 \, dy \int_0^1 K_2 (x)^2 \, dx \left(\int_0^1 f^{(a,m)} (u,\, 0) \, du \right) \left(\int_0^1 f^{(l,n)} (v,\, 0) \, dv \right). \tag{S.25}$$

Similarly, the limit of the term involving $\Gamma^{(a,n)}_{(h-j)/T} (s-h-j) \, \Gamma^{(l,m)}_{h/T} (s-h+k)$ is the same as (S.25) but with m and n interchanged. The commutation-tensor product formula follows from the same argument as in Lemma S.A.7. The proof of part (ii) of the theorem follows from that of the scalar case with minor changes. Since part (iii) simply uses part (i)-(ii), it follows that

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{MSE} \left(Tb_{1,T} b_{2,T},\, \widetilde{J}_T,\, W \right) \\
&\quad = \lim_{T\to\infty} \gamma b_{1,T}^{-2q} \mathbb{E} \left(\widetilde{J}_T - J_T \right)' W \mathbb{E} \left(\widetilde{J}_T - J_T \right) + \lim_{T\to\infty} Tb_{1,T} b_{2,T} \, \mathrm{tr} W \, \mathrm{Var} \left(\mathrm{vec} \left(\widetilde{J}_T \right) \right),
\end{aligned}$$

converges to the desired limit. □

S.A.2.3 Proof of Theorem 3.2

Under Assumption 3.2, $|| \int_0^1 f^{(0)} (u,\, 0) \, || < \infty$. In view of $K_{1,0} = 0$, Theorem 3.1-(i,ii) [with $q = 0$ in part (ii)] implies $\widetilde{J}_T - J_T = o_{\mathbb{P}} (1)$. Noting that $\widehat{J}_T - \widetilde{J}_T = o_{\mathbb{P}} (1)$ if and only if $b' \widehat{J}_T b - b' \widetilde{J}_T b = o_{\mathbb{P}} (1)$ for arbitrary $b \in \mathbb{R}^p$ we shall provide the proof only for the scalar case. We first show that $\sqrt{T} b_{1,T} (\widehat{J}_T - \widetilde{J}_T) = O_{\mathbb{P}} (1)$ under Assumption 3.3. Let $\widetilde{J}_T (\beta)$ denote the estimator that uses $\{V_{t,T} (\beta)\}$. A mean-value expansion of $\widetilde{J}_T(\widehat{\beta}) (= \widehat{J}_T)$ about β_0 yields

$$\begin{aligned}
\sqrt{T} b_{1,T} (\widehat{J}_T - \widetilde{J}_T) &= b_{1,T} \frac{\partial}{\partial \beta'} \widetilde{J}_T (\bar{\beta}) \sqrt{T} (\widehat{\beta} - \beta_0) \\
&= b_{1,T} \sum_{k=-T+1}^{T-1} K_1 (b_{1,T} k) \, \frac{\partial}{\partial \beta'} \widehat{\Gamma} (k) \,|_{\beta=\bar{\beta}} \sqrt{T} (\widehat{\beta} - \beta_0),
\end{aligned} \tag{S.26}$$

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

$$\begin{aligned}
&\left\| \frac{\partial}{\partial \beta'} \widehat{c} \, (rn_T/T,\, k) \right\| |_{\beta=\bar{\beta}} \\
&= \left\| (Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2^* \left(\frac{(r+1) \, n_T - (s - k/2)}{Tb_{2,T}} \right) \right. \\
&\quad \left. \times \left(V_s (\beta) \, \frac{\partial}{\partial \beta'} V_{s-k} (\beta) + \frac{\partial}{\partial \beta'} V_s (\beta) \, V_{s-k} (\beta) \right) \right\| |_{\beta=\bar{\beta}} \\
&\leq 2 \left((Tb_{2,T})^{-1} \sum_{s=1}^{T} K_2^* \left(\frac{(r+1) \, n_T - (s - k/2)}{Tb_{2,T}} \right)^2 \sup_{s\geq 1} \sup_{\beta\in\Theta} (V_s (\beta))^2 \right)^{1/2}
\end{aligned} \tag{S.27}$$

$$
\times \left((Tb_{2,T})^{-1} \sum_{s=1}^{T} K_2^* \left(\frac{(r+1)\, n_T - (s-k/2)}{Tb_{2,T}} \right)^2 \sup_{s \geq 1} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta'} V_s(\beta) \right\|^2 \right)^{1/2}
$$
$$
= O_{\mathbb{P}}(1),
$$

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

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

Then, (S.26) becomes

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

where the last equality uses $b_{1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{1,T} k)| \rightarrow \int |K_1(x)|\, dx < \infty$. This concludes the proof of part (i) of Theorem 3.2 because $\sqrt{T} b_{1,T} \rightarrow \infty$ by assumption.

The next step is to show that $\sqrt{T b_{1,T}} (\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$ under the assumptions of Theorem 3.2-(ii). A second-order Taylor's expansion gives

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

Proceeding as in (S.27) but now using Assumption 3.4-(ii),

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

and thus,

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

since $Tb_{1,T} \to \infty$. Next, we want to show that $G_T = o_{\mathbb{P}}(1)$. We apply the results of Theorem 3.1-(i,ii) to $\widetilde{J}_T$ where the latter is constructed using $(V_t', \partial V_t/\partial\beta' - \mathbb{E}(\partial V_t/\partial\beta'))'$ rather than just with V_t. The first row and column of the off-diagonal elements of $\widetilde{J}_T$ are now (written as column vectors)

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

By Theorem 3.1-(i,ii), each expression above is $O_{\mathbb{P}}(1)$. Since

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

where

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

it remains to show that A_3 is $o_{\mathbb{P}}(1)$. Note that

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

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

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

This implies $G_T = o_{\mathbb{P}}(1)$. It follows that $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$ which concludes the proof of part (ii) because $\sqrt{Tb_{1,T}b_{2,T}}(\widetilde{J}_T - J_T) = O_{\mathbb{P}}(1)$ by Theorem 3.1-(iii).

Finally, we need to consider part (iii). Let

$$\xi_T \triangleq Tb_{1,T}\left(\mathrm{vec}\left(\widehat{J}_T - J_T\right)' W \mathrm{vec}\left(\widehat{J}_T - J_T\right) - \mathrm{vec}\left(\widetilde{J}_T - J_T\right)' W \mathrm{vec}\left(\widetilde{J}_T, -J_T\right)\right).$$

By part (ii), we know that $\sqrt{Tb_{1,T}}(\widehat{J}_T - J_T) = O_{\mathbb{P}}(1)$ and $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathbb{P}}(1)$. This implies

$$Tb_{1,T}\left(\mathrm{vec}\left(\widehat{J}_T - J_T\right)' W_T \mathrm{vec}\left(\widehat{J}_T - J_T\right) - \mathrm{vec}\left(\widetilde{J}_T - J_T\right)' W_T \mathrm{vec}\left(\widetilde{J}_T, -J_T\right)\right) \xrightarrow{\mathbb{P}} 0.$$

Then, using Assumption 3.5, $\xi_T = o_{\mathbb{P}}(1)$ and since $|\xi_T|$ is bounded we have $\mathbb{E}\left(\xi_T\right) \to 0$ by Lemma A1 in Andrews (1991). □

References

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

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

Fuller, W. A. (1995): *Introduction to Statistical Time Series*. New York: Wiley.

Supplemental Material not for Publication to

Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models

Alessandro Casini

Department of Economics and Finance
University of Rome Tor Vergata

26th July 2021

Abstract

This supplemental material is not for publication and is structured as follows. Section N.A reviews how to apply the proposed DK-HAC estimator in GMM and IV contexts. Section N.B contains the proofs of the results of Section 2 and 4-5.

N.A Implementation of DK-HAC in GMM and IV Models

Section N.A.1 reviews the DK-HAC estimation in GMM models while Section N.A.2 considers IV models.

N.A.1 GMM

We begin with the GMM setup [cf. Hansen (1982)]. For a k-vector β_* of unknown parameters, we have the moment condition $\mathbb{E}m_t(\beta_*)=0$ where $m_t(\beta)$ is a p-vector of functions of the data and parameters where $p\geq k$. The GMM estimator $\widehat{\beta}$ is defined as the solution to $\min_\beta m_T(\beta)'\widehat{W}_{2,T}m_T(\beta)$, where $m_T(\beta)=T^{-1}\sum_{t=1}^T m_t(\beta)$ is the sample average of the vector of sample moments $m_t(\beta)$ and $\widehat{W}_{2,T}$ is a (possibly) random, symmetric weighting matrix. The asymptotic covariance matrix of $\widehat{\beta}$ is given by $H=\lim_{T\to\infty}H_T$ where

$$H_T=\left(L_T'W_{2,T}L_T\right)^{-1}L_T'W_{2,T}J_TW_{2,T}L_T\left(L_T'W_{2,T}L_T\right)^{-1},$$

where $L_T=T^{-1}\sum_{t=1}^T\mathbb{E}m_{t\beta}(\beta_*)$ and $m_{t\beta}(\beta)$ is the $p\times k$ matrix of partial derivatives of $m_t(\beta)$, $W_{2,T}$ is a nonrandom matrix such that $\widehat{W}_{2,T}-W_{2,T}\xrightarrow{\mathbb{P}}0$, and $J_T=T^{-1}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}(m_t(\beta_*)\,m_s(\beta_*))'$. Let $J=\lim_{T\to\infty}J_T$. The consistent estimation of H boils down to the consistent estimation of J since the estimation of L_T and $W_{2,T}$ is straightforward. $\widehat{W}_{2,T}$ is a natural estimator of $W_{2,T}$ while under regularity conditions $L_T-T^{-1}\sum_{t=1}^T m_{t\beta}(\widehat{\beta})\xrightarrow{\mathbb{P}}0$. In place of the classical HAC estimators we now estimate J by

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

where

$$\widehat{c}_T(rn_T/T,\,k)\triangleq\begin{cases}(Tb_{2,T})^{-1}\sum_{s=k+1}^T K_2^*\left(\frac{((r+1)n_T-(s+k/2))/T}{b_{2,T}}\right)\widehat{m}_s\widehat{m}_{s-k}', & k\geq0\\ (Tb_{2,T})^{-1}\sum_{s=-k+1}^T K_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{2,T}}\right)\widehat{m}_{s+k}\widehat{m}_s', & k<0\end{cases},$$

and $\widehat{m}_s=m_s(\widehat{\beta})$. We can implement $\widehat{J}_T$ with the data-dependent methods for selecting $b_{1,T}$ and $b_{2,T}$, and choose K_1 and K_2 on the basis of the optimality results of Section 4. For K_1 one can use the QS kernel while for K_2 one can choose $K_2=6x(1-x)$ for $0\leq x\leq1$ and 0 otherwise as suggested in Section 4. From the results in Section 5,

$$\widehat{b}_{1,T}=0.6828\left(\widehat{\phi}(2)\,T\widehat{\overline{b}}_{2,T}\right)^{-1/5}$$
$$\widehat{b}_{2,T}(u_r)=1.6786\left(\widehat{D}_1(u_r)\right)^{-1/5}\left(\widehat{D}_2(u_r)\right)^{1/5}T^{-1/5},\qquad u_r=rn_T/T,$$

where the expressions for $\widehat{\phi}(2)$, $\widehat{D}_1(u_r)$ and $\widehat{D}_2(u_r)$ are given in the same section.

N.A.2 IV

Consider the linear model $y_t=x_t'\beta_0+e_t$ $(t=1,\dots,T)$, where $\beta_0\in\Theta\subset\mathbb{R}^p$, y_t is an observation on the dependent variable, x_t is a p-vector of regressors and e_t is an unobserved disturbance potentially autocorrelated. Suppose the regressor is endogenous: $\mathbb{E}(x_t e_t) \neq 0$. The IV estimator $\widehat{\beta}_{\mathrm{IV}}$ is given by $\widehat{\beta}_{\mathrm{IV}} = (Z'X)^{-1} Z'Y$, where $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$ and $Z = (z_1, \ldots, z_T)'$ where z_t is a p-vector of instruments. The asymptotic variance of the IV estimator is given by the limit of $\mathrm{Var}(\sqrt{T}(\widehat{\beta}_{\mathrm{IV}} - \beta_0)) = Q_{ZX}^{-1} J_T Q_{ZX}^{-1}$ where $Q_{ZX} = T^{-1}\sum_{t=1}^{T} z_t x_t'$ and $J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathbb{E}(e_s z_s (e_t z_t)')$. A natural estimator of $\lim_{T\to\infty} Q_{ZX}$ is $T^{-1}\sum_{t=1}^{T} z_t x_t'$. Let $J = \lim_{T\to\infty} J_T$. J can be consistently estimated by $\widehat{J}_T$ as given in (N.1) where $\widehat{m}_t$ is replaced by $\widehat{e}_t z_t$ where $\widehat{e}_t = y_t - x_t'\widehat{\beta}_{\mathrm{IV}}$.

N.B Appendix: Proofs of the Results of Section 2 and 4-5

N.B.1 Proofs of the Results of Section 2.1

N.B.1.1 Proof of Theorem 2.1

For $Tu \notin \mathcal{T}$ we use the arguments in the proof of Theorem 2.2 in Dahlhaus (1997). Without loss of generality, assume $T_{j-1}^0 < Tu < T_j^0$ for some $1 \leq j \leq m_0 + 1$. Then,

$$f_{j,T}(u, \omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s) A_{j,\lfloor Tu - s/2 \rfloor, T}^0 (\eta) \overline{A_{j,\lfloor Tu + s/2 \rfloor, T}^0 (\eta)} d\eta,$$

and

$$f_j(u, \omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s) A_j(u, \eta) \overline{A_j(u, \eta)} d\mu.$$

We have, in virtue of standard orthogonality relations,

$$\begin{aligned}
&\int_{-\pi}^{\pi} |f_{j,T}(u, \omega) - f_j(u, \omega)|^2 d\omega \\
&\quad = \int_{-\pi}^{\pi} \Bigg| \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \\
&\qquad \times \left[\int_{-\pi}^{\pi} \exp(i\eta s) \left(A_{j,\lfloor Tu - s/2 \rfloor, T}^0 (\eta) \overline{A_{j,\lfloor Tu + s/2 \rfloor, T}^0 (\eta)} - A_j(u, \eta) \overline{A_j(u, \eta)} \right) d\eta \right] \Bigg|^2 d\omega \\
&\quad = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |c_{s,j}|^2 + o(1),
\end{aligned}$$

where $c_{s,j} = \int_{-\pi}^{\pi} \exp(i\eta s) G_j(s/2T, \eta) d\eta$ and

$$G_j\left(\frac{s}{2T}, \eta\right) = A_j\left(u - \frac{s}{2T}, \eta\right) A_j\left(u + \frac{s}{2T}, -\eta\right) - A_j(u, \eta) A_j(u, -\eta).$$

By well-known results on Fourier coefficients [cf. Bary (1964), Chapter 2.3], $|c_{s,j}| \leq Cs^{-\vartheta}$ and thus $\sum_{s=n}^{\infty} |c_{s,j}|^2 = O(n^{1-2\vartheta})$. Let $\Delta_s(\omega) = \sum_{r=0}^{s-1} \exp(-i\omega r)$. Applying summation by parts yields

$$\begin{aligned}\sum_{s=0}^{n-1} |c_{s,j}|^2 &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{s=0}^{n-1} \exp(-i(\omega-\eta)s)\, G_j\left(\frac{s}{2T},\,\omega\right)\overline{G_j\left(\frac{s}{2T},\,\eta\right)}d\omega d\eta \\ &\leq \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Bigg|-\sum_{s=0}^{n-1}[G_j\left(\frac{s}{2T},\,\omega\right)\overline{G_j\left(\frac{s}{2T},\,\eta\right)} - G_j\left(\frac{s-1}{2T},\,\omega\right)\overline{G_j\left(\frac{s-1}{2T},\,\eta\right)}]\Delta_s(\eta-\omega) \\ &\quad + G_j\left(\frac{n-1}{2T},\,\omega\right)\overline{G_j\left(\frac{n-1}{2T},\,\eta\right)}\Delta_n(\eta-\omega)\Bigg| d\omega d\eta \\ &= O\left(\frac{n\ln n}{T^{\vartheta}}\right).\end{aligned}$$

A similar bound holds for $\sum_{s=n}^{\infty} |c_{-s,j}|^2$. The result for $Tu \notin \mathcal{T}$ follows by choosing n appropriately. Next, suppose $Tu \in \mathcal{T}$ and $u = T_j^0/T$. Then, we have

$$f_{j,T}(u,\,\omega) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-i\omega s)\int_{-\pi}^{\pi} \exp(i\eta s)\, A^0_{j,\lfloor Tu-3|s|/2\rfloor,T}(\eta)\,\overline{A^0_{j,\lfloor Tu-|s|/2\rfloor,T}(\eta)}d\eta$$

and

$$f_j(u,\,\omega) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-i\omega s)\int_{-\pi}^{\pi} \exp(i\eta s)\, A_j(u,\,\eta)\,\overline{A_j(u,\,\eta)}d\eta.$$

Proceeding as above,

$$\begin{aligned}&\int_{-\pi}^{\pi} |f_T(u,\,\omega) - f(u,\,\omega)|^2 d\omega \\ &\quad = \int_{-\pi}^{\pi}\Bigg|\frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-i\omega s) \\ &\qquad \left[\int_{-\pi}^{\pi} \exp(i\eta s)\, A^0_{j,\lfloor uT-3|s|/2\rfloor,T}(\eta)\,\overline{A^0_{j,\lfloor uT-|s|/2\rfloor,T}(\eta)}d\eta - \int_{-\pi}^{\pi} \exp(i\eta s)\, A_j(u,\,\eta)\,\overline{A_j(u,\,\eta)}d\eta\right]\Bigg|^2 d\omega \\ &\quad = \int_{-\pi}^{\pi}\Bigg|\frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-i\omega s) \\ &\qquad \left[\int_{-\pi}^{\pi} \exp(i\eta s)\left(A^0_{j,\lfloor Tu-3|s|/2\rfloor,T}(\eta)\,\overline{A^0_{j,\lfloor Tu-|s|/2\rfloor,T}(\eta)} - A_j(u,\,\eta)\,\overline{A_j(u,\,\eta)}\right)d\eta\right]\Bigg|^2 d\omega \\ &\quad = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} |c_{s,j}|^2 + o(1),\end{aligned}$$

with $c_{s,j}=\int_{-\pi}^{\pi}\exp(i\eta s)\,G_j(s/2T,\,\eta)\,d\eta$ and

$$G_j\left(\frac{s}{2T},\,\eta\right)=A_j\left(u-\frac{3\,|s|}{2T},\,\eta\right)A_j\left(u-\frac{|s|}{2T},\,-\eta\right)-A_j(u,\,\eta)\,A_j(u,\,-\eta)\,.$$

Using the definition of $\Delta_s(\omega)$ and the above-mentioned properties of $c_{s,j}$ which continue to hold, summation by parts and the Lipschitz continuity of $A_j(u,\,\cdot)$ then imply $\sum_{s=0}^{n-1}|c_{s,j}|^2=O(n\ln n/T^{\vartheta})$. Since the same bound applies to $\sum_{s=n}^{\infty}|c_{-s,j}|^2$, we can choose an appropriate n to yield the result for $Tu\in\mathcal{T}$. □

N.B.2 Proofs of the Results of Section 4

N.B.2.1 Proof of Proposition 4.1

We first need to show that $\sqrt{Tb_{2,T}}\left(\widehat{c}_T(rn_T/T,\,k)-\widetilde{c}(rn_T/T,\,k)\right)=o_{\mathbb{P}}(1)$. Without loss of generality, we can focus on the scalar case. From (S.27), $\left\|\frac{\partial}{\partial\beta'}\widehat{c}_T(rn_T/T,\,k)\right\||_{\beta=\bar{\beta}}=O_{\mathbb{P}}(1)$. A mean-value Taylor's expansion gives

$$\begin{aligned}\sqrt{Tb_{2,T}}\left(\widehat{c}_T(rn_T/T,\,k)-\widetilde{c}_T(rn_T/T,\,k)\right)&=\sqrt{b_{2,T}}\frac{\partial}{\partial\beta'}\widehat{c}_T(rn_T/T,\,k)\,|_{\beta=\bar{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\\&\leq\sqrt{b_{2,T}}\sup_{r\geq1}\left\|\frac{\partial}{\partial\beta'}\widehat{c}(rn_T/T,\,k)\right\||_{\beta=\bar{\beta}}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\\&=\sqrt{b_{2,T}}O_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)\,.\end{aligned}$$

Thus,

$$\xi_T=\mathrm{vec}\left(\widehat{c}_T(rn_T/T,\,k)-\widetilde{c}(rn_T/T,\,k)\right)'\,\widetilde{W}_T\mathrm{vec}\left(\widehat{c}_T(rn_T/T,\,k)-\widetilde{c}(rn_T/T,\,k)\right)\xrightarrow{\mathbb{P}}0.$$

Since ξ_T is a bounded sequence, $\mathbb{E}(\xi_T)\xrightarrow{\mathbb{P}}0$. Hence, given that $\widetilde{W}_T\xrightarrow{\mathbb{P}}\widetilde{W}$, we have MSE$(1,\,\widehat{c}_T(u_0,\,k)\,,\,\widetilde{W}_T)=$ MSE$(1,\,\widetilde{c}_T(u_0,\,k)\,,\,\widetilde{W})+o_{\mathbb{P}}(1)$. By using the results of Lemma S.A.4, the MSE of $\widehat{c}_T(u_0,\,k)$ for any $u_0\in(0,\,1)$ and any integer k, is given by

$$\begin{aligned}&\mathbb{E}\left[\widehat{c}_T(u_0,\,k)-c(u_0,\,k)\right]^2\\&\quad=\frac{1}{4}b_{2,T}^4\left(\int_0^1x^2K_2(x)\,dx\right)^2\left(\frac{\partial^2}{\partial^2u}c(u_0,\,k)\right)^2\\&\qquad+\frac{1}{Tb_{2,T}}\int_0^1K_2^2(x)\,dx\sum_{l=-\infty}^{\infty}c(u_0,\,l)\left[c(u_0,\,l)+c(u_0,\,l+2k)\right]\\&\qquad+\frac{1}{Tb_{2,T}}\int_0^1K_2^2(x)\,dx\sum_{h_1=-\infty}^{\infty}\kappa_{V,\lfloor Tu_0\rfloor}(-k,\,h_1,\,h_1-k)+o\left(b_{2,T}^4\right)+o\left(1/\left(b_{2,T}T\right)\right)\\&\quad\triangleq g\left(K_2,\,b_{2,T}\right)+o\left(b_{2,T}^4\right)+o\left(1/\left(b_{2,T}T\right)\right).\end{aligned}\tag{N.1}$$

Then $g\left(K_{2},\, b_{2,T}\right) = 4^{-1} b_{2,T}^{4} H\left(K_{2}\right) D_{1}\left(u_{0}\right) + \left(T b_{2,T}\right)^{-1} F\left(K_{2}\right)\left(D_{2}\left(u_{0}\right)+D_{3}\left(u_{0}\right)\right)$. The minimum of $g\left(K_{2},\, b_{2,T}\right)$ in $b_{2,T}$ is determined by the equation

$$\frac{\partial}{\partial b_{2,T}} g\left(K_{2},\, b_{2,T}\right) = b_{2,T}^{3} H\left(K_{2}\right) D_{1}\left(u_{0}\right) - \frac{1}{T b_{2,T}^{2}} F\left(K_{2}\right)\left(D_{2}\left(u_{0}\right)+D_{3}\left(u_{0}\right)\right) = 0.$$

The minimum is achieved at

$$b_{2,T}^{\mathrm{opt}} = \left[H\left(K_{2}\right) D_{1}\left(u_{0}\right)\right]^{-1/5} \left(F\left(K_{2}\right)\left(D_{2}\left(u_{0}\right)+D_{3}\left(u_{0}\right)\right)\right)^{1/5} T^{-1/5}.$$

If $V_{t,T}$ is Gaussian, then the term involving $\kappa_{V,\lfloor T u_{0}\rfloor}$ in (N.1) is equal to zero and so $D_{3}\left(u_{0}\right)=0$ in $b_{2,T}^{\mathrm{opt}}$. Next, we minimize $g(K_{2},\, b_{2,T}^{\mathrm{opt}})$ with respect to the class of kernels $K_{2}:\mathbb{R}\rightarrow[0,\,\infty]$ that are centered at $x=1/2$ with

$$\int_{\mathbb{R}} K_{2}\left(x\right) dx = 1, \tag{N.2}$$

$$K_{2}\left(x\right) = K_{2}\left(1-x\right). \tag{N.3}$$

We use arguments similar to those in Chapter 7 of Priestley (1981) and in Dahlhaus and Giraitis (1998). Let

$$\sqrt{K_{2\sigma}\left(x\right)} = \frac{1}{\sqrt{\sigma}}\left(K_{2}\left(\frac{x-1/2}{\sigma}+\frac{1}{2}\right)\right)^{1/2}, \qquad \text{where}\, \sigma\in(0,\,\infty)\,.$$

We have $F\left(K_{2\sigma}\right)=(1/\sigma)\, F\left(K_{2}\right)$ and $H\left(K_{2\sigma}\right)=\sigma^{4} H\left(K_{2}\right)$ (with the integrals in the definition of F and H extended to $\mathbb{R}$ and with the variable of integration x subtracted by 1/2). Then, $b_{2,K_{2\sigma},T}^{\mathrm{opt}}=\sigma^{-1} b_{2,T}^{\mathrm{opt}}$ where $b_{2,K_{2\sigma},T}^{\mathrm{opt}}$ is the optimal bandwidth associated with the kernel $K_{2\sigma}$. Also, $g(K_{2\sigma},\, b_{2,K_{2\sigma},T}^{\mathrm{opt}}) = g(K_{2},\, b_{2,T}^{\mathrm{opt}})$. We can thus restrict our attention to K_{2} satisfying

$$\int_{\mathbb{R}}\left(x-\frac{1}{2}\right)^{2} K_{2}\left(x\right) dx = \int_{\mathbb{R}}\left(x-\frac{1}{2}\right)^{2} K_{2}^{\mathrm{opt}}\left(x\right) dx, \tag{N.4}$$

where $K_{2}^{\mathrm{opt}}\left(x\right)=6x\left(1-x\right)$ for $x\in[0,\,1]$ and $K_{2}^{\mathrm{opt}}\left(x\right)=0$ for $x\notin[0,\,1]$. Therefore, we have to show that, for any K_{2} that satisfies (N.2)-(N.3),

$$\int_{\mathbb{R}/[0,\,1]} K_{2}^{2}\left(x\right) dx + \int_{0}^{1} K_{2}^{2}\left(x\right) dx = \int_{\mathbb{R}} K_{2}^{2}\left(x\right) dx \geq \int_{\mathbb{R}}\left(K_{2}^{\mathrm{opt}}\left(x\right)\right)^{2} dx = \int_{0}^{1}\left(K_{2}^{\mathrm{opt}}\left(x\right)\right)^{2} dx.$$

This is implied by

$$\int_{0}^{1} K_{2}^{2}\left(x\right) dx \geq \int_{0}^{1}\left(K_{2}^{\mathrm{opt}}\left(x\right)\right)^{2} dx.$$

Let $K_2(x) = K_2^{\mathrm{opt}}(x) + \varepsilon(x)$, $x \in \mathbb{R}$, where $\varepsilon > 0$. Since $\int_{\mathbb{R}} \varepsilon^2(x)\, dx \geq 0$ and K_2^{opt} vanishes outside $[0,\, 1]$, it is sufficient to prove that $\int_0^1 \left(K_2^{\mathrm{opt}}(x)\, \varepsilon(x)\right) dx \geq 0$ because

$$\int_0^1 K_2^2(x)\, dx = \int_0^1 \left(K_2^{\mathrm{opt}}(x) + \varepsilon(x)\right)^2 dx \geq \int_0^1 \left(K_2^{\mathrm{opt}}(x)\right)^2 + 2\int_0^1 \left(K_2^{\mathrm{opt}}(x)\, \varepsilon(x)\right) dx.$$

By (N.2), we have $\int_{\mathbb{R}} \varepsilon(x)\, dx = 0$, while $\int_{\mathbb{R}} \varepsilon(x)\,(x^2 - x)\, dx = 0$ in view of

$$\begin{aligned} 0 = \int_{\mathbb{R}} \left(K_2(x) - K_2^{\mathrm{opt}}(x)\right)\left(x - \frac{1}{2}\right)^2 dx &= \int_{\mathbb{R}} \left(K_2(x) - K_2^{\mathrm{opt}}(x)\right)\left(x^2 - x\right) dx + \frac{1}{4}\int_{\mathbb{R}} \varepsilon(x)\, dx \\ &= \int_{\mathbb{R}} \left(K_2(x) - K_2^{\mathrm{opt}}(x)\right)\left(x^2 - x\right) dx. \end{aligned}$$

Note that $(x^2 - x) = x(x-1)$. Therefore, we deduce

$$6\int_{\mathbb{R}/[0,\,1]} x(1-x)\,\varepsilon(x)\, dx + 6\int_0^1 x(1-x)\,\varepsilon(x)\, dx = 0.$$

Rearranging the last expression yields,

$$\int_0^1 K_2^{\mathrm{opt}}(x)\,\varepsilon(x)\, dx = 6\int_{\mathbb{R}/[0,\,1]} x(x-1)\,\varepsilon(x)\, dx \geq 0,$$

because $\varepsilon(x) \geq 0$ and $x(x-1) \geq 0$ for $x \notin [0,\, 1]$. □

N.B.2.2 Proof of Theorem 4.1

Without loss of generality, we provide the proof for the scalar case. By Theorem 3.2-(iii), if $Tb_{1,T}^{2q+1} b_{2,T} \to \gamma_2 \in (0,\, \infty)$ for some $q \in [0,\, \infty)$ for which $K_{1,q}$, $|\int_0^1 f^{(q)}(u,\, 0)\, du| \in [0,\, \infty)$, then

$$\begin{aligned} &\lim_{T\to\infty} \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T(b_{1,T,K_1}),\, 1\right) \\ &\quad = 4\pi^2 \left[\gamma_2 K_{1,q}^2 \left(\int_0^1 f^{(q)}(u,\, 0)\, du\right)^2 + \int K_1^2(y)\, dy \int_0^1 (K_2(x))^2\, dx \left(\int_0^1 f(u,\, 0)\, du\right)^2\right]. \end{aligned}$$

We have $Tb_{1,T}^5 b_{2,T} \to \gamma$ by assumption. Thus, we apply Theorem 3.2-(iii) with $q = 2$, K_1 and b_{1,T,K_1}. Then, $Tb_{1,T,K_1}^5 b_{2,T} \to \gamma / \left(\int K_1^2(x)\, dx\right)^5$ and

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

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

$$\lim_{T\to\infty}\left(\mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T(b_{1,T,K_1}),\, 1\right) - \mathrm{MSE}\left(Tb_{1,T}b_{2,T},\, \widehat{J}_T^{\mathrm{QS}}(b_{1,T}),\, 1\right)\right)$$

$$= 4\gamma\pi^2 \left(\int_0^1 f^{(q)}(u, 0)\, du \right)^2 \int_0^1 (K_2(x))^2\, dx \left[K_{1,2}^2 \left(\int K_1^2(y)\, dy \right)^4 - \left(K_{1,2}^{\mathrm{QS}} \right)^2 \right].$$

Let $\widetilde{K}_1(\cdot)$ and $\widetilde{K}_1^{\mathrm{QS}}(\cdot)$ denote the spectral window generators of $K_1(\cdot)$ and $K_1^{\mathrm{QS}}(\cdot)$, respectively. They have the following properties: $K_{1,2} = \int_{-\infty}^{\infty} \omega^2 \widetilde{K}_1(\omega)\, d\omega$, $K_1(0) = \int_{-\infty}^{\infty} \widetilde{K}_1(\omega)\, d\omega$, and $\int_{-\infty}^{\infty} K_1^2(x)\, dx = \int_{-\infty}^{\infty} \widetilde{K}_1^2(\omega)\, d\omega$. As in Andrews (1991), the result of the theorem follows if we can show the following inequality,

$$K_{1,2}^2 \left(\int K_1^2(x)\, dx \right)^4 \geq \left(K_{1,2}^{\mathrm{QS}} \right)^2 \qquad \text{for all } K_1(\cdot) \in \widetilde{\boldsymbol{K}}_1. \tag{N.5}$$

Priestley (1981, Ch. 7.5) showed that $\widetilde{K}_1^{\mathrm{QS}}(\cdot)$ minimizes

$$\int_{-\infty}^{\infty} \omega^2 \widetilde{K}_1(\omega)\, d\omega \left(\int_{-\infty}^{\infty} \widetilde{K}_1^2(\omega)\, d\omega \right)^2, \tag{N.6}$$

subject to (a) $\int_{-\infty}^{\infty} \widetilde{K}_1(\omega)\, d\omega = 1$, (b) $\widetilde{K}_1(\omega) \geq 0, \forall\, \omega \in \mathbb{R}$, and (c) $\widetilde{K}_1(\omega) = \widetilde{K}_1(-\omega), \forall\, \omega \in \mathbb{R}$, where $K_1^{\mathrm{QS}}(\omega) = (5/8\pi)(1 - \omega^2/c^2)$ for $|\omega| \leq c$ for $c = 6\pi/5$. and $K_1^{\mathrm{QS}}(\omega) = 0$ otherwise. Note that the inequality (N.5) holds if and only if $\widetilde{K}_1^{\mathrm{QS}}(\cdot)$ minimizes (N.6). This proves the inequality of the theorem. Strict inequality holds when $K_1^{\mathrm{QS}}(x) \neq K_1(x)$ with positive Lebesgue measure. □

N.B.2.3 Proof of Corollary 4.1

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

$$\begin{aligned} &\lim_{T\to\infty} \mathrm{MSE}\left(T^{\frac{2q}{2q+1}} b_{2,T}^{\frac{2q}{2q+1}},\, \widehat{J}_T(b_{1,T},\, b_{2,T}),\, W_T \right) \\ &\quad = \gamma^{-1/(2q+1)} 4\pi^2 \left[\gamma K_{1,q}^2 \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right)' W \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right) \right. \\ &\qquad \left. + \int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx\, \mathrm{tr} W \left(I_{p^2} - C_{pp} \right) \left(\int_0^1 f(u, 0)\, du \right) \otimes \left(\int_0^1 f(v, 0)\, dv \right) \right]. \end{aligned} \tag{N.7}$$

Minimizing this with respect to γ gives

$$\begin{aligned} &\gamma^{2q/(2q+1)} K_{1,q}^2 \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right)' W \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right) \\ &\quad = \gamma^{-1/(2q+1)} \int K_1^2(y)\, dy \int K_2^2(x)\, dx\, \mathrm{tr} W \left(I_{p^2} - C_{pp} \right) \left(\int_0^1 f(u, 0)\, du \right) \otimes \left(\int_0^1 f(v, 0)\, dv \right), \end{aligned}$$

or

$$\gamma^{\mathrm{opt}} = \frac{1}{2q} \frac{\int K_1^2(y)\, dy \int K_2^2(x)\, dx\, \mathrm{tr} W \left(I_{p^2} + C_{pp} \right) \left(\int_0^1 f(u, 0)\, du \right) \otimes \left(\int_0^1 f(v, 0)\, dv \right)}{K_{1,q}^2 \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right)' W \mathrm{vec} \left(\int_0^1 f^{(q)}(u, 0)\, du \right)}$$

$$= \left(2qK_{1,q}^{2}\phi\left(q\right)\right)^{-1}\left(\int K_{1}^{2}\left(y\right)dy\int_{0}^{1}K_{2}^{2}\left(x\right)dx\right).$$

Note that $\gamma^{\mathrm{opt}} > 0$ provided that $0 < \phi(q) < \infty$ and W is positive definite. Hence, $\{b_{1,T}\}$ is optimal in the sense that $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma^{\mathrm{opt}}$ if and only if $b_{1,T} = b_{1,T}^{\mathrm{opt}} + o((Tb_{2,T})^{-1/(2q+1)})$ where $b_{2,T} = O(b_{2,T}^{\mathrm{opt}})$. □

N.B.3 Proofs of the Results of Section 5

N.B.3.1 Proof of Theorem 5.1

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), $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})-J_T) = O_{\mathbb{P}}(1)$. It remains to establish the second result of Theorem 5.1-(ii). Let $S_T = \left\lfloor b_{\theta_1,T}^{-r}\right\rfloor$ where

$$\begin{aligned}r \in &(\max\{(8b-5-2q)/8(b-1),\, 1.25,\, (b/2-1/4)/(b-1),\, q/(l-1),\, (8b-7-6q)/8(b-1)\\ &(b-3/4-q/2)/(b-1)\},\, \min\{13q/24+49/48,\, 46/48+20q/48,\, 7/8+3q/4,\, (6+4q)/8,\, 2\}),\end{aligned}$$

with $b > 1 + 1/q$. We will use the following decomposition

$$\begin{aligned}\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T}) &= (\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, \widehat{\overline{b}}_{2,T}))\\ &\quad + (\widehat{J}_T(b_{\theta_1,T},\, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})).\end{aligned} \tag{N.8}$$

Let

$$\begin{aligned}N_1 &\triangleq \{-S_T,\, -S_T+1, \ldots,\, -1,\, 1, \ldots,\, S_T-1,\, S_T\}\,,\\ N_2 &\triangleq \{-T+1, \ldots,\, -S_T-1,\, S_T+1, \ldots,\, T-1\}\,.\end{aligned}$$

Let us consider the first term of (N.8). We have

$$\begin{aligned}T^{8q/10(2q+1)}&(\widehat{J}_T(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, \widehat{\overline{b}}_{2,T}))\\ &= T^{8q/10(2q+1)}\sum_{k\in N_1}(K_1(\widehat{b}_{1,T}k) - K_1(b_{\theta_1,T}k))\widehat{\Gamma}\left(k\right)\\ &\quad + T^{8q/10(2q+1)}\sum_{k\in N_2}K_1(\widehat{b}_{1,T}k)\widehat{\Gamma}\left(k\right)\\ &\quad - T^{8q/10(2q+1)}\sum_{k\in N_2}K_1(b_{\theta_1,T}k)\widehat{\Gamma}\left(k\right)\\ &\triangleq A_{1,T} + A_{2,T} - A_{3,T}.\end{aligned} \tag{N.9}$$

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

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

for some $C < \infty$. By Assumption 5.1-(ii) ($\widetilde{n}_T \left|\widehat{\phi}(q) - \phi_{\theta^*}\right| = O_{\mathbb{P}}(1)$) and, using the delta method, it suffices to show that $B_{1,T} + B_{2,T} + B_{3,T} \xrightarrow{\mathbb{P}} 0$, where

$$\begin{aligned}B_{1,T} &= \widehat{\bar{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right|, \\ B_{2,T} &= \widehat{\bar{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|, \\ B_{3,T} &= \widehat{\bar{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_T^{-1} \sum_{k=1}^{S_T} k \left|\Gamma_T(k)\right|,\end{aligned} \tag{N.11}$$

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

$$\begin{aligned}B_{1,T} &\leq \widehat{\bar{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_T^{-1} T^{-1/2} \sum_{k=1}^{S_T} k \left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}\right)\sqrt{T}\left(\widehat{\beta} - \beta_0\right)\right| \\ &\leq C\widehat{\bar{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)-1/2} \left(Tb_{\theta_2,T}\right)^{2r/(2q+1)} \widetilde{n}_T^{-1} \sup_{k\geq 1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}\right\| \sqrt{T}\left\|\widehat{\beta} - \beta_0\right\| \\ &\leq C\widehat{\bar{b}}_{2,T}^{(-1+2r)/(2q+1)} T^{(8q-10)/10(2q+1)-1/2+2r/(2q+1)} \widetilde{n}_T^{-1} \sup_{k\geq 1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}\right\| \sqrt{T}\left\|\widehat{\beta} - \beta_0\right\| \xrightarrow{\mathbb{P}} 0,\end{aligned} \tag{N.12}$$

since $\widetilde{n}_T/T^{1/3} \to \infty$, $r < 13q/24 + 49/48$, $\sqrt{T}||\widehat{\beta} - \beta_0|| = O_{\mathbb{P}}(1)$, and $\sup_{k\geq 1} ||(\partial/\partial\beta)\,\widehat{\Gamma}(k)\,|_{\beta=\overline{\beta}}|| = O_{\mathbb{P}}(1)$ using (S.27) and Assumption 3.3-(ii,iii). In addition,

$$\begin{aligned}\mathbb{E}\left(B_{2,T}^2\right) &\leq \mathbb{E}\left(\widehat{\bar{b}}_{2,T}^{-2/(2q+1)} T^{(8q-10)/5(2q+1)} \widetilde{n}_T^{-2} \sum_{k=1}^{S_T}\sum_{j=1}^{S_T} kj \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right| \left|\widetilde{\Gamma}(j) - \Gamma_T(j)\right|\right) \\ &\leq \widehat{\bar{b}}_{2,T}^{-2/(2q+1)-1} T^{(8q-10)/5(2q+1)-2/3-1} S_T^4 \sup_{k\geq 1} T\widehat{\bar{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \\ &\leq \widehat{\bar{b}}_{2,T}^{-2/(2q+1)-1} T^{(8q-10)/5(2q+1)-2/3-1} \left(Tb_{\theta_2,T}\right)^{4r/(2q+1)} \sup_{k\geq 1} T\widehat{\bar{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \\ &\leq T^{1/5} T^{2/5(2q+1)} T^{(8q-10)/5(2q+1)-2/3-1} T^{4r/(2q+1)} T^{-4r/5(2q+1)} \sup_{k\geq 1} T\widehat{\bar{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \to 0,\end{aligned} \tag{N.13}$$

given that $\sup_{k\geq 1} T\widehat{\overline{b}}_{2,T}\mathrm{Var}(\widetilde{\Gamma}(k)) = O(1)$ using Lemma S.A.5 and $r < 46/48 + 20q/48$. Assumption 5.1-(iii) and $\sum_{k=1}^{\infty} k^{1-l} < \infty$ for $l > 2$ yield

$$B_{3,T} \leq \widehat{\overline{b}}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_T^{-1} C_3 \sum_{k=1}^{\infty} k^{1-l} \to 0, \tag{N.14}$$

where we have used $\widetilde{n}_T/T^{3/10} \to \infty$ and $q < 34/4$. Combining (N.10)-(N.14), 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

$$L_{1,T} = T^{8q/10(2q+1)} \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}k\right)\left(\widehat{\Gamma}(k) - \widetilde{\Gamma}(k)\right), \tag{N.15}$$

$$L_{2,T} = L_{2,T}^A + L_{2,T}^B = T^{8q/10(2q+1)} \left(\sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)} \rfloor} + \sum_{k=\lfloor D_T T^{8/5(2q+1)} \rfloor + 1}^{T-1}\right) K_1\left(\widehat{b}_{1,T}k\right)\left(\widetilde{\Gamma}(k) - \Gamma_T(k)\right),$$

$$L_{3,T} = T^{8q/10(2q+1)} \sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}k\right)\Gamma_T(k).$$

We apply a mean-value expansion and use $\sqrt{T}(\widehat{\beta} - \beta_0) = O_{\mathbb{P}}(1)$ as well as (S.27) to obtain

$$\begin{aligned}
|L_{1,T}| &= T^{8q/10(2q+1)-1/2} \sum_{k=S_T+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta} - \beta_0\right)\right| \\
&= T^{8q/10(2q+1)-1/2+4b/5(2q+1)} \sum_{k=S_T+1}^{T-1} C_1 k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta} - \beta_0\right)\right| \\
&= T^{8q/10(2q+1)-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}(k)\right)|_{\beta=\overline{\beta}}\sqrt{T}\left(\widehat{\beta} - \beta_0\right)\right| \\
&= T^{8q/10(2q+1)-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)} O_{\mathbb{P}}(1)\, O_{\mathbb{P}}(1),
\end{aligned} \tag{N.16}$$

which goes to zero since $r > (8b - 5 - 2q)/8(b-1)$. Let us now consider $L_{2,T}$. We have

$$\begin{aligned}
\left|L_{2,T}^A\right| &= T^{(8q-1)/10(2q+1)} \sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)} \rfloor} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right| \\
&= C_1\left(2qK_{1,q}^2\widehat{\phi}(q)\right)^{b/(2q+1)} T^{8q/10(2q+1)+b/(2q+1)-1/2}\widehat{\overline{b}}_{2,T}^{b/(2q+1)-1/2}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)} \rfloor} k^{-b}\right) \\
&\quad \times \sqrt{T\widehat{\overline{b}}_{2,T}}\left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|.
\end{aligned} \tag{N.17}$$

Note that

$$
\begin{aligned}
&\mathbb{E}\left(T^{8q/10(2q+1)+b/(2q+1)-1/2}\widehat{\overline{b}}_{2,T}^{b/(2q+1)-1/2}\sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)}\rfloor} k^{-b}\sqrt{T\widehat{\overline{b}}_{2,T}}\left|\widetilde{\Gamma}\left(k\right)-\Gamma_T\left(k\right)\right|\right)^2 \qquad \text{(N.18)}\\
&\leq T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)}\rfloor} k^{-b}\sqrt{T\widehat{\overline{b}}_{2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^2\\
&= T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1}\left(\sum_{k=S_T+1}^{\lfloor D_T T^{8/5(2q+1)}\rfloor} k^{-b}\right)^2 O\left(1\right)\\
&= T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1} S_T^{2(1-b)}O\left(1\right)\rightarrow 0,
\end{aligned}
$$

since $r>1.25$ and $T\widehat{\overline{b}}_{2,T}\mathrm{Var}(\widetilde{\Gamma}\left(k\right))=O\left(1\right)$ as above. Next,

$$
\begin{aligned}
\left|L_{2,T}^{B}\right| &= T^{(8q-1)/10(2q+1)}\sum_{k=\lfloor D_T T^{8/5(2q+1)}\rfloor+1}^{T-1} C_1\left(\widehat{b}_{1,T}k\right)^{-b}\left|\widetilde{\Gamma}\left(k\right)-\Gamma_T\left(k\right)\right| \qquad \text{(N.19)}\\
&= C_1\left(2qK_{1,q}^2\widehat{\phi}\left(q\right)\right)^{b/(2q+1)} T^{8q/10(2q+1)+b/(2q+1)-1/2}\widehat{\overline{b}}_{2,T}^{b/(2q+1)-1/2}\left(\sum_{k=\lfloor D_T T^{8/5(2q+1)}\rfloor+1}^{T-1} k^{-b}\right)\\
&\quad\times\sqrt{T\widehat{\overline{b}}_{2,T}}\left|\widetilde{\Gamma}\left(k\right)-\Gamma_T\left(k\right)\right|.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\mathbb{E}\left(T^{8q/10(2q+1)+b/(2q+1)-1/2}\widehat{b}_{2,T}^{b/(2q+1)-1/2}\sum_{k=\lfloor D_T T^{8/5(2q+1)}\rfloor+1}^{T-1} k^{-b}\sqrt{T\widehat{\overline{b}}_{2,T}}\left|\widetilde{\Gamma}\left(k\right)-\Gamma_T\left(k\right)\right|\right)^2 \qquad \text{(N.20)}\\
&\leq T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{b}_{2,T}^{2b/(2q+1)-1}\left(\sum_{k=\lfloor D_T T^{8/5(2q+1)}\rfloor+1}^{T-1} k^{-b}\sqrt{T\widehat{\overline{b}}_{2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^2\\
&= T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{b}_{2,T}^{2b/(2q+1)-1}\left(\sum_{k=\lfloor D_T T^{8/5(2q+1)}\rfloor+1}^{T-1} k^{-b}\right)^2 O\left(1\right)\\
&= T^{8q/5(2q+1)+2b/(2q+1)-1}\widehat{b}_{2,T}^{2b/(2q+1)-1} S_T^{2(1-b)}T^{16(1-b)/5(2q+1)}D_T^2 O\left(1\right)\rightarrow 0,
\end{aligned}
$$

since $r > (b/2 - 1/4)/(b-1)$. Combining (N.17) and (N.20) yields $L_{2,T} \xrightarrow{\mathbb{P}} 0$, since $\widehat{\phi}(q) = O_{\mathbb{P}}(1)$. Let us turn to $L_{3,T}$. By Assumption 5.1-(iii) and $|K_1(\cdot)| \leq 1$, we have,

$$\begin{aligned} |L_{3,T}| &\leq T^{8q/10(2q+1)} \sum_{k=S_T}^{T-1} C_3 k^{-l} \leq T^{8q/10(2q+1)} C_3 S_T^{1-l} \\ &\leq C_3 T^{8q/10(2q+1)} T^{-4r(l-1)/5(2q+1)} \rightarrow 0, \end{aligned} \tag{N.21}$$

since $r > q/(l-1)$. In view of (N.15)-(N.21), we deduce that $A_{2,1,T} \xrightarrow{\mathbb{P}} 0$. Applying the same argument to $A_{2,2,T}$, we have $A_{2,T} \xrightarrow{\mathbb{P}} 0$. Using similar arguments, one has $A_{3,T} \xrightarrow{\mathbb{P}} 0$. It remains to show that $T^{8q/10(2q+1)}(\widehat{J}_T(b_{\theta_1,T},\, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. 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 $k \geq 0$,

$$\begin{aligned} &\widehat{c}_T(rn_T/T,\, k) - \widehat{c}_{\theta_2,T}(rn_T/T,\, k) \\ &= \left(T\overline{b}_{\theta_2,T}\right)^{-1} \sum_{s=k+1}^{T} \left(K_2\left(\frac{((r+1)n_T - (s-k/2))/T}{\widehat{b}_{2,T}((r+1)n_T/T)}\right) - K_2\left(\frac{((r+1)n_T - (s-k/2))/T}{b_{\theta_2,T}((r+1)n_T/T)}\right)\right) \widehat{V}_s \widehat{V}_{s-k} \\ &\quad + O_{\mathbb{P}}\left(1/T\overline{b}_{\theta_2,T}\right). \end{aligned} \tag{N.22}$$

Given Assumption 5.1-(v,vi,vii) and using the delta method, we have for $s \in \{Tu - \lfloor Tb_{\theta_2,T} \rfloor, \ldots, Tu + \lfloor Tb_{\theta_2,T} \rfloor\}$:

$$\begin{aligned} &K_2\left(\frac{(Tu-(s-k/2))/T}{\widehat{b}_{2,T}(u)}\right) - K_2\left(\frac{(Tu-(s-k/2))/T}{b_{\theta_2,T}(u)}\right) \\ &\leq C_4 \left|\frac{Tu-(s-k/2)}{T\widehat{b}_{2,T}(u)} - \frac{Tu-(s-k/2)}{Tb_{\theta_2,T}(u)}\right| \\ &\leq CT^{-4/5-2/5}T^{2/5} \left|\left(\frac{\widehat{D}_2(u)}{\widehat{D}_1(u)}\right)^{-1/5} - \left(\frac{D_2(u)}{D_{1,\theta}(u)}\right)^{-1/5}\right| |Tu-(s-k/2)| \\ &\leq CT^{-4/5-2/5} O_{\mathbb{P}}(1) |Tu-(s-k/2)|. \end{aligned} \tag{N.23}$$

Therefore,

$$\begin{aligned} &T^{8q/10(2q+1)}\left(\widehat{J}_T\left(b_{\theta_1,T},\, \widehat{\overline{b}}_{2,T}\right) - \widehat{J}_T(b_{\theta_1,T},\, b_{\theta_2,T})\right) \\ &= T^{8q/10(2q+1)} \sum_{k=-T+1}^{T-1} K_1(b_{\theta_1,T}k) \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} (\widehat{c}(rn_T/T,\, k) - \widehat{c}_{\theta_2,T}(rn_T/T,\, k)) \\ &\leq T^{8q/10(2q+1)} C \sum_{k=-T+1}^{T-1} |K_1(b_{\theta_1,T}k)| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T\overline{b}_{\theta_2,T}} \\ &\quad \times \sum_{s=k+1}^{T} \left|K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}_{2,T}((r+1)n_T/T)}\right) - K_2\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{\theta_2,T}((r+1)n_T/T)}\right)\right| \end{aligned} \tag{N.24}$$

$$\times \left| \left(\widehat{V}_s \widehat{V}_{s-k} - V_s V_{s-k} \right) + \left(V_s V_{s-k} - \mathbb{E}\left(V_s V_{s-k} \right) \right) + \mathbb{E}\left(V_s V_{s-k} \right) \right|$$
$$\triangleq H_{1,T} + H_{2,T} + H_{3,T}.$$

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 restricted to $k = 1, \ldots, S_T$ (resp., $k = S_T + 1, \ldots, T$). By a mean-value expansion, using (N.23),

$$\begin{aligned}
|H_{1,1,T}| \leq{} & CT^{8q/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} \left| K_1 \left(b_{\theta_1,T} k \right) \right| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T \overline{b}_{\theta_2,T}} \\
& \sum_{s=k+1}^{T} \left| K_2 \left(\frac{\left((r+1) n_T - (s - k/2) \right)/T}{\widehat{\overline{b}}_{2,T} \left((r+1) n_T / T \right)} \right) - K_2 \left(\frac{\left((r+1) n_T - (s - k/2) \right)/T}{b_{\theta_2,T} \left((r+1) n_T / T \right)} \right) \right| \\
& \times \left\| V_s \left(\overline{\beta} \right) \frac{\partial}{\partial \beta} V_{s-k} \left(\overline{\beta} \right) + V_{s-k} \left(\overline{\beta} \right) \frac{\partial}{\partial \beta} V_s \left(\overline{\beta} \right) \right\| \sqrt{T} \left\| \widehat{\beta} - \beta_0 \right\| \\
\leq{} & CT^{8q/10(2q+1)} \overline{b}_{\theta_2,T}^{-1} T^{-1/2-2/5} S_T \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} O_{\mathbb{P}}(1) \\
& \times \left(\left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} V_s^2 (\beta) \right)^2 \left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} V_s (\beta) \right\|^2 \right)^{1/2} \right) \sqrt{T} \left\| \widehat{\beta} - \beta_0 \right\|.
\end{aligned}$$

Using Assumption 3.3 the right-hand side above is such that

$$CT^{8q/10(2q+1)} T^{-1/2-2/5} b_{\theta_2,T}^{-1} S_T \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,$$

since $r < 7/8 + 3q/4$. Next,

$$\begin{aligned}
|H_{1,2,T}| \leq{} & CT^{8q/10(2q+1)} T^{-1/2} \sum_{k=S_T+1}^{T-1} \left(b_{\theta_1,T} k \right)^{-b} \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \frac{1}{T b_{\theta_2,T}} \\
& \times \sum_{s=k+1}^{T} \left| K_2 \left(\frac{\left((r+1) n_T - (s - k/2) \right)/T}{\widehat{b}_{2,T} \left((r+1) n_T / T \right)} \right) - K_2 \left(\frac{\left((r+1) n_T - (s - k/2) \right)/T}{b_{\theta_2,T} \left((r+1) n_T / T \right)} \right) \right| \\
& \times \left\| V_s \left(\overline{\beta} \right) \frac{\partial}{\partial \beta} V_{s-k} \left(\overline{\beta} \right) + V_{s-k} \left(\overline{\beta} \right) \frac{\partial}{\partial \beta} V_s \left(\overline{\beta} \right) \right\| \sqrt{T} \left\| \widehat{\beta} - \beta_0 \right\| \\
\leq{} & CT^{8q/10(2q+1)} b_{\theta_2,T}^{-1} T^{-1/2-2/5} b_{\theta_1,T}^{-b} \sum_{k=S_T+1}^{T-1} k^{-b} \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} O_{\mathbb{P}}(1) \\
& \times \left(\left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} V_s^2 (\beta) \right)^2 \left(T^{-1} \sum_{s=1}^{T} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} V_s (\beta) \right\|^2 \right)^{1/2} \right) \sqrt{T} \left\| \widehat{\beta} - \beta_0 \right\| \\
\leq{} & CT^{8q/10(2q+1)} b_{\theta_2,T}^{-1} T^{-1/2-2/5} b_{\theta_1,T}^{-b} \sum_{k=S_T+1}^{T-1} k^{-b} O_{\mathbb{P}}(1)
\end{aligned}$$

$$
\begin{aligned}
&\leq CT^{8q/10(2q+1)} b_{\theta_2,T}^{-1} T^{-1/2-2/5} b_{\theta_1,T}^{-b} S_T^{1-b} O_{\mathbb{P}}(1) \\
&\leq CT^{8q/10(2q+1)} b_{\theta_2,T}^{-1} T^{-1/2-2/5} b_{\theta_1,T}^{-b} b_{\theta_1,T}^{-r(1-b)} O_{\mathbb{P}}(1) \\
&\leq CT^{8q/10(2q+1)} b_{\theta_2,T}^{-1} T^{-1/2-2/5} b_{\theta_1,T}^{-b} T^{4r(1-b)/5(2q+1)} O_{\mathbb{P}}(1) \to 0,
\end{aligned}
$$

since $r > (8b - 7 - 6q)/8(b-1)$. 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 restricted to $k = 1, \ldots, S_T$ (resp., $k = S_T + 1, \ldots, T$). We have

$$
\begin{aligned}
\mathbb{E}\left(H_{2,1,T}^2\right) \leq\ & T^{8q/5(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} K_1(b_{\theta_1,T} k) K_1(b_{\theta_1,T} j) \left(\frac{n_T}{T}\right)^2 \sum_{r_1=0}^{\lfloor T/n_T \rfloor} \sum_{r_2=0}^{\lfloor T/n_T \rfloor} \frac{1}{(T b_{\theta_2,T})^2} \\
&\times \sum_{s=k+1}^{T} \sum_{t=j+1}^{T} \left| K_2\left(\frac{((r_1+1) n_T - (s-k/2))/T}{\widehat{b}_{2,T}((r_1+1) n_T/T)}\right) - K_2\left(\frac{((r_1+1) n_T - (s-k/2))/T}{b_{\theta_2,T}((r_1+1) n_T/T)}\right) \right| \\
&\times \left| K_2\left(\frac{((r_2+1) n_T - (t-j/2))/T}{\widehat{b}_{2,T}((r_2+1) n_T/T)}\right) - K_2\left(\frac{((r_2+1) n_T - (t-j/2))/T}{b_{\theta_2,T}((r_2+1) n_T/T)}\right) \right| \\
&\times \left| \mathbb{E}\left(V_s V_{s-k} - \mathbb{E}(V_s V_{s-k})\right)\left(V_t V_{t-k} - \mathbb{E}(V_t V_{t-k})\right) \right| \\
\leq\ & CT^{8q/5(2q+1)} S_T^2 T^{-2/5} (T b_{\theta_2,T})^{-1} \sup_{k \geq 1} T b_{\theta_2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) O_{\mathbb{P}}(1) \\
\leq\ & CT^{(8q+8r)/5(2q+1)-2/5-1} O_{\mathbb{P}}\left(b_{\theta_2,T}^{-1}\right) \to 0,
\end{aligned}
\tag{N.25}
$$

where we have used Lemma S.A.5 and $r < (6+4q)/8$. Turning to $H_{2,2,T}$,

$$
\begin{aligned}
\mathbb{E}\left(H_{2,2,T}^2\right) &\leq T^{8q/5(2q+1)-2/5} (T b_{\theta_2,T})^{-1} b_{\theta_1,T}^{-2b} \left(\sum_{k=S_T+1}^{T-1} k^{-b} \sqrt{T b_{\theta_2,T}} \left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2} O(1) \right)^2 \\
&\leq T^{8q/5(2q+1)} T^{-2/5-1} b_{\theta_2,T}^{-1} b_{\theta_1,T}^{-2b} \left(\sum_{k=S_T+1}^{T-1} k^{-b} \sqrt{T b_{\theta_2,T}} \left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2} \right)^2 \\
&\leq T^{8q/5(2q+1)} T^{-2/5-1} b_{\theta_2,T}^{-1} b_{\theta_1,T}^{-2b} \left(\sum_{k=S_T+1}^{T-1} k^{-b} O(1) \right)^2 \\
&\leq T^{8q/5(2q+1)} T^{-2/5-1} b_{\theta_2,T}^{-1} b_{\theta_1,T}^{-2b} S_T^{2(1-b)} \to 0,
\end{aligned}
\tag{N.26}
$$

since $r > (b - 3/4 - q/2)/(b-1)$. Combining (N.25)-(N.26) yields $H_{2,T} \xrightarrow{\mathbb{P}} 0$. Let $H_{3,1,T}$ (resp., $H_{3,2,T}$) be defined as $H_{3,T}$ but with the sum over k restricted to $k = 1, \ldots, S_T$ (resp., $k = S_T + 1, \ldots, T$). Given $|K_1(\cdot)| \leq 1$ and (N.23), we have

$$
\begin{aligned}
|H_{3,1,T}| &\leq CT^{8q/10(2q+1)} T^{-2/5} \sum_{k=1}^{S_T} |\Gamma_T(k)| \\
&\leq CT^{8q/10(2q+1)} T^{-2/5} \sum_{k=1}^{\infty} k^{-l} \to 0,
\end{aligned}
$$

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

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

This completes the proof of part (ii).

We now move to part (i). For some arbitrary $\phi_{\theta^*} \in (0, \infty)$, $\widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T}) - J_T = o_{\mathbb{P}}(1)$ by Theorem 3.2-(i) since $b_{\theta_2,T} = O(T^{-1/5})$ and $q > 1/2$ imply that $\sqrt{T} b_{1,T} \to \infty$ holds. Hence, it remains to show that $\widehat{J}_T(b_{\theta_1,T}, b_{\theta_2,T}) - \widehat{J}_T(\widehat{b}_{1,T}, \widehat{b}_{2,T}) = o_{\mathbb{P}}(1)$. Note that this result differs from that of part (ii) only because the scale factor $T^{8q/10(2q+1)}$ does not appear, Assumption 5.1-(ii) is replaced by part (i) of the same assumption, Assumption 5.1-(iii, v, vi) is not imposed, and $q > 1/2$. Let S_T be defined as in part (ii) and

$$\begin{aligned}
r \in &(\max\{(8b - 10q - 5)/8(b-1), (b - 1/2 - q)/(b-1)\}, \\
&\min\{13/16 + 5q/8, (3+2q)/4, 1\}),
\end{aligned}$$

with $b > 1 + 1/q$. We will use the decomposition in (N.8), and N_1 and N_2 as defined after (N.8). Let $A_{1,T}$, $A_{2,T}$ and $A_{3,T}$ be as in (N.9) without the scale factor $T^{8q/10(2q+1)}$. Proceeding as in (N.10), we have

$$\begin{aligned}
|A_{1,1,T}| &\leq \sum_{k=1}^{S_T} C_2 \left| \widehat{b}_{1,T} - b_{\theta_1,T} \right| k \left| \widehat{\Gamma}(k) \right| \qquad \text{(N.27)} \\
&\leq C \left| \widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*}^{1/(2q+1)} \right| \left(\widehat{\phi}(q) \phi_{\theta^*} \right)^{-1/(2q+1)} \left(T \widehat{\widehat{b}}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \widehat{\Gamma}(k) \right|,
\end{aligned}$$

for some $C < \infty$. By Assumption 5.1-(i),

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

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

$$\begin{aligned}
B_{1,T} &= \left(T \widehat{\widehat{b}}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \widehat{\Gamma}(k) - \widetilde{\Gamma}(k) \right|, \qquad \text{(N.28)} \\
B_{2,T} &= \left(T \widehat{\widehat{b}}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \widetilde{\Gamma}(k) - \Gamma_T(k) \right|, \\
B_{3,T} &= \left(T \widehat{\widehat{b}}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \Gamma_T(k) \right|.
\end{aligned}$$

By a mean-value expansion, we have

$$
\begin{aligned}
B_{1,T} &\leq \left(T\widehat{\overline{b}}_{2,T}\right)^{-1/(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k) |_{\beta=\overline{\beta}} \right) \sqrt{T} \left(\widehat{\beta} - \beta_0 \right) \right| \\
&\leq C \left(T\widehat{\overline{b}}_{2,T}\right)^{-1/(2q+1)} \left(Tb_{\theta_{2,T}}\right)^{2r/(2q+1)} T^{-1/2} \sup_{k\geq 1} \left\| \frac{\partial}{\partial \beta} \widehat{\Gamma}(k) |_{\beta=\overline{\beta}} \right\| \sqrt{T} \left\| \widehat{\beta} - \beta_0 \right\|,
\end{aligned} \tag{N.29}
$$

since $r < 13/16 + 5q/8$, and $\sup_{k\geq 1} || \left(\partial/\partial\beta\right) \widehat{\Gamma}(k) |_{\beta=\overline{\beta}} || = O_{\mathbb{P}}(1)$ using (S.27) and Assumption 3.3-(ii,iii). In addition,

$$
\begin{aligned}
\mathbb{E}\left(B_{2,T}^2\right) &\leq \mathbb{E}\left(\left(T\widehat{\overline{b}}_{2,T}\right)^{-2/(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} kj \left| \widetilde{\Gamma}(k) - \Gamma_T(k) \right| \left| \widetilde{\Gamma}(j) - \Gamma_T(j) \right| \right) \\
&\leq \mathbb{E}\left(\left(T\widehat{\overline{b}}_{2,T}\right)^{-2/(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} kj \left| \widetilde{\Gamma}(k) - \Gamma_T(k) \right| \left| \widetilde{\Gamma}(j) - \Gamma_T(j) \right| \right) \\
&\leq \left(T\widehat{\overline{b}}_{2,T}\right)^{-2/(2q+1)-1} S_T^4 \sup_{k\geq 1} T\widehat{\overline{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \\
&\leq \left(T\widehat{\overline{b}}_{2,T}\right)^{-2/(2q+1)-1} \left(Tb_{2,T}\right)^{4r/(2q+1)} \sup_{k\geq 1} T\widehat{\overline{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \\
&\leq \widehat{\overline{b}}_{2,T}^{-2/(2q+1)-1} T^{-1-2/(2q+1)} T^{16r/5(2q+1)} \sup_{k\geq 1} T\widehat{\overline{b}}_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}(k)\right) \to 0,
\end{aligned} \tag{N.30}
$$

given that $\sup_{k\geq 1} T\widehat{\overline{b}}_{2,T} \mathrm{Var}(\widetilde{\Gamma}(k)) = O(1)$ by Lemma S.A.5 and $r < (3+2q)/4$. The bound in (N.14) is replaced by

$$
\begin{aligned}
B_{3,T} &\leq \left(T\widehat{\overline{b}}_{2,T}\right)^{-1/(2q+1)} S_T \sum_{k=1}^{\infty} |\Gamma_T(k)| \\
&\leq \left(T\widehat{\overline{b}}_{2,T}\right)^{(r-1)/(2q+1)} O_{\mathbb{P}}(1) \to 0,
\end{aligned} \tag{N.31}
$$

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 (N.15) without the factor $T^{8q/10(2q+1)}$. Equation (N.16) is then replaced by

$$
\begin{aligned}
|L_{1,T}| &= T^{-1/2} \sum_{k=S_T+1}^{T-1} C_1 \left(\widehat{b}_{1,T} k\right)^{-b} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k) \right) |_{\beta=\overline{\beta}} \sqrt{T} \left(\widehat{\beta} - \beta_0 \right) \right| \\
&= T^{-1/2+4b/5(2q+1)} \sum_{k=S_T+1}^{T-1} C_1 k^{-b} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k) \right) |_{\beta=\overline{\beta}} \sqrt{T} \left(\widehat{\beta} - \beta_0 \right) \right| \\
&= T^{-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}(k) \right) |_{\beta=\overline{\beta}} \sqrt{T} \left(\widehat{\beta} - \beta_0 \right) \right| \\
&= T^{-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)} O(1) O_{\mathbb{P}}(1),
\end{aligned} \tag{N.32}
$$

which converges to zero since $r > (8b - 10q - 5)/8(b-1)$. Also, (N.17) is replaced by

$$
\begin{aligned}
|L_{2,T}| &= \sum_{k=S_T+1}^{T-1} C_1 \left(\widehat{b}_{1,T} k\right)^{-b} \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right| \qquad \text{(N.33)} \\
&= C_1 \left(q K_{1,q}^2 \widehat{\phi}(q)\right)^{b/(2q+1)} T^{b/(2q+1)-1/2} \widehat{b}_{2,T}^{b/(2q+1)-1/2} \left(\sum_{k=S_T+1}^{T-1} k^{-b}\right) \sqrt{T \widehat{b}_{2,T}} \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|,
\end{aligned}
$$

and the bound in (N.18) is replaced by,

$$
\begin{aligned}
&\mathbb{E}\left(T^{b/(2q+1)-1/2} \widehat{\overline{b}}_{2,T}^{b/(2q+1)-1/2} \sum_{k=S_T}^{T-1} k^{-b} \sqrt{T \widehat{\overline{b}}_{2,T}} \left|\widetilde{\Gamma}(k) - \Gamma_T(k)\right|\right)^2 \qquad \text{(N.34)} \\
&\leq T^{2b/(2q+1)-1} \widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1} \left(\sum_{k=S_T}^{T-1} k^{-b} \sqrt{T \widehat{\overline{b}}_{2,T}} \left(\mathrm{Var}\left(\widetilde{\Gamma}(k)\right)\right)^{1/2}\right)^2 \\
&= T^{2b/(2q+1)-1} \widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1} \left(\sum_{k=S_T}^{T-1} k^{-b}\right)^2 O(1) \\
&= T^{2b/(2q+1)-1} \widehat{\overline{b}}_{2,T}^{2b/(2q+1)-1} S_T^{2(1-b)} O(1) \to 0,
\end{aligned}
$$

since $r > (b - 1/2 - q)/(b-1)$ and $T b_{2,T} \mathrm{Var}(\widetilde{\Gamma}(k)) = O(1)$, as above. Combining (N.33)-(N.34) yields $L_{2,T} \xrightarrow{\mathbb{P}} 0$ since $\widehat{\phi}(q) = O_{\mathbb{P}}(1)$. Let us turn to $L_{3,T}$. We have (N.21) replaced by,

$$
\begin{aligned}
\left|\sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T} k\right) \Gamma_T(k)\right| &\leq \sum_{k=S_T+1}^{T-1} \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} |c(r n_T/T, k)| \qquad \text{(N.35)} \\
&\leq \sum_{k=S_T+1}^{T-1} \sup_{u \in [0,1]} |c(u, k)| \to 0.
\end{aligned}
$$

Equations (N.32)-(N.35) 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 $(\widehat{J}_T(b_{\theta_1,T}, \widehat{\overline{b}}_{2,T}) - \widehat{J}_T(b_{\theta_1,T}, \overline{b}_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. Let $\widehat{c}_{\theta_2,T}(r n_T/T, k)$ be defined as in part (ii). We have (N.22), and (N.23) is replaced by

$$
\begin{aligned}
&K_2\left(\frac{(Tu - (s-k/2)/T)}{\widehat{b}_{2,T}(u)}\right) - K_2\left(\frac{(Tu - (s-k/2)/T)}{b_{\theta_2,T}(u)}\right) \qquad \text{(N.36)} \\
&\leq C_4 \left|\frac{Tu - (s-k/2)}{T \widehat{b}_{2,T}(u)} - \frac{Tu - (s-k/2)}{T b_{\theta_2,T}(u)}\right| \\
&\leq C_4 T^{-1} \left|\frac{Tu - (s-k/2)\left(\widehat{b}_{2,T}(u) - b_{\theta_2,T}(u)\right)}{\widehat{b}_{2,T}(u)\, b_{\theta_2,T}(u)}\right|
\end{aligned}
$$

$$
\begin{aligned}
&= C_4 T^{-4/5}\left(\left(\frac{\widehat{D}_1(u)}{\widehat{D}_2(u)}\right)\left(\frac{D_{1,\theta}(u)}{D_2(u)}\right)\right)^{1/5}\left|\left(\frac{\widehat{D}_2(u)}{\widehat{D}_1(u)}\right)^{1/5}-\left(\frac{D_2(u)}{D_{1,\theta}(u)}\right)^{1/5}\right|\left|Tu-(s-k/2)\right| \\
&= CT^{-4/5}\left|Tu-(s-k/2)\right|,
\end{aligned}
$$

for $s \in \{Tu-\lfloor Tb_{\theta_2,T}(u)\rfloor, \ldots, Tu+\lfloor Tb_{\theta_2,T}(u)\rfloor\}$, where $u=(r+1)\,n_T/T$. Therefore,

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

We have to show that $H_{1,T}+H_{2,T}+H_{3,T}\overset{\mathbb{P}}{\to}0$. By a mean-value expansion, using (N.36),

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

Using Assumption 3.3 and (N.36), the right-hand side above is such that

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

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 restricted to $k = 1, \ldots, S_T$ (resp., $k = S_T + 1, \ldots, T$). We have

$$\begin{aligned}
\mathbb{E}\left(H_{2,1,T}^2\right) &\leq \sum_{k=1}^{S_T}\sum_{j=1}^{S_T} K_1\left(b_{\theta_1,T}k\right) K_1\left(b_{\theta_1,T}j\right) \qquad \text{(N.37)}\\
&\times \left(\frac{n_T}{T}\right)^2 \sum_{r_1=0}^{\lfloor T/n_T \rfloor}\sum_{r_2=0}^{\lfloor T/n_T \rfloor} \frac{1}{\left(Tb_{\theta_2,T}\right)^2} \sum_{s=k+1}^{T}\sum_{t=j+1}^{T} \\
&\times \left|K_2\left(\frac{\left(\left(r_1+1\right)n_T - \left(s-k/2\right)\right)/T}{\widehat{b}_{2,T}\left(\left(r_1+1\right)n_T/T\right)}\right) - K_2\left(\frac{\left(\left(r_1+1\right)n_T - \left(s-k/2\right)\right)/T}{b_{\theta_2,T}\left(\left(r_1+1\right)n_T/T\right)}\right)\right| \\
&\times \left|K_2\left(\frac{\left(\left(r_2+1\right)n_T - \left(t-j/2\right)\right)/T}{\widehat{b}_{2,T}\left(\left(r_2+1\right)n_T/T\right)}\right) - K_2\left(\frac{\left(\left(r_2+1\right)n_T - \left(t-j/2\right)\right)/T}{b_{\theta_2,T}\left(\left(r_2+1\right)n_T/T\right)}\right)\right| \\
&\times \left|V_sV_{s-k} - \mathbb{E}\left(V_sV_{s-k}\right)\left(V_tV_{t-k} - \mathbb{E}\left(V_tV_{t-k}\right)\right)\right|. \\
&\leq CS_T^2\left(Tb_{\theta_2,T}\right)^{-1} \sup_{k\geq 1} Tb_{\theta_2,T}\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right) O_{\mathbb{P}}\left(1\right) \\
&\leq CT^{8r/5(2q+1)} O_{\mathbb{P}}\left(T^{-1}b_{\theta_2,T}^{-1}\right) \to 0,
\end{aligned}$$

where we have used Lemma S.A.5, (N.36) and $r < 3/2$. Turning to $H_{2,2,T}$,

$$\begin{aligned}
\mathbb{E}\left(H_{2,2,T}^2\right) &\leq \left(Tb_{\theta_2,T}\right)^{-1} b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1} k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2} O\left(1\right)\right)^2 \qquad \text{(N.38)}\\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1} k^{-b}\sqrt{Tb_{\theta_2,T}}\left(\mathrm{Var}\left(\widetilde{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^2 \\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b}\left(\sum_{k=S_T+1}^{T-1} k^{-b} O\left(1\right)\right)^2 \\
&\leq T^{-1}b_{\theta_2,T}^{-1}b_{\theta_1,T}^{-2b} S_T^{2(1-b)} \to 0,
\end{aligned}$$

since $r > \left(b-q-1/2\right)/\left(b-1\right)$. Combining (N.37)-(N.38) yields $H_{2,T} \xrightarrow{\mathbb{P}} 0$. Given $\left|K_1\left(\cdot\right)\right| \leq 1$ and (N.36), 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 5.1-(i,ii). □

References

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

BARY, N. (1964): *Treatise on Trigonometric Series*, vol. I. Pergamon, Oxford.

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

DAHLHAUS, R., AND L. GIRAITIS (1998): "On the Optimal Segment Length for Parameter Estimates for Locally Stationary Time Series," *Journal of Time Series Analysis*, 19(6), 629–655.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

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