Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference*

Alessandro Casini[†] University of Rome Tor Vergata

Taosong Deng[‡] Hunan University

Pierre Perron[§] Boston University

22nd August 2023

Abstract

We establish theoretical results about the low frequency contamination (i.e., long memory effects) induced by general nonstationarity for estimates such as the sample autocovariance and the periodogram, and deduce consequences for heteroskedasticity and autocorrelation robust (HAR) inference. We present explicit expressions for the asymptotic bias of these estimates. We distinguish cases where this contamination only occurs as a small-sample problem and cases where the contamination continues to hold asymptotically. We show theoretically that nonparametric smoothing over time is robust to low frequency contamination. Our results allow a better understanding of long-run variance (LRV) estimation. Existing LRV estimators tend to be inflated when the data are nonstationary. Thus, HAR tests can be undersized and exhibit dramatic power losses. Our theory indicates that long bandwidths or fixed-b HAR tests suffer more from low frequency contamination relative to HAR tests based on HAC estimators, whereas recently introduced double kernel HAC estimators do not suffer from this problem. We present second-order Edgeworth expansions under nonstationarity about the distribution of HAC and DK-HAC estimators and about the corresponding t-test in the regression model. The results show that the low frequency contamination can be induced by time variation in the second moments even when there is no break in the mean.

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

Keywords: Edgeworth expansions, Fixed-b, HAC standard errors, HAR, Long memory, Long-run variance, Low frequency contamination, Nonstationarity, Outliers, Segmented locally stationary.

*We are grateful to Anna Mikusheva and the referees for useful suggestions. We thank Whitney Newey and Tim Vogelsang for discussions and Andrew Chesher, Adam McCloskey, Zhongjun Qu and Daniel Whilem for comments.

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

[‡]College of Finance and Statistics, Hunan University, 109 Shijiachong Road, Yuelu District, Changsha, Hunan 41006, China. Email: tsdeng@hnu.edu.cn.

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

1 Introduction

Economic and financial time series are highly nonstationary [see, e.g., Perron (1989), Stock and Watson (1996), Ng and Wright (2013), and Giacomini and Rossi (2015)]. We develop theoretical results about the behavior of the sample autocovariance ($\widehat{\Gamma}(k), k \in \mathbb{Z}$) and the periodogram ($I_T(\omega), \omega \in [-\pi, \pi]$) for a short memory nonstationary process. This means processes that have non-constant moments and whose sum of absolute autocovariances is finite. The latter rules out processes with unbounded second moments (e.g., unit root). We show that nonstationarity (e.g., time-variation in the mean or autocovariances) induces low frequency contamination, meaning that the sample autocovariance and the periodogram share features that are similar to those of a long memory series. We present explicit expressions for the asymptotic bias of these estimates, showing that it is always positive and increases with the degree of heterogeneity in the data.

The low frequency contamination can be explained as follows. For a short memory series, the autocorrelation function (ACF) displays exponential decay and vanishes as the lag length $k \to \infty$, and the periodogram is finite at the origin. Under general forms of nonstationarity, we show theoretically that $\widehat{\Gamma}(k) = \lim_{T\to\infty} \Gamma_T(k) + d^*$, where $\Gamma_T(k) = T^{-1}\sum_{t=k+1}^{T} \mathbb{E}(V_t V_{t-k})$, $k \geq 0$ and $d^* > 0$ is independent of k. Assuming positive dependence for simplicity (i.e., $\lim_{T\to\infty} \Gamma_T(k) > 0$), this means that each sample autocovariance overestimates the true dependence in the data. The bias factor $d^* > 0$ depends on the type of nonstationarity and in general does not vanish as $T \to \infty$. In addition, since short memory implies $\Gamma_T(k) \to 0$ as $k \to \infty$, it follows that d^* generates long memory effects since $\widehat{\Gamma}(k) \approx d^* > 0$ as $k \to \infty$. As for the periodogram, $I_T(\omega)$, we show that under nonstationarity $\mathbb{E}(I_T(\omega)) \to \infty$ as $\omega \to 0$, a feature also shared by long memory processes.

This low frequency contamination holds as an asymptotic approximation. We verify analytically the quality of the approximation to finite-sample situations. We distinguish cases where this contamination only occurs as a small-sample problem and cases where it continues to hold asymptotically. The former involves $d^* \approx 0$ asymptotically but a consistent estimate of d^* satisfies $\widehat{d}^* > 0$ in finite-sample. We show analytically that using $\widehat{d}^*$ in place of d^* provides a good approximation in this context. This helps us to explain the long memory effects in situations where asymptotically no low frequency contamination should occur. An example is a t-test on a regression coefficient in a correctly specified model with nonstationary errors that are serially correlated. Other examples include t-tests in the linear model with mild forms of misspecification that do not undermine the conditions for consistency of the least-squares estimator. Further, similar issues arise if one applies some prior de-trending techniques where the fitted model is not correctly specified (e.g., the data follow a nonlinear trend but one removes a linear trend). Yet another example for which our results are relevant is the case of outliers. In all these examples, $d^* \approx 0$ asymptotically but one observes enhanced persistence in finite-samples that can affect the properties of heteroskedasticity and autocorrelation robust (HAR) inference. Most of the HAR inference in applied work (besides the t- and F-test in regression models) are characterized by nonstationary alternative hypotheses for which $d^* > 0$ even asymptotically. This class of tests is very large. Tests for forecast evaluation [e.g., Casini (2018), Diebold and Mariano (1995), Giacomini and Rossi (2009, 2010), Giacomini and White (2006), Perron and Yamamoto (2021) and West (1996)], tests and inference for structural change models [e.g., Andrews (1993), Bai and Perron (1998), Casini and Perron (2020, 2021b, 2021c), Elliott and Müller (2007), and Qu and Perron (2007)], tests and inference in time-varying parameters models [e.g., Cai (2007) and Chen and Hong (2012)], tests and inference for regime switching models [e.g., Hamilton (1989) and Qu and Zhuo (2020)] and others are part of this class.

Recently, Casini (2022c) proposed a new HAC estimator that applies nonparametric smoothing over time in order to account flexibly for nonstationarity. We show theoretically that nonparametric smoothing over time is robust to low frequency contamination and prove that the resulting sample local autocovariance and the local periodogram do not exhibit long memory features. Nonparametric smoothing avoids mixing highly heterogeneous data coming from distinct nonstationary regimes as opposed to what the sample autocovariance and the periodogram do.

Our work is different from the literature on spurious persistence caused by the presence of level shifts or other deterministic trends. Perron (1990) showed that the presence of breaks in mean often induces spurious non-rejection of the unit root hypothesis, and that the presence of a level shift asymptotically biases the estimate of the AR coefficient towards one. Bhattacharya, Gupta and Waymire (1983) demonstrated that certain deterministic trends can induce the spurious presence of long memory. In other contexts, similar issues were discussed by Christensen and Varneskov (2017), Diebold and Inoue (2001), Demetrescu and Salish (2020), Lamoureux and Lastrapes (1990), Hillebrand (2005), Granger and Hyung (2004), McCloskey and Hill (2017), Mikosch and Stărica (2004), Müller and Watson (2008) and Perron and Qu (2010). Our results are different from theirs in that we consider a more general problem and we allow for more general forms of nonstationarity using the segmented locally stationary framework of Casini (2022c). Importantly, we provide a general solution to these problems and show theoretically its robustness to low frequency contamination. Moreover, we discuss in detail the implications of our theory for HAR inference.

HAR inference relies on estimation of the long-run variance (LRV). The latter, from a time domain perspective, is equivalent to the sum of all autocovariances while from a frequency domain perspective, is equal to 2π times an integrated time-varying spectral density at the zero frequency.

From a time domain perspective, estimation involves a weighted sum of the sample autocovariances, while from a frequency domain perspective estimation is based on a weighted sum of the periodogram ordinates near the zero frequency. Therefore, our results on low frequency contamination for the sample autocovariances and the periodogram can have important implications.

There are two main approaches in HAR inference, one based on traditional asymptotics and the other based on fixed-smoothing asymptotics. The classical approach relies on a LRV estimator using a small bandwidth [cf. the HAC estimators of Newey and West (1987, 1994) and Andrews (1991)]. Inference is standard because HAR test statistics follow asymptotically standard distributions. It was shown early that HAC standard errors can result in oversized tests when there is substantial temporal dependence. This stimulated a second approach based on a LRV estimator that keeps the bandwidth at a fixed fraction of the sample size and that converges weakly to a random variable [cf. Kiefer, Vogelsang and Bunzel (2000)]. Inference is then based on a nonstandard reference distribution and it is shown that fixed-b achieves high-order refinements [e.g., Sun, Phillips and Jin (2008)] and reduces the oversize problem of HAR tests.[1] However, unlike the classical approach, current fixed-b HAR inference is only valid under stationarity [cf. Casini (2022b)]. More recently, a variant of the fixed-b approach [see, e.g., Sun (2014b) and Lazarus et al. (2018)] considered the use of small-b asymptotics in conjunction with fixed-b critical values. These bandwidths are typically larger than the MSE-optimal bandwidths used for the HAC estimators.

Recent work by Casini (2022c) questioned the performance of HAR inference tests under nonstationarity from a theoretical standpoint. Simulation evidence of serious (e.g., non-monotonic) power or related issues in specific HAR inference contexts were documented by Altissimo and Corradi (2003), Casini (2018), Casini and Perron (2019, 2020, 2021c), Chan (2022a, 2022b), Crainiceanu and Vogelsang (2007), Deng and Perron (2006), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Otto and Breitung (2021), Perron and Yamamoto (2021), Shao and Zhang (2010), Vogelsang (1999) and Zhang and Lavitas (2018) among others]. Our theoretical results show that these issues occur because the unaccounted nonstationarity alters the spectrum at low frequencies. Each sample autocovariance is upward biased ($d^* > 0$) and the resulting LRV estimators tend to be inflated. When these estimators are used to normalize test statistics, the latter lose power. Interestingly, d^* is independent of k so that the more lags are included the more severe is the problem. Further, by virtue of weak dependence, we have that $\Gamma_T(k) \to 0$ as $k \to \infty$

[1]See Dou (2019), Hwang and Sun (2017), Ibragimov, Kattuman and Skrobotov (2021), Ibragimov and Müller (2010), Jansson (2004), Kiefer and Vogelsang (2002, 2005), Lazarus, Lewis and Stock (2021), Lazarus et al. (2018) Müller (2007, 2014), Phillips (2005), Politis (2011), Pötscher and Preinerstorfer (2016, 2018, 2019), Robinson (1998), Sun (2013, 2014a, 2014b) and Zhang and Shao (2013).

but $d^* > 0$ across k. We show formally that long bandwidths/fixed-b LRV estimators are expected to suffer most from power losses because they use many/all lagged autocovariances.

To precisely analyze the theoretical properties of the HAR tests under the null hypothesis, we present second-order Edgeworth expansions under nonstationarity for the distribution of the HAC and DK-HAC estimator and for the distribution of the corresponding t-test in the linear regression model. Under stationarity the results concerning the HAC estimator were provided by Velasco and Robinson (2001). We show that the order of the approximation error of the expansion is the same as under stationarity from which it follows that the error in rejection probability (ERP) is also the same. The ERP of the t-test based on the DK-HAC estimator is slightly larger than that of the t-test based on the HAC estimator due to the double smoothing. High-order asymptotic expansions for spectral and other estimates were studied by Bhattacharya and Ghosh (1978), Bentkus and Rudzkis (1982), Janas (1994), Phillips (1977, 1980) and Taniguchi and Puri (1996). The asymptotic expansions of the fixed-b HAR tests under stationarity were developed by Jansson (2004) and Sun et al. (2008). Casini (2022b) showed that under nonstationarity the ERP of the fixed-b HAR tests can be larger than that of HAR tests based on HAC and DK-HAC estimators thereby controverting the conclusion in the literature that the original fixed-b HAR tests have superior null rejection rates relative to HAR tests based on traditional LRV estimators. Overall, our finite-sample results based on Edgeworth expansions show that the low frequency contamination can arise from time variation in the second moments even when the mean is constant.

The paper is organized as follows. Section 2 presents the statistical setting and Section 3 establishes the theoretical results on low frequency contamination. Section 4 presents the Edgeworth expansions of HAR tests based on the HAC and DK-HAC estimators. The implications of our results for HAR inference are analyzed analytically and computationally through simulations in Section 5. Section 6 concludes. The supplemental materials [cf. Casini, Deng and Perron (2022)] contain some additional examples and all mathematical proofs.

2 Statistical Framework for Nonstationarity

Suppose $\{V_{t,T}\}_{t=1}^{T}$ is defined on a 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 analyze time series models that have a time-varying spectrum it is useful to introduce an infill asymptotic setting whereby we rescale the original discrete time horizon $[1, T]$ by dividing each t by T. Letting $u = t/T$ we define a new time scale $u \in [0, 1]$ on which as $T \to \infty$ we observe more and more realizations of $V_{t,T}$

close to time t. As a notion of nonstationarity, we use the concept of segmented local stationarity (SLS) introduced in Casini (2022c). This extends the locally stationary processes [cf. Dahlhaus (1997)] to allow for structural change and regime switching-type models. SLS processes allow for a finite number of discontinuities in the spectrum over time. We collect the break dates in the set $\mathcal{T} \triangleq \{T_1^0, \ldots, T_m^0\}$. Let $i \triangleq \sqrt{-1}$. A function $G(\cdot, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C}$ is said to be left-differentiable at u_0 if $\partial G(u_0, \omega) / \partial_- u \triangleq \lim_{u \to u_0^-} (G(u_0, \omega) - G(u, \omega)) / (u_0 - u)$ exists for any $\omega \in \mathbb{R}$. Let $m_0 \geq 0$ be a finite integer.

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

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

for $j = 1, \ldots, m_0 + 1$, where by convention $T_0^0 = 0$ and $T_{m_0+1}^0 = T$. The following technical conditions are also assumed to hold: (i) $\xi(\lambda)$ is a process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and

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

where $\operatorname{cum}\{\cdots\}$ denotes the cumulant spectra of r-th order, $g_1 = 0$, $g_2(\omega) = 1$, $|g_r(\omega_1, \ldots, \omega_{r-1})| \leq M_r$ for all r with $M_r < \infty$ that may depend on r, and $\zeta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function $\delta(\cdot)$; (ii) There exists a $K < \infty$ and a piecewise continuous function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0 + 1$, there exists a 2π-periodic function $A_j : (\lambda_{j-1}^0, \lambda_j^0] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$, $\lambda_j^0 \triangleq T_j^0 / T$ and for all T,

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

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

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

Assumption 2.1. *(i) $\{V_{t,T}\}$ is a 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 bounded derivatives $(\partial/\partial u) A(u, \cdot)$ and $(\partial^2/\partial u^2) A(u, \cdot)$; (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 bounded derivatives $(\partial/\partial_- u) A(u, \cdot)$ and $(\partial^2/\partial_- u^2) A(u, \cdot)$ and has piecewise Lipschitz continuous derivative*

$(\partial^2/\partial_- u^2)\, A(u,\, \cdot)$; *(v)* $A(u,\, \omega)$ *is Lipschitz continuous in* ω.

We define the time-varying spectral density as $f_j(u,\, \omega) \triangleq (2\pi)^{-1}|A_j(u,\, \omega)|^2$ for $T^0_{j-1}/T < u = t/T \leq T^0_j/T$. Then we can define the local covariance of $V_{t,T}$ at the 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 $c(u,\, k) \triangleq \lim_{T\to\infty} \int_{-\pi}^{\pi} e^{i\omega k} A(u,\, \omega)\, A(u - k/T,\, -\omega)\, d\omega$.

Next, we impose conditions on the temporal dependence (we omit the second subscript T when it is clear from the context). Let

$$
\begin{aligned}
&\kappa^{(a_1,a_2,a_3,a_4)}_{V,t}(u,\, v,\, w) \\
&\qquad \triangleq \kappa^{(a_1,a_2,a_3,a_4)}(t,\, t+u,\, t+v,\, t+w) - \kappa^{(a_1,a_2,a_3,a_4)}_{\mathscr{N}}(t,\, t+u,\, t+v,\, t+w) \\
&\qquad \triangleq \mathbb{E}\left(V^{(a_1)}_t - \mathbb{E}V^{(a_1)}_t\right)\left(V^{(a_2)}_{t+u} - \mathbb{E}V^{(a_2)}_{t+u}\right)\left(V^{(a_3)}_{t+v} - \mathbb{E}V^{(a_3)}_{t+v}\right)\left(V^{(a_4)}_{t+w} - \mathbb{E}V^{(a_4)}_{t+w}\right) \\
&\qquad\quad - \mathbb{E}\left(V^{(a_1)}_{\mathscr{N},t} - \mathbb{E}V^{(a_1)}_{\mathscr{N},t}\right)\left(V^{(a_2)}_{\mathscr{N},t+u} - \mathbb{E}V^{(a_2)}_{\mathscr{N},t+u}\right)\left(V^{(a_3)}_{\mathscr{N},t+v} - \mathbb{E}V^{(a_3)}_{\mathscr{N},t+v}\right)\left(V^{(a_4)}_{\mathscr{N},t+w} - \mathbb{E}V^{(a_4)}_{\mathscr{N},t+w}\right),
\end{aligned}
$$

where $\{V_{\mathscr{N},t}\}$ is a Gaussian sequence with the same mean and covariance structure as $\{V_t\}$, $\kappa^{(a_1,a_2,a_3,a_4)}_{V,t}(u,\, v,\, w)$ is the time-t fourth-order cumulant of $(V^{(a_1)}_t,\, V^{(a_2)}_{t+u},\, V^{(a_3)}_{t+v},\, V^{(a_4)}_{t+w})$ while $\kappa^{(a_1,a_2,a_3,a_4)}_{\mathscr{N}}$ $(t,\, t+u,\, t+v,\, t+w)$ is the time-t centered fourth moment of V_t if V_t were Gaussian.

Assumption 2.2. *(i)* $\sum_{k=-\infty}^{\infty} \sup_{u\in[0,\,1]} \|c(u,\, k)\| < \infty$ *and* $\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \sup_{u\in[0,\,1]} |\kappa^{(a_1,a_2,a_3,a_4)}_{V,\lfloor Tu\rfloor}$ $(k,\, j,\, l)| < \infty$ *for all* $a_1, a_2, a_3, a_4 \leq p$. *(ii) For all* $a_1, a_2, a_3, a_4 \leq p$ *there exists a function* $\widetilde{\kappa}_{a_1,a_2,a_3,a_4} : [0,\, 1] \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ *such that* $\sup_{1\leq j\leq m_0+1} \sup_{\lambda^0_{j-1} < u \leq \lambda^0_j} |\kappa^{(a_1,a_2,a_3,a_4)}_{V,\lfloor Tu\rfloor}(k,\, s,\, l) - \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\, k,\, s,\, l)| \leq LT^{-1}$ *for some constant* L*; the function* $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\, k,\, s,\, l)$ *is twice differentiable in* u *at all* $u \neq \lambda^0_j$ $(j = 1, \ldots, m_0 + 1)$ *with bounded derivatives* $(\partial/\partial u)\, \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, \cdot, \cdot, \cdot)$ *and* $(\partial^2/\partial u^2)\, \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, \cdot, \cdot, \cdot)$*, and twice left-differentiable in* u *with bounded derivatives* $(\partial/\partial_- u)\, \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, \cdot, \cdot, \cdot)$ *and* $(\partial^2/\partial_- u^2)\, \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, \cdot, \cdot, \cdot)$*, and piecewise Lipschitz continuous derivative* $(\partial^2/\partial_- u^2)\, \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u, \cdot, \cdot, \cdot)$.

If $\{V_t\}$ is stationary then the cumulant condition of Assumption 2.2-(i) reduces to the standard one used in the time series literature [see Andrews (1991)]. Note that α-mixing and some moment conditions imply that the cumulant condition of Assumption 2.2 holds. Part (ii) extends the smoothness conditions on $A(u,\, \omega)$ in Assumption 2.1 to the fourth-order cumulant. These smoothness conditions are not particularly restrictive.

Consider the following time-varying AR(1) process with one break at mid-sample $\lambda^0_1 = 0.5$,

$$V_{t,T} = \rho(t/T)\, V_{t-1,T} + \sigma(t/T)\, u_t,$$

$$\rho(u) = \begin{cases} \rho_1(u), & u \leq 0.5 \\ \rho_2(u), & u > 0.5 \end{cases},$$

where $\rho(u)$ and $\sigma(u)$ are piecewise Lipschitz continuous and $\{u_t\}$ are i.i.d. random variables with mean zero and unit variance. Then, $V_{t,T}$ is a SLS process with $A(u, \omega) = \sigma(u)(1 + \rho(u)\exp(i\omega))$. If $\rho(u)$ and $\sigma(u)$ satisfy the same smoothness conditions in u required for $A(u, \omega)$ in Assumption 2.1, and $\sup_{u\in[0,1]} |\rho(u)| < 1$ and $\sup_{u\in[0,1]} \sigma(u) < \infty$, then $V_{t,T}$ fulfills Assumption 2.1-2.2.

3 Theoretical Results on Low Frequency Contamination

In this section we establish theoretical results about the low frequency contamination induced by nonstationarity, misspecification and outliers. We first consider the asymptotic proprieties of two key quantities for inference in time series contexts, i.e., the sample autocovariance and the periodogram. These are defined, respectively, by

$$\widehat{\Gamma}(k) = T^{-1} \sum_{t=|k|+1}^{T} \left(V_t - \overline{V}\right)\left(V_{t-|k|} - \overline{V}\right), \tag{3.1}$$

where $\overline{V}$ is the sample mean and

$$I_T(\omega) = \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \exp(-i\omega t) V_t \right|^2, \qquad \omega \in [0, \pi],$$

which is evaluated at the Fourier frequencies $\omega_j = (2\pi j)/T \in [0, \pi]$. In the context of autocorrelated data, hypotheses testing and construction of confidence intervals require estimation of the so-called long-run variance. Traditional HAC estimators are weighted sums of sample autocovariances while frequency domain estimators are weighted sums of the periodograms. Casini (2022c) considered an alternative estimate for the sample autocovariance to be used in the DK-HAC estimators,[2] namely,

$$\widehat{\Gamma}_{\mathrm{DK}}(k) \triangleq \frac{n_T}{T} \sum_{r=1}^{\lfloor T/n_T \rfloor} \widehat{c}_T(rn_T/T, k),$$

[2]The DK-HAC estimators are defined in Section 5.1.

where $k \in \mathbb{Z}$, $n_T \to \infty$ satisfying the conditions given below, and

$$\widehat{c}_T\left(rn_T/T,\, k\right) = n_{2,T}^{-1} \sum_{s=0}^{n_{2,T}-1} \left(V_{rn_T+\lfloor |k/2| \rfloor - n_{2,T}/2+s+1} - \overline{V}_{rn_T,T}\right)\left(V_{rn_T-\lfloor |k/2| \rfloor - n_{2,T}/2+s+1} - \overline{V}_{rn_T,T}\right), \tag{3.2}$$

with $\overline{V}_{rn_T,T} = n_{2,T}^{-1} \sum_{s=0}^{n_{2,T}-1} V_{rn_T - n_{2,T}/2+s+1}$ and $n_{2,T} \to \infty$ such that $n_{2,T}/T \to 0$. For notational simplicity we assume that n_T and $n_{2,T}$ are even. $\widehat{c}_T\left(rn_T/T,\, k\right)$ is an estimate of the autocovariance at time rn_T and lag k, i.e., $\mathrm{cov}(V_{rn_T},\, V_{rn_T-k})$. One could use a smoothed or tapered version; the estimate $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ is an integrated local sample autocovariance. It extends $\widehat{\Gamma}\left(k\right)$ to better account for nonstationarity. Similarly, the DK-HAC estimator does not relate to the periodogram but to the local periodogram defined by

$$I_{\mathrm{L},T}\left(u,\, \omega\right) \triangleq \left| \frac{1}{\sqrt{n_T}} \sum_{s=0}^{n_T-1} V_{\lfloor Tu \rfloor - n_T/2+s+1,T} \exp\left(-i\omega s\right) \right|^2,$$

where $I_{\mathrm{L},T}\left(u,\, \omega\right)$ is the (untapered) periodogram over a segment of length n_T with midpoint $\lfloor Tu \rfloor$. We also consider the statistical properties of both $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ and $I_{\mathrm{L},T}\left(u,\, \omega\right)$ under nonstationarity. Define $r_j = (\lambda_j^0 - \lambda_{j-1}^0)$ for $j = 1, \ldots,\, m_0+1$ with $\lambda_0^0 = 0$ and $\lambda_{m_0+1}^0 = 1$. Note that $\lambda_j^0 = \sum_{s=0}^{j} r_s$.

The low frequency bias is generated by breaks in the mean function. For the sample autocovariance, the bias factor is given by $d^* = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\overline{\mu}_{j_2} - \overline{\mu}_{j_1})^2$ where

$$\overline{\mu}_j = r_j^{-1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \mu_j\left(u\right) du, \qquad \text{for } j = 1, \ldots,\, m_0+1,$$

with $\mu_j\left(\cdot\right)$ defined in (2.1) and we use $\sum_{j_1 \neq j_2}$ as a shorthand for $\sum_{\{j_1,\, j_2=1,\ldots,\, m_0+1,\, j_1\neq j_2\}}$. When the mean is constant in each regime $\mu_j\left(t/T\right) = \mu_j$. Then, $\overline{\mu}_j = \mu_j$ and $d^* = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\mu_{j_2} - \mu_{j_1})^2$. If the mean is constant across regimes, then there is no low frequency bias and $d^* = 0$.

In Section 3.1 we generalize the results in the literature on low frequency contamination for the sample autocovariance and the periodogram. In Section 3.2 we show that the local sample autocovariance and the local periodogram are in general robust to low frequency contamination.

3.1 The Sample Autocovariance and the Periodogram Under Nonstationarity

Mikosch and Stărica (2004) established some results on the low frequency bias for the sample autocovariance and periodogram under the assumption that V_t is stationary in each regime and that the regimes are independent. In Section S.A in the supplement we extend these results by allowing time-varying mean and autocovariace function in each regime and weak dependence across regimes. Here we present a brief summary of these results. Theorem S.A.1 shows that for $\{V_{t,T}\}$ that satisfies Definition 2.1 and Assumption 2.1-2.2, we have

$$\widehat{\Gamma}\left(k\right) \geq \int_{0}^{1} c\left(u,\, k\right) du + d^{*} + o_{\mathrm{a.s.}}\left(1\right), \tag{3.3}$$

and as $k \to \infty$, $\widehat{\Gamma}\left(k\right) \geq d^{*}$ $\mathbb{P}$-a.s. The result suggests that $\widehat{\Gamma}\left(k\right)$ is asymptotically the sum of two terms. The first is the true autocovariance of $\{V_t\}$ at lag k. The second, d^{*}, is always positive and increases with the difference in the mean across regimes. Thus, the time-varying mean induces a positive bias. In Section 5 we shall discuss cases in which this bias arises as a finite-sample problem and cases where the bias remains even asymptotically. The result that $\widehat{\Gamma}\left(k\right) \geq d^{*}$ $\mathbb{P}$-a.s. as $k \to \infty$ implies that unaccounted nonstationarity generates long memory effects. The intuition is straightforward. A long memory SLS process satisfies $\sum_{k=-\infty}^{\infty} |\Gamma\left(u,\, k\right)| \to \infty$ for some $u \in (0,\, 1)$, similar to a stationary long memory process.[3] The theorem shows that $\widehat{\Gamma}\left(k\right)$ exhibits a similar property and $\widehat{\Gamma}\left(k\right)$ decays more slowly than for a short memory stationary process for small lags and approaches a positive constant d^{*} for large lags.

Theorem S.A.2 in the supplement analyzes the properties of the periodogram $I_T\left(\omega_l\right)$ as $\omega \to 0$ when the mean is time-varying. The result states that as $\omega \to 0$ $\mathbb{E}\left(I_T\left(\omega\right)\right)$ generally takes unbounded values except for some ω for which $\mathbb{E}\left(I_T\left(\omega\right)\right)$ is bounded below by $2\pi \int_0^1 f\left(u,\, \omega\right) du > 0$. A SLS process with long memory has an unbounded local spectral density $f\left(u,\, \omega\right)$ as $\omega \to 0$ for some $u \in [0,\, 1]$. Since $f\left(\cdot,\, \cdot\right)$ cannot be negative, it follows that $\int_0^1 f\left(u,\, \omega\right) du$ is also unbounded as $\omega \to 0$. Theorem S.A.2 suggests that nonstationarity consisting of time-varying first moment results in a periodogram sharing features of a long memory series.

This discussion suggests that certain deviations from stationarity can generate a long memory component that leads to overestimation of the true autocovariance. It follows that the LRV is also overestimated. Since the LRV is used to normalize test statistics, this has important consequences for many HAR inference tests characterized by deviations from stationarity under the alternative hypothesis. These include tests for forecast evaluation, tests and inference for structural change models, time-varying parameters models and regime-switching models. In the linear regression model, V_t corresponds to the regressors multiplied by the fitted residuals. Unaccounted

[3] In Section S.A.1 in the supplement we define long memory SLS processes that are characterized by the property $\sum_{k=-\infty}^{\infty} |\rho_V\left(u,\, k\right)| = \infty$ for some $u \in [0,\, 1]$ where $\rho_V\left(u,\, k\right) \triangleq \mathrm{Corr}(V_{\lfloor Tu \rfloor},\, V_{\lfloor Tu \rfloor + k})$ and $\vartheta\left(u\right) \in (0,\, 1/2)$ is the long memory parameter at time u.

nonlinearities and outliers can contaminate the mean of V_t and therefore contribute to d^*.

3.2 The Sample Local Autocovariance and Local Periodogram Under Nonstationarity

We now consider the behavior of $\widehat{c}_T(rn_T/T,\,k)$ defined in (3.2) for fixed k as well as for $k\to\infty$. For notational simplicity we assume that k is even. For $u\in(0,\,1)$ define $\mathbf{S}(u,\,k,\,n_{2,T})=\{\lfloor Tu\rfloor+k/2-n_{2,T}/2+1,\ldots,\,\lfloor Tu\rfloor+k/2+n_{2,T}/2\}$, $n_{j,L}(u,\,k,\,n_{2,T})=(T_j^0-(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1))$, and $n_{j,R}(u,\,k,\,n_{2,T})=((\lfloor Tu\rfloor+k/2+n_{2,T}/2+1)-T_j^0)$. $\mathbf{S}(u,\,k,\,n_{2,T})$ denotes a window of length $n_{2,T}$ around $\lfloor Tu\rfloor$, $n_{j,L}(u,\,k,\,n_{2,T})$ (resp. $n_{j,R}(u,\,k,\,n_{2,T})$) denotes the distance between the left (resp. right) end point of $\mathbf{S}(u,\,k,\,n_{2,T})$ and T_j^0.

Theorem 3.1. *Assume that $\{V_{t,T}\}$ satisfies Definition 2.1, $n_T,\,n_{2,T}\to\infty$ with $n_T/T\to0$, $n_{2,T}/T\to0$ and $n_T/n_{2,T}\to0$. Under Assumption 2.1-2.2,*

(i) for $u\in(0,\,1)$ such that $T_j^0\notin\mathbf{S}(u,\,k,\,n_{2,T})$ for all $j=1,\ldots,\,m_0$, $\widehat{c}_T(u,\,k)=c(u,\,k)+o_{\mathbb{P}}(1)$;

(ii) for $u\in(0,\,1)$ such that $T_j^0\in\mathbf{S}(u,\,k,\,n_{2,T})$ for some $j=1,\ldots,\,m_0$, we have two sub-cases: (a) if $n_{j,L}(u,\,k,\,n_{2,T})/n_{2,T}\to\gamma$ or $n_{j,R}(u,\,k,\,n_{2,T})/n_{2,T}\to\gamma$ with $\gamma\in(0,\,1)$, then

$$\widehat{c}_T(u,\,k)\geq\gamma c\left(\lambda_j^0,\,k\right)+(1-\gamma)\,c(u,\,k)+\gamma(1-\gamma)\left(\mu_j\left(\lambda_j^0\right)-\mu_{j+1}(u)\right)^2+o_{\mathbb{P}}(1)\,.$$

(b) if $n_{j,L}(u,\,k,\,n_{2,T})/n_{2,T}\to0$ or $n_{j,R}(u,\,k,\,n_{2,T})/n_{2,T}\to0$, then $\widehat{c}_T(u,\,k)=c(u,\,k)+o_{\mathbb{P}}(1)$.

Further, if there exists an $r=1,\ldots,\,\lfloor T/n_T\rfloor$ such that there exists a $j=1,\ldots,\,m_0$ with $T_j^0\in\mathbf{S}(rn_T,\,k,\,n_{2,T})$ satisfying (ii-a), then, as $k\to\infty$, $\widehat{\Gamma}_{\mathrm{DK}}(k)\geq d_T^*$ $\mathbb{P}$-a.s., where $d_T^*=(n_{2,T}/T)\,\gamma(1-\gamma)\,(\mu_j(\lambda_j^0)-\mu_{j+1}(u))^2>0$ and $d_T^*\to0$ as $T\to\infty$.

The theorem shows that the behavior of $\widehat{c}_T(u,\,k)$ depends on whether a change in mean is present, and if so whether it is close enough to $\lfloor Tu\rfloor$. For a given $u\in(0,\,1)$ and $k\in\mathbb{Z}$, if the condition of part (i) of the theorem holds, then $\widehat{c}_T(u,\,k)$ is consistent for $\mathrm{cov}(V_{\lfloor Tu\rfloor}V_{\lfloor Tu\rfloor-k})=c(u,\,k)+O(T^{-1})$ [see Casini (2022c)]. If a change-point falls close to either boundary of the window $\mathbf{S}(u,\,k,\,n_{2,T})$, as specified in case (ii-b), then $\widehat{c}_T(u,\,k)$ remains consistent. The only case in which a non-negligible bias arises is when the change-point falls in a neighborhood around $\lfloor Tu\rfloor$ sufficiently far from either boundary. This represents case (ii-a), for which a biased estimate results. However, the bias vanishes asymptotically. Since $\widehat{\Gamma}_{\mathrm{DK}}(k)$ is an average of $\widehat{c}_T(rn_T,\,k)$ over blocks $r=1,\ldots,\,\lfloor T/n_T\rfloor$, if case (ii-a) holds then $\widehat{\Gamma}_{\mathrm{DK}}(k)\geq d_T^*$ as $k\to\infty$ but $d_T^*\to0$ as $T\to\infty$. Thus, comparing this result with the discussion above on $\widehat{\Gamma}(k)$ (see also Theorem S.A.1), in practice the long memory effects are unlikely to occur when using $\widehat{\Gamma}_{\mathrm{DK}}(k)$. Furthermore, one can reduce this problem by appropriately choosing the blocks $r=1,\ldots,\lfloor T/n_T\rfloor$. A procedure was proposed in Casini (2022c) using the methods developed in Casini and Perron (2021a).

We now study the asymptotic properties of $I_{\mathrm{L},T}(u,\,\omega)$ as $\omega\to 0$ for $u\in[0,\,1]$. We consider the Fourier frequencies $\omega_l=2\pi l/n_T\in(-\pi,\,\pi)$ for an integer $l\neq 0$ (mod n_T). We need the following high-level conditions. Part (i) corresponds to Assumption S.A.1 while part (ii) requires additional smoothness.

Assumption 3.1. *(i) For each ω_l and $u\in[0,\,1]$ with $T_j^0\in\mathbf{S}(u,\,0,\,n_T)$ there exist $B_j\in\mathbb{R}$, $j=1,\ldots,m_0$ with $B_{j_1}\neq B_{j_2}$ for $j_1\neq j_2$ such that*

$$\left|\sum_{s=0}^{n_T-1}\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\exp\left(-i\omega_l s\right)\right|^2\geq$$
$$\left|B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)+B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\right|^2.$$

(ii) $\sup_{u\in[0,\,1],\,u\neq\lambda_0^j,\,j=1,\ldots,m_0}\left(\partial^2/\partial u^2\right)f(u,\,\omega)$ is continuous in ω.

Theorem 3.2. *Assume that $\{V_{t,T}\}$ satisfies Definition 2.1 and that $n_T\to\infty$ with $n_T/T\to 0$. Under Assumption 2.1-2.2, S.A.1-(ii) and 3.1,*

(i) for any $u\in(0,\,1)$ such that $T_j^0\notin\mathbf{S}(u,\,0,\,n_T)$ for all $j=1,\ldots,m_0$, $\mathbb{E}\left(I_{\mathrm{L},T}(u,\,\omega_l)\right)\geq f(u,\,\omega_l)$ as $\omega_l\to 0$;

(ii) for any $u\in(0,\,1)$ such that $T_j^0\in\mathbf{S}(u,\,0,\,n_T)$ for some $j=1,\ldots,m_0$ we have two subcases: (a) if $n_{j,L}(u,\,0,\,n_T)/n_T\to\gamma$ or $n_{j,R}(u,\,0,\,n_T)/n_T\to\gamma$ with $\gamma\in(0,\,1)$, and $n_T\omega_l^2\to 0$ as $T\to\infty$, then $\mathbb{E}\left(I_{\mathrm{L},T}(u,\,\omega)\right)\to\infty$ for many values in the sequence $\{\omega_l\}$ as $\omega_l\to 0$; (b) if $n_{j,L}(u,\,0,\,n_T)/n_T\to 0$ or $n_{j,R}(u,\,0,\,n_T)/n_T\to 0$, then $\mathbb{E}\left(I_{\mathrm{L},T}(u,\,\omega_l)\right)\geq f(u,\,\omega_l)$ as $\omega_l\to 0$.

It is useful to compare Theorem 3.2 with the discussion above about the periodogram (see also Theorem S.A.2). Unlike the periodogram, the asymptotic behavior of the local periodogram as $\omega_l\to 0$ depends on the vicinity of u to λ_j^0 ($j=1,\ldots,m_0$). Since $I_{\mathrm{L},T}(u,\,\omega_l)$ uses observations in the window $\mathbf{S}(u,\,0,\,n_T)$, if no discontinuity in the mean occurs in this window then $I_{\mathrm{L},T}(u,\,\omega_l)$ is asymptotically unbiased for the spectral density $f(u,\,\omega_l)$. More complex is its behavior if some T_j^0 falls in $\mathbf{S}(u,\,0,\,n_T)$. The theorem shows that if T_j^0 is close to the boundary, as indicated in case (ii-b), then $I_{\mathrm{L},T}(u,\,\omega_l)$ is bounded below by $f(u,\,\omega_l)$, similarly to case (i). If instead T_j^0 falls sufficiently close to the mid-point $\lfloor Tu \rfloor$, as indicated in case (ii-a), then $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\,\omega\right)\right) \to \infty$ for many values in the sequence $\{\omega_l\}$ as $\omega_l \to 0$ provided it satisfies $n_T\omega_l^2 \to 0$ as $T \to \infty$. Hence, unless $T\lambda_j^0$ is close to $\lfloor Tu \rfloor$, the local periodogram $I_{\mathrm{L},T}\left(u,\,\omega_l\right)$ behaves very differently from the periodogram $I_T\left(\omega_l\right)$. Accordingly, nonstationarity is unlikely to generate long memory effects if one uses the local periodogram. As for $\widehat{c}_T\left(u,\,k\right)$, if one uses preliminary inference procedures [cf. Casini and Perron (2021a)] for the detection and estimation of the discontinuities in the spectrum and for the estimation of their locations, then one can construct the window efficiently and avoid T_j^0 being too close to $\lfloor Tu \rfloor$.

4 Edgeworth Expansions for HAR Tests Under Nonstationarity

We now consider Edgeworth expansions for the distribution of the t-statistic in the location model based on the HAC and DK-HAC estimator where $\{V_t\}$ is assumed to have zero-mean and time-varying second moments. This is useful for analyzing the theoretical properties of the null rejection probabilities of the HAR tests under nonstationarity. As in the literature, we make use of the Gaussianity assumption for mathematical convenience.[4] We relax the stationarity assumption used in the literature [cf. Jansson (2004), Sun et al. (2008) and Velasco and Robinson (2001)] which has important consequences for the nature of the results. The results concerning the t-test based on the HAC estimator are presented in Section 4.1 while those based on the DK-HAC estimator are presented in Section 4.2.

Let $\{V_t\}$ be a zero-mean Gaussian SLS process satisfying Assumption 2.1-(i-iv). Let

$$h_1 \triangleq \frac{\sqrt{T}\,\overline{V}}{\sqrt{J_T}} \sim \mathscr{N}\left(0,\,1\right), \tag{4.1}$$

which is valid for all T such that $J_T > 0$ where $J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(V_sV_t)$.

4.1 HAC-based HAR Tests

The classical HAC estimator is defined as

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

[4] This can be relaxed by considering distributions with Gram-Charlier representations at the expenses of more complex derivations.

where $K_1(\cdot)$ is a kernel and $b_{1,T}$ a bandwidth parameter. Under appropriate conditions on $b_{1,T}$, we have $\widehat{J}_{\mathrm{HAC},T} - J_T \xrightarrow{\mathbb{P}} 0$ from which it follows that

$$Z_T \triangleq \frac{\sqrt{T}\,\overline{V}}{\sqrt{\widehat{J}_{\mathrm{HAC},T}}} \xrightarrow{d} \mathscr{N}(0,\,1).$$

Let $\mathbf{V} = (V_1, \ldots, V_T)'$. Note that $\widehat{J}_{\mathrm{HAC},T} = \mathbf{V}' W_{b_1} \mathbf{V}/T$ where W_{b_1} has $(r,\,s)$th element

$$W_{b_1}^{(r,s)} = w\left(b_{1,T}(r-s)\right) = \int_{\Pi} \widetilde{K}_{b_1}(\omega)\, e^{i(r-s)\omega} d\omega, \tag{4.2}$$

such that $\widetilde{K}_{b_1}(\omega)$ is a kernel with smoothing number $b_{1,T}^{-1}$ and $\Pi = (-\pi,\,\pi]$. For an even function K that integrates to one, we define

$$\widetilde{K}_{b_1}(\omega) = b_{1,T}^{-1} \sum_{j=-\infty}^{\infty} K\left(b_{1,T}^{-1}(\omega + 2\pi j)\right).$$

Note that $\widetilde{K}_{b_1}(\omega)$ is periodic of period 2π, even and satisfies $\int_{-\pi}^{\pi} \widetilde{K}_{b_1}(\omega)\, d\omega = 1$. It follows that $w(r) = \int_{-\infty}^{\infty} e^{irx} K(x)\, dx$ and $\widehat{J}_{\mathrm{HAC},T} = 2\pi \int_{\Pi} \widetilde{K}_{b_1}(\omega)\, I_T(\omega)\, d\omega$. $\widetilde{K}_{b_1}(\omega)$ is the so-called spectral window generator. We refer to Brillinger (1975) for a review of these introductory concepts.

We now analyze the joint distribution of $\overline{V}$ and $\widehat{J}_{\mathrm{HAC},T}$. Let $\mathsf{B}_T = \mathbb{E}(\widehat{J}_{\mathrm{HAC},T})/J_T - 1$ and $\mathsf{V}_T^2 = \mathrm{Var}(\sqrt{Tb_{1,T}}\widehat{J}_{\mathrm{HAC},T}/J_T)$ denote the relative bias and variance, respectively, of $\widehat{J}_{\mathrm{HAC},T}$. It is convenient to work with standardized statistics with zero mean and unit variance. Write

$$Z_T = Z_T(\mathbf{h}) = h_1 \left(1 + \mathsf{B}_T + \mathsf{V}_T h_2 \left(Tb_{1,T}\right)^{-1/2}\right)^{-1/2}, \qquad h_2 = \sqrt{Tb_{1,T}} \left(\frac{\widehat{J}_{\mathrm{HAC},T} - \mathbb{E}\left(\widehat{J}_{\mathrm{HAC},T}\right)}{J_T \mathsf{V}_T} \right),$$

where $\mathbf{h} = (h_1,\, h_2)'$. Note that $h_2 = \mathbf{V}' Q_T \mathbf{V} - \mathbb{E}\left(\mathbf{V}' Q_T \mathbf{V}\right)$ is a centered quadratic form in a Gaussian vector where $Q_T = W_{b_1}(\sqrt{T/b_{1,T}}\mathsf{V}_T J_T)^{-1}$. The joint characteristic function of $\mathbf{h}$ is

$$\psi_T(\mathbf{t}) = \psi_T(t_1,\, t_2) = \left|I - 2it_2 \Sigma_V Q_T\right|^{-1/2} \exp\left(-2^{-1} t_1^2 \xi_T' \left(I - 2it_2 \Sigma_V Q_T\right)^{-1} \Sigma_V \xi_T - it_2 \Upsilon_T\right),$$

where $\Upsilon_T = \mathbb{E}\left(\mathbf{V}' Q_T \mathbf{V}\right) = \mathrm{Tr}\left(\Sigma_V Q_T\right)$, $\Sigma_V = \mathbb{E}\left(\mathbf{V}\mathbf{V}'\right)$, and $\xi_T = \mathbf{1}/\sqrt{TJ_T}$ with $\mathbf{1}$ being the $T \times 1$ vector $(1, 1, \ldots, 1)'$. The cumulant generating function of $\mathbf{h}$ is

$$\mathrm{K}_T(t_1,\, t_2) = \log \psi_T(t_1,\, t_2) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \kappa_T(r,\, s) \frac{(it_1)^r}{r!} \frac{(it_2)^r}{s!},$$

where $\kappa_T(r, s)$ is the cumulant of $\mathbf{h}$. Phillips (1980) considered the distribution of linear and quadratic forms under Gaussianity. From his derivations, the nonzero bivariate cumulants are

$$\kappa_T(0, s) = 2^{s-1}(s-1)!\mathrm{Tr}\left((\Sigma_V Q_T)^s\right), \qquad s > 1,$$
$$\kappa_T(2, s) = 2^s s! \xi_T' (\Sigma_V Q_T)^s \Sigma_V \xi_T, \qquad s > 0.$$

We introduce the following assumptions about $\{V_t\}$ and $f(u, 0)$.

Assumption 4.1. *For all $u \in [0, 1]$, $0 < f(u, 0) < \infty$ and $f(u, \omega)$ has d_f continuous derivatives $(d_f \geq 2)$ $f^{(d_f)}(u, \omega)$ in a neighborhood of $\omega = 0$ and the d_fth derivative satisfies a Lipschitz condition of order ϱ with $\varrho \in (0, 1]$.*

Assumption 4.2. *For all u, $f(u, \omega) \in L_p$ for some $p > 1$, i.e., $\|f(u, \cdot)\|_p^p = \int_\Pi f^p(u, \omega)\, d\omega < \infty$.*

Assumption 4.3. *$|K(x)| < \infty$, $K(x) = K(-x)$, $K(x) = 0$ for $x \notin \Pi$ and $\int_\Pi K(x)\, dx = 1$.*

Assumption 4.4. *$K(x)$ satisfies a uniform Lipschitz condition of order 1 in $[-\pi, \pi]$.*

Assumption 4.5. *For $j = 0, 1, \ldots, d_f$, $d_f \geq 2$ and $r = 1, 2, \ldots$*

$$\mu_j(K^r) \triangleq \int_\Pi x^j (K(x))^r\, dx = \begin{cases} = 0, & j < d_f,\ r = 1; \\ \neq 0, & j = d_f,\ r = 1. \end{cases}$$

Assumption 4.6. *$b_{1,T} + (Tb_{1,T})^{-1} \to 0$ as $T \to \infty$.*

Assumption 4.7. *$b_{1,T} = CT^{-q}$ where $0 < q < 1$ and $0 < C < \infty$.*

Assumption 4.3-4.7 about the kernel and bandwidth are the same as in Velasco and Robinson (2001) in which a discussion can be found. They are satisfied by most kernels used in practice. The bandwidth condition in Assumption 4.6 is sufficient for the consistency of $\widehat{J}_{\mathrm{HAC},T}$ and is strengthened in Assumption 4.7, for some parts of the proofs, which is satisfied by popular MSE-optimal bandwidths [cf. Andrews (1991), Casini (2022a), Belotti et al. (2023) and Whilelm (2015)].

Assumption 4.1-4.2 impose conditions on the smoothness and boundedness of the spectral density. Assumption 4.1 is implied by $\sum_{k=-\infty}^{\infty} |k|^{d_f+\varrho} \sup_t |\mathbb{E}V_t V_{t-k}| < \infty$ but it is stronger than necessary because it extends the smoothness restriction to all frequencies. Assumption 4.2 does impose some restrictions on $f(u, \cdot)$ beyond the origin, though it is not particularly restrictive since any $p > 1$ arbitrarily close to 1 will suffice.

We now analyze the asymptotic distribution of $\widehat{J}_{\mathrm{HAC},T}$. Under stationarity this was discussed by Bentkus and Rudzkis (1982) and Velasco and Robinson (2001). From Lemma S.B.11-S.B.12 in the supplement we yield

$$\mathsf{B}_T = \bar{c}_1 b_{1,T}^{d_f} + O\left(b_{1,T}^{d_f+\varrho} + T^{-1}\log T\right), \qquad \text{where} \qquad \bar{c}_1 = \frac{\mu_{d_f}(K)\int_0^1 f^{(d_f)}(u,0)\,du}{d_f!\int_0^1 f(u,0)\,du}.$$

We present a second-order Edgeworth expansion to approximate the distribution of $\mathbf{h}$, with error $o((Tb_{1,T})^{-1/2})$ and including terms up to order $(Tb_{1,T})^{-1/2}$ to correct the asymptotic normal distribution. This will imply the validity of that expansion for the distribution of $\widehat{J}_{\mathrm{HAC},T}$. For $\mathbf{B} \in \mathscr{B}^2$, where $\mathscr{B}^2$ is any class of Borel sets in $\mathbb{R}^2$, let $\mathbb{Q}_T^{(2)}(\mathbf{B}) = \int_{\mathbf{B}} \varphi_2(\mathbf{h})\, q_T^{(2)}(\mathbf{h})\, d\mathbf{h}$, where $\varphi_2(\mathbf{h}) = (2\pi)^{-1}\exp\{-(1/2)\|\mathbf{h}\|^2\}$ is the density of the bivariate standard normal distribution,

$$q_T^{(2)}(\mathbf{h}) = 1 + (1/3!)\,(Tb_{1,T})^{-1/2}\left(\Xi_0(0,\,3)\mathcal{H}_3(h_2) + \Xi_0(2,\,1)\mathcal{H}_2(h_1)\,\mathcal{H}_1(h_2)\right),$$

where $\mathcal{H}_j(\cdot)$ are the univariate Hermite polynomials of order j, and $\Xi_0(0,\,3) = (4\pi)^{1/2}\,3!\int_{\Pi} K^3(\omega)\,d\omega\,\|K\|_2^{-3}$ and $\Xi_0(2,\,1) = (4\pi)^{1/2} K(0)\,\|K\|_2^{-1}$ (see Lemma S.B.3-S.B.4). Let $(\partial\mathbf{B})^{\phi}$denote a neighborhood of radius ϕ of the boundary of a set $\mathbf{B}$. Let $\mathbb{P}_T$ denote the probability measure of $\mathbf{h}$.

Theorem 4.1. *Let Assumption 4.1, 4.2* $(p > 1)$*, 4.3-4.4 and 4.7* $(0 < q < 1)$ *hold. For* $\phi_T = (Tb_{1,T})^{-\varpi}$ *with* $1/2 < \varpi < 1$*, we have*

$$\sup_{\mathbf{B}\in\mathscr{B}^2}\left|\mathbb{P}_T(\mathbf{B}) - \mathbb{Q}_T^{(2)}(\mathbf{B})\right| = o\left((Tb_{1,T})^{-1/2}\right) + (4/3)\sup_{\mathbf{B}\in\mathscr{B}^2}\mathbb{Q}_T^{(2)}\left((\partial\mathbf{B})^{2\phi_T}\right). \tag{4.3}$$

Theorem 4.1 shows that $\mathbb{Q}_T^{(2)}$ is a valid second-order Edgeworth expansion for the measure $\mathbb{P}_T$. The method of proof is the same as in Velasco and Robinson (2001). We first approximate the true characteristic function and then apply a smoothing lemma [cf. Lemma S.B.2 in the supplement which is from Bhattacharya and Rao (1975)]. The leading term of the approximation error is of order $o((Tb_{1,T})^{-1/2})$ as the second term on the right hand side of (4.3) is negligible if $\mathbf{B}$ is convex because ϕ_T decreases as a power of T. This is the same order obtained for the corresponding leading term under stationarity. Since the higher-order correction terms in $q_T^{(2)}$ depend only on $K(\cdot)$ but not on $f(\cdot,\cdot)$, they are equal to the one obtained under stationarity.

Next, we focus on Z_T, i.e., a t-statistic for the mean. Proceeding as in Velasco and Robinson (2001), we first derive a linear stochastic approximation to $Z_T(\mathbf{h})$ and show that its distribution is the same as that of Z_T up to order $o((Tb_{1,T})^{-1/2})$. Then, we show that the asymptotic approximation for the distribution of the linear stochastic approximation is valid also for Z_T with the same error $o((Tb_{1,T})^{-1/2})$. Using Lemma S.B.12-S.B.13 in the supplement we can substitute out B_T and V_T in Z_T and, by only focusing on the leading terms, we define the following linear stochastic approximation,

$$\widetilde{Z}_T \triangleq h_1 \left(1 - 2^{-1}\overline{c}_1 b_{1,T}^{d_f} - 2^{-1}\sqrt{4\pi}\left\|K_2\right\| h_2 \left(Tb_{1,T}\right)^{-1/2}\right).$$

The next theorem presents a valid Edgeworth expansion for the distribution of $\widetilde{Z}_T$ from that of $\mathbf{h}$.

Theorem 4.2. *Let Assumption 4.1, 4.2 ($p > 1$), 4.3-4.5 and 4.7 ($q = 1/\left(1 + 2d_f\right)$) hold. For a convex Borel set $\mathbf{C}$, we have, for $r_2\left(x\right) = -\overline{c}_1\left(x^2 - 1\right)/2$,*

$$\sup_{\mathbf{C}} \left|\mathbb{P}\left(Z_T \in \mathbf{C}\right) - \int_{\mathbf{C}} \varphi\left(x\right)\left(1 + r_2\left(x\right) b_{1,T}^{d_f}\right) dx\right| = o\left(\left(Tb_{1,T}\right)^{-1/2}\right). \tag{4.4}$$

Theorem 4.2 shows the form of the correction term to the standard normal distribution, i.e., $b_{1,T}^{d_f}\int_{\mathbf{C}}\varphi\left(x\right) r_2\left(x\right) dx$. The error of the approximation is of order $o((Tb_{1,T})^{-1/2})$ which is the same as the one obtained under stationarity by Velasco and Robinson (2001).

Let $\Phi\left(\cdot\right)$ denote the distribution function of the standard normal. Setting $\mathbf{C} = (-\infty, z]$, and integrating and Taylor expanding $\Phi\left(\cdot\right)$, we obtain, uniformly in z,

$$\begin{aligned}\mathbb{P}\left(Z_T \leq z\right) &= \Phi\left(z\right) + \frac{1}{2}\overline{c}_1 z\varphi\left(z\right) b_{1,T}^{d_f} + o\left(\left(Tb_{1,T}\right)^{-1/2}\right) \\ &= \Phi\left(z\left(1 + \frac{1}{2}\overline{c}_1 b_{1,T}^{d_f}\right)\right) + o\left(\left(Tb_{1,T}\right)^{-1/2}\right) = \Phi\left(z\right) + O\left(\left(Tb_{1,T}\right)^{-1/2}\right).\end{aligned} \tag{4.5}$$

This shows that under the conditions of Theorem 4.2, the standard normal approximation is correct up to order $O((Tb_{1,T})^{-1/2})$. Consider the location model $y_t = \beta + V_t$ $\left(t = 1, \ldots, T\right)$. For the null hypothesis $\mathbb{H}_0 :\ \beta = \beta_0$, consider the following t-test,

$$t_{\mathrm{HAC}} = \frac{\sqrt{T}\left(\widehat{\beta} - \beta_0\right)}{\sqrt{\widehat{J}_{\mathrm{HAC},T}}},$$

where $\widehat{\beta}$ is the least-squares estimator of β. Theorem 4.2 and (4.5) imply that

$$\mathbb{P}\left(t_{\mathrm{HAC}} \leq z\right) = \Phi\left(z\right) + p\left(z\right)\left(Tb_{1,T}\right)^{-1/2} + o\left(\left(Tb_{1,T}\right)^{-1/2}\right), \tag{4.6}$$

for any $z \in \mathbb{R}$, where $p(z)$ is an odd function. Thus, the error in rejection probability (ERP) of t_{HAC} is of order $O((Tb_{1,T})^{-1/2})$. If $\{V_t\}$ is second-order stationary, the results in Velasco and Robinson (2001) imply that the ERP of t_{HAC} is also of order $O((Tb_{1,T})^{-1/2})$. Below we establish the corresponding ERP when the t-statistic is instead normalized by $\widehat{J}_{\mathrm{DK},T}$ and also discuss the ERP of the t-test under fixed-b asymptotics.

4.2 DK-HAC-based HAR Tests

We now consider the Edgeworth expansion for tests based on the DK-HAC estimator. In order to simplify some parts of the proof here we consider an asymptotically equivalent version of the DK-HAC estimator discussed in Section 5. Let

$$\widehat{J}^{*}_{\mathrm{DK},T}=\sum_{k=-T+1}^{T-1}K_1\left(b_{1,T}k\right)\widehat{\Gamma}^{*}_{\mathrm{DK}}\left(k\right),\qquad \widehat{\Gamma}^{*}_{\mathrm{DK}}\left(k\right)\triangleq\int_0^1\widehat{c}_{\mathrm{DK},T}\left(r,\,k\right)dr,$$

where $b_{1,T}$ is a bandwidth sequence and

$$\widehat{c}_{\mathrm{DK},T}\left(r,\,k\right)=\left(Tb_{2,T}\right)^{-1}\sum_{s=|k|+1}^{T}K_2\left(\frac{\left(Tr-\left(s-|k|/2\right)\right)/T}{b_{2,T}}\right)V_sV_{s-|k|},$$

with K_2 a kernel and $b_{2,T}$ a bandwidth. Note that $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ and $\widehat{\Gamma}^{*}_{\mathrm{DK}}\left(k\right)$ are asymptotically equivalent and $\widehat{c}_T$ is a special case of $\widehat{c}_{\mathrm{DK},T}$ with K_2 being a rectangular kernel and $n_{2,T}=Tb_{2,T}$.

Assumption 4.8. *$K_2(\cdot):\mathbb{R}\to[0,\,\infty]$, $K_2(x)=K_2(1-x)$, $\int_0^1K_2(x)\,dx=1$, $K_2(x)=0$ for $x\notin[0,\,1]$ and $K_2(\cdot)$ is continuous. The bandwidth sequence $\{b_{2,T}\}$ satisfies $b_{2,T}\to0$, $b_{2,T}^2/b_{1,T}^{q_2}\to\bar{b}\in[0,\,\infty)$ and $1/Tb_{1,T}b_{2,T}\to0$ where q_2 is the index of smoothness of $K_1(\cdot)$ at 0.*

Under Assumption 4.3-4.4, 4.6 and 4.8 it holds that $\widehat{J}^{*}_{\mathrm{DK},T}-J_T\xrightarrow{\mathbb{P}}0$ [cf. Casini (2022c)] and

$$U_T\triangleq\frac{\sqrt{T}\,\overline{V}}{\sqrt{\widehat{J}^{*}_{\mathrm{DK},T}}}\xrightarrow{d}\mathscr{N}\left(0,\,1\right).\tag{4.7}$$

Note that $\widehat{J}^{*}_{\mathrm{DK},T}=\int_0^1\widetilde{\mathbf{V}}\left(r\right)'W_{b_1}\widetilde{\mathbf{V}}\left(r\right)dr/(Tb_{2,T})$ where $\widetilde{\mathbf{V}}\left(r\right)=(\widetilde{V}_1\left(r\right),\,\widetilde{V}_2\left(r\right),\ldots,\,\widetilde{V}_T\left(r\right))'$ with $\widetilde{V}_j\left(r\right)=\sqrt{K_2\left(\left(r-j\right)/Tb_{2,T}\right)}V_j$ and W_{b_1} defined in (4.2). Let

$$\widetilde{I}_T\left(r,\,\omega\right)=\frac{1}{2\pi Tb_{2,T}}\left|\sum_{t=1}^{T}\exp\left(-i\omega t\right)\widetilde{V}_t\left(r\right)\right|^2.$$

$\widetilde{I}_T(r,\,\omega)$ is the local periodogram of $\{\widetilde{\mathbf{V}}(r)\}$. Then, $\widehat{J}^*_{\mathrm{DK},T}=2\pi\int_0^1\int_\Pi \widetilde{K}_{b_1}(\omega)\,\widetilde{I}_T(r,\,\omega)\,d\omega dr$.

We begin by analyzing the joint distribution of $\overline{V}$ and $\widehat{J}^*_{\mathrm{DK},T}$. Let $\mathsf{B}_{2,T}=\mathbb{E}(\widehat{J}^*_{\mathrm{DK},T})/J_T-1$ and $\mathsf{V}^2_{2,T}=\mathrm{Var}(\sqrt{Tb_{1,T}b_{2,T}}\widehat{J}^*_{\mathrm{DK},T}/J_T)$ denote the relative bias and variance of $\widehat{J}^*_{\mathrm{DK},T}$, respectively. It is convenient to work with standardized statistics with zero mean and unit variance. Write

$$U_T=U_T(\mathbf{v})=v_1\left(1+\mathsf{B}_{2,T}+\mathsf{V}_{2,T}v_2\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right)^{-1/2},\quad v_2=\sqrt{Tb_{1,T}b_{2,T}}\left(\frac{\widehat{J}^*_{\mathrm{DK},T}-\mathbb{E}\left(\widehat{J}^*_{\mathrm{DK},T}\right)}{J_T\mathsf{V}_{2,T}}\right),$$

where $\mathbf{v}=(v_1,\,v_2)'$ with $v_1=h_1$. Note that $v_2=\int_0^1(\widetilde{\mathbf{V}}(r)'\,Q_{2,T}\widetilde{\mathbf{V}}(r)-\mathbb{E}(\widetilde{\mathbf{V}}(r)'\,Q_{2,T}\widetilde{\mathbf{V}}(r)))dr$ is a centered quadratic form in a Gaussian vector where $Q_{2,T}=W_{b_1}(\sqrt{Tb_{2,T}/b_{1,T}}\mathsf{V}_{2,T}J_T)^{-1}$. The joint characteristic function of $\mathbf{v}$ is

$$\psi_{2,T}(t_1,\,t_2)=\left|I-2it_2\Sigma_{\widetilde{V}}Q_{2,T}\right|^{-1/2}\exp\left\{-2^{-1}t_1^2\xi'_{2,T}\left(I-2it_2\Sigma_{\widetilde{V}}Q_{2,T}\right)^{-1}\Sigma_{\widetilde{V}}\xi_{2,T}-it_2\Upsilon_{2,T}\right\},$$

where $\Upsilon_{2,T}=\mathbb{E}(\int_0^1(\widetilde{\mathbf{V}}(r)'\,Q_{2,T}\widetilde{\mathbf{V}}(r))dr)=\mathrm{Tr}(\Sigma_{\widetilde{V}}Q_{2,T})$, $\Sigma_{\widetilde{V}}=\mathbb{E}(\int_0^1(\widetilde{\mathbf{V}}(r)\,\widetilde{\mathbf{V}}(r)')dr)$ and $\xi_{2,T}=\mathbf{1}/\sqrt{Tb_{2,T}J_T}$. The cumulant generating function of $\mathbf{v}$ is

$$\mathrm{K}_{2,T}(t_1,\,t_2)=\log\psi_{2,T}(t_1,\,t_2)=\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\kappa_{2,T}(r,\,s)\,\frac{(it_1)^r}{r!}\frac{(it_2)^r}{s!},$$

where $\kappa_{2,T}(r,\,s)$ is the cumulant of $\mathbf{v}$. To obtain more precise bounds in some parts of the proofs we use the following assumption on the cross-partial derivatives of $f(u,\,\omega)$. Let $\widetilde{\mathbf{C}}$ denote the set of continuity points of $f(u,\,\omega)$ in u, i.e., $\widetilde{\mathbf{C}}=\{[0,\,1]/\{\lambda^0_j,\,j=1,\ldots,\,m_0\}\}$.

Define

$$\Delta_f(\omega)=\sum_{j=1}^{m_0}\int_0^1\left(\frac{\partial}{\partial u_-}f\left(\lambda^0_j,\,\omega\right)\int_0^{1-s}xK_2(x)\,dx+\frac{\partial}{\partial u_+}f\left(\lambda^0_j,\,\omega\right)\int_{1-s}^1xK_2(x)\,dx\right)ds,$$

where

$$\frac{\partial}{\partial u_-}f\left(\lambda^0_j,\,\omega\right)=\lim_{h\uparrow 0}\frac{f\left(\lambda^0_j+h,\,\omega\right)-f\left(\lambda^0_j,\,\omega\right)}{h},\qquad\frac{\partial}{\partial u_+}f\left(\lambda^0_j,\,\omega\right)=\lim_{h\downarrow 0}\frac{f\left(\lambda^0_j+h,\,\omega\right)-f\left(\lambda^0_j,\,\omega\right)}{h}.$$

Assumption 4.9. *For $u\in\widetilde{\mathbf{C}}$, $(\partial^2/\partial u^2)\,f(u,\,\omega)$ has d_f continuous derivatives in ω in a neighborhood of $\omega=0$, the d_f derivative satisfying a Lipschitz condition of order $\varrho_2\in(0,\,1]$.*
For $u\notin\widetilde{\mathbf{C}}$, $(\partial/\partial u_-)\,f(u,\,\omega)$ and $(\partial/\partial u_+)\,f(u,\,\omega)$ have d_f continuous derivatives in ω in a neighborhood of $\omega=0$, the d_f derivative satisfying a Lipschitz condition of order $\varrho_2\in(0,\,1]$.

From Lemma S.B.11 and S.B.17 the relative bias of $\widehat{J}^{*}_{\mathrm{DK},T}$ is

$$\mathsf{B}_{2,T}=\overline{c}_1 b_{1,T}^{d_f}+\overline{c}_2 b_{2,T}^{2}+O\left(b_{1,T}^{d_f+\varrho}+T^{-1}\log T+(Tb_{2,T})^{-1}\right)+o\left(b_{2,T}^{2}\right),$$

where

$$\overline{c}_1=\frac{f^{(d_f)}\left(u,\,0\right)\mu_{d_f}\left(K\right)}{d_f!\int_0^1 f\left(u,\,0\right)du},\qquad \overline{c}_2=\frac{2^{-1}\int_0^1 x^2K_2\left(x\right)dx\int_{\widetilde{\mathbf{C}}}\frac{\partial^2}{\partial u^2}f\left(u,\,0\right)du+\Delta_f\left(0\right)}{\int_0^1 f\left(u,\,0\right)du}.$$

We now present a second-order Edgeworth expansion to approximate the distribution of $\mathbf{v}$ with error $o((Tb_{1,T}b_{2,T})^{-1/2})$. The expansion includes terms up to order $(Tb_{1,T}b_{2,T})^{-1/2}$ to correct the asymptotic normal distribution. This implies the validity of that expansion for the distribution of $\widehat{J}^{*}_{\mathrm{DK},T}$. For $\mathbf{B}\in\mathscr{B}^2$, let $\mathbb{Q}^{(2)}_{2,T}(\mathbf{B})=\int_{\mathbf{B}}\varphi_2\left(\mathbf{v}\right)q^{(2)}_{2,T}\left(\mathbf{v}\right)d\mathbf{v}$, where

$$q^{(2)}_{2,T}\left(\mathbf{v}\right)=1+\left(1/3!\right)\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\left\{\Xi_{2,0}(0,\,3)\mathcal{H}_{2,3}\left(v_2\right)+\Xi_{2,0}(2,\,1)\mathcal{H}_{2,2}\left(v_1\right)\mathcal{H}_{2,1}\left(v_1\right)\right\},$$

$\mathcal{H}_{2,j}\left(\cdot\right)$ are the univariate Hermite polynomials of order j and $\Xi_{2,0}(0,\,3)$ and $\Xi_{2,0}(2,\,1)$ are bounded and depend on K, K_2 and on $f\left(u,\,0\right)$ (see Lemma S.B.5-S.B.6).

Theorem 4.3. *Let Assumption 4.1, 4.2 ($p>1$), 4.3-4.4, 4.7 ($0<q<1$), 4.8-4.9 hold. For* $\phi_T=(Tb_{1,T}b_{2,T})^{-\varpi}$ *with* $1/2<\varpi<1$*, and every class* $\mathscr{B}^2$ *of Borel sets in* $\mathbb{R}^2$*, we have*

$$\sup_{\mathbf{B}\in\mathscr{B}^2}\left|\mathbb{P}_T\left(\mathbf{B}\right)-\mathbb{Q}^{(2)}_{2,T}\left(\mathbf{B}\right)\right|=o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right)+\left(4/3\right)\sup_{\mathbf{B}\in\mathscr{B}^2}\mathbb{Q}^{(2)}_{2,T}\left(\left(\partial\mathbf{B}\right)^{2\phi_T}\right). \tag{4.8}$$

Theorem 4.3 shows that $\mathbb{Q}^{(2)}_{2,T}$ is a valid second-order Edgeworth expansion for the probability measure $\mathbb{P}_T$ of $\mathbf{v}$. The correction $q^{(2)}_{2,T}\left(\mathbf{v}\right)$ differs from $q^{(2)}_{T}\left(\mathbf{h}\right)$ in Theorem 4.1. This difference depends on the smoothing over time, i.e., on $b_{2,T}$ and $K_2\left(\cdot\right)$. The theorem also suggests that the leading term of the error of the approximation is order of $o((Tb_{1,T}b_{2,T})^{-1/2})$.

Next, we focus on U_T defined in (4.7), i.e., a t-statistic based on $\widehat{J}^{*}_{\mathrm{DK},T}$, and present the Edgeworth expansion. We need the following assumption, replacing Assumption 4.6-4.7, that controls the rate of smoothing over lagged autocovariances and time implied by the bandwidths $b_{1,T}$ and $b_{2,T}$, respectively. It requires that the bias due to smoothing over frequency and over time is of the same order as the correction term obtained in $\mathbb{Q}^{(2)}_{2,T}\left(\mathbf{B}\right)$ or as the standard deviation of $\widehat{J}^{*}_{\mathrm{DK},T}$. The assumption is satisfied by, for example, the MSE-optimal DK-HAC estimators proposed by Belotti et al. (2023) and Casini (2022c).

Assumption 4.10. *The bandwidths* $b_{1,T} \to 0$ *and* $b_{2,T} \to 0$ *satisfy* $0 < b_{1,T}^{d_f} \left(Tb_{1,T}b_{2,T}\right)^{-1/2} < \infty$ *and* $0 < b_{2,T}^{2} \left(Tb_{1,T}b_{2,T}\right)^{-1/2} < \infty$.

Theorem 4.4. *Let Assumption 4.1, 4.2* $(p > 1)$*, 4.3-4.5, and 4.8-4.10 hold. For convex Borel sets* $\mathbf{C}$*, we have, for* $r_2(x) = -\bar{c}_1 (x^2 - 1)/2$ *and* $r_3(x) = -\bar{c}_2 (x^2 - 1)/2$*,*

$$\sup_{\mathbf{C}} \left| \mathbb{P}\left(U_T \in \mathbf{C}\right) - \int_{\mathbf{C}} \varphi(x) \left(1 + r_2(x)\, b_{1,T}^{d_f} + r_3(x)\, b_{2,T}^{2}\right) dx \right| = o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right). \tag{4.9}$$

Theorem 4.4 shows that the correction term to the standard normal distribution, i.e., $\int_{\mathbf{C}} \varphi(x)$ $(r_2(x)\, b_{1,T}^{d_f} + r_3(x)\, b_{2,T}^{2})dx$, depends on both smoothing directions. The error of the approximation is of order $o((Tb_{1,T}\ b_{2,T})^{-1/2})$ which can be larger than that obtained in Theorem 4.2 for the HAC estimators. Similarly to (4.5), we obtain uniformly in z,

$$\mathbb{P}\left(U_T \leq z\right) = \Phi(z) + O\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right), \tag{4.10}$$

where $\mathbf{C} = (-\infty,\ z]$, which suggests that the standard normal approximation is correct up to order $O((Tb_{1,T}b_{2,T})^{-1/2})$. Returning to the location model, consider the t-statistic based on $\widehat{J}_{\mathrm{DK},T}^{*}$,

$$t_{\mathrm{DK}} = \frac{\sqrt{T}\left(\widehat{\beta} - \beta_0\right)}{\sqrt{\widehat{J}_{\mathrm{DK},T}^{*}}}.$$

Theorem 4.4 and (4.10) imply that

$$\mathbb{P}\left(t_{\mathrm{DK}} \leq z\right) = \Phi(z) + p_2(z) \left(Tb_{1,T}b_{2,T}\right)^{-1/2} + o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right), \tag{4.11}$$

for any $z \in \mathbb{R}$, where $p_2(z)$ is an odd function different from $p(z)$ in (4.6). Thus, the ERP of t_{DK} can be larger than that of t_{HAC}, though the margin is very small. This follows from the fact that $\widehat{J}_{\mathrm{DK},T}^{*}$ applies smoothing over two directions. The smoothing over time is useful to more flexibly account for nonstationarity. Its benefits appear explicitly under the alternative hypothesis as we show in the next section whereas the ERP refers to the null hypothesis. One can also show that the ERP of t_{DK} and t_{HAC} remain unchanged if prewhitening is applied, though the proofs are omitted since they are similar.

We can further compare the ERP of t_{HAC} and t_{DK} to that of the corresponding t-test under the fixed-b asymptotics. Casini (2022b) showed that the limiting distribution of the original fixed-b HAR test statistics under nonstationarity is not pivotal as it depends on the true data-generating process of the errors and regressors. This contrasts to the stationarity case for which the fixed-b limiting distribution is pivotal and the ERP is of order $O(T^{-1})$ [see Jansson (2004) and Sun et al. (2008)]. Based on an ERP of smaller magnitude relative to that of HAR tests based on HAC estimators [cf. $O(T^{-1}) < O((Tb_{1,T})^{-1/2})$], the literature has long suggested that the original fixed-b HAR tests are superior to HAR tests based on HAC estimators. However, this break downs under nonstationarity as shown by Casini (2022b) who established that (i) the ERP of the original fixed-b HAR tests does not converge to zero because under nonstationarity the fixed-b limiting distribution is different; (ii) for fixed-b HAR tests that use the critical values from the non-pivotal fixed-b limiting distribution the ERP increases by an order of magnitude relative to the stationary case [i.e., from $O(T^{-1})$ to $O(T^{-\eta})$ with $\eta \in (0, 1/2)$]. Therefore, fixed-b HAR tests can have an ERP larger than that of t_{HAC} and t_{DK}. Overall, the finite-sample results based on Edgeworth expansions show that the distortions on the null rejection rates of the HAR tests can arise from time variation in the second moments even when the mean is constant. Thus, these results complement the asymptotic bias results induced by breaks in the mean function discussed in Section 3.

5 Consequences for HAR Inference

In this section we discuss the implications of the theoretical results from Section 3-4. In Section 5.1, we first present a review of HAR inference methods and their connection to the estimates considered in Section 3. We separate the discussion into two parts. In Section 5.2, we discuss HAR inference tests for which the issues of low frequency contamination induce a finite-sample problem, and in Section 5.3 when they persist even asymptotically. For the latter case, in Section 5.4 we provide theoretical results about the power of the tests.

5.1 HAR Inference Methods

There are two main approaches for HAR inference. Classical HAC standard errors [cf. Newey and West (1987, 1994) and Andrews (1991)] require estimation of the LRV defined as $J \triangleq \lim_{T\to\infty} J_T$ where J_T is defined after (4.1). The form of $\{V_t\}$ depends on the specific problem under study. For example, for a t-test on a regression coefficient in the linear model $y_t = x_t\beta_0 + e_t$ $(t = 1, \ldots, T)$

we have $V_t = x_t e_t$. Classical HAC estimators take the following form,

$$\widehat{J}_{\mathrm{HAC},T} = \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right)\widehat{\Gamma}\left(k\right),$$

where $\widehat{\Gamma}(k)$ is given in (3.1) with $\widehat{V}_t = x_t\widehat{e}_t$ where $\{\widehat{e}_t\}$ are the least-squares residuals, $K_1(\cdot)$ is a kernel and $b_{1,T}$ is bandwidth. One can use the the Bartlett kernel, advocated by Newey and West (1987), the quadratic spectral kernel as suggested by Andrews (1991), or any other kernel suggested in the literature, see e.g. de Jong and Davidson (2000) and Ng and Perron (1996). Under $b_{1,T} \to 0$ at an appropriate rate, we have $\widehat{J}_{\mathrm{HAC},T} \xrightarrow{\mathbb{P}} J$. Hence, equipped with $\widehat{J}_{\mathrm{HAC},T}$, HAR inference is standard and simple because HAR test statistics follow asymptotically standard distributions.

HAC standard errors can result in oversized tests when there is substantial temporal dependence [e.g., Andrews (1991)]. This stimulated a second approach based on LRV estimators that keep the bandwidth at some fixed fraction of T [cf. Kiefer et al. (2000)], e.g., using all autocovariances, so that $\widehat{J}_{\mathrm{KVB},T} \triangleq T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1-|t-s|/T\right)\widehat{V}_t\widehat{V}_s$ which is equivalent to the Newey-West estimator with $b_{1,T} = T^{-1}$. Under fixed-b asymptotics the reference distribution of HAR test statistics is nonstandard. The validity of fixed-b inference rests on stationarity [cf. Casini (2022b)]. Many authors have considered various versions of $\widehat{J}_{\mathrm{KVB},T}$. However, the one that leads to HAR inference tests that are least oversized is the original $\widehat{J}_{\mathrm{KVB},T}$ [see Casini and Perron (2021d) for simulation results]. For comparison we also report the equally-weighted cosine (EWC) estimator of Lazarus et al. (2021). It is an orthogonal series estimators that use long bandwidths,

$$\widehat{J}_{\mathrm{EWC},T} \triangleq B^{-1}\sum_{j=1}^{B}\Lambda_j^2, \qquad \text{where} \quad \Lambda_j = \sqrt{\frac{2}{T}}\sum_{t=1}^{T}\widehat{V}_t\cos\left(\pi j\left(\frac{t-1/2}{T}\right)\right)$$

with B some fixed integer. Assuming B satisfies some conditions, under fixed-b asymptotics a t-statistic normalized by $\widehat{J}_{\mathrm{EWC},T}$ follows a t_B distribution where B is the degree of freedom.

Recently, a new HAC estimator was proposed in Casini (2022c). Motivated by the power impact of low frequency contamination of existing LRV estimators, he proposed a double kernel HAC (DK-HAC) estimator, defined by

$$\widehat{J}_{\mathrm{DK},T} \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right)\widehat{\Gamma}_{\mathrm{DK}}\left(k\right),$$

where $b_{1,T}$ is a bandwidth sequence and $\widehat{\Gamma}_{\text{DK}}(k)$ defined in Section 3 with $\widehat{c}_T(\cdot, k)$ replaced by

$$\widehat{c}_{\text{DK},T}(rn_T/T, k) = (Tb_{2,T})^{-1} \sum_{s=|k|+1}^{T} K_2\left(\frac{(rn_T - (s - |k|/2))/T}{b_{2,T}}\right) \widehat{V}_s \widehat{V}_{s-|k|},$$

with K_2 a kernel and $b_{2,T}$ a bandwidth. Note that $\widehat{c}_{\text{DK},T}$ and $\widehat{c}_T$ are asymptotically equivalent and the results of Section 3 continue to hold for $\widehat{c}_{\text{DK},T}$. More precisely, $\widehat{c}_T$ is a special case of $\widehat{c}_{\text{DK},T}$ with K_2 being a rectangular kernel and $n_{2,T} = Tb_{2,T}$. This approach falls in the first category of standard inference $\widehat{J}_{\text{DK},T} \xrightarrow{\mathbb{P}} J$ and HAR test statistics normalized by $\widehat{J}_{\text{DK},T}$ follows standard distribution asymptotically. Additionally, Casini and Perron (2021d) proposed prewhitened DK-HAC ($\widehat{J}_{\text{pw,DK},T}$) estimator that improves the size control of HAR tests. The estimator $\widehat{J}_{\text{pw,DK},T}$ applies a prewhitening transformation to the data before constructing $\widehat{J}_{\text{DK},T}$ and enjoys the same asymptotic properties. Due to their ability to more flexibly account for nonstationarity, Casini (2022c) and Casini and Perron (2021d) demonstrated via simulations that tests based on $\widehat{J}_{\text{DK},T}$ and $\widehat{J}_{\text{pw,DK},T}$ have superior power properties relative to tests based on the other estimators. In terms of size, the simulation results showed that tests based on $\widehat{J}_{\text{pw,DK},T}$ performs better than those based on $\widehat{J}_{\text{HAC},T}$ and $\widehat{J}_{\text{DK},T}$, and is competitive with $\widehat{J}_{\text{KVB},T}$ when the latter works well. We include $\widehat{J}_{\text{DK},T}$ and $\widehat{J}_{\text{pw,DK},T}$ in our simulations below. We report the results only for the DK-HAC estimators that do not use the pre-test for discontinuities in the spectrum [cf. Casini and Perron (2021a)] because we do not want the results to be affected by such pre-test.

5.2 Small-Sample Low Frequency Contamination

We now discuss situations in which the low frequency contamination arises as a small-sample problem. These comprise situations where $d^* \approx 0$ asymptotically but a consistent estimate of d^* satisfies $\widehat{d}^* > 0$ in finite-sample. In these situations, one can expect the low frequency contamination to have an effect in small samples.

We begin with a simple model involving a zero-mean SLS process with changes in persistence. The combination of nonstationarity and serial dependence generates long memory effects because $\widehat{d}^* > 0$. We specify $\{V_t\}$ as a SLS process given by a two-regimes zero-mean time-varying AR(1), labeled model M1. That is, $V_t = 0.9V_{t-1} + u_t$, $V_0 \sim \mathscr{N}(0, 1)$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $t = 1, \ldots, T_1^0$ with $T_1^0 = T\lambda_1^0$, and $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.3(\cos(1.5 - \cos(t/T)))$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.5)$ for $t = T_1^0 + 1, \ldots, T$. Note that $\rho(\cdot)$ varies between 0.172 and 0.263. We set $\lambda_1^0 = 0.1$ and $T = 200$. A plot of $\{V_t\}$ is reported in Figure 1. Note that $\mathbb{E}(V_t) = 0$ for all t and so $d^* = 0$. However, if we replace $\overline{\mu}_1$ and $\overline{\mu}_2$ in d^* by $\overline{V}_1$ and $\overline{V}_2$, respectively, where

$\overline{V}_1 = (T_1^0)^{-1} \sum_{t=1}^{T_1^0} V_t = 1.27$ and $\overline{V}_2 = (T - T_1^0)^{-1} \sum_{t=T_1^0+1}^{T} V_t = -0.03$, then the estimate $\widehat{d}^*$ of d^* is different from zero and generates a finite-sample bias which can give rise to long memory effects.

We first look at the behavior of the sample autocovariance $\widehat{\Gamma}(k)$. We compare it with the theoretical autocovariance $\Gamma_T(k) = T^{-1} \sum_{t=k+1}^{T} \mathbb{E}(V_t V_{t-k})$. The latter is equal to $\Gamma_T(k) \approx \lambda_1^0 0.9^k / (1 - 0.81) + \int_{\lambda_1^0}^{1} c(u,\, k)\, du$. We can compute $\Gamma_T(k)$ numerically using,

$$c(u,\, k) = \int_{-\pi}^{\pi} e^{ik\omega} f(u,\, \omega)\, d\omega = \int_{-\pi}^{\pi} e^{ik\omega} \frac{0.5}{2\pi} \left(1 + \rho(u)^2 - 2\rho(u)\cos(\omega)\right)^{-1} d\omega.$$

Table 1 reports $\Gamma_T(k)$, $\widehat{\Gamma}(k)$, $\widehat{\Gamma}(k) - \widehat{d}^*$ and $\widehat{\Gamma}_{\mathrm{DK}}(k)$ for several values of k using 5,000 replications. Despite noisy autocovariance estimates, it is still possible to discern some patterns. For all k, $\widehat{\Gamma}(k)$ largely overestimates $\Gamma_T(k)$. This is consistent with Theorem S.A.1 which suggests that this is due to the bias $d^* > 0$. This is also supported by the bias-corrected estimate $\widehat{\Gamma}(k) - \widehat{d}^*$ which is more accurate in approximating $\Gamma_T(k)$. This is especially so for small k. In general, Theorem S.A.1 provides good approximations confirming that $\widehat{\Gamma}(k)$ suffers from low frequency contaminations. Theorem 3.1 suggests that this issue should not occur for $\widehat{\Gamma}_{\mathrm{DK}}(k)$. In fact, $\widehat{\Gamma}_{\mathrm{DK}}(k)$ is more accurate than $\widehat{\Gamma}(k)$. For $k \geq 20$ the form of $\Gamma_T(k)$ is different, because $T_1^0 = 20$, and is simply given by the autocovariance of V_t for $t \geq 21$ (i.e., the second regime). Thus, $\Gamma_T(k) \approx 0$ for $k \geq 20$ whereas $\widehat{\Gamma}(k)$ is often small but positive. In contrast, $\widehat{\Gamma}_{\mathrm{DK}}(k) \approx 0$ for $k \geq 20$ thereby confirming that $\widehat{\Gamma}_{\mathrm{DK}}(k)$ does not suffer from low frequency contamination. These results are confirmed in Figure 2 which plots the autocorrelation function (ACF) of $V_{1,t} = V_t$ $(t = 1, \ldots, 20)$, $V_{2,t} = V_t$ $(t = 21, \ldots, 200)$ and V_t $(t = 1, \ldots, 200)$. 5,000 repetitions are used. Although the ACF of V_t should be a weighted average of the ACF of $V_{1,t}$ and of $V_{2,t}$, the ACF of V_t in the bottom panel shows much higher persistence than either ACF in the top $(V_{1,t})$ and mid $(V_{2,t})$ panels. This is odd since $V_{1,t}$ is a highly persistent series. Further, it shows that the dependence is essentially always positive. These features are consistent with our theory which suggests that nonstationarity makes V_t appear more persistent and that the bias is positive. Other examples with similar features involve V_t given by the least-squares residuals under mild forms of misspecification that do not undermine the conditions for consistency of the least-squares estimator, e.g., exclusion of a relevant regressor uncorrelated with the included reggressors, or inclusion of an irrelevant regressor. Another example involves V_t obtained after applying some de-trending techniques where the fitted model is not correctly specified (e.g., the data follow a nonlinear trend but one removes a linear trend). A final example is the case of outliers because they influence the mean of V_t and therefore d^*.

What is especially relevant is whether this evidence of long memory feature has any consequence for HAR inference. We obtain the empirical size and power for a t-test on the intercept normalized by several LRV estimators for the model $y_t = \delta + V_t$ with $\delta = 0$ under the null and $\delta > 0$ under the alternative hypothesis. In addition to M1, we consider other models: M2 involves a time-varying AR(1) with a break in volatility $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.7(\cos(1.5t/T))$, $u_t \sim \mathscr{N}(0, \sigma_t^2)$, $\sigma_t^2 = 5$ for $t \leq 4$ and $\sigma_t^2 = 0.25$ for $t > 4$, $V_0 \sim \mathscr{N}(0, 5)$; M3 involves $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.8(\cos(1.5t/T))$, $u_t \sim \mathscr{N}(0, 0.25)$, $V_0 = 0$ with outliers $V_t \sim \text{Uniform}(\underline{c}, 5\underline{c})$ for $t = T/2$, $3T/4$ where $\underline{c} = -1/(\sqrt{2}\text{erfc}^{-1}(3/2))\text{med}(|V - \text{med}(V)|)$ with erfc^{-1} the inverse complementary error function, $\text{med}(\cdot)$ is the median and $V = (V_t)_{t=1}^T$;[5] M4 involves a time varying AR(1) with periods of strong persistence where $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.95(\cos(1.5t/T))$, $u_t \sim \text{i.i.d.}\,\mathscr{N}(0, 0.4)$ and $V_0 \sim \mathscr{N}(0, 4)$. $\rho(\cdot)$ varies between 0.7 and 0.05 in M2, between 0.05 and 0.8 in M3 and between 0.95 and 0.07 in M4.

We consider the DK-HAC estimators with and without prewhitening ($\widehat{J}_{\text{DK},T}$, $\widehat{J}_{\text{DK,pw,SLS},T}$, $\widehat{J}_{\text{DK,pw,SLS},\mu,T}$) of Casini (2022c) and Casini and Perron (2021d), respectively; Andrews' (1991) HAC estimator with and without the prewhitening procedure of Andrews and Monahan (1992); Newey and West's (1987) HAC estimator with the popular rule to select the number of lags (i.e., $b_{1,T} = (4(T/100)^{2/9})^{-1}$; Newey-West with the fixed-b method of Kiefer et al. (2000) with $b = 1$ (labeled KVB); and the Equally-Weighted Cosine (EWC) of Lazarus et al. (2018) with the bandwidth choice reccomended by the authors. For the DK-HAC estimators we use the data-dependent methods for the bandwidths, kernels and choice of n_T as proposed in Casini (2022c) and Casini and Perron (2021d), which are optimal under mean-squared error (MSE). Let $\widehat{V}_t$ denote the least-squares residual based on $\widehat{\delta}$ where the latter is the least-squares estimate of δ. We set $\widehat{b}_{1,T} = 0.6828(\widehat{\phi}(2)\, T\widehat{\overline{b}}_{2,T})^{-1/5}$ where

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

with $\widehat{a}_1(u) = \frac{\sum_{j=t-n_T+1}^{t}\widehat{V}_j\widehat{V}_{j-1}}{\sum_{j=t-n_T+1}^{t}(\widehat{V}_{j-1})^2}$ and $\widehat{\sigma}(u) = (\sum_{j=t-n_T+1}^{t}(\widehat{V}_j-\widehat{a}_1(u)\,\widehat{V}_{j-1})^2)^{1/2}$ and $\widehat{\overline{b}}_{2,T} = (n_T/T)\sum_{r=1}^{\lfloor T/n_T\rfloor-1}\widehat{b}_{2,T}(rn_T/T)$, $\widehat{b}_{2,T}(u) = 1.6786(\widehat{D}_1(u))^{-1/5}(\widehat{D}_2(u))^{1/5}T^{-1/5}$ where $\widehat{D}_2(u) \triangleq 2\sum_{l=-\lfloor T^{4/25}\rfloor}^{\lfloor T^{4/25}\rfloor}\widehat{c}_{\text{DK},T}(u, l)^2$

[5] We follow the literature on outlier detection for continuous functions and use the median absolute deviation to generate the outlier. This notion used in this literature does not deem a value smaller than $\underline{c}$ as an outlier.

and

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

with $[S_\omega]$ being the cardinality of S_ω and $\omega_{s+1} > \omega_s$, $\omega_1 = -\pi$, $\omega_{[S_\omega]} = \pi$. We set $n_T = T^{0.6}$, $S_\omega = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\}$. $K_1(\cdot)$ is the QS kernel and $K_2(x) = 6x(1-x)$ for $x \in [0, 1]$.

Table 2-5 report the results. The t-test normalized by Newey and West's (1987) and Andrews' (1991) prewhitened HAC estimators are excessively oversized.[6] Andrews' (1991) HAC-based test is slightly undersized while KVB's fixed-b and EWC-based tests are severely undersized.[7] These outcomes arise in part from nonstationarity and are consistent with our theoretical analysis of $\widehat{\Gamma}(k)$. Since $\widehat{\Gamma}(k) > 0$ for many large lags k, the KVB's fixed-b and EWC's estimators that include many lags are inflated and reduce the magnitude of the test statistic even under the null hypotheses. The fact that KVB's fixed-b and EWC-based tests have larger size distortions than other tests is also consistent with the results in Section 4 which suggest that they have a larger ERP. For the t-test on the intercept, $\widehat{J}_{\mathrm{DK},T}$ can lead to tests that are oversized when there is strong dependence, as shown in Table 2. However, the prewhitened DK-HAC estimators $\widehat{J}_{\mathrm{DK,pw,SLS},T}$ and $\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$ lead to tests that show more accurate rejection rates. Overall, inspection of the size properties suggests that the DK-HAC-based tests do not suffer from low frequency contamination which, in contrast, affects the tests based on LRV estimators that rely on the full sample estimates $\widehat{\Gamma}(k)$ or equivalently on $I_T(\omega)$ (e.g., the EWC). The same issue also affects the power properties of the tests. The KVB's fixed-b and EWC-based estimators suffer from relatively large power losses. The power of tests normalized by Newey and West's (1987) and Andrews' (1991) prewhitened HAC are not comparable because they are significantly oversized. The DK-HAC-based tests have the best

[6]This is not in contradiction with our theoretical results. The prewhitening of Andrews and Monahan (1992) is unstable when there is nonstationarity as shown by Casini and Perron (2021d). The reason is that there is a bias both in the whitening and the recoloring stages. The biases have opposite signs so that here the underestimation of the LRV dominates. Newey-West uses a fixed rule for determining the number of lags. The number of included lags is small. This estimator is known to be largely oversized when the data are stationary with high dependence. Our results say that the included sample autocovariances may be inflated if there is nonstationarity. However, given that the fixed rule selects a small number of lags then nonstationarity results in a smaller oversize problem than under stationarity.

[7]In general, Andrews' (1991) HAC estimator leads to tests that are oversized when the data are stationary with strong dependence. Here, nonstationarity reduces the oversize problem. It follows from a similar argument as for the Newey-West estimator even though Andrews' (1991) HAC estimator uses a data-dependent method for the selection of the number of lags. Thus, it selects more lags than the ones suggested by the fixed rule. Consequently, more sample autocovariances are overestimated and this helps to reduce its oversize problem.

power, the second best being Andrews' (1991) HAC-based test.

Turning to M2, Table 3 shows some size distortions and power losses for KVB's fixed-b and EWC-based tests. The prewhitened DK-HAC-based tests display accurate size control and good power. Newey and West's (1987) and Andrews' (1991) prewhitened HAC-based tests are again excessively oversized. Andrews' (1991) HAC-based test and the DK-HAC-based test show a similar performance. For model M3-M4, Table 4-5 show that all methods lead to oversized tests except prewhitened DK-HAC and KVB's fixed-b. However, KVB's fixed-b-based tests show substantial underejection that has consequences for power whereas the prewhitened DK-HAC-based-tests show accurate null rejection rates and good power.

Overall, Table 2-5 suggest that there are offsetting effects on the rejection rates. On the one hand, strong dependence (either under stationarity or nonstationarity) tends to increase the rejection rates and this likely makes the HAR tests oversized. On the other hand, nonstationarity generates upward bias in the LRV estimator and this likely reduces the rejection rates.[8] In addition, there is the issue of the accuracy of the reference distribution used to construct the critical values, though the direction of the distortion in the rejection rates under nonstationarity is unknown. Which of these multiple effects prevails depend on the true data-generating process. Nonetheless, it is possible to discern some pattern. The oversize problem due to strong dependence becomes less severe under nonstationarity as one can see from the results. In fact, the HAC-based tests display smaller over-rejections than those under stationarity that one can find in the literature. The method that suffers mostly from under-rejection is the fixed-b-based HAR test. This reconciles with the theoretical results in two mutual ways. First, the fixed-b-based tests use the critical value from the incorrect fixed-b distribution. Thus, the theory suggests that the ERP of the fixed-b-based tests should be larger than that of the HAC and DK-HAC-based tests. Second, the fixed-b-based tests use a larger number of sample autocovariances. Thus, the theory suggests that the fixed-b LRV estimator is more inflated than the other LRV estimators. Finally, the simulations show that the null rejection rates of HAC- and DK-HAC-based tests are not very far from each other, thereby confirming that the their respective ERP are close as shown in Section 4.

[8]In model M1-M4 the low frequency contamination occurs as a small-sample problem where $d^* = 0$ but $\widehat{d}^* > 0$. Although this bias can be substantial as shown in Table 1 which is based on model M1, it does not prevent any HAR test from having monotonic power. Non-monotonic power occurs when the low frequency contamination continues to hold asymptotically, i.e., when $d^* > 0$ (see Section 5.3).

5.3 General Low Frequency Contamination

We now discuss HAR inference tests for which the low frequency contamination results of Section 3 hold even asymptotically. This means that $d^* > 0$ for all T and as $T \to \infty$. This comprises the class of HAR tests that admit a nonstationary alternative hypotheses. This class is very large and include most HAR tests as discussed in the Introduction. Here we consider the Diebold-Mariano test for the sake of illustration and remark that similar issues apply to other HAR tests.

The Diebold-Mariano test statistic is defined as $t_{\mathrm{DM}} \triangleq T_n^{1/2}\overline{d}_L/\sqrt{\widehat{J}_{d_L,T}}$, 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 LRV of the the loss differential series and T_n is the number of observations in the out-of-sample. We use the quadratic loss. We consider an out-of-sample forecasting exercise with a fixed forecasting scheme where, given a sample of T observations, $0.5T$ observations are used for the in-sample and the remaining half is used for prediction [see Perron and Yamamoto (2021) for recommendations on using a fixed scheme in the presence of breaks]. The DGP under the null hypotheses is given by $y_t = 1 + \beta_0 x_{t-1}^{(0)} + e_t$ 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 $\beta_0 = 1$ and $T = 400$. 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 $L_t^{(j)}$ $(j = 1,\,2)$ 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 hypotheses of equal predictive ability when $\overline{d}_L$ is sufficiently far from zero. Under the alternative hypothesis, the two competing forecast models are as follows: the first uses $x_t^{(1)} = x_t^{(0)} + u_{X_1,t}$ where $u_{X_1,t} \sim$ i.i.d. $\mathscr{N}(0,\,1)$ while the second uses $x_t^{(2)} = x_t^{(0)} + 0.2z_t + 2u_{X_2,t}$ for $t \in [1, \ldots, 3T/4-1,\, 3T/4+21, \ldots T]$ and $x_t^{(2)} = \delta(t/T) + 0.2z_t + 2u_{X_2,t}$ for $t = 3T/4, \ldots, 3T/4+20$ with $u_{X_2,t} \sim$ i.i.d. $\mathscr{N}(0,\,1)$, where z_t has the same distribution as $x_t^{(0)}$.

We consider four specifications for $\delta(\cdot)$. In the first $x_t^{(2)}$ is subject to an abrupt break in the mean $\delta(t/T) = \delta > 0$; in the second $x_t^{(2)}$ is locally stationary with time-varying mean $\delta(t/T) = \delta(\sin(t/T - 3/4))$; in the third specification $x_t^{(2)} = x_t^{(0)} + 0.2z_t + 2u_{X_2,t}$ for $t \in [1, \ldots, T/2-30,\, T/2 +21, \ldots T]$ and $x_t^{(2)} = \delta(t/T) + 0.2z_t + 2u_{X_2,t}$ for $t = T/2-30, \ldots, T/2+20$ with $\delta(t/T) = \delta(\sin(t/T - 1/2 - 30/T))$; in the fourth $x_t^{(2)}$ is the same as in the second with in addition two outliers $x_t^{(2)} \sim \mathrm{Uniform}(|\underline{c}|,\, 5|\underline{c}|)$ for $t = 6T/10,\, 8T/10$ where $\underline{c} = -1/(\sqrt{2}\mathrm{erfc}^{-1}(3/2))\mathrm{med}(|x^{(2)} - \mathrm{med}(x^{(2)})|)$ where $x^{(2)} = (x_t^{(2)})_{t=1}^{T}$. That is, in the second model $x_t^{(2)}$ is locally stationary only in the out-of-sample, in the third it is locally stationary in both the in-sample and out-of sample and in the fourth model $x_t^{(2)}$ has two outliers in the out-of-sample. The location of the outliers is irrelevant for the results; they can also occur in the in-sample.

Table 6 reports the null rejection rate and the power of the various tests for all models. We begin with the case $\delta(t/T) = \delta > 0$ (top panel). The null rejection rate of the test using the DK-HAC estimators is accurate while the test using other LRV estimators are oversized with the exception of the KVB's fixed-b method for which the rejection rate is equal to zero. The HAR tests using existing LRV estimators have lower power relative to that obtained with the DK-HAC estimators for small values of δ. When δ increases the tests standardized by the HAC estimators of Andrews (1991) and Newey and West (1987), and by the KVB's fixed-b and EWC LRV estimators display non-monotonic power gradually converging to zero as the alternative gets further away from the null value. In contrast, when using the DK-HAC estimators the test has monotonic power that reaches and maintains unit power. The results for the other models are even stronger. In general, except when using the DK-HAC estimators, all tests display serious power problems. Thus, either form of nonstationarity or outliers leads to similar implications, consistent with our theoretical results.

In order to further assess the theoretical results from Section 3, Figure 3 reports the plots of d_t, its sample autocovariances and its periodogram, for $\delta = 1$. Figure S.1-S.2 in the supplement report the corresponding plots for $\delta = 2, 5$, respectively. We only consider the case $\delta_t = \delta > 0$. The other cases lead to the same conclusions. For $\delta = 1$, Figure 3 (mid panel) shows that $\widehat{\Gamma}(k)$ decays slowly. As δ increases, Figure S.1 and S.2 (mid panels), $\widehat{\Gamma}(k)$ decays even more slowly at a rate far from the typical exponential decay of short memory processes. This suggests evidence of long memory. However, the data are short memory with small temporal dependence. What is generating the spurious long memory effect is the nonstationarity present under the alternative hypotheses. This is visible in the top panels which present plots of d_t for the first specification. The shift in the mean of d_t for $t = 3T/4, \ldots, 3T/4 + 20$ is responsible for the long memory effect. This corresponds to the second term of (S.A.7) in Theorem S.A.1. The overall behavior of the sample autocovariance is as predicted by Theorem S.A.1. For small lags, $\widehat{\Gamma}(k)$ shows a power-like decay and it is positive. As k increases to medium lags, the autocovariances turn negative because the sum of all sample autocovariances has to be equal to zero [cf. Percival (1992)]. Next, we move to the bottom panels which plot the periodogram of $\{d_t\}$. It is unbounded at frequencies close to $\omega = 0$ as predicted by Theorem S.A.2 and as would occur if long memory was present. It also explains why the Diebold-Mariano test normalized by Newey-West's, Andrews', KVB's fixed-b and EWC's LRV estimators have serious power problems. These LRV estimators are inflated and consequently the tests lose power. The figures show that as we raise δ the more severe these issues and the power losses so that the power eventually reaches zero. This is consistent with our theory since d^* is increasing in δ (cf. $d^* \approx 0.1 \cdot 0.9\delta^2$).

We now verify the results about the local sample autocovariance $\widehat{c}_T(u, k)$ and the local periodogram from Theorem 3.1-3.2. We set $n_{2,T} = T^{0.6} = 36$ following the MSE criterion of Casini (2022c). We consider (i) $u = 236/T$, (ii-a) $u = T_1^0/T = 3/4$ and (ii-b) $u = 264/T$. Note that cases (i)-(ii-b) correspond to parts (i)-(ii-b) in Theorem 3.1-3.2. We consider $\delta = 1$, 2 and 5. According to Thereon 3.1-3.2, we should expect long memory features only for case (ii-a). Figure 4-5 and S.3-S.6 in the supplement confirm this. The results pertaining to case (ii-a) are plotted in the middle panels. Figures 4, S.3 and S.5 show that the local autocovariance displays slow decay similar to the pattern discussed above for $\widehat{\Gamma}(k)$ and that this problem becomes more severe as δ increases. Such long memory features also appear for $I_{\mathrm{L}}(3/4, \omega)$. The middle panels in Figure 5, S.4 and S.6 show that the local periodogram at $u = 3/4$ and at a frequency close to $\omega = 0$ are extremely large. The latter result is consistent with Theorem 3.2-(ii-a) which suggests that $I_{\mathrm{L},T}(3/4, \omega) \to \infty$ as $\omega \to 0$. For case (i) and (ii-b) both figures show that the local autocovariance and the local periodogram do not display long memory features. Indeed, they have forms similar to those of a short memory process, a result consistent with Theorem 3.1-3.2 also for cases (i) and (ii-b).

It is noteworthy to explain why HAR inference based on the DK-HAC estimators does not suffer from the low frequency contamination even if case (ii-a) occurs. The DK-HAC estimator computes an average of the local spectral density over time blocks. If one of these blocks contains a discontinuity in the spectrum, then as in case (ii-a) some bias would arise for the local spectral density estimate corresponding to that block. However, by virtue of the time-averaging over blocks that bias becomes negligible. Hence, nonparametric smoothing over time asymptotically cancels the bias, so that inference based on the DK-HAC estimators is robust to nonstationarity.

5.4 Theoretical Results about the Power

We present theoretical results about the power of t_{DM} for the case of general low frequency contamination discussed in Section 5.3. In particular, we focus on specification (1) (i.e., $\delta > 0$). The same intuition and qualitative theoretical results apply to the other specifications of $\delta(\cdot)$.

Let $t_{\mathrm{DM},i} = T_n^{1/2}\overline{d}_L/\sqrt{\widehat{J}_{d_L,i,T}}$ denote the DM test statistic where $i =$ DK, pwDK, KVB, EWC, A91, pwA91, NW87 and pwNW87 with $\widehat{J}_{\mathrm{A91},T}$ and $\widehat{J}_{\mathrm{NW87},T}$ being $\widehat{J}_{\mathrm{HAC},T}$ using the quadratic spectral and Bartlett kernel, respectively. Define the power of $t_{\mathrm{DM},i}$ as $\mathbb{P}_\delta(|t_{\mathrm{DM},i}| > z_\alpha)$ where z_α is the two-sided standard normal critical value and $\alpha \in (0, 1)$ is the significance level. To avoid repetitions we present the results only for $i =$ DK, KVB and NW87. The results concerning the prewhitening DK-HAC estimator are the same as those corresponding to the DK-HAC estimator while the results concerning the EWC estimator are similar to those corresponding to the KVB's fixed-b estimator, though for the latter the non-monotonic power is more pronounced. The results pertaining to Andrews' (1991) HAC estimator (with and without prewhitening) are the same as those corresponding to Newey and West's (1987) estimator. Let $n_\delta = T - T_b - 2$ denote the length of the regime in which $x_t^{(2)}$ exhibits a shift δ in the mean. The deviation from the null hypothesis depends on the shift magnitude δ and on n_δ.

Theorem 5.1. *Let $\{d_t - \mathbb{E}(d_t)\}_{t=1}^{T_n}$ be a SLS process satisfying Assumption 2.1-(i-iv) and 2.2. Let Assumption 4.3-4.4 hold and $n_\delta = O(T_n^{1/2+\zeta})$ where $\zeta \in (0,\, 1/2)$ such that $T_n^\zeta b_{1,T}^{1/2} \to 0$ and $T_n^\zeta (\widehat{b}_{1,T})^{1/2} \to 0$. Then, we have:*

(i) Under Assumption 4.6, $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}| > z_\alpha) \to 0$. If Assumption 4.6 is replaced by Assumption 4.7 with $q = 1/3$, then $|t_{\mathrm{DM,NW87}}| = O_{\mathbb{P}}(T_n^{\zeta-1/6})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}| > z_\alpha) \to 0$.

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

(iii) Under Assumption 4.8, $|t_{\mathrm{DM,DK}}| = \delta^2 O_{\mathbb{P}}(T_n^\zeta)$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,DK}}| > z_\alpha) \to 1$.

Note that Assumption 4.7 with $q = 1/3$ refers to the MSE-optimal bandwidth for the Newey and West's (1987) estimator. The conditions $T_n^\zeta b_{1,T}^{1/2} \to 0$ and $T_n^\zeta (\widehat{b}_{1,T})^{1/2} \to 0$ mean that the length of the regime in which $x_t^{(2)}$ exhibits a shift δ in the mean increases to infinity at a slower rate than T. Theorem 5.1 shows that when the HAC estimators or the fixed-b LRV estimators are used, the DM test is not consistent and its power approaches zero. The theorem also implies that the power functions corresponding to tests based on HAC estimators lie above the power functions corresponding to those based on fixed-b/EWC LRV estimators. This follows from $|t_{\mathrm{DM,KVB}}| \ll |t_{\mathrm{DM,NW87}}|$. Another interesting feature is that $|t_{\mathrm{DM,NW87}}|$ and $|t_{\mathrm{DM,KVB}}|$ do not increase in magnitude with δ because δ appears in both the numerator and denominator [δ enters the denominator through the low frequency contamination term d^* that accounts for the bias in the HAC and fixed-b estimators [cf. Theorem S.A.1]]. Part (iii) of the theorem suggests that these issues do not occur when DK-HAC estimator is used since the test is consistent and its power increases with δ and with the sample size as it should be. These results match the empirical results in Table 6 discussed above, thereby confirming the relevance of Theorem 5.1.

6 Conclusions

Economic time series are highly nonstationary and models might be misspecified. If nonstationary is not accounted for properly, parameter estimates and, in particular, asymptotic LRV estimates can be largely biased. We establish results on the low frequency contamination induced by nonstationarity and misspecification for the sample autocovariance and the periodogram under general conditions. These estimates can exhibit features akin to long memory when the data are nonstationary short memory. We distinguish cases where this contamination only implies a small-sample problem and cases where the problem remains asymptotically. We show, using theoretical arguments, that nonparametric smoothing is robust. Since the autocovariances and the periodogram are basic elements for HAR inference, our results allow a better understanding of LRV estimation. Existing LRV estimators tend to be inflated when the data are nonstationarity. This results in HAR tests that can be undersized and exhibit dramatic power losses or even no power. Long bandwidths/fixed-b HAR tests suffer more from low frequency contamination relative to HAR tests based on HAC estimators, whereas the DK-HAC estimators do not suffer from this problem.

Supplemental Materials

The supplement for online publication [cf. Casini et al. (2022)] introduces the notion of long memory segmented locally stationary processes, presents the theoretical results referenced in Section 3, contains the proofs of the results in the paper and additional figures.

References

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

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

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

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

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

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

Bentkus, R.Y., Rudzkis, R.A., 1982. On the distribution of some statistical estimates of spectral density. Theory of Probability and Its Applications 27, 795–814.

Bhattacharya, R., Gupta, V., Waymire, E., 1983. The Hurst effect under trends. Journal of Applied Probability 20, 649–662.

Bhattacharya, R.N., Ghosh, J.K., 1978. On the validity of the formal Edgeworth expansion. Annals of Statistics 6, 434–451.

Bhattacharya, R.N., Rao, R.R., 1975. Normal Approximation and Asymptotic Expansion. New York: Wiley.

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

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

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

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

Casini, A., 2022b. The Fixed-b limiting distribution and the ERP of HAR tests under nonstationarity. arXiv preprint arXiv 2111.14590.

Casini, A., 2022c. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Journal of Econometrics, 235, 372–392.

Casini, A., Deng, T., Perron, P., 2022. Supplement to "Theory of low frequency contamination from nonstationarity and misspecification: consequences for HAR inference". arXiv preprint arXiv:2103.01604 .

Casini, A., Perron, P., 2019. Structural breaks in time series, in: Oxford Research Encyclopedia of Economics and Finance,. Oxford University Press.

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

Casini, A., Perron, P., 2021a. Change-point analysis of time series with evolutionary spectra. arXiv preprint arXiv 2106.02031.

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

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

Casini, A., Perron, P., 2021d. Prewhitened long-run variance estimation robust to nonstattionarity. arXiv preprint arXiv:2103.02235.

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

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

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

Christensen, B.J., Varneskov, R.T., 2017. Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination. Journal of Econometrics 97, 218–244.

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

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

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

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

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

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

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

Elliott, G., Müller, U.K., 2007. Confidence sets for the date of a single break in linear time series regressions. Journal of Econometrics 141, 1196–1218.

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

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

Giacomini, R., Rossi, B., 2015. Forecasting in nonstationary environments: What works and what doesn't in reduced-form and structural models. Annual Review of Economics 7, 207–229.

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

Granger, C.W.J., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the SP 500 absolute stock returns. Journal of Empirical Finance 11, 399–421.

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

Hillebrand, E., 2005. Neglecting parameter changes in GARCH models. Journal of Econometrics 129, 121–138.

Hwang, J., Sun, Y., 2017. Asymptotic F and t tests in an efficient GMM setting. Journal of Econometrics 198, 277–295.

Ibragimov, R., Kattuman, P., Skrobotov, A., 2021. Robust inference on income inequality: t-statistic based approaches. arXiv preprint arXiv:2105.05335.

Ibragimov, R., Müller, U.K., 2010. t-statistic based correlation and heterogeneity robust inference. Journal of Business and Economic Statistics 28, 453–468.

Janas, D., 1994. Edgeworth expansions for spectral mean estimates with applications to Whittle estimates. Annals of the Institute of Statistical Mathematics 46, 667–682.

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

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

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

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

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

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

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

Lamoureux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change, and the GARCH model. Journal of Business and Economic Statistics 8, 225–234.

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

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

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

McCloskey, A., Hill, J.B., 2017. Parameter estimation robust to low frequency contamination. Journal of Business and Economic Statistics 35, 598–610.

Mikosch, T., Stărica, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economic and Statistics 86, 378–390.

Müller, U.K., Watson, M.W., 2008. Testing models of low-frequency variability. Econometrica 76, 979–1016.

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

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

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

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

Ng, S., Perron, P., 1996. The exact error in estimating the spectral density at the origin. Journal of Time Series Analysis 17, 379–408.

Ng, S., Wright, J.H., 2013. Facts and challenges from the great recession for forecasting and macroeconomic modeling. Journal of Economic Literature 51, 1120–54.

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

Percival, D., 1992. Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean. The American Statistician 47, 274–276.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 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, 153–162.

Perron, P., Qu, Z., 2010. Long-memory and level shifts in the volatility of stock market return indices. Journal of Business and Economic Statistics 28, 275–290.

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

Phillips, P.C.B., 1977. Approximations to some finite sample distributions associated with a first-order stochastic difference equation. Econometrica 45, 463–485.

Phillips, P.C.B.., 1980. Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume. Review of Economic Studies 47, 183–224.

Phillips, P.C.B., 2005. HAC estimation by automated regression. Econometric Theory 21, 116–142.

Politis, D.M., 2011. Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. Econometric Theory 27, 703–744.

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

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

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

Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459–502.

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

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

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

Stock, J.H., Watson, M.W., 1996. Evidence on structural stability in macroeconomic time series. Journal of Business and Economic Statistics 14, 11–30.

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

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

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

Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. Econometrica 76, 175–194.

Taniguchi, M., Puri, M.L., 1996. Valid Edgeworth expansions of M-estimators in regression models with weakly dependent residuals. Econometric Theory 12, 331–346.

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

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

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

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

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

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

A Appendix

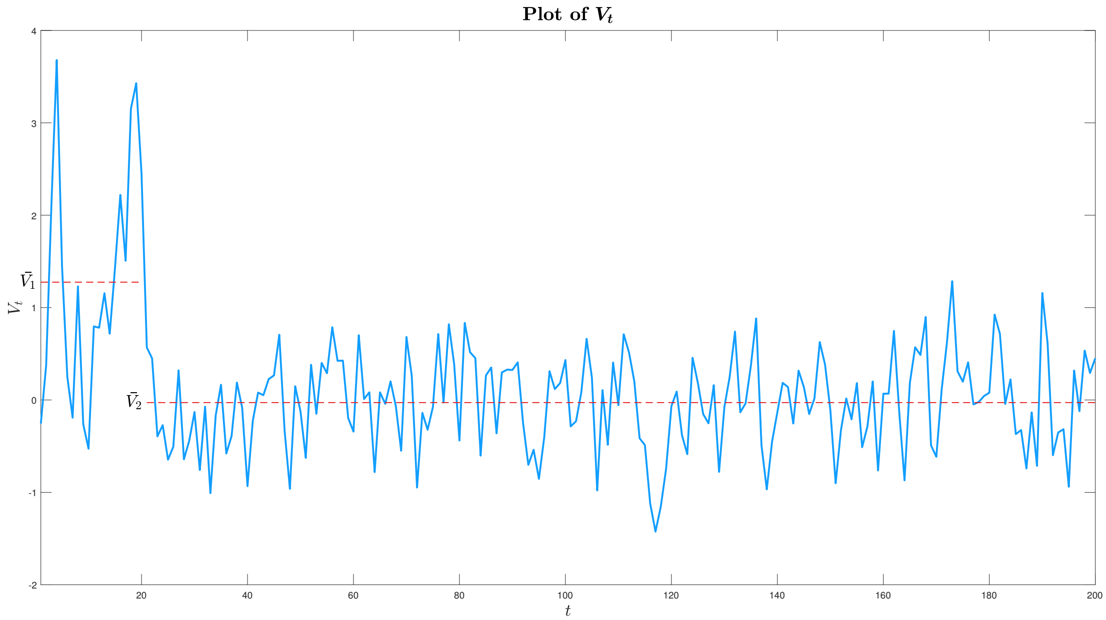

Figure 1: Plot of V_t for model M1. The sample size is $T = 200$. $T_1^0 = 20$. Also reported in red dashed lines are the sample averages in the two regimes with $\overline{V}_1 = 1.27$ and $\overline{V}_2 = -0.03$.

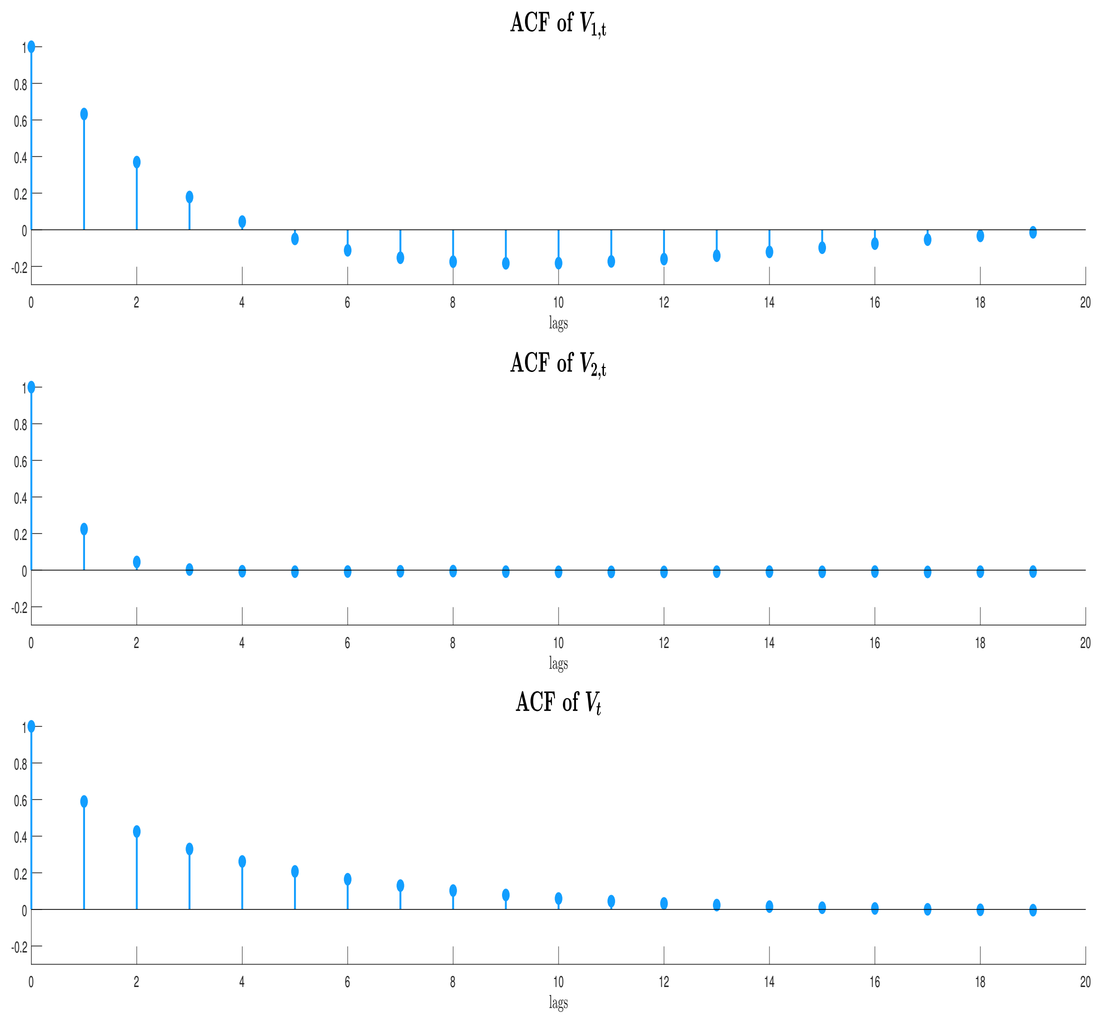

Figure 2: ACF of $V_{1,t}$ (top panel), ACF of $V_{2,t}$ (mid panel) and ACF of V_t (bottom panel) for model M1.

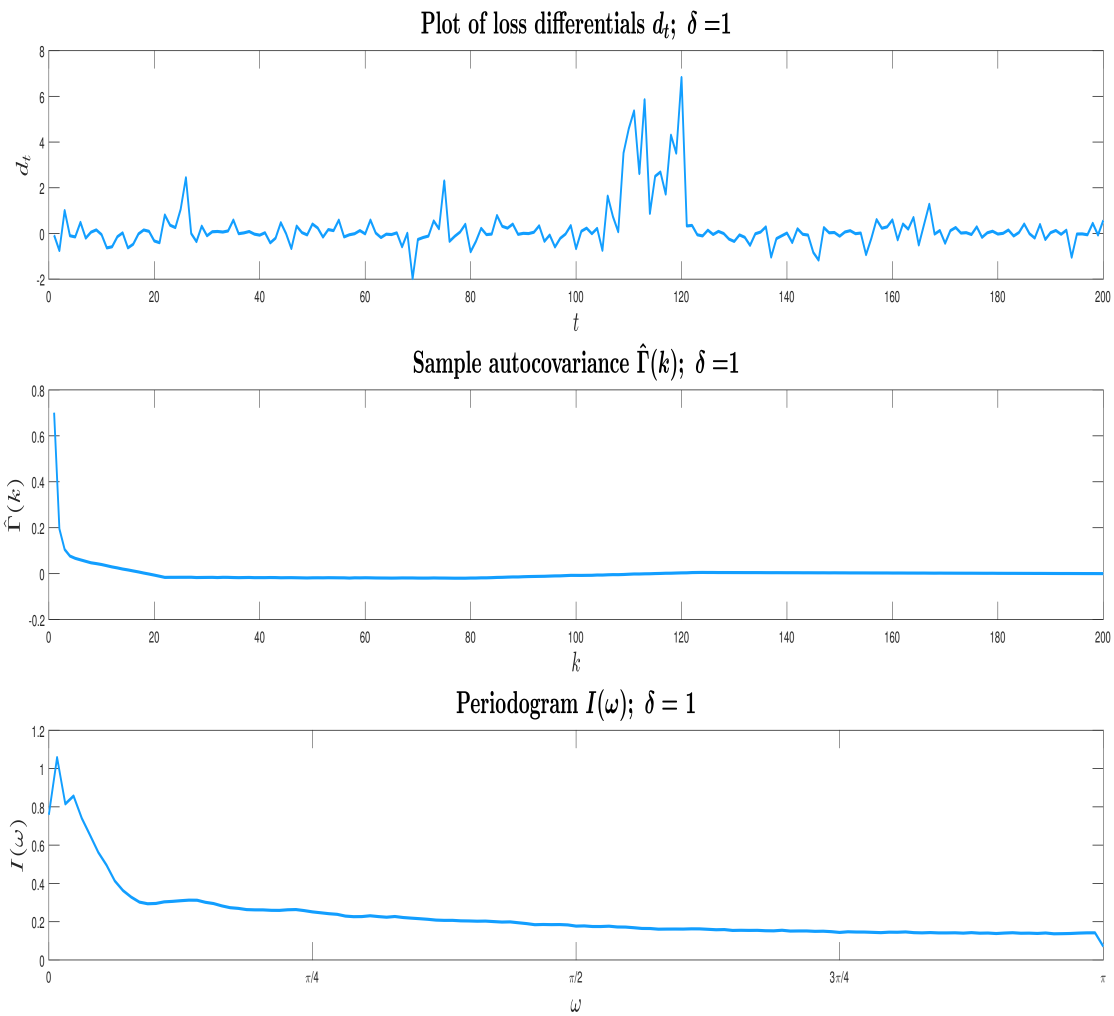

Figure 3: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 1$.

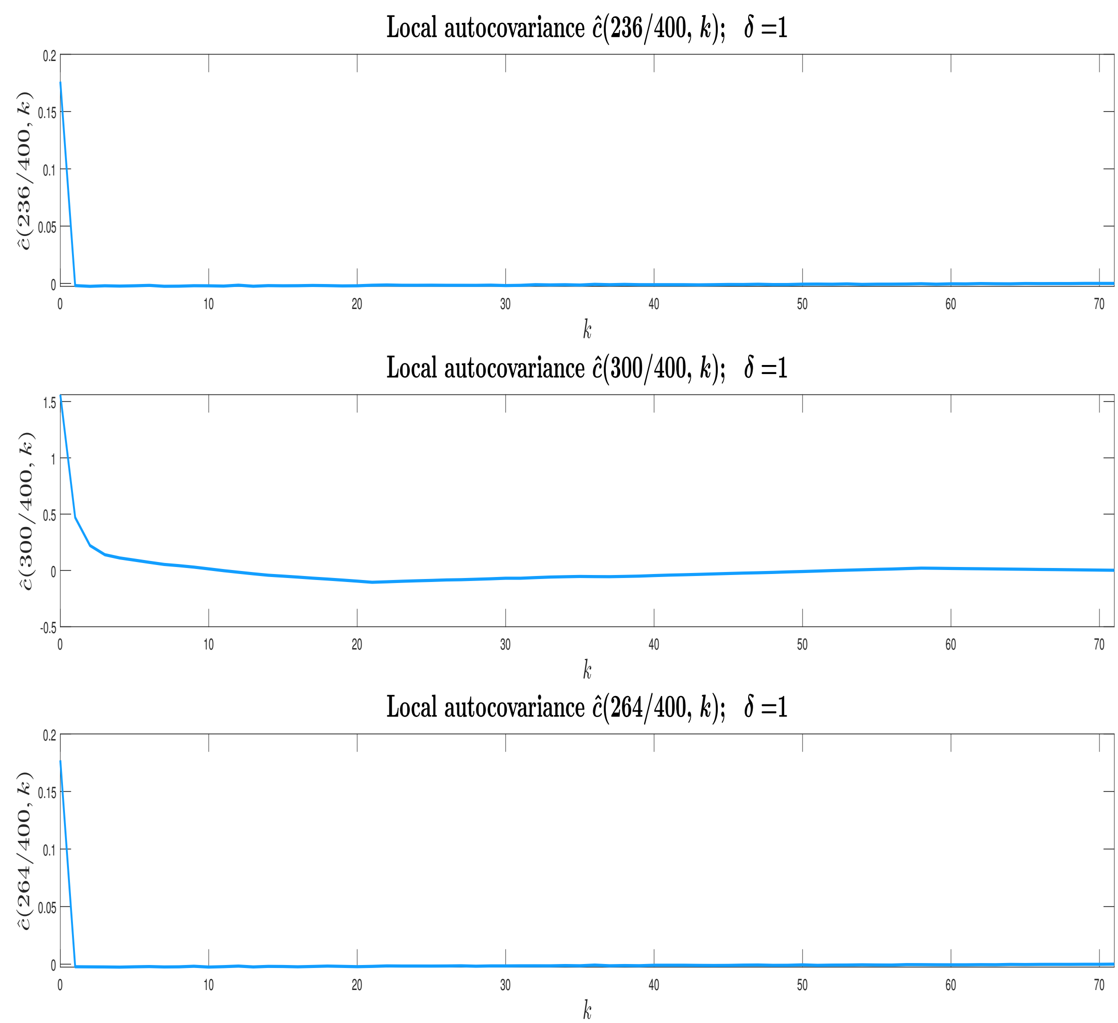

Figure 4: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 1$.

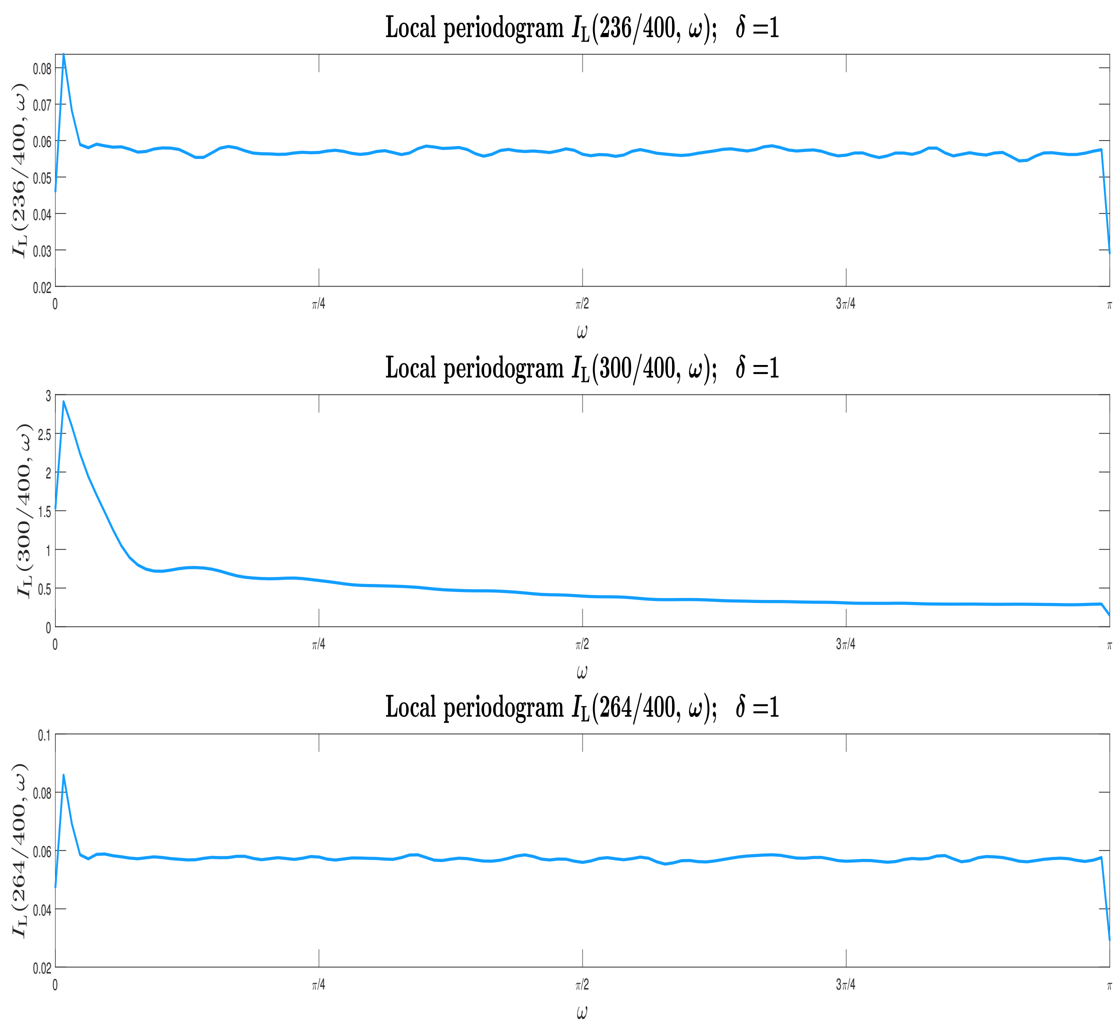

Figure 5: The figure plots $I_L(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 1$.

Table 1: Comparison between the theoretical autocovariance and the sample estimates

	$\Gamma_T(k)$	$\widehat{\Gamma}(k)$	$\widehat{\Gamma}(k) - \widehat{d}^*$	$\widehat{\Gamma}_{\mathrm{DK}}(k)$
$k = 0$	0.37	0.73	0.46	0.62
$k = 1$	0.13	0.49	0.21	0.35
$k = 2$	0.01	0.38	0.11	0.05
$k = 5$	0.00	0.22	-0.03	0.00
$k = 10$	-0.03	0.08	-0.13	0.00

The table reports the values of a) the theoretical autocovariance $\Gamma_T(k)$; b) the sample autocovariance $\widehat{\Gamma}(k)$; c) the corrected sample autocovariance $\widehat{\Gamma}(k) - \widehat{d}$; d) the double kernel sample autocovariance $\widehat{\Gamma}_{\mathrm{DK}}(k)$, for $k = 0$, 1, 2, 5 and 10, with the process $\{V_t\}$ as in model M1 with $T = 200$. The estimates in b)-d) were computed using 5,000 replications.

Table 2: Empirical small-sample size and power of t-test for model M1

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.25$	$\delta = 1$	$\delta = 1.5$
$\widehat{J}_{\mathrm{DK},T}$	0.068	0.189	0.286	0.460	0.661	0.992	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.045	0.085	0.199	0.332	0.612	0.976	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.046	0.090	0.202	0.333	0.613	0.977	1.000
Andrews (1991)	0.039	0.095	0.185	0.383	0.623	0.968	0.999
Andrews (1991), prewhite	0.115	0.168	0.304	0.447	0.650	0.988	0.999
Newey-West (1987)	0.209	0.272	0.398	0.516	0.689	0.997	1.000
KVB fixed-b	0.004	0.018	0.063	0.139	0.301	0.870	0.969
EWC	0.011	0.038	0.137	0.273	0.539	0.978	0.999

Table 3: Empirical small-sample size and power of the of t-test for model M2

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.3$	$\delta = 1$
$\widehat{J}_{\mathrm{DK},T}$	0.080	0.132	0.257	0.423	0.842	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.059	0.098	0.190	0.310	0.736	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.055	0.088	0.187	0.306	0.735	1.000
Andrews (1991)	0.081	0.133	0.266	0.433	0.838	1.000
Andrews (1991), prewhite	0.094	0.141	0.268	0.438	0.842	1.000
Newey-West (1987)	0.137	0.190	0.336	0.510	0.881	1.000
KVB fixed-b	0.014	0.036	0.078	0.203	0.561	0.990
EWC	0.032	0.064	0.157	0.299	0.712	1.000

Table 4: Empirical small-sample size and power of the of t-test for model M3

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.3$	$\delta = 1$
$\widehat{J}_{\mathrm{DK},T}$	0.117	0.363	0.537	0.928	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.049	0.227	0.384	0.865	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.052	0.223	0.374	0.855	1.000
Andrews (1991)	0.106	0.334	0.515	0.917	1.000
Andrews (1991), prewhite	0.122	0.351	0.524	0.928	1.000
Newey-West (1987)	0.169	0.412	0.596	0.948	1.000
KVB fixed-b	0.024	0.165	0.309	0.712	0.999
EWC	0.058	0.245	0.400	0.858	1.000

Table 5: Empirical small-sample size and power of the of t-test for model M4

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.1$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 1$	$\delta = 3$
$\widehat{J}_{\mathrm{DK},T}$	0.154	0.146	0.496	0.706	0.953	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.037	0.050	0.168	0.459	0.771	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.041	0.079	0.198	0.477	0.795	1.000
Andrews (1991)	0.127	0.162	0.398	0.623	0.905	0.999
Andrews (1991), prewhite	0.197	0.226	0.439	0.653	0.922	1.000
Newey-West (1987)	0.397	0.423	0.584	0.758	0.969	1.000
KVB fixed-b	0.005	0.012	0.135	0.339	0.709	0.964
EWC	0.115	0.147	0.367	0.681	0.906	0.999

Table 6: Empirical small-sample size and power of the DM (1995) test

		(1) $\delta > 0$				
$\alpha = 0.05$, $T = 200$	(null rejection)	$\delta = 0.2$	$\delta = 0.5$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$	0.033	0.312	0.551	0.997	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.042	0.322	0.563	0.999	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.046	0.348	0.573	0.998	1.000	1.000
Andrews (1991)	0.085	0.254	0.305	0.114	0.000	0.000
Andrews (1991), prewhite	0.085	0.246	0.293	0.401	0.045	0.000
Newey-West (1987)	0.083	0.246	0.299	0.612	0.817	0.782
KVB fixed-b	0.002	0.212	0.185	0.000	0.000	0.000
EWC	0.083	0.252	0.268	0.045	0.000	0.000
		(2) $\delta\,(t/T)$ locally stationary				
$\alpha = 0.05$, $T = 200$		$\delta = 0.2$	$\delta = 0.5$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.278	0.297	0.592	0.889	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.301	0.363	0.634	0969	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.327	0.368	0.642	0.969	1.000
Andrews (1991)		0.255	0.259	0.255	0.110	0.005
Andrews (1991), prewhite		0.249	0.243	0.268	0.188	0.031
Newey-West (1987)		0.281	0.282	0.313	0.268	0.078
KVB fixed-b		0.203	0.202	0.178	0.025	0.000
EWC		0.244	0.252	0.219	0.045	0.000
		(3) $\delta\,(t/T)$ locally stationary				
$\alpha = 0.05$, $T = 200$		$\delta = 0.2$	$\delta = 1$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.540	0.862	0.992	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.396	0.664	0.988	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.412	0.724	0.987	1.000	1.000
Andrews (1991)		0.328	0.234	0.235	0.241	0.777
Andrews (1991), prewhite		0.342	0.315	0.512	0.296	0.882
Newey-West (1987)		0.381	0.384	0.720	0.972	0.999
KVB fixed-b		0.100	0.032	0.000	0.002	0.040
EWC		0.312	0.152	0.142	0.296	0.852
		(4) case (2) with outliers				
$\alpha = 0.05$, $T = 400$		$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.694	0.733	0.822	0.981	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.724	0.777	0.846	0.982	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.727	0.771	0.847	0.981	1.000
Andrews (1991)		0.192	0.242	0.245	0.203	0.022
Andrews (1991), prewhite		0.182	0.233	0.243	0.288	0.114
Newey-West (1987)		0.222	0.271	0.245	0.345	0.225
KVB fixed-b		0.203	0.222	0.212	0.075	0.000
EWC		0.186	0.221	0.174	0.062	0.000

Supplement to “Theory of Low Frequency Contamination from Nonstationarity and Misspecication: Consequences for HAR Inference”

Alessandro Casini
University of Rome Tor Vergata

Taosong Deng
Hunan University

Pierre Perron
Boston University

22nd August 2023

Abstract

This supplemental material is for online publication only. Section S.A introduces the notion of long memory segmented locally stationary processes and presents the theoretical results referenced in Section 3. Section S.B contains the proofs of the results in the paper and Section S.C contains additional figures.

S.A Results on Low Frequency Bias for the Sample Autocovariance and the Periodogram

In Section S.A.1 we define the long memory SLS processes. In Section S.A.2 and S.A.3 we present results on the low frequency bias for the sample autocovariance and the periodogram, respectively.

S.A.1 Long Memory Segmented Locally Stationary Processes

Define the backward difference operator $\Delta V_t = \Delta^1 V_t = V_t - V_{t-1}$ and $\Delta^l V_t$ recursively. Long memory features can be expressed as a "pole" in the spectral density at frequency zero. That is, for a stationary process, long memory implies that $f(\omega) \sim \omega^{-2\vartheta}$ as $\omega \to 0$ where $\vartheta \in (0, 1/2)$ is the long memory parameter. In what follows, l is some non-negative integer.

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

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

for $j = 1, \ldots, m_0+1$, where by convention $T^0_0 = 0$ and $T^0_{m_0+1} = T$, (i) and (iii) of Definition 2.1 hold, and (ii) of Definition 2.1 is replaced by

(ii) There exists two constants $L_2 > 0$ and $D < 1/2$ (which depend on j) and a piecewise continuous function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0+1$, there exists a 2π-periodic function $A_j : (\lambda^0_{j-1}, \lambda^0_j] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$,

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

$$\sup_{1\leq j\leq m_0+1}\ \sup_{T^0_{j-1}<t\leq T^0_j,\,\omega} \left|A^0_{j,t,T}(\omega) - A_j(t/T, \omega)\right| \leq L_2 T^{-1} |\omega|^{-D}, \tag{S.A.3}$$

and

$$\sup_{0\leq v\leq u\leq 1,\, u\neq \lambda^0_j\, (j=1,\ldots,m_0+1,),\, \omega} |A(u, \omega) - A(v, \omega)| \leq L_2 |u-v|\, |\omega|^{-D}. \tag{S.A.4}$$

The spectral density of $\{V_{t,T}\}$ is given by $f_j(u, \omega) = |1-\exp(-i\omega)|^{-2l} |A_j(u, \omega)|^{-2}$ for $j = 1, \ldots, m_0+1$. We say that the process $\{V_{t,T}\}$ has local memory parameter $\vartheta(u) \in (-\infty, l+1/2)$ at time $u \in [0, 1]$ if it satisfies (S.A.1)-(S.A.4), and its generalized spectral density $f_j(u, \omega)$ $(j = 1, \ldots, m_0+1)$ satisfies the following condition,

$$f_j(u, \omega) = \left|1 - e^{-i\omega}\right|^{-2\vartheta_j(u)} f^*_j(u, \omega), \tag{S.A.5}$$

with $f^*_j(u, \omega) > 0$ and

$$\left|f^*_j(u, \omega) - f^*_j(u, 0)\right| \leq L_4 f^*_j(u, \omega) |\omega|^{\nu}, \qquad \omega \in [-\pi, \pi], \tag{S.A.6}$$

where $L_4 > 0$ and $\nu \in (0, 2]$.

Definition S.A.1 extends Definition 2.1 and Assumption 2.1 by requiring the bound on the smoothness of $A(\cdot, \omega)$ to depend also on $|\omega|^{-D}$ thereby allowing a singularity at $\omega = 0$. Casini (2022c) showed that $f_j(u, \omega) = |A_j(u, \omega)|^2$ for $j = 1, \ldots, m_0 + 1$. Using similar arguments, we obtain the form $f_j(u, \omega)$ given in (S.A.5). See Roueff and von Sachs (2011) for a definition of long memory local stationarity. Definition S.A.1 extends their definition to allow for m_0 discontinuities. We have assumed that breaks in the long memory parameter occur at the same locations as the breaks in the spectrum. This can be relaxed but would provide no added value in this paper.

Example S.A.1. A time-varying AR fractionally integrated moving average (p, ϑ, q) process with m_0 structural breaks satisfies Definition S.A.1 with $\vartheta_j : [0, 1] \to (-\infty, l + 1/2)$, $\sigma_j : [0, 1] \to \mathbb{R}_+$, $\phi_j = [\phi_1, \ldots, \phi_p]' : [0, 1] \to \mathbb{R}^q$ and $\theta_j = [\theta_1, \ldots, \theta_q]' : [0, 1] \to \mathbb{R}^p$ are left-Lipschitz functions for each $j = 1, \ldots, m_0 + 1$ such that $1 - \sum_{k=1}^{p} \phi_{j,k}(u) z^k$ does not vanish for all $u \in [0, 1]$ and $z \in \mathbb{C}$ such that $|z| \leq 1$. Using the latter condition, the local transfer function $A_j(u; \cdot)$ defines for each j a causal autoregressive fractionally integrated moving average (ARFIMA$(p, \vartheta(u) - l, q)$ process whose spectral density satisfies the conditions (S.A.5) and (S.A.6) with $\nu = 2$. Using Lemma 3 in Roueff and von Sachs (2011), condition (S.A.4) holds with with $D > \sup_{1 \leq j \leq m_0+1} \sup_{\lambda_{j-1}^0 < u \leq \lambda_j^0, \omega} \vartheta_j(u) - l$.

Definition S.A.1 implies that $\rho_V(u, k) \triangleq \mathrm{Corr}(V_{\lfloor Tu \rfloor}, V_{\lfloor Tu \rfloor + k}) \sim C k^{2\vartheta_j(u) - 1}$ for $\lambda_{j-1}^0 < u < \lambda_j^0$ and large k where $C > 0$. This means that the rescaled time-u autocorrelation function (ACF(u)) has a power law decay which implies $\sum_{k=-\infty}^{\infty} |\rho_V(u, k)| = \infty$ if $\vartheta_j(u) \in (0, 1/2)$.

S.A.2 The Sample Autocovariance Under Nonstationarity

We now establish some asymptotic properties of the sample autocovariance under nonstationarity. We consider the case $k \geq 0$ only; the case $k < 0$ is similar.

Theorem S.A.1. *Assume that $\{V_{t,T}\}$ satisfies Definition 2.1. Under Assumption 2.1-2.2,*

$$\widehat{\Gamma}(k) \geq \int_0^1 c(u, k)\, du + d^* + o_{\text{a.s.}}(1), \tag{S.A.7}$$

where $d^* = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\overline{\mu}_{j_2} - \overline{\mu}_{j_1})^2$. Further, as $k \to \infty$, $\widehat{\Gamma}(k) \geq d^*$ $\mathbb{P}$-a.s. If in addition it holds that $\mu_j(t/T) = \mu_j$ for $j = 1, \ldots, m_0 + 1$, then

$$\widehat{\Gamma}(k) = \int_0^1 c(u, k)\, du + d^*_{\text{Sta}} + o_{\text{a.s.}}(1),$$

where $d^*_{\text{Sta}} = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\mu_{j_2} - \mu_{j_1})^2$ and, as $k \to \infty$, $\widehat{\Gamma}(k) = d^*_{\text{Sta}} + o_{\text{a.s.}}(1)$.

S.A.3 The Periodogram Under Nonstationarity

Classical LRV estimators are weighted averages of periodogram ordinates around the zero frequency. Thus, it is useful to study the behavior of the periodogram as the frequency ω approaches zero. We now establish some properties of the asymptotic bias of the periodogram under nonstationarity. We consider the Fourier frequencies $\omega_l = 2\pi l/T \in (-\pi, \pi)$ for an integer $l \neq 0 \pmod T$ and exclude $\omega_l = 0$ for mathematical convenience.

Assumption S.A.1. *(i) For each $j = 1, \ldots, m_0 + 1$ there exists a $B_j \in \mathbb{R}$ such that*

$$\left| \sum_{j=1}^{m_0+1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor + 1}^{\lfloor T\lambda_j^0 \rfloor} \mu_j (t/T) \exp(-i\omega_l t) \right|^2 \geq \left| \sum_{j=1}^{m_0+1} B_j \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor + 1}^{\lfloor T\lambda_j^0 \rfloor} \exp(-i\omega_l t) \right|^2, \qquad \omega_l \in (-\pi, \pi),$$

where $B_{j_1} \neq B_{j_2}$ for $j_1 \neq j_2$; (ii) $|\Gamma(u, k)| = C_{u,k} k^{-m}$ for all $u \in [0, 1]$ and all $k \geq C_3 T^{\kappa}$ for some $C_3 < \infty$, $C_{u,k} < \infty$ (which depends on u and k), $0 < \kappa < 1/2$, and $m > 2$.

Part (i) is easily satisfied (e.g., the special case with $\mu_j(t/T) = \mu_j$). Part (ii) is satisfied if $\{V_t\}$ is strong mixing with mixing parameters of size $-2\nu/(\nu - 1/2)$ for some $\nu > 1$ such that $\sup_{t\geq 1} \mathbb{E}|V_t|^{4\nu} < \infty$. This is less stringent than the size condition $-3\nu/(\nu - 1)$ for some $\nu > 1$ sufficient for Assumption 2.2-(i).

Theorem S.A.2. *Assume that $\{V_{t,T}\}$ satisfies Definition 2.1. Under Assumption 2.1-2.2 and S.A.1,*

$$\mathbb{E}(I_T(\omega_l)) = 2\pi \int_0^1 f(u, \omega_l)\, du + \frac{1}{T\omega_l^2} \left| \left[B_1 - B_{m_0+1} - \sum_{j=1}^{m_0} (B_j - B_{j+1}) \exp\left(-2\pi i l \lambda_l^0\right) \right] \right|^2 + o(1). \tag{S.A.8}$$

Under Assumption 2.1-2.2 and S.A.1-(ii), if $\mu_j(t/T) = \mu_j$ for each $j = 1, \ldots, m_0 + 1$, then

$$\mathbb{E}(I_T(\omega_l)) = 2\pi \int_0^1 f(u, \omega_l)\, du + \frac{1}{T\omega_l^2} \left| \left[\mu_j - \mu_{m_0+1} - \sum_{j=1}^{m_0} (\mu_j - \mu_{j+1}) \exp\left(-2\pi i l \lambda_j^0\right) \right] \right|^2 + o(1).$$

In either case, if $T\omega_l^2 \to 0$ as $T \to \infty$ then $\mathbb{E}(I_T(\omega_l)) \to \infty$ for many values in $\{\omega_l\}$ as $\omega_l \to 0$.

The theorem suggests that for small frequencies ω_l close to 0, the periodogram attains very large values. This follows because the first term of (S.A.8) is bounded for all ω_j. Since $B_1, \ldots, B_{m_0+1}$ are fixed, the order of the second term of (S.A.8) is $O((T\omega_j^2)^{-1})$. Note that as $\omega_l \to 0$ there are some values l for which the corresponding term involving $|\cdot|^2$ on the right-hand side of (S.A.8) is equal to zero. In such cases, $\mathbb{E}(I_T(\omega_l)) \geq 2\pi \int_0^1 f(u, \omega_l)\, du > 0$. For other values of $\{l\}$ as $\omega_l \to 0$, the second term of (S.A.8) diverges to infinity. Thus, considering the behavior of $\{\mathbb{E}(I_T(\omega_l))\}$ as $\omega_l \to 0$, it generally takes unbounded values except for some ω_l for which $\mathbb{E}(I_T(\omega_l))$ is bounded below by $2\pi \int_0^1 f(u, \omega_l)\, du > 0$. A SLS process with long memory has an unbounded local spectral density $f(u, \omega)$ as $\omega \to 0$ for some $u \in [0, 1]$. Since $f(\cdot, \cdot)$ cannot be negative, it follows that $\int_0^1 f(u, \omega)\, du$ is also unbounded as $\omega \to 0$. Theorem S.A.2 suggests that nonstationarity consisting of time-varying first moment results in a periodogram sharing features of a long memory series.

S.B Mathematical Appendix

S.B.1 Proofs of the Results in Section 3 and 3

S.B.1.1 Proof of Theorem S.A.1

Let $\overline{V}_j = (Tr_j)^{-1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t$, $\mu_{2,j}(u) = \mathbb{E}(V_{\lfloor Tu \rfloor})^2$ for $T_{j-1}^0 \leq Tu \leq T_j^0$ and $\overline{\mu}_{2,j} = r_j^{-1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \mu_{2,j}(u)\, du$. By Assumption 2.1-2.2-(i), the latter implying ergodicity, it follows for fixed $k \geq 0$ that

$$
\begin{aligned}
\widehat{\Gamma}(k) &= \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} V_t V_{t-k} - \left(\sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t \right)^2 \\
&= \sum_{j=1}^{m_0+1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mathbb{E}(V_t)\, \mathbb{E}(V_{t-k}) \\
&\quad - \left(\sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1) \\
&= \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mathbb{E}(V_t)\, \mathbb{E}(V_{t-k}) \\
&\quad - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1) \\
&= \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mu^2(t/T) - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1),
\end{aligned}
$$

where we have used $\mathbb{E}(V_{t-k}) - \mathbb{E}(V_t) = O(k/T)$ by local stationarity in the third equality. Note that by ergodicity and an approximation to Riemann sums, we have

$$
\begin{aligned}
\sum_{j=1}^{m_0+1} r_j \overline{V}_j - \sum_{j=1}^{m_0+1} r_j \overline{\mu}_j &= \sum_{j=1}^{m_0+1} r_j \overline{V}_j - \sum_{j=1}^{m_0+1} r_j \mathbb{E}\left(\overline{V}_j\right) + \sum_{j=1}^{m_0+1} r_j \mathbb{E}\left(\overline{V}_j\right) - \sum_{j=1}^{m_0+1} r_j \overline{\mu}_j \\
&= o_{\text{a.s.}}(1) + O\left(T^{-1}\right).
\end{aligned} \tag{S.B.1}
$$

Basic manipulations show that

$$
\begin{aligned}
&\sum_{j_2 \neq j_1} r_{j_1} r_{j_2} \left(\overline{\mu}_{j_2} - \overline{\mu}_{j_1}\right)^2 \\
&\quad = \sum_{j_2 \neq j_1} r_{j_1} r_{j_2} \left(\overline{\mu}_{j_2}^2 + \overline{\mu}_{j_1}^2 - 2\overline{\mu}_{j_2}\overline{\mu}_{j_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{1\le j_2\le m_0+1} r_{j_2}\overline{\mu}_{j_2}^2\left(1-r_{j_2}\right) + \sum_{1\le j_1\le m_0+1} r_{j_1}\overline{\mu}_{j_1}^2\left(1-r_{j_1}\right) - 2\sum_{j_1\neq j_2} r_{j_1}r_{j_2}\overline{\mu}_{j_2}\overline{\mu}_{j_1} \\
&= 2\sum_{1\le j\le m_0+1} r_j\overline{\mu}_j^2 - 2\sum_{1\le j\le m_0+1} r_j^2\overline{\mu}_j^2 - 2\sum_{j_1\neq j_2} r_{j_1}r_{j_2}\overline{\mu}_{j_2}\overline{\mu}_{j_1}. \qquad \text{(S.B.2)}
\end{aligned}
$$

Note that

$$
\left(Tr_j-k\right)\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu^2\left(t/T\right) \ge \left(\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu\left(t/T\right)\right)^2. \qquad \text{(S.B.3)}
$$

Thus,

$$
\begin{aligned}
\sum_{j=1}^{m_0+1} r_j\frac{1}{Tr_j}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu^2\left(t/T\right) &= \sum_{j=1}^{m_0+1} r_j\frac{1}{Tr_j\left(Tr_j-k\right)}\left(Tr_j-k\right)\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu^2\left(t/T\right) \\
&\ge \sum_{j=1}^{m_0+1} r_j\frac{1}{Tr_j\left(Tr_j-k\right)}\left(\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu\left(t/T\right)\right)^2 \\
&= \sum_{1\le j\le m_0+1} r_j\overline{\mu}_j^2 + o\left(1\right). \qquad \text{(S.B.4)}
\end{aligned}
$$

Using (S.B.1)-(S.B.4) we have,

$$
\begin{aligned}
\widehat{\Gamma}\left(k\right) &= \int_0^1 c\left(u,\,k\right)du + \sum_{j=1}^{m_0+1} r_j\frac{1}{Tr_j}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1+k}^{\lfloor T\lambda_j^0\rfloor}\mu^2\left(t/T\right) - \left(\sum_{j=1}^{m_0+1} r_j\overline{V}_j\right)^2 + o_{\mathrm{a.s.}}\left(1\right) \\
&\ge \int_0^1 c\left(u,\,k\right)du + \sum_{j=1}^{m_0+1} r_j\overline{\mu}_{2,j} - \left(\sum_{j=1}^{m_0+1} r_j\overline{V}_j\right)^2 + O\left(T^{-1}\right) + o_{\mathrm{a.s.}}\left(1\right) \\
&= \int_0^1 c\left(u,\,k\right)du + 2^{-1}\sum_{j_1\neq j_2} r_{j_1}r_{j_2}\left(\overline{\mu}_{j_2}-\overline{\mu}_{j_1}\right)^2 + O\left(T^{-1}\right) + o_{\mathrm{a.s.}}\left(1\right). \qquad \text{(S.B.5)}
\end{aligned}
$$

The claim that $\widehat{\Gamma}\left(k\right) \ge d$ $\mathbb{P}$-a.s. as $k\to\infty$ follows from Assumption 2.2-(i) since this implies that $c\left(u,\,k\right)\to 0$ as $k\to\infty$ and from the fact that the second term on the right-hand side of (S.B.5) does not depend on k. If in addition it holds that $\mu_j\left(t/T\right)=\mu_j$ for $j=1,\ldots,m_0+1$, then (S.B.3) holds with equality and the result follows as a special case of (S.B.5). □

S.B.1.2 Proof of Theorem S.A.2

Lemma S.B.1. *Assume that $\{V_{t,T}\}$ satisfies Definition 2.1. Under Assumption 2.1-2.2 and S.A.1-(ii),*

$$\sum_{j_1 \neq j_2} \frac{1}{T} \sum_{t=\lfloor T\lambda^0_{j_1-1}\rfloor+1}^{\lfloor T\lambda^0_{j_1}\rfloor} \sum_{s=\lfloor T\lambda^0_{j_2-1}\rfloor+1}^{\lfloor T\lambda^0_{j_2}\rfloor} \mathbb{E}\left(\left(V_t - \mu\left(t/T\right)\right)\left(V_s - \mu\left(s/T\right)\right)\right)\exp\left(-i\omega_l\left(t-s\right)\right) = o\left(1\right).$$

Proof. Let $\overline{r}_{j_1,j_2} = \max\{r_{j_1}, r_{j_2}\}$ and $\underline{r}_{j_1,j_2} = \min\{r_{j_1}, r_{j_2}\}$. We consider the case of adjacent regimes (i.e., $j_2 = j_1 + 1$) which also provides an upper bound for non-adjacent regimes due to the short memory property. For any $k = s - t = 1, \ldots, \lfloor T\underline{r}_{j_1,j_2}\rfloor$ there are k pairs in the above sum. The double sum above (over t and s) can be split into

$$\begin{aligned} & T^{-1}\sum_{k=1}^{\lfloor CT^{\kappa}\rfloor}\left|\Gamma_{\{1:\lfloor CT^{\kappa}\rfloor\}}\left(\cdot,\,k\right)\right| + T^{-1}\sum_{k=\lfloor CT^{\kappa}\rfloor+1}^{\lfloor hT\rfloor}\left|\Gamma_{\{\lfloor CT^{\kappa}\rfloor+1:\lfloor hT\rfloor\}}\left(\cdot,\,k\right)\right| \\ & \quad + T^{-1}\sum_{k=\lfloor hT\rfloor+1}^{\lfloor T\underline{r}_{j_1,j_2}\rfloor-1}\left|\Gamma_{\left\{\lfloor hT\rfloor+1:\lfloor T\underline{r}_{j_1,j_2}\rfloor-1\right\}}\left(\cdot,\,k\right)\right| + T^{-1}\sum_{k=\lfloor T\underline{r}_{j_1,j_2}\rfloor}^{\lfloor T\overline{r}_{j_1,j_2}\rfloor}\left|\Gamma_{\left\{\underline{r}_{j_1,j_2}:\underline{r}_{j_1,j_2}\right\}}\left(\cdot,\,k\right)\right| \end{aligned} \tag{S.B.6}$$

where $C > 0$, $0 < h < 1$ with $\lfloor hT\rfloor < \lfloor T\underline{r}_{j_1,j_2}\rfloor - 1$, and $\Gamma_S(\cdot,\,k)$ is the sum of the autocovariances at lag k computed at the time points corresponding to $k \in S$. Note that the term $|\exp(-i\omega_l(\pm k))|$ can be bounded by some constant. The sums run over only $k > 0$ because by symmetry $\Gamma_u(k) = \Gamma_{u-k/T}(-k)$. Consider the first sum in (S.B.6). This is of order $O\left(T^{-1}T^{2\kappa}\right)$ which goes to zero given $\kappa < 1/2$. The second sum is also negligible using the following arguments. By Assumption S.A.1-(ii), $|\Gamma(u,\,k)| = C_{u,k}k^{-m}$ with $m > 2$ and choosing C large enough yields that the second sum of (S.B.6) converges to zero. In the third sum, the number of summands grows at rate $O(T)$ and for each lag k there are $O(T)$ autocovariances. However, by Assumption S.A.1-(ii) each autocovariance is $O\left(T^{-m}\right)$. Thus, the bound is $O\left(T^{-1}T^{2-m}\right)$ which goes to zero as $T \to \infty$. The difference between the arguments used for the third sum and fourth sums is that now we do not have $O(T)$ autocovariances for each lag k. Thus, the bound for the fourth sum cannot be greater than the bound for the third sum. Thus, the fourth sum also converges to zero. □

Proof of Theorem S.A.2. We have,

$$\begin{aligned} I_T\left(\omega_l\right) &= \left|\frac{1}{\sqrt{T}}\sum_{j=1}^{m_0+1}\sum_{t=\lfloor T\lambda^0_{j-1}\rfloor+1}^{\lfloor T\lambda^0_j\rfloor}\exp\left(-i\omega_l t\right)V_t\right|^2 \\ &= \left|\frac{1}{\sqrt{T}}\sum_{j=1}^{m_0+1}\sum_{t=\lfloor T\lambda^0_{j-1}\rfloor+1}^{\lfloor T\lambda^0_j\rfloor}\left(X_t - \mu\left(t/T\right)\right)\exp\left(-i\omega_l t\right) + \frac{1}{\sqrt{T}}\sum_{j=1}^{m_0+1}\sum_{t=\lfloor T\lambda^0_{j-1}\rfloor+1}^{\lfloor T\lambda^0_j\rfloor}\mu\left(t/T\right)\exp\left(-i\omega_l t\right)\right|^2. \end{aligned}$$

From Assumption S.A.1,

$$\begin{aligned}
&\left|\sum_{j=1}^{m_0+1}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1}^{\lfloor T\lambda_j^0\rfloor}\mu(t/T)\exp(-i\omega_l t)\right|^2\\
&\quad\geq\left|\sum_{j=1}^{m_0+1}B_j\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1}^{\lfloor T\lambda_j^0\rfloor}\exp(-i\omega_l t)\right|^2\\
&\quad=\left|\sum_{j=1}^{m_0+1}B_j\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor+1\right)\right)\sum_{t=0}^{\lfloor T\lambda_j^0\rfloor-\lfloor T\lambda_{j-1}^0\rfloor-1}\exp(-i\omega_l t)\right|^2\\
&\quad=\left|\frac{\exp(-i\omega_l)}{1-\exp(-i\omega_l)}\sum_{j=1}^{m_0+1}B_j\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)\left(1-\exp\left(-i\omega_l\left(\left\lfloor T\lambda_j^0\right\rfloor-\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)\right)\right|^2\\
&\quad=\left|\frac{\exp(-i\omega_l)}{1-\exp(-i\omega_l)}\sum_{j=1}^{m_0+1}B_j\left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)-\exp\left(-i\omega_l\left\lfloor T\lambda_j^0\right\rfloor\right)\right)\right|^2,
\end{aligned}$$

using the formula for the first n-th terms of a geometric series $\sum_{k=0}^{n-1}ar^k=a\sum_{k=0}^{n-1}r^k=a(1-r^n)/(1-r)$. Then, using summation by parts,

$$\begin{aligned}
&\frac{\exp(-i\omega_j)}{1-\exp(-i\omega_j)}\sum_{j=1}^{m_0+1}B_j\left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)-\exp\left(-i\omega_l\left\lfloor T\lambda_j^0\right\rfloor\right)\right)\\
&\qquad\qquad=\frac{\exp(-i\omega_j)}{1-\exp(-i\omega_j)}\left[B_1-B_{m_0+1}-\sum_{j=1}^{m_0}(B_j-B_{j+1})\exp\left(-i\omega_l\left\lfloor T\lambda_j^0\right\rfloor\right)\right].
\end{aligned}$$

By Lemma S.B.1, it is sufficient to consider the cross-products within each regime j,

$$\begin{aligned}
\mathbb{E}(I_T(\omega_l))\geq&\sum_{j=1}^{m_0+1}r_j\frac{1}{Tr_j}\mathbb{E}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1}^{\lfloor T\lambda_j^0\rfloor}\sum_{s=\lfloor T\lambda_{j-1}^0\rfloor+1}^{\lfloor T\lambda_j^0\rfloor}(V_t-\mu(t/T))(V_s-\mu(s/T))\exp(-i\omega_l(t-s))\\
&+\sum\sum_{j_1\neq j_2}\frac{1}{T}\mathbb{E}\sum_{t=\lfloor T\lambda_{j_1-1}^0\rfloor+1}^{\lfloor T\lambda_{j_1}^0\rfloor}\sum_{s=\lfloor T\lambda_{j_2-1}^0\rfloor+1}^{\lfloor T\lambda_{j_2}^0\rfloor}(V_t-\mu(t/T))(V_s-\mu(s/T))\exp(-i\omega_l(t-s))\\
&+\left|\frac{1}{\sqrt{T}}\frac{\exp(-i\omega_l)}{1-\exp(-i\omega_l)}\sum_{j=1}^{m_0+1}B_j\left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)-\exp\left(-i\omega_l\left\lfloor T\lambda_j^0\right\rfloor\right)\right)\right|^2+o(1)\\
=&\sum_{j=1}^{m_0+1}\left(\mathbb{E}\frac{1}{T}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+1}^{\lfloor T\lambda_j^0\rfloor}(V_t-\mu(t/T))^2+\frac{2}{Tr_j}\sum_{k=1}^{\lfloor Tr_j\rfloor-1}\sum_{t=\lfloor T\lambda_{j-1}^0\rfloor+k+1}^{\lfloor T\lambda_j^0\rfloor}\Gamma_{t/T}(k)\exp(-i\omega_l k)\right)
\end{aligned}$$

$$+\left|\frac{1}{\sqrt{T}}\frac{\exp\left(-i\omega_l\right)}{1-\exp\left(-i\omega_l\right)}\sum_{j=1}^{m_0+1}B_j\left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0\right\rfloor\right)\right)-\exp\left(-i\omega_l\left\lfloor T\lambda_j^0\right\rfloor\right)\right)\right|^2+o\left(1\right).$$

Next, using the definition of $f\left(u,\,\omega_l\right)$, $e^{-2i\omega_l}=1$ by Euler's formula and letting $\omega_l\to 0$ we have,

$$\begin{aligned}\mathbb{E}\left(I_T\left(\omega_l\right)\right)\geq&\sum_{j=1}^{m_0+1}\left(\int_{\lambda_{j-1}^0}^{\lambda_j^0}c\left(u,\,0\right)du+2\sum_{k=1}^{\infty}\int_{\lambda_{j-1}^0}^{\lambda_j^0}c\left(u,\,k\right)\exp\left(-i\omega_lk\right)du\right)\\&+\frac{1}{T}\frac{1}{\left|1-\exp\left(-i\omega_l\right)\right|^2}\left|\left[B_1-B_{m_0+1}-\left(1+o\left(1\right)\right)\sum_{j=1}^{m_0}\left(B_j-B_{j+1}\right)\exp\left(-2\pi il\lambda_j^0\right)\right]\right|^2+o\left(1\right)\\=&\,2\pi\sum_{j=1}^{m_0+1}\int_{\lambda_{j-1}^0}^{\lambda_j^0}f\left(u,\,\omega_l\right)du\\&+\frac{1}{T}\frac{1}{\left|1-\exp\left(-i\omega_l\right)\right|^2}\left|\left[B_1-B_{m_0+1}-\left(1+o\left(1\right)\right)\sum_{j=1}^{m_0}\left(B_j-B_{j+1}\right)\exp\left(-2\pi il\lambda_j^0\right)\right]\right|^2+o\left(1\right)\\=&\,2\pi\int_0^1f\left(u,\,\omega_l\right)du+\frac{1}{T\omega_l^2}\left|\left[B_1-B_{m_0+1}-\sum_{j=1}^{m_0}\left(B_j-B_{j+1}\right)\exp\left(-2\pi il\lambda_j^0\right)\right]\right|^2+o\left(1\right).\end{aligned}\tag{S.B.7}$$

By Assumption 2.1-(ii), the first term of (S.B.7) is bounded for all frequencies ω_j. Since $B_1,\ldots,\,B_{m_0+1}$ are fixed, if $T\omega_l^2\to 0$ then the order of the second term of (S.B.7) is $O((T\omega_l^2)^{-1})$. Note that as $\omega_l\to 0$ there are some values of l for which the corresponding term involving $|\cdot|^2$ on the right-hand side of (S.B.7) is equal to zero [see the argument in Mikosch and Stărica (2004)]. In such a case, $\mathbb{E}\left(I_T\left(\omega_l\right)\right)\geq 2\pi\int_0^1 f\left(u,\,\omega_l\right)du>0$. For the other values of $\{l\}$ as $\omega_l\to 0$, the second term of (S.B.7) diverges to infinity. The outcome is that there are frequencies close to $\omega_l=0$ for which $\mathbb{E}\left(I_T\left(\omega_l\right)\right)\to\infty$. □

S.B.1.3 Proof of Theorem 3.1

We consider the case $k\geq 0$. The case $k<0$ follows similarly. Consider any $u\in(0,\,1)$ such that $T_j^0\notin\mathbf{S}\left(u,\,k,\,n_{2,T}\right)$ for all $j=1,\ldots,\,m_0$. Theorem S.B.3 in Casini (2022c) showed that

$$\mathbb{E}\left[\widehat{c}_T\left(u,\,k\right)\right]=c\left(u_0,\,k\right)+\frac{1}{2}\left(n_{2,T}/T\right)^2\left[\frac{\partial^2}{\partial^2u}c\left(u,\,k\right)\right]+o\left(\left(n_{2,T}/T\right)^2\right)+O\left(1/n_{2,T}\right).\tag{S.B.8}$$

Since $n_{2,T}\to\infty$ and $n_{2,T}/T\to 0$, $\mathbb{E}\left[\widehat{c}_T\left(u,\,k\right)\right]=c\left(u_0,\,k\right)+o\left(1\right)$. The same aforementioned theorem shows that $n_{2,T}\mathrm{Var}\left[\widehat{c}_T\left(u,\,k\right)\right]=O_{\mathbb{P}}\left(1\right)$. This combined with (S.B.8) yields part (i) of the theorem.

Next, we consider case (ii-a) with $n_{j,L}\left(u,\,k,\,n_{2,T}\right)/n_{2,T}\to\gamma\in(0,\,1)$. We have,

$$\widehat{c}_T\left(u,\,k\right)=n_{2,T}^{-1}\sum_{s=0}^{n_{2,T}}V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}-\left(n_{2,T}^{-1}\sum_{s=0}^{n_{2,T}}V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1}\right)^2$$

$$
\begin{aligned}
&= n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k} \\
&\quad + n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k} \\
&\quad - \left(n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1} \right. \\
&\quad \left. + n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1} \right)^2 \\
&= n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} \left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k} \right. \\
&\quad \left. - \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right) \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}\right) \right) \\
&\quad + n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} \left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1} V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k} \right. \\
&\quad \left. - \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right) \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}\right) \right) \\
&\quad + n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right) \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}\right) \\
&\quad + n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right) \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}\right) \\
&\quad - \left(n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1} \right. \qquad \text{(S.B.9)} \\
&\quad \left. + n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1} \right)^2 + o_{\mathbb{P}}(1) \\
&\geq \gamma c\left(\lambda_j^0,\, k\right) + (1-\gamma)\, c(u,\, k) + \gamma \mu_j \left(\lambda_j^0\right)^2 + (1-\gamma)\, \mu_{j+1}(u)^2 \\
&\quad - \left(\gamma \mu_j\left(\lambda_j^0\right) + (1-\gamma)\, \mu_{j+1}(u)\right)^2 + o_{\mathbb{P}}(1) \\
&= \gamma c\left(\lambda_j^0,\, k\right) + (1-\gamma)\, c(u,\, k) + \gamma(1-\gamma)\left(\mu_j\left(\lambda_j^0\right) - \mu_{j+1}(u)\right)^2 + o_{\mathbb{P}}(1). \qquad \text{(S.B.10)}
\end{aligned}
$$

Consider the case (ii-b) with $n_{j,L}(u,\, k,\, n_{2,T})/n_{2,T} \to 0$. The other sub-case follows by symmetry. Eq. (S.B.9) continues to hold. The first term, third term and the first summation of the last term on the right-hand side of (S.B.9) are negligible. Thus, using ergodicity, implied by Assumption 2.1-2.2-(i),

$$\begin{aligned}\widehat{c}_T(u,\,k) &= c(u,\,k) + n_{2,T}^{-1}\sum_{s=T_j^0-(\lfloor Tu\rfloor+k/2-n_{2,T}/2)}^{n_{2,T}}\mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right)\mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right)\\ &\quad-\mu_{j+1}(u)^2+o_{\mathbb{P}}(1)\\ &= c(u,\,k)+\mu_{j+1}(u)^2-\mu_{j+1}(u)^2+o_{\mathbb{P}}(1)=c(u,\,k)+o_{\mathbb{P}}(1),\end{aligned}$$

where we have used the smoothness of $\mathbb{E}(V_t)$ implied by local stationarity. The second claim of the lemma follows from Assumption 2.2-(i) since this implies that $\sup_{u\in[0,\,1]} c(u,\,k)\to 0$ as $k\to\infty$ and the fact that the third term on the right-hand side of (S.B.10) does not depend on k. Thus, $\widehat{\Gamma}_{\mathrm{DK}}(k)\geq d_T^*+o_{\mathbb{P}}(1)$ where $d_T^*=(n_{2,T}/T)\,\gamma(1-\gamma)\left(\mu_j\left(\lambda_j^0\right)-\mu_{j+1}(u)\right)^2>0$ and $d_T^*\to 0$ since $n_{2,T}/T\to 0$. The factor $n_{2,T}/T$ in d_T^* follows because the neighborhood $(\lambda_j^0-n_{2,T}/T,\,\lambda_j^0+n_{2,T}/T)$ includes $O(n_{2,T}/n_T)$ blocks which are then averaged out. □

S.B.1.4 Proof of Theorem 3.2

Consider first any $u\in(0,\,1)$ such that $T_j^0\notin\mathbf{S}(u,\,0,\,n_T)$ for all $j=1,\ldots,\,m_0$. Theorem 3.3 in Casini and Perron (2021a) shows that

$$\begin{aligned}\mathbb{E}\left(I_{\mathrm{L},T}(u,\,\omega_l)\right) &= \left|\frac{1}{\sqrt{n_T}}\sum_{s=0}^{n_T-1}V_{\lfloor Tu\rfloor-n_T/2+s+1,T}\exp(-i\omega_l s)\right|^2\\ &= f(u,\,\omega_l)+\frac{1}{6}\left(\frac{n_T}{T}\right)^2\frac{\partial^2}{\partial u^2}f(u,\,\omega_l)+o\left(\left(\frac{n_T}{T}\right)^2\right)+O\left(\frac{\log(n_T)}{n_T}\right).\end{aligned}\tag{S.B.11}$$

By Assumption 2.1 the absolute value of the first term on the right-hand side is bounded for all frequencies ω_l. By Assumption 3.1-(iii) $\left|\left(\partial^2/\partial u^2\right)f(u,\,\omega_l)\right|$ is bounded and, since $n_T/T\to 0$, the second term converges to zero. Similarly, the third and fourth terms are negligible. Thus, $\mathbb{E}\left(I_{\mathrm{L},T}(u,\,\omega_l)\right)$ is bounded below by $f(u,\,\omega_l)>0$ as $\omega_l\to 0$ which establishes part (i). Now we consider to part (ii). We begin with case (a). We only focus on the sub-case $n_{j,L}(u,\,0,\,n_T)/n_T\to\gamma$ with $\gamma\in(0,\,1)$. We have

$$\begin{aligned}&I_{\mathrm{L},T}(\omega_l)=\\ &\left|\frac{1}{\sqrt{n_T}}\left(\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}V_{\lfloor Tu\rfloor-n_T/2+s+1,T}\exp(-i\omega_l s)+\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}V_{\lfloor Tu\rfloor-n_T/2+s+1,T}\exp(-i\omega_l s)\right|^2\right.\\ &=\frac{1}{n_T}\left|\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\left(V_{\lfloor Tu\rfloor-n_T/2+s+1,T}-\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\right)\exp(-i\omega_l s)\right.\\ &\quad+\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\left(V_{\lfloor Tu\rfloor-n_T/2+s+1,T}-\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\right)\exp(-i\omega_l s)\\ &\quad\left.+\sum_{s=0}^{n_T-1}\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\exp(-i\omega_l s)\right|^2.\end{aligned}\tag{S.B.12}$$

Using Assumption 3.1, we have

$$\left|\sum_{s=0}^{n_T-1}\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\exp\left(-i\omega_l s\right)\right|^2\geq$$
$$\left|B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)+B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\right|^2. \tag{S.B.13}$$

Note that

$$\begin{aligned}&B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)+B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\\&\quad=B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)\\&\qquad+B_{j+1}\exp\left(-i\omega_l\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)\right)\sum_{s=0}^{n_T-1-\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)}\exp\left(-i\omega_l s\right).\end{aligned} \tag{S.B.14}$$

Focusing on the second term on the right-hand side above,

$$\begin{aligned}&n_T^{-1}\left|B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\right|^2\\&=n_T^{-1}\left|B_{j+1}\exp\left(-i\omega_l\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)\right)\sum_{s=0}^{n_T-1-\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)}\exp\left(-i\omega_l s\right)\right|^2\\&=n_T^{-1}\left|B_{j+1}\exp\left(-i\omega_l\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)\right)\frac{1-\exp\left(-i\omega_l\left(n_T-\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)\right)\right)}{1-\exp\left(-i\omega_l\right)}\right|^2\\&=n_T^{-1}\left|B_{j+1}\frac{\exp\left(-i\omega_l\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)\right)-\exp\left(-i\omega_l n_T\right)}{1-\exp\left(-i\omega_l\right)}\right|^2.\end{aligned} \tag{S.B.15}$$

We show that the above equation diverges to infinity as $\omega_l\to 0$ with $n_T\omega_l^2\to 0$. If $n_T\omega_l\to a\in(0,\infty)$ then $\mathrm{Re}\left(\exp\left(-i\omega_l n_T\right)\right)\neq 1$ and the order is determined by the denominator. As in the proof of Theorem S.A.2, $|1-\exp(-i\omega_l)|^2=\omega_l^2$. Since $n_T\omega_l^2\to 0$, the right-hand side above diverges. If $n_T\omega_l\to 0$, we apply L'Hôpital's rule to obtain

$$n_T^{-1}\left|B_{j+1}\frac{-i\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)+in_T}{i}\right|^2$$

$$
\begin{aligned}
&= n_T^{-1} B_{j+1}^2 \left(-\left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right)^2 + n_T^2 - \left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right) n_T \right) \\
&= O\left(n_T^2/n_T\right) = O\left(n_T\right),
\end{aligned}
$$

which shows that the right-hand side of (S.B.15) diverges. A similar argument can be applied to the first term on the right-hand side of (S.B.14) and to the product of the latter term and the complex conjugate of the second term on the right-hand side of (S.B.14).

It remains to consider case (b) and the sub-case $n_{j,L}(u,\, 0,\, n_T)/n_T \to 0$. The other sub-case follows by symmetry. We have (S.B.12) and (S.B.13). Note that,

$$
\begin{aligned}
&\left| \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=T_j^0-(\lfloor Tu \rfloor - n_T/2)}^{n_T-1} \exp\left(-i\omega_l s\right) \right|^2 \\
&= \left| \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{n_T-1} \exp\left(-i\omega_l s\right) - \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{T_j^0-(\lfloor Tu \rfloor - n_T/2)-1} \exp\left(-i\omega_l s\right) \right|^2 \\
&= \left| -\frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{T_j^0-(\lfloor Tu \rfloor - n_T/2)-1} \exp\left(-i\omega_l s\right) \right|^2 \to 0.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}\left(I_{\mathrm{L}T}\left(\omega_l\right)\right) = \frac{1}{n_T} \Bigg| &\left(\sum_{s=0}^{T_j^0-(\lfloor Tu \rfloor - n_T/2+1)} \left(V_{\lfloor Tu \rfloor - n_T/2+s+1,T} - \mu\left(\left(\lfloor Tu \rfloor - n_T/2 + s + 1\right)/T\right)\right) \exp\left(-i\omega_l s\right) \right) \\
&+ \sum_{s=T_j^0-(\lfloor Tu \rfloor - n_T/2)}^{n_T-1} \left(V_{\lfloor Tu \rfloor - n_T/2+s+1,T} - \mu\left(\left(\lfloor Tu \rfloor - n_T/2 + s + 1\right)/T\right)\right) \exp\left(-i\omega_l s\right) \Bigg|^2 + o\left(1\right).
\end{aligned}
$$

Note that the first sum above involves at most $C < \infty$ summands. So the first term is negligible. The expectation of the product of the first term and the conjugate of the second term is negligible by using arguments similar to the proof in Lemma S.B.1 with n_T in place of T. Thus, the limit of $\mathbb{E}\left(I_T\left(\omega_l\right)\right)$ is equal to the right-hand side of (S.B.11) plus additional $o\left(1\right)$ terms. □

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

We first introduce the multiple Fejér kernel as in Velasco and Robinson (2001),

$$
\Psi_T^{(n)}\left(x_1, \ldots, x_n\right) = \frac{1}{\left(2\pi\right)^{n-1} T} \sum_{t_1 \cdots t_n = 1}^{T} \exp\left\{ i \sum_{j=1}^{n} t_j x_j \right\},
$$

with $x_n = -\sum_{j=1}^{n-1} x_j$. Velasco and Robinson (2001) discussed the following properties. $\Psi_T^{(n)}(x_1, \ldots, x_n)$ is integrable in Π^{n-1} and integrates to one for all T. For $\delta > 0$ and $T \geq 1$, we have

$$\int_{\mathbf{D}^c} \left| \Psi_T^{(n)}(x_1, \ldots, x_n) \right| dx_1 \ldots dx_{n-1} = O\left(\frac{\log^{n-1} T}{T \sin \delta/2} \right), \tag{S.B.16}$$

where $\mathbf{D}^c$ is the complement in Π^{n-1} of the set $\mathbf{D} = \{x \in \Pi^{n-1} : |x_j| \leq \delta, j = 1, \ldots, n-1\}$. For $j = 1, \ldots, n-1$,

$$\int_{\Pi} \cdots \int_{\Pi} |x_j| |\Psi_T^{(n)}(x_1, \ldots, x_n)| dx_1 \cdots dx_n = O\left(T^{-1} \log^{n-1} T \right). \tag{S.B.17}$$

Recall that the Dirichlet kernel is defined as $D_T(x) = \sum_{t=1}^{T} \exp(itx)$. It satisfies the following two relations,

$$|D_T(x)| \leq \min\left\{ T, 2|x|^{-1} \right\}; \qquad \int_{\Pi} |D_T(x)| dx = O(\log T). \tag{S.B.18}$$

Eq. (S.B.16)-(S.B.17) follow from

$$\left| \Psi_T^{(n)}(x_1, \ldots, x_n) \right| \leq \frac{1}{(2\pi)^{n-1} T} |D_T(x_1)| |D_T(x_2)| \cdots |D_T(x_n)| dx_1 \cdots dx_n. \tag{S.B.19}$$

S.B.2.1 Preliminary Lemmas

Lemma S.B.2. *(Bhattacharya and Rao, 1975, pp. 97-98, 113). Let $\mathbb{Q}_1$ and $\mathbb{Q}_2$ be probability measures on $\mathbb{R}^2$ and $\mathscr{B}^2$ the class of all Borel subsets of $\mathbb{R}^2$. Let ϕ be a positive number. Then there exists a kernel probability measure $\mathbb{G}_\phi$ such that*

$$\sup_{\mathbf{B} \in \mathscr{B}^2} |\mathbb{Q}_1(\mathbf{B}) - \mathbb{Q}_2(\mathbf{B})| \leq \frac{2}{3} \|(\mathbb{Q}_1 - \mathbb{Q}_2) \bullet \mathbb{G}_\phi\| + \frac{4}{3} \sup_{\mathbf{B} \in \mathscr{B}^2} \mathbb{Q}_2\left((\partial \mathbf{B})^{2\phi} \right),$$

where $\mathbb{G}_\phi$ satisfies

$$\mathbb{G}_\phi(\mathbf{B}(0, r)^c) = O\left(\left(\frac{\phi}{r} \right)^3 \right), \tag{S.B.20}$$

and its Fourier transform $\widehat{\mathbb{G}}_\phi$ satisfies

$$\widehat{\mathbb{G}}_\phi(\mathbf{t}) = 0 \qquad \text{for} \quad \|\mathbf{t}\| \geq 8 \times 2^{4/3}/\pi^{1/3}\phi. \tag{S.B.21}$$

Here $(\partial \mathbf{B})^{2\phi}$ is a neighborhood of radius 2ϕ of the boundary of $\mathbf{B}$, $\|\cdot\|$ is the variation norm, and $\bullet$ means convolution.

Lemma S.B.3. *Let Assumption 4.1, 4.3-4.4 hold. For $s \geq 2$ with $\epsilon_T(2s) \to 0$, we have*

$$\operatorname{Tr}\left((\Sigma_V W_{b_1})^s \right) = T (2\pi)^{2s-1} \sum_{j=0}^{d_f} L_j(s) b_{1,T}^{1+j-s} + O\left(T b_{1,T}^{1-s} \epsilon_T(2s) \right),$$

where $\epsilon_T(2s) = (Tb_{1,T})^{-1}\log^{2s-1}T$, $L_j(s) = (1/j)!\mu_j(K^s)(d^j/d\omega^j)(f(u,0)\,du)^s$ *with* $|L_j(s)| < \infty$ *and* $L_j(s)$ *differs from zero only for* j *even* $(j = 0, \ldots, d_f)$.

Proof of Lemma S.B.3. Let $r_{2s+1} = r_1$ and note that

$$
\begin{aligned}
&\mathrm{Tr}\left((\Sigma_V W_{b_1})^s\right)\\
&\quad = \sum_{1\le r_1,\ldots,r_{2s}\le T}\prod_{j=1}^{s}\mathbb{E}\left(V_{r_{2j-1}}V_{r_{2j}}\right)w\left(b_{1,T}\left(r_{2j}-r_{2j+1}\right)\right)\\
&\quad = \sum_{1\le r_1,\ldots,r_{2s}\le T}\prod_{j=1}^{s}\int_{\Pi} f\left(r_{2j-1}/T,\,\omega_{2j-1}\right)e^{i(r_{2j-1}-r_{2j})\omega_{2j-1}}\int_{\Pi}\widetilde{K}_{b_1}\left(\omega_{2j}\right)e^{i(r_{2j}-r_{2j+1})\omega_{2j}}d\omega\\
&\quad = \sum_{k_2,k_4,\ldots,k_{2s}=-T+1}^{T-1}\ \sum_{r_1=|k_2|+1}^{T}\ \sum_{r_3=|k_4|+1}^{T}\cdots\sum_{r_{2s-1}=|k_{2s}|+1}^{T}\prod_{j=1}^{s}\int_{\Pi} f\left(r_{2j-1}/T,\,\omega_{2j-1}\right)e^{ik_{2j}(\omega_{2j-1}-\omega_{2j})}\\
&\quad\quad \times\int_{\Pi}\widetilde{K}_{b_1}\left(\omega_{2j}\right)e^{i((-k_{2j}-k_{2j+2})\omega_{2j})}d\omega\\
&\quad = \sum_{k_2,k_4,\ldots,k_{2s}=-T+1}^{T-1}\ \prod_{j=1}^{s}\left(T-|k_{2j}|\right)\int_{\Pi}\int_0^1 f\left(u_{2j-1},\,\omega_{2j-1}\right)e^{ik_{2j}(\omega_{2j-1}-\omega_{2j})}\\
&\quad\quad \times\int_{\Pi}\widetilde{K}_{b_1}\left(\omega_{2j}\right)e^{i((-k_{2j}-k_{2j+2})\omega_{2j})}du\,d\omega + O\left(T^{-1}\right)\\
&\quad = \sum_{1\le r_1,\ldots,r_{2s}\le T}\prod_{j=1}^{s}\left(T-|k_{2j}|\right)\int_{\Pi}\int_0^1 f\left(u_{2j-1},\,\omega_{2j-1}\right)\int_{\Pi}\widetilde{K}_{b_1}\left(\omega_{2j}\right)\exp\left\{i\sum_{j=1}^{2s}\omega_j\left(r_j-r_{j+1}\right)\right\}du\,d\omega + O\left(T^{-1}\right)\\
&\quad = T\left(2\pi\right)^{2s-1}\int_{\Pi^{2s}} H_{b_1}\left(\omega,\,\mu\right)\widetilde{K}_{b_1}\left(\omega\right)\Psi_T^{(2s)}\left(\mu\right)d\omega\,d\mu + O\left(T^{-1}\right),
\end{aligned}
\tag{S.B.22}
$$

where $\Psi_T^{(2s)}(\mu) = \Psi_T^{(2s)}(\mu_1, \ldots, \mu_{2s})$,

$$
\begin{aligned}
H_{b_1}(\omega,\,\mu) = &\int_0^1\cdots\int_0^1 f\left(u_1,\,\omega-\mu_2-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_3-\ldots-\mu_{2s}\right)\\
&\times f\left(u_3,\,\omega-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_4-\ldots-\mu_{2s}\right)\ldots f\left(u_{2s-1},\,\omega-\mu_{2s}\right)du,
\end{aligned}
$$

$d\mu = d\mu_2, \ldots, d\mu_{2s}$, $d\omega = d\omega_1, \ldots, \omega_{2s}$, $du = du_1, du_3, \ldots, du_{2s-1}$, and we have made the change in variables

$$
\begin{cases}
\mu_1 = \omega_1 - \omega_2\\
\mu_2 = \omega_2 - \omega_1\\
\cdots\\
\mu_{2s} = \omega_{2s} - \omega_{2s-1}
\end{cases}
\qquad
\begin{cases}
\omega_{2s-1} = \omega - \mu_{2s}\\
\omega_{2s-2} = \omega - \mu_{2s} - \mu_{2s-1}\\
\cdots\\
\omega_1 = \omega - \mu_{2s} - \ldots - \mu_s = \omega - \mu_1
\end{cases}
$$

with $\sum_{j=1}^{2s}\mu_j = 0$, setting $\omega = \omega_{2s}$, and expressing all the ω_j in terms of ω and μ_j, $j = 2, \ldots, 2s$.

Let

$$
B = \left|\mathrm{Tr}\left((\Sigma_V W_{b_1})^s\right) - T\left(2\pi\right)^{2s-1}\int_{\Pi}\left(\int_0^1 f\left(u,\,\omega\right)du\right)^s\widetilde{K}_{b_1}^{s-1}\left(\omega\right)d\omega\right|.
$$

Using (S.B.22) we have

$$B \leq T\,(2\pi)^{2s-1} \int_{\Pi^{2s}} \left| H_{b_1}(\omega,\, \mu) - \left(\int_0^1 f(u,\, \omega)\, du \right)^s \widetilde{K}_{b_1}^{s-1}(\omega) \right| \left| \widetilde{K}_{b_1}(\omega)\, \Psi_T^{(2s)}(\mu) \right| d\omega d\mu + O\left(T^{-1}\right). \tag{S.B.23}$$

We split the integral in (S.B.23) into two sets, for small and for large μ_j. Define the set $\mathbf{M} = \{\mu \in \Pi^{2s-1} : \sup_j |\mu_j| \leq b_{1,T}/(2s)\}$. Since $K(\omega)$ takes small values for $|\omega| > \pi b_{1,T}$, for all u all functions $f(u,\, \omega)$ are boundedly differentiable in ω in the set $\mathbf{M}$. We use the following inequality,

$$|A_1 \cdots A_r - B_1 \cdots B_r| \leq \sum_{q=0}^{r-1} |B_1 \cdots B_q|\, |B_{q+1} - A_{q+1}|\, |A_{q+2} \cdots A_r|\,, \tag{S.B.24}$$

and $\sup_\omega |\widetilde{K}_{b_1}(\omega)| = O(b_{1,T}^{-1})$ to bound the integral in (S.B.23) over $\mathbf{M}$ by

$$O\left(Tb_{1,T}^{-s+1}\right) \sum_{q=0}^{s-1} \int_\Pi \int_{\mathbf{M}} \int_0^1 |f(u_{2q+1},\, \omega - \mu_{2+2q} - \ldots - \mu_{2s}) - f(u_{2q+1},\, \omega)| \left| \widetilde{K}_{b_1}(\omega)\, \Psi_T^{(2s)}(\mu) \right| du_{2q+1} d\mu d\omega \tag{S.B.25}$$

$$+O\left(Tb_{1,T}^{-s+1}\right) \sum_{q=0}^{s-2} \int_\Pi \int_{\mathbf{M}} \left| \widetilde{K}_{b_1}(\omega - \mu_{3+2q} - \ldots - \mu_{2s}) - \widetilde{K}_{b_1}(\omega) \right| \left| \Psi_T^{(2s)}(\mu) \right| d\mu d\omega. \tag{S.B.26}$$

We apply the mean value theorem in (S.B.25) to yield,

$$\begin{aligned}
&O\left(Tb_{1,T}^{1-s}\right) \int_\Pi \left| \widetilde{K}_{b_1}(\omega) \right| d\omega \sum_{q=0}^{2s} \int_{\mathbf{M}} |\mu_q| |\Psi_T^{(2s)}(\mu)|\, d\mu \\
&\qquad \leq O\left(Tb_{1,T}^{1-s}\right) \int_\Pi \left| \widetilde{K}_{b_1}(\omega) \right| d\omega \sum_{q=0}^{2s} \int_{\Pi^{2s-1}} |\mu_q| |\Psi_T^{(2s)}(\mu)|\, d\mu \\
&\qquad = O\left(b_{1,T}^{1-s} \log^{2s-1} T\right),
\end{aligned}$$

where the equality follows from (S.B.17). Using the Lipschitz property of K (cf. Assumption 4.4), the expression in (S.B.26) is of order $O(b_{1,T}^{-s} \log^{2s-1} T)$.

Let $\mathbf{M}^c$ denote the complement of $\mathbf{M}$ in Π^{2s-1}. We now study the contribution to B corresponding to the set $\mathbf{M}^c$. This is bounded by

$$T\,(2\pi)^{2s-1} \int_\Pi \int_{\mathbf{M}^c} \left| H_{b_1}(\omega,\, \mu)\, \widetilde{K}_{b_1}(\omega) \right| \left| \Psi_T^{(2s)}(\mu) \right| d\omega d\mu \tag{S.B.27}$$

$$+\, T\,(2\pi)^{2s-1} \int_\Pi \left| \left(\int_0^1 f(u,\, \omega)\, du \right)^s \widetilde{K}_{b_1}^s(\omega) \right| d\omega \int_{\mathbf{M}^c} \left| \Psi_T^{(2s)}(\mu) \right| d\mu. \tag{S.B.28}$$

The expression in (S.B.28) is $O(b_{1,T}^{-s}\log^{2s-1}T)$ using (S.B.16) and

$$\int_{\Pi}\left|\left(\int_0^1 f(u,\,\omega)du\right)^s \widetilde{K}_{b_1}^s\,(\omega)\right|d\omega = O\left(b_{1,T}^{-s}\right).$$

Applying (S.B.19) the expression in (S.B.27) is bounded by

$$\int_{\mathbf{M}'}\prod_{j=1}^{s}\int_0^1\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)\widetilde{K}_{b_1}\left(\omega_{2j}\right)D_T\left(\omega_{2j}-\omega_{2j-1}\right)D_T\left(\omega_{2j+1}-\omega_{2j}\right)\right|du_{2j-1}d\omega_{2j}d\omega_{2j-1},\quad \text{(S.B.29)}$$

where $\mathbf{M'}=\{|\omega_2-\omega_1|>\nu_T\}\cup\{|\omega_3-\omega_2|>\nu_T\}\cup\ldots\cup\{|\omega_{2s}-\omega_{2s-1}|>\nu_T\}$ with $\nu_T=b_{1,T}/\left(2s\right)$ and $2s+1$ is to be interpreted as 1. Note that the integral in (S.B.29) differs from zero only if $|\omega_2|,\,|\omega_4|,\ldots,\,|\omega_{2s}|\leq b_{1,T}\pi$. Without loss of generality, we consider only the case where just one of the events in $\mathbf{M'}$ is satisfied, $|\omega_{2j}-\omega_{2j-1}|>\nu_T$, say, the other cases can be handled similarly.

From (S.B.18) it follows that $|D_T(\omega_{2j}-\omega_{2j-1})|=O(b_{1,T}^{-1})$ since $|\omega_{2j}-\omega_{2j-1}|>\nu_T=b_{1,T}/\left(2s\right)$, and $\int_{\Pi}|D_T(\omega_{2j}-\omega_{2j-1})\widetilde{K}_{b_1}(\omega_{2j})|d\omega_{2j}=O(b_{1,T}^{-1}\log T)$. For $\epsilon>0$, consider the following decomposition

$$\begin{aligned}&\int_{\Pi}\int_0^1\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_T\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1} \qquad \text{(S.B.30)}\\ &\quad=\int_{|\omega_{2j-1}|\leq\epsilon}\int_0^1\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_T\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1}\\ &\qquad+\int_{|\omega_{2j-1}|>\epsilon}\int_0^1\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_T\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1}.\end{aligned}$$

By Assumption 4.1 $f(u_{2j-1},\,\omega_{2j-1})$ is bounded if $|\omega_{2j-1}|\leq\epsilon$. Then, the integral over $|\omega_{2j-1}|\leq\epsilon$ above is of order $O\left(\log T\right)$. On the other hand, if $|\omega_{2j-1}|>\epsilon$ (and recall that $|\omega_{2j-1}|\leq b_{1,T}\pi$), we yield as $T\rightarrow\infty$ $|\omega_{2j-1}-\omega_{2j-2}|>\epsilon/2$, say. Then, $|D_T(\omega_{2j-1}-\omega_{2j-2})|=O\left(1\right)$ by (S.B.18) and the second summand of (S.B.30) is finite in view of the integrability of $f\left(u,\,\omega\right)$ by Assumption 4.2. It follows that (S.B.30) is $O\left(\log T\right)$. There are other $s-1$ integrals of this type that can be handled in the same way. The remaining integral is of the form

$$\int_{\Pi}\int_{\Pi}\int_0^1\left|\widetilde{K}_{b_1}\left(\omega_{2s}\right)f\left(u_{2s-1},\,\omega_1\right)D_T\left(\omega_1-\omega_{2s}\right)\right|du_{2s-1}d\omega_1d\omega_{2s}=O\left(\log T\right),$$

where $\omega_1=\omega_{2s+1}$ and we have used the same argument as in (S.B.30) to show that the integral in ω_1 is $O\left(\log T\right)$ for all ω_{2s} and that $\int_{\Pi}|\widetilde{K}_{b_1}(\omega_{2s})|d\omega_{2s}=O\left(1\right)$. Thus, (S.B.29) is $O(b_{1,T}^{-s}\log^{2s-1}T)$ and $B=O(b_{1,T}^{1-s}\log^{2s-1}T+b_{1,T}^{-s}\log^{2s-1}T+T^{-1})=O(Tb_{1,T}^{1-s}\epsilon_T\left(2s\right))$.

Define $R_{b_1}\left(s\right)=\sum_{j=0}^{d_f}L_j\left(s\right)b_{1,T}^{1+j-s}$. Using the Lipschitz property of $f^{(d_f)}\left(u,\,\omega\right)$ for all u,

$$\begin{aligned}&\left|\int_{\Pi}\widetilde{K}_{b_1}^s\left(\omega\right)\left(\int_0^1 f\left(u,\,\omega\right)du\right)^s d\omega-R_{b_1}\left(s\right)\right|\\ &\quad\leq\int_{\Pi}\left|\widetilde{K}_{b_1}\left(\omega\right)\right|^{s-1}\left|\left(\int_0^1 f\left(u,\,\omega\right)du\right)^s-\sum_{j=0}^{d_f}\frac{1}{j!}\left(\frac{d}{d\omega}\right)^j\left(\int_0^1 f\left(u,\,0\right)du\right)^s\omega^j\right|\left|\widetilde{K}_{b_1}\left(\omega\right)\right|d\omega\end{aligned}$$

$$= O\left(\sup_{\omega\in\Pi}\left|\widetilde{K}_{b_1}(\omega)\right|^{s-1}\left|\int_{\Pi}|\omega|^{d_f+\varrho}\right|\left|\widetilde{K}_{b_1}(\omega)\right|d\omega\right) = O\left(b_{1,T}^{d_f+\varrho-s+1}\right),$$

where we have used $\sup_{\omega\in\Pi}|\widetilde{K}_{b_1}(\omega)| = O(b_{1,T}^{-1})$.

Note that $L_j(s)$ differs from zero for j even because $L_j(s)$ depends on $\mu_j(K^s)$. $\square$

Lemma S.B.4. *Let Assumption 4.1 and 4.3-4.4 hold. For $s\geq 1$ with $\epsilon_T(2s+2)\to 0$, we have*

$$\mathbf{1}'(\Sigma_V W_{b_1})^s\Sigma_V\mathbf{1} = T(2\pi)^{2s+1}\left(\int_0^1 f(u,0)\,du\right)^{s+1}\left(\widetilde{K}_{b_1}(0)\right)^s + O\left(b_{1,T}^{-1-s}\log^{2s+1}T + T^{-1}\right).$$

Proof of Lemma S.B.4. We first write $\mathbf{1}'(\Sigma_V W_{b_1})^s\Sigma_V\mathbf{1}$ using an argument similar to the one used to derive (S.B.22), the only difference being that we also have the summation over two additional indexes. We write

$$\begin{aligned}
&\sum_{0\leq r_1,\ldots,r_{2s+2}\leq T}\mathbb{E}\left(V_{r_{2s+1}}V_{r_{2s+2}}\right)\Pi_{j=1}^s\left\{\mathbb{E}\left(V_{r_{2j-1}}V_{r_{2j}}\right)w\left(b_{1,T}\left(r_{2j}-r_{2j+1}\right)\right)\right\}\\
&\qquad=\sum_r\int_\Pi f\left(r_{2s+1}/T,\,\omega_{2s+1}\right)e^{i(r_{2s+1}-r_{2s+2})\omega_{2s+1}}\Pi_{j=1}^s\\
&\qquad\quad\times\left\{f\left(r_{2j-1}/T,\,\omega_{2j-1}\right)e^{i(r_{2j-1}-r_{2j})\omega_{2j-1}}\int_\Pi\widetilde{K}_{b_1}(\lambda_{2j})\,e^{i(r_{2j}-r_{2j+1})\lambda_{2j}}\right\}d\lambda d\omega\\
&\qquad=T(2\pi)^{2s+1}\int_{\Pi^{2s+1}}S_{b_1}(\mu)\,\Psi_T^{(2s+2)}(\mu)\,d\mu+O\left(T^{-1}\right),
\end{aligned}\tag{S.B.31}$$

using a change of variable, where $\Psi_T^{(2s+2)}(\mu)=\Psi_T^{(2s+2)}(\mu_1,\ldots,\mu_{2s+1},-\sum_{j=1}^{2s+1}\mu_j)$,

$$S_{b_1}(\mu)=\int_0^1\cdots\int_0^1 f(u_1,\mu_1)\,\widetilde{K}_{b_1}(\mu_1+\mu_2)\ldots\widetilde{K}_{b_1}(\mu_1+\ldots+\mu_{2s})\,f(u_{2s+1},\mu_1+\ldots+\mu_{2s+1})\,du,$$

and $d\mu=d\mu_1\ldots d\mu_{2s+1}$, $du=du_1\ldots du_{2s+1}$ and $d\omega=d\omega_1\ldots d\omega_{2s+1}$. Proceeding as in the proof of Lemma S.B.3, we divide the range of integration in (S.B.31), Π^{2s+1}, into two sets, $\mathbf{M}$ and its complement $\mathbf{M}^c$, where $\mathbf{M}=\{|\mu_j|\leq\pi b_{1,T}/(2s+2),\,j=1,\ldots,2s+1\}$. We have

$$\begin{aligned}
&\left|\int_{\mathbf{M}}S_{b_1}(\mu)\,\Psi_T^{(2s+2)}(\mu)\,d\mu-\int_{\mathbf{M}}\left(\int_0^1 f(u,0)\,du\right)^{s+1}\widetilde{K}_{b_1}^s(0)\,\Psi_T^{(2s+2)}(\mu)\,d\mu\right|\\
&\qquad=O\left(b_{1,T}^{-s-1}\right)\int_{\Pi^{2s+1}}\sum_{j=2}^{2s}|\mu_j|\left|\Psi_T^{(2s+2)}(\mu)\right|d\mu\\
&\qquad=O\left(b_{1,T}^{-s-1}T^{-1}\log^{2s+1}T\right),
\end{aligned}\tag{S.B.32}$$

using (S.B.17), (S.B.24), Assumption 4.1 and 4.4. On the other hand, the contribution from $\mathbf{M}^c$ is less than or equal to

$$\int_{\mathbf{M}^c}\left|S_{b_1}(\mu)\right|\left|\Psi_T^{(2s+2)}(\mu)\right|d\mu+O\left(b_{1,T}^{-s-1}T^{-1}\log^{2s+1}T\right),\tag{S.B.33}$$

where we have used (S.B.16). Using the same argument used for (S.B.29), the integral in (S.B.33) is less than or equal to

$$\frac{1}{T(2\pi)^{2s+1}}\int_{\mathbf{M}'}\prod_{j=1}^{s}\int_0^1\int_0^1\left[f\left(u_{2j-1},\,\omega_{2j-1}\right)\widetilde{K}_{b_1}\left(\omega_{2j}\right)D_T\left(\omega_{2j}-\omega_{2j-1}\right)\right.$$
$$\left.\times D_T\left(\omega_{2j+1}-\omega_{2j}\right)f\left(u_{2s+1},\,\omega_{2s+1}\right)D_T\left(\omega_1\right)D_T\left(-\omega_{2s-1}\right)\right]dud\omega, \tag{S.B.34}$$

where

$$\mathbf{M}'=\left\{|\omega_1|>\pi b_{1,T}/\left(2s+2\right)\right\}\cup\left\{|\omega_2-\omega_1|>\pi b_{1,T}/\left(2s+2\right)\right\}\cup\ldots\cup\left\{|\omega_{2s-1}-\omega_{2s}|>\pi b_{1,T}/\left(2s+2\right)\right\},$$

and (S.B.34) is nonzero only if $|\omega_2|,\,|\omega_4|,\ldots,\,|\omega_{2s}|\leq\pi b_{1,T}$.

If $|\omega_{j+1}-\omega_j|>\pi b_{1,T}/\left(2s+2\right)$ for at least one index $j\in\{1,\ldots,\,2s\}$ we can obtain a bound of order $(T^{-1}b_{1,T}^{-s-1}\log^{2s+1}T)$ for (S.B.34) as in Lemma S.B.3. The same bound is obtained for the case $|\omega_1|>\pi b_{1,T}/\left(2s+2\right)$ with a similar argument. Combining these results with (S.B.31)-(S.B.33) concludes the proof. □

Lemma S.B.5. *Let Assumption 4.1, 4.3-4.4 and 4.8-4.9 hold. For $s\geq 2$ with $\epsilon_{Tb_{2,T}}\left(2s\right)\to 0$, we have*

$$\mathrm{Tr}\left(\left(\Sigma_{\widetilde{V}}W_{b_1}\right)^s\right)=Tb_{2,T}\left(2\pi\right)^{2s-1}\left(\sum_{j=0}^{d_f}L_j\left(s\right)b_{1,T}^{1+j-s}+b_{2,T}^2\sum_{j=0}^{d_f}\left(\left(L_{2,j}\left(s\right)+L_{3,j}\left(s\right)\right)b_{1,T}^{1+j-s}\right)\right)$$
$$+O\left(Tb_{2,T}b_{1,T}^{1-s}\epsilon_{Tb_{2,T}}\left(2s\right)+b_{1,T}^{-s}\frac{\log^{2s}\left(Tb_{2,T}\right)}{Tb_{2,T}}\right),$$

where $\epsilon_{Tb_{2,T}}\left(2s\right)=\left(Tb_{2,T}\right)^{-1}\log^{2s-1}\left(Tb_{2,T}\right)$, $L_j\left(s\right)=(1/j)!\mu_j(K^s)\int_0^1K_2^s\left(x\right)dx\,(d^j/d\omega^j)\,(\int_0^1f\left(u,\,0\right)du)^s$ with $|L_j\left(s\right)|<\infty$, $L_j\left(s\right)$ differs from zero only for j even, $L_{2,j}\left(s\right)$ depends on $\frac{\partial^2}{\partial u^2}\int_{\widetilde{\mathbf{C}}}f\left(u,\,\omega\right)du$, K_2, $\widetilde{K}_{b_1}$ and s with $|L_{2,j}\left(s\right)|<\infty$, and $L_{3,j}\left(s\right)$ depends on $\Delta_f\left(\cdot\right)$, $\widetilde{K}_{b_1}$ and s with $|L_{3,j}\left(s\right)|<\infty$.

Proof of Lemma S.B.5. Let $r_{2s+1}=r_1$ and note that

$$\begin{aligned}\mathrm{Tr}\left(\left(\Sigma_{\widetilde{V}}W_{b_1}\right)^s\right)&=\int_0^1\cdots\int_0^1\sum_{1\leq r_1,\ldots,r_{2s}\leq T}\prod_{j=1}^{s}\mathbb{E}\left(\widetilde{V}_{r_{2j-1}}\left(u_j\right)\widetilde{V}_{r_{2j}}\left(u_j\right)\right)w\left(b_{1,T}\left(r_{2j}-r_{2j+1}\right)\right)du\\&=\int_0^1\cdots\int_0^1\sum_{1\leq r_1,\ldots,r_{2s}\leq T}\prod_{j=1}^{s}K_2\left(\frac{\left(Tu_j-\left(r_{2j-1}-\left(r_{2j}-r_{2j-1}\right)/2\right)\right)/T}{b_{2,T}}\right)\\&\quad\times\int_\Pi f\left(r_{2j-1}/T,\,\omega\right)e^{i(r_{2j-1}-r_{2j})\omega_{2j-1}}d\omega\int_\Pi\widetilde{K}_{b_1}\left(\omega_{2j}\right)e^{i(r_{2j}-r_{2j+1})\omega_{2j}}d\omega du\\&=\sum_{k_2,\,k_4,\ldots,\,k_{2s}=-\lfloor Tb_{2,T}\rfloor+1}^{\lfloor Tb_{2,T}\rfloor-1}\int_0^1\cdots\int_0^1\int_{\Pi^2}\prod_{j=1}^{s}\left(Tb_{2,T}-|k_{2j}|\right)f\left(u_{2j-1},\,\omega_{2j-1}\right)e^{i(\omega_{2j-1}-\omega_{2j})k_{2j}}\\&\quad\times\widetilde{K}_{b_1}\left(\omega_{2j}\right)e^{i(-k_{2j}-k_{2j+2})\omega_{2j}}d\omega du+O\left(b_{2,T}^2\right)+O\left(\frac{\log\left(Tb_{2,T}\right)}{Tb_{2,T}}\right)\end{aligned}$$

$$= Tb_{2,T}\left(2\pi\right)^{2s-1}\int_{\Pi^{2s}}\left(H_{b_1}\left(\omega,\,\mu\right)\int_0^1 K_2^s\left(x\right)dx + H_{2,b_1}\left(\omega,\,\mu\right) + H_{3,b_1}\left(\omega,\,\mu\right)\right) \tag{S.B.35}$$
$$\times \widetilde{K}_{b_1}\left(\omega\right)\Psi^{(2s)}_{Tb_{2,T}}\left(\mu\right)d\omega d\mu + O\left(b_{2,T}^2 b_{1,T}^{-s}\log^{2s-1}\left(Tb_{2,T}\right)\right) + O\left(b_{1,T}^{-s}\frac{\log^{2s}\left(Tb_{2,T}\right)}{Tb_{2,T}}\right),$$

where $H_{b_1}\left(\omega,\,\mu\right)$, $d\omega$ and $d\mu$ are defined as in (S.B.22), $\Psi^{(2s)}_{Tb_{2,T}}\left(\mu\right) = \Psi^{(2s)}_{Tb_{2,T}}\left(\mu_1,\ldots,\,\mu_{2s}\right)$,

$$H_{2,b_1}\left(\omega,\,\mu\right) = b_{2,T}^2\left(\int_0^1 x^2 K_2\left(x\right)dx\right)\left(\int_0^1 K_2^{s-1}\left(x\right)dx\right)$$
$$\times \sum_{j\in\mathbf{J}}\frac{\partial^2}{\partial u_j^2}\int_{\widetilde{\mathbf{C}}}\cdots\int_{\widetilde{\mathbf{C}}} f\left(u_1,\,\omega-\mu_2-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_3-\ldots-\mu_{2s}\right)$$
$$\times f\left(u_3,\,\omega-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_4-\ldots-\mu_{2s}\right)\ldots f\left(u_{2s-1},\,\omega-\mu_{2s}\right)du_1\cdots du_{2s-1},$$

with $\mathbf{J}=\{1,\,3,\ldots,\,2s-1\}$, and $H_{3,b_1}\left(\omega,\,\mu\right)$ depends on the discontinuity points, i.e.,

$$H_{3,b_1}\left(\omega,\,\mu\right) = b_{2,T}^2\left(\int_0^1 K_2^{s-1}\left(x\right)dx\right)\left(\mathbf{1}\left\{u_1=\lambda_j^0,\,j=1,\ldots,\,m_0\right\}\Delta_{f,j}\left(\omega-\mu_2-\ldots-\mu_{2s}\right)\right)$$
$$\times \widetilde{K}_{b_1}\left(\omega-\mu_3-\ldots-\mu_{2s}\right)f\left(u_3,\,\omega-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_4-\ldots-\mu_{2s}\right)\ldots f\left(u_{2s-1},\,\omega-\mu_{2s}\right)$$
$$\vdots$$
$$+ b_{2,T}^2\left(\int_0^1 K_2^{s-1}\left(x\right)dx\right)f\left(u_1,\,\omega-\mu_2-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_3-\ldots-\mu_{2s}\right)$$
$$\times f\left(u_3,\,\omega-\ldots-\mu_{2s}\right)\widetilde{K}_{b_1}\left(\omega-\mu_4-\ldots-\mu_{2s}\right)\ldots$$
$$\times \mathbf{1}\left\{u_{2s-1}=\lambda_j^0,\,j=1,\ldots,\,m_0\right\}\Delta_{f,j}\left(\omega-\mu_{2s}\right),$$

with

$$\Delta_{f,j}\left(\omega\right) = \int_0^1\left(\frac{\partial}{\partial u_-}f\left(\lambda_j^0,\,\omega\right)\int_0^{1-s} xK_2\left(x\right)dx + \frac{\partial}{\partial u_+}f\left(\lambda_j^0,\,\omega\right)\int_{1-s}^1 xK_2\left(x\right)dx\right)ds. \tag{S.B.36}$$

Let

$$B = \left|Tb_{2,T}\left(2\pi\right)^{2s-1}\int_0^1 K_2^s\left(x\right)dx\int_{\Pi^{2s}}\left(H_{b_1}\left(\omega,\,\mu\right)\widetilde{K}_{b_1}\left(\omega\right)\Psi^{(2s)}_{Tb_{2,T}}\left(\mu\right) - \left(\int_0^1 f\left(u,\,\omega\right)du\right)^s\widetilde{K}^s_{b_1}\left(\omega\right)\right)d\omega d\mu\right|.$$

Using (S.B.35) we have

$$B \le Tb_{2,T}\left(2\pi\right)^{2s-1}\int_0^1 K_2^s\left(x\right)dx\int_{\Pi^{2s}}\left|H_{b_1}\left(\omega,\,\mu\right) - \left(\int_0^1 f\left(u,\,\omega\right)du\right)^s\widetilde{K}^{s-1}_{b_1}\left(\omega\right)\right|\left|\widetilde{K}_{b_1}\left(\omega\right)\Psi^{(2s)}_{Tb_{2,T}}\left(\mu\right)\right|d\omega d\mu. \tag{S.B.37}$$

We split the integral in (S.B.37) into two sets, for small and for large μ_j. Define the set $\mathbf{M}=\{\mu\in\Pi^{2s-1}:$

$\sup_j |\mu_j| \leq b_{1,T}/(2s)\}$. Proceeding as in (S.B.25)-(S.B.26), we have

$$O\left(Tb_{2,T}b_{1,T}^{-s+1}\right)\sum_{q=0}^{s-1}\int_{\Pi}\int_{\mathbf{M}}\int_0^1 |f(u,\,\omega-\mu_{2+2q}-\ldots-\mu_{2s})-f(u,\,\omega)|\left|\widetilde{K}_{b_1}(\omega)\,\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| dud\omega d\mu \tag{S.B.38}$$

$$+O\left(Tb_{2,T}b_{1,T}^{-s+1}\right)\sum_{q=0}^{s-2}\int_{\Pi}\int_{\mathbf{M}}\left|\widetilde{K}_{b_1}(\omega-\mu_{2+2q}-\ldots-\mu_{2s})-\widetilde{K}_{b_1}(\omega)\right|\left|\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| d\omega d\mu. \tag{S.B.39}$$

We apply the mean value theorem in (S.B.38) and use (S.B.17) to yield,

$$\begin{aligned}
&O\left(Tb_{2,T}b_{1,T}^{-s+1}\right)\int_{\Pi}\left|\widetilde{K}_{b_1}(\omega)\right| d\omega\sum_{q=0}^{2s}\int_{\mathbf{M}}|\mu_q|\left|\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| d\mu\\
&\qquad\leq O\left(Tb_{2,T}b_{1,T}^{-s+1}\right)\int_{\Pi}\left|\widetilde{K}_{b_1}(\omega)\right| d\omega\sum_{q=0}^{2s}\int_{\Pi^{2s-1}}|\mu_q|\left|\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| d\mu\\
&\qquad= O\left(b_{1,T}^{-s+1}\log^{2s-1}(Tb_{2,T})\right).
\end{aligned}$$

On the other hand, using the Lipschitz property of K (cf. Assumption 4.4), the expression in (S.B.39) is of order $O(b_{1,T}^{-s}\log^{2s-1}(Tb_{2,T}))$.

Let $\mathbf{M}^c$ denote the complement of $\mathbf{M}$ in Π^{2s-1}. The contribution to B corresponding to the set $\mathbf{M}^c$ is bounded by

$$Tb_{2,T}(2\pi)^{2s-1}\int_{\Pi}\int_{\mathbf{M}^c}\left|H_{b_1}(\omega,\,\mu)\,\widetilde{K}_{b_1}(\omega)\right|\left|\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| d\omega d\mu \tag{S.B.40}$$

$$+\,Tb_{2,T}(2\pi)^{2s-1}\int_{\Pi}\left|\left(\int_0^1 f(u,\,\omega)\,du\right)^s\widetilde{K}_{b_1}^s(\omega)\right| d\omega\int_{\mathbf{M}^c}\left|\Psi_{Tb_{2,T}}^{(2s)}(\mu)\right| d\mu. \tag{S.B.41}$$

The expression in (S.B.41) is $O(b_{1,T}^{-s}\log^{2s-1}(Tb_{2,T}))$ using (S.B.16) and

$$\int_{\Pi}\left|\left(\int_0^1 f(u,\,\omega)\right)^s\widetilde{K}_{b_1}^s(\omega)\right| d\omega = O\left(b_{1,T}^{-s}\right).$$

The expression in (S.B.40) is bounded by

$$\int_{\mathbf{M}'}\prod_{j=1}^{s}\int_0^1\left|f(u_{2j-1},\,\omega_{2j-1})\,\widetilde{K}_{b_1}(\omega_{2j})\,D_{Tb_{2,T}}(\omega_{2j}-\omega_{2j-1})\,D_{Tb_{2,T}}(\omega_{2j+1}-\omega_{2j})\right| du_{2j-1}d\omega_{2j}d\omega_{2j-1}, \tag{S.B.42}$$

where $\mathbf{M}'$ is defined after (S.B.29).

From (S.B.18) it follows that $|D_{Tb_{2,T}}(\omega_{2j}-\omega_{2j-1})| = O(b_{1,T}^{-1})$ since $|\omega_{2j}-\omega_{2j-1}| > \nu_T = b_{1,T}/(2s)$, and $\int_{\Pi}|D_{Tb_{2,T}}(\omega_{2j}-\omega_{2j+1})\,\widetilde{K}_{b_1}(\omega_{2j})|d\omega_{2j} = O(b_{1,T}^{-1}\log(Tb_{2,T}))$. For $\epsilon > 0$, consider the following decomposition

$$\begin{aligned}
&\int_{\Pi}\int_{0}^{1}\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_{Tb_{2,T}}\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1} \qquad \text{(S.B.43)}\\
&\quad=\int_{|\omega_{2j-1}|\leq\epsilon}\int_{0}^{1}\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_{Tb_{2,T}}\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1}\\
&\qquad+\int_{|\omega_{2j-1}|>\epsilon}\int_{0}^{1}\left|f\left(u_{2j-1},\,\omega_{2j-1}\right)D_{Tb_{2,T}}\left(\omega_{2j-1}-\omega_{2j-2}\right)\right|du_{2j-1}d\omega_{2j-1}.
\end{aligned}$$

By Assumption 4.1 $f(u_{2j-1},\,\omega_{2j-1})$ is bounded if $|\omega_{2j-1}|\leq\epsilon$. Then the integral over $|\omega_{2j-1}|\leq\epsilon$ above is of order $O(\log(Tb_{2,T}))$. On the other hand, if $|\omega_{2j-1}|>\epsilon$ we have $|D_{Tb_{2,T}}(\omega_{2j-1}-\omega_{2j-2})|=O\left(1\right)$ by (S.B.18) and the second summand of (S.B.43) is finite in view of the integrability of $f\left(u,\,\omega\right)$ by Assumption 4.2. It follows that (S.B.43) is $O(\log(Tb_{2,T}))$. There are other $s-1$ integrals of this type that can be handled in the same way. The remaining integral is of the form

$$\int_{\Pi}\int_{\Pi}\int_{0}^{1}\left|\widetilde{K}_{b_1}\left(\omega_{2s}\right)f\left(u_{2s-1},\,\omega_{1}\right)D_{Tb_{2,T}}\left(\omega_{1}-\omega_{2s}\right)\right|du_{2s-1}d\omega_{1}d\omega_{2s}=O\left(\log\left(Tb_{2,T}\right)\right),$$

where $\omega_1=\omega_{2s+1}$ and we have used the same argument as in (S.B.43) to show that the integral in ω_1 is $O(\log(Tb_{2,T}))$ for all ω_{2s} and that $\int_{\Pi}|\widetilde{K}_{b_1}\left(\omega_{2s}\right)|d\omega_{2s}=O\left(1\right)$. Thus, (S.B.42) is $O(b_{1,T}^{-s}\log^{2s-1}Tb_{2,T})$ and $B=O(b_{1,T}^{1-s}\log^{2s-1}(Tb_{2,T})+b_{1,T}^{-s}\log^{2s-1}(Tb_{2,T}))=O(Tb_{2,T}b_{1,T}^{1-s}\epsilon_{Tb_{2,T}}\left(2s\right))$.

Next, let

$$B_{2}=Tb_{2,T}\left(2\pi\right)^{2s-1}\int_{\Pi^{2s}}\left|H_{2,b_1}\left(\omega,\,\mu\right)-b_{2,T}^{2}\Lambda_{2}\left(f'',\,\widetilde{\mathbf{C}},\,s\right)\widetilde{K}_{b_1}^{s-1}\left(\omega\right)\right|\left|\widetilde{K}_{b_1}\left(\omega\right)\Psi_{Tb_{2,T}}^{(2s)}\left(\mu\right)\right|d\omega d\mu,$$

where $\Lambda_2(f'',\,\widetilde{\mathbf{C}},\,s)$ depends on $f\left(u,\,\omega\right)$, the second partial derivative of $f\left(u,\,\omega\right)$ in u at the continuity points in $\widetilde{\mathbf{C}}$ and s. By Assumption 4.9, for $j\in\mathbf{J}$ and $u_j\in\widetilde{\mathbf{C}}$ $(\partial^2/\partial u_j^2)f\left(u_j,\,\omega_j\right)$ has similar smoothness properties in ω_j to those of $f\left(u_j,\,\omega_j\right)$. Thus, the proof used above to bound B can be repeated which then results in $B_2=O(Tb_{2,T}^{3}b_{1,T}^{1-s}\epsilon_{Tb_{2,T}}\left(2s\right))$.

Let

$$\begin{aligned}
B_{3}=&\,Tb_{2,T}\left(2\pi\right)^{2s-1}\int_{\Pi^{2s}}\left|H_{3,b_1}\left(\omega,\,\mu\right)-b_{2,T}^{2}\Lambda_{3}\left(f',\,\left\{\lambda_{j}^{0},\,j=1,\ldots,\,m_{0}\right\},\,s\right)\widetilde{K}_{b_1}^{s-1}\left(\omega\right)\right|\\
&\times\left|\widetilde{K}_{b_1}\left(\omega\right)\Psi_{Tb_{2,T}}^{(2s)}\left(\mu\right)\right|d\omega d\mu,
\end{aligned}$$

where $\Lambda_3(f',\,\{\lambda_j^0,\,j=1,\ldots,\,m_0\},\,s)$ depends on $f\left(u,\,\omega\right),\Delta_f\left(\cdot\right)$ and s. By Assumption 4.9, $\left(\partial/\partial u_{-}\right)f\left(u,\,\omega\right)$ and $\left(\partial/\partial u_{+}\right)f\left(u,\,\omega\right)$ for u a discontinuity point have similar smoothness properties in ω to those of $f\left(u,\,\omega\right)$. Thus, the proof used above to bound B can be repeated which then results in $B_3=O(Tb_{2,T}^{3}b_{1,T}^{1-s}\,\epsilon_{Tb_{2,T}}\left(2s\right))$.

The rest of the proof follows from the same arguments used in the last part of the proof of Proposition S.B.3. □

Lemma S.B.6. *Let Assumption 4.1, 4.3-4.4 and 4.8-4.9 hold. For $s \geq 1$ with $\epsilon_T(2s+2) \to 0$, we have*

$$\mathbf{1}'\left(\Sigma_{\widetilde{V}}W_{b_1}\right)^s\Sigma_{\widetilde{V}}\mathbf{1} = Tb_{2,T}(2\pi)^{2s+1}\left(\left(\left(\int_0^1 f(u,0)\,du\right)^{s+1}\int_0^1 K_2^{s+1}(x)\,dx\right.\right.$$
$$\left.\left.+\, b_{2,T}^2\left(\widetilde{\Lambda}_2\left(f'',\widetilde{\mathbf{C}},s\right)+\widetilde{\Lambda}_3\left(f',\left\{\lambda_j^0,\,j=1,\ldots,m_0\right\},s\right)\right)\right)\left(\widetilde{K}_{b_1}(0)\right)^s\right.$$
$$+\,O\left(b_{1,T}^{1-s}\log^{2s+1}(Tb_{2,T})+b_{1,T}^{-s}\frac{\log^{2s+1}(Tb_{2,T})}{Tb_{2,T}}\right),$$

where $\Lambda_2(f'',\widetilde{\mathbf{C}},s)$ depends on $f(u,\omega)$, the second partial derivative of $f(u,\omega)$ in u at the continuity points in $\widetilde{\mathbf{C}}$ and s, and $\widetilde{\Lambda}_3(f',\{\lambda_j^0,\,j=1,\ldots,m_0\},s)$ depends on $f(u,\omega)$, $\Delta_f(\cdot)$ and s.

Proof of Lemma S.B.6. We first write $\mathbf{1}'(\Sigma_{\widetilde{V}}W_{b_1})^s\Sigma_{\widetilde{V}}\mathbf{1}$ using an argument similar to the one used to derive (S.B.31),

$$\int_0^1\sum_{1\leq r_1,\ldots,r_{2s+2}\leq T}\mathbb{E}\left(\widetilde{V}_{r_{2s+1}}(u_{s+1})\widetilde{V}_{r_{2s+2}}(u_{s+1})\right)\int_0^1\cdots\int_0^1\Pi_{j=1}^s$$
$$\times\left\{\mathbb{E}\left(\widetilde{V}_{r_{2j-1}}(u_j)\widetilde{V}_{r_{2j}}(u_j)\right)w\left(b_{1,T}(r_{2j}-r_{2j+1})\right)\right\}du$$
$$=Tb_{2,T}\sum_{k_{2s+2}=-\lfloor Tb_{2,T}\rfloor+1}^{\lfloor Tb_{2,T}\rfloor-1}\int_0^1\int_\Pi f(u_{s+1}/T,\omega_{2s+1})\,e^{-ik_{2s+2}\omega_{2s+1}}\Pi_{j=1}^s\int_0^1\cdots\int_0^1$$
$$\times\left\{f(u_{2j-1}/T,\omega_{2j-1})\sum_{k_2,k_4,\ldots,k_{2s}=-\lfloor Tb_{2,T}\rfloor+1}^{\lfloor Tb_{2,T}\rfloor-1}\frac{Tb_{2,T}-|k_{2j}|}{Tb_{2,T}}\int_\Pi\widetilde{K}_{b_1}(\omega_{2j})\,e^{i(k_{2j}+k_{2j+1})\omega_{2j}}\right\}d\omega du$$
$$=Tb_{2,T}(2\pi)^{2s+1}\int_{\Pi^{2s+1}}\left(S_{b_1}(\mu)\int_0^1K_2^{s+1}(x)\,dx+S_{2,b_1}(\mu)+S_{3,b_1}(\mu)\right)\Psi_{Tb_{2,T}}^{(2s+2)}(\mu)\,d\mu \tag{S.B.44}$$
$$+\,O\left(b_{2,T}^2b_{1,T}^{-s}\log^{2s-1}(Tb_{2,T})\right)+O\left(b_{1,T}^{-s}\frac{\log^{2s}(Tb_{2,T})}{Tb_{2,T}}\right),$$

where $\Psi_{Tb_{2,T}}^{(2s+2)}(\mu)$, $S_{b_1}(\mu)$ and $d\mu=d\mu_1\ldots d\mu_{2s+1}$ are defined as in (S.B.31),

$$S_{2,b_1}(\mu)=b_{2,T}^2\left(\int_0^1x^2K_2(x)\,dx\right)\int_0^1K_2^s(x)\,dx\sum_{j\in\mathbf{J}}\frac{\partial^2}{\partial u_j^2}\int_{\widetilde{\mathbf{C}}}\cdots\int_{\widetilde{\mathbf{C}}}f(u_1,\mu_1)\,\widetilde{K}_{b_1}(\mu_1+\mu_2)\ldots$$
$$\times\,\widetilde{K}_{b_1}(\mu_1+\ldots+\mu_{2s})\,f(u_{2s+1},\mu_1+\ldots+\mu_{2s+1})\,du,$$

with $\mathbf{J}=\{1,3,\ldots,2s+1\}$ and $S_{3,b_1}(\omega,\mu)$ depends on the discontinuity points, i.e.,

$$S_{3,b_1}(\mu)=b_{2,T}^2\int_0^1K_2^s(x)\,dx\left(\mathbf{1}\left\{u_1=\lambda_j^0,\,j=1,\ldots,m_0\right\}\Delta_{f,j}(\mu_1)\right)\widetilde{K}_{b_1}(\mu_1+\mu_2)$$
$$\ldots\widetilde{K}_{b_1}(\mu_1+\ldots+\mu_{2s})\,f(u_{2s-1},\mu_1+\ldots+\mu_{2s+1})$$

$$
\begin{aligned}
&\vdots \\
&+ b_{2,T}^{2} \int_{0}^{1} K_{2}^{s}(x)\, dx f(u_{1},\, \omega - \mu_{2} - \ldots - \mu_{2s}) \widetilde{K}_{b_{1}}(\omega - \mu_{3} - \ldots - \mu_{2s}) \\
&\times \widetilde{K}_{b_{1}}(\mu_{1} + \ldots + \mu_{2s}) \mathbf{1}\left\{u_{2s-1} = \lambda_{j}^{0},\, j = 1, \ldots,\, m_{0}\right\} \Delta_{f,j}(\mu_{1} + \ldots + \mu_{2s+1}),
\end{aligned}
$$

with $\Delta_{f,j}(\omega)$ defined in (S.B.36). Proceeding as in the proof of Lemma S.B.4, we divide the range of integration of the integral involving $S_{b_1}(\mu)$ in (S.B.44), Π^{2s+1}, into two sets, $\mathbf{M}$ and its complement $\mathbf{M}^c$, where $\mathbf{M} = \{|\mu_j| \leq \pi b_{1,T}/(2s+2),\, j = 1, \ldots,\, 2s+1\}$. We have

$$
\begin{aligned}
&\left| \int_{\mathbf{M}} S_{b_1}(\mu)\, \Psi_{Tb_{2,T}}^{(2s+2)}(\mu)\, d\mu - \int_{\mathbf{M}} \left(\int_{0}^{1} f(u,\, 0)\, du \right)^{s+1} \widetilde{K}_{b_1}^{s}(0)\, \Psi_{Tb_{2,T}}^{(2s+2)}(\mu)\, d\mu \right| \\
&\quad = O\left(b_{1,T}^{-s-1}\right) \int_{\Pi^{2s+1}} \sum_{j=2}^{2s} |\mu_j| \left| \Psi_{Tb_{2,T}}^{(2s+2)}(\mu) \right| d\mu \\
&\quad = O\left(b_{1,T}^{-s-1} (Tb_{2,T})^{-1} \log^{2s+1}(Tb_{2,T})\right), \qquad \text{(S.B.45)}
\end{aligned}
$$

using (S.B.17), (S.B.24), Assumption 4.1 and 4.4. On the other hand, the contribution from $\mathbf{M}^c$ is less than or equal to

$$
Tb_{2,T} (2\pi)^{2s+1} \int_{\mathbf{M}^c} |S_{b_1}(\mu)| \left| \Psi_{Tb_{2,T}}^{(2s+2)}(\mu) \right| d\mu + O\left(b_{1,T}^{-s} \log^{2s+1}(Tb_{2,T})\right), \qquad \text{(S.B.46)}
$$

where we have used (S.B.16). Using the same argument used for (S.B.42), the expression in (S.B.46) is less than or equal to

$$
\begin{aligned}
&\int_{\mathbf{M}'} \prod_{j=1}^{s} \int_{0}^{1} \int_{0}^{1} \Big| f(u_{2j-1},\, \lambda_{2j-1}) \widetilde{K}_{b_1}(\lambda_{2j}) D_{Tb_{2,T}}(\lambda_{2j} - \lambda_{2j-1}) \qquad \text{(S.B.47)} \\
&\quad \times D_{Tb_{2,T}}(\lambda_{2j+1} - \lambda_{2j}) f(u_{2s+1},\, \lambda_{2s+1}) D_{Tb_{2,T}}(\lambda_1) D_{Tb_{2,T}}(-\lambda_{2s-1}) \Big| du_{2s+1} du_{2j-1} d\lambda,
\end{aligned}
$$

where $\mathbf{M}' = \{|\lambda_1| > \pi b_{1,T}/(2s+2)\} \cup \{|\lambda_2 - \lambda_1| > \pi b_{1,T}/(2s+2)\} \cup \ldots \cup \{|\lambda_{2s-1} - \lambda_{2s}| > \pi b_{1,T}/(2s+2)\}$ and (S.B.47) is nonzero only if $|\lambda_2|, |\lambda_4|, \ldots, |\lambda_{2s}| \leq \pi b_{1,T}$.

If $|\lambda_{j+1} - \lambda_j| > \pi b_{1,T}/(2s+2)$ for at least one index $j \in \{1, \ldots,\, 2s\}$ we can obtain a bound of order $((Tb_{2,T})^{-1} b_{1,T}^{-s-1} \log^{2s+1}(Tb_{2,T}))$ for (S.B.47) as in Lemma S.B.5.

Next, we have

$$
\begin{aligned}
&Tb_{2,T} (2\pi)^{2s+1} \bigg| \int_{\Pi^{2s}} (S_{b_2}(\mu) + S_{b_3}(\mu))\, \Psi_{Tb_{2,T}}^{(2s+2)}(\mu)\, d\mu \qquad \text{(S.B.48)} \\
&\quad - b_{2,T}^{2} \int_{\Pi^{2s}} \left(\widetilde{\Lambda}_2\left(f'',\, \widetilde{\mathbf{C}},\, s\right) + \widetilde{\Lambda}_3\left(f',\, \left\{\lambda_j^0,\, j = 1, \ldots,\, m\right\},\, s\right) \right) \widetilde{K}_{b_1}^{s}(0)\, \Psi_{Tb_{2,T}}^{(2s+2)}(\mu)\, d\mu \bigg|.
\end{aligned}
$$

By Assumption 4.9, $(\partial^2/\partial u^2) f(u,\, \omega)$ for $u \in \widetilde{\mathbf{C}}$, $(\partial/\partial u_-) f(u,\, \omega)$ and $(\partial/\partial u_+) f(u,\, \omega)$ for u a discontinuity point have similar smoothness properties in ω to those of $f(u,\, \omega)$. Thus, the proof used above to bound (S.B.45) can be repeated which then results in (S.B.48) being $O(b_{2,T}^2 b_{1,T}^{-s-1} \log^{2s+1}(Tb_{2,T})$. □

Lemma S.B.7. *Let Assumption 4.1, 4.2* ($p>1$)*, 4.3-4.4 and 4.7* ($0<q<1$) *hold. Then,* $||\Sigma_V W_{b_1}|| \leq C_1 \nu_{2,T}$ *where* C_1 *depends on* $f(\cdot,\,\cdot)$ *and* K*,* $0<C_1<\infty$ *and* $\nu_{2,T}=\max\{b_{1,T}^{-1}\log^2 T,\, T^{(2-p)/2p} b_{1,T}^{-1/2}\log^2 T)\}$ $\to \infty$.

Proof of Lemma S.B.7. We have

$$\begin{aligned}
\|\Sigma_V W_{b_1}\| &= \sup_{\|x\|=1}\left|\sum_{j,h=1}^{T} x_j x_h \sum_{t=1}^{T}\sum_{s=1}^{T}\int_{\Pi^2} f(t/T,\,\lambda)\,\widetilde{K}_{b_1}(\omega)\,e^{it\lambda}e^{-is\omega}e^{i(h\omega-j\lambda)}d\lambda d\omega\right| + O\left(T^{-1}\right)\\
&= \sup_{\|x\|=1}\left|\sum_{t=1}^{T} f(t/T,\,\lambda)\,e^{it\lambda}\sum_{j,h} x_j x_h \int_{\Pi^2}\widetilde{K}_{b_1}(\omega)\,D_T(-\omega)\,e^{i(h\omega-j\lambda)}d\lambda d\omega\right| + O\left(T^{-1}\right)\\
&\leq \sup_{\|x\|=1}\left|\int_{\omega\leq\epsilon}\int_{\lambda}\sum_{t=1}^{T} f(t/T,\,\lambda)\,e^{it\lambda}D_T(-\omega)\sum_{j,h} x_j x_h \widetilde{K}_{b_1}(\omega)\,e^{i(h\omega-j\lambda)}d\lambda d\omega\right|\\
&\quad + \sup_{\|x\|=1}\left|\int_{\omega>\epsilon}\int_{\lambda}\sum_{t=1}^{T} f(t/T,\,\lambda)\,e^{it\lambda}D_T(-\omega)\sum_{j,h} x_j x_h \widetilde{K}_{b_1}(\omega)\,e^{i(h\omega-j\lambda)}d\lambda d\omega\right| + O\left(T^{-1}\right)\\
&\triangleq A_1 + o(1) + O\left(T^{-1}\right).
\end{aligned} \tag{S.B.49}$$

Let $L_{2,T}:\mathbb{R}\to\mathbb{R}$ be the periodic extension with period 2π of

$$L_{2,T}(\omega)=\begin{cases} T, & |\omega|\leq 1/T,\\ 1/|\omega|, & 1/T\leq|\omega|\leq|\pi|.\end{cases}$$

Lemma S.A.1-2 in Casini and Perron (2021a) showed that

$$\left|\sum_{t=1}^{T} f(t/T,\,\lambda)\,e^{-it\lambda}\right| \leq L_{2,T}(\lambda), \tag{S.B.50}$$

and $\int_{\Pi} L_{2,T}(\lambda)\,d\lambda \leq C_L \log T$ for $T>1$ and $C_L>0$ being a constant independent of T. Let $X_T(\omega)=\sum_{j=1}^{T} x_j e^{ij\omega}$. Then, the contribution to A_1 from $|\lambda|\leq\epsilon$ is bounded by

$$\begin{aligned}
&\sup_{\|x\|=1}\int_{\omega\leq\epsilon}\int_{\lambda}\left|\sum_{t=1}^{T} f(t/T,\,\lambda)\,e^{it\lambda}\right| |D_T(-\omega)|\,|X_T(\omega)|\,|X_T(\lambda)|\left|\widetilde{K}_{b_1}(\omega)\right| d\lambda d\omega\\
&\quad\leq \sup_{\|x\|=1} b_{1,T}^{-1}\sup_{\omega\in\Pi}|K(\omega)|\int_{\Pi} L_{2,T}(\lambda)\left(\int_{\Pi}|D_T(-\omega)|\,|X_T(\omega)|\,|X_T(\lambda)|\right)d\lambda d\omega\\
&\quad\leq \sup_{\|x\|=1} b_{1,T}^{-1}\sup_{\omega\in\Pi}|K(\omega)|\left(\int_{\Pi} L_{2,T}(\lambda)^2\,d\lambda\right)^{1/2}\left(\int_{\Pi}|X_T(\lambda)|^2\,d\lambda\right)^{1/2}\\
&\qquad\times\left(\int_{\Pi}|D_T(-\omega)|^2\,d\omega\right)^{1/2}\left(\int_{\Pi}|X_T(\omega)|^2\,d\omega\right)^{1/2}\\
&\quad\leq 2\pi C_2 b_{1,T}^{-1}\sup_{\omega\in\Pi}|K(\omega)|\log^2 T,
\end{aligned} \tag{S.B.51}$$

where $0 < C_2 < \infty$ and we have used $\sup_{\omega \in \Pi} |K(\omega)| = O(b_{1,T}^{-1})$, $(\int_\omega |X_T(\omega)|^2 d\omega) = 2\pi$ and (S.B.50). For $|\lambda| > \epsilon$ the contribution to A_1 is bounded by

$$\begin{aligned}
&\sup_{\|x\|=1} \int_{\omega \leq \epsilon} \sum_{t=1}^{T} \left(\int_\Pi (f(t/T, \lambda))^p d\lambda \right)^{1/p} \left(\int_\Pi |e^{it\lambda} X_T(\lambda)|^{\frac{p}{p-1}} d\lambda \right)^{(p-1)/p} \left| D_T(-\omega) X_T(\omega) \widetilde{K}_{b_1}(\omega) \right| d\omega d\omega \\
&\leq C_2 \sup_{\|x\|=1} \sum_{t=1}^{T} \left(\int_\Pi |e^{it\lambda} X_T(\lambda)|^{\frac{p}{p-1}} d\lambda \right)^{(p-1)/p} \int_{\omega \leq \epsilon} \left| D_T(-\omega) X_T(\omega) \widetilde{K}_{b_1}(\omega) \right| d\omega \\
&\leq C_2 \sup_{\|x\|=1} \sum_{t=1}^{T} \left(\int_\Pi |e^{it\lambda}|^{\frac{p}{p-1}} d\lambda \right)^{(p-1)/p} \int_{\omega \leq \epsilon} \left(\int_\Pi |X_T(\lambda)|^{\frac{p}{p-1}} d\lambda \right)^{(p-1)/p} \\
&\quad \times \left(\int_\Pi |D_T(-\omega)| d\omega \right) \left(\int_\Pi |X_T(\omega)|^2 d\omega \right)^{1/2} \left(\int_\Pi \left| \widetilde{K}_{b_1}(\omega) \right|^2 d\omega \right)^{1/2} \\
&\leq \sqrt{2\pi} C_2 \left(\sup_\omega |K(\omega)| \right)^{1/2} \|K\|_1 (2\pi)^{(p-1)/p} T^{\frac{2-p}{2p}} b_{1,T}^{-1} \log^2 T,
\end{aligned} \tag{S.B.52}$$

where $0 < C_2 < \infty$ and we have used $\sup_{x,\lambda} |X_T(\lambda)| \leq \sqrt{T}$ and

$$\begin{aligned}
\left(\int_\Pi |X_T(\lambda)|^{\frac{p}{p-1}} d\lambda \right)^{(p-1)/p} &= \left(\int_\Pi |X_T(\lambda)|^{2 + \frac{2-p}{p-1}} d\lambda \right)^{(p-1)/p} \\
&= \left(\int_\Pi |X_T(\lambda)|^2 |X_T(\lambda)|^{\frac{2-p}{p-1}} d\lambda \right)^{(p-1)/p} \\
&\leq \left(\int_\Pi |X_T(\lambda)|^2 T^{\frac{1}{2}\left(\frac{2-p}{p-1}\right)} d\lambda \right)^{(p-1)/p} \\
&\leq (2\pi)^{(p-1)/p} T^{\frac{2-p}{2p}}.
\end{aligned}$$

From (S.B.51)-(S.B.52) we have $A_1 \leq C_1 \nu_{2,T}$ for some C_1 such that $0 < C_1 < \infty$. □

Lemma S.B.8. *Let Assumption 4.1, 4.2 (for some $p > 1$), 4.3, 4.4 and $b_{1,T} + T^{-1} b_{1,T}^{-1} \log^3 T \to 0$ hold. Then, there exists $c_2 > 0$ such for $\|\mathbf{t}\| > c_1 m_T$ with $c_1 > 0$ we have $|\psi(\mathbf{t})| \leq \exp\{-c_2 m_T^2\}$, where $m_T = \min\{(Tb_{1,T})^{-1/2} \log T, T^{(p-1)/p}\} \to \infty$.*

Proof of Lemma S.B.8. The proof is similar to the proof of Lemma 15 in Velasco and Robinson (2001) with the difference that reference to Lemma 16 there is changed to reference to Lemma S.B.7. □

Lemma S.B.9. *Let Assumption 4.1, 4.2 ($p > 1$), 4.3-4.4, 4.7 ($0 < q < 1$) and 4.8-4.9 hold. Then, $\|\Sigma_{\widetilde{V}} W_{b_1}\| \leq C_1 \nu_{2,T}$ where C_1 depends on $f(u, \omega)$ and K, $0 < C_1 < \infty$ and $\nu_{2,T} = \max\{b_{1,T}^{-1} \log(Tb_{2T}), (Tb_{2,T})^{(2-p)/2p}$ $b_{1,T}^{-1/2})\} \to \infty$.*

Proof of Lemma S.B.9. The proof is similar to the proof of Lemma S.B.7. □

Lemma S.B.10. *Let Assumption 4.1, 4.2 ($p > 1$), 4.3-4.4, 4.8-4.9 and $b_{1,T} + (Tb_{1,T} b_{2,T})^{-1} \log^3 T \to 0$ hold. Then, there exists a $c_4 > 0$ such for $\|\mathbf{t}\| > c_3 m_{2,T}$ with $c_3 > 0$ we have $|\psi(t_1, t_2)| \leq \exp(-c_4 m_{2,T}^2)$, where $m_{2,T} = \min\{(Tb_{2,T} b_{1,T})^{1/2} / \log(Tb_{2,T}), (Tb_{2,T})^{(p-1)/p}\} \to \infty$.*

Proof of Lemma S.B.10. Following Bentkus and Rudzkis (1982) and Velasco and Robinson (2001) we first study the characteristic function of $\widehat{J}_{\mathrm{DK},T}$. Define $\tau(t_2) = \mathbb{E}(\exp(it_2 v_2)) = \tau'(t_2)\exp(-it_2\Upsilon_{2,T})$, where

$$\tau'(t_2) = \left| I - \frac{2it_2}{\sqrt{Tb_{2,T}/b_{1,T}}\mathsf{V}_{2,T}J_T}\Sigma_{\widetilde{V}}W_{b_1}\right|^{-1/2} = \prod_{j=1}^{T}\left(1 - 2it_2\frac{\widetilde{\lambda}_j}{\sqrt{Tb_{2,T}/b_{1,T}}\mathsf{V}_{2,T}J_T}\right)^{-1/2},$$

and $\widetilde{\lambda}_j$ are the eigenvalues of $\Sigma_{\widetilde{V}}W_{b_1}$. Note that

$$1 = \mathrm{Var}(v_2) = \frac{b_{1,T}}{Tb_{2,T}}\frac{1}{\mathsf{V}_{2,T}^2J_T^2}2\mathrm{Tr}\left[(\Sigma_{\widetilde{V}}W_{b_1})^2\right] = \frac{b_{1,T}}{Tb_{2,T}}\frac{2}{\mathsf{V}_{2,T}^2J_T^2}\sum_{j=1}^{T}\widetilde{\lambda}_j^2,$$

where we have used the normality of $\{V_t\}$ and the relationship between the trace and the eigenvalues. Rearranging yields $\sum_{j=1}^{T}\widetilde{\lambda}_j^2 = 2^{-1}b_{1,T}^{-1}Tb_{2,T}\mathsf{V}_{2,T}^2J_T^2 = O(b_{1,T}^{-1}Tb_{2,T})$. Further, we have $\max_j|\widetilde{\lambda}_j| = \sup_{\|x\|=1}|\Sigma_{\widetilde{V}}W_{b_1}x,\, x| = ||\Sigma_{\widetilde{V}}W_{b_1}||$. We can apply Lemma S.B.9 to yield

$$\max_j\left|\widetilde{\lambda}_j\right| \leq C_1\nu_{2,T}, \qquad \nu_{2,T} = \max\left\{b_{1,T}^{-1}\log(Tb_{2T}),\, (Tb_{2,T})^{(2-p)/2p}b_{1,T}^{-1/2}\right\} \to \infty,$$

where $C_1 > 0$ is such that $C_1 < \infty$. Let $g_j = \widetilde{\lambda}_j(C_1\nu_{2,T})^{-1}$ and note that for T large enough we have $|g_j| \leq 1$. Using $\sum_{j=1}^{T} g_j^2 = (2C_1^2\nu_{2,T}^2)\mathsf{V}_{2,T}^2J_T^2b_{1,T}^{-1}Tb_{2,T}$ we yield

$$\begin{aligned}
|\tau(t_2)| &\leq \prod_{j=1}^{T}\left(1 + 4t^2\frac{C_1^2\nu_{2,T}^2}{b_{1,T}^{-1}Tb_{2,T}\mathsf{V}_{2,T}^2J_T^2}\right)^{-(1/4)g_j^2} \\
&= \left(1 + t_2^2\frac{\nu_{2,T}^2}{b_{1,T}^{-1}Tb_{2,T}}\frac{4C_1^2}{\mathsf{V}_{2,T}^2J_T^2}\right)^{-(1/8)C_1^{-2}\mathsf{V}_{2,T}^2J_T^2b_{1,T}^{-1}Tb_{2,T}\nu_{2,T}^{-2}} \\
&= \left(1 + t_2^2\frac{\nu_{2,T}^2}{b_{1,T}^{-1}Tb_{2,T}}\left[C_2 + O\left(b_{1,T}^2 + \epsilon_{Tb_{2,T}}(2)\right)\right]\right)^{-(1/2)\left(C_2^{-1} + O\left(b_{1,T}^2 + \epsilon_{Tb_{2,T}}(2)\right)\right)Tb_{2,T}b_{1,T}^{-1}\nu_{2,T}^{-2}},
\end{aligned}$$

where $C_2 = C_1^2/(\pi^3 4(\int_0^1 f(u,0)\,du)^2\,\|K\|_2^2\,\|K_2\|_2^2)$ and we have applied $(1+at) \geq (1+t)^a$ which is valid for $t \geq 0$ and $0 \leq a \leq 1$. Thus, for all $\eta > 0$, we have

$$|\tau(t_2)| \leq \left(1 + \eta_1^2\right)^{-\eta_2\left(Tb_{2,T}b_{1,T}^{-1}\nu_{2,T}^{-2}\right)}, \tag{S.B.53}$$

for $|t_2| > \eta\sqrt{Tb_{2,T}b_{1,T}^{-1}}\nu_{2,T}^{-1}$ and for $\eta_1 > 0$ and $\eta_2 > 0$ depending on η.

Next, we consider the joint characteristic function $\psi_T(t_1,\, t_2)$. Its modulus is equal to

$$|\psi_T(t_1,\, t_2)| = |\tau(t_2)|\exp\left(-\frac{1}{2}t_1^2\xi_{2,T}'\mathscr{R}\left(I - 2it_2\Sigma_{\widetilde{V}}Q_{2,T}\right)^{-1}\Sigma_{\widetilde{V}}\xi_{2,T}\right), \tag{S.B.54}$$

where $\mathscr{R}(A)$ stands for the real part of A. From Anderson (1958, p. 161) $\mathscr{R}(\Sigma_{\widetilde{V}}^{-1} - 2it_2Q_{2,T})^{-1} = \mathscr{R}(I - 2it_2Q_{2,T})^{-1}\Sigma_{\widetilde{V}}$ is positive definite since $t_2Q_{2,T}$ is real. Then $\xi_{2,T}'\mathscr{R}(I - 2it_2\Sigma_{\widetilde{V}}Q_{2,T})^{-1}\Sigma_{\widetilde{V}}\xi_{2,T} > 0$ for all $t_2 \in \mathbb{R}$. Thus, $|t_2| \leq d\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}$ for all $d>0$ and $\xi_{2,T}'\mathscr{R}(I-2it_2\Sigma_{\widetilde{V}}Q_{2,T})^{-1}\Sigma_{\widetilde{V}}\xi_{2,T} > \epsilon$ for some $\epsilon>0$ depending on d because $||\Sigma_{\widetilde{V}}Q_{2,T}|| = O(Tb_{2,T}b_{1,T}^{-1})^{-1/2}||\Sigma_{\widetilde{V}}W_{b_1}|| = (O(Tb_{2,T}b_{1,T}^{-1})^{-1/2}\nu_{2,T})$, and $||\xi_{2,T}|| = (\sqrt{Tb_{2,T}J_T})^{-1}\sqrt{1^2+1^2+\ldots+1^2} = 1/\sqrt{b_{2,T}J_T}$, with $J_T \to 2\pi\int_0^1 f(u,0)\,du$, $0<f(u,0)<\infty$ for all u by Assumption 4.1. Then, for $|t_1|\sqrt{2} > d_1\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}$ and $|t_2|\sqrt{2} \leq d_1\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}$ and some $\epsilon_1>0$ depending on d_1,

$$\exp\left(-\frac{1}{2}t_1^2\xi_{2,T}'\mathscr{R}\left(I-2it_2\Sigma_{\widetilde{V}}Q_{2,T}\right)^{-1}\Sigma_{\widetilde{V}}\xi_{2,T}\right) \leq \exp\left(-\frac{1}{2}t_1^2\epsilon_1\right) \leq \exp\left(-\frac{1}{4}d_1^2\epsilon_1\frac{Tb_{2,T}b_{1,T}^{-1}}{\nu_{2,T}^2}\right). \quad \text{(S.B.55)}$$

From (S.B.53)-(S.B.55), there exists a $d_2>0$ such that $|\psi_T(\mathbf{t})| \leq \exp(-d_2(Tb_{2,T}b_{1,T}^{-1}/\nu_{2,T}^2))$ for $\{\mathbf{t} : ||\mathbf{t}|| > d_1\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}\} \subset \mathbf{B}_1 \cup \mathbf{B}_2$ where $\mathbf{B}_1 = \{\mathbf{t}\in\mathbb{R}^2 : |t_2| > (d_1/\sqrt{2})\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}\}$ and $\mathbf{B}_2 = \{\mathbf{t}\in\mathbb{R}^2 : |t_2| \leq (d_1/\sqrt{2})\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}$ and $|t_1| > (d_1/\sqrt{2})\sqrt{Tb_{2,T}b_{1,T}^{-1}}/\nu_{2,T}\}$, and the lemma follows because $Tb_{2,T}b_{1,T}^{-1}/\nu_{2,T}^2 = m_{2,T}^2 \to \infty$. □

S.B.2.2 Additional Lemmas Used for the Proofs of Theorem 4.1-4.2

We first present a result about the limit of J_T and a result about the bias of $\widehat{J}_{\mathrm{HAC},T}$.

Lemma S.B.11. *Let Assumption 4.1 with $d_f=1$ and $\varrho=0$ hold. Then, $J_T - 2\pi\int_0^1 f(u,0)\,du = O(T^{-1}\log T)$. If in addition Assumption 2.2-(i) holds, then the order is $O(T^{-1})$.*

Lemma S.B.12. *Let Assumption 4.1, 4.3, 4.5, and 4.6 hold. Then,*

$$\mathbb{E}\left(\widehat{J}_{\mathrm{HAC},T}\right) - 2\pi\int_0^1 f(u,0)\,du - 2\pi\frac{\int_0^1 f^{(d_f)}(u,0)\,du}{d_f!}\mu_{d_f}(K)\,b_{1,T}^{d_f} = O\left(T^{-1}\log T + b_{1,T}^{d_f+\varrho}\right).$$

We now study the cumulants of the normalized spectral estimate h_2.

Lemma S.B.13. *Let Assumption 4.1, 4.3-4.4 hold. For $s>2$ with $\epsilon_T(s) = b_{1,T}^{d_f+\varrho} + T^{-1}b_{1,T}\log^{2s-1}T \to 0$, we have*

$$\overline{\kappa}_T(0,s) \triangleq \kappa_T(0,s)\left(\frac{T}{b_{1,T}}\right)^{(s-2)/2} = \sum_{j=0}^{d_f}\Xi_j(0,s)\,b_{1,T}^j + O(\epsilon_T(s)),$$

where $\Xi_j(0,s)$ is bounded and depends on K and $f^{(j)}(u,0)$ $(j=0,\ldots,d_f)$.

A few examples of $\Xi_j(0,s)$ are $\Xi_0(0,s) = (4\pi)^{(s-2)/2}s!\int_\Pi K^s(\omega)\,d\omega\,\|K\|_2^{-s}$ and $\Xi_1(2,s)=0$. If $(\partial/\partial\omega)(\int_0^1 f(u,\omega)\,du)|_{\omega=0} = 0$ then $\Xi_j(0,s)=0$ for $j\geq 1$. In order to develop an Edgeworth expansion to approximate the distribution of $\mathbf{h}$, we need to study the cross-cumulants of $\mathbf{h}$.

Lemma S.B.14. *Let Assumption 4.1 and 4.3-4.4 hold. For $s>0$ with $\epsilon_T(s+2)\to 0$, we have*

$$\overline{\kappa}_T(2,s) \triangleq \kappa_T(2,s)\,(Tb_{1,T})^{s/2} = \sum_{j=0}^{d_f}\Xi_j(2,s)b_{1,T}^j + O(\epsilon_T(s+2)),$$

where $\Xi_j(2,\, s)$ is bounded and depends on K and $f^{(j)}(u,\, 0)$ $(j = 0, \ldots,\, d_f)$.

For example, we have $\Xi_0(2,\, s) = (4\pi)^{s/2}\, s! K^s(0)\, \|K\|_2^{-s}$ and $\Xi_1(2,\, s) = 0$. Using Lemma S.B.12-S.B.13 we can substitute out B_T and V_T in Z_T and, by only focusing on the leading terms, we define the following linear stochastic approximation,

$$\widetilde{Z}_T \triangleq h_1 \left(1 - 2^{-1}\overline{c}_1 b_{1,T}^{d_f} - 2^{-1}\sqrt{4\pi}\, \|K_2\|\, h_2\,(Tb_{1,T})^{-1/2}\right).$$

Lemma S.B.15. *Let Assumption 4.1, 4.2 ($p > 1$), 4.3-4.5 and 4.7 ($q = 1/(1 + 2d_f)$) hold. Then, Z_T has the same Edgeworth expansion as $\widetilde{Z}_T$ uniformly for convex Borel sets up to order $O((Tb_{1,T})^{-1/2})$.*

Note that the condition $q = 1/(1+2d_f)$ is sufficient for the consistency of $\widehat{J}_{\mathrm{HAC},T}$. Indeed, for $d_f = 2$ it implies that $b_{1,T} = T^{-1/5}$ which coincides with the MSE-optimal bandwidth choice for the quadratic spectral kernel [cf. Andrews (1991)].[9]

S.B.2.3 Proof of Lemma S.B.11

Note that $J_T = \sum_{k=-T+1}^{T-1} \Gamma_T(k)$ where $\Gamma_T(k) = T^{-1} \sum_{t=|k|+1}^{T} \mathbb{E}(V_t V_{t-|k|})$. We have

$$\begin{aligned}
J_T &= \sum_{k=-T+1}^{T-1} \frac{1}{T} \sum_{t=|k|+1}^{T} \int_{\Pi} f(t/T,\, \omega)\, e^{ik\omega} d\omega \\
&= \sum_{k=-T+1}^{T-1} \frac{T-|k|}{T} \int_{|k|/T}^{1} \int_{\Pi} f(u,\, \omega)\, e^{ik\omega} d\omega du + O\left(T^{-1}\right) \\
&= 2\pi \int_0^1 \int_{\Pi} f(u,\, \omega)\, \Psi_T^{(2)}(\omega)\, d\omega du + O\left(T^{-1}\right).
\end{aligned}$$

Since $\int_{\Pi} \Psi_T^{(2)}(\omega)\, d\omega = 1$, we can apply the mean value theorem for $f(u,\, \omega)$ in a small interval $[-\epsilon,\, \epsilon]$, $\epsilon > 0$, for some $|\eta| \leq 1$ depending on ω,

$$\begin{aligned}
\left| J_T - 2\pi \int_0^1 f(u,\, 0)\, du \right| &\leq 2\pi \left(\int_{|\omega| \leq \epsilon} + \int_{|\omega| > \epsilon} \right) \int_0^1 \int_{\Pi} |f(u,\, \omega) - f(u,\, 0)| \left|\Psi_T^{(2)}(\omega)\right| d\omega du + O\left(T^{-1}\right) \\
&= O\left(\int_{|\omega| \leq \epsilon} \int_0^1 |\omega|\, |f^{(1)}(u,\, \omega\eta)|\, \left|\Psi_T^{(2)}(\omega)\right| du d\omega \right. \\
&\quad \left. + \left(\int_0^1 (||f(u,\, \omega)||_1 + f(u,\, 0))\, du \right) T^{-1} \right) + O\left(T^{-1}\right) \\
&= O\left(T^{-1} \log T\right) + O\left(T^{-1}\right),
\end{aligned}$$

where we have used Assumption 4.1,

$$\left|\Psi_T^{(2)}(\omega)\right| \leq \frac{1}{2\pi T} \left|D_T(\omega)\right| \left|D_T(-\omega)\right| \leq \frac{1}{\pi T} \left|\omega^{-2}\right|,$$

[9] Note that the MSE bounds under nonstationarity in Section 8 in Andrews (1991), which are used to determine the optimal bandwidth, are not correctly stated [cf. Casini (2022a)].

from (S.B.18)-(S.B.19) and $|\Psi_T^{(2)}(\omega)| \leq O((T)^{-1})$ if $|\omega| > \epsilon$.

For the second result in the lemma, note that

$$J_T = \sum_{k=-T+1}^{T-1} T^{-1} \sum_{t=|k|+1}^{T} \mathbb{E}\left(V_t V_{t-|k|}\right) = -\sum_{k=-T+1}^{T-1} T^{-1} \sum_{t=1}^{|k|} \mathbb{E}\left(V_t V_{t-|k|}\right) + \sum_{k=-T+1}^{T-1} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(V_t V_{t-|k|}\right).$$

Then,

$$\begin{aligned}\left|J_T - 2\pi \int_0^1 f(u,\,0)\,du\right| &\leq \left|\sum_{k=-T+1}^{T-1} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(V_t V_{t-|k|}\right) - 2\pi \int_0^1 f(u,\,0)\,du\right| + \left|\sum_{k=-T+1}^{T-1} T^{-1} \sum_{t=1}^{k} \mathbb{E}\left(V_t V_{t-|k|}\right)\right|, \\ &= O\left(T^{-1}\right),\end{aligned}$$

using Assumption 2.2-(i). □

S.B.2.4 Proof of Lemma S.B.12

We can write $\widehat{J}_{\mathrm{HAC},T} = 2\pi \int_\Pi \widetilde{K}_{b_1}(\omega)\, I_T(\omega)\, d\omega$. Note that

$$\mathbb{E}\left(I_T(\omega)\right) = \int_0^1 \int_\Pi f(u,\,\lambda)\, \Psi_T^{(2)}(\omega - \lambda)\, d\lambda du + O\left(T^{-1}\right).$$

Thus, we obtain

$$\mathbb{E}\left(\widehat{J}_{\mathrm{HAC},T}\right) = 2\pi \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 \int_\Pi f(u,\,\alpha+\omega)\, \Psi_T^{(2)}(\alpha)\, d\alpha du d\omega + O\left(T^{-1}\right).$$

Then, using $\int_\Pi \Psi_T^{(2)}(\omega)\, d\omega = 1$ and $\int_\Pi \widetilde{K}_{b_1}(\omega)\, d\omega = 1$ we have

$$\begin{aligned}&\mathbb{E}\left(\widehat{J}_{\mathrm{HAC},T}\right) - 2\pi \int_0^1 f(u,\,0)\,du - 2\pi b_{1,T}^{d_f} \mu_{d_f}(K) \int_0^1 \frac{f^{(d_f)}(u,\,0)}{d_f!} du \\ &\quad = 2\pi \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 \int_\Pi \Psi_T^{(2)}(\alpha)\left(f(u,\,\omega+\alpha) - f(u,\,\omega)\right) d\alpha du d\omega \\ &\qquad + \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 \left[f(u,\,\omega) - f(u,\,0) - b_{1,T}^{d_f} \mu_{d_f}(K) \frac{f^{(d_f)}(u,\,0)}{d_f!}\right] du d\omega + O\left(T^{-1}\right) \\ &\quad \triangleq A_1 + A_2 + O\left(T^{-1}\right).\end{aligned}$$

For $\epsilon > 0$, we introduce the sets $\mathbf{A} = \{|\alpha|,\, |\omega| \leq \epsilon/2\}$ and its complement $\mathbf{A}^c$, both defined in Π^2. Let A_{11} and A_{12} be the contributions to A_1 corresponding to $\mathbf{A}$ and $\mathbf{A}^c$, respectively. Then, applying the mean value theorem we have

$$\begin{aligned}|A_{11}| &= 2\pi \int_{|\omega| \leq \epsilon/2} \left|\widetilde{K}_{b_1}(\omega)\, d\omega\right| d\omega \int_{|\alpha| \leq \epsilon/2} \left|\Psi_T^{(2)}(\alpha)\right| |\alpha|\, d\alpha \int_0^1 \sup_{|\omega| \leq \epsilon} \left|f^{(1)}(u,\,\omega)\right| du \\ &= O\left(T^{-1} \log T\right),\end{aligned}$$

where we have used (S.B.18)-(S.B.19) and Assumption 4.1. Let $\mathbf{B}_1 = \{|\alpha| > \epsilon/2\}$ and $\mathbf{B}_2 = \{|\omega| > \epsilon/2,\, |\alpha| \le \epsilon/2\}$ and note that $\mathbf{A}^c \subset \{\mathbf{B}_1 \cup \mathbf{B}_2\}$. The contribution to A_{12} from $\mathbf{B}_1$ is

$$\begin{aligned}
&\left|\int_{|\alpha|>\epsilon/2} \Psi_T^{(2)}(\alpha) \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 (f(u,\,\omega+\alpha) - f(u,\,\omega))\, du d\omega d\alpha\right| \\
&\quad = O\left(T^{-1} \int_{\Pi^2}\int_0^1 \left|\widetilde{K}_{b_1}(\omega)(f(u,\,\omega+\alpha) - f(u,\,\omega))\right| du d\omega d\alpha\right) \\
&\quad = O\left(T^{-1}\left(1 + \int_{|\omega|\le\epsilon}\int_0^1 \left|\widetilde{K}_{b_1}(\omega) f(u,\,\omega)\right| du d\omega\right)\right) \\
&\quad = O\left(T^{-1}\int_\Pi \left|\widetilde{K}_{b_1}(\omega)\right| d\omega\right),
\end{aligned} \tag{S.B.56}$$

using (S.B.18)-(S.B.19) and Assumption 4.1. Since $\widetilde{K}_{b_1}(\omega)$ is of reduced magnitude for $\omega > \epsilon/2$, the contribution to A_{12} from $\mathbf{B}_2$ is, for large T,

$$\left|\int_{|\omega|>\epsilon/2}\int_{|\alpha|\le\epsilon/2} \widetilde{K}_{b_1}(\omega)\Psi_T^{(2)}(\alpha)\int_0^1 (f(u,\,\omega+\alpha) - f(u,\,\omega))\, du d\alpha d\omega\right| = 0, \tag{S.B.57}$$

This implies that $A_{12} = O(T^{-1})$.

As for A_2 we apply a Taylor's expansion of $f(u,\,\omega)$ around $\omega = 0$ and we split the integral into two parts for $|\omega| \le \epsilon$ and $|\omega| > \epsilon$, denoted as A_{21} and A_{22}, respectively. We have for $|\eta| \le 1$ depending on ω,

$$\begin{aligned}
A_{21} &= \int_{|\omega|\le\epsilon} \widetilde{K}_{b_1}(\omega)\int_0^1\left(\sum_{j=1}^{d_f-1} f^{(j)}(u,\,0)\frac{\omega^j}{j!} + f^{(d_f)}(u,\,\eta\omega)\frac{\omega^{d_f}}{d_f!} - \frac{f^{(d_f)}(u,\,0)}{d_f!}\mu_{d_f}(K)\, b_{1,T}^{d_f}\right) du d\omega \\
&= \sum_{j=1}^{d_f-1}\int_\Pi \omega^j \widetilde{K}_{b_1}(\omega)\, d\omega \int_0^1 f^{(j)}(u,\,0)\frac{1}{j!}du \\
&\quad + d_f^{-1}\int_{|\omega|\le b_{1,T}\pi} \omega^{d_f}\widetilde{K}_{b_1}(\omega)\int_0^1\left(f^{(d_f)}(u,\,\eta\omega) - f^{(d_f)}(u,\,0)\right) du d\omega \\
&= O\left(\int_{|\omega|\le b_{1,T}\pi}\left|\widetilde{K}_{b_1}(\omega)\right| |\omega|^{d_f+\varrho}\, d\omega\right) = O\left(b_{1,T}^{d_f+\varrho}\right),
\end{aligned}$$

where we have used Assumption 4.5 and the fact that as $b_{1,T} \to 0$ the integration is within $[-\epsilon,\,\epsilon]$ and that by Assumption 4.1 $f^{(d_f)}(u,\,0)$ is Lipschitz continuous of order ϱ for all $u \in [0,\,1]$. We can use the same argument used for A_{12} to show that $A_{22} = 0$. □

S.B.2.5 Proof of Lemma S.B.13

From the definition of Q_T, we have

$$\kappa_T(0,\,s) = 2^{s-1}(s-1)!\,(\mathsf{V}_T J_T)^{-s}(T/b_{1,T})^{-s/2}\,\mathrm{Tr}((\Sigma_V W_{b_1})^s),$$

for $s>1$. By Lemma S.B.3,

$$\overline{\kappa}_T(0,\,s)=\kappa_T(0,\,s)\,(b_{1,T}T)^{(s-2)/2}=\frac{2^{s-1}\,(s-1)!\,(2\pi)^{2s-1}}{(\mathsf{V}_TJ_T)^s}\left(\sum_{j=0}^{d_f}L_j\,(s)\,b_{1,T}^j+O\,(\epsilon_T\,(2s))\right). \tag{S.B.58}$$

Using again Lemma S.B.3 with $s=2$ to evaluate V_T^2 yields

$$\begin{aligned}\mathsf{V}_T^2\frac{J_T^2}{4\pi^2}&=\frac{1}{4\pi^2}Tb_{1,T}\mathrm{Var}\left(\widehat{J}_{\mathrm{HAC},T}\right)=\frac{1}{4\pi^2}Tb_{1,T}\mathrm{Var}\left(\mathbf{V}'\frac{W_{b_1}}{T}\mathbf{V}\right)\\&=\frac{2b_{1,T}}{4\pi^2T}\mathrm{Tr}\left(W_{b_1}^2\Sigma_V^2\right)=\frac{2b_{1,T}}{4\pi^2T}\left(T\,(2\pi)^3\sum_{j=0}^{d_f}L_j\,(2)\,b_{1,T}^{j-1}+Tb_{1,T}^{-1}\epsilon_T\,(2)\right)\\&=4\pi\sum_{j=0}^{d_f}L_j\,(2)\,b_{1,T}^j+\epsilon_T\,(2)\,,\end{aligned}$$

where we have use the normality of V_t. Lemma S.B.3 implies that $0<L_0\,(2)<\infty$ and $L_j\,(2)$ are fixed constants independent of T. Then

$$\left(\mathsf{V}_T\frac{J_T}{2\pi}\right)^{-s}=(4\pi)^{-s/2}\sum_{j=0}^{d_f}H_j\,(s)\,b_{1,T}^j+O\,(\epsilon_T\,(s))\,, \tag{S.B.59}$$

where $H_0\,(s)=L_0\,(2)^{-s/2}$ and so on. Denoting $c\,(0,\,s)=(4\pi)^{(s-2)/2}\,(s-1)!$ and using (S.B.58)-(S.B.59) we yield the following expression for the cumulants, $\overline{\kappa}_T(0,\,s)=c\,(0,\,s)\sum_{j=0}^{d_f}P_j\,(s)\,b_{1,T}^j+O\,(\epsilon_T\,(s))$, where $P_j\,(s)=\sum_{t=0}^{j}H_t\,(s)\,L_{j-t}\,(s)$ are constants not depending on T with $P_1\,(s)=0$, $P_2\,(s)=H_0\,(s)\,L_2\,(s)+J_2\,(s)\,L_0\,(s)$, and so on. Setting $\Xi_j\,(0,\,s)=c\,(0,\,s)\,P_j\,(s)$ the lemma follows. $\square$

S.B.2.6 Proof of Lemma S.B.14

Note that for $s>0$ we have

$$\kappa_T(2,\,s)=2^ss!\xi_T'\,(\Sigma_VQ_T)^s\,\Sigma_V\xi_T=2^ss!\frac{1}{TJ_T}\frac{b_{1,T}^{s/2}}{T^{s/2}\mathsf{V}_T^sJ_T^s}\mathbf{1}'\,(W_{b_1}\Sigma_V)^s\,\Sigma_V\mathbf{1}.$$

From Lemma S.B.4,

$$\begin{aligned}\overline{\kappa}_T(2,\,s)&=(Tb_{1,T})^{s/2}\,2^ss!\frac{1}{TJ_T}\frac{b_{1,T}^{s/2}}{T^{s/2}\mathsf{V}_T^sJ_T^s}\mathbf{1}'\,(W_{b_1}\Sigma_V)^s\,\Sigma_V\mathbf{1}\\&=(Tb_{1,T})^{s/2}\,2^ss!\frac{1}{TJ_T}\frac{b_{1,T}^{s/2}}{T^{s/2}\mathsf{V}_T^sJ_T^s}\left(T\,(2\pi)^{2s+1}\left(\int_0^1f\,(u,\,0)\,du\right)^{s+1}\left(\widetilde{K}_{b_1}\,(0)\right)^s\right.\\&\quad\left.+\,O\left(b_{1,T}^{-1-s}\log^{2s+1}T\right)\right)\\&=\left(\frac{2\pi}{J_T\mathsf{V}_T}\right)^s\frac{2\pi\int_0^1f\,(u,\,0)\,du}{J_T}\,(4\pi)^s\,s!\left(\int_0^1f\,(u,\,0)\,du\right)^sK\,(0)^s+O\,(\epsilon_T\,(s+2))\,,\end{aligned}$$

where we have used the fact that $\widetilde{K}_{b_1}(0) = b_{1,T}^{-1} K(0)$. Using Lemma S.B.11 and eq. (S.B.59), we yield

$$
\begin{aligned}
\overline{\kappa}_T(2,\, s) &= \left(\frac{2\pi}{J_T \mathsf{V}_T}\right)^s \left(1 + O\left(T^{-1}\log T\right)\right)(4\pi)^s\, s! \left(\int_0^1 f(u,\, 0)\, du\right)^s K(0)^s + O\left(\epsilon_T(s+2)\right) \\
&= (4\pi)^{-s/2} (4\pi)^s\, s! \left(\int_0^1 f(u,\, 0)\, du\right)^s K(0)^s \sum_{j=0}^{d_f} H_j(s)\, b_{1,T}^j + O\left(\epsilon_T(s+2)\right),
\end{aligned}
$$

where the $H_s(j)$ are as in the proof of Lemma S.B.13. The lemma follows by setting $\Xi_j(2,\, s) = (4\pi)^{-s/2} (4\pi)^s\, s!\, (\int_0^1 f(u,\, 0)\, du)^s K(0)^s\, H_j(s)$. ☐

S.B.2.7 Proof of Theorem 4.1

We first construct the approximation for $\psi_T(\mathbf{t})$. It follows from Velasco and Robinson (2001) and Taniguchi and Puri (1996) that only the cumulants $\kappa_T(0,\, s)$ and $\kappa_T(2,\, s)$ are nonzero, and that the cumulant generating function is given by

$$
\log \psi_T(\mathbf{t}) = \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} \frac{(Tb_{1,T})^{(2-s)/2}}{s!} \sum_{|\mathbf{r}|=s} \frac{s!}{r_1! r_2!} \overline{\kappa}_T(r_1,\, r_2)\,(it_1)^{r_1} (it_2)^{r_2} + R_T(\tau), \tag{S.B.60}
$$

where $\mathbf{r} = (r_1,\, r_2)'$ with $r_1 \in \{0,\, 2\}$ and $|\mathbf{r}| = r_1 + r_2$, and

$$
\begin{aligned}
R_T(\tau) &= (Tb_{1,T})^{-\tau/2}\left(R_{0,\tau+2}\,(it_2)^{\tau+2} + R_{2,\tau}\,(it_1)^2 (it_2)^{\tau}\right), && \tau \text{ even}, \\
R_T(\tau) &= (Tb_{1,T})^{-\tau/2} \frac{1}{(\tau+2)!}\left(\overline{\kappa}_T(0,\, \tau+2)\,(it_2)^{\tau+2} + \frac{(\tau+2)(\tau+1)}{2}\overline{\kappa}_T(2,\, \tau)\,(it_1)^2 (it_2)^{\tau}\right) \\
&\quad + (Tb_{1,T})^{-\tau/2}\left(R_{0,\tau+3}\,(it_2)^{\tau+3} + R_{2,\tau+1}\,(it_1)^2 (it_2)^{\tau+1}\right), && \tau \text{ odd},
\end{aligned}
$$

where the $R_{0,j}$ and $R_{2,j}$ are bounded. Using Lemma S.B.13-S.B.14, we have

$$
\begin{aligned}
\log \psi_T(\mathbf{t}) &= \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} \frac{(Tb_{1,T})^{(2-s)/2}}{s!}\left(\overline{\kappa}_T(0,\, s)\,(it_2)^s + \frac{s(s-1)}{2}\overline{\kappa}_T(2,\, s-2)\,(it_1)^2 (it_2)^{s-2}\right) + R_T(\tau) \\
&= \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} (Tb_{1,T})^{(2-s)/2}\left(B_T(s,\, \mathbf{t}) + \left\{(it_2)^s + (it_1)^2 (it_2)^{s-2}\right\} O\left(\epsilon_T(s)\right)\right) + R_T(\tau),
\end{aligned}
$$

where

$$
B_T(s,\, \mathbf{t}) = \frac{1}{s!}\sum_{j=0}^{d_f} b_{1,T}^j \left\{\Xi_j(0,\, s)\,(it_2)^s + \frac{s(s-1)}{2}\Xi_j(2,\, s-2)\,(it_1)^2 (it_2)^{s-2}\right\}.
$$

The approximation of the characteristic function of $\mathbf{u}$ using its cumulant generating function is

$$\mathcal{A}_T(\mathbf{t},\,\tau)=\exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1+\sum_{j=3}^{\tau+1}(Tb_{1,T})^{(2-j)/2}\sum_{\mathbf{r}}\prod_{n=3}^{\tau+1}\left[B_T(n,\,\mathbf{t})\right]^{r_n}\frac{1}{r_3!\cdots r_{\tau+1}!}\right],$$

where $\mathbf{r}=(r_3,\ldots,r_{\tau+1})'$, $r_n\in\{0,1,\ldots\}$, and the summation is over all $\mathbf{r}$ satisfying $\sum_{n=3}^{\tau+1}(n-2)\,r_n=j-2$. To obtain a second-order Edgeworth expansion we set $\tau=2$ and we include in $\mathcal{A}_T(\mathbf{t},\,2)$ terms up to order $(Tb_{1,T})^{-1/2}$,

$$\mathcal{A}_T(\mathbf{t},\,2)=\exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left(1+\overline{B}_T(3,\,\mathbf{t})\,(Tb_{1,T})^{-1/2}\right),\tag{S.B.61}$$

where in $\overline{B}_T(3,\,\mathbf{t})$ includes only the leading term in $b_{1,T}^j$ $(j=0)$ in the expansion for the culumant of order three. Note that the characteristic function of $\mathbb{Q}_T^{(2)}(\cdot)$ is $\mathcal{A}_T(\mathbf{t},\,2)$.

The rest of the proof consists of studying the distance between the true distribution and its Edgeworth approximation. Lemma S.B.16 studies the Edgeworth approximation for the characteristic function for $\|\mathbf{t}\|\leq c_1\sqrt{Tb_{1,T}}$, whereas Lemma S.B.8 analyzes its tail behavior. The desired result follows from the same steps as in Theorem 1 of Velasco and Robinson (2001) which relies on Lemma S.B.2. □

Lemma S.B.16. *Let Assumption 4.1, 4.3, 4.4 and $b_{1,T}+(Tb_{1,T})^{-1}\log^5 T\to 0$ hold. There exists $\delta_1>0$ such that, for $\|\mathbf{t}\|\leq\delta_1\sqrt{Tb_{1,T}}$ and a number $d_1>0$,*

$$|\psi_T(\mathbf{t})-A_T(\mathbf{t},\,2)|\leq\exp\left\{-d_1\|\mathbf{t}\|^2\right\}\widetilde{F}(\|\mathbf{t}\|)\,O\left((Tb_{1,T})^{-1/2}\left(b_{1,T}^2+\epsilon_T(3)\right)+\frac{1}{Tb_{1,T}}\right),$$

where $\widetilde{F}(\|\mathbf{t}\|)$ is a polynomial in $\mathbf{t}$ with bounded coefficients and $\mathcal{A}_T(\mathbf{t},\,2)$ is defined as in (S.B.61).

Proof of Lemma S.B.16. It is similar to the proof of Lemma 14 in Velasco and Robinson (2001). □

S.B.2.8 Proof of Lemma S.B.15

It is similar to the proof of Lemma 5 in Velasco and Robinson (2001). □

S.B.2.9 Proof of Theorem 4.2

Consider the transformation $\mathbf{s}=(s_1,\,s_2)'=(\widetilde{Z}_T(h_1,\,h_2),\,h_2)'=\Delta_T(\mathbf{h})$ say, and its inverse $\mathbf{h}=\Delta_T^{-1}(\mathbf{s})=(h_1^\dagger(s_1,\,s_2),\,s_2)'$. Let $\mathbf{L}_T=\{\mathbf{h}:\,|h_i|<l_1T^\gamma,\,0<\gamma<d_f/(3(1+2d_f)),\,i=1,\,2\}$, where l_i are some fixed constants. Using $(1+x)^{-1}=1-x+x^2-x^3+\ldots$ for $|x|<1$, we have uniformly in the set $\mathbf{L}_T$,

$$h_1^\dagger(\mathbf{s})=s_1\left[1+\frac{1}{2}\bar{c}_1b_{1,T}^{d_f}+\frac{1}{2}\sqrt{4\pi}\,\|K_2\|\,s_2\,(Tb_{1,T})^{-1/2}\right]+o\left((Tb_{1,T})^{-1/2}\right).$$

We have $\mathbb{P}(Z_T\in\mathbf{C})=\mathbb{P}(\mathbf{h}\in\Delta_T^{-1}(\mathbf{C}\times\mathbb{R}))$ and from Theorem 4.1,

$$\sup_{\mathbf{C}}\left|\mathbb{P}\left(\mathbf{h}\in\Delta_T^{-1}(\mathbf{C}\times\mathbb{R})\right)-\mathbb{Q}_T^{(2)}\left(Z_T^{-1}(\mathbf{C}\times\mathbb{R})\right)\right|=o\left((Tb_{1,T})^{-1/2}\right)+\text{cost}\sup_{\mathbf{C}}\mathbb{Q}_T^{(2)}\left(\left(\partial\Delta_T^{-1}(\mathbf{C}\times\mathbb{R})\right)^{2\phi_T}\right),$$

where $\phi_T = (Tb_{1,T})^{-\varpi}$ with $1/2 < \varpi < 1$. The rest of the proof is similar to the proof of Theorem 2 in Velasco and Robinson (2001). □

S.B.3 Additional Lemmas Used for the Proofs of Theorem 4.3-4.4

Lemma S.B.17. *Let Assumption 4.1, 4.3, 4.5-4.6 and 4.8-4.9 hold. Then,*

$$\begin{aligned}\mathbb{E}\left(\widehat{J}^{*}_{\mathrm{DK},T}\right) &- 2\pi\int_0^1 f(u,0)\,du - 2\pi\frac{\int_0^1 f^{(d_f)}(u,0)\,du}{d_f!}\mu_{d_f}(K)\,b_{1,T}^{d_f}\\ &- \pi b_{2,T}^2\int_0^1 x^2K_2(x)\,dx\int_{\widetilde{\mathbf{C}}}\frac{\partial^2}{\partial u^2}f(u,0)\,du - 2\pi b_{2,T}^2\Delta_f(0)\\ &= O\left(b_{1,T}^{d_f+\varrho} + (Tb_{2,T})^{-1}\log(Tb_{2,T})\right) + o\left(b_{2,T}^2\right).\end{aligned}$$

The term $2\pi b_{2,T}^2\Delta_f(0)$ in Lemma S.B.17 is the contribution to the bias due to the local time-smoothing in the neighborhoods involving a discontinuity point.

We now consider the cumulants of the normalized spectral estimate v_2.

Lemma S.B.18. *Let Assumption 4.1, 4.3-4.4 and 4.8-4.9 hold. For $s > 2$ with $\epsilon_{Tb_{2,T}}(s) = b_{1,T}^{d_f+\varrho} + (Tb_{2,T}b_{1,T})^{-1}\log^{2s-1}(Tb_{2,T}) \to 0$, we have*

$$\begin{aligned}\overline{\kappa}_{2,T}(0,s) &\triangleq \kappa_{2,T}(0,s)\,(Tb_{1,T}b_{2,T})^{(s-2)/2}\\ &= \sum_{j=0}^{d_f}\Xi_{2,j}(0,s)\,b_{1,T}^j + b_{2,T}^2\sum_{j=0}^{d_f}\left(\widetilde{\Xi}_{2,j}(0,s) + \widetilde{\widetilde{\Xi}}_{3,j}(0,s)\right)b_{1,T}^j + O\left(\epsilon_{Tb_{2,T}}(s)\right),\end{aligned}$$

where $\Xi_{2,j}(0,s)$ is bounded and depends on K, K_2 and on $f^{(j)}(u,0)$ $(j = 0,\ldots,d_f)$, $\widetilde{\Xi}_{2,j}(0,s)$ is bounded and depends on K, K_2, $f^{(j)}(u,0)$ and $(\partial^2/\partial u^2)\,f(u,\omega)$ and $\widetilde{\widetilde{\Xi}}_{3,j}(0,s)$ is bounded and depends on K, K_2, $f^{(j)}(u,0)$ and $\Delta_f(\omega)$.

We now consider the cross-cumulants of $\mathbf{v}$.

Lemma S.B.19. *Let Assumption 4.1, 4.3-4.4 and 4.8-4.9 hold. For $s > 0$ with $\epsilon_{Tb_{2,T}}(s+2) \to 0$,*

$$\begin{aligned}\overline{\kappa}_{2,T}(2,s) \triangleq \kappa_{2,T}(2,s)\,(Tb_{2,T}b_{1,T})^{s/2} = \sum_{j=0}^{d_f}\left(\Xi_{2,j}(2,s) + b_{2,T}^2\left(\widetilde{\Xi}_{2,j}(2,s) + \widetilde{\widetilde{\Xi}}_{3,j}(2,s)\right)\right)b_{1,T}^j\\ + O\left(\epsilon_{Tb_{2,T}}(s+2)\right),\end{aligned}$$

where $\Xi_{2,j}(2,s)$ is bounded and depends on K, K_2 and $f^{(j)}(u,0)$ $(j = 0,\ldots,d_f)$, $\widetilde{\Xi}_{2,j}(2,s)$ is bounded and depends on K, K_2, $f^{(j)}(u,0)$ and $(\partial^2/\partial u^2)\,f(u,\omega)$, and $\widetilde{\widetilde{\Xi}}_{3,j}(2,s)$ is bounded and depends on K, K_2, $f^{(j)}(u,0)$ and $\Delta_f(\omega)$.

S.B.3.1 Proof of Lemma S.B.17

For $r \in \widetilde{\mathbf{C}}$, using a second-order Taylor's expansion as in the proof of Theorem 7.3 in Casini and Perron (2021a), we yield

$$\begin{aligned}
\mathbb{E}\left(\widetilde{I}_T\left(r,\,\omega\right)\right) &= \mathbb{E}\left(\frac{1}{2\pi T b_{2,T}}\left|\sum_{t=1}^{T}\exp\left(-i\omega t\right)\widetilde{V}_t\left(r\right)\right|^2\right)\\
&= \frac{1}{2\pi}\frac{1}{Tb_{2,T}}\sum_{k=-\lfloor Tb_{2,T}\rfloor+1}^{\lfloor Tb_{2,T}\rfloor-1}\sum_{t=|k|+1}^{T}\int_{\Pi}K_2\left(\frac{\left(Tr-\left(t-k/2\right)\right)/T}{b_{2,T}}\right)f\left(\left(t+k/2\right)/T,\,\lambda\right)e^{ik(\omega-\lambda)}d\lambda\\
&\quad+O\left(\left(Tb_{2,T}\right)^{-1}\log\left(Tb_{2,T}\right)\right)\\
&= \int_{\Pi}f\left(r,\,\lambda\right)\Psi_{Tb_{2,T}}^{(2)}\left(\omega-\lambda\right)d\lambda\\
&\quad+\frac{b_{2,T}^2}{2}\int_0^1 x^2K_2\left(x\right)dx\frac{\partial^2}{\partial u^2}f\left(u,\,\omega\right)|_{u=r}+o\left(b_{2,T}^2\right)+O\left(\left(Tb_{2,T}\right)^{-1}\log\left(Tb_{2,T}\right)\right).
\end{aligned}$$

In a neighborhood of a break point λ_j^0, let $r=\lambda_j^0+sb_{2,T}$ for some $s\in(0,\,1)$. Then,

$$\begin{aligned}
\mathbb{E}\left(\widetilde{I}_T\left(r,\,\omega\right)\right) &= \int_{\Pi}f\left(r,\,\lambda\right)\Psi_{Tb_{2,T}}^{(2)}\left(\omega-\lambda\right)d\lambda\\
&\quad+b_{2,T}\left(\int_0^{1-s}xK_2\left(x\right)dx\frac{\partial}{\partial u_-}f\left(\lambda_j^0,\,\omega\right)+\int_{1-s}^1 xK_2\left(x\right)dx\frac{\partial}{\partial u_+}f\left(\lambda_j^0,\,\omega\right)\right).
\end{aligned}$$

When integrating the last term above over r we have

$$b_{2,T}^2\sum_{j=1}^{m_0}\int_0^1\left(\frac{\partial}{\partial u_-}f\left(\lambda_j^0,\,\omega\right)\int_0^{1-s}xK_2\left(x\right)dx+\frac{\partial}{\partial u_+}f\left(\lambda_j^0,\,\omega\right)\int_{1-s}^1 xK_2\left(x\right)dx\right)ds.$$

Thus, we obtain

$$\begin{aligned}
\mathbb{E}\left(\widehat{J}_{\mathrm{DK},T}^{*}\right) &= 2\pi\int_{\Pi}\widetilde{K}_{b_1}\left(\omega\right)\int_0^1\int_{\Pi}f\left(u,\,\alpha+\omega\right)\Psi_T^{(2)}\left(\alpha\right)d\lambda du d\omega\\
&\quad+\pi b_{2,T}^2\int_0^1 x^2K_2\left(x\right)dx\int_{\Pi}\widetilde{K}_{b_1}\left(\omega\right)\int_{\widetilde{\mathbf{C}}}\frac{\partial^2}{\partial u^2}f\left(u,\,\omega\right)du d\omega\\
&\quad+2\pi b_{2,T}^2\int_{\Pi}\widetilde{K}_{b_1}\left(\omega\right)\Delta_f\left(\omega\right)d\omega+o\left(b_{2,T}^2\right)+O\left(\left(Tb_{2,T}\right)^{-1}\log\left(Tb_{2,T}\right)\right).
\end{aligned}$$

Then, using $\int_{\Pi}\Psi_T^{(2)}\left(\omega\right)d\omega=1$, $\int_{\Pi}\widetilde{K}_{b_1}\left(\omega\right)d\omega=1$, Assumption 4.9 and similar arguments as in the proof of Lemma S.B.12 applied to the terms involving $\frac{\partial^2}{\partial u^2}f\left(u,\,\omega\right)$ and $\Delta_f\left(\omega\right)$, we have

$$\mathbb{E}\left(\widehat{J}_{\mathrm{DK},T}^{*}\right)-2\pi\int_0^1 f\left(u,\,0\right)du-2\pi b_{1,T}^{d_f}\mu_{d_f}\left(K\right)\int_0^1\frac{f^{(d_f)}\left(u,\,0\right)}{d_f!}du$$

$$
\begin{aligned}
& -\pi b_{2,T}^2 \int_0^1 x^2 K_2(x)\, dx \int_{\widetilde{\mathbf{C}}} \frac{\partial^2}{\partial u^2} f(u,\,0)\, du - 2\pi b_{2,T}^2 \Delta_f(0) \\
&= 2\pi \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 \int_\Pi \Psi_T^{(2)}(\alpha)\left(f(u,\,\omega+\alpha) - f(u,\,\omega)\right) d\alpha du d\omega \\
&\quad + 2\pi \int_\Pi \widetilde{K}_{b_1}(\omega) \int_0^1 \left[f(u,\,\omega) - f(u,\,0) - b_{1,T}^{d_f} \mu_d(K) \frac{f^{(d_f)}(u,\,0)}{d_f!} \right] du d\omega \\
&\quad + o\left(b_{2,T}^2\right) + O\left((Tb_{2,T})^{-1} \log(Tb_{2,T})\right) + o\left(b_{2,T}^2 b_{1,T}^{q_2}\right) \\
&\triangleq A_1 + A_2 + o\left(b_{2,T}^2\right) + O\left((Tb_{2,T})^{-1} \log(Tb_{2,T})\right).
\end{aligned}
$$

To conclude the proof, note that by Lemma S.B.12 we have $|A_1| + |A_2| = O\left(T^{-1}\log T\right) + O(b_{1,T}^{d_f+\varrho})$. $\square$

S.B.3.2 Proof of Lemma S.B.18

We have

$$
\kappa_{2,T}(0,\,s) = 2^{s-1}(s-1)!\,(\mathsf{V}_{2,T} J_T)^{-s}\,(Tb_{2,T}/b_{1,T})^{-s/2}\,\mathrm{Tr}((\Sigma_{\widetilde{V}} W_{b_1})^s),
$$

for $s > 1$. By Lemma S.B.5,

$$
\begin{aligned}
\overline{\kappa}_{2,T}(0,\,s) &= \kappa_{2,T}(0,\,s)\,(Tb_{1,T} b_{2,T})^{(s-2)/2} \qquad \text{(S.B.62)} \\
&= \frac{2^{s-1}(s-1)!\,(2\pi)^{2s-1}}{(\mathsf{V}_{2,T} J_T)^s} \left(\sum_{j=0}^{d_f} L_j(s)\, b_{1,T}^j + b_{2,T}^2 \sum_{j=0}^{d_f} \left((L_{2,j}(s) + L_{3,j}(s))\, b_{1,T}^j \right) + O\left(\epsilon_{Tb_{2,T}}(s)\right) \right).
\end{aligned}
$$

Using Lemma S.B.5 to evaluate $\mathsf{V}_{2,T}^2$ yields

$$
\begin{aligned}
\mathsf{V}_{2,T}^2 \frac{J_T^2}{4\pi^2} &= \frac{1}{4\pi^2} Tb_{1,T} b_{2,T} \mathrm{Var}\left(\widehat{J}_{\mathrm{DK},T}^*\right) = Tb_{1,T} b_{2,T} \mathrm{Var}\left(\int_0^1 \widetilde{\mathbf{V}}(r)' \frac{W_{b_1}}{Tb_{2,T}} \widetilde{\mathbf{V}}(r)\, dr \right) \\
&= \frac{2b_{1,T}}{4\pi^2 Tb_{2,T}} \mathrm{Tr}\left(W_{b_1}^2 \Sigma_{\widetilde{V}}^2 \right) \\
&= \frac{2b_{1,T}}{4\pi^2} (2\pi)^3 \left(\sum_{j=0}^{d_f} L_j(2)\, b_{1,T}^{j-1} + b_{2,T}^2 \sum_{j=0}^{d_f} \left((L_{2,j}(s) + L_{3,j}(s))\, b_{1,T}^{j-1} \right) \right) + Tb_{2,T} b_{1,T}^{-1} O\left(\epsilon_{Tb_{2,T}}(2)\right) \\
&= 4\pi \left(\sum_{j=0}^{d_f} L_j(2)\, b_{1,T}^j + b_{2,T}^2 \sum_{j=0}^{d_f} \left((L_{2,j}(s) + L_{3,j}(s))\, b_{1,T}^j \right) \right) + O\left(\epsilon_{Tb_{2,T}}(2)\right),
\end{aligned}
$$

where we have use the normality of $\{V_t\}$. Since Lemma S.B.5 implies that $0 < L_0(2) < \infty$ and $L_j(2)$ are fixed constants independent of T, we then have

$$
\left(\mathsf{V}_{2,T} \frac{J_T}{2\pi} \right)^{-s} = (4\pi)^{-s/2} \sum_{j=0}^{d_f} H_j(2)\, b_{1,T}^j + O\left(\epsilon_{Tb_{2,T}}(2)\right), \qquad \text{(S.B.63)}
$$

where $H_0(s) = L_0(2)^{-s/2}$ and so on. Using (S.B.62)-(S.B.63) we yield

$$\overline{\kappa}_{2,T}(0,\, s) = c\,(0,\, s) \left(\sum_{j=0}^{d_f} P_{2,j}\,(s)\, b_{1,T}^{j} + b_{2,T}^{2} \sum_{j=0}^{d_f} \left(\left(\widetilde{P}_{2,j}\,(s) + \widetilde{P}_{3,j}\,(s) \right) b_{1,T}^{j} \right) \right) + O\left(\epsilon_{Tb_{2,T}}\,(2) \right),$$

where $c\,(0,\, s) = (4\pi)^{(s-2)/2}\,(s-1)!$, $P_{2,j}\,(s) = \sum_{t=0}^{j} H_t\,(s)\, L_{j-t}\,(s)$ are constants not depending on T with $P_{2,1}\,(s) = 0$, $P_{2,2}\,(s) = H_0\,(s)\, L_2\,(s) + H_2\,(s)\, L_0\,(s)$ and so on, and $\widetilde{P}_{2,j}\,(s) = \sum_{t=0}^{j} H_t\,(s)\, L_{2,j-t}\,(s)$ and $\widetilde{P}_{3,j}\,(s) = \sum_{t=0}^{j} H_t\,(s)\, L_{3,j-t}\,(s)$. The lemma follows from setting $\Xi_{2,j}(0,\, s) = c\,(0,\, s)\, P_{2,j}\,(s)$, $\widetilde{\Xi}_{2,j}(0,\, s) = c\,(0,\, s)\, \widetilde{P}_{2,j}\,(s)$ and $\widetilde{\widetilde{\Xi}}_{2,j}(0,\, s) = c\,(0,\, s)\, \widetilde{P}_{3,j}\,(s)$. $\square$

S.B.3.3 Proof of Lemma S.B.19

For $s > 0$ we have

$$\kappa_{2,T}(2,\, s) = 2^s s! \xi_T' \left(\Sigma_{\widetilde{V}} Q_{2,T} \right)^s \Sigma_{\widetilde{V}} \xi_T = 2^s s! \frac{1}{Tb_{2,T} J_T} \frac{b_{1,T}^{s/2}}{(Tb_{2,T})^{s/2} \mathsf{V}_{2,T}^{s} J_T^{s}} \mathbf{1}' \left(W_{b_1} \Sigma_{\widetilde{V}} \right)^s \Sigma_{\widetilde{V}} \mathbf{1}.$$

From Lemma S.B.6, we have

$$\begin{aligned}
\overline{\kappa}_{2,T}(2,\, s) &= (Tb_{1,T} b_{2,T})^{s/2}\, 2^s s! \frac{1}{Tb_{2,T} J_T} \frac{b_{1,T}^{s/2}}{(Tb_{2,T})^{s/2} \mathsf{V}_{2,T}^{s} J_T^{s}} \mathbf{1}' \left(W_{b_1} \Sigma_{\widetilde{V}} \right)^s \Sigma_{\widetilde{V}} \mathbf{1} \\
&= (Tb_{1,T} b_{2,T})^{s/2}\, 2^s s! \frac{1}{Tb_{2,T} J_T} \frac{b_{1,T}^{s/2}}{(Tb_{2,T})^{s/2} \mathsf{V}_{2,T}^{s} J_T^{s}} \\
&\quad \times \left(Tb_{2,T}\, (2\pi)^{2s+1} \left(\left(\left(\int_0^1 f\,(u,\, 0)\, du \right)^{s+1} \int_0^1 K_2^{s+1}\,(x)\, dx + b_{2,T}^2 \widetilde{\Lambda}_2 \left(f'',\, \widetilde{\mathbf{C}},\, s \right) \right. \right. \right. \\
&\quad \left. \left. + b_{2,T}^2 \widetilde{\Lambda}_3 \left(f',\, \left\{ \lambda_j^0,\, j = 1, \ldots,\, m_0 \right\},\, s \right) \right) \left(\widetilde{K}_{b_1}\,(0) \right)^s \\
&\quad \left. + O\left(b_{1,T}^{1-s} \log^{2s+1}(Tb_{2,T}) + b_{1,T}^{-s} \frac{\log^{2s+1}(Tb_{2,T})}{Tb_{2,T}} \right) \right) \\
&= \left(\frac{2\pi}{J_T \mathsf{V}_{2,T}} \right)^s \frac{2\pi \int_0^1 f\,(u,\, 0)\, du}{J_T} (4\pi)^s s! \left(\left(\int_0^1 f\,(u,\, 0)\, du \right)^s \int_0^1 K_2^{s+1}\,(x)\, dx + b_{2,T}^2 \left(\widetilde{\Lambda}_2^* + \widetilde{\Lambda}_3^* \right) \right) K\,(0)^s \\
&\quad + O\left(\epsilon_{Tb_{2,T}}\,(s+2) \right),
\end{aligned}$$

where $\widetilde{\Lambda}_2^*$ and $\widetilde{\Lambda}_3^*$ are equal to $\widetilde{\Lambda}_2$ and $\widetilde{\Lambda}_3$, respectively, without the factor $\int_0^1 f\,(u,\, 0)\, du$, and we have used $\widetilde{K}_{b_1}\,(0) = b_{1,T}^{-1} K\,(0)$. Using Lemma S.B.11 and (S.B.63), we yield

$$\begin{aligned}
\overline{\kappa}_{2,T}(2,\, s) &= \left(\frac{J_T \mathsf{V}_{2,T}}{2\pi} \right)^{-s} \left(1 + O\left((Tb_{2,T})^{-1} \log(Tb_{2,T}) \right) \right) \\
&\quad \times (4\pi)^s s! \left(\left(\int_0^1 f\,(u,\, 0)\, du \right)^s \int_0^1 K_2^{s+1}\,(x)\, dx + b_{2,T}^2 \left(\widetilde{\Lambda}_2^* + \widetilde{\Lambda}_3^* \right) \right) K\,(0)^s + O\left(\epsilon_{Tb_{2,T}}\,(s+2) \right)
\end{aligned}$$

$$= (4\pi)^{-s/2} (4\pi)^{s} s! \left(\left(\int_0^1 f(u,\,0)\,du \right)^{s} \int_0^1 K_2^{s+1}(x)\,dx + b_{2,T}^2 \left(\widetilde{\Lambda}_2^* + \widetilde{\Lambda}_3^* \right) \right) K(0)^s \sum_{j=0}^{d_f} H_j(s)\, b_{1,T}^j$$
$$+ O\left(\epsilon_{Tb_{2,T}}(s+2) \right),$$

where the $H_j(s)$ are as in (S.B.63). Letting

$$\Xi_{2,j}(2,\,s) = (4\pi)^{-s/2} (4\pi)^{s} s! \left(\int_0^1 f(u,\,0)\,du \right)^{s} K(0)^s \int_0^1 K_2^{s+1}(x)\,dx H_j(s)$$
$$\widetilde{\Xi}_{2,j}(2,\,s) = (4\pi)^{-s/2} (4\pi)^{s} s! \widetilde{\Lambda}_2^* K(0)^s \int_0^1 K_2^{s}(x)\,dx H_j(s)$$
$$\widetilde{\Xi}_{3,j}(2,\,s) = (4\pi)^{-s/2} (4\pi)^{s} s! \widetilde{\Lambda}_3^* K(0)^s \int_0^1 K_2^{s}(x)\,dx H_j(s),$$

the lemma follows. □

S.B.3.4 Proof of Theorem 4.3

It follows from Velasco and Robinson (2001) and Taniguchi (1987) that only the cumulants $\kappa_{2,T}(0,\,s)$ and $\kappa_{2,T}(2,\,s)$ are nonzero, and that the cumulant generating function is given by

$$\log \psi_T(\mathbf{t}) = \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} \frac{(Tb_{1,T}b_{2,T})^{(2-s)/2}}{s!} \sum_{|\mathbf{r}|=s} \frac{s!}{r_1! r_2!} \overline{\kappa}_{2,T}(r_1,\,r_2)\,(it_1)^{r_1}(it_2)^{r_2} + R_T^*(\tau), \qquad \text{(S.B.64)}$$

where $\mathbf{r} = (r_1,\,r_2)'$, with $r_1 \in \{0,\,2\}$ and $|\mathbf{r}| = r_1 + r_2$, and

$$R_T^*(\tau) = (Tb_{1,T}b_{2,T})^{-\tau/2} \left[R'_{0,\tau+2}(it_2)^{\tau+2} + R'_{2,\tau}(it_1)^2(it_2)^{\tau} \right], \qquad \tau \text{ even},$$
$$R_T^*(\tau) = (Tb_{1,T}b_{2,T})^{-\tau/2} \frac{1}{(\tau+2)!} \left[\overline{\kappa}_{2,T}(0,\,\tau+2)(it_2)^{\tau+2} + \frac{(\tau+2)(\tau+1)}{2} \overline{\kappa}_{2,T}(2,\,\tau)(it_1)^2(it_2)^{\tau} \right]$$
$$+ (Tb_{1,T}b_{2,T})^{-\tau/2} \left[R'_{0,\tau+3}(it_2)^{\tau+3} + R'_{2,\tau+1}(it_1)^2(it_2)^{\tau+1} \right], \qquad \tau \text{ odd},$$

where the $R'_{0,j}$ and $R_{2,j}$ are bounded. Using Lemma S.B.18-S.B.19, we have

$$\log \psi_T(\mathbf{t}) = \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} \frac{(Tb_{1,T}b_{2,T})^{(2-s)/2}}{s!} \left(\overline{\kappa}_{2,T}(0,\,s)(it_2)^s + \frac{s(s-1)}{2} \overline{\kappa}_{2,T}(2,\,s-2)(it_1)^2(it_2)^{s-2} \right)$$
$$+ R_T^*(\tau)$$
$$= \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1} (Tb_{1,T}b_{2,T})^{(2-s)/2} \left[B_{2,T}(s,\,\mathbf{t}) + \left\{ (it_2)^s + (it_1)^2(it_2)^{s-2} \right\} O(\epsilon_T(s)) \right] + R_T^*(\tau),$$

where

$$B_{2,T}(s,\,\mathbf{t}) = \frac{1}{s!}\sum_{j=0}^{d_f} b_{1,T}^{j}\Big\{\left(\Xi_{2,j}(0,\,s) + b_{2,T}^{2}\left(\widetilde{\Xi}_{2,j}(0,\,s) + \widetilde{\Xi}_{3,j}(0,\,s)\right)\right)(it_2)^{s}$$
$$+ \frac{s\,(s-1)}{2}\left(\Xi_{2,j}(2,\,s-2) + b_{2,T}^{2}\left(\widetilde{\Xi}_{2,j}(2,\,s-2) + \widetilde{\Xi}_{3,j}(2,\,s-2)\right)\right)(it_1)^{2}\,(it_2)^{s-2}\Big\}.$$

The approximation of the characteristic function of $\mathbf{v}$ using its cumulant generating function is

$$\mathcal{A}_{2,T}(\mathbf{t},\,\tau) = \exp\left(\frac{1}{2}\|i\mathbf{t}\|^{2}\right)\left[1 + \sum_{j=3}^{\tau+1}(Tb_{1,T}b_{2,T})^{(2-j)/2}\sum_{\mathbf{r}}\prod_{n=3}^{\tau+1}(B_{2,T}(n,\,\mathbf{t}))^{r_n}\frac{1}{r_3!\ldots r_{\tau+1}!}\right],$$

where $\mathbf{r} = (r_3,\ldots,\,r_{\tau+1})'$, $r_n \in \{0,\,1,\ldots\}$, and the summation is over all $\mathbf{r}$ satisfying $\sum_{n=3}^{\tau+1}(n-2)\,r_n = j-2$. To obtain a second-order Edgeworth expansion we set $\tau = 2$ and we include in $\mathcal{A}_{2,T}(\mathbf{t},\,2)$ the terms up to order $(Tb_{1,T}b_{2,T})^{-1/2}$,

$$\mathcal{A}_{2,T}(\mathbf{t},\,2) = \exp\left(\frac{1}{2}\|i\mathbf{t}\|^{2}\right)\left[1 + \overline{B}_{2,T}(3,\,\mathbf{t})\,(Tb_{1,T}b_{2,T})^{-1/2}\right], \tag{S.B.65}$$

where $\overline{B}_{2,T}(3,\,\mathbf{t})$ includes only the leading term in $b_{1,T}^{j}$ $(j=0)$ in the expansion for the culumant of order three. Note that the characteristic function of $\mathbb{Q}_{2,T}^{(2)}(\cdot)$ is $\mathcal{A}_{2,T}(\mathbf{t},\,2)$. We use Lemma S.B.2 with kernel $\mathbb{G}$ to bound the distance between $\mathbb{P}_T$ and $\mathbb{Q}_{2,T}^{(2)}$. First,

$$\left\|\left(\mathbb{P}_T - \mathbb{Q}_{2,T}^{(2)}\right)\bullet\mathbb{G}_{\phi_T}\right\|_{\mathrm{TV}} \le 2\sup_{\mathbf{B}\subset\mathbf{B}(0,\,r_T)}\left|\left(\mathbb{P}_T - \mathbb{Q}_{2,T}^{(2)}\right)\bullet\mathbb{G}_{\phi_T}\right| + 2\sup_{\mathbf{B}\subset\mathbf{B}(0,\,r_T)^c}\left|\left(\mathbb{P}_T - \mathbb{Q}_{2,T}^{(2)}\right)\bullet\mathbb{G}_{\phi_T}\right|,$$

where $\mathbf{B}(0,\,r_T)$ is a neighborhood around 0 with radius r_T, $r_T = (Tb_{1,T}b_{2,T})^{a}$ with $a > 0$, and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. For $\mathbf{B} \subset \mathbf{B}(0,\,r_T)^c$ we have uniformly

$$\begin{aligned}\left|\left(\mathbb{P}_T - \mathbb{Q}_{2,T}^{(2)}\right)\bullet\mathbb{G}_{\phi_T}\right| &\le \left|\mathbb{P}_T\bullet\mathbb{G}_{\phi_T}\right| + \left|\mathbb{Q}_{2,T}^{(2)}\bullet\mathbb{G}_{\phi_T}\right| \\ &\le \mathbb{P}\left(\|\mathbf{v}\| \ge r_T/2\right) + 2\mathbb{G}_{\phi_T}\left(\mathbf{B}(0,\,r_T/2)^c\right) + 2\mathbb{Q}_{2,T}^{(2)}\left(\mathbf{B}(0,\,r_T/2)^c\right).\end{aligned}$$

By definition of $q_{2,T}^{(2)}(\mathbf{v})$ it follows that $\mathbb{Q}_{2,T}^{(2)}(\mathbf{B}(0,\,r_T/2)^c) = o((Tb_{1,T}b_{2,T})^{-1/2})$. In view of the definition of v_2, we have $\mathbb{P}\{\|\mathbf{v}\| \ge r_T/2\} = o((Tb_{1,T}b_{2,T})^{-1/2})$. By Lemma S.B.2,

$$\mathbb{G}_{\phi_T}\left(\mathbf{B}(0,\,r_T/2)^c\right) = O\left((\phi_T/r_T)^{3}\right) = O\left((Tb_{1,T}b_{2,T})^{-3(\varpi+a)}\right) = o\left((Tb_{1,T}b_{2,T})^{-1/2}\right).$$

For $\mathbf{B} \subset \mathbf{B}(0,\,r_T)$ we have by Fourier inversion

$$\left|\left(\mathbb{P}_T - \mathbb{Q}_{2,T}^{(2)}\right)\bullet\mathbb{G}_{\phi_T}\right| \le (2\pi)^{-1}\,\pi r_T^{2}\int\left|\left(\widehat{\mathbb{P}}_T - \widehat{\mathbb{Q}}_{2,T}^{(2)}\right)(\mathbf{t})\,\widehat{\mathbb{G}}_{\phi_T}(\mathbf{t})\right|d\mathbf{t}, \tag{S.B.66}$$

where $\widehat{\mathbb{P}}_T$ denotes the characteristic function of $\mathbb{P}_T$ (i.e., $\widehat{\mathbb{P}}_T = \psi_T(\mathbf{t})$) and $\widehat{\mathbb{Q}}^{(2)}_{2,T} = \mathcal{A}_{2,T}(\mathbf{t}, 2)$. Let $a' = 8 \times 2^{4/3}\pi^{-1/3}$. Using Lemma S.B.20, a bound for (S.B.66) is given by

$$O\left((Tb_{1,T}b_{2,T})^{2a-1/2}\right)\left[b_{1,T}^2 + \epsilon_{Tb_{2,T}}(3)\right]\int_{\|\mathbf{t}\|\leq c_2\sqrt{Tb_{1,T}b_{2,T}}}\left|e^{-d_2\|\mathbf{t}\|^2}F(\|\mathbf{t}\|)\right|\left|\widehat{\mathbb{G}}_{\phi_T}(\|\mathbf{t}\|)\right|d\mathbf{t} \tag{S.B.67}$$

$$+ O(Tb_{1,T}b_{2,T})^{2a}\int_{c_2\sqrt{Tb_{1,T}b_{2,T}}<\|\mathbf{t}\|\leq a'(Tb_{1,T}b_{2,T})^{\varpi}}\int\left|\left(\widehat{\mathbb{P}}_T - \widehat{\mathbb{Q}}^{(2)}_{2,T}\right)(\mathbf{t})\,\widehat{\mathbb{G}}_{\phi_T}(\mathbf{t})\right|d\mathbf{t}. \tag{S.B.68}$$

The integral over $\|\mathbf{t}\| > a'(Tb_{1,T}b_{2,T})^{\varpi}$ is equal to zero from (S.B.21). Choosing $a \leq 1/4$ (S.B.67) is $o(((Tb_{1,T}b_{2,T}))^{-1/2})$.

By Lemma S.B.10, for $c_2 m_{2,T} < \|\mathbf{t}\|$ the expression in (S.B.68) is bounded by

$$O\left((Tb_{1,T}b_{2,T})^{2a}\right)\int_{c_2\sqrt{Tb_{1,T}b_{2,T}}<\|\mathbf{t}\|\leq a'(Tb_{1,T}b_{2,T})^{\varpi}} e^{-d_3 m_{2,T}^2}d\mathbf{t} + o\left((Tb_{1,T}b_{2,T})^{-1/2}\right),$$

for some $d_3 > 0$. This implies that (S.B.68) is bounded by $O(((Tb_{1,T}b_{2,T})^{2(\varpi+a)}e^{-d_3 m_{2,T}^2}) + o((Tb_{1,T}b_{2,T})^{-1/2})$ since by Assumption 4.7-4.8 it holds $m_{2,T} \geq \epsilon(Tb_{2,T})^{\epsilon}$ for some $\epsilon > 0$ depending on q and p. □

Lemma S.B.20. *Let Assumption 4.1, 4.3-4.4, 4.8-4.9 and* $b_{1,T} + (Tb_{1,T}b_{2,T})^{-1}\log^5(Tb_{2,T}) \to 0$ *hold. Then there exists a* $c_2 > 0$ *such that, for* $\|\mathbf{t}\| \leq c_2\sqrt{Tb_{1,T}b_{2,T}}$ *and a* $d_2 > 0$,

$$|\psi_T(\mathbf{t}) - \mathcal{A}_{2,T}(\mathbf{t}, 2)| \leq \exp\left(-d_2\|\mathbf{t}\|^2\right)\widetilde{F}(\|\mathbf{t}\|)\,O\left((Tb_{1,T}b_{2,T})^{-1/2}\left(b_{1,T}^2 + \epsilon_{Tb_{2,T}}(3)\right) + \frac{1}{Tb_{1,T}b_{2,T}}\right),$$

where $\widetilde{F}(\|\mathbf{t}\|)$ *is a polynomial in* $\mathbf{t}$ *with bounded coefficients and* $\mathcal{A}_{2,T}(\mathbf{t}, 2)$ *is defined in* (S.B.65).

Proof of Lemma S.B.20. From Feller (1971, p. 535) for complex α and β it holds that $|e^a - 1 - b| \leq e^{\gamma}(|a-b| + |b|^2/2)$, where $\gamma = \max\{|a|, |b|\}$. We set

$$a = \log\psi(\mathbf{t}) - \frac{1}{2}\|i\mathbf{t}\|^2 = (Tb_{1,T}b_{1,T})^{-1/2}\sum_{|\mathbf{r}|=3}\frac{s!}{r_1!r_2!}\overline{\kappa}_{2,T}(r_1, r_2)(it_1)^{r_1}(it_2)^{r_2} + R_T^*(2),$$

where the right-hand side follows from (S.B.64). Let $b = (Tb_{1,T}b_{1,T})^{-1/2}\overline{B}_{2,T}(3, \mathbf{t})$ where $\overline{B}_{2,T}(3, \mathbf{t})$ is defined after (S.B.65). Using Lemma S.B.18-S.B.19 for $s = 3$ we have

$$\begin{aligned}|a - b| &\leq \Bigg|(Tb_{1,T}b_{1,T})^{-1/2}O\left(b_{1,T}^2 + \epsilon_{Tb_{2,T}}(3)\right)\left((it_2)^3 + (it_1)^2(it_2)\right) \\ &\quad + \frac{1}{Tb_{1,T}b_{2,T}}\left(R'_{0,4}(it_2)^4 + R'_{2,2}(it_1)^2(it_1)^2\right)\Bigg| \\ &\leq P_1(\|\mathbf{t}\|)\,O\left((Tb_{1,T}b_{1,T})^{-1/2}\left(b_{1,T}^2 + \epsilon_{Tb_{2,T}}(3)\right) + \frac{1}{Tb_{1,T}b_{2,T}}\right),\end{aligned} \tag{S.B.69}$$

where P_1 is a polynomial of degree of 4. Note that $|b|^2/2 \leq P_2(\|\mathbf{t}\|)\,O(Tb_{1,T}b_{1,T})^{-1})$ where P_2 is a polynomial of degree 6. Then, for some polynomial P

$$|a-b|+\frac{|b|^2}{2}\leq P\left(\|\mathbf{t}\|\right)O\left(\left(Tb_{1,T}b_{1,T}\right)^{-1/2}\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)+\frac{1}{Tb_{1,T}b_{2,T}}\right).$$

Next, we need to find a bound for $\gamma=\max\left\{|a|,\,|b|\right\}$. For $\|\mathbf{t}\|\leq c_b\sqrt{Tb_{1,T}b_{2,T}}$ with $c_b>0$ we have

$$\begin{aligned}|b|&=\left|\left(Tb_{1,T}b_{1,T}\right)^{-1/2}\overline{B}_{2,T}\left(3,\,\mathbf{t}\right)\right|\leq\|\mathbf{t}\|^2\left\{\frac{1}{3!}\left(Tb_{1,T}b_{1,T}\right)^{-1/2}\left[\left|\Xi_{2,0}(0,\,3)\right|+3\left|\Xi_{2,0}(2,\,1)\right|\|\mathbf{t}\|\right]\right\}\\&\leq\|\mathbf{t}\|^2\left\{\frac{c_b}{3!}\left(\left|\Xi_{2,0}(0,\,3)\right|+3\left|\Xi_{2,0}(2,\,1)\right|\right)\right\}\leq\|\mathbf{t}\|^2T_b,\end{aligned}\tag{S.B.70}$$

where $0<T_b<1/4$ by choosing c_b sufficiently small. For a given a we can choose a $c_a>0$ sufficiently small such that, for $\|\mathbf{t}\|\leq c_a\sqrt{Tb_{1,T}b_{1,T}}$,

$$\begin{aligned}|a|&\leq\|\mathbf{t}\|^2\left\{\frac{1}{3!}\left(Tb_{1,T}b_{1,T}\right)^{-1/2}\left[\left|\Xi_{2,0}(0,\,3)\right|+3\left|\Xi_{2,1}(2,\,1)\right|+O\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)\right]\right.\\&\qquad\left.\times\|\mathbf{t}\|+\left(Tb_{1,T}b_{1,T}\right)^{-1}\left[\left|R'_{0,4}\right|+\left|R'_{2,2}\right|\right]\|\mathbf{t}\|^2\right\}\\&\leq\|\mathbf{t}\|^2\left\{\frac{c_a}{3!}\left[\left|\Xi_{2,0}(0,\,3)\right|+3\left|\Xi_{2,0}(2,\,1)\right|+O\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)\right]+c_a^2\left[\left|R'_{0,4}\right|+\left|R'_{2,2}\right|\right]\right\}\\&\leq\|\mathbf{t}\|^2\left\{\frac{1}{4}+O\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)\right\}.\end{aligned}\tag{S.B.71}$$

From (S.B.70)-(S.B.71) we have for $\|\mathbf{t}\|\leq c_2\sqrt{Tb_{1,T}b_{1,T}}$ with $c_2=\min\left\{c_a,\,c_b\right\}$,

$$\exp\left(\gamma\right)\leq\exp\left\{\|\mathbf{t}\|^2\left[\frac{1}{4}+O\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)\right]\right\},$$

or

$$\exp\left\{-\frac{1}{2}\mathbf{t}^2+\gamma\right\}\leq\exp\left\{\|\mathbf{t}\|^2\left[-\frac{1}{4}+O\left(b_{1,T}^2+\epsilon_{Tb_{2,T}}\left(3\right)\right)\right]\right\}\leq\exp\left\{-d_2\|\mathbf{t}\|^2\right\},\tag{S.B.72}$$

for some $d_2>0$. Note that $\psi\left(\mathbf{t}\right)=\exp\{\frac{1}{2}\|i\mathbf{t}\|^2+a\}$ and $\mathcal{A}_{2,T}\left(\mathbf{t},\,2\right)=\exp\{\frac{1}{2}\|i\mathbf{t}\|^2\}(1+b)$. Using (S.B.69)-(S.B.72) the result of the lemma follows. $\square$

S.B.3.5 Proof of Theorem 4.4

Consider the following linear stochastic approximation to U_T,

$$\widetilde{U}_T\triangleq v_1\left(1-\frac{1}{2}\overline{c}_1b_{1,T}^{d_f}-\frac{1}{2}\sqrt{4\pi}\left\|K\right\|_2\left\|K_2\right\|_2v_2\left(Tb_{1,T}b_{2,T}\right)^{-1/2}-\frac{1}{2}\overline{c}_2b_{2,T}^2\right).\tag{S.B.73}$$

Consider the transformation $\mathbf{s} = (s_1,\, s_2)' = (\widetilde{U}_T (h_1,\, v_2)\, ,\, v_2)' = \Delta_T (\mathbf{v})$ say, and its inverse $\mathbf{v} = \Delta_T^{-1} (\mathbf{s}) = (h_1^{\dagger} (s_1,\, s_2)\, ,\, s_2)'$. Let $\gamma > 0$ be such that

$$\frac{T^{3\gamma}}{(Tb_{1,T} b_{2,T})^{3/2}} \to 0,$$

and define $\mathbf{L}_T = \{\mathbf{v} : |v_i| < l_i T^{\gamma},\, i = 1,\, 2\}$, where l_i are some fixed constants. Using $(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots$ for $|x| < 1$, we have uniformly in the set $\mathbf{L}_T$,

$$h_1^{\dagger} (\mathbf{s}) = s_1 \left[1 + \frac{1}{2} \overline{c}_1 b_{1,T}^{d_f} + \frac{1}{2} \sqrt{4\pi}\, \|K_2\|\, \|K_2\|_2\, s_2\, (Tb_{1,T} b_{2,T})^{-1/2} + \frac{1}{2} \overline{c}_2 b_{2,T}^2\right] + o\left((Tb_{1,T} b_{2,T})^{-1/2}\right).$$

We have $\mathbb{P}(U_T \in \mathbf{C}) = \mathbb{P}(\mathbf{v} \in \Delta_T^{-1} (\mathbf{C} \times \mathbb{R}))$ and from Theorem 4.1,

$$\begin{aligned} \sup_{\mathbf{C}} \left|\mathbb{P}\left(\mathbf{v} \in \Delta_T^{-1} (\mathbf{C} \times \mathbb{R})\right) - \mathbb{Q}_{2,T}^{(2)} \left(\Delta_T^{-1} (\mathbf{C} \times \mathbb{R})\right)\right| \\ = o\left((Tb_{1,T} b_{2,T})^{-1/2}\right) + \text{cost} \sup_{\mathbf{C}} \mathbb{Q}_{2,T}^{(2)} \left(\left(\partial \Delta_T^{-1} (\mathbf{C} \times \mathbb{R})\right)^{2\phi_T}\right), \end{aligned} \tag{S.B.74}$$

where $\phi_T = (Tb_{1,T} b_{2,T})^{-\rho}$, $1/2 < \rho < 1$. From the continuity of Δ_T, we can obtain, for some $c > 0$,

$$Q_{2,T}^{(2)} \left(\left(\partial \Delta_T^{-1} (\mathbf{C} \times \mathbb{R})\right)^{2\phi_T}\right) \leq Q_{2,T}^{(2)} \left(\Delta_T^{-1} (\partial \mathbf{C})^{c\phi_T} \times \mathbb{R}\right), \tag{S.B.75}$$

and

$$\begin{aligned} Q_{2,T}^{(2)} \left(\Delta_T^{-1} (\mathbf{C} \times \mathbb{R})\right) &= \int_{\mathbf{L}_T \cap \Delta_T^{-1} (\mathbf{C} \times \mathbb{R})} \varphi_2 (\mathbf{x})\, q_{2,T}^{(2)} (\mathbf{x})\, d\mathbf{x} + o\left((Tb_{1,T} b_{2,T})^{-1/2}\right) \\ &= \int_{\mathbf{L}_T^* \cap \{\mathbf{C} \times \mathbb{R}\}} \varphi_2 \left(\Delta_T^{-1} (\mathbf{s})\right) q_{2,T}^{(2)} \left(\Delta_T^{-1} (\mathbf{s})\right) |\mathcal{J}|\, d\mathbf{s} + o\left((Tb_{1,T} b_{2,T})^{-1/2}\right), \end{aligned}$$

where $\varphi_2 (\cdot)$ is the bivariate standard normal density, $\mathbf{L}_T^* = \Delta_T (\mathbf{L}_T)$, and $|\mathcal{J}|$ is the Jacobian of the transformation. Neglecting the terms that contribute $o((Tb_{1,T} b_{2,T})^{-1/2})$ to the integrals, we yield

$$\varphi_2 \left(\Delta_T^{-1} (\mathbf{s})\right) = \varphi (s_1)\, \varphi (s_2) \left(1 - \frac{1}{2} s_1^2 \left[\overline{c}_1 b_{1,T}^{d_f} + \frac{1}{2} \sqrt{4\pi}\, \|K\|_2\, \|K_2\|_2\, s_2\, (Tb_{1,T} b_{2,T})^{-1/2} + \frac{1}{2} \overline{c}_2 b_{2,T}^2\right]\right), \tag{S.B.76}$$

and

$$q_{2,T}^{(2)} (\mathbf{v}) = 1 + \frac{1}{3!} (Tb_{1,T} b_{2,T})^{-1/2} \left(\Xi_{2,0} (0,\, 3)\, \mathcal{H}_3 (v_2) + \Xi_{2,0} (2,\, 1)\, \mathcal{H}_2 (h_1)\, \mathcal{H}_1 (v_2)\right), \tag{S.B.77}$$

where

$$|\mathcal{J}| = 1 + \frac{1}{2} \overline{c}_1 b_{1,T}^{d_f} + \frac{1}{2} \sqrt{4\pi}\, \|K_2\|\, \|K_2\|_2\, s_2\, (Tb_{1,T} b_{2,T})^{-1/2} + \frac{1}{2} \overline{c}_2 b_{2,T}^2.$$

For $j = 1, 2, 3$ let $p_j(\mathbf{s})$ denote polynomials not depending on T. We have

$$\begin{aligned} Q_{2,T}^{(2)}\left(\Delta_T^{-1}(\mathbf{C}\times\mathbb{R})\right) &= \int_{\mathbf{C}}\varphi(s_1)\left\{\int_{\mathbb{R}}\left[1+p_1(\mathbf{s})(Tb_{1,T}b_{2,T})^{-1/2}+p_2(\mathbf{s})b_{1,T}^{d_f}+p_3(\mathbf{s})b_{2,T}^2\right]\varphi(s_2)\,ds_2\right\}ds_1 \\ &\quad + o\left((Tb_{1,T}b_{2,T})^{-1/2}\right) \\ &= \int_{\mathbf{C}}\varphi(s_1)\left[1+r_1(s_1)(Tb_{1,T}b_{2,T})^{-1/2}+r_2(s_1)b_{1,T}^{d_f}+r_3(s_1)b_{2,T}^2\right]ds_1 \\ &\quad + o\left((Tb_{1,T}b_{2,T})^{-1/2}\right), \end{aligned} \tag{S.B.78}$$

where $r_j(s_1)$ are polynomials in s_1 for $j = 1, 2, 3$ with bounded coefficients. Integration with respect to s_2 in $\mathbb{R}$ yields $r_1(x) = 0$, $r_2(x) = -2^{-1}\bar{c}_1(x^2-1)$ and $r_3(x) = -2^{-1}\bar{c}_2(x^2-1)$. Using (S.B.74)-(S.B.78) provides the second-order Edgeworth expansion for the linear stochastic approximation $\widetilde{U}_T$. Since Lemma S.B.21 below shows that $\widetilde{U}_T$ and U_T have the same Edgeworth expansion, the proof is concluded. □

Lemma S.B.21. *Let Assumption 4.1, 4.2 ($p > 1$) and 4.3-4.5, 4.8-4.10 hold. Then, U_T has the same Edgeworth expansion as $\widetilde{U}_T$ uniformly for convex Borel sets up to the order $O((Tb_{1,T}b_{2,T})^{-1/2})$.*

Proof of Lemma S.B.21. We first expand $U_T(\mathbf{v})$ around $\mathbf{0}$ in $\mathbf{L}_T$ with $|\eta_2| \le 1$,

$$U_T = d_T h_1 - \frac{1}{2}d_T^3\mathsf{V}_{2,T}h_1 v_2(Tb_{1,T}b_{2,T})^{-1/2} + U_{1,T}^*(Tb_{1,T}b_{2,T})^{-1}, \tag{S.B.79}$$

where $d_T = (1+\mathsf{B}_{2,T})^{-1/2}$ and

$$U_{1,T}^* = \frac{3}{8}\left(1+\mathsf{B}_{2,T}+\eta_2\mathsf{V}_{2,T}v_2(Tb_{1,T}b_{2,T})^{-1/2}\right)^{-5/2}\mathsf{V}_{2,T}^2h_1v_2^2.$$

We now express U_T in terms of $\widetilde{U}_T$ where the latter is defined in (S.B.73). Substituting for $\mathsf{B}_{2,T}$ and $\mathsf{V}_{2,T}$ in (S.B.79), we yield $U_T = \widetilde{U}_T + U_T^*(Tb_{1,T}b_{2,T})^{-1}$ where $U_T^* = \sum_{i=1}^3 U_{i,T}^*$,

$$U_{2,T}^* = h_1\left(O\left((b_{1,T}b_{2,T})^{-1}\log T + Tb_{2,T}b_{1,T}^{1+d_f+\varrho}\right) + o\left(Tb_{2,T}^3b_{1,T}\right)\right)$$

and

$$U_{3,T}^* = h_1v_2O\left((Tb_{1,T}b_{2,T})^{1/2}\left(b_{1,T}^2+\epsilon_T(2)\right)\right).$$

We now show that $U_T^*(Tb_{1,T}b_{2,T})^{-1}$ can be neglected with error $o((Tb_{1,T}b_{2,T})^{1/2})$. This follows from Theorem 2 in Chibisov (1972) provided that the following condition holds,

$$\mathbb{P}\left(|U_T^*| > \gamma_T\sqrt{Tb_{1,T}b_{2,T}}\right) \le \sum_{i=1}^3\mathbb{P}\left(\left|U_{i,T}^*\right| > \frac{1}{3}\gamma_T\sqrt{Tb_{1,T}b_{2,T}}\right) = o\left((Tb_{1,T}b_{2,T})^{-1/2}\right), \tag{S.B.80}$$

for some positive sequence $\{\gamma_T\}$ such that $\gamma_T \to 0$ and $\gamma_T\sqrt{Tb_{1,T}b_{2,T}} \to \infty$. Note that

$$(Tb_{1,T}b_{2,T})^{-1/2}U_{2,T}^* = h_1O\left((Tb_{2,T})^{1/2}b_{1,T}^{-3/2}(Tb_{2,T})^{-1}\log T + (Tb_{2,T}b_{1,T})^{1/2}b_{1,T}^{d_f+\varrho}\right).$$

By Assumption 4.10 the right-hand side above is $O((Tb_{2,T}b_{1,T})^{-\upsilon})$ for some $\upsilon > 0$. Further,

$$\left(Tb_{1,T}b_{2,T}\right)^{-1/2} U_{3,T}^{*} = h_1 v_2 O\left(b_{1,T}^{2} + \epsilon_T\left(2\right)\right) = O((Tb_{2,T}b_{1,T})^{-\upsilon}),$$

for some $\upsilon > 0$. Since h_1 and v_2 have finite moments of all orders, we can take $\gamma_T = 1/\log T$ and apply Chebyshev's inequality to establish $\mathbb{P}(|U_{i,T}^{*}| > 3^{-1}\gamma_T\sqrt{Tb_{1,T}b_{2,T}}) = o((Tb_{1,T}b_{2,T})^{-1/2})$ for $i = 2,\, 3$.

It remains to show $\mathbb{P}(|U_{1,T}^{*}| > 3^{-1}\gamma_T\sqrt{Tb_{1,T}b_{2,T}}) = o((Tb_{1,T}b_{2,T})^{-1/2})$. We have

$$\begin{aligned}
&\mathbb{P}\left(\left|U_{1,T}^{*}\right| > \frac{1}{3}\gamma_T\sqrt{Tb_{1,T}b_{2,T}}\right)\\
&< \mathbb{P}\left(\left|\frac{3}{8}\mathsf{V}_{2,T}^{2}h_1v_2^{2}\right|\left(Tb_{1,T}b_{2,T}\right)^{-1/4} > \gamma_T^{1/2}\right)\\
&\quad + \mathbb{P}\left(\left|1 + \mathsf{B}_{2,T} + \eta_2\mathsf{V}_{2,T}v_2\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right|\left(Tb_{1,T}b_{2,T}\right)^{-1/4} > \gamma_T^{1/2}\right).\\
&\triangleq A_1 + A_2.
\end{aligned}$$

Using Chebyshev's inequality $A_1 = o((Tb_{1,T}b_{2,T})^{-1/2})$. Using $\left(Tb_{1,T}b_{2,T}\right)^{-1/10}\gamma_T^{-1/5} \to 0$ we yield

$$A_2 < C_2\mathbb{P}\left(\left|v_2\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right| > c_2\right) = o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right),$$

where C_2 and c_2 are some positive constants and we have used Chebyshev's inequality. □

S.B.4 Proof of the Results of Section 5

S.B.4.1 Proof of Theorem 5.1

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

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

for some $J_{\mathrm{DM}} \in (0,\, \infty)$ where n_δ depends on the length of the segment where the mean of $x_t^{(2)}$ shifts by δ. The factor δ^2 follows from the quadratic loss.

Next, we focus on the expansion of the denominator of $t_{\mathrm{DM},i}$ which hinges on which LRV estimator is used. We begin with part (i). Under Assumption 4.6 $b_{1,T} \to 0$ as $T \to \infty$. Using Theorem S.A.1,

$$\begin{aligned}
\widehat{J}_{d_L,\mathrm{NW87},T} &= \sum_{k=-\left\lfloor b_T^{-1}\right\rfloor}^{\left\lfloor b_T^{-1}\right\rfloor}\left(1 - \left|b_{1,T}k\right|\right)\widehat{\Gamma}\left(k\right)\\
&= \sum_{k=-\left\lfloor b_{1,T}^{-1}\right\rfloor}^{\left\lfloor b_{1,T}^{-1}\right\rfloor}\left(1 - \left|b_{1,T}k\right|\right)\int_0^1 c\left(u,\, k\right)du
\end{aligned}$$

$$+\sum_{k=-\lfloor b_{1,T}^{-1}\rfloor}^{\lfloor b_{1,T}^{-1}\rfloor}(1-|b_{1,T}k|)\left(2^{-1}\left(\frac{T_b-T_m-1}{T_n}\right)\left(\frac{T_n-T_b-2}{T_n}\right)\delta^4+o_{\mathbb{P}}(1)\right)$$
$$=CJ_{\mathrm{DM}}+\sum_{k=-\lfloor b_{1,T}^{-1}\rfloor}^{\lfloor b_{1,T}^{-1}\rfloor}(1-|b_{1,T}k|)\left(2^{-1}\left(\frac{T_b-T_m-1}{T_n}\right)\left(\frac{T_n-T_b-2}{T_n}\right)\delta^4+o_{\mathbb{P}}(1)\right),$$

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

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

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

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

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

$$|t_{\mathrm{DM,NW87}}|\leq\frac{\delta^2O_{\mathbb{P}}\left(T_n^{-1/2}n_\delta\right)+O_{\mathbb{P}}(1)}{\left(\delta^4O\left(b_{1,T}^{-1}\right)\right)^{1/2}}\tag{S.B.81}$$
$$=\frac{\delta^2O\left(T_n^{\zeta}\right)}{\delta^2O\left(b_{1,T}^{-1/2}\right)}=O\left(T_n^{\zeta}b_{1,T}^{1/2}\right),$$

which implies $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}|>z_\alpha)\to0$.

Under Assumption 4.7 with $q=1/3$, similar derivations yield $|t_{\mathrm{DM,NW87}}|=O(T_n^{\zeta-1/6})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}|>z_\alpha)\to0$.

In part (ii), $b_{1,T}=T^{-1}$. Proceeding as in (S.B.81) we have $|t_{\mathrm{DM,KVB}}|=O(T_n^{\zeta-1})$ and $\mathbb{P}_\delta(|t_{\mathrm{DM,KVB}}|>z_\alpha)\to0$ since $T_n^{\zeta-1}\to0$.

Finally, we consider part (iii). Using Theorem 3.1, we have

$$\widehat{J}_{d_L,\mathrm{DK},T}=\sum_{k=-T_n+1}^{T_n-1}K_1\left(\widehat{b}_{1,T}k\right)\frac{n_T}{T_n}\sum_{r=1}^{\lfloor T_n/n_T\rfloor}\widehat{c}_{\mathrm{DK},T}\left(rn_T/T,\,k\right)$$
$$=\sum_{k=-T_n+1}^{T_n-1}K_1\left(\widehat{b}_{1,T}k\right)\frac{n_T}{T_n}\sum_{r=1}^{\lfloor T_n/n_T\rfloor}\Big(c\left(rn_T/T,\,k\right)$$
$$+\delta^2\mathbf{1}\left\{\left(|rn_T+k/2+n_{2,T}/2+1)-T_j^0|/n_{2,T}\right)\in(0,\,1)\right\}\Big)+o_{\mathbb{P}}(1)$$

$$= J_{\mathrm{DM}} + \delta^2 O_{\mathbb{P}} \left(\widehat{b}_{1,T}^{-1} \frac{T \widehat{\overline{b}}_{2,T}}{n_T} \frac{n_T}{T_n} \right) + o_{\mathbb{P}}(1) .$$

It follows that

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

and so $\mathbb{P}_\delta(|t_{\mathrm{DM,DK}}| > z_\alpha) \to 1$ since $T_n^{\zeta} \to \infty$. $\square$

S.C Figures

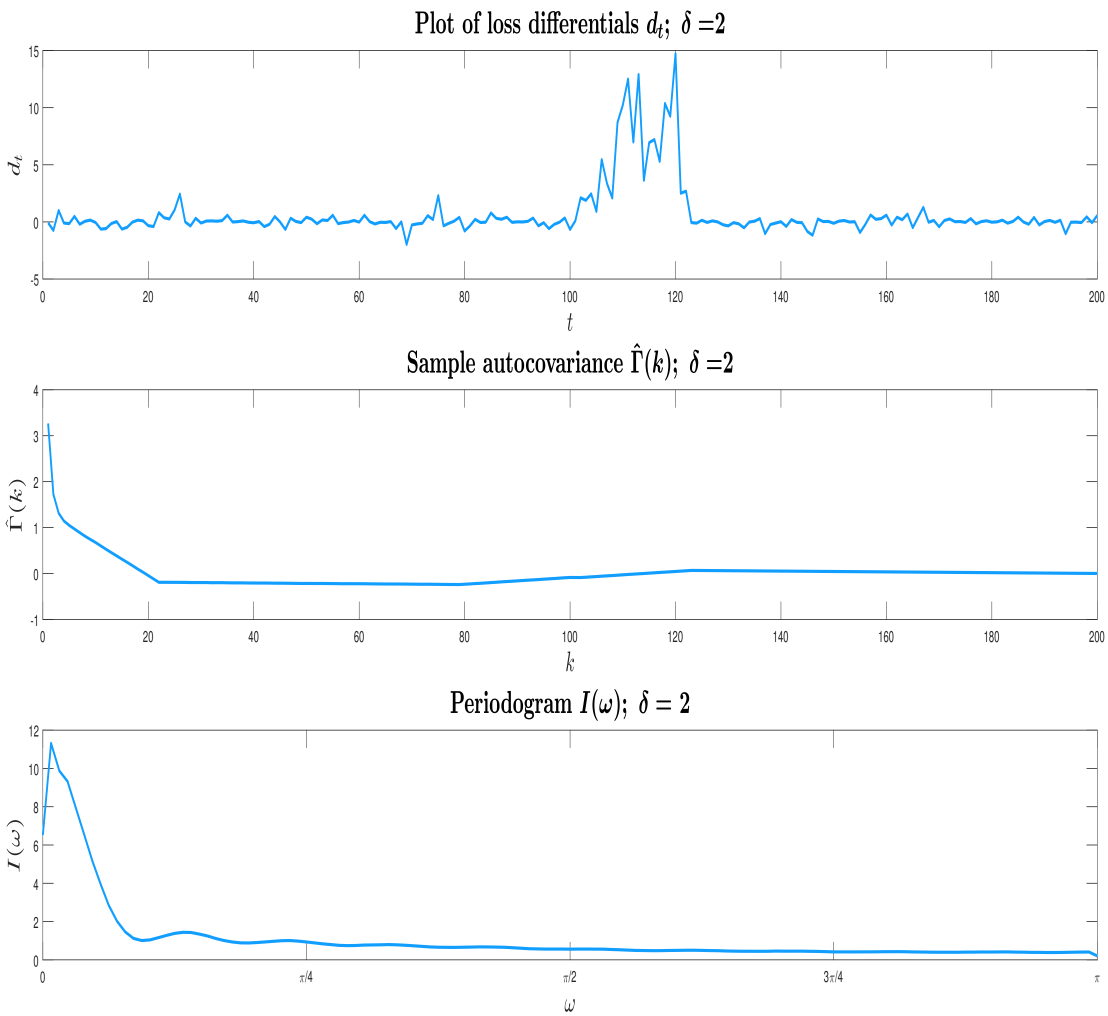

Figure S.1: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 2$.

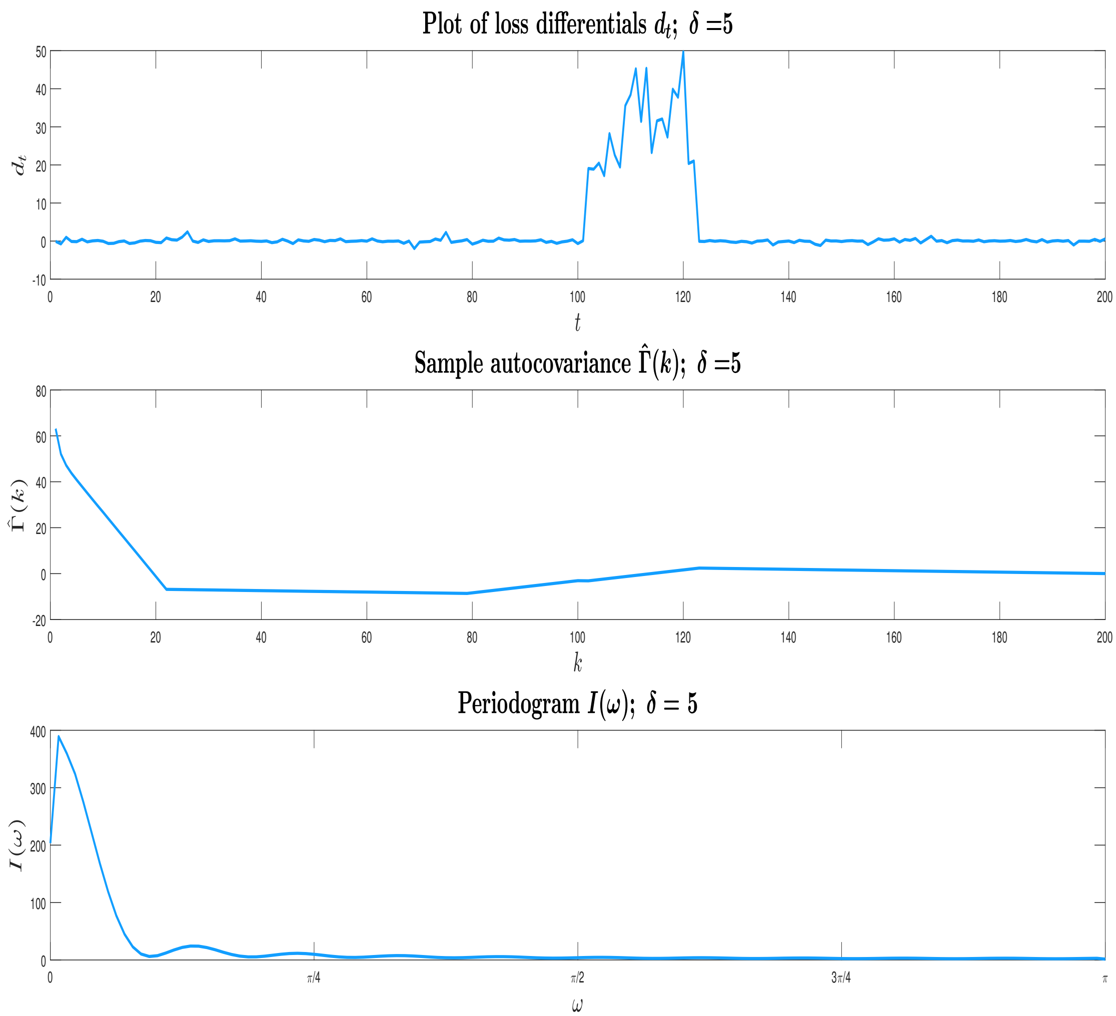

Figure S.2: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 5$.

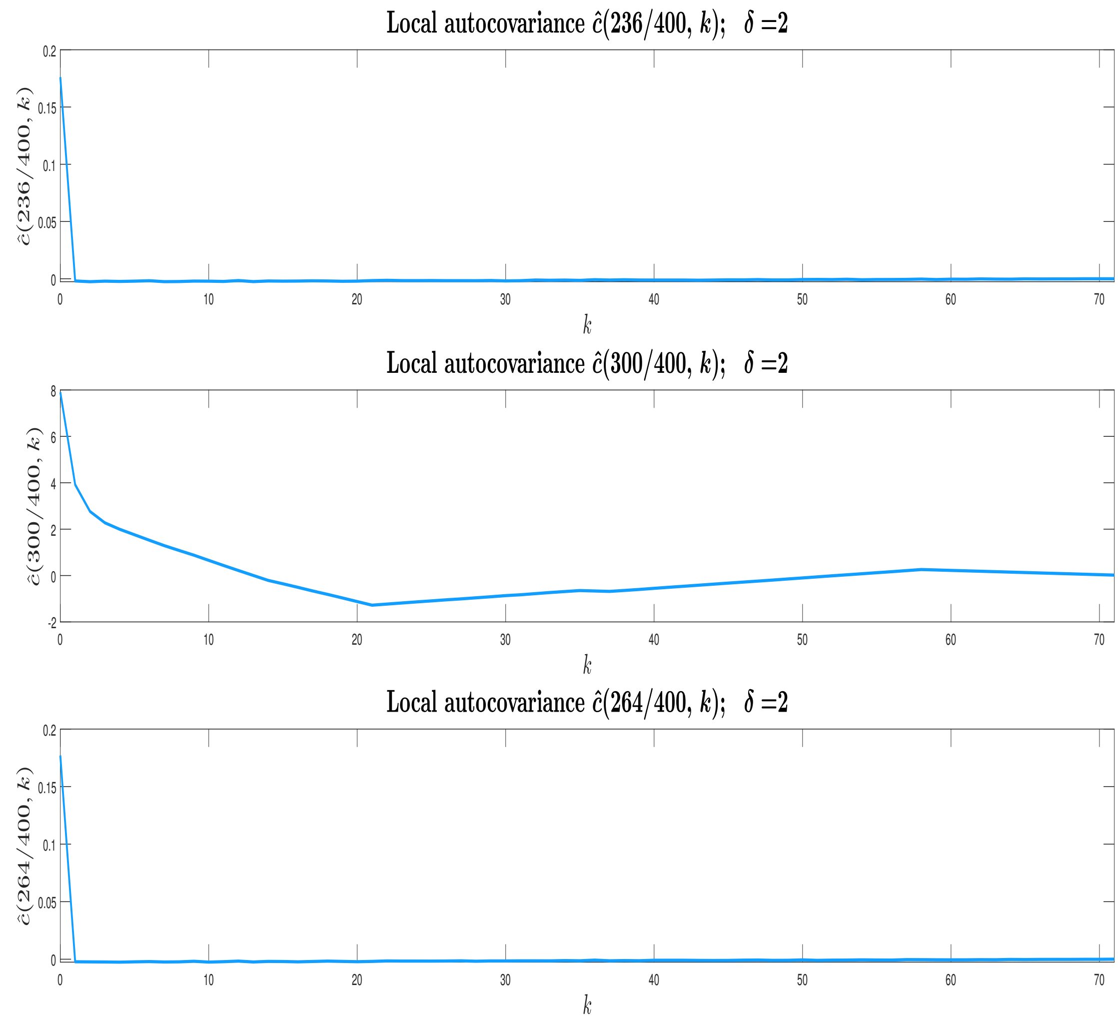

Figure S.3: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 2$.

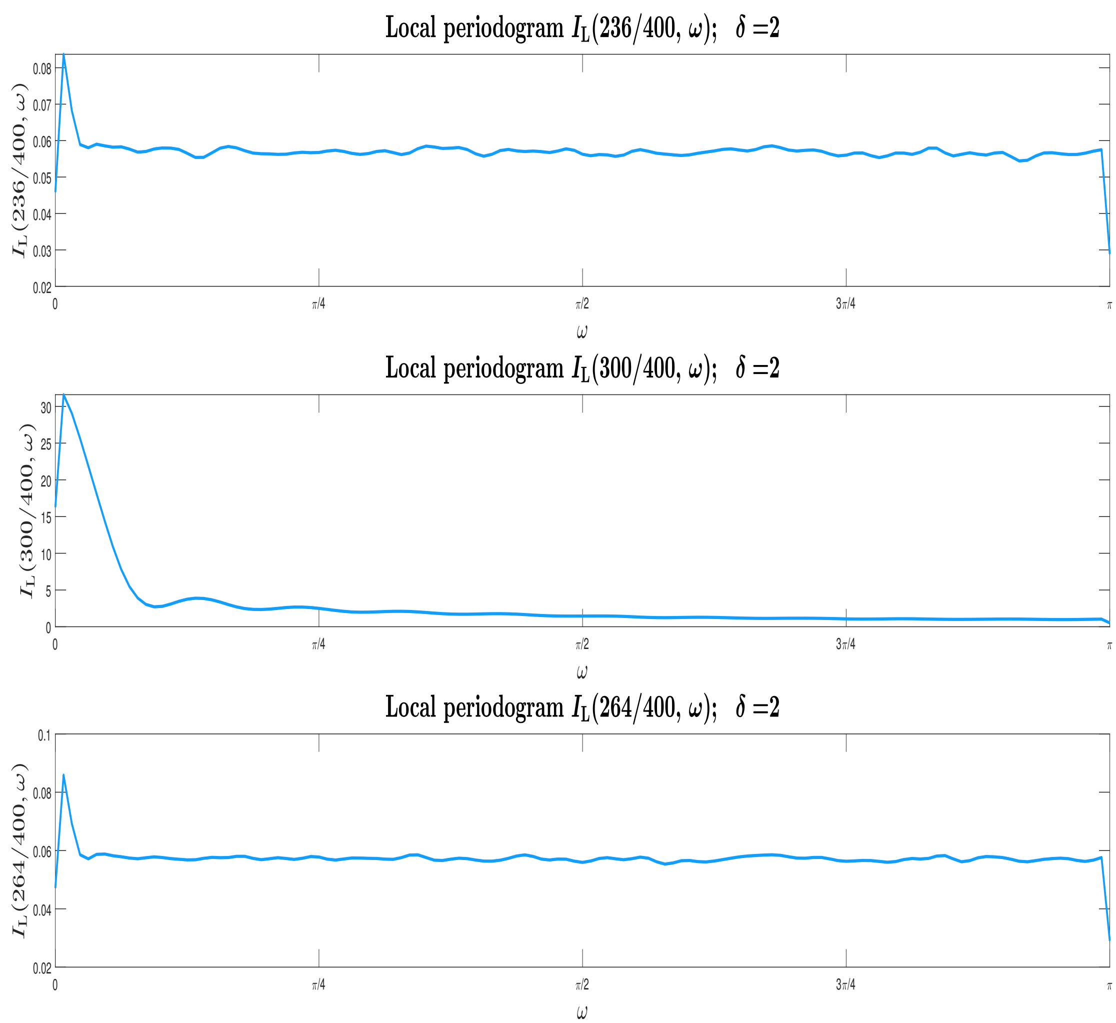

Figure S.4: The figure plots $I_{\mathrm{L}}(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 2$.

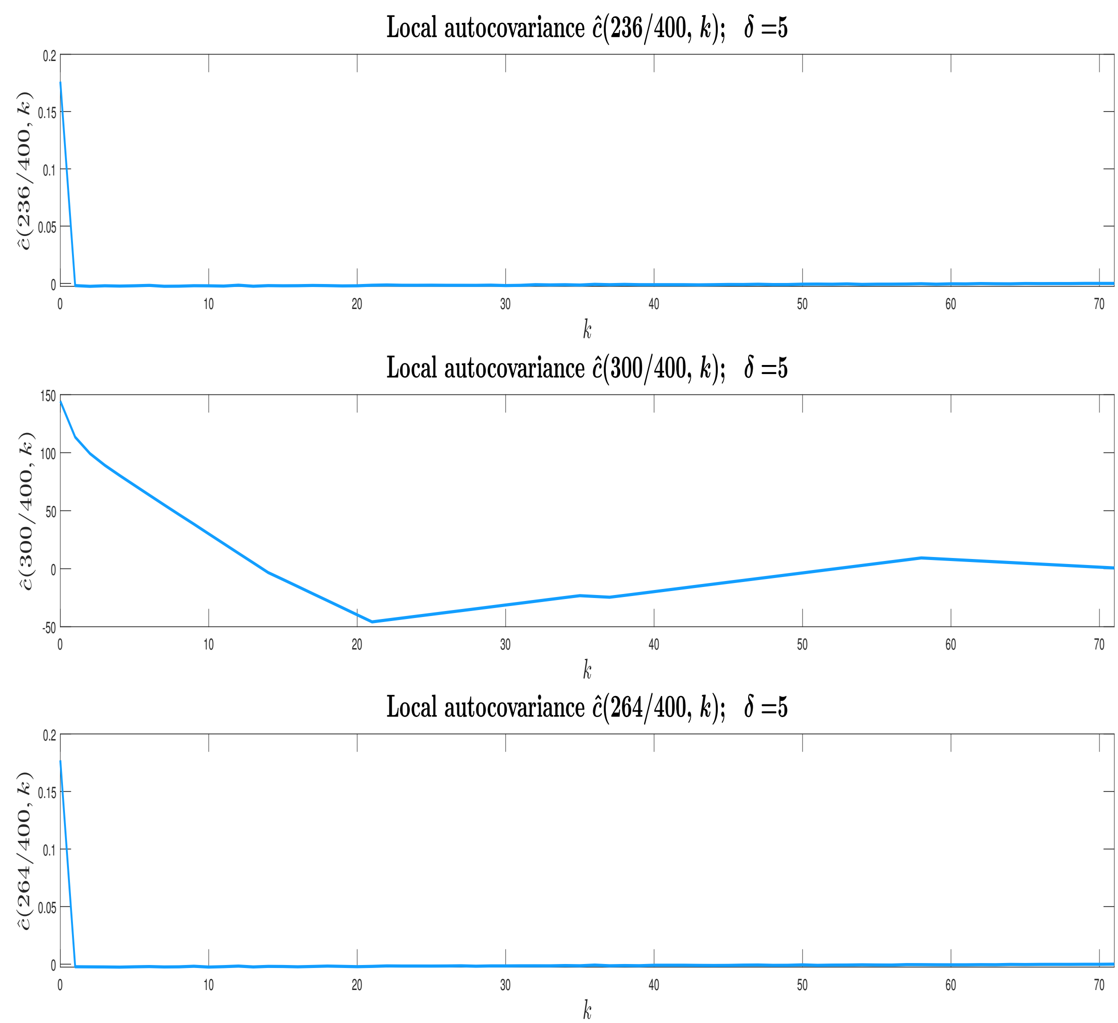

Figure S.5: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 5$.

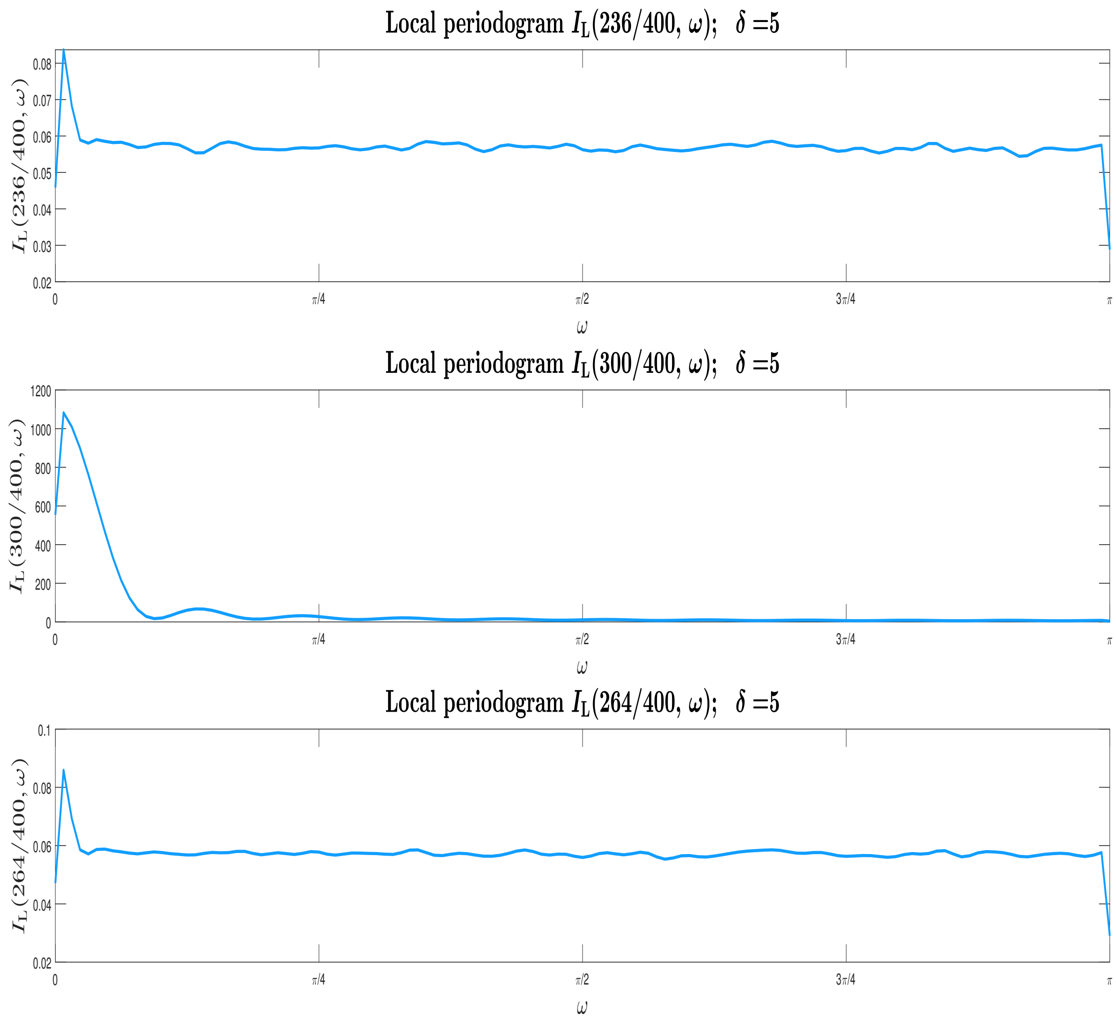

Figure S.6: The figure plots $I_{\mathrm{L}}(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 5$.

References

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Bentkus, R.Y., Rudzkis, R.A., 1982. On the distribution of some statistical estimates of spectral density. Theory of Probability and Its Applications 27, 795–814.

Casini, A., 2022. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Journal of Econometrics, forthcoming.

Casini, A., Perron, P., 2021. Change-point analysis of time series with evolutionary spectra. arXiv preprint arXiv 2106.02031.

Chibisov, D.M., 1972. An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. Theory of Probability and Its Applications 17, 621–630.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications. New York: Wiley.

Mikosch, T., Stărica, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economic and Statistics 86, 378–390.

Roueff, F., von Sachs, R., 2011. Locally stationary long memory estimation. Stochastic Processes and their Applications 121, 813–844.

Taniguchi, M., 1987. Validity of edgeworth expansions of minimum contrast estimators for gaussian ARMA processes. Journal of Multivariate Analysis 21, 1–28.

Taniguchi, M., Puri, M.L., 1996. Valid Edgeworth expansions of M-estimators in regression models with weakly dependent residuals. Econometric Theory 12, 331–346.

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