Econometric Theory, **0**, 2024, 1–42.
doi:10.1017/S026646662400032X

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

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 show theoretically that nonparametric smoothing over time is robust to low frequency contamination. Nonstationarity can have consequences for both the size and power of HAR tests. Under the null hypothesis there are larger size distortions than when data are stationary. Under the alternative hypothesis, existing LRV estimators tend to be inflated and HAR tests can 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 distortions in the rejection rates can be induced by time variation in the second moments even when there is no break in the mean.

We are grateful to Peter C.B. Phillips, 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.; e-mail: alessandro.casini@uniroma2.it. College of Finance and Statistics, Hunan University, 109 Shijiachong Road, Yuelu District, Changsha, Hunan 41006, China. e-mail: tsdeng@hnu.edu.cn. Dep. of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, US. e-mail: perron@bu.edu.

1. INTRODUCTION

Many economic and financial time series have nonstationary characteristics that need to be accounted for in inference [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 time-variation in the mean 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 involving changes in the mean, 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., $\lim_{T\to\infty} \Gamma_T(k) > 0$), that 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.

Several HAR inference problems 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 changes [e.g., Andrews (1993), Bai and Perron (1998), Casini and Perron (2021, 2022a, 2022b), 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 (2023) 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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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 et al. (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 (2024), 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 (2023). 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 an 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 an LRV estimator that keeps the bandwidth at a fixed fraction of the sample size and that converges weakly to a random variable [cf. Kiefer et al. (2000)]. Inference is then based on a nonstandard reference distribution and it is shown that fixed-b achieves high-order refinements [e.g., Sun et al. (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 (2024)] as the fixed-b

[1] See Dou (2024), Hwang and Sun (2017), Ibragimov et al. (2021), Ibragimov and Müller (2010), Jansson (2004), Kiefer and Vogelsang (2002, 2005), Lazarus et al. (2020), 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).

limiting distribution of the *t*/*F* statistic is non-pivotal under nonstationarity. 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* or t/F critical values. These bandwidths are typically larger than the MSE-optimal bandwidths used for the HAC estimators.

Recently, Casini (2023) questioned the performance of HAR inference 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, 2021, 2022b), 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 (2023), Perron (1991), 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 $k \to \infty$ 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 (2024) 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. Casini (2024) also developed fixed-*b* methods that are valid under nonstationarity and in fact provide better null rejection rates in finite-sample.

The Monte Carlo results suggest that under the null hypothesis nonstationarity can generate larger size distortions than what one finds under stationarity. In particular, fixed-smoothing methods can exhibit under-rejections whereas HAC and DK-HAC methods can exhibit over-rejections when there is strong persistence. For the latter problem, our second-order Edgeworth expansions could be used to construct corrections to the standard normal critical value. We relegate this opportunity to future research.

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 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 (2023). 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 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*

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

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

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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

where $\mathrm{cum}\{\cdots\}$ *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* $C < \infty$ *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,*

$$A(u, \omega) = A_j(u, \omega) \text{ for } \lambda_{j-1}^0 < u \leq \lambda_j^0, \tag{2}$$

$$\sup_{1 \leq j \leq m_0+1} \sup_{T_{j-1}^0 < t \leq T_j^0,\, \omega} \left| A_{j,t,T}^0(\omega) - A_j(t/T, \omega) \right| \leq CT^{-1}; \tag{3}$$

(iii) $\mu_\cdot(\cdot)$ *is piecewise Lipschitz continuous.*

Definition 1 states that $V_{t,T}$ has a time-varying spectral representation where both the mean $\mu_\cdot(\cdot)$ and transfer function $A_{\cdot,\cdot,T}^0(\omega)$ are piecewise continuous. Since the transfer function depends on the parameters that enter the second moments of $V_{t,T}$, the smoothness properties of $\mu_\cdot(\cdot)$ and A guarantee that $V_{t,T}$ has a piecewise locally stationary behavior. We require additional smoothness properties for A and an example is presented at the end of this section.

Assumption 1. (i) $\{V_{t,T}\}$ is an 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 $T_{j-1}^0/T < u = t/T \leq T_j^0/T$. 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

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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

Assumption 2. (i) $\sum_{k=-\infty}^{\infty}\sup_{u\in[0,1]}\|c(u,k)\| < \infty$ and $\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sup_{u\in[0,1]}|\kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}(k,j,l)| < \infty$ for all $a_1,a_2,a_3,a_4 \leq p$. (ii) For all $a_1,a_2,a_3,a_4 \leq p$ there exists a function $\widetilde{\kappa}_{a_1,a_2,a_3,a_4} : [0,1]\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\to\mathbb{R}$ such that $\sup_{1\leq j\leq m_0+1}\sup_{\lambda_{j-1}^0<u\leq\lambda_j^0}|\kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}(k,s,l)-\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,k,s,l)| \leq LT^{-1}$ 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 $u\neq\lambda_j^0 (j=1,\ldots,m_0+1)$ 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-(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 holds. Part (ii) extends the smoothness conditions on $A(u,\omega)$ in Assumption 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 $\lambda_1^0 = 0.5$,

$$V_{t,T} = \rho(t/T)V_{t-1,T} + \sigma(t/T)u_t, \tag{4}$$
$$\rho(u) = \begin{cases}\rho_1(u), & u\leq 0.5\\ \rho_2(u), & u>0.5\end{cases},$$

where $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are Lipschitz continuous, $\sigma(\cdot)$ is piecewise Lipschitz continuous and $\{u_t\}$ are i.i.d. random variables with mean zero and unit variance. Then, $V_{t,T}$ is an 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 1, $\sup_{u\in[0,1]}|\rho(u)| < 1$ and $\sup_{u\in[0,1]}\sigma(u) < \infty$, then $V_{t,T}$ fulfills Assumptions 1–2.

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

$$\widehat{\Gamma}(k) = T^{-1} \sum_{t=|k|+1}^{T} \left(V_t - \overline{V}\right)\left(V_{t-|k|} - \overline{V}\right), \tag{5}$$

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 (2023) considered an alternative estimate for the sample autocovariance to be used in the DK-HAC estimators, defined in Section 5.1, namely,

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

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

$$\begin{aligned} \widehat{c}_T(rn_T/T, k) = n_{2,T}^{-1} \sum_{s=0}^{n_{2,T}-1} & \left(V_{rn_T+\lfloor |k/2| \rfloor - n_{2,T}/2+s+1} - \overline{V}_{rn_T,T}\right) \\ & \times \left(V_{rn_T-\lfloor |k/2| \rfloor - n_{2,T}/2+s+1} - \overline{V}_{rn_T,T}\right), \end{aligned} \tag{6}$$

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. $\widehat{c}_T(rn_T/T, k)$ 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 $\widehat{\Gamma}_{\mathrm{DK}}(k)$ is an integrated local sample autocovariance. It extends $\widehat{\Gamma}(k)$ to better account for nonstationarity. Similarly, the DK-HAC estimator does not relate to the periodogram but to the local periodogram defined by

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

where $I_{\mathrm{L},T}(u, \omega)$ is the (untapered) periodogram over a segment of length n_T with midpoint $\lfloor Tu \rfloor$. We also consider the statistical properties of both $\widehat{\Gamma}_{\mathrm{DK}}(k)$ and $I_{\mathrm{L},T}(u, \omega)$ 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 $\lambda_j^0 = \sum_{s=0}^{j} r_s$.

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

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

with $\mu_j(\cdot)$ defined in (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 $\mu_j(t/T) = \mu_j$. 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.1 shows that for $\{V_{t,T}\}$ that satisfies Definition 1 and Assumptions 1–2, we have

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

and as $k \to \infty, \widehat{\Gamma}(k) \geq d^*$ $\mathbb{P}$-a.s. This suggests that $\widehat{\Gamma}(k)$ is asymptotically the sum of two terms. The first is the 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. The result that $\widehat{\Gamma}(k) \geq d^* \mathbb{P}$-a.s. as $k \to \infty$ implies that unaccounted nonstationarity generates long memory effects. The intuition is straightforward. A long memory SLS process satisfies $\sum_{k=-\infty}^{\infty} |\Gamma(u, k)| \to \infty$ for some $u \in (0, 1)$, similar to a stationary long memory process.[2] The theorem shows that $\widehat{\Gamma}(k)$ exhibits a similar property and $\widehat{\Gamma}(k)$ decays more slowly than for a short memory stationary process for small lags and approaches a constant $d^* > 0$ for large lags.

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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

Theorem S.2 in the supplement analyzes the properties of the periodogram $I_T(\omega_l)$ as $\omega \to 0$ when the mean is time-varying. The result states that as $\omega \to 0$ $\mathbb{E}(I_T(\omega))$ generally takes unbounded values except for some ω for which $\mathbb{E}(I_T(\omega))$ is bounded below by $2\pi\int_0^1 f(u,\omega)\,du > 0$. An 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.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 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 (6) 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 1. *Assume that* $\{V_{t,T}\}$ *satisfies Definition 1,* $n_T, n_{2,T} \to \infty$ *with* $n_T/T \to 0$, $n_{2,T}/T \to 0$ *and* $n_T/n_{2,T} \to 0$. *Under Assumptions 1- and 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*

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

(b) *if* $n_{j,L}(u,k,n_{2,T})/n_{2,T} \to 0$ *or* $n_{j,R}(u,k,n_{2,T})/n_{2,T} \to 0$, *then* $\widehat{c}_T(u,k) = c(u,k) + o_{\mathbb{P}}(1)$.

Further, if there exists an $r = 1, \dots, \lfloor T/n_T \rfloor$ *such that there exists a* $j = 1, \dots, m_0$ *with* $T_j^0 \in \mathbf{S}\left(rn_T, k, n_{2,T}\right)$ *satisfying (ii-a), then, as* $k \to \infty$, $\widehat{\Gamma}_{\text{DK}}(k) \geq d_T^*$ $\mathbb{P}$*-a.s., where* $d_T^* = \left(n_{2,T}/T\right) \gamma (1-\gamma) \left(\mu_j(\lambda_j^0) - \mu_{j+1}(u)\right)^2 > 0$ *and* $d_T^* \to 0$ *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 $\text{cov}(V_{\lfloor Tu \rfloor} V_{\lfloor Tu \rfloor - k}) = c(u,k) + O\left(T^{-1}\right)$ [see Casini (2023)]. If a change-point falls close to either boundary of the window $\mathbf{S}\left(u, k, n_{2,T}\right)$, 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 $\widehat{\Gamma}_{\text{DK}}(k)$ is an average of $\widehat{c}_T(rn_T, k)$ over blocks $r = 1, \dots, \lfloor T/n_T \rfloor$, if case (ii-a) holds then $\widehat{\Gamma}_{\text{DK}}(k) \geq d_T^*$ as $k \to \infty$ but $d_T^* \to 0$ as $T \to \infty$. Thus, comparing this result with the discussion above on $\widehat{\Gamma}(k)$ (see also Theorem S.1), in practice the long memory effects are unlikely to occur when using $\widehat{\Gamma}_{\text{DK}}(k)$. Furthermore, one can reduce this problem by appropriately choosing the blocks $r = 1, \dots, \lfloor T/n_T \rfloor$. A procedure was proposed in Casini (2023) using the methods developed in Casini and Perron (2024a).

We now study the asymptotic properties of $I_{\text{L},T}(u,\omega)$ 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 $l \neq 0 \pmod{n_T}$. We need the following high-level conditions. Part (i) corresponds to Assumption S.14, 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 $\sup_{t\geq 1} \mathbb{E}|V_t|^{4\nu} < \infty$, while part (iii) requires additional smoothness.

Assumption 3. (i) For each ω_l and $u \in [0,1]$ with $T_j^0 \in \mathbf{S}(u, 0, n_T)$ there exist $B_j \in \mathbb{R}$, $j = 1, \dots, m_0$ with $B_{j_1} \neq B_{j_2}$ for $j_1 \neq j_2$ such that

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

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

THEOREM 2. *Assume that* $\left\{V_{t,T}\right\}$ *satisfies Definition 1 and that* $n_T \to \infty$ *with* $n_T/T \to 0$. *Under Assumptions 1–3,*

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

(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 sub-cases: (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 2 with the discussion above about the periodogram (see also Theorem S.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 $\mathbb{E}\left(I_{\mathrm{L},T}(u, \omega)\right) \to \infty$ 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 $I_{\mathrm{L},T}(u, \omega_l)$ behaves very differently from the periodogram $I_T(\omega_l)$. Accordingly, nonstationarity is unlikely to generate long memory effects if one uses the local periodogram. As for $\widehat{c}_T(u, k)$, if one uses preliminary inference procedures [cf. Casini and Perron (2024a)] 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.[3] 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.

[3] This can be relaxed by considering distributions with Gram–Charlier representations at the expense of more complex derivations.

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

Let $\{V_t\}$ be a zero-mean Gaussian SLS process satisfying Assumption 1-(i-iv). Let

$$h_1 \triangleq \frac{\sqrt{T}\,\overline{V}}{\sqrt{J_T}} \sim \mathscr{N}(0,1), \tag{8}$$

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

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

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

$$W_{b_1}^{(r,s)} = w\left(b_{1,T}(r-s)\right) = \int_{\Pi} \widetilde{K}_{b_1}(\omega)\, e^{i(r-s)\omega} d\omega, \tag{9}$$

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

$$Z_T = Z_T(\mathbf{h}) = h_1\left(1 + \mathsf{B}_T + \mathsf{V}_T h_2\left(Tb_{1,T}\right)^{-1/2}\right)^{-1/2},$$

$$h_2 = \sqrt{Tb_{1,T}}\left(\frac{\widehat{J}_{\mathrm{HAC},T} - \mathbb{E}\left(\widehat{J}_{\mathrm{HAC},T}\right)}{J_T\mathsf{V}_T}\right),$$

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

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

where $\Upsilon_T = \mathbb{E}\left(\mathbf{V}'Q_T\mathbf{V}\right) = \mathrm{Tr}(\Sigma_V Q_T),\ \Sigma_V = \mathbb{E}\left(\mathbf{V}\mathbf{V}'\right)$, 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

$$\begin{aligned}\kappa_T(0, s) &= 2^{s-1}(s-1)!\,\mathrm{Tr}((\Sigma_V Q_T)^s), && s > 1,\\ \kappa_T(2, s) &= 2^s s!\,\xi_T'(\Sigma_V Q_T)^s\Sigma_V\xi_T, && s > 0.\end{aligned}$$

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

Assumption 4. For all $u \in [0, 1]$, $0 < f(u, 0) < \infty$ and $f(u, \omega)$ has d_f continuous derivatives $\left(d_f \geq 2\right) f^{(d_f)}(u, \omega)$ in a neighborhood of $\omega = 0$ and the d_fth derivative satisfies a Lipschitz condition of order ϱ with $\varrho \in (0, 1]$.

Assumption 5. 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 6. $|K(x)| < \infty$, $K(x) = K(-x)$, $K(x) = 0$ for $x \notin \Pi$ and $\int_\Pi K(x)\,dx = 1$.

Assumption 7. $K(x)$ satisfies a uniform Lipschitz condition of order 1 in $[-\pi, \pi]$.

Assumption 8. 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 9. $b_{1,T} + (Tb_{1,T})^{-1} \to 0$ as $T \to \infty$.

Assumption 10. $b_{1,T} = CT^{-q}$ where $0 < q < 1$ and $0 < C < \infty$.

Assumptions 6–10 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 9 is sufficient for the consistency of $\widehat{J}_{\mathrm{HAC},T}$ and is strengthened in Assumption 10, for some parts of the proofs, which is satisfied by popular MSE-optimal bandwidths [cf. Andrews (1991), Casini (2022), Belotti et al. (2023) and Whilelm (2015)].

Assumptions 4 and 5 impose conditions on the smoothness and boundedness of the spectral density. Assumption 4 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 5 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 Lemmas S.11–S.12 in the supplement we obtain

$$\mathsf{B}_T = \bar{c}_1 b_{1,T}^{d_f} + O\left(b_{1,T}^{d_f+\varrho} + T^{-1}\log T\right), \quad \text{where} \quad \bar{c}_1 = \frac{\mu_{d_f}(K)\int_0^1 f^{(d_f)}(u,0)\,du}{d_f!\int_0^1 f(u,0)\,du}. \tag{10}$$

The order of the asymptotic bias $b_{1,T}^{d_f}$ depends on the smoothness of the spectral density at $\omega = 0$ [cf. Assumption 4]. The constant $\bar{c}_1$ depends on the moment of order d_f of the kernel K and on the smoothness of $f(u, \omega)$ at $\omega = 0$. For example, for the time-varying AR(1) in (4),

$$f^{(2)}(u,0) = -\frac{\sigma^2(u)\rho(u)}{\pi\left(1+\rho(u)^2 - 2\rho(u)\right)^2}. \tag{11}$$

If there is positive dependence at time u, then $\rho(u) > 0$ and $f^{(2)}(u, 0) < 0$. Suppose $K(x) \geq 0$ for all x so that $\mu_2(K) > 0$. Then the sign of the bias is determined by the sign of $\int_0^1 f^{(2)}(u,0)\,du$. A positive local AR(1) coefficient contributes negative bias which corresponds to the well-known downward bias of the LRV estimator when there is positive dependence. Conversely, with anti-persistence $\rho(u) < 0$ and $f^{(2)}(u, 0) > 0$. Since $\rho(\cdot)$ is time-varying, whether the bias is positive or negative depends on the path of $\rho(\cdot)$. The smoother the spectral density is at frequency zero, the smoother the kernel and the slower $b_{1,T}$ can be. The factor $\int_0^1 f(u,0)\,du$ in the denominator follows by definition because B_T is the relative bias.

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,

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

where $\mathcal{H}_j(\cdot)$ are the univariate Hermite polynomials of order j, and $\Xi_0(0,3) = (4\pi)^{1/2}\, 2! \int_\Pi K^3(\omega) \quad d\omega \|K\|_2^{-3}$ and $\Xi_0(2,1) = (4\pi)^{1/2} K(0)\|K\|_2^{-1}$ (see

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

Lemmas S.13 and S.14). 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 3. *Let Assumptions 4, 5* ($p > 1$)*, 6–7 and 10* ($0 < q < 1$) *hold. For* $\phi_T = (Tb_{1,T})^{-\varpi}$ *with* $1/2 < \varpi < 1$*, we have*

$$\sup_{\mathbf{B}\in\mathscr{B}^2}\left|\mathbb{P}_T(\mathbf{B}) - \mathbb{Q}_T^{(2)}(\mathbf{B})\right| = o\left(\left(Tb_{1,T}\right)^{-1/2}\right) + (4/3)\sup_{\mathbf{B}\in\mathscr{B}^2}\mathbb{Q}_T^{(2)}\left((\partial\mathbf{B})^{2\phi_T}\right). \tag{12}$$

Theorem 3 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.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 (12) 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 Lemmas S.13 and S.14 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,

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

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

THEOREM 4. *Let Assumptions 4, 5* ($p > 1$)*, 6–8, and 10* ($q = 1/\left(1+2d_f\right)$) *hold. For a convex Borel set* $\mathbf{C}$*, we have, for* $r_2(x) = -\overline{c}_1\left(x^2-1\right)/2$*,*

$$\sup_{\mathbf{C}}\left|\mathbb{P}(Z_T\in\mathbf{C}) - \int_{\mathbf{C}}\varphi(x)\left(1 + r_2(x)\,b_{1,T}^{d_f}\right)dx\right| = o\left(\left(Tb_{1,T}\right)^{-1/2}\right). \tag{13}$$

Theorem 4 shows the form of the correction term to the standard normal distribution, i.e., $b_{1,T}^{d_f}\int_{\mathbf{C}}\varphi(x)\,r_2(x)\,dx$. 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).

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

$$\begin{aligned}\mathbb{P}(Z_T \leq z) &= \Phi(z) + \frac{1}{2}\overline{c}_1 z\varphi(z)\, b_{1,T}^{d_f} + o\left(\left(Tb_{1,T}\right)^{-1/2}\right) \qquad \textbf{(14)}\\ &= \Phi\left(z\left(1+\frac{1}{2}\overline{c}_1 b_{1,T}^{d_f}\right)\right) + o\left(\left(Tb_{1,T}\right)^{-1/2}\right) = \Phi(z) + O\left(\left(Tb_{1,T}\right)^{-1/2}\right).\end{aligned}$$

This shows that under the conditions of Theorem 4, the standard normal approximation is correct up to order $O((Tb_{1,T})^{-1/2})$. Eq (14) has an immediate interpretation. Consider the time-varying AR(1) example in (4) and suppose $K(x) \geq 0$ for all x so that $\mu_2(K) \geq 0$. Given (11) we know that with local positive persistence (i.e., $\rho(u) > 0$) $f(u, \omega)$ has a peak at $\omega = 0$. If the pattern of $\rho(u)$ is such that $\int_0^1 f^{(2)}(u, 0)\, du < 0$ so that the positive persistence dominates, then $\overline{c}_1 < 0$ and as is well-known the HAC estimator underestimates the true LRV and the corresponding HAC-based test over-rejects. The approximation in (14) tends to correct this problem as it follows that one uses $\Phi(z(1+\gamma_T))$ where $\gamma_T \leq 0$, so for a given significance level the critical value z is larger in absolute value than the corresponding standard normal critical value. Conversely, if there is anti-persistence, then $\overline{c}_1 > 0$ and the implied critical value is smaller than the corresponding standard normal critical value. For $d_f > 2$ the reasoning is the same but one has to take into account the sign of $\mu_{d_f}(K)$.

Consider the location model $y_t = \beta + V_t\, (t = 1, \ldots, T)$. 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 and (14) imply that

$$\mathbb{P}(t_{\mathrm{HAC}} \leq z) = \Phi(z) + p(z)\left(Tb_{1,T}\right)^{-1/2} + o\left(\left(Tb_{1,T}\right)^{-1/2}\right), \qquad \textbf{(15)}$$

for any $z \in \mathbb{R}$, where $p(z)$ is an odd function. When $q = 1/\left(1+2d_f\right)$, we have $p(z) = 2^{-1}\overline{c}_1 z\varphi(z)\, C^{d_f+1/2}$ where C is defined in Assumption 10. 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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

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

with K_2 a kernel and $b_{2,T}$ a bandwidth. Note that $\widehat{\Gamma}_{\mathrm{DK}}(k)$ and $\widehat{\Gamma}^{*}_{\mathrm{DK}}(k)$ 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 11. $K_2(\cdot):\mathbb{R}\to[0,\infty]$, $K_2(x)=K_2(1-x)$, $\int_0^1 K_2(x)\,dx=1$, $K_2(x)=0$ for $x\notin[0,1]$ and $K_2(\cdot)$ is continuous. The bandwidth sequence $\{b_{2,T}\}$ satisfies $b_{2,T}\to 0$, $b^{2}_{2,T}/b^{q_2}_{1,T}\to\overline{b}\in[0,\infty)$ and $1/Tb_{1,T}b_{2,T}\to 0$ where q_2 is the index of smoothness of $K_1(\cdot)$ at 0.

Under Assumptions 67, 9, and 11 it holds that $\widehat{J}^{*}_{\mathrm{DK},T}-J_T\xrightarrow{\mathbb{P}}0$ [cf. Casini (2023)] and

$$U_T\triangleq\frac{\sqrt{T}\,\overline{V}}{\sqrt{\widehat{J}^{*}_{\mathrm{DK},T}}}\xrightarrow{d}\mathscr{N}(0,1).\tag{16}$$

Note that $\widehat{J}^{*}_{\mathrm{DK},T}=\int_0^1\widetilde{\mathbf{V}}(r)'W_{b_1}\widetilde{\mathbf{V}}(r)\,dr/(Tb_{2,T})$ where $\widetilde{\mathbf{V}}(r)=(\widetilde{V}_1(r),\widetilde{V}_2(r),\ldots,\widetilde{V}_T(r))'$ with $\widetilde{V}_j(r)=\sqrt{K_2\left((r-j)/Tb_{2,T}\right)}V_j$ and W_{b_1} defined in (9). Let

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

$\widetilde{I}_T(r,\omega)$ 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

$$U_T=U_T(\mathbf{v})=v_1\left(1+\mathsf{B}_{2,T}+\mathsf{V}_{2,T}v_2\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right)^{-1/2},$$

$$v_2=\sqrt{Tb_{1,T}b_{2,T}}\left(\frac{\widehat{J}^{*}_{\mathrm{DK},T}-\mathbb{E}\left(\widehat{J}^{*}_{\mathrm{DK},T}\right)}{J_T\mathsf{V}_{2,T}}\right),$$

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

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

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., $\widetilde{\mathbf{C}} = \{[0, 1]/\{\lambda_j^0, j = 1, \ldots, m_0\}\}$. Define

$$\Delta_f(\omega)$$
$$= \sum_{j=1}^{m_0} \int_0^1 \left(\frac{\partial}{\partial u_-} f\left(\lambda_j^0, \omega\right) \int_0^{1-s} x K_2(x)\, dx + \frac{\partial}{\partial u_+} f\left(\lambda_j^0, \omega\right) \int_{1-s}^1 x K_2(x)\, dx \right) ds,$$

where

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

Assumption 12. For $u \in \widetilde{\mathbf{C}}$, $\left(\partial^2/\partial u^2\right) f(u, \omega)$ 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 Lemmas S.11 and S.17, the relative bias of $\widehat{J}_{\mathrm{DK},T}^*$ is

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

where

$$\overline{c}_1 = \frac{\mu_{d_f}(K) \int_0^1 f^{(d_f)}(u, 0)\, du}{d_f! \int_0^1 f(u, 0)\, du},$$

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

$$\bar{c}_2 = \frac{2^{-1}\int_0^1 x^2 K_2(x)\,dx \int_{\widetilde{\mathbf{C}}} \frac{\partial^2}{\partial u^2} f(u,0)\,du + \Delta_f(0)}{\int_0^1 f(u,0)\,du}.$$

The factor $\bar{c}_1$ in the relative bias $\mathsf{B}_{2,T}$ also enters B_T and we already discussed it. The second factor, $\bar{c}_2$, includes two elements. The first depends on the second moment of the kernel K_2 and on the smoothness over time of the spectral density $f(u,0)$. The second element in $\bar{c}_2$ is $\Delta_f(0)$ which depends on the right and left first partial derivatives of $f(u,0)$ with respect to u at the discontinuity points. The more nonstationary is the data the more complex is $\bar{c}_2$, and in fact the larger in magnitude are $\partial^2 f(u,0)/\partial u^2$ and $\Delta_f(0)$. For the special case of stationary data, $\bar{c}_2 = 0$. The more nonstationary is the data, the smaller $b_{2,T}$ should be chosen so as to weight more the data locally. The smoothing over sample autocovariances is needed to achieve consistency while the time-smoothing is introduced to more flexibly account for the time-varying properties of the data. The disadvantage of the time-smoothing is that it reduces the effective sample size thereby making accounting for strong dependence more difficult.

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 $\mathbb{Q}^{(2)}_{2,T}(\mathbf{B}) = \int_{\mathbf{B}} \varphi_2(\mathbf{v})\, q^{(2)}_{2,T}(\mathbf{v})\, d\mathbf{v}$, where

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

$\mathcal{H}_{2,j}(\cdot)$ 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 $f(u,0)$ (see Lemmas S.5 and S.6).

THEOREM 5. *Let Assumptions 4, 5* ($p > 1$)*, 6, 7, 10* ($0 < q < 1$)*, 11, 12 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*

$$\sup_{\mathbf{B}\in\mathscr{B}^2}\left|\mathbb{P}_T(\mathbf{B}) - \mathbb{Q}^{(2)}_{2,T}(\mathbf{B})\right| = o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right) + (4/3)\sup_{\mathbf{B}\in\mathscr{B}^2}\mathbb{Q}^{(2)}_{2,T}\left((\partial\mathbf{B})^{2\phi_T}\right). \tag{17}$$

Theorem 5 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 $q^{(2)}_{2,T}(\mathbf{v})$ differs from $q^{(2)}_T(\mathbf{h})$ in Theorem 3. This difference depends on the smoothing over time, i.e., on $b_{2,T}$ and $K_2(\cdot)$. The theorem also suggests that the leading term of the error of the approximation is of order $o((Tb_{1,T}b_{2,T})^{-1/2})$.

Next, we focus on U_T defined in (16), i.e., a t-statistic based on $\widehat{J}^*_{\mathrm{DK},T}$, and present the Edgeworth expansion. We need the following assumption, replacing

Assumptions 9 and 10, 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 $\mathbb{Q}_{2,T}^{(2)}(\mathbf{B})$ 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 (2023).

Assumption 13. 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 6. *Let Assumptions 4, 5* ($p > 1$)*, 6–8, and 11–13 hold. For convex Borel sets* $\mathbf{C}$*, we have, for* $r_2(x) = -\overline{c}_1\left(x^2-1\right)/2$ *and* $r_3(x) = -\overline{c}_2\left(x^2-1\right)/2$*,*

$$\sup_{\mathbf{C}}\left|\mathbb{P}\left(U_T \in \mathbf{C}\right) - \int_{\mathbf{C}} \varphi(x)\left(1 + r_2(x)\, b_{1,T}^{d_f} + r_3(x)\, b_{2,T}^{2}\right) dx\right| = o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right). \tag{18}$$

Theorem 6 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 for the HAC estimators. Similar to (14), we obtain uniformly in z,

$$\mathbb{P}\left(U_T \le z\right) = \Phi\left(z\left(1 + \frac{1}{2}\overline{c}_1 b_{1,T}^{d_f} + \frac{1}{2}\overline{c}_2 b_{2,T}^{2}\right)\right) + O\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right), \tag{19}$$

where $\mathbf{C} = (-\infty, z]$, which suggests that the standard normal approximation is correct up to order $O((Tb_{1,T}b_{2,T})^{-1/2})$. Eq (19) has a similar interpretation to (14). Consider the time-varying AR(1) example in (4) and suppose $\rho(u) > 0$ for all u. Then, $\overline{c}_1 < 0$. However, the sign of $\overline{c}_2$ is not easily determined even for this simple model. For the special case $\rho(u) = \sin(u\pi/10)$, no break and $\sigma^2(u) = \sigma^2$ we have $\overline{c}_2 < 0$. Then, the implied critical value from the approximation is larger than the standard normal critical value. In general, however, the correction to strong persistence might be either attenuated or strengthened by the correction to nonstationarity depending on the true data-generating process.

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 6 and (19) imply that

$$\mathbb{P}\left(t_{\mathrm{DK}} \le z\right) = \Phi(z) + p_2(z)\left(Tb_{1,T}b_{2,T}\right)^{-1/2} + o\left(\left(Tb_{1,T}b_{2,T}\right)^{-1/2}\right), \tag{20}$$

for any $z \in \mathbb{R}$, where $p_2(z)$ is an odd function. Under the conditions of Theorem 6, $p_2(z) = 2^{-1}((C^{d_f+1/2}\overline{c}_1 + C_2\overline{c}_2)z\varphi(z))$, where C is defined in Assumption 10,

$C_2 = (\overline{b}C^{d_f+1/2})^{1/2}$ and $\overline{b}$ is defined in Assumption 11. Thus, the ERP of t_{DK} can be larger than that of t_{HAC}, though the margin is small. This follows from the fact that $\widehat{J}^*_{\mathrm{DK},T}$ applies smoothing over two directions. The smoothing over time is useful to flexibly account for nonstationarity. Its benefits appear explicitly under the alternative hypothesis as we show in Section 5 whereas the ERP refers to the null hypothesis. One can 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 (2024) 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 breaks down under nonstationarity as shown by Casini (2024) 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 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.

5. CONSEQUENCES FOR HAR INFERENCE

In this section, we discuss the implications of the theoretical results from Section 3 and 4. In Section 5.1, we first present a review of HAR inference methods and their connection to the estimates considered in Section 3. In Section 5.2, we present evidence that the HAR inference tests can suffer from larger size distortions under nonstationarity than under stationarity. In Section 5.3, we show the consequences of low frequency contamination for the power of the HAR tests and we provide the corresponding theoretical results in Section 5.4.

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 (8). The form of

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

$\{V_t\}$ 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,

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

where $\widehat{\Gamma}(k)$ is given in (5) 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 $\widehat{J}_{\mathrm{HAC},T} \overset{\mathbb{P}}{\to} J$. 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 keeps the bandwidth at some fixed fraction of T [cf. Kiefer et al. (2000)], e.g., using all autocovariances, so that $\widehat{J}_{\mathrm{KVB},T} \triangleq T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(1-|t-s|/T)\,\widehat{V}_t\widehat{V}_s$ 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 (2024)]. 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 (2024b) for simulation results]. For comparison we also report the equally-weighted cosine (EWC) estimator of Lazarus et al. (2020). 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 (2023). Motivated by the power impact of low frequency contamination of existing LRV estimators, he proposed a double kernel HAC (DK-HAC) estimator, defined by

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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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

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

with K_2 a kernel and $b_{2,T}$ a bandwidth. Note that $\widehat{c}_{\mathrm{DK},T}$ and $\widehat{c}_T$ are asymptotically equivalent and the results of Section 3 continue to hold for $\widehat{c}_{\mathrm{DK},T}$. More precisely, $\widehat{c}_T$ is a special case of $\widehat{c}_{\mathrm{DK},T}$ with K_2 being a rectangular kernel and $n_{2,T} = Tb_{2,T}$. This approach falls in the first category of standard inference $\widehat{J}_{\mathrm{DK},T} \xrightarrow{\mathbb{P}} J$ and HAR test statistics normalized by $\widehat{J}_{\mathrm{DK},T}$ follows standard distribution asymptotically. The DK-HAC estimator involves two kernels: K_1 smooths the lagged sample autocovariances, akin to the classical HAC estimators, while K_2 applies smoothing over time. The latter feature is useful to avoid the low frequency contamination. Additionally, Casini and Perron (2024b) proposed prewhitened DK-HAC ($\widehat{J}_{\mathrm{pw,DK},T}$) estimator that improves the size control of HAR tests and enjoys the same asymptotic properties of $\widehat{J}_{\mathrm{DK},T}$. Casini (2023) and Casini and Perron (2024b) demonstrated via simulations that tests based on $\widehat{J}_{\mathrm{DK},T}$ and $\widehat{J}_{\mathrm{pw,DK},T}$ have superior power properties relative to tests based on the other estimators. In terms of size, the simulation results showed that tests based on $\widehat{J}_{\mathrm{pw,DK},T}$ perform better than those based on $\widehat{J}_{\mathrm{HAC},T}$ and $\widehat{J}_{\mathrm{DK},T}$, and is competitive with $\widehat{J}_{\mathrm{KVB},T}$ when the latter works well. We include $\widehat{J}_{\mathrm{DK},T}$ and $\widehat{J}_{\mathrm{pw,DK},T}$ 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 (2024a)] because we do not want the results to be affected by such pre-test.

5.2. Null Rejection Rates and Power in Finite-Sample

In order to better understand the effect of nonstationarity on the null rejection rates of HAR tests we first conduct a Monte Carlo analysis where we compare a nonstationary model with a stationary one that has either the same spectral density at frequency zero or the same average dependence. Consider the following four AR(1) data-generating processes (DGPs). DGP 1 is given by

$$V_t = 0.26V_{t-1} + e_t, \qquad t = 1, \ldots, T,$$

where $e_t \sim \mathscr{N}(0, 1)$ for all t. The LRV of DGP 1 is $J = 1.826$. DGP 2 is

$$V_t = 0.7817V_{t-1} + e_t, \qquad t = 1, \ldots, T,$$

where $e_t \sim \mathscr{N}(0, 1)$ for all t. Its LRV is $J = 20.988$. We now introduce two nonstationary DGPs. DGP 3 takes the following form

$$V_t = \begin{cases} 0.9V_{t-1} + e_t, & 1 \le t \le 0.2T \\ 0.1V_{t-1} + e_t, & 0.2T < t \le T, \end{cases}$$

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

where $e_t \sim \mathscr{N}(0, 1)$. Note that the spectral density at frequency zero of V_t is given by the weighted average of the spectral densities of V_t in the two regimes:

$$f(0) = \int_0^1 f(u, 0)\, du$$
$$= 0.2\frac{1}{2\pi\left(1-2\cdot 0.9+0.9^2\right)} + 0.8\frac{1}{2\pi\left(1-2\cdot 0.1+0.1^2\right)} = 3.342.$$

Thus, the LRV of V_t is $J = 2\pi \int_0^1 f(u, 0)\, du = 20.988$ which takes the same value as the LRV of DGP 2. Further, DGP 3 has the same average dependence as DGP 1, meaning that the AR(1) coefficient in DGP 1 is equal to the weighted average of the AR(1) coefficients of DGP 3 in the two regimes, i.e., $\overline{\rho} = 0.2 \cdot 0.9 + 0.8 \cdot 0.1 = 0.26$. We also want to verify whether the location of the break in persistence in DGP 3 is important for the bias. Thus, we consider DGP 4:

$$V_t = \begin{cases} 0.1V_{t-1} + e_t, & 1 \leq t \leq 0.5T \\ 0.9V_{t-1} + e_t, & 0.5T < t \leq 0.5T + 0.2T \\ 0.1V_{t-1} + e_t, & 0.5T + 0.2T < t \leq T, \end{cases}$$

where $e_t \sim \mathscr{N}(0, 1)$ for all t. While in DGP 3 the regime with strong persistence occurs in the first 20% of the sample, in DGP 4 it occurs between the 50% and 70% of the sample. The LRV of DGP 4 is the same as that of DGP 3.

For each DGP we consider three different initial conditions: (a) $V_0 = 0$; (b) $V_0 \sim \mathscr{N}(0, 1)$; (c) $V_0 \sim \mathscr{N}(0, 4)$. This is useful in order to verify whether the initial condition has any effect on the bias generated by changes in the second-order properties. DGP 3(a) should exhibit a smaller bias due to nonstationarity than DGP 3(b,c) and 4. To see this, note that in DGP 3(a) the initial condition is $V_0 = 0$. Thus, the process starts from zero. Since there is strong persistence in the first 20% of the sample, the process is more likely to stay close to zero in the first regime than when the initial condition is $V_0 \sim \mathscr{N}(0, 1)$ or $V_0 \sim \mathscr{N}(0, 4)$. In DGP 4 the different specifications of the initial condition should not lead to any differences in the bias due to nonstationarity because the regime with strong dependence occurs about mid-sample.

To summarize, we have four DGPs. DGP 1 and 2 are stationary while DGP 3 and 4 are nonstationary. Since DGP 2 has a LRV that takes the same value as that of DGP 3 and 4, this allows us to better separate the effect of persistence from that of nonstationarity in the second moments on the following quantities: $\widehat{J}_{\mathrm{HAC}}$, $-\widehat{c}_1 b_{1,T}$ and $\widehat{\Gamma}(k)$ for $k = 0, 1, 5, 10$. In the simulations below $\widehat{J}_{\mathrm{HAC}}$ is the Newey–West estimator based on a predetermined number of lagged sample autocovariances following the rule $4(T/100)^{2/9}$ [cf. Lazarus et al. (2018)]. We compare $\widehat{\Gamma}(k)$ to the theoretical value $\Gamma_T(k)$ corresponding to each DGP which can be computed by hand given the simple form of the DGPs. In fact, for the nonstationary DGPs, $\Gamma_T(k)$ is a weighed average of the theoretical autocovariances corresponding to each regime. Here, $\widehat{c}_1$ is an estimate of $\overline{c}_1$ in (10) that enters the asymptotic bias of $\widehat{J}_{\text{HAC}}$. In order to compute $\widehat{c}_1$, we recall that the asymptotic bias of the LRV estimator based on the Bartlett kernel is given by

$$\lim_{T\to\infty} b_{1,T}^{-1}\mathbb{E}\left(\widehat{J}_{\text{HAC}} - J_T\right) = -2\pi K_{\text{BT},1}\int_0^1 f^{(1)}(u,0)\,du,$$

where

$$K_{\text{BT},q} = \lim_{x\to 0}\frac{1-K_{\text{BT}}(x)}{|x|^q}$$

denotes the index of smoothness of the kernel at zero and $f^{(1)}(u,0)$ is the index of smoothness of the local spectral density at time u and frequency zero. For the Bartlett kernel $K_{\text{BT},q} = 0$ if $q < 1$, $K_{\text{BT},q} = 1$ if $q = 1$ and $K_{\text{BT},q} = \infty$ if $q > 1$. The Parzen characteristic exponent is the largest q such that $K_{\text{BT},q}$ is finite. Thus, the relative bias is

$$\lim_{T\to\infty} b_{1,T}^{-1}\mathbb{E}\left(\widehat{J}_{\text{HAC}}/J_T - 1\right) = -K_{\text{BT},1}\frac{\int_0^1 f^{(1)}(u,0)\,du}{\int_0^1 f(u,0)\,du} = -\overline{c}_1,$$

using $K_{\text{BT},1} = 1$. The index of smoothness of $f(u,\omega)$ at $\omega = 0$ is defined as

$$f^{(1)}(u,0) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}|k|\Gamma(u,k).$$

For an AR(1) process with parameters $\rho(u)$ and $\sigma_e^2(u)$, we have $\Gamma(u,k) = \sigma_e^2(u)\rho(u)^{|k|}/(1-\rho(u)^2)$. It follows that

$$f^{(1)}(u,0) = -\frac{1}{2\pi}\frac{2\rho(u)\sigma_e^2(u)}{(\rho(u)-1)^3(1+\rho(u))}.$$

Based on this result we can obtain $\overline{c}_1$ for each model. In particular, for model DGP 1, 2, 3, and 4 we have $\overline{c}_1 = 0.55, 3.92, 9.04$ and 9.05, respectively.

We estimate $\overline{c}_1$ as follows. For DGP 1, we obtain the OLS residuals $\widehat{V}_t$ and estimate ρ and σ_e^2 from the autoregression

$$\widehat{V}_t = \rho\widehat{V}_{t-1} + e_t, \qquad t = 1,\ldots,T,$$

where σ_e^2 is the variance of e_t. Let these estimates be denoted by $\widehat{\rho}$ and $\widehat{\sigma}_e^2$, respectively. Then, the estimate of $\overline{c}_1$ is defined as

$$\widehat{c}_1 = -\frac{2\widehat{\rho}\widehat{\sigma}_e^2}{\widehat{J}_{\text{HAC}}(\widehat{\rho}-1)^3(1+\widehat{\rho})}.$$

The same applies to DGP 2. For DGP 3, we obtain the estimate of the autoregressive coefficient of V_t and of the variance of the innovations by estimating the autoregression in the two regimes separately. That is, we obtain

$$\widehat{V}_t = \begin{cases}\widehat{\rho}_1\widehat{V}_{t-1} + \widehat{e}_t, & 1 \le t \le 0.2T\\ \widehat{\rho}_2\widehat{V}_{t-1} + \widehat{e}_t, & 0.2T < t \le T,\end{cases}$$

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

where we also compute $\widehat{\sigma}_{1,e}^2$ and $\widehat{\sigma}_{2,e}^2$ which are the sample variances of the residuals $\widehat{e}_t$ in the two regimes, respectively. Then, the estimate of $\overline{c}_1$ is defined as

$$\widehat{c}_1 = -0.2 \frac{2\widehat{\rho}_1 \widehat{\sigma}_{1,e}^2}{\widehat{J}_{\text{HAC}} (\widehat{\rho}_1 - 1)^3 (1 + \widehat{\rho}_1)} - 0.8 \frac{2\widehat{\rho}_2 \widehat{\sigma}_{2,e}^2}{\widehat{J}_{\text{HAC}} (\widehat{\rho}_2 - 1)^3 (1 + \widehat{\rho}_2)}.$$

The same applies to DGP 4 with the difference that the autoregressive coefficient and the variance of the innovations are estimated separately in each of the three distinct regimes.

We consider the sample size $T = 100, 200,$ and 1000, and 50,000 repetitions were used for each DGP. The results are reported in Table 1. Let us first discuss the finite-sample properties of $\widehat{J}_{\text{HAC}}$. The results clearly suggest that $\widehat{J}_{\text{HAC}}$ deviates substantially from J when the data are nonstationary. $\widehat{J}_{\text{HAC}}$ underestimates J for all DGPs but it does so much more when the DGP is nonstationary. The difference between the values of $\widehat{J}_{\text{HAC}}$ in DGP 2 and those in DGP 3-4 is about one half, e.g., $\widehat{J}_{\text{HAC}} = 6.775$ in DGP 2(a) and $\widehat{J}_{\text{HAC}} = 3.142$ in DGP 3(a). As the sample size increases the downward bias becomes smaller, though $\widehat{J}_{\text{HAC}}$ still underestimates J for $T = 1000$. The downward bias continues to remain larger in DGP 3-4 than in DGP 2 even when $T = 1000$. Thus, this evidence based on $\widehat{J}_{\text{HAC}}$ already points out that basic forms of nonstationarity generate bias in the LRV estimator. This bias adds to the well-known bias generated by strong persistence in stationary data documented in the literature.

Let us discuss the relative bias $-\overline{c}_1 b_{1,T}$ and its estimate $-\widehat{c}_1 b_{1,T}$. First note that $-\overline{c}_1 b_{1,T} < 0$ and $-\widehat{c}_1 b_{1,T} < 0$ for all DGPs and sample sizes considered. This confirms the downward bias of $\widehat{J}_{\text{HAC}}$ observed above. For a given model, the asymptotic relative bias $-\overline{c}_1 b_{1,T}$ and its estimate increase with the sample size. The downward bias is much larger for the nonstationary DGP 3-4 than for the stationary DGP 1-2. The estimates $-\widehat{c}_1 b_{1,T}$ of the relative bias $-\overline{c}_1 b_{1,T}$ significantly underestimate $-\overline{c}_1 b_{1,T}$ in DGP 3-4 while in DGP 1-2 the deviations are much smaller. The large deviations of $-\widehat{c}_1 b_{1,T}$ from $-\overline{c}_1 b_{1,T}$ continue to hold even for $T = 1000$.

We now move to discuss the finite-sample properties of $\widehat{\Gamma}(k)$. When the data are stationary, $\widehat{\Gamma}(k)$ is close to $\Gamma_T(k)$ even when $T = 100$ and it approaches $\Gamma_T(k)$ when $T = 1000$. For nonstationary data, $\widehat{\Gamma}(k)$ is much farther from $\Gamma_T(k)$. For example, in DGP 2(a) $\widehat{\Gamma}(0) = 2.507$ and $\Gamma_T(0) = 2.571$ whereas in DGP 3(a) $\widehat{\Gamma}(0) = 1.589$ and $\Gamma_T(0) = 1.861$. Thus, $\widehat{\Gamma}(k)$ has larger bias (in general downward) when the data are nonstationary. This result is present even when $T = 200$. As T increases, $\widehat{\Gamma}(k)$ approaches $\Gamma_T(k)$ for all DGPs, though the downward bias remains larger in DGP 3-4 than in DGP 1-2.

We repeated this exercise for other DGPs and the conclusions were the same. The results suggest that under nonstationarity the bias in the LRV estimator is affected by multiple factors. In addition to the downward bias arising from strong persistence which is also present under stationarity there is bias generated by the time-varying properties of the process. Under the null hypothesis this time variation occurs in the autocovariance structure of the process. For example, in

TABLE 1. Average estimates of $\widehat{J}_{\mathrm{HAC}}, \widehat{c}_1$ and $\widehat{\Gamma}(k)$, $k = 0, 1, 5, 10$.

$T = 100$

DGP	J	$\widehat{J}_{\mathrm{HAC}}$	$-\bar{c}_1 b_{1,T}$	$-\widehat{c}_1 b_{1,T}$	$\Gamma_T(0)$	$\widehat{\Gamma}(0)$	$\Gamma_T(1)$	$\widehat{\Gamma}(1)$	$\Gamma_T(5)$	$\widehat{\Gamma}(5)$	$\Gamma_T(10)$	$\widehat{\Gamma}(10)$
1(a)	1.826	1.483	−0.138	−0.169	1.072	1.062	0.279	0.273	0.001	0.002	0.000	0.000
1(b)	1.826	1.499	−0.138	−0.165	1.072	1.072	0.279	0.276	0.001	0.001	0.000	0.000
1(c)	1.826	1.549	−0.138	−0.160	1.072	1.105	0.279	0.285	0.001	0.001	0.000	0.000
2(a)	20.988	6.755	−0.980	−2.685	2.571	2.507	2.009	1.940	0.751	0.696	0.219	0.195
2(b)	20.988	6.830	−0.980	−2.617	2.571	2.533	2.009	1.961	0.751	0.702	0.219	0.195
2(c)	20.988	7.038	−0.980	−2.622	2.571	2.609	2.009	2.019	0.751	0.725	0.219	0.206
3(a)	20.988	3.142	−2.260	−40.480	1.861	1.589	1.028	0.736	0.622	0.312	0.367	0.100
3(b)	20.988	3.301	−2.260	−38.312	1.861	1.635	1.028	0.781	0.622	0.338	0.367	0.113
3(c)	20.988	3.761	−2.260	−35.695	1.861	1.790	1.028	0.920	0.622	0.427	0.367	0.161
4(a)	20.988	3.437	−2.260	−37.756	1.861	1.670	1.028	0.829	0.622	0.373	0.367	0.133
4(b)	20.988	3.448	−2.260	−37.145	1.861	1.680	1.028	0.830	0.622	0.373	0.367	0.134
4(c)	20.988	3.472	−2.260	−35.472	1.861	1.711	1.028	0.834	0.622	0.373	0.367	0.134

$T = 200$

DGP	J	$\widehat{J}_{\mathrm{HAC}}$	$-\bar{c}_1 b_{1,T}$	$-\widehat{c}_1 b_{1,T}$	$\Gamma_T(0)$	$\widehat{\Gamma}(0)$	$\Gamma_T(1)$	$\widehat{\Gamma}(1)$	$\Gamma_T(5)$	$\widehat{\Gamma}(5)$	$\Gamma_T(10)$	$\widehat{\Gamma}(10)$
1(a)	1.826	1.569	−0.110	−0.127	1.072	1.067	0.279	0.276	0.001	0.001	0.000	0.000
1(b)	1.826	1.577	−0.110	−0.128	1.072	1.071	0.279	0.277	0.001	0.001	0.000	0.000
1(c)	1.826	1.602	−0.110	−0.124	1.072	1.089	0.279	0.281	0.001	0.001	0.000	0.000
2(a)	20.988	8.388	−0.784	−1.862	2.571	2.539	2.009	1.975	0.751	0.722	0.219	0.207
2(b)	20.988	8.449	−0.784	−1.839	2.571	2.553	2.009	1.988	0.751	0.728	0.219	0.207
2(c)	20.988	8.555	−0.784	−1.821	2.571	2.588	2.009	2.013	0.751	0.737	0.219	0.211
3(a)	20.988	4.354	−1.808	−30.914	1.861	1.723	1.028	0.883	0.622	0.465	0.367	0.229
3(b)	20.988	4.459	−1.808	−30.284	1.861	1.749	1.028	0.903	0.622	0.479	0.367	0.237
3(c)	20.988	4.771	−1.808	−30.321	1.861	1.823	1.028	0.978	0.622	0.526	0.367	0.265
4(a)	20.988	4.548	−1.808	−28.901	1.861	1.766	1.028	0.929	0.622	0.496	0.367	0.247
4(b)	20.988	4.552	−1.808	−29.944	1.861	1.770	1.028	0.931	0.622	0.496	0.367	0.248
4(c)	20.988	4.569	−1.808	−29.132	1.861	1.786	1.028	0.932	0.622	0.499	0.367	0.248

$T = 1000$

DGP	J	$\widehat{J}_{\mathrm{HAC}}$	$-\bar{c}_1 b_{1,T}$	$-\widehat{c}_1 b_{1,T}$	$\Gamma_T(0)$	$\widehat{\Gamma}(0)$	$\Gamma_T(1)$	$\widehat{\Gamma}(1)$	$\Gamma_T(5)$	$\widehat{\Gamma}(5)$	$\Gamma_T(10)$	$\widehat{\Gamma}(10)$
1(a)	1.826	1.667	−0.079	−0.088	1.072	1.071	0.279	0.278	0.001	0.001	0.000	0.000
1(b)	1.826	1.669	−0.079	−0.087	1.072	1.073	0.279	0.279	0.001	0.000	0.000	0.000
1(c)	1.826	1.673	−0.079	−0.087	1.072	1.076	0.279	0.279	0.001	0.002	0.000	0.000
2(a)	20.988	10.904	−0.560	−1.097	2.571	2.565	2.009	2.003	0.751	0.743	0.219	0.216
2(b)	20.988	10.934	−0.560	−1.084	2.571	2.571	2.009	2.008	0.751	0.749	0.219	0.219
2(c)	20.988	10.935	−0.560	−1.084	2.571	2.574	2.009	2.009	0.751	0.746	0.219	0.217
3(a)	20.988	6.510	−1.291	−20.845	1.861	1.834	1.028	1.001	0.622	0.592	0.367	0.339
3(b)	20.988	6.541	−1.291	−20.449	1.861	1.841	1.028	1.001	0.622	0.595	0.367	0.343
3(c)	20.988	6.629	−1.291	−20.475	1.861	1.857	1.028	1.021	0.622	0.605	0.367	0.349
4(a)	20.988	6.543	−1.291	−20.854	1.861	1.840	1.028	0.838	0.622	0.595	0.367	0.344
4(b)	20.988	6.555	−1.291	−20.361	1.861	1.843	1.028	1.009	0.622	0.598	0.367	0.347
4(c)	20.988	6.559	−1.291	−20.551	1.861	1.846	1.028	1.011	0.622	0.598	0.367	0.347

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

DGP 3, one has $0.2T$ observations to estimate $2\pi \int_0^{0.2} f(u, 0)\, du = 0.4\pi f(0)$ where $f(0) = 1/(2\pi(1-2\rho+\rho^2))$ with $\rho = 0.9$, and $0.8T$ observations to estimate $2\pi \int_{0.2}^{1} f(u, 0)\, du = 1.6\pi f(0)$ where $f(0) = 1/(2\pi(1-2\rho+\rho^2))$ with $\rho = 0.1$. This is more difficult than estimating $2\pi f(0) = 1/(2\pi(1-2\rho+\rho^2))$ with $\rho = 0.7817$ using T observations, which applies to DGP 2. Even if the total sample size is T in both DGP 2 and 3, nonstationarity reduces the effective sample size making the estimation of the LRV in DGP 3 effectively based on a smaller number of observations. For example, $\widehat{\Gamma}(k)$ involves an average on $\{\widehat{V}_t\widehat{V}_{t-k}\}$ for $t = k+1, \ldots, T$. Some of these pairs $\{\widehat{V}_t\widehat{V}_{t-k}\}$ are such that $\widehat{V}_t$ and $\widehat{V}_{t-k}$ belong to two different regimes, and so contribute bias to the estimation of $\Gamma_T(k)$. Under stationarity all the pairs $\{\widehat{V}_t\widehat{V}_{t-k}\}$ are such that $\widehat{V}_t$ and $\widehat{V}_{t-k}$ belong to the same regime leading to more precise estimates of $\widehat{\Gamma}(k)$ and LRV. In addition, changes in persistence over short regimes share features similar to shifts in the mean, at least graphically. While the former is consistent with the null hypothesis, the latter is not. This is likely to generate some bias where changes in persistence are confounded with shifts in the mean even when the unconditional mean of the series has not changed. The downward bias due to strong persistence and the bias due to time-varying second-order properties are likely to influence each other making the estimation problem even harder.

We now investigate the consequence of nonstationarity for HAR inference. We obtain the empirical size and power for a two-tailed 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. Model M1 involves an SLS process: $V_t = 0.9V_{t-1} + u_t$, $V_0 \sim \mathscr{N}(0, 1)$, $u_t \sim$ 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 $\lambda_1^0 = 0.1$. 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 $V = (V_t)_{t=1}^T$;[4] M4 involves a time varying AR(1) with periods of strong persistence where $V_t = \rho(t/T)V_{t-1} + u_t$, $\rho(t/T) = 0.95(\cos(1.5t/T))$, $u_t \sim$ 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 (2023) and Casini and Perron (2024b), respectively; Andrews' (1991) HAC estimator with and without the prewhitening procedure of Andrews and Monahan (1992); Newey and West's (1987)

[4] In this literature, values smaller than $\underline{c}$ are not classified as outliers.

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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 recommended 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 (2023) and Casini and Perron (2024b), 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 $\widehat{b}_{1,T} = 0.6828(\widehat{\phi}(2)\,T\widehat{\widehat{b}}_{2,T})^{-1/5}$ 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

$$\widehat{a}_1(u) = \frac{\sum_{j=t-n_T+1}^{t}\widehat{V}_j\widehat{V}_{j-1}}{\sum_{j=t-n_T+1}^{t}(\widehat{V}_{j-1})^2}, \qquad \text{and} \qquad \widehat{\sigma}(u) = (\sum_{j=t-n_T+1}^{t}(\widehat{V}_j-\widehat{a}_1(u)\widehat{V}_{j-1})^2)^{1/2},$$

and $\widehat{\widehat{b}}_{2,T} = (n_T/T)\sum_{r=1}^{\lfloor T/n_T\rfloor-1}\widehat{b}_{2,T}(rn_T/T)$, $\widehat{b}_{2,T}(u) = 1.6786(\widehat{D}_1(u))^{-1/5}(\widehat{D}_2(u))^{1/5}T^{-1/5}$ where $\widehat{D}_2(u) \triangleq 2\sum_{l=-\lfloor T^{4/25}\rfloor}^{\lfloor T^{4/25}\rfloor}\widehat{c}_{\mathrm{DK},T}(u,l)^2$ and

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

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 reports the results using 5,000 replications. The t-test based on Newey and West's (1987) and Andrews' (1991) prewhitened HAC estimators are excessively oversized. Andrews' (1991) HAC-based test is slightly undersized while the KVB's fixed-b and EWC-based tests are severely undersized. The fact that the KVB's fixed-b and EWC-based tests have larger size distortions than other tests is 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. 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 having more accurate rejection rates. Nonstationarity affects the power of the tests

TABLE 2. Empirical small-sample null rejection rates and power of t-test for model M1-M4.

M1					
$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.25$	$\delta = 1.5$
$\widehat{J}_{\mathrm{DK},T}$	0.068	0.189	0.286	0.661	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.045	0.085	0.199	0.612	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.046	0.090	0.202	0.613	1.000
Andrews (1991)	0.039	0.095	0.185	0.623	0.999
Andrews (1991), prewhite	0.115	0.168	0.304	0.650	0.999
Newey-West (1987)	0.209	0.272	0.398	0.689	1.000
KVB fixed-b	0.004	0.018	0.063	0.301	0.969
EWC	0.011	0.038	0.137	0.539	0.999

M2					
$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.3$	$\delta = 1$
$\widehat{J}_{\mathrm{DK},T}$	0.080	0.132	0.257	0.842	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.059	0.098	0.190	0.736	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.055	0.088	0.187	0.735	1.000
Andrews (1991)	0.081	0.133	0.266	0.838	1.000
Andrews (1991), prewhite	0.094	0.141	0.268	0.842	1.000
Newey-West (1987)	0.137	0.190	0.336	0.881	1.000
KVB fixed-b	0.014	0.036	0.078	0.561	0.990
EWC	0.032	0.064	0.157	0.712	1.000

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

M4					
$\alpha = 0.05$, $T = 200$	$\delta = 0$ (null rejection)	$\delta = 0.1$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 3$
$\widehat{J}_{\mathrm{DK},T}$	0.154	0.146	0.496	0.706	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.037	0.050	0.168	0.459	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.041	0.079	0.198	0.477	1.000
Andrews (1991)	0.127	0.162	0.398	0.623	0.999
Andrews (1991), prewhite	0.197	0.226	0.439	0.653	1.000
Newey-West (1987)	0.397	0.423	0.584	0.758	1.000
KVB fixed-b	0.005	0.012	0.135	0.339	0.964
EWC	0.115	0.147	0.367	0.681	0.999

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

based on LRV estimators that rely on $\widehat{\Gamma}(k)$ or equivalently on $I_T(\omega)$ (e.g., the EWC). The KVB's fixed-b and EWC-based tests 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 power, the second best being Andrews' (1991) HAC-based test.

Turning to M2, Table 2 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 2 shows that all methods lead to oversized tests except prewhitened DK-HAC and KVB's fixed-b. However, the KVB's fixed-b-based tests show substantial unde-rejection that has consequences for power whereas the prewhitened DK-HAC-based-tests show accurate null rejection rates and good power. 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 their respective ERP are close as shown in Section 4.

5.3. General Low Frequency Contamination

We now discuss HAR inference tests for which the low frequency contamination results of Section 3 hold 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 hypothesis. This class is very large and includes 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 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 hypothesis 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 with respect to 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 hypothesis 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 $t \in [1, \dots, 3T/4 - 1, 3T/4 + 21, \dots T]$ and $x_t^{(2)} = \delta(t/T) + 0.2z_t + 2u_{X_2,t}$ for $t = 3T/4, \dots, 3T/4 + 20$ 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 $t \in [1, \dots, T/2 - 30, T/2 + 21, \dots T]$ and $x_t^{(2)} = \delta(t/T) + 0.2z_t + 2u_{X_2,t}$ for $t = T/2 - 30, \dots, T/2 + 20$ 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 $x_t^{(2)} \sim \text{Uniform}\left(|\underline{c}|, 5|\underline{c}|\right)$ for $t = 6T/10, 8T/10$ where $\underline{c} = -1/(\sqrt{2}\text{erfc}^{-1}(3/2))\text{med}(|x^{(2)} - \text{med}(x^{(2)})|)$ 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 3 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 tests 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 1 (top panel) reports the plots of d_t, its sample autocovariances and its periodogram, for $\delta = 1$. Figures S.1 and S.2 (top panels) 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 1 (top panel) shows that $\widehat{\Gamma}(k)$ decays slowly. As δ increases, from Figures S.1 and S.2 (top 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 hypothesis.

TABLE 3. Empirical small-sample null rejection rates 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) $\delta\,(t/T)$ segmented 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

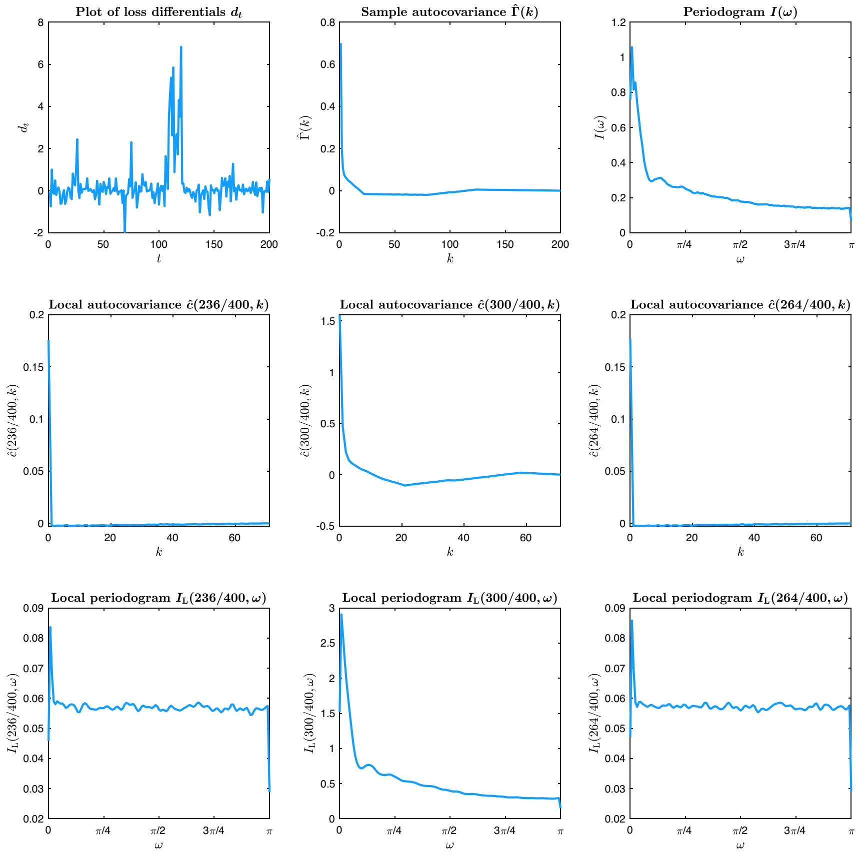

FIGURE 1. Plots of loss differentials d_t, sample autocovariance $\widehat{\Gamma}(k)$, periodogram $I(\omega)$, sample local autocovariance $\widehat{c}(u, k)$ and local periodogram $I_{\mathrm{L}}(u, \omega)$. In all panels $\delta = 1$.

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.7 in Theorem S.1. The overall behavior of the sample autocovariance is as predicted by Theorem S.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.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.

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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 Theorems 1–2. We set $n_{2,T} = T^{0.6} = 36$ following the MSE criterion of Casini (2023). 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 Theorems 1 and 2. We consider $\delta = 1$, 2 and 5. According to Theorems 1 and 2, we should expect long memory features only for case (ii-a). Figures 1 and S.1–S.2 in the supplement confirm this. The results pertaining to case (ii-a) are plotted in the middle panels. They 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 bottom panels in Figures 1 and S.1–S.2 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 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 Theorems 1 and 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 for case (ii-a). 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 $\mathbb{P}_\delta(|t_{\mathrm{DM},i}| > z_{1-\alpha/2})$ where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal for a two-sided test with significance level $\alpha \in (0, 1)$. 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

https://doi.org/10.1017/S026646662400032X Published online by Cambridge University Press

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 7. *Let* $\{d_t - \mathbb{E}(d_t)\}_{t=1}^{T_n}$ *be an SLS process satisfying Assumptions 1-(i-iv) and 2. Let Assumptions 6–7 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 9,* $\mathbb{P}_\delta(|t_{\mathrm{DM,NW87}}| > z_\alpha) \to 0$. *If Assumption 9 is replaced by Assumption 10 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 11,* $|t_{\mathrm{DM,DK}}| = \delta^2 O_{\mathbb{P}}(T_n^\zeta)$ *and* $\mathbb{P}_\delta(|t_{\mathrm{DM,DK}}| > z_\alpha) \to 1$.

Note that Assumption 10 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 7 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.8). Part (iii) of the theorem suggests that these issues do not occur when the 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 3 discussed above, thereby confirming the relevance of Theorem 7.

6. CONCLUSIONS

Economic time series often display nonstationary features that are usefully addressed in testing by allowing for some misspecification in standard model formulations. If nonstationarity 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 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. Under the null hypothesis there are larger size distortions than when the data are stationary. Under the alternative hypothesis, existing LRV estimators tend to be inflated and HAR tests can exhibit dramatic power losses. 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

Casini, A., T. Deng and P. Perron (2024): Supplement to "Theory of low frequency contamination from nonstationarity and misspecification: consequences for HAR inference", Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S026646662400032X

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, 821–856.

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. (Eds.). (1975). *Normal approximation and asymptotic expansion*. Wiley.

Brillinger, D. (1975). *Time series data analysis and theory*. 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. Preprint, arXiv:1803.10883.

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

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

Casini, A. (2024). The fixed-b limiting distribution and the ERP of HAR tests under nonstationarity. *Journal of Econometrics*, 238, 105625.

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. (2021). Continuous record Laplace-based inference about the break date in structural change models. *Journal of Econometrics*, 224, 3–21.

Casini, A., & Perron, P. (2022a). Continuous record asymptotics for change-point models. Preprint, arXiv:1803.10881.

Casini, A., & Perron, P. (2022b). Generalized Laplace inference in multiple change-points models. *Econometric Theory*, 38, 35–65.

Casini, A., & Perron, P. (2024a). Change-point analysis of time series with evolutionary spectra. *Journal of Econometrics*, 242, 105811.

Casini, A., & Perron, P. (2024b). Prewhitened long-run variance estimation robust to nonstationarity. *Journal of Econometrics*, 242, 105794.

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.

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

Demetrescu, M., & Salish, N. (2024). (Structural) VAR models with ignored changes in mean and volatility. *International Journal of Forecasting*, 40, 840–854.

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, 253–263.

Dou, L. (2024). Optimal HAR inference. *Quantitative Economics*, forthcoming.

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. Preprint, arXiv:2105.05335.

Ibragimov, R., & Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & 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.

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. (2020). 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 Economics and Statistics*, 86, 378–390.

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.

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

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, 1120–1154.

Otto, S., & Breitung, J. (2023). Backward CUSUM for testing and monitoring structural change. *Econometric Theory*, 39, 659–692.

Percival, D. (1992). Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean. *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. (1991). *A test for changes in a polynomial trend function for a dynamic time series.* Research Memorandum No. 363. Econometrics Research Program, Princeton University.

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.