Simultaneous bandwidths determination for DK-HAC estimators and long-run variance estimation in nonparametric settings

Federico Belotti, Alessandro Casini, Leopoldo Catania, Stefano Grassi, and Pierre Perron

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

1. Introduction

Long-run variance (LRV) estimation has a long history in econometrics and statistics since it plays a key role for heteroscedasticity and autocorrelation robust (HAR) inference. The classical approach in HAR inference relies on consistent estimation of the LRV. Newey and West (1987) and Andrews (1991) proposed kernel heteroscedasticity and autocorrelation consistent (HAC) estimators and showed their consistency. However, recent work by Casini (2022c) showed that, in both the linear regression model and other contexts, their results do not provide accurate approximations in that test statistics normalized by classical HAC estimators may exhibit size distortions and substantial power losses. Issues with the power have been shown for a variety of HAR testing problems outside the regression model (e.g., Altissimo & Corradi, 2003; Casini, 2018; Casini & Perron, 2019, 2022a, 2021c; Chan, 2022; Chang & Perron, 2018; Crainiceanu & Vogelsang, 2007; Deng & Perron, 2006; Juhl & Xiao, 2009; Kim & Perron, 2009; Martins & Perron, 2016; Perron &

CONTACT Alessandro Casini alessandro.casini@uniroma2.it Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, Italy.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07474938.2023.2178138.
We thank Zhongjun Qu and Daniel Whilelm as well as two referees for useful comments.
© 2023 Taylor & Francis Group, LLC
Yamamoto, 2021; Vogelsang, 1999). Casini et al. (2022) showed theoretically that such power issues are generated by low-frequency contamination induced by nonstationarity. More specifically, nonstationarity biases upward each sample autocovariance. Thus, LRV estimators are inflated and HAR test statistics lose power. These issues can also be provoked by misspecification, nonstationary alternative hypotheses and outliers. They also showed that LRV estimators that rely on fixed-\(b\) or versions thereof suffer more from these problems than classical HAC estimators since the former use a larger number of sample autocovariances.\(^1\) Finally, Casini (2022b) showed that current fixed-\(b\) methods are not theoretically valid in general because under nonstationarity the asymptotic distribution of HAR tests is not pivotal. He showed that the error in rejection probability (ERP) associated to fixed-\(b\) HAR tests is an order of magnitude larger than that under stationarity and is also larger than that of HAR tests based on HAC estimators.

In order to flexibly account for nonstationarity, Casini (2022c) introduced a double kernel HAC (DK-HAC) estimator that applies kernel smoothing over two directions. In addition to the usual smoothing over lagged autocovariances used in classical HAC estimators, the DK-HAC estimator uses a second kernel that applies smoothing over time. The latter accounts for time variation in the covariance structure of time series which is a relevant feature in economics and finance. Since the DK-HAC uses two kernels and bandwidths, one cannot rely on the theory of Andrews (1991) or Newey and West (1994) for selecting the bandwidths. Casini (2022c) considered a sequential MSE criterion that determines the optimal bandwidth controlling the number of lags as a function of the optimal bandwidth controlling the smoothing over time. Thus, the latter influences the former but not vice-versa. However, each smoothing affects the bias-variance tradeoff so that the two bandwidths should affect each other’s optimal value. Consequently, it is useful to consider an alternative criterion to select the bandwidths. In this paper, we consider simultaneous bandwidths determination obtained by jointly minimizing the global MSE of the DK-HAC estimator. We obtain the asymptotic optimal formula for the two bandwidths and use the plug-in approach to replace unknown quantities by consistent estimates. Our results are established under the nonstationary framework characterized by segmented locally stationary processes [cf. Casini, 2022c]. The latter extends the locally stationary framework of Dahlhaus (1997) to allow for discontinuities in the spectrum. Thus, the class of segmented locally stationary processes includes structural break models (see, e.g., Bai & Perron, 1998; Casini & Perron, 2021b), time-varying parameter models [see e.g., Cai (2007)] and regime switching (cf. Hamilton, 1989).

We establish the consistency, rate of convergence and asymptotic MSE results for the DK-HAC estimators with data-dependent simultaneous bandwidths. The optimal bandwidths have the same order \(O(\frac{T}{C^1_{0.6}})\) whereas under the sequential criterion the optimal bandwidths smoothing over time has an order \(O(\frac{T}{C^1_{0.5}})\) and the optimal bandwidth smoothing the lagged autocovariances has an order \(O(\frac{T}{C^4_{25}})\). Thus, asymptotically, the joint (or global) MSE criterion implies the use of (marginally) more lagged autocovariances and a longer segment length for the smoothing over time relative to the sequential criterion. Hence, the former should control more accurately the variance due to nonstationarity while the latter should control better the bias. If the degree of nonstationarity is high then the theory suggests that one should expect the sequential criterion to perform marginally better. The difference in the smoothing over lags is very minor between the order of the corresponding bandwidths implied by the two criteria. Our simulation analysis supports this view as we show that the joint MSE criterion performs better especially when the degree of nonstationarity is not too high.

---

\(^1\)The fixed-\(b\) literature is extensive. Pioneering contribution of Kiefer et al. (2000) and Kiefer and Vogelsang (2002; 2005) introduced the fixed-\(b\) LRV estimators. Additional contributions can be found in Dou (2019), Lazarus et al. (2020), Lazarus et al. (2018), Gonçalves and Vogelsang (2011), de Jong and Davidson (2000), Ibragimov and Müller (2010), Jansson (2004), Müller (2007, 2014), Phillips (2005), Politis (2011), Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018; 2019), Robinson (1998), Sun (2013; 2014a; 2014c), Sun et al. (2008), Velasco and Robinson (2001) and Zhang and Shao (2013). We refer to Casini (2019, 2022c) and Casini et al. (2022) for discussions and comparisons between our approach based on DK-HAC and the fixed-\(b\) approach.
Overall, we find that HAR tests normalized by DK-HAC estimators strike the best balance between empirical size and power among the existing LRV estimators and we also find that using the bandwidths selected from the joint MSE criterion yields tests that perform better than the sequential criterion in terms of size control. The optimal rate $O(T^{-1/6})$ is also found by Neumann and von Sachs (1997) and Dahlhaus (2012) in the context of local spectral density estimates under local stationarity. Under both sequential and joint MSE criterion the optimal kernels are found to be the same, i.e., the quadratic spectral kernel for smoothing over autocovariance lags (similar to Andrews, 1991) and a parabolic kernel (cf. Epanechnikov, 1969) for smoothing over time.

Another contribution of the paper is developed asymptotic results for consistent LRV estimation in nonparametric parameter estimation settings. Newey and West (1987) and Andrews (1991) established the consistency of HAC estimators for the LRV of some series $\{V_t(\hat{\beta})\}$ where $\hat{\beta}$ is a semiparametric estimator of $\beta_0$ having the usual parametric rate of convergence $\sqrt{T}$ [i.e., they assumed that $\sqrt{T}(\hat{\beta} - \beta_0) = O_p(1)$]. For example, in the linear regression model estimated by least-squares, $V_t(\hat{\beta}) = \hat{e}_t x_t$, where $\{\hat{e}_t\}$ are the least-squares residuals and $\{x_t\}$ is a vector of regressors. Unfortunately, the condition $\sqrt{T}(\hat{\beta} - \beta_0) = O_p(1)$ does not hold for nonparametric estimators $\hat{\beta}_{np}$ since they satisfy $T^{\theta}(\hat{\beta}_{np} - \beta_0) = O_p(1)$ for some $\theta \in (0, 1/2)$. For example, for tests for forecast evaluation often forecasters use nonparametric kernel methods to obtain the forecasts [i.e., $\{V_t(\hat{\beta})\} = L(e_t(\hat{\beta}_{np}))$ where $L(\cdot)$ is a forecast loss, $e_t(\cdot)$ is a forecast error and $\hat{\beta}_{np}$ is, e.g., a rolling window estimate of a parameter that is used to construct the forecasts]. Given the widespread use of nonparametric methods in applied work, it is useful to extend the theoretical results of HAC and DK-HAC estimators for these settings. We establish the validity of HAC and DK-HAC estimators including their versions based on data-dependent bandwidths.

The DK-HAC estimators can result in HAR tests that are oversized when there is strong dependence in the data, a well-known problem for all methods, though for ours these distortions are relatively minor compared to, e.g., the methods of Newey and West (1987) and Andrews (1991). Still, in order to improve the size control of HAR tests, Casini and Perron (2022b) proposed a nonparametric nonlinear VAR prewhitened DK-HAC estimators. This form of prewhitening differs from those discussed previously (e.g., Andrews and Monahan (1992), Rho and Shao (2013) and Xiao and Linton (2002)) in that it accounts explicitly for nonstationarity. As shown in Casini and Perron (2022b), HAR tests based on prewhitened DK-HAC estimators have size control competitive to fixed-$b$ HAR tests even with strong dependence when the latter work well (i.e., under stationarity). Thus, we also apply the prewhitening to the DK-HAC estimators with simultaneous bandwidths.

The remainder of the article is organized as follows. Section 2 introduces the statistical setting and the joint MSE criterion. Section 3 presents consistency, rates of convergence, asymptotic MSE results, and optimal kernels and bandwidths for the DK-HAC estimators using the joint MSE criterion. Section 4 develops a data-dependent method for simultaneous bandwidth parameters selection and its asymptotic properties are then discussed. Section 5 presents theoretical results for LRV estimation in nonparametric parameter estimation. Section 6 presents Monte Carlo results about the small-sample size and power of HAR tests based on the DK-HAC estimators using the proposed automatic simultaneous bandwidths. We also provide comparisons with a variety of other approaches. Section 7 presents an empirical application and Section 8 concludes the article. The supplemental material (Belotti et al., 2022) contains the mathematical proofs and additional simulations.

2. The statistical environment

We consider the estimation of the LRV $J \triangleq \lim_{T \to \infty} J_T$ where $J_T = T^{-1} \sum_{t=1}^T \sum_{\beta \in \Theta} \mathbb{E}(V_t(\beta_0) V_t(\beta_0)^\prime)$ with $V_t(\beta)$ being a random $p$-vector for each $\beta \in \Theta$. For example, for the linear model $V_t(\beta) = (y_t - x_t^\prime \beta)x_t$. The classical approach for inference in the context of serially correlated data is based on
consistent estimation of $J$. Newey and West (1987) and Andrews (1991) considered the class of kernel HAC estimators, where the subscript \text{Cla} stands for classical,

$$
\hat{J}_{\text{Cla}, T} = \hat{J}_{\text{Cla}, T}(b_1, T) \triangleq \frac{T}{T - p} \sum_{k=-T+1}^{T-1} K_1(b_1, T k) \hat{\Gamma}_{\text{Cla}}(k),
$$

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

where \( \hat{V}_t = V_t(\hat{\beta}) \), \( K_1(\cdot) \) is a real-valued kernel in the class \( K_1 \) defined below and \( b_1, T \) is a bandwidth sequence. The factor \( T/(T - p) \) is an optional small-sample degrees of freedom adjustment. For the Newey–West estimator \( K_1 \) corresponds to the Bartlett kernel while for Andrews’ (1991) \( K_1 \) corresponds to the quadratic spectral (QS) kernel. Data-dependent methods for the selection of \( b_1, T \) were proposed by Newey and West (1994) and Andrews (1991), respectively. Under appropriate conditions on \( b_1, T \to 0 \) they showed that \( \hat{J}_{\text{Cla}, T} \xrightarrow{P} J \). When \( \{V_t\} \) is second-order stationary, \( J = 2\pi f(0) \) where \( f(0) \) is the spectral density of \( \{V_t\} \) at frequency zero. Most of the LRV estimation literature has focused on the stationarity assumption for \( \{V_t\} \) (e.g., Kiefer et al. 2000; Müller, 2007; Lazarus et al., 2020). Unlike the HAC estimators, fixed-\( b \) (and versions thereof) LRV estimators require stationarity of \( \{V_t\} \). The latter assumption is restrictive for economic and financial time series. The properties of \( J \) under nonstationarity were studied recently by Casini (2022c) who showed that if \( \{V_t\} \) is either locally stationary or segmented locally stationary (SLS), then \( J = 2\pi \int_0^1 f(u, 0) du \) where \( f(u, 0) \) is the time-varying spectral density at rescaled time \( u = t/T \) and frequency zero. For locally stationary processes, \( f(u, 0) \) is smooth in \( u \) while for SLS processes \( f(u, 0) \) can in addition contain a finite-number of discontinuities. The number of discontinuities can actually grow to infinity with unchanged results though at the expense of slightly more complex derivations. Since the assumption of a finite number of discontinuities captures well the idea that a finite number of regimes or structural breaks is enough to account for structural changes (or big events) in economic time series we maintain this assumption here. The latter is relaxed by Casini and Perron (2022b).

Under nonstationarity Casini (2022c) argued that an extension of the classical HAC estimators can actually account flexibly for the time-varying properties of the data. He proposed the class of double kernel HAC (DK-HAC) estimators,

$$
\hat{J}_T = \hat{J}_T(b_1, T, b_2, T) \triangleq \frac{T}{T - p} \sum_{k=-T+1}^{T-1} K_1(b_1, T k) \hat{\Gamma}_T(k),
$$

with

$$
\hat{\Gamma}_T(k) \triangleq \frac{n_T}{T - n_T} \sum_{r=0}^{[T-n_T]/n_T} \hat{c}_T(r n_T/T, k),
$$

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

$$
\hat{c}_T(r n_T/T, k) \triangleq \begin{cases} 
(T b_2, T)^{-1} \sum_{s=k+1}^{T} K_2^s \left( (r+1) n_T - (s-k)/2 \right) / b_{2, T} \hat{V}_s \hat{V}_{s-k}' & k \geq 0 \\
(T b_2, T)^{-1} \sum_{s=-k+1}^{s} K_2^s \left( (r+1) n_T - (s+k)/2 \right) / b_{2, T} \hat{V}_{s+k} \hat{V}_{s}' & k < 0,
\end{cases}
\)

with \( K_2^s \) being a real-valued kernel and \( b_{2, T} \) is a bandwidth sequence. \( \hat{c}_T(u, k) \) is an estimate of the local autocovariance \( c(u, k) = \mathbb{E}(V_{[u]_T}, V_{[u+k]_T}) + O(T^{-1}) \) [under regularity conditions; see Casini, 2022c] at lag \( k \) and time \( u = r n_T/T \). Estimation of \( c(u, k) \) for locally stationary processes was considered by Dahlhaus (2012). \( \hat{\Gamma}_T(k) \) estimates the local autocovariance across blocks of length \( n_T \) and then takes an average over the blocks. The estimator \( \hat{J}_T \) involves two kernels: \( K_1 \)
smoothes the lagged autocovariances—akin to the classical HAC estimators—while $K_2$ applies smoothing over time. The smoothing over time better account for nonstationarity and makes $J_{DK,T}$ robust to low frequency contamination. See Casini et al. (2022) who showed theoretically that existing LRV estimators are contaminated by nonstationarity so that they become inflated with consequent large power losses when the estimators are used to normalize HAR test statistics. Dahlhaus (2012) discussed how to estimate $f(u, \omega)$ for the scalar case under smoothness in both arguments using the smoothed local periodogram. Our goals are to estimate $J$ using a time-domain method and to relax the smoothness assumption in $u$.

Casini (2022c) considered adaptive estimators $J_{DK,T}$ for which $b_{1,T}$ and $b_{2,T}$ are data-dependent. Observe that the optimal $b_{2,T}$ actually depends on the properties of $\{V_i,T\}$ in any given block [i.e., $b_{2,T} = b_{2,T}(t/T)$]. Let

$$\text{MSE}(b_{2,T}^{-4}, \hat{c}_T(u_0, k), \hat{W}_T) = b_{2,T}^{-4}\text{E}[\text{vec}(\hat{c}_T(u_0, k) - c(u_0, k))\hat{W}_T[\text{vec}(\hat{c}_T(u_0, k) - c(u_0, k))],$$

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

$$\text{MSE}\left(Tb_{1,T}^{opt}, \tilde{b}_{2,T}^{opt}, \hat{J}_T^{opt}, \hat{W}_T\right) \leq \text{MSE}\left(Tb_{1,T}^{opt}, \hat{b}_{2,T}^{opt}, \hat{J}_T^{opt}, \tilde{b}_{2,T}^{opt}, \hat{W}_T\right)$$

where $\hat{b}_{2,T}^{opt} = \int_0^1 b_{2,T}^{opt}(u)du$

and $b_{2,T}^{opt}(u) = \arg\min_{b_{2,T}} \text{MSE}\left(b_{2,T}^{-4}, \hat{c}_T(u_0, k) - c(u_0, k), \hat{W}_T\right).$

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

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

$$\lim_{T \to \infty} \text{ReMSE}\left(Tb_{1,T}b_{2,T}, \tilde{J}_T(b_{1,T}, b_{2,T})J_{1}^{-1}, \hat{W}_T\right) = \lim_{T \to \infty} \text{MSE}\left(Tb_{1,T}b_{2,T}^{opt}, \tilde{J}_T^{opt}(b_{1,T}, b_{2,T}^{opt})J_{1}^{-1}, \hat{W}_T\right),$$

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

This is a more reasonable property. In Section 3 we solve for the sequences $(\tilde{b}_{1,T}^{opt}, \tilde{b}_{2,T}^{opt})$ that minimize (2.3). We propose a data-dependent method for $(\tilde{b}_{1,T}^{opt}, \tilde{b}_{2,T}^{opt})$ in Section 4. Besides Andrews (1991) and Newey and West (1987) in the context of LRV estimation, the MSE-optimality criterion was also used more recently by Whilem (2015) in a GMM context to determine the optimal bandwidth of the nonparametric estimator of the optimal weighting matrix.\footnote{Note that the MSE bounds in Section 8 in Andrews (1991) are not correct. See Casini (2022a) for details.}
an important case, the recent increasing use of nonparametric methods suggests that the case
where \( \hat{\beta} \) enjoys a nonparametric rate of convergence slower than \( \sqrt{T} \) is of potential interest. Hence, in Section 5 we consider consistent LRV estimation under the latter framework and develop corresponding results for the classical HAC as well as the DK-HAC estimators.

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

\[
K_1 = \left\{ K_1(\cdot) : \mathbb{R} \to [-1, 1] : K_1(0) = 1, \ K_1(x) = K_1(-x), \ \forall x \in \mathbb{R} \right\}
\]

\[
\int_{-\infty}^{\infty} K_1^2(x)dx < \infty, \ \text{\( K_1(\cdot) \) is continuous at 0 and at all but finite numbers of points} \right\}.
\]

(2.4)

\[
K_2 = \left\{ K_2(\cdot) : \mathbb{R} \to [0, \infty] : K_2(x) = K_2(1 - x), \ \int K_2(x)dx = 1, \ \right. \]

\[
K_2(x) = 0, \ \text{for } x \notin [0, 1], \ \text{\( K_2(\cdot) \) is continuous} \right\}.
\]

(2.5)

The class \( K_1 \) was also considered by Andrews (1991). Examples of kernels in \( K_1 \) include the Truncated, Bartlett, Parzen, Quadratic Spectral (QS) and Tukey-Hanning kernels. The QS kernel was shown to be optimal for \( J_{\text{Cla}}^{(p)} \) under the MSE criterion by Andrews (1991) and for \( J_T^{(p)} \) under a sequential MSE criterion by Casini (2022c),

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

The class \( K_2 \) was also considered by, for example, Dahlhaus and Giraitis (1998).

Throughout we adopt the following notational conventions. The \( j \)th element of a vector \( x \) is indicated by \( x^{(j)} \) while the \( (j, l) \)th element of a matrix \( X \) is indicated as \( X^{(j,l)} \). \( \text{tr}(\cdot) \) denotes the trace function and \( \otimes \) denotes the tensor (or Kronecker) product operator. The \( p^2 \times p^2 \) matrix \( C_{pp} \) is a commutation matrix that transforms \( \text{vec}(A) \) into \( \text{vec}(A') \), i.e., \( C_{pp} = \sum_{j=1}^{p} \sum_{l=1}^{p} t_j t_l' \otimes t_l t_j' \), where \( t_j \) is the \( j \)th elementary \( p \)-vector. \( \lambda_{\text{max}}(A) \) denotes the largest eigenvalue of the matrix \( A \). \( W \) and \( \tilde{W} \) are used for \( p^2 \times p^2 \) weight matrices. \( \mathbb{C} \) is used for the set of complex numbers and \( \bar{A} \) for the complex conjugate of \( A \in \mathbb{C} \). Let \( 0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m < \lambda_{m+1} = 1 \). A function \( G(\cdot, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C} \) is said to be piecewise (Lipschitz) continuous with \( m + 1 \) segments if it is (Lipschitz) continuous within each segment. For example, it is piecewise Lipschitz continuous if for each segment \( j = 1, \ldots, m + 1 \) it satisfies \( \sup_{u \neq v} |G(u, \omega) - G(v, \omega)| \leq K|u - v| \) for any \( \omega \in \mathbb{R} \) with \( \lambda_{j-1} < u, v \leq \lambda_j \) for some \( K < \infty \). We define \( G_j(u, \omega) = G(u, \omega) \) for \( \lambda_{j-1} < u \leq \lambda_j \), so \( G_j(u, \omega) \) is Lipschitz continuous for each \( j \). A function \( G(\cdot, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C} \) is said to be left-differentiable at \( u_0 \) if \( \partial G(u_0, \omega) / \partial_-. u \) exists for any \( \omega \in \mathbb{R} \). We use \( [\cdot] \) to denote the largest smaller integer function. The symbol \( \check{\ldots} \) is for definitional equivalence.

### 3. Simultaneous bandwidths for DK-HAC estimators

In Section 3.1 we present the consistency, rate of convergence and asymptotic MSE properties of predetermined bandwidths for the DK-HAC estimators. We use the MSE results to determine the optimal bandwidths and kernels in Section 3.2. We use the framework for nonstationarity introduced in Casini (2022c). That is, we assume that \( \{V_i\}_{i=1}^{T} \) is segmented locally stationary (SLS).

Suppose \( \{V_i\}_{i=1}^{T} \) is defined on an abstract probability space \((\Omega, \mathcal{F}, \mathbb{P})\) where \( \Omega \) is the sample space, \( \mathcal{F} \) is the \( \sigma \)-algebra and \( \mathbb{P} \) is a probability measure. We use an infill asymptotic setting and
rescale the original discrete time horizon \([1, T]\) by dividing each \(t\) by \(T\). Letting \(u = t/T\) and \(T \to \infty\), this defines a new time scale \(u \in [0, 1]\). Let \(i = \sqrt{-1}\).

**Definition 3.1.** 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 \(\mu\) if there exists a representation

\[
V_{t,T} = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(iot)A^0_{j,t,T}(\omega)d\zeta(\omega), \quad (t = T_{j-1}^0 + 1, \ldots, T_j^0),
\]

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

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

\[
\text{cum}\{d\zeta(\omega_1), \ldots, d\zeta(\omega_r)\} = \varphi \left( \sum_{j=1}^{r} \omega_j \right) g_r(\omega_1, \ldots, \omega_{r-1})d\omega_1 \ldots d\omega_r,
\]

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

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

\[
A(u, \omega) = A_j(u, \omega) \quad \text{for} \quad \lambda_{j-1}^0 \leq u \leq \lambda_j^0,
\]

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

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

Observe that this representation is similar to the spectral representation of stationary processes (see Anderson, 1971; Brillinger, 1975; Hannan, 1970; Priestley, 1981) for introductory concepts). The main difference is that \(A(t/T, \omega)\) and \(\mu(t/T)\) are not constant in \(t\). Dahlhaus (1997) used the time-varying spectral representation to define the so-called locally stationary processes which are characterized, broadly speaking, by smoothness conditions on \(\mu(\cdot)\) and \(A(\cdot, \cdot)\). Locally stationary processes are often referred to as time-varying parameter processes (see e.g., Cai, 2007; Chen & Hong, 2012). However, the smoothness restrictions exclude many prominent models that account for time variation in the parameters. For example, structural change and regime switching-type models do not belong to this class because parameter changes occur suddenly at a particular time. We have explicitly assumed a finite number of breaks for practical reasons as mentioned above. In between any two breaks, local stationarity captures features such as transitory dynamics or changes that take some time to take effect. Thus, the class of SLS processes is more general and likely to be more useful. Stationarity and local stationarity are recovered as special cases of the SLS definition.\(^3\)

Let \(\mathcal{T} = \{T_0^0, \ldots, T_{m_0}^0\}\). The spectrum of \(V_{t,T}\) is defined (for fixed \(T\)) as follows:

\[
\hat{f}_{j,T}(u, \omega) \triangleq \begin{cases} 
(2\pi)^{-1} \sum_{s=\infty}^{\infty} \text{Cov}(V_{[uT-s|/2],T}, V_{[uT-s|/2],T}) \exp(-i\omega s), & Tu \in \mathcal{T}, \ u = T_j^0/T \\
(2\pi)^{-1} \sum_{s=\infty}^{\infty} \text{Cov}(V_{[uT-s/2],T}, V_{[uT+s/2],T}) \exp(-i\omega s), & Tu \notin \mathcal{T}, T_{j-1}^0/T < u < T_j^0/T
\end{cases}
\]

\(^3\)Some authors have used alternative notions of local stationarity that allow for discontinuities and have established some results in other contexts (cf. Dahlhaus (2009) and Zhou (2013)).
Let $\hat{J}_T$ denote the pseudo-estimator identical to $J_T$ but based on $\{V_{t,T}\} = \{V_{t,T}(\beta_0)\}$ rather than on $\{\hat{V}_{t,T}\} = \{V_{t,T}(\hat{\beta})\}$.

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

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

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

Following Parzen (1957), we define $K_{1,q} \triangleq \lim_{x \to 0} (1 - K_1(x))/|x|^q$ for $q \in [0, \infty)$; $q$ increases with the smoothness of $K_1(\cdot)$ with the largest value being such that $K_{1,q} < \infty$. When $q$ is an even integer, $K_{1,q} = -(d^2K_1(x)/dx^2)|_{x=0}/q!$ and $K_{1,q} < \infty$ if and only if $K_1(x)$ is $q$ times differentiable at zero. We define the index of smoothness of $f(u, \omega)$ at $\omega = 0$ by $f^{(q)}(u, 0) \triangleq (2\pi)^{-1/2} \sum_{k = -\infty}^{\infty} |k|^q c(u, k)$, with $A_{t,l,T}(\omega) = A_t(0, \omega)$ for $t < 1$ and $A_{m+1,l,T}(\omega) = A_{m+1}(1, \omega)$ for $t > T$. Casini (2022c) showed that $f_j(u, \omega)$ tends in mean-squared to $f_j(u, \omega) \triangleq |A_j(u, \omega)|^2$ for $T_{j+1}/T < u = t/T \leq T_j/T$, which is the spectrum that corresponds to the spectral representation. Therefore, we call $f_j(u, \omega)$ the time-varying spectral density matrix of the process. Given $f(u, \omega)$, we can define the local covariance of $V_{t,T}$ at rescaled time $u$ with $Tu \notin \mathcal{T}$ and lag $k \in \mathbb{Z}$ as $c(u, k) \triangleq \int_{-\infty}^{\infty} e^{iku} f(u, \omega) \, d\omega$. 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 $c(u, k) \triangleq \lim_{T \to -\infty} \int_{-\infty}^{\infty} e^{iku} A(u, \omega) A(u - k/T, -\omega) \, d\omega$.  

### 3.1. Asymptotic MSE properties of DK-HAC Estimators

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

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

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

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

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

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

Let $\check{C}$ denote the set of continuity points of $f(u, \omega)$ in $u$, i.e., $\check{C} = \{0, 1\}/\{x_j^q, j = 1, \ldots, m_0\}$. Define

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

where

$$\frac{\partial}{\partial u^-} f(x_j^0, \omega) = \lim_{h \downarrow 0} \frac{f(x_j^0 + h, \omega) - f(x_j^0, \omega)}{h}, \quad \frac{\partial}{\partial u^+} f(x_j^0, \omega) = \lim_{h \downarrow 0} \frac{f(x_j^0 + h, \omega) - f(x_j^0, \omega)}{h}.$$

**Theorem 3.1.** Suppose $K_1(\cdot) \in K_1$, $K_2(\cdot) \in K_2$, Assumption 3.1-3.2 hold, $b_{1,T}$, $b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ and $1/Tb_{1,T}b_{2,T} \to 0$. We have: (i) \[ \lim_{T \to \infty} Tb_{1,T}b_{2,T}\text{Var}\left[\text{vec}(\bar{J}_T)\right] = 4\pi^2 \int K_1^2(y)dy \int_0^1 K_2^2(x)dx (I + C_{pp}) \left( \int_0^1 f(u, 0)du \right) \otimes \left( \int_0^1 f(v, 0)dv \right). \]

(ii) If $1/Tb_{1,T}b_{2,T} \to 0$, $n_T/Tb_{1,T}^2 \to 0$ and $b_{2,T}^2/b_{1,T}^q \to \nu \in (0, \infty)$ for some $q \in [0, \infty)$ then $\lim_{T \to \infty} b_{1,T}^{-q}E(\bar{J}_T - J_T) = B_1 + B_2$ where $B_1 = -2\pi K_{1,q} \int_0^1 f^{(q)}(u, 0)du$ and $B_2 = 2^{-1}\nu \int_0^1 x^2 K_2(x) \int_{-\infty}^x (\partial^2/\partial u^2) \jmath(x, k)du + 2\nu \Delta_f(0)$.

(iii) If $n_T/Tb_{1,T} \to 0$, $b_{2,T}^2/b_{1,T}^q \to \nu$ and $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0, \infty)$ for some $q \in [0, \infty)$ for which $K_{1,q}$, $\int_0^1 f^{(q)}(u, 0)du \in (0, \infty)$ then $\lim_{T \to \infty} \text{MSE}(Tb_{1,T}b_{2,T}, \bar{J}_T, W) = 4\pi^2 \left[ \gamma(4\pi^2)^{-1}\text{vec}(B_1 + B_2)'W\text{vec}(B_1 + B_2) \right. \left. + \int K_1^2(y)dy \int K_2^2(x)dx \text{tr}W(L_{p^2} + C_{pp}) \left( \int_0^1 f(u, 0)du \right) \otimes \left( \int_0^1 f(v, 0)dv \right) \right].$ \]

The bias expression in part (ii) of Theorem 3.1 is different from the corresponding one in Casini (2022c) because $b_{2,T}^2/b_{1,T}^q \to \nu \in (0, \infty)$ replaces $b_{2,T}^2/b_{1,T}^q \to 0$ there. The extra term is $B_2$. This means that both $b_{1,T}$ and $b_{2,T}$ affect the bias as well as the variance. It is therefore possible to consider a joint minimization of the asymptotic MSE with respect to $b_{1,T}$ and $b_{2,T}$. Note that $B_2 = 0$ when $\int_{-\infty}^x (\partial^2/\partial u^2) \jmath(x, k)du = 0$ and $m_0 = 0$. The latter two conditions occur when the process is stationary. We now move to the results concerning $\bar{J}_T$.

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

Assumption 3.3(i)–(iii) is the same as Assumption B in Andrews (1991). Part (i) is satisfied by standard (semi)parametric estimators. In Section 5 we relax this assumption and consider
nonparametric estimators that satisfy $T^4(\hat{\beta} - \beta_0) = O_p(1)$ where $\vartheta \in (0, 1/2)$. In order to obtain rate of convergence results we replace Assumption 3.2 with the following assumptions.

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

$$
\left( V'_{[Tu]}, \text{vec}\left( \frac{\partial}{\partial \theta} V_{[Tu]}(\beta_0) \right) - \mathbb{E}\left( \frac{\partial}{\partial \theta} V_{[Tu]}(\beta_0) \right) \right)'.
$$

(ii) $\sup_{\omega \in [0, 1]} \mathbb{E}(\sup_{\beta \in \Theta} \left| \left( \frac{\partial^2}{\partial \theta \partial \theta} V_{[Tu]}(\beta) \right)^2 \right|) < \infty$ for all $a = 1, \ldots, p$.

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

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

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

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

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

$$
\lim_{T \to \infty} \text{MSE}(Tb_{1,T}b_{2,T}, J_T, W_T) = \lim_{T \to \infty} \text{MSE}(Tb_{1,T}b_{2,T}, \hat{J}_T, W).
$$

Part (ii) yields the consistency of $\hat{J}_T$ with $b_{1,T}$ only required to be $o(T b_{2,T})$. This rate is slower than the corresponding rate $o(T)$ of the classical kernel HAC estimators as shown by Andrews (1991) in his Theorem 1-(b). However, this property is of little practical import because optimal growth rates typically are less than $T^{1/2}$—for the QS kernel the optimal growth rate is $T^{1/3}$ while it is $T^{1/3}$ for the Bartlett. Part (ii) of the theorem presents the rate of convergence of $\hat{J}_T$ which is $\sqrt{Tb_{2,T} b_{1,T}}$, the same rate shown by Casini (2022c) when $b_{2,T}/b_{1,T}^q \to 0$. Thus, the presence of the bias term $B_2$ does not alter the rate of convergence. In Section 3.2, we compare the rate of convergence of $\hat{J}_T$ with optimal bandwidths $(\hat{b}_{1,T}^{\text{opt}}, \hat{b}_{2,T}^{\text{opt}})$ from the joint MSE criterion (2.3) with that using $(\hat{b}_{1,T}^{\text{opt}}, \hat{b}_{2,T}^{\text{opt}})$ from the sequential MSE criterion (2.2), and with that of the classical HAC estimators when the corresponding optimal bandwidths are used.

### 3.2. Optimal bandwidths and kernels

Under $m_0 = 0$ we consider the optimal bandwidths $(\hat{b}_{1,T}^{\text{opt}}, \hat{b}_{2,T}^{\text{opt}})$ and kernels $\hat{K}_1^{\text{opt}}$ and $\hat{K}_2^{\text{opt}}$ that minimize the global asymptotic relative MSE (2.3) given by

$$
\lim_{T \to \infty} \text{ReMSE}(\hat{J}_T(b_{1,T}, b_{2,T}) J^{-1}, W_T),
$$

$$
= \mathbb{E}\left( \text{vec}(\hat{J}_T J^{-1} - I_p)' W_T \text{vec}(\hat{J}_T J^{-1} - I_p) \right).
$$

Let $\Xi_{1,1} = -K_{1,q}$, $\Xi_{1,2} = (4\pi)^{-1} \int_0^1 x^2 K_2(x) dx$, $\Xi_2 = \int K_2^2(y) dy \int_0^1 K_2^2(x) dx,$
\[ \Delta_{1,0} = \int_0^1 f(u, 0) du \left( \int_0^1 f(u, 0) du \right)^{-1} \]
\[ \Delta_{1,2} = \sum_{k=-\infty}^{\infty} \int_0^1 \left( \frac{\partial^2}{\partial u^2} \right) c(u, k) du \left( \int_0^1 f(u, 0) du \right)^{-1}. \]

For the determination of the optimal \( K(\cdot) \) we need to restrict attention to a subset of \( K_1 \). Let \( \tilde{K}_1 = \{ K(\cdot) \in K_1 | \tilde{K}(\omega) \geq 0 \ \forall \ \omega \in \mathbb{R} \} \) where \( \tilde{K}(\omega) = (2\pi)^{-1} \int_\mathbb{R} K(x)e^{-i\omega x} dx \). The function \( \tilde{K}(\omega) \) is referred to as the spectral window generator. The set \( \tilde{K}_1 \) contains all kernels \( K \) that necessarily generate positive semidefinite estimators in finite samples.

**Theorem 3.3.** Suppose \( K(\cdot) \in K_1, K(\cdot) \in K_2 \), Assumption 3.1, 3.3–3.5 hold, \( m_0 = 0, \int_0^1 ||f^{(2)}(u, 0)||du < \infty, \text{vec}(\Delta_{1,0})W \text{vec}(\Delta_{1,0}) > 0, \text{vec}(\Delta_{1,2})W \text{vec}(\Delta_{1,2}) > 0 \) and \( W \) is positive definite. Then, \( \lim_{T \to \infty} \text{ReMSE} \left( \hat{f}_T(b_{1,T}, b_{2,T})^{-1}, W_T \right) \) is jointly minimized by

\[ \tilde{b}_{opt}^{1,T} = 0.46 \left( \frac{\text{vec}(\Delta_{1,2})W \text{vec}(\Delta_{1,2})}{(\text{vec}(\Delta_{1,0})W \text{vec}(\Delta_{1,0}))^5} \right)^{1/24} T^{-1/6}, \]
\[ \tilde{b}_{opt}^{2,T} = 3.56 \left( \frac{\text{vec}(\Delta_{1,0})W \text{vec}(\Delta_{1,0})}{(\text{vec}(\Delta_{1,2})W \text{vec}(\Delta_{1,2}))^5} \right)^{1/24} T^{-1/6}. \]

Furthermore, the optimal kernels are given by \( K_{1, opt}^{1} = K_{1, opt}^{2} \) and \( K_{2, opt}(x) = 6x(1-x) \) for \( x \in [0, 1] \), where \( K_{1, opt} \) is optimal among the kernels in \( \tilde{K}_1 \).

The requirement \( \int_0^1 ||f^{(2)}(u, 0)||du < \infty \) is not stringent and reduces to the one used by Andrews (1991) when \( \{ V_t \} \) is stationary. Note that \( \Delta_{1,1,0} \) accounts for the relative variation of \( \int_0^1 f(u, \omega) du \) around \( \omega = 0 \) whereas \( \Delta_{1,2} \) accounts for the relative time variation (i.e., nonstationarity). The theorem states that as \( \Delta_{1,1,0} \) increases \( \tilde{b}_{opt}^{1,T} \) becomes smaller while \( \tilde{b}_{opt}^{2,T} \) becomes larger. This is intuitive. With more variation around the zero frequency, more smoothing is required over the frequency direction and less over the time direction. Conversely, the more nonstationary is the data the more smoothing is required over the time direction (i.e., \( \tilde{b}_{opt}^{2,T} \) is smaller and the optimal block length \( T\tilde{b}_{opt}^{2,T} \) smaller) relative to the frequency direction. Both optimal bandwidths \( (\tilde{b}_{opt}^{1,T}, \tilde{b}_{opt}^{2,T}) \) have the same order \( O(T^{-1/6}) \). We can compare it with \( b_{opt}^{1,T} = O(T^{-4/25}) \) and \( \tilde{b}_{opt}^{2,T} = O(T^{-1/5}) \) resulting from the sequential MSE criterion in Casini (2022c). The latter leads to a slightly smaller block length relative to the joint criterion (2.3) [i.e., \( O(T\tilde{b}_{opt}^{2,T}) < O(T\tilde{b}_{opt}^{2,T}) \)]. Since \( K_2 \) applies overlapping smoothing, a smaller block length is beneficial if there is substantial nonstationarity. On the same note, a smaller block length is less exposed to low frequency contamination since it allows to better account for nonstationarity. The rate of convergence when the optimal bandwidths are used is \( O(T^{1/3}) \) which is slightly faster than the corresponding rate of convergence with \( (b_{opt}^{1,T}, \tilde{b}_{opt}^{2,T}) \). The latter is \( O(T^{0.32}) \), so the difference is small.

### 4. Data-dependent bandwidths

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

4.1. Implementation

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

$$\phi_1 \triangleq \frac{\text{vec}(\Delta_{1,2})'W\text{vec}(\Delta_{1,2})}{(\text{vec}(\Delta_{1,1,0})'W\text{vec}(\Delta_{1,1,0}))^3}, \quad \phi_2 \triangleq \frac{\text{vec}(\Delta_{1,1,0})'W\text{vec}(\Delta_{1,1,0})}{(\text{vec}(\Delta_{1,2})'W\text{vec}(\Delta_{1,2}))^3}.$$ 

In a third step, we replace the unknown parameters in $\phi_1$ and $\phi_2$ with corresponding estimates. Such estimates $\hat{\phi}_1$ and $\hat{\phi}_2$ are then substituted into the expression for $\hat{b}_{1,T}^{opt}$ and $\hat{b}_{2,T}^{opt}$ to yield

$$\hat{b}_{1,T} = 0.46\hat{\phi}_1^{1/24}T^{-1/6}, \quad \hat{b}_{2,T} = 3.56\hat{\phi}_2^{1/24}T^{-1/6}. \quad (4.1)$$

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

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

$$\phi_{1,1} = \sum_{r=1}^{p} \sum_{k=-\infty}^{\infty} \left[ \int_{0}^{1} \left( \int_{0}^{1} f^{(r,r)}(u, k)du \right)^2 \left( \int_{0}^{1} f^{(r,r)}(u, k)du \right)^2 \right],$$

$$\phi_{1,2} = \sum_{r=1}^{p} \sum_{k=-\infty}^{\infty} \left[ \int_{0}^{1} f^{(r,r)}(u, k)du \right]^2 \left( \int_{0}^{1} f^{(r,r)}(u, k)du \right)^2.$$ 

The usual choice is $W^{(r,r)} = 1$ for all $r$. An estimate of $f^{(r,r)}(u, k) = (2\pi)^{-1}(\hat{\sigma}(r)(u))^2(1-\hat{\alpha}(r)(u))^{-2}$ while $f^{(2)(r,r)}(u, 0)$ can be estimated by $\hat{f}^{(2)(r,r)}(u, 0) = 3\pi^{-1}(\hat{\sigma}(r)(u))^2(\hat{\alpha}(r)(u))^{-4}$ where $\hat{\alpha}(r)(u)$ and $\hat{\sigma}(r)(u)$ are the LS estimates computed using local data to the left of $u = t/T$:

$$\hat{\alpha}(r)(u) = \frac{\sum_{j=t-n_2,T+1}^{t} \hat{V}_j(u) \hat{V}_{j-1}(u)}{\sum_{j=t-n_2,T+1}^{t} \hat{V}_j^2(u)}, \quad \hat{\sigma}(r)(u) = \left( \sum_{j=t-n_2,T+1}^{t} \left( \hat{V}_j - \hat{\alpha}(r)(u)\hat{V}_{j-1}(u) \right)^2 \right)^{1/2}, \quad (4.2)$$

where $n_{2,T} \to \infty$. More complex is the estimation of $\hat{\Delta}_{1,2,1} \triangleq \sum_{k=-\infty}^{\infty} \int_{0}^{1} (\partial^2/\partial u^2)c(u, k)du$ because it involves the second partial derivative of $c(u, k)$. We need a further parametric assumption. We assume that the parameters of the approximating time-varying AR(1) models change slowly such that the smoothness of $f(\cdot, \cdot)$ and thus of $c(\cdot, \cdot)$ is the same to the one that would arise if $a_1(u) = 0.8(\cos 1.5 + \cos 4\pi u)$ and $\sigma(u) = \sigma = 1$ for all $u \in [0, 1]$ [cf. Dahlhaus (2012)]. Then, $\Delta^{(r,r)}(u, k) \triangleq (\partial^2/\partial u^2)c^{(r,r)}(u, k)$ can be computed analytically:
\[ \Delta_{r,1,2}(u, k) = \int_{-\pi}^{\pi} e^{ik\omega} \left[ \frac{3}{\pi} \left( 1 + 0.8 (\cos 1.5 + \cos 4\pi u) \exp(-i\omega) \right)^{-4} (0.8(-4\pi \sin(4\pi u)) \exp(-i\omega) \right] \frac{1}{\pi} [1 + 0.8(\cos 1.5 + \cos 4\pi u) \exp(-i\omega)]^{-3} (0.8(-16\pi^2 \cos(4\pi u)) \exp(-i\omega)] d\omega. \]

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

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

where \( |S_{\omega}| \) is the cardinality of \( S_{\omega} \) and \( \omega_{k+1} > \omega_k \) with \( \omega_1 = -\pi, \omega_{|S_{\omega}|} = \pi \). In our simulations, we use \( S_{\omega} = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\} \). We can average over \( u \) and sum over \( k \) to obtain an estimate of \( \hat{\Delta}_{1,2,1} : \hat{\Delta}_{1,2,1} = \sum_{k=-|S_{\omega}|}^{\max(1,2)} \frac{n_T}{T} \sum_{j=0}^{\lfloor T/|S_{\omega}| \rfloor} \hat{\Delta}_{1,2,1}(j_T/n_T, k) \) where the number of summands over \( k \) grows at the same rate as \( 1/b_{1,T}^{\text{opt}} \); a different choice is allowed as long as it grows at a slower rate than \( T^{2/5} \) but our sensitivity analysis does not indicate significant changes.

Then, \( \hat{b}_{1,T} = 0.46 \hat{b}_{1,T}^{1/24} T^{-1/6} \) and \( \hat{b}_{2,T} = 3.56 \hat{b}_{2,T}^{1/24} T^{-1/6} \) where \( \hat{b}_1 = \hat{b}_{1,1}/\hat{b}_{1,2}, \hat{b}_2 = \hat{b}_{1,2}/\hat{b}_{1,1}, \)

\[ \hat{b}_{1,1} = (4\pi)^{-2} \sum_{r=1}^{\rho} W^{(r,r)} \left( \hat{\Delta}_{1,2,1} \right)^2 / \left( n_T^{T/n_T} \sum_{j=0}^{\lfloor T/n_T \rfloor} (\hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^2 (1 - \hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^{-2} \right)^2, \]

\[ \hat{b}_{1,2} = 36 \sum_{r=1}^{\rho} W^{(r,r)} \left( \frac{n_T \sum_{j=0}^{\lfloor T/n_T \rfloor} (\hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^2 (1 - \hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^{-2} \right)^2 / \left( n_T^{T/n_T} \sum_{j=0}^{\lfloor T/n_T \rfloor} (\hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^2 (1 - \hat{\Delta}_{1,2,1}(j_T/n_T) + 1)^{-2} \right). \]

For most of the results below we can take \( n_T = n_{2,T} = n_T \).

### 4.2. Theoretical results

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

\[ K_3 = \{ K_3(\cdot) \in K_1 : (i) |K_1(x)| \leq C_1|x|^{-b} \text{ with } b > \max(1 + 1/q, 3) \text{ for } |x| \in [\bar{x}_L, D_T h_T \bar{x}_U], \]

\[ b_{1,T}^{1/7} h_T \rightarrow \infty, D_T > 0, \bar{x}_L, \bar{x}_U \in \mathbb{R}, 1 \leq \bar{x}_L < \bar{x}_U, \text{ and} \]

with \( b > 1 + 1/q \) for \( |x| \not\in [\bar{x}_L, D_T h_T \bar{x}_U] \), and some \( C_1 < \infty \),

where \( q \in (0, \infty) \) is such that \( K_1, \hat{K}_1 \in (0, \infty) \), (ii) \( |K_1(x) - K_1(y)| \leq C_2|x - y| \forall x, y \in \mathbb{R} \) for some constant \( C_2 < \infty \).

Let \( \hat{\theta} \) denote the estimator of the parameter of the approximate (time-varying) parametric model(s) introduced above [i.e., \( \hat{\theta} = (\int_0^1 \hat{a}_1(u) du, \int_0^1 \hat{\sigma}_1^2(u) du, \cdots, \int_0^1 \hat{a}_p^2(u) du, \int_0^1 \hat{\sigma}_p^2(u) du)' \)]. Let \( \theta^{*} \) denote the probability limit of \( \hat{\theta} \). \( \hat{\phi}_1 \) and \( \hat{\phi}_2 \) are the values of \( \phi_1 \) and \( \phi_2 \), receptively, with \( \hat{\theta} \) instead of \( \theta \). The probability limits of \( \hat{\phi}_1 \) and \( \hat{\phi}_2 \) are denoted by \( \phi_{1,\theta^{*}} \) and \( \phi_{2,\theta^{*}} \), respectively.
Assumption 4.1. (i) \( \hat{\phi}_1 = O_p(1) \), \( 1/\hat{\phi}_1 = O_p(1) \), \( \hat{\phi}_2 = O_p(1) \), and \( 1/\hat{\phi}_2 = O_p(1) \); (ii) \( \inf \{ T/n_3, T \} \sqrt{\| n_2 \|^2} \{(\hat{\phi}_1 - \phi_{1,0^*}), (\hat{\phi}_2 - \phi_{2,0^*})\}' = O_p(1) \) for some \( \hat{\phi}_{1,0^*}, \hat{\phi}_{2,0^*} \in (0, \infty) \) where \( n_{2,T} + n_{3,T} \to 0 \), \( n_{2,T}/T \to \sigma_2, n_{3,T}/T \to \sigma_3 \), \( n_{3,T}/T \to \sigma_3, n_{3,T}/T \to \sigma_3 \) with \( 0 < \sigma_2, \sigma_3 < \infty \); (iii) \( \sup_{u \in [0,1]} l_{\max}(\Gamma_u(k)) \leq C_3 k^{-l} \) for all \( k \geq 0 \) for some \( C_3 < \infty \) and some \( l > 3 \), where \( q \) is as in \( K_q \); (iv) \( |\omega_{x+1} - \omega_{x}| = O(T^{-1}) \) and \( |S_0| = O(T) \); (v) \( K_2 \) includes kernels that satisfy \( |K_2(x) - K_2(y)| \leq C_4 |x - y| \) for all \( x, y \in \mathbb{R} \) and some constant \( C_4 < \infty \).

Parts (i) and (ii) are the nonparametric analogue to Assumptions E & F, respectively, in Andrews (1991). Part (iii) is satisfied if \( \{ V_t \} \) is strong mixing with mixing numbers that are less stringent than those sufficient for the cumulant condition in Assumption 3.2(i). Part (iv) is needed to apply the convergence of Riemann sums. Part (v) requires \( K_2 \) to satisfy Lipschitz continuity. Parts (i) and (v) are sufficient for the consistency of \( \hat{J}_T(b_{1,T}, \hat{b}_{2,T}) \). Parts (ii)-(iii) and (iv)-(v) are required for the rate of convergence and MSE results. Note that \( \hat{\phi}_{1,0^*} \) and \( \hat{\phi}_{2,0^*} \) coincide with the optimal values \( \phi_1 \) and \( \phi_2 \), respectively, only when the approximate parametric model indexed by \( \theta^* \) corresponds to the true data-generating mechanism.

Let \( b_{0_1,T} = 0.46\phi_{1,0^*}^{1/24}T^{-1/6} \) and \( b_{0_2,T} = 3.56\phi_{2,0^*}^{1/24}T^{-1/6} \). The asymptotic properties of \( \hat{J}_T(b_{1,T}, \hat{b}_{2,T}) \) are shown to be equivalent to those of \( \hat{J}_T(b_{0_1,T}, b_{0_2,T}) \) where the theoretical properties of the latter follow from Theorem 3.2.

**Theorem 4.1.** Suppose \( K_1(\cdot) \in K_3 \), \( q \) is as in \( K_3 \), \( K_2(\cdot) \in K_2 \), \( n_T \to \infty \), \( n_T/T b_{0_1,T} \to 0 \), and \( \| f^q(u, 0) du \| < \infty \). Then, we have:

(i) If Assumption 3.1-3.3 and 4.1-(i,v) hold and \( n_{3,T} = n_{2,T} = n_T \), then \( \hat{J}_T(b_{1,T}, \hat{b}_{2,T}) \to 0 \).

(ii) If Assumption 3.1, 3.3-3.4 and 4.1-(ii,iii,iv,v) hold \( q \leq 2 \) and \( n_T/T b_{0_1,T}^2 \to 0 \), then \( \sqrt{T b_{0_1,T}^2} (\hat{J}_T(b_{1,T}, \hat{b}_{2,T}) - J_T) = O_p(1) \) and \( \sqrt{T b_{0_1,T}^2} (\hat{J}_T(b_{1,T}, \hat{b}_{2,T}) - J_T(b_{0_1,T}, b_{0_2,T})) = O_p(1) \).

(iii) If Assumption 3.1, 3.3-3.5 and 4.1-(ii,iii,iv,v) hold and \( m_0 = 0 \), then

\[
\lim_{T \to \infty} \text{MSE}(\hat{J}_T(b_{1,T}, \hat{b}_{2,T}, W_T)) = \lim_{T \to \infty} \text{MSE}((b_{0_1,T} b_{0_2,T}, \hat{J}_T(b_{0_1,T}, b_{0_2,T}, W_T))
\]

When the chosen parametric model indexed by \( \theta \) is correct, it follows that \( \phi_{1,0^*} = \phi_1, \phi_{2,0^*} = \phi_2, \hat{\phi}_1 \to \phi_1 \) and \( \hat{\phi}_2 \to \phi_2 \). The theorem then implies that \( \hat{J}_T(b_{1,T}, \hat{b}_{2,T}) \) exhibits the same optimality properties presented in Theorem 3.3.

5. Consistent LRV in the context of nonparametric parameter estimates

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

5.1. Classical HAC estimators

We show that the classical HAC estimators that use the data-dependent bandwidths suggested in Andrews (1991) remain valid when \( \sqrt{T} (\hat{\beta} - \beta_0) = O_P(1) \) is replaced by \( T^\vartheta (\hat{\beta}_{np} - \beta_0) = O_P(1) \) for some \( \vartheta \in (0, 1/2) \). We work under the same assumptions as in Andrews (1991). Under stationarity we have \( \Gamma_u(k) = \Gamma(k) \) and \( \kappa_{V_t, T_u} (k, s, l) = \kappa_{V, 0} (k, s, l) \) for any \( u \in [0, 1] \).

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

Assumption 5.2. (i) \( \sup_{t \geq 1} \mathbb{E} \mathbb{E} \mathbb{E} [\mathbb{E} |V_t|^2 < \infty ; (ii) \sup_{t \geq 1} \mathbb{E}[(\partial / \partial \beta) V_t(\beta)]^2 < \infty ; (iv) \} K(y)dy < \infty \).

Assumption 5.3. (i) Assumption 5.1 holds with \( V_t \) replaced by
\[
(V_t', \text{ vec} \left( \left( \frac{\partial}{\partial \beta} V_t(\beta_0) - \mathbb{E} \left( \frac{\partial}{\partial \beta} V_t(\beta_0) \right) \right) \right),)
\]

(ii) \( \sup_{t \geq 1} \mathbb{E} \mathbb{E} \mathbb{E} [\mathbb{E} |(\partial^2 / \partial \beta \partial \beta') V_t(a) \beta^2] < \infty \) for all \( a = 1, ..., p \).

Let \( \hat{b}_{t, K} = (q K_t^2 q^{-2/2q-1} T / \int K_t^2(x) dx) \). The form of \( \hat{q}(q) \) depends on the approximating parametric model for \( \{V_t(r)\} \). Andrews (1991) considered stationarity AR(1) models for \( \{V_t(r)\} \), which result in
\[
\hat{q}(2) = \sum_{r=1}^{p} W(r, r, r) 4 \left( \hat{a}(r)^{2} \right) \left( \hat{a}(r)^{4} \right) / \sum_{r=1}^{p} W(r, r, r) \left( \hat{a}(r)^{4} \right)
\]

and
\[
\hat{q}(1) = \sum_{r=1}^{p} W(r, r, r) 4 \left( \hat{a}(r)^{2} \right) \left( \hat{a}(r)^{4} \right) / \sum_{r=1}^{p} W(r, r, r) \left( \hat{a}(r)^{4} \right)
\]

Let
\[
K_{t, 2} = \{K_t(\cdot) \in K_t : \text{ (i) } |K_t(y)| \leq C_1 |y|^{-b} \text{ for some } b > 1 + 1/q \text{ and some } C_1 < \infty, \text{ where } q \in (0, \infty) \text{ is such that } K_t(1, q) \in (0, \infty), \text{ and (ii) } |K_t(1) - K_t(y)| \leq C_2 |x - y| \forall x, y \in \mathbb{R} \text{ for some constant } C_2 < \infty \}.
\]

Assumption 5.4. \( \hat{q}(q) = O_P(1) \) and \( 1/ \hat{q}(q) = O_P(1) \).

Assumption 5.4 corresponds to Assumption E in Andrews (1991).

Theorem 5.1. Suppose \( K_t(\cdot) \in K_{3, 1} \), \( ||f^{(q)}|| < \infty, q > (1/ \vartheta - 1)/2 \), and Assumption 5.1-5.4 hold, then \( \int_{C_{\text{clt}}} T \hat{b}_{t, T} - J_T \to 0 \).
5.2. Dk-HAC Estimators

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

**Assumption 5.5.** (i) \( T^{\varphi}(\hat{\beta}_{np} - \beta_0) = O_p(1) \) for some \( \varphi \in (0, 1/2) \); (ii)–(iv) from Assumption 3.3 continue to hold.

**Theorem 5.2.** Suppose \( K_1(\cdot) \in K_3, \ K_2(\cdot) \in K_2, \ T^{\varphi}b_{\theta_1, T}b_{\theta_2, T} \rightarrow \infty, \ n_T \rightarrow \infty, \ n_T/Tb_{\theta_1, T} \rightarrow 0, \) and \( ||\int_0^1 f^{(q)}(u, 0)du|| < \infty. \) If Assumption 3.1-3.2, 4.1-(i,v), 5.5 hold and \( n_{3,T} = n_{2,T} = n_T, \) then \( \hat{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T}) - J_T \xrightarrow{p} 0. \)

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

6. Small-sample evaluations

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

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

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

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

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

- **M1:** \( e_t = 0.4e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 0.5), \ x_t \sim \text{i.i.d. } \mathcal{N}(1, 1), \ \beta_0^{(1)} = 0 \) and \( \beta_0^{(2)} = 1. \)
- **M2:** \( e_t = 0.4e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \ x_t \sim \text{i.i.d. } \mathcal{N}(1, 1), \) and \( \beta_0^{(1)} = \beta_0^{(2)} = 0. \)
- **M3:** segmented locally stationary errors \( e_t = \rho_t e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \) \( \rho_t = \max\{0, -1 \ (\cos(1.5 - \cos(5t/T)))\}^4 \) for \( t \not\in \langle 4T/5 + 1, \ 4T/5 + h \rangle \) and \( e_t = 0.99e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \) for \( t \in \langle 4T/5 + 1, \ 4T/5 + h \rangle \) where \( h = 10 \) for \( T = 200 \) and \( h = 30 \) for \( T = 400, \) and \( x_t = 1 + 0.6x_{t-1} + u_{X,t}, \ u_{X,t} \sim \text{i.i.d. } \mathcal{N}(0, 1). \)

Finally, we consider model M4 which we use to investigate the performance of Rossi’s (2009) test for forecast breakdown. Suppose we want to forecast a variable \( y_t \) generated by \( y_t = \ldots \)
For a given forecast model and forecasting scheme, the test of Giacomini and Rossi (2009) detects a forecast breakdown when the average of the out-of-sample losses differs significantly from the average of the in-sample losses. The in-sample is used to obtain estimates of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ which are in turn used to construct out-of-sample forecasts $\hat{y}_t = \hat{\beta}_0^{(1)} + \hat{\beta}_0^{(2)} x_{t-1}$. We set $\beta_0^{(1)} = \beta_0^{(2)} = 1$.

We consider a fixed forecasting scheme. Granger’s (2009) test statistic is defined as

$$t^{GR} = \sqrt{T_n} \frac{\hat{\sigma}_e}{\sqrt{\hat{\sigma}_e^2}}$$

where $\hat{\sigma}_e$ is the OLS estimator. We restrict attention to one-step ahead forecasts (i.e., $T = 1$) and $T = 7$.

We set $n_T = 0.66$ as explained in Casini (2022c) and $n_{2,T} = n_{3,T} = n_T$. Simulation results for models involving ARMA, ARCH and heteroskedastic errors are not discussed here because the results are qualitatively equivalent. The significance level is $\alpha = 0.05$ throughout.

### 6.1. Empirical sizes of HAR Inference Tests

Tables 1 and 2 reported the rejection rates for model M1-M4. As a general pattern, we confirm previous evidence that Newey–West’s (1987) HAC estimator leads to $t$-tests that are oversized when the data are stationary and there is substantial dependence [cf. model M1-M2]. This is a long-discussed issue in the literature. Newey–West with prewhitening is often effective in reducing the oversize problem under stationarity. However, the simulation results below and in the literature show that the prewhitened Newey–West-based tests can be oversized when there is high serial dependence. Among the existing methods, the rejection rates of the Newey–West-based tests with KVB’s fixed-$b$ or with Sun’s (2014b) fixed-smoothing are accurate in model M1-M2, with Sun’s (2014b) the most accurate. EWC performs similarly to KVB’s fixed-$b$. Among the recently introduced DK-HAC estimators, Table 1 reports evidence that the non-prewhitened DK-HAC from Casini (2022c) leads to HAR tests that are a bit oversized whereas the tests based on the new DK-HAC with simultaneous data-dependent bandwidths, $J_l^*(\hat{b}_{1,T}, \hat{b}_{2,T})$, are more accurate. The results also show that the tests based on the prewhitened DK-HAC estimators are

---

5We have excluded Andrews’s (1991) HAC estimator since its performance is similar to that of the Newey-West estimator.
competitive with those based on KVB’s fixed-\(b\) in controlling the empirical size. For results involving data-generating processes with stronger serial dependence see Casini and Perron (2022b). They showed that prewhitened DK-HAC estimators are competitive to KVB’s fixed-\(b\) in controlling the empirical size even when the data are stationary with strong dependence. The tests based on \(\hat{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T})\) with prewhitening are more accurate than those using the prewhitened DK-HAC with sequential data-dependent bandwidths. Since \(\hat{J}_T, \text{pw}\), uses a stationarity VAR model to whiten the data, it works as well as \(\hat{J}_T, \text{pw}, \text{SLS}\) and \(\hat{J}_T, \text{pw}, \text{SLS, } \mu\) when stationarity actually holds, as documented in Table 1.

Turning to nonstationary data and to the GR test, Table 2 casts concerns about the finite-sample performance of existing methods in this context. For both model M3 and M4, existing LRV estimators lead to HAR tests that have either empirical size equal or close to zero. The methods that use long bandwidths (i.e., many lagged autocovariances) such as KVB’s fixed-\(b\), Sun’s (2014b) fixed-smoothing, and EWC suffer most from this problem relative to using the Newey–West estimator.\(^6\)

\(^6\)The method of Sun (2014b) suffers more from this problem than KVB’s fixed-\(b\) since the LRV estimator is the same but the critical values of Sun’s (2014b) are in practice larger than the KVB’s fixed-\(b\) critical values.
This is demonstrated in Casini et al. (2022) who showed theoretically that nonstationarity induces positive bias for each sample autocovariance. That bias is constant across lag orders. Since existing LRV estimators are weighted sum of sample autocovariances, the more lags are included the larger is the positive bias. Thus, LRV estimators are inflated and HAR tests have lower rejection rates than the significance level. As we show below, this mechanism has consequences for power as well. In model M3-M4, tests based on the non-prewhitened DK-HAC \( \hat{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T}) \) performs well although tests based on the prewhitened DK-HAC are more accurate. \( \hat{J}_{T,pw,1} \) leads to tests that are slightly less accurate because it uses stationarity and when the latter is violated its performance is affected. In model M4, KVB’s fixed-b and prewhitened DK-HAC are associated to rejection rates relatively close to the significance level.

In summary, the prewhitened DK-HAC estimators yield \( t \)-test in regression models with rejection rates that are relatively close to the exact size. The DK-HAC with simultaneous bandwidths developed in Section 4 performs better (i.e., the associated null rejection rates are closer to the significance level and approach it from below) than the corresponding DK-HAC estimators with sequential bandwidths when the data are stationary. This is in accordance with our theoretical results. Also, for nonstationary data the simultaneous bandwidths perform in general better than the sequential bandwidths, though the margin is smaller. This highlights the tradeoff between rate of convergence and size control. The joint method leads to a faster rate of convergence in general. However, under nonstationarity it uses a larger neighborhood length relative to the sequential method which may explain why the finite-sample performance becomes closer under nonstationarity (a smaller neighborhood is supposed to work better under nonstationarity since the parameters of the DGP change locally). The non-prewhitened DK-HAC can lead to oversized \( t \)-test on the intercept if there is high dependence. Our results confirm the oversize problem induced by the use of the Newey–West estimators documented in the literature under stationarity. Fixed-b/Fixed-smoothing HAR tests control the size well when the data are stationary but can be severely undersized under nonstationarity, a problem that also affects tests based on the Newey–West. Thus, prewhitened DK-HAC estimators are competitive to fixed-b methods under stationarity and they perform well also when the data are nonstationary.

6.2. Empirical power of HAR inference tests

For model M1-M4 we report the power results in Table 3-6. The sample size is \( T=200 \). For model M1, tests based on the Newey and West’s (1987) HAC and on the non-prewhitened DK-HAC estimators have the highest power but they were more oversized than the tests based on other methods. KVB’s fixed-b and Sun’s (2014b) fixed-smoothing lead to \( t \)-tests that sacrifice some power relative to using the prewhitened DK-HAC estimators while EWC-based tests have lower power locally to \( \delta = 0 \) (i.e., \( \delta = 0.1 \) and \( 0.2 \)). In model M2, a similar pattern holds. HAR tests normalized by either classical HAC or DK-HAC estimators have similarly good power while HAR tests based on KVB’s fixed-b or Sun’s (2014b) fixed-smoothing have relatively less power. In model M3, the best power is achieved with Newey–West’s (1987) HAC estimator followed by \( \hat{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T}) \) and EWC. Using KVB’s fixed-b or Sun’s (2014b) fixed-smoothing leads to large power losses. In model M4, it appears that all versions of the classical HAC estimators of Newey and West (1987), the KVB’s fixed-b, Sun’s (2014b) fixed-smoothing and EWC lead to \( t \)-tests that have, essentially, zero power for all \( \delta \). In contrast, the \( t \)-test standardized by the DK-HAC estimators have good power. Among the latter DK-HAC estimators, the ones that use the sequential bandwidths have slightly higher power but they margin is very small. This follows from the usual size-power tradeoff since the simultaneous bandwidths led to tests that have more accurate size control.
The severe power problems of tests based on classical HAC estimators, KVB’s fixed-b, Sun’s (2014b) fixed-smoothing and EWC can be simply reconciled with the fact that under the alternative hypotheses the spectrum of $V_t$ is not constant. Existing estimators estimate an average of a time-varying spectrum. Because of this instability in the spectrum, they overestimate the dependence in $V_t$. Casini et al. (2022) showed that nonstationarity/misspecification alters the low frequency components of a time series making the latter appear as more persistent. Since classical HAC estimators are a weighted sum of an infinite number of low frequency periodogram ordinates, these estimates tend to be inflated. Similarly, LRV estimators using long bandwidths are weighted sum of a large number of sample autocovariances. Each sample autocovariance is biased upward so that the latter estimates are even more inflated than the classical HAC estimators. This explains why KVB’s fixed-b, Sun’s (2014b) fixed-smoothing and EWC HAR tests have large power problems, even though classical HAC estimators are also affected.

Casini et al. (2022) showed that the introduction of the smoothing over time in the DK-HAC estimators avoids such low frequency contamination. This follows because observations belonging to different regimes do not overlap when computing sample autocovariances. This guarantees excellent power properties also under nonstationarity/misspecification or under nonstationary alternative hypotheses (e.g., GR test discussed above). Simulation evidence suggests that tests

| Table 3. Empirical small-sample rejection rates of the $t_1$-test for model M1. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| $\alpha = 0.05$, $T = 200$    | $\delta = 0.1$  | $\delta = 0.2$  | $\delta = 0.4$  | $\delta = 0.8$  |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$ | 0.218           | 0.589           | 0.980           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite | 0.132           | 0.465           | 0.960           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite, SLS | 0.172           | 0.553           | 0.958           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite, SLS, $\mu$ | 0.174           | 0.544           | 0.958           | 1.000           |
| $J_r$, Casini (2022c)         | 0.291           | 0.620           | 0.980           | 1.000           |
| $J_r$, prewhite, CP           | 0.191           | 0.518           | 0.949           | 1.000           |
| $J_r$, prewhite, SLS, CP      | 0.161           | 0.509           | 0.969           | 1.000           |
| $J_r$, prewhite, SLS, $\mu$ CP | 0.165          | 0.508           | 0.970           | 1.000           |
| Newey and West (1987)         | 0.248           | 0.629           | 0.987           | 1.000           |
| Newey and West (1987), prewhite | 0.197           | 0.576           | 0.979           | 1.000           |
| Newey and West (1987), fixed-b (KVB) | 0.141           | 0.373           | 0.844           | 0.998           |
| Newey and West, Sun (2014b)   | 0.121           | 0.339           | 0.811           | 0.998           |
| EWC                            | 0.150           | 0.493           | 0.963           | 1.000           |

CP stands for Casini and Perron (2022b).

| Table 4. Empirical small-sample rejection rates of the $t_2$-tests for model M2. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| $\alpha = 0.05$, $T = 200$    | $\delta = 0.1$  | $\delta = 0.2$  | $\delta = 0.4$  | $\delta = 0.8$  |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$ | 0.263           | 0.642           | 0.988           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite | 0.191           | 0.532           | 0.968           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite, SLS | 0.221           | 0.592           | 0.982           | 1.000           |
| $J_T(\hat{b}_{1,T}, \hat{b}_{2,T})$, prewhite, SLS, $\mu$ | 0.221           | 0.597           | 0.983           | 1.000           |
| $J_r$, Casini (2022c)         | 0.276           | 0.653           | 0.988           | 1.000           |
| $J_r$, prewhite, CP           | 0.237           | 0.611           | 0.986           | 1.000           |
| $J_r$, prewhite, SLS, CP      | 0.225           | 0.598           | 0.982           | 1.000           |
| $J_r$, prewhite, SLS, $\mu$ CP | 0.165           | 0.598           | 0.988           | 1.000           |
| Newey and West (1987)         | 0.268           | 0.332           | 0.992           | 1.000           |
| Newey and West (1987), prewhite | 0.258           | 0.374           | 0.990           | 1.000           |
| Newey and West (1987), fixed-b (KVB) | 0.199           | 0.463           | 0.914           | 1.000           |
| Newey and West, Sun (2014b)   | 0.160           | 0.422           | 0.884           | 1.000           |
| EWC                            | 0.193           | 0.571           | 0.978           | 1.000           |

CP stands for Casini and Perron (2022b).
based on the DK-HAC with simultaneous bandwidths are robust to low frequency contamination and overall performs better than tests based on the DK-HAC with sequential bandwidths especially with respect to size control.

7. Empirical application

We consider the stability of the predictive ability of the Phillips curve when used as a forecast model for inflation. We consider the $t$-test for forecast failure of Giacomini and Rossi (2009) normalized by different LRV estimators. Let $\pi_t^\tau = (1200/\tau) \ln (P_t/P_{t-\tau})$ denote the $\tau$-period inflation in the price level $P_t$ reported at an annual rate, and $u_t$ denote the unemployment gap (i.e., the difference between the unemployment rate and a measure of the NAIRU). The Phillips curve relates changes in inflation to past values of the unemployment gap and to past changes in inflation:

$$\pi_t^{\tau+\tau} - \pi_t = \theta_0 + \theta_1(L)u_t + \theta_2(L)(\pi_t - \pi_{t-1}) + \epsilon_{t+\tau},$$

(7.1)

where $\pi_t \equiv \pi_t^\tau = (1200)\ln (P_t/P_{t-1})$, where $\theta_1(L)$ and $\theta_2(L)$ are lag polynomials with $q_u$ and $q_\pi$ lags, respectively. The literature suggests that the forecasting ability of the Phillips curve is unstable. In particular, Fisher et al. (2002) documented that the Phillips curve appeared to forecast well 12-month ahead during the 1977–1984 period but not during the period 1993–2000.
The same concerns about changes in the performance of Phillips curve for forecasting inflation were expressed by Giacomini and Rossi (2009) and Perron and Yamamoto (2021).

We assume that the researcher generates a sequence of $\tau$-step-ahead forecasts of $Y_{t+\tau} = \pi_{t+\tau}$ using an out-of-sample procedure. That is, we divide the sample size $T$ into an in-sample window of size $m$ and an out-of-sample window of size $n = T - m - \tau + 1$. We consider: (1) a fixed forecasting scheme, where the in-sample window includes observations indexed 1, ..., $m$; and (2) a rolling forecasting scheme, where the in-sample window at time $t$ contains observations indexed $t - m + 1, ..., t$. Let $\beta^* \triangleq (\theta_0, \theta_1', \theta_2')'$ and $f_t(\hat{\beta}_t)$ be the time-$t$ forecast produced by estimating a model over the in-sample window at time $t$, with $\hat{\beta}_t$ indicating the least-squares estimate of $\beta^*$. Each time-$t$ forecast corresponds to a sequence of in-sample fitted values $y_j(\hat{\beta}_t)$, with $j$ varying over the in-sample window.

We evaluate the forecasts using the quadratic loss $L(\cdot)$. Each time-$t$ out-of-sample loss $L_{t+\tau}(\hat{\beta}_t) = L(Y_{t+\tau}, f_t(\hat{\beta}_t))$ corresponds to in-sample losses $L_{j}(\hat{\beta}_t) = L(Y_j, y_j(\hat{\beta}_t))$. Let $X_t$ collect the set of regressors at time $t$ of model (7.1). We have $\hat{\beta}_t = \left(\sum_{j=1}^{m} X_j X_j'\right)^{-1} \sum_{j=1}^{m} X_j Y_{t+\tau}$ for the fixed scheme and $\hat{\beta}_t = \left(\sum_{j=t-m+1}^{t} X_j X_j'\right)^{-1} \sum_{j=t-m+1}^{t} X_j Y_{t+\tau}$ for the rolling scheme. The out-of-sample loss corresponding to the forecast at time $t$ is $L_{t+\tau}(\hat{\beta}_t) = L(Y_{t+\tau}, X_t'\hat{\beta}_t)$ and the corresponding in-sample losses are $L_{j}(\hat{\beta}_t) = L(Y_j, X_j'\hat{\beta}_t)$, where $j = \tau + 1, ..., m$ for the fixed scheme and $j = t - m + 1, ..., t$ for the rolling scheme.

We verify the presence of forecast failure for the Phillips curve using the forecast breakdown test of Giacomini and Rossi (2009). This relies on the sequence of so-called surprise losses. The surprise loss at time $t + \tau$ is defined as the difference between the out-of-sample loss at time $t + \tau$ and the average in-sample loss:

$$\text{SL}_{t+\tau}(\hat{\beta}_t) = L_{t+\tau}(\hat{\beta}_t) - \bar{L}_t(\hat{\beta}_t),$$

for $t = m, ..., T - \tau$, (7.2)

where $\bar{L}_t(\hat{\beta}_t)$ is the average in-sample loss computed over the in-sample window implied by the forecasting scheme. The null hypotheses is $H_0: \mathbb{E}(n^{-1} \sum_{t=n}^{T-\tau} \text{SL}_{t+\tau}(\beta^*)) = 0$. The forecast breakdown test statistic of Giacomini and Rossi (2009) is given by $t_{m,n,\tau}^{GR} = n^{1/2} \overline{\Sigma}^{-1} \tilde{\sigma}_{m,n}$, where $\tilde{\sigma}_{m,n} = \tilde{\sigma}_{m,n}$ and (1) $\lambda = 1 + n/m$ for the fixed scheme, (2) $\lambda = 1 - 1/3(n/m)^2$ for the rolling scheme with $n < m$, (3) $\lambda = 2m/3n$ for the rolling scheme with $n \geq m$, and $\overline{\Sigma}$ is the sample variance of the squared losses if the sequence of squared losses are i.i.d. or an HAC estimator otherwise. A level $\alpha$ test rejects the null hypothesis whenever $|t_{m,n,\tau}^{GR}| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of a standard normal distribution.

The Breusch–Godfrey test for serial correlation in the squared forecast losses suggests the presence of serial dependence. Thus, we use the HAC estimators in place of $\overline{\Sigma}$. Here, we report the results only for DK-HAC estimator $\tilde{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T})$ (with and without prewhitening), the Newey and West’s (1987) estimator with automatic bandwidth and the Newey and West’s (1987) estimator with the fixed-$b$ method of Kiefer et al. (2000).  

We use the same data as in Perron and Yamamoto (2021). We use monthly CPI (consumer price index; revised version), and the unemployment gap for the period 1959:01 to 2004:06. We choose $q_u = 3$ and $q_\pi = 3$. We consider several sizes for the in-sample windows ranging from $m = 150$ (1971:06) to 175 (1973:07). We consider $\tau = 1$ and 12.

Table 7 shows strong rejections of no change in forecasting accuracy for $\tau = 1$ when $t_{m,n,\tau}^{GR}$ uses $\tilde{J}_T(\hat{b}_{1,T}, \hat{b}_{2,T})$. The $t_{m,n,\tau}^{GR}$ tests that use Newey and West’s (1987) (with or without fixed-$b$)

\footnote{For the test normalized by the latter estimator, we use the fixed-$b$ critical values.}

\footnote{The results for $q_\pi = 0$ and the other combinations of $q_u$ and $q_\pi$ are similar and not reported.}
essentially display little evidence for rejection of the hypotheses of no forecast breakdown with the fixed scheme. Indeed, there are cases in which the tests based on fixed-\( b \) are not able to reject the null hypotheses at the 10% significance even when the test \( t^{GR}_{m,n,\tau} \) rejects at the 5% significance level. For the rolling scheme, all tests show evidence against the null hypothesis, though the tests based on \( \bar{J}_T \) show stronger rejections. A clearer pattern holds for the case \( \tau = 12 \) where now also for the rolling scheme the tests based on \( \bar{J}_T \) show much stronger evidence against the null. Hence, the classical HAC standard errors are shown to be often unreliable in the sense that a researcher would misleadingly conclude that the forecasting performance of the Phillips curve is stable which, however, contrasts the empirical findings in the literature. The results are in accordance with the study of Martins and Perron (2016) who discussed the poor power properties of the Giacomini and Rossi’s (2009) test when a standard estimate of the LRV is used. When \( \bar{J}_T \) is used, inference based on \( t^{GR}_{m,n,\tau} \) confirms the evidence of changes in the forecasting performance of the Phillips curve over time as suggested by the literature.

8. Conclusions

We considered the derivation of data-dependent simultaneous bandwidths for double kernel heteroskedasticity and autocorrelation consistent (DK-HAC) estimators. We obtained the optimal bandwidths that jointly minimize the global asymptotic MSE criterion and discussed the tradeoff between bias and variance with respect to smoothing over lagged autocovariances and over time. We highlighted how the derived MSE bounds are influenced by nonstationarity unlike the MSE bounds in Andrews (1991). We compared the DK-HAC estimators with simultaneous bandwidths to the DK-HAC estimators with bandwidths from the sequential MSE criterion. The new method

| \( \tau = 1 \) | Fixed | Rolling |
|---|---|---|
| \( J_T(b_{1,T}, b_{2,T}) \) | 2.08** | 3.53*** |
| \( J_T(b_{1,T}, b_{2,T}), \) prewhite | 2.07** | 3.77*** |
| Newey and West (1987) | 1.89* | 3.33*** |
| Newey and West (1987), fixed-\( b \) (KVB) | 1.37 | 3.53*** |

| \( \tau = 12 \) | Fixed | Rolling |
|---|---|---|
| \( J_T(b_{1,T}, b_{2,T}) \) | 2.16** | 3.05*** |
| \( J_T(b_{1,T}, b_{2,T}), \) prewhite | 2.15** | 3.10*** |
| Newey and West (1987) | 1.92* | 2.91*** |
| Newey and West (1987), fixed-\( b \) (KVB) | 1.35 | 2.75*** |

Table 7. Giacomini and Rossi (2009) \( t \)-test.

Note: \( m + 1 \) refers to the start date of the out-of-sample period.
leads to HAR tests that perform better in terms of size control, especially with stationary and close to stationary data. Finally, we considered LRV estimation where the relevant observations are a function of a nonparametric estimator and established the validity of the HAC and DK-HAC estimators in this setting. Hence, we also extended the consistency results in Andrews (1991) and Newey and West (1987) to nonparametric estimation settings.

References

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

Anderson, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

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

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. (2022). Supplement to “simultaneous bandwidths determination for double-kernel HAC estimators and long-run variance estimation in nonparametric settings. *arXiv*: 2103.00060.

Brillinger, D. (1975). *Time series data analysis and theory*. New York: Holt, Rinehart and Winston.

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

Casini, A. (2019). Improved methods for statistical inference in the context of various types of parameter variation (Ph. D Dissertation), Boston University.

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

Casini, A. (2022a). Comment on andrews (1991)” heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 90:1–2.

Casini, A. (2022b). The fixed-b limiting distribution and the ERP of HAR tests under nonstationarity. *arXiv*: 2111.14590.

Casini, A. (2022c). Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *Journal of Econometrics, Forthcoming*

Casini, A., Deng, T., & Perron, P. (2022). Theory of low frequency contamination from nonstationarity and misspecification: consequences for HAR inference. *arXiv Preprint arXiv:2103.01604*

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

Casini, A., & Perron, P. (2021a). Change-point analysis of time series with evolutionary spectra. *arXiv*: 2106.02031.

Casini, A., & Perron, P. (2021b). Continuous record asymptotics for change-point models. *arXiv*: 1803.10881.

Casini, A., & Perron, P. (2021c). Continuous record laplace-based inference about the break date in structural change models. *Journal of Econometrics* 224:3–21.

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

Casini, A., & Perron, P. (2022b). Prewhitened long-run variance estimation robust to nonstationarity. *arXiv Preprint arXiv*: 2103.02235.

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

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

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

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

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

Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics* 151:101–112.

Dahlhaus, R. (2012). Locally stationary processes. In Subba Rao, T., Subba Rao, S., Rao, C. (Eds.), *Handbook of Statistics* (Vol. 30, pp. 351–413), Elsevier.
Dahlhaus, R., & Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis* 19:629–655.

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

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

Dou, L. (2019). *Optimal HAR inference* (Unpublished manuscript). Department of Economics, Princeton University.

Epanechnikov, V. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications* 14:153–158.

Fisher, J., Liu, C., Zhu, R. (2002). *When can we forecast inflation?* Economic Perspective Federal Reserve Bank of Chicago.

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

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

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

Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.

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

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.

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.

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.

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

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.

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

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

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.
Priestley, M. B. (1981). *Spectral analysis and time series* (Vols. I and II). New York: Academic Press.

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

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

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). Fixed-smoothing asymptotics in the presence of strong autocorrelation. *Advances in Econometrics: Essays in Honor of Peter C.B. Phillips* 33:23–63.

Sun, Y. (2014c). 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.

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.

Whileelm, D. (2015). Optimal bandwidth selection for robust generalized methods of moments estimation. *Econometric Theory* 31:1054–1077.

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

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

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association* 103:726–740.