Inference in Time Series Models using Smoothed-Clustered Standard Errors

Seunghwa Rho
Department of Quantitative Theory and Methods
Emory University

Timothy J. Vogelsang*
Department of Economics
Michigan State University

June 6, 2020

Abstract

This paper proposes a long run variance estimator for conducting inference in time series regression models that combines the nonparametric approach with a cluster approach. The basic idea is to divide the time periods into non-overlapping clusters. The long run variance estimator is constructed by first aggregating within clusters and then kernel smoothing across clusters or applying the nonparametric series method to the clusters with Type II discrete cosine transform. We develop an asymptotic theory for test statistics based on these "smoothed-clustered" long run variance estimators. We derive asymptotic results holding the number of clusters fixed and also treating the number of clusters as increasing with the sample size. For the kernel smoothing approach, these two asymptotic limits are different whereas for the cosine series approach, the two limits are the same. When clustering before kernel smoothing, we find that the "fixed-number-of-clusters" asymptotic approximation works well whether the number of clusters is small or large. Finite sample simulations suggest that the naive *i.i.d.* bootstrap mimics the fixed-number-of-clusters critical values. The simulations also suggest that clustering before kernel smoothing can reduce over-rejections caused by strong serial correlation although at a cost of power. When there is a natural way of clustering, clustering can reduce over-rejection problems and achieve small gains in power for the kernel approach. In contrast, the cosine series approach does not benefit from clustering.

Keywords: Fixed-b Asymptotics, Equally Weighted Cosines, Systematic Missing Data, Heteroskedasticity Autocorrelation Robust Inference

*We thank Serena Ng and an anonymous referee for helpful comments and suggestions. Correspondence: Tim Vogelsang, Department of Economics, 486 W. Circle Drive, 110 Marshall-Adams Hall, Michigan State University East Lansing, MI 48824-1038. Phone: 517-353-4582, Fax: 517–432-1068, email: tjv@msu.edu

1 Introduction

This paper proposes long run variance estimators for conducting inference in time series regression models that combines the traditional nonparametric kernel smoothing approach (Newey and West (1987) and Andrews (1991)) or equally weighted cosine (EWC) series approach (Grenander and Rosenblatt (1953), Phillips (2005), Müller (2007), Sun (2013) and Lazarus, Lewis, Stock and Watson (2018)) with a dependent clusters approach (Bester, Conley and Hansen (2011)). We label this combined long run variance estimator the "smoothed-clustered" long run variance estimator.

The basic idea is to divide the time periods into non-overlapping clusters with equal number of observations. From a practical perspective, dividing the data into non-overlapping clusters of equal size is a straightforward mechanical process because of the natural ordering of time series data. Applicability is wide. In some cases, data structures naturally lend themselves to equal sized clustering. For example, consider time series data for markets that are open on weekdays but are closed on weekends.[1] It is natural to cluster by week in which case each cluster has five observations. One could also naturally cluster by two week periods or other integer groupings of weeks.

The smoothed-clustered long run variance estimator is constructed by first aggregating within clusters and then kernel smoothing across clusters or applying nonparametric series methods to these aggregated series with Type II discrete cosine transform. For the kernel smoothing case, the approach is similar in spirit to the approach proposed by Driscoll and Kraay (1998) in panel settings. We develop asymptotic theory for test statistics based on the smoothed-clustered long run variance estimator under the assumption that the time series data is weakly dependent and covariance stationary. We obtain results under two asymptotic approaches that are commonly used in the cluster inference literature. The first approach treats the number of observations per cluster as fixed as the number of clusters increases with the sample size. The second approach holds the number of clusters fixed as the number of observations per cluster increases with the sample size.

The "large-number-of-clusters" framework has received some attention in the recent econometrics literature. Hansen and Lee (2019) provide a comprehensive asymptotic distribution theory for large numbers of independent clusters. Related work by Djogbenou, MacKinnon and Nielsen (2019) establishes conditions under which the wild bootstrap allows valid inference with a large number of independent clusters. Both frameworks allow general dependence within clusters, substantial heterogeneity across clusters, and the number of observations within clusters can be fixed or increasing. In contrast, because of our weakly dependent and covariance stationary time series setting, clusters

[1] Of course, other data structures may imply clusters with unequal observations such as clusters based on business cycle timing. If the underlying time series data is covariance stationary and weakly dependent across recessions and booms, then the timing of clusters won't matter as long as unequal cluster sizes are taken into account. We conjecture that valid inference can proceed using unequal cluster sizes however with more tedious technical details. If the covariance structure depends on whether the economy is in a recession or not, then timing of clusters matters. If recession timing is latent and the wrong dates are used, this could complicate inference. We leave an analysis of latent cluster timing to future work.

are dependent but they are homogeneous.

The "fixed-number-of-clusters" framework has been used by Hansen (2007), Bester et al. (2011) and Ibragimov and Müller (2010) among others. Hansen (2007) obtained results in panel models with large time dimensions and Ibragimov and Müller (2010) proposed a statistic based on aggregation of subsample estimators. Bester et al. (2011) obtain results for cluster inference in spatial settings that include weakly dependent covariance stationary time series as a special case. Without smoothing, our approach falls within the framework of Bester et al. (2011) as a special case. With smoothing, our approach is an extension of Bester et al. (2011) to a case where dependent clusters have a distance measure (time) related to strength of correlation across clusters.

For the kernel smoothing approach, the large-number-of-clusters results we develop are closely linked to the fixed-b results obtained by Vogelsang (2012) for Driscoll and Kraay (1998) statistics in panel settings. We show that in the large number of clusters setting, robust test statistics follow the standard fixed-b limits obtained by Kiefer and Vogelsang (2005) assuming that the kernel bandwidth is treated as a fixed proportion of the number of clusters. In contrast, in the fixed-number-of-clusters setting, we obtain a different asymptotic limit that depends on the number of clusters. For the EWC approach, we show that the large-number-of-clusters and the fixed-number-of-clusters limits are the same when the number of cosine basis functions is held fixed. One might expect the relative accuracy of the two asymptotic approximations to depend on the number of clusters relative to the sample size in the kernel smoothing method. However, we find in a simulation study that the fixed-number-of-clusters asymptotic approximation works well whether the number of clusters is small or large as does the common limit for the EWC approach. The simulations also suggest that the naive *i.i.d.* bootstrap mimics the fixed-number-of-clusters critical values of the kernel smoothing approach.

Outside of data structures that suggest natural clustering, the motivation for clustering before kernel or EWC series smoothing is as follows. Aggregating within clusters works well when serial correlation is relatively strong within clusters. Under a weak dependence and covariance stationarity assumption, cluster averages will be asymptotically independent of each other. However, in finite samples, the cluster averages will be correlated and taking this into account by smoothing can help reduce finite sample over-rejection problems. In our finite sample simulations, clustering before kernel smoothing does reduce over-rejections caused by strong serial correlation but, not surprisingly, at a cost of power. In contrast for the EWC approach, clustering does not further reduce over-rejections. In fact clustering may induce some small additional over-rejections in the presence of strong serial correlation. For cases where the data has a natural cluster structure, clustering that matches the structure in the data can help reduce over-rejection problems and deliver some modest gains in power for the kernel approach. In contrast, clustering does not improve the performance of the EWC approach.

The rest of the paper is organized as follows. In the next section the model is given and it lays out the inference problem with long run variance estimators and the relevant test statistics. Section

3 provides asymptotic results for test statistics based on the smoothed-clustered long run variance estimators. Section 4 explores the finite sample properties of the test statistics in a simple location model. For the kernel smoothing approach, we use both asymptotic and bootstrap critical values. Section 5 discusses some data dependent bandwidth approaches focusing on mean square error (MSE) optimal bandwidths (Andrews (1991)) and the test-optimal bandwidths (Sun, Phillips and Jin (2008)). Section 6 concludes. Key proofs are given in an appendix. Theory for the case where the number of clusters does not evenly divide the sample is provided in Supplemental Appendix A along with derivations for the data dependent bandwidths. Tables of asymptotic critical values for kernel tests for the fixed-number-of-clusters case are given in Supplemental Appendix B.

2 Clustered Smoothed Standard Errors and Test Statistics

Consider the time series regression model,

$$y_t = x_t'\beta + u_t, t = 1, \ldots, T,$$

where β is a $(k \times 1)$ vector of regression parameters, x_t is a $(k \times 1)$ vector of regressors, and u_t is a mean zero error process and T is the sample size. The ordinary least squares (OLS) estimator of β is

$$\widehat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t.$$

Suppose we divide the time series into G contiguous, non-overlapping clusters of equal size n_{G} so that $T = n_{\mathrm{G}} G$.[2] The OLS estimator can be rewritten using cluster notation as

$$\widehat{\beta} = \left(\sum_{g=1}^{G} \sum_{t=(g-1)n_{\mathrm{G}}+1}^{g\,n_{\mathrm{G}}} x_t x_t'\right)^{-1} \sum_{g=1}^{G} \sum_{t=(g-1)n_{\mathrm{G}}+1}^{g\,n_{\mathrm{G}}} x_t y_t.$$

Conceptually, this way of rewriting $\widehat{\beta}$ can be viewed as the outcome of rearranging the data into G time periods with n_{G} "cross-section" units per time period resulting in an artificial panel data structure. From this artificial panel perspective, $\widehat{\beta}$ is the pooled OLS estimator of β. Plugging in for y_t and centering around β gives

$$\widehat{\beta} - \beta = \left(\sum_{g=1}^{G} S_g^{xx}\right)^{-1} \sum_{g=1}^{G} \overline{v}_g,$$

where

$$\overline{v}_g = \sum_{t=(g-1)n_{\mathrm{G}}+1}^{g\,n_{\mathrm{G}}} v_t \quad \text{and} \quad S_g^{xx} = \sum_{t=(g-1)n_{\mathrm{G}}+1}^{g\,n_{\mathrm{G}}} x_t x_t'$$

[2] Cases where G does not evenly divide T is easily handled but notation is more tedious. See Supplemental Appendix A.

with $v_t = x_t u_t$. Note that $\overline{v}_g$ and S_g^{xx} are within cluster sums.

The kernel smoothed-clustered long run variance estimator of v_t is constructed as follows. Let $\widehat{v}_t = x_t \widehat{u}_t$, where $\widehat{u}_t = y_t - x_t' \widehat{\beta}$ are the OLS residuals. Define the within cluster sums of $\widehat{v}_t$ as

$$\widehat{\overline{v}}_g = \sum_{t=(g-1)n_{\mathrm{G}}+1}^{g n_{\mathrm{G}}} \widehat{v}_t, \quad g = 1, \ldots, G.$$

Using $\widehat{\overline{v}}_g$, the autocovariance matrix estimator is computed as

$$\widehat{\overline{\Gamma}}_j = G^{-1} \sum_{g=j+1}^{G} \widehat{\overline{v}}_g \widehat{\overline{v}}_{g-j}' \quad \text{for} \quad j \geq 0.$$

Let $\mathcal{K}(x)$ be a kernel function such that $\mathcal{K}(x) = \mathcal{K}(-x)$, $\mathcal{K}(0) = 1$, $|\mathcal{K}(x)| \leq 1$, $\mathcal{K}(x)$ be continuous at $x = 0$, and $\int_{-\infty}^{\infty} \mathcal{K}^2(x) < \infty$. Let M_G be the bandwidth parameter. The clustered heteroskedasticity autocorrelation robust (CHAC) variance estimator of $\overline{v}_g$ is defined as

$$\widehat{\overline{\Omega}}^{CHAC} = \widehat{\overline{\Gamma}}_0 + \sum_{j=1}^{G-1} \mathcal{K}\left(\frac{j}{M_G}\right)\left(\widehat{\overline{\Gamma}}_j + \widehat{\overline{\Gamma}}_j'\right) = \frac{1}{G} \sum_{g=1}^{G} \sum_{h=1}^{G} \mathcal{K}\left(\frac{|g-h|}{M_G}\right) \widehat{\overline{v}}_g \widehat{\overline{v}}_h'.$$

Notice that the CHAC estimator gives full weight for observations within clusters, a feature that the usual nonparametric kernel HAC estimator does not have. Smoothing across clusters accounts for finite sample serial correlation across clusters and is a generalization of the cluster estimator proposed by Bester et al. (2011). The Bester et al. (2011) estimator is obtained when $\widehat{\overline{\Omega}}^{CHAC} = \widehat{\overline{\Gamma}}_0$, i.e. when no smoothing is used across clusters. Also note that when $G = T$ and $n_{\mathrm{G}} = 1$, the CHAC estimator becomes the usual kernel HAC estimator. Therefore, the CHAC estimator nests the traditional kernel approach and the time series cluster approach.

The second long run variance estimator we consider is the EWC estimator (Müller (2007)) applied to the clusters and is defined as

$$\widehat{\overline{\Omega}}^{CEWC} = \frac{1}{B} \sum_{j=1}^{B} \widehat{\overline{\Omega}}_j, \quad \widehat{\overline{\Omega}}_j = \widehat{\Lambda}_j \widehat{\Lambda}_j', \quad \widehat{\Lambda}_j = \sqrt{\frac{2}{G}} \sum_{g=1}^{G} \cos\left(\frac{(g-0.5)}{G}\pi j\right) \widehat{\overline{v}}_g,$$

where CEWC denotes "cluster before using equally weighted cosine" estimator. The CEWC estimator was proposed by Müller (2007) and is a special case of the orthonormal series estimator of Sun (2013). It has been recommended in practice in a recent paper by Lazarus et al. (2018).

Suppose we are testing a linear hypothesis about β of the form $H_0 : R\beta = r$ against $H_1 : R\beta \neq r$, where R is a $m \times k$ matrix of known constants with full rank and r is a $m \times 1$ vector of known constants. Define Wald statistics for $l \in \{CHAC, CEWC\}$ as

$$W_l = \left(R\widehat{\beta} - r\right)' \left[R\widehat{V}_l R'\right]^{-1} \left(R\widehat{\beta} - r\right),$$

where

$$\widehat{V}_l = G\left(\sum_{g=1}^{G} S_g^{xx}\right)^{-1} \widehat{\overline{\Omega}}^l \left(\sum_{g=1}^{G} S_g^{xx}\right)^{-1}.$$

For the case of $m = 1$, we can define a t-statistic as

$$t_l = \frac{\left(R\widehat{\beta} - r\right)}{\sqrt{R\widehat{V}_l R'}}.$$

For the analysis of data dependent bandwidth approaches, it is useful to note that while $\widehat{\overline{\Omega}}^l$ is an estimator of the long-run variance of $\overline{v}_g$, it is easy to verify that $n_{\text{G}}^{-1}\widehat{\overline{\Omega}}^l$ is an estimator of the long run variance of v_t. Using $\sum_{g=1}^{G} S_g^{xx} = \sum_{t=1}^{T} x_t x_t'$ and $T = n_{\text{G}} G$, we can rewrite $\widehat{V}_l$ in the conventional form

$$\widehat{V}_l = T\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \widehat{\Omega}^l \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}$$

where $\widehat{\Omega}^l = n_{\text{G}}^{-1}\widehat{\overline{\Omega}}^l$.

3 Asymptotic Theory

We obtain asymptotic results for the CHAC and CEWC statistics using two distinct asymptotic nestings for G and n_{G}. The first approach is to let G increase with the sample size, T, but hold n_{G} fixed, i.e. large-G, fixed-n_{G} asymptotics. The second approach is to hold G fixed and let n_{G} increase with T, i.e. fixed-G, large-n_{G} asymptotics. Results for the two approaches are treated separately as they require slightly different regularity conditions. Throughout, the symbol "$\Rightarrow$"denotes weak convergence of a sequence of stochastic process to a limiting stochastic process.

3.1 Large-G, fixed-n_G case

In this section we assume that $G \to \infty$ and n_{G} is held fixed as $T \to \infty$. By definition, $n_{\text{G}} = T/G$, so we are implicitly assuming that G is a fixed proportion of the sample size. Vogelsang (2012) developed fixed-b results for the Driscoll and Kraay (1998) panel analogues to W_{CHAC} and t_{CHAC} for the case of large number of time periods and fixed number of cross-section units. Vogelsang (2012) provided conditions under which the fixed-b limits are equivalent to the standard fixed-b limits obtained by Kiefer and Vogelsang (2005). Given the natural similarities between W_{CHAC} or t_{CHAC} and the panel statistics, it is not surprising that the large-G, fixed-n_{G} limits of W_{CHAC} and t_{CHAC} follow the standard fixed-b limits under suitable regularity conditions. The asymptotic theory in Vogelsang (2012) mainly relies on weak dependence and covariance stationarity in the time dimension of the panel. In our model, because we divide the pure time series into non-overlapping clusters, as long as the original time series satisfies weak dependence and covariance stationarity, the regularity conditions used by Vogelsang (2012) hold here as well.

For the CEWC statistics, Sun (2013) provides relevant assumptions to obtain results with the number of cosine terms, B, held constant, i.e. fixed-B limits. The assumptions used by Sun (2013) are weaker than those required for the fixed-b kernel smoothing tests. This is because the limit of the CEWC test statistics are based on a multivariate central limit theorem (CLT) which is implied by the functional central limit theorem (FCLT) required for fixed-b asymptotic theory.

The following assumptions are sufficient to obtain results in the large-G, fixed-n_G case.

Assumption A *1. n_G is fixed and $G \to \infty$ as $T \to \infty$.*

2. For $r \in (0,1]$, $G^{-1}\sum_{g=1}^{[rG]}\sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' \Rightarrow rQ_c$, where Q_c is non-singular.

3. $E(\overline{v}_g) = 0$ and $G^{-1/2}\sum_{g=1}^{[rG]}\overline{v}_g \Rightarrow \Lambda_c \mathcal{W}_k(r)$, where $\mathcal{W}_k(r)$ is an $k \times 1$ vector of independent standard Wiener processes and $\Lambda_c\Lambda_c' = \Omega_c$ is the $k \times k$ long run variance matrix (2π times the zero frequency spectral density matrix) of $\overline{v}_g$.

Assumptions A2 and A3 are the usual high level assumptions used to obtain fixed-b asymptotic results. Note that

$$\frac{1}{G}\sum_{g=1}^{[rG]}\sum_{t=(g-1)n_G+1}^{gn_G} x_t x_t' = \frac{1}{G}\sum_{t=1}^{[rG]n_G} x_t x_t' = \frac{n_G}{T}\sum_{t=1}^{[\frac{r}{n_G}T]n_G} x_t x_t',$$

where the second equality is obtained by plugging in $G = T/n_G$. If the second moment of x_t satisfies a law of large numbers (LLN) uniformly in r, i.e. $T^{-1}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ$, then Assumption A2 is satisfied with $Q_c = n_G Q$ because $(n_G/T)\sum_{t=1}^{[(r/n_G)T]n_G} x_t x_t'$ is asymptotically equivalent to $(n_G/T)\sum_{t=1}^{[rT]} x_t x_t'$. Assumption A3 states that a FCLT holds for the scaled partial sums of $\overline{v}_g$. As with Assumption A2, we can show that $n_G^{1/2}T^{-1/2}\sum_{t=1}^{[\frac{r}{n_G}T]n_G} v_t$ is asymptotically equivalent to $n_G^{1/2}T^{-1/2}\sum_{t=1}^{[rT]} v_t$ and it follows that

$$\Omega_c = n_G \Omega$$

where Ω is the long run variance of v_t.

Under primitive assumptions for a FCLT such as v_t being a mean zero δ-order (for some $\delta > 2$) covariance stationary process that is α-mixing of size $-\nu/(\nu-2)$,[3] then $\overline{v}_g$ is also a mean zero δ-order (for some $\delta > 2$) covariance stationary process that is α-mixing of the same size because finite sums ($n_G < \infty$) of α-mixing processes are also α-mixing with the same size. See White (2001). Therefore, if a FCLT holds for the scaled partial sums of v_t, then it will hold for the scaled partial sums of $\overline{v}_g$. In general, Assumptions A2 and A3 are slightly weaker than assumptions usually used to obtain fixed-b results and are sufficient for the following theorem. The following theorem gives the asymptotic behavior of OLS, W_{CHAC}, and W_{CEWC}. The proof is provided in the Appendix.

[3]Phillips and Durlauf (1986) provide sufficient conditions for v_t to satisfy a FCLT.

Theorem 1 *Suppose that Assumption A is satisfied. Then, the following holds as $T \to \infty$.*

(a) Asymptotic normality of OLS:

$$\sqrt{G}\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{G}\sum_{g=1}^{G} S_g^{xx}\right)^{-1} G^{-1/2}\sum_{g=1}^{G}\overline{v}_g \Rightarrow (Q_c)^{-1}\Lambda_c \mathcal{W}_k(1).$$

(b) CHAC result: Let $\mathcal{K}_b^*(r,s) = \mathcal{K}\left(\frac{r-s}{b}\right) - \int_0^1 \mathcal{K}\left(\frac{r-\tau}{b}\right)d\tau - \int_0^1 \mathcal{K}\left(\frac{t-s}{b}\right)dt + \int_0^1\int_0^1 \mathcal{K}\left(\frac{t-\tau}{b}\right)dtd\tau$. *Assume* $M_G = bG$ *where* $b \in (0,1]$ *is fixed. Then,*

$$\widehat{\Omega}^{CHAC} \Rightarrow \Lambda_c \int_0^1\int_0^1 \mathcal{K}_b^*(r,s)d\mathcal{W}_k(r)d\mathcal{W}_k(s)'\Lambda_c',$$

and under H_0,

$$W_{CHAC} \Rightarrow \mathcal{W}_m(1)'\left[\int_0^1\int_0^1 \mathcal{K}_b^*(r,s)d\mathcal{W}_m(r)d\mathcal{W}_m(s)'\right]^{-1}\mathcal{W}_m(1).$$

In the case of $m = 1$,

$$t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{\int_0^1\int_0^1 \mathcal{K}_b^*(r,s)d\mathcal{W}_1(r)d\mathcal{W}_1(s)}}.$$

(c) CEWC result: Let $\xi_j^{(d)} \overset{i.i.d.}{\sim} N(0, I_d)$. *Assume* B *is held fixed. Then,*

$$\widehat{\Omega}^{CEWC} = \frac{1}{B}\sum_{j=1}^{B}\widehat{\Omega}_j \Rightarrow \Lambda_c \frac{1}{B}\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)\prime}\Lambda_c',$$

and under H_0,

$$F_{CEWC} = \frac{B-m+1}{mB}W_{CEWC} \Rightarrow F_{m,B-m+1},$$

where $F_{m,B-m+1}$ *is the F distribution with degrees of freedom* $(m, B-m+1)$*. In the case of* $m = 1$,

$$t_{CEWC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{\frac{1}{B}\sum\limits_{j=1}^{B}\left(\xi_j^{(1)}\right)^2}} \overset{d}{=} t_B,$$

where t_B *is the t-distribution with degrees of freedom* B.

When no smoothing is used, the results in Theorem 1(b) share some similarities with the large-G results in Hansen and Lee (2019) and Djogbenou et al. (2019) although there are key differences in the assumptions used in those papers. As $b \to 0$ or $B \to \infty$, no smoothing is used after clustering, and it is well known in the fixed-smoothing literature that the limiting random variables in Theorem 1(b) become standard chi-square and normal.

While the cluster-only limits are the same as in Hansen and Lee (2019) and Djogbenou et al. (2019), the assumptions used here and in those papers have key differences. First, Assumption A allows dependence across homogeneous clusters. In contrast, Hansen and Lee (2019) and Djogbenou et al. (2019) assume clusters are independent but within cluster dependence can be strong and heterogenous across clusters. Second, Assumption A makes cluster sizes equal and holds the cluster size fixed. In contrast, Hansen and Lee (2019) and Djogbenou et al. (2019) use assumptions that allow clusters to have different sizes that potentially increase with the sample size. Theorem 1(b) suggests the results of Hansen and Lee (2019) and Djogbenou et al. (2019) could continue to hold with weak dependence across clusters at least for the case of clusters with equal and fixed size.

An interesting theoretical question is whether Theorem 1(b) continues to hold with unequal cluster sizes or cluster sizes that increase with G. While the technical details are likely to be complicated given the reliance on a FCLT for the kernel smoothing case, as long as the suitably scaled $\overline{v}_g$ converge to nondegenerate random variables, it is reasonable to conjecture that a version of Theorem 1(b) continues to hold if cluster sizes increase with G.

3.2 Fixed-G, large-n_G case

Now suppose we flip the asymptotic nesting so that G is held fixed as $T \to \infty$ in which case $n_G \to \infty$. In this case, the number of observations per cluster is a fixed proportion of the sample size. With the number of clusters fixed, the LLN, FCLT and multivariate CLT work within the clusters rather than across the clusters. If the limit theorems hold for the original time series, this implies that the limit theorems hold within clusters. The following assumptions are sufficient to obtain results in the fixed-G, large-n_G case.

Assumption B *1. G is fixed and $n_G \to \infty$ as $T \to \infty$.*

2. For $r \in (0,1]$, $T^{-1} \sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ$ where Q is non-singular.

3. For $r \in (0,1]$, $T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda \mathcal{W}_k(r)$, where $\Omega = \Lambda\Lambda'$ is the $k \times k$ long run variance matrix of v_t.

Assumptions B2 and B3 state that a LLN applies to $T^{-1} \sum_{t=1}^{[rT]} x_t x_t'$ uniformly in r and a FCLT applies to the scaled partial sum of v_t. The following theorem gives the asymptotic behavior of OLS, W_{CHAC}, and W_{CEWC} and the proof is provided in the Appendix.

Theorem 2 *Under Assumption B, the following holds as $T \to \infty$.*

(a) Asymptotic normality of OLS:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{T}\sum_{g=1}^{G} S_g^{xx}\right)^{-1} T^{-1/2} \sum_{g=1}^{G} \overline{v}_g \Rightarrow Q^{-1}\Lambda\mathcal{W}_k(1).$$

(b) *CHAC result: Assume $M_G = bG$ where $b \in (0,1]$ is fixed. Then*

$$\frac{1}{n_G}\widehat{\Omega}^{CHAC} \Rightarrow \Lambda P_k(G,b)\Lambda',$$

where

$$P_k(G,b) = \int_0^1\int_0^1 \mathcal{K}\left(\frac{\sum_{j=1}^{G} j\mathbb{1}\left[\frac{(j-1)+1}{G} \le r \le \frac{j}{G}\right] - \sum_{j=1}^{G} j\mathbb{1}\left[\frac{(j-1)+1}{G} \le s \le \frac{j}{G}\right]}{bG}\right) d\widetilde{\mathcal{W}}_k(r)d\widetilde{\mathcal{W}}_k(s)',$$

with $d\widetilde{\mathcal{W}}_k(r) = d\mathcal{W}_k(r) - dr\mathcal{W}_k(1)$, and

$$W_{CHAC} \Rightarrow \mathcal{W}_m(1)'\left[P_m(G,b)\right]^{-1}\mathcal{W}_m(1).$$

In the case of $m = 1$,

$$t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{P_1(G,b)}}.$$

(c) *CEWC result: Assume B is held fixed. Then,*

$$\frac{G}{T}\widehat{\Omega}^{CEWC} = \frac{1}{B}\sum_{j=1}^{B}\widehat{\Omega}_j \Rightarrow \Lambda\frac{1}{B}\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)\prime}\Lambda',$$

and under H_0,

$$F_{CEWC} = \frac{B-m+1}{mB}W_{CEWC} \Rightarrow F_{m,B-m+1},$$

In the case of $m = 1$,

$$t_{CEWC} \Rightarrow \frac{\mathcal{W}_1(1)}{\sqrt{\frac{1}{B}\sum_{j=1}^{B}\left(\xi_j^{(1)}\right)^2}} \overset{d}{=} t_B.$$

The fixed-G, large-n_G asymptotic limits of W_{CHAC} and t_{CHAC} in Theorem 2(b) are different from the standard fixed-b asymptotic limits found in Theorem 1(b). The limits in Theorem 2(b) depend on both G and b. Therefore, different asymptotic critical values are needed across b for each value of G. Table B in the Supplemental Appendix B tabulates asymptotic critical values for t_{CHAC} with the Bartlett kernel for a range of values for G. When G is small, the critical values that correspond to a given value of b are substantially different from the standard fixed-b critical values and have fatter tails. This makes sense because using a small value of G is equivalent to using a large bandwidth. As G increases, clustering is reduced and critical values approach the standard fixed-b critical values. A simple way to implement the fixed-G, fixed-b critical values is to use the i.i.d. bootstrap following Gonçalves and Vogelsang (2011). Finite sample simulations reported in the next section indicate that the i.i.d. bootstrap works well in the simple location model for both small and large values of G.

When no smoothing is used, the limits of the CHAC statistics simplify to the scaled t and F limits obtained by Bester et al. (2011) for case with strong homogeneity across clusters. This is expected given that Assumption B implies homogeneity across clusters and Assumption B is essentially a time series version of the assumptions used by Bester et al. (2011) in their spatial setting. Bester et al. (2011) also provide results that allow heterogeneity across clusters and they appeal to a result in Ibragimov and Müller (2010) to provide critical values based on a bounding argument. It is not obvious how those arguments would work when there is smoothing across clusters and we leave such an investigation for future work.

The limit of the CEWC statistics is the same in the fixed-G, large-n_{G} case as in the large-G, fixed-n_{G} case. This suggests that the critical values from the F and t distributions will perform similarly in practice regardless of whether G is small or large. Our finite sample simulations in the next section show that this is indeed the case unless serial correlation is very strong.

4 Finite Sample Performance

In this section, we examine the finite sample performance of the test statistics based on the CHAC and CEWC estimators using a simple location model. The data generating process (DGP) we consider is

$$y_t = \beta + u_t,$$
$$u_t = \rho u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1},$$

where $u_0 = \varepsilon_0 = 0,\ \varepsilon_t \sim i.i.d.\, N(0,1)$ with $\rho \in \{-0.5, 0, 0.5, 0.8, 0.9\}$, $\theta \in \{-0.5, 0, 0.5\}$. Results are given for the sample size $T = 60$ with number of clusters $G \in \{2, 3, 4, 5, 6, 10, 12, 15, 60\}$ that are factors of 60 so that clusters evenly divide the sample. With this DGP, we test the null hypothesis $H_0 : \beta = 0$ against the alternative $H_1 : \beta \neq 0$ at a nominal level of 5%. When computing the CHAC t-statistic, we use the Bartlett, QS and Daniell kernels with $M \in \{1, 2, \ldots, 9, 10, 12, 15, 30, 40, 50, 60\}$. When computing the CEWC t-statistic, we consider $B \in \{1, \ldots, 59\}$. Here we focus on representative results for $\rho \in \{0, 0.5, 0.8\}$, $\theta \in \{0\}$ and we exclude the Daniell kernel given the very similar results to the QS kernel. Tables with a full set of empirical null rejections and size-adjusted power are available upon request.

In this simple location model, the CHAC and CEWC t-statistics are computed as

$$t_l = \frac{\widehat{\beta}}{\sqrt{G\left(T^{-1}\widehat{\widehat{\Omega}}^{l}T^{-1}\right)}} = \frac{\sqrt{T}\widehat{\beta}}{\sqrt{\frac{\widehat{\widehat{\Omega}}^{l}}{n_{\text{G}}}}}, \quad l \in \{CHAC, CEWC\},$$

where

$$\widehat{\widehat{\Omega}}^{CHAC} = \frac{1}{G}\sum_{g=1}^{G}\sum_{h=1}^{G} k\left(\frac{|g-h|}{M_G}\right)\widehat{\widehat{v}}_g\widehat{\widehat{v}}_h$$

and

$$\widehat{\Omega}^{CEWC} = \frac{1}{B}\sum_{j=1}^{B}\widehat{\Omega}_j, \quad \widehat{\Omega}_j = \widehat{\Lambda}_j^2, \quad \widehat{\Lambda}_j = \sqrt{\frac{2}{G}}\sum_{g=1}^{G}\cos\left(\frac{(g-0.5)}{G}\pi j\right)\widehat{v}_g$$

with $\widehat{v}_g = \sum_{t=(g-1)n_G+1}^{gn_G}\widehat{v}_t$, $\widehat{v}_t = y_t - \widehat{\beta}$, and $\widehat{\beta} = T^{-1}\sum_{t=1}^{T} y_t$.

4.1 Empirical Null Rejections

In this section, we examine empirical null rejection probabilities of the CHAC and CEWC test statistics. Note that when $G = T$, it follows that $n_G = 1$ and the CHAC and CEWC estimators simplify to the usual HAC and EWC variance estimators without clustering. For the CHAC approach the pure time series clustering approach of Bester et al. (2011) is obtained when $M = 1$.

We compute empirical null rejection probabilities using 10,000 replications. We reject the null hypothesis whenever $|t_l| > t_c$, $l = CEWC, CHAC$, where t_c is a critical value. For the CEWC approach, regardless of whether G is considered as fixed or $G \to \infty$, the critical value is the 97.5% percentile of the t_B distribution (Theorem 1(c) and 2(c)). On the other hand, for the CHAC approach, the limiting distributions of the test statistic differ depending on whether G is considered fixed or $G \to \infty$. When $G \to \infty$, the asymptotic critical value is the 97.5% percentile of the standard fixed-b asymptotic distribution with $b = M_G/G$ (Theorems 1(b)). For the fixed-G case the critical value is the 97.5% percentile of the distribution given in Theorem 2(b).

These nonstandard asymptotic critical values are obtained using standard simulation methods. Given that the asymptotic critical values in the fixed-G case depend on both G and M_G, a convenient alternative is to use the bootstrap to obtain critical values. We use the naive *i.i.d.* bootstrap critical values and the overlapping moving block bootstrap with the block length $l = n_G$, thereby matching the block size with the number of observations per cluster. Gonçalves and Vogelsang (2011) showed that the naive moving block bootstrap with block length fixed (including $l = 1$) or increasing but slower than the sample size ($l^2/T \to 0$) has the same limiting distribution as the fixed-b asymptotic distribution for statistics like the CHAC statistics as long as the fixed-b limit is asymptotically pivotal. It is not obvious whether the bootstrap distribution will mimic the large-G or the fixed-G limit given that the results of Gonçalves and Vogelsang (2011) apply to both asymptotic nestings for G. Intuitively, we should expect the bootstrap to mimic the fixed-G limit when G is small but to mimic the large-G limit for large values of G. Because the small-G limit critical values approach the large-G critical values as G increases, a reasonable conjecture is that the bootstrap will mimic the small-G critical values. As the simulations results show, this is indeed the case.

Table 1 reports empirical null rejections for t_{CHAC} using the Bartlett kernel for large-G and fixed-G asymptotic critical values. Similar results were obtained for other kernels and are omitted. The results are arranged in the table to hold the amount of smoothing, $b = M_G/G$, the same across values of G (across rows). The table has two panels because of the way values of b correspond to the integer values of G. Combining the panels would result in blank table entries making it more difficult to see patterns clearly.

For the $\rho = 0$ case, rejection rates suggest that the fixed-G asymptotic critical values (right panel) work better, as expected, than the large-G critical values (left panel) when G is small. Both critical values work well when G is large. For $\rho = 0.5, 0.8$, there are three distinct patterns. First, as ρ approaches 1, over-rejections occur and become more pronounced. This is well known. Second, for a given value of G, increasing b tends to reduce over-rejections caused by positive serial correlation. This is also well known and expected. Third, for a given b, using a small number of clusters helps to reduce over-rejections. This is a benefit of using time series clustering and the finding intuitively makes sense. There is no down-weighting across autocovariances within clusters which helps accommodate stronger serial correlation. The smaller the value of G, the larger the number of observations per cluster and the greater robustness to serial correlation.

Tables 2-3 report empirical null rejections for t_{CHAC} for the Bartlett and QS kernels using bootstrap critical values. The left panels report rejection probabilities using the overlapping n_{G} block bootstrap whereas the right panels report rejections using the *i.i.d* bootstrap. The first obvious pattern is that *i.i.d.* bootstrap rejections for the Bartlett kernel in Table 2 are nearly identical to the fixed-G rejections in Table 1 even when G is large. This confirms the conjecture that the *i.i.d.* bootstrap mimics the fixed-G asymptotic distribution and is a convenient way to obtain fixed-G critical values. The performance of the block bootstrap depends on the strength of the serial correlation and the size of blocks. The middle sized blocks, corresponding to moderate values of G, can result in less over-rejections than the *i.i.d.* bootstrap. However for small values of G (large block size) we see substantial under-rejections. This is caused by the block length being too large relative to the sample size. As long as G is not too small, the block bootstrap with $l = n_{\text{G}}$ works reasonably well. If we compare rejections across the two kernels, we see that the QS kernel tends to suffer less from over-rejections than the Bartlett kernel. This is well known in the fixed-b literature.

Empirical null rejections for t_{CEWC} are reported in Table 4. Similar to the t_{CHAC} tables, the rejections are reported with the amount of smoothing (B) held fixed in each row. It is important to keep in mind that $1/B$ roughly corresponds to b for the t_{CHAC} statistics. Therefore, small (large) values of B are equivalent to large (small) bandwidths. With no serial correlation in the data ($\rho = 0$), rejections are close to zero regardless of the values of B and G. With positive serial correlation, we see that for a given value of G, increasing B (equivalent to a decrease in b) leads to over-rejections as expected. For given values of B, rejections are stable and close to 5% even for $\rho = 0.8$ regardless of the value of G. Therefore, clustering does not matter much when B is small. For large values of B, there are over-rejections that are similar in magnitude to those of t_{CHAC} with the QS kernel when $1/B$ is matched with b. This makes sense given that the $CEWC$ variance estimator is closely related to the QS CHAC estimator (see Lazarus et al. (2018)). However, the impact of G is different between t_{CEWC} and t_{CHAC}. Consider the case of $B = 3$ with $\rho = 0.8$. Increasing G leads to less over-rejections for t_{CEWC}. This is in contrast to t_{CHAC} with both the Bartlett and QS kernels where, with $b = 0.33$, increasing G tends to increase over-rejections. This increase is more pronounced for the Bartlett kernel. While the contrast between t_{CHAC} and t_{CEWC} with respect to G is difficult to understand intuitively, what is clear from Table 4 is that clustering either doesn't have an impact on null rejections for t_{CEWC} or can inflate over-rejections when serial correlation is strong. There do not appear to be benefits of clustering before smoothing for the EWC approach.

4.2 Size-Adjusted Power

It is well established in the fixed-b literature that there is a trade-off between size distortions and power with respect to the amount of smoothing used for the variance estimator. Given that clustering can reduce over-rejections for a given value of b for t_{CHAC}, one would expect there to be cost in terms of power. This is indeed the case. Tables 5 and 6 report size-adjusted power for the t_{CHAC} and t_{CEWC} statistics. Power is averaged (integrated) across $\beta \in (0, 5]$. We see the expected relationship between smoothing and power. As the bandwidth increases, power of t_{CHAC} tends to decrease. Similarly, as B decreases ($1/B$ increases), power of t_{CEWC} decreases. For a given value of b, clustering by decreasing G tends to reduce the power of the t_{CHAC} statistics. As expected, the reductions of over-rejections delivered by clustering result in reduced power. In contrast, clustering has very little impact on power of t_{CEWC} again confirming there are no benefits of clustering with EWC approach.

4.3 Weekends Missing Example

Our finite sample simulations results suggest that in the simple location model, clustering can be used to reduce over-rejections problems of t_{CHAC} caused by strong serial correlation but this reduction comes at the price of reduced power. In contrast, there is no material impact on t_{CEWC} from clustering. We now investigate a simple data structure where clustering is natural to see whether our finite sample results continue to hold. Suppose we have daily data but observations for the weekends are systematically missing (markets could be closed on the weekends). Here, the data can naturally be divided into clusters with five observations, or more generally, into clusters with a number of observations that are evenly divisible by five.

While there are multitudes of ways to generate daily data with missing weekends, we chose a simple specification. We use the DGP from the previous simulations and generate samples with 84 observations, i.e. twelve seven-day weeks. We then drop every 6^{th} and 7^{th} observation to match a missing weekends specification giving $T = 60$ observations. Given our $AR(1)$ structure, adjacent observations within a week have correlation ρ whereas adjacent end of week and beginning of week observations have correlation ρ^3. We can think of the data as being composed of 12 weeks with 5 observations per week. Using $G = 12$ becomes natural and matches the correlation structure of the data.

Tables 7 and 8 report empirical null rejections for t_{CHAC} for the Bartlett and QS kernels respectively. We no longer hold smoothing constant across values of G. Instead we report results for values of M_G (not b) in each row. This will permit us to see how lining up the choice of G with the cluster structure of the data matters. We only report results for $\rho = 0.5$ and 0.8. Results for $\rho = 0$ are not interesting in the missing weekend case.

For a given value of M_G, there is a general pattern of over-rejections becoming more severe as G increases. This intuitively makes sense because larger values of G include more down-weighting when computing the kernel HAC variance estimator. However, this pattern is not monotonic in G especially for small values of M_G. While rejections tend to increase as G increases, rejections tend to decrease when G increases from 5 to 6 and from 10 to 12. It is exactly when $G = 12$ that the clustering in the variance estimation matches the cluster structure of the data. The case of $G = 6$ has clusters with exactly two weeks of data. These results show that matching the clustering of t_{CHAC} to the cluster structure of the data can reduce over-rejections relative to the clustering that does not match the data.

Because increasing G for a given value of M_G tends to increase power, one might conjecture that moving from $G = 10$ to $G = 12$ not only reduces size distortions but does so without a cost in terms of power. This is indeed the case as Table 9 shows. The average size-adjusted power is generally increasing in G and specifically increases when G goes from 5 to 6 or from 10 to 12. Therefore, at least for our simple weekend missing data structure, it is advantageous to match the variance estimator clustering with the cluster structure of the data in terms of both size distortions and power.

Weekends missing results for t_{CEWC} are given in Tables 10 and 11 for null rejections and size-adjusted power respectively. Similar to the t_{CHAC} statistics, we see reductions in over-rejections with G going from 5 to 6 and from 10 to 12 especially for the larger values of B in the table. While null rejections are less distorted with $G = 12$ relative to $G = 10$, null rejections with $G = 60$ (no clustering) are essentially the same as $G = 12$. Furthermore, average size-adjusted power for t_{CEWC} with $G = 12$ is essentially the same as with $G = 60$. Again, there is no advantage of clustering for the EWC approach.

5 Data Dependent Bandwidths for the CHAC Approach

The finite sample simulations suggest that clustering before smoothing can be useful for the CHAC approach if a researcher wants to reduce size distortion caused by strong serial correlation or if the time series has a natural cluster structure like the missing weekends case. In this section we briefly examine the extent to which existing data dependent bandwidths methods can be used to choose the bandwidth and/or cluster size for the CHAC approach. The results we sketch here are appropriate for the large-G, fixed-n_G case. It is not obvious how to extend existing results in the literature to the fixed-G, large-n_G case and we leave such theoretical developments to future research.

We consider both the MSE-optimal (Andrews (1991)) and test-optimal (Sun et al. (2008), Sun (2014)) bandwidth approaches. For simplicity of exposition, we continue to focus on the simple location model, i.e. the case where x_t only contains an intercept regressor. We provide calculations for the widely used autoregressive lag one ($AR(1)$) plug-in method. Derivations are provided in Supplemental Appendix A.

Recall that in the large-G case, $\widehat{\Omega}^{CHAC}$ is an estimator of Ω_c, the long run variance of $\overline{v}_g$. When the time series is covariance stationary, $n_{\mathrm{G}}^{-1}\widehat{\Omega}^{CHAC}$ is an estimator of Ω, the long run variance of v_t because $\Omega_c = n_{\mathrm{G}}\Omega$. We apply existing bandwidth results to $n_{\mathrm{G}}^{-1}\widehat{\Omega}^{CHAC}$.

According to the $AR(1)$ plug-in approach, v_t is approximated by the $AR(1)$ process $v_t = \rho v_{t-1} + \varepsilon_t$. It then follows from Amemiya and Wu (1972) that $\overline{v}_g$ is an $ARMA(1,1)$ process. We show in Supplemental Appendix A that

$$\Omega_c^{(1)} = \Omega^{(1)}, \tag{1}$$

$$\Omega_c^{(2)} = \Omega^{(2)} \frac{(1+\rho^{n_{\mathrm{G}}})(1-\rho)}{(1-\rho^{n_{\mathrm{G}}})(1+\rho)}. \tag{2}$$

Here, $\Omega_c^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \Gamma_{cj}$ and $\Omega^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j$, where Γ_{cj} and Γ_j are the autocovariance functions of $\overline{v}_g$ and v_t respectively.

5.1 MSE-optimal Bandwidth

Following Andrews (1991):

$$MSE\left(\frac{1}{n_{\mathrm{G}}}\widehat{\Omega}^{CHAC}\right) = \frac{1}{n_G^2} MSE\left(\widehat{\Omega}^{CHAC}\right) \approx \frac{1}{n_G^2}\left[\left(\frac{k_q \Omega_c^{(q)}}{M_G}\right)^2 + 2c_2 \Omega_c^2 \frac{M_G}{G}\right],$$

where M_G is the bandwidth and $q \in [0,\infty)$ is the largest integer such that $k_q = lim_{x\to 0}\frac{1-\mathcal{K}(x)}{|x|^q} < \infty$, and $c_2 = \int \mathcal{K}(x)^2 dx$. Replacing Ω_c with $n_{\mathrm{G}}\Omega$, plugging in for $\Omega_c^{(q)}$ using (1) and (2), and using $T = n_{\mathrm{G}}G$ gives

$$MSE\left(\frac{1}{n_{\mathrm{G}}}\widehat{\Omega}^{CHAC}\right) = \begin{cases} \left(\frac{k_1\Omega^{(1)}}{n_{\mathrm{G}}M_G}\right)^2 + 2c_1 T^{-1}\Omega^2 n_{\mathrm{G}} M_G & q=1 \\ \left(\frac{k_2\Omega^{(2)}}{n_{\mathrm{G}}M_G^2}\frac{(1+\rho^{n_{\mathrm{G}}})(1-\rho)}{(1-\rho^{n_{\mathrm{G}}})(1+\rho)}\right)^2 + 2c_2 T^{-1}\Omega^2 n_{\mathrm{G}} M_G & q=2. \end{cases}$$

In the case of $q = 1$ (Bartlett kernel), the MSE formula depends on n_{G} and M_G only through the product $n_{\mathrm{G}}M_G$. Therefore, minimization of the MSE can only determine the product but not n_{G} and M_G individually. Notice also in the $q = 1$ case that if we replace $n_{\mathrm{G}}M_G$ with M_T we obtain the MSE formula for the case of no clustering. Therefore, if we let M_T^* denote the MSE-optimal bandwidth for the case of no clustering, then it immediately follows for a given cluster size, n_{G}, that $n_{\mathrm{G}}M_G^* = M_T^*$ or $M_G^* = M_T^*/n_{\mathrm{G}}$.

A practical recommendation for the Bartlett kernel can be made from this result. First, compute M_T^*, the MSE-optimal bandwidth without clustering. Once the cluster size has been chosen, perhaps based on the cluster structure of the data, use the bandwidth $M_G^* = M_T^*/n_{\text{G}}$ for the CHAC estimator.

The case of $q = 2$ (QS kernel) is more complicated because of the $\frac{(1+\rho^{n_{\text{G}}})(1-\rho)}{M_G(1-\rho^{n_{\text{G}}})(1+\rho)}$ term in the MSE formula. We show in Supplemental Appendix A that, for the empirically relevant case of positive autocorrelation ($\rho > 0$), the MSE minimization has a corner solution with $n_{\text{G}}^* = 1$ in which case no clustering is used and the usual bandwidth formula for M is obtained. Should an empirical researcher decide to use a cluster size different from 1, the MSE-optimal bandwidth can be computed as

$$M_G^* = M_T^* \left[\left(\frac{(1+\widehat{\rho}^{n_{\text{G}}})(1-\widehat{\rho})}{(1-\widehat{\rho}^{n_{\text{G}}})(1+\widehat{\rho})} \right)^2 \frac{1}{{n_{\text{G}}}^3} \right]^{1/5},$$

where $\widehat{\rho}$ is the same estimated value of ρ used to calculate M_T^*.

5.2 Test-optimal Bandwidth

Following Sun et al. (2008) (SPJ), the test-optimal bandwidth minimizes the SPJ objective function, which is a weighted average of the approximate type I and the type II errors of the CHAC test statistic. Without going into details, the SPJ objective function shares the same essential features as the MSE objective function with respect to n_{G} and M_G. In the $q = 1$ case, the SPJ objective function depends on n_{G} and M_G only through the product $n_{\text{G}} M_G$. In the $q = 2$ case, $n_{\text{G}}^* = 1$ is obtained as a corner solution. Should an empirical researcher decide to use a given cluster size, n_{G}, the test-optimal bandwidths are given by

$$M_G^* = \begin{cases} M_T^*/n_{\text{G}} & q = 1 \\ M_T^* \left(\frac{1}{{n_{\text{G}}}^2} \frac{(1+\widehat{\rho}^{n_{\text{G}}})(1-\widehat{\rho})}{(1-\widehat{\rho}^{n_{\text{G}}})(1+\widehat{\rho})} \right)^{1/3} & q = 2, \end{cases}$$

where M_T^* is the test-optimal bandwidth without clustering and $\widehat{\rho}$ is the same estimated value of ρ used to calculate M_T^*. For the derivation, see Supplemental Appendix A.

6 Conclusion

This paper proposes a long run variance estimator for conducting inference in time series regression models that combines the nonparametric approach with a cluster approach. The basic idea is to divide the time periods into non-overlapping clusters. The long run variance estimator is constructed by first aggregating within clusters and then kernel smoothing across clusters or applying the nonparametric series method to the clusters with Type II discrete cosine transform. We develop an asymptotic theory for test statistics based on these "smoothed-clustered" long run variance estimators. We derive asymptotic results holding the number of clusters fixed and also treating the clusters as increasing with the sample size. For the kernel approach, these two asymptotic limits are different and nonstandard whereas for the cosine series approach, the two limits are the same and have standard t or F distributions. When clustering before kernel smoothing, we find that the “fixed-number-of-clusters” asymptotic approximation works well whether the number of clusters is small or large. The moving blocks bootstrap (including the naive *i.i.d.* bootstrap) is a convenient way to obtain critical values that are asymptotically equivalent to critical values from the “fixed-number-of-clusters” limiting distribution.

Finite sample simulations for the simple location model suggest that clustering before kernel smoothing can reduce over-rejections caused by strong serial correlation although at a cost of power as typical. In contrast, clustering before using the cosine series approach does not tend to reduce over-rejection problems. When there is a natural way of clustering, such as weekly data with missing weekends, then clustering can reduce over-rejection problems with some potential gains in power for the kernel approach. In contrast, there are no gains to clustering for the cosine series approach.

For the kernel approach we analyze data dependent bandwidth approaches configured for the $AR(1)$ plug-in approach. For the Bartlett kernel, both MSE-optimal and test-optimal approaches only determine the product, $n_{\mathrm{G}}M_G$, and not the cluster size and kernel bandwidth separately. For kernels in the same class as the QS kernels, both bandwidth approaches give $n_{\mathrm{G}} = 1$ in which case no clustering is used. An empirical researcher using the Bartlett kernel should use clustering if either there is a desire to reduce over-rejections caused by strong serial correlation or there is a natural cluster structure to the data. For the QS kernel clustering has no distinct advantage except when the data has a natural cluster structure. Once the number of clusters has been chosen, data dependent bandwidths can be computed as a simple functions of the non-clustered data dependent bandwidths.

References

Amemiya, T. and Wu, R. Y.: (1972), The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association* **67**, 628–632.

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

Bester, C. A., Conley, T. G. and Hansen, C. B.: (2011), Inference with dependent data using cluster covariance estimators, *Journal of Econometrics* **165**, 137–151.

Djogbenou, A. A., MacKinnon, J. G. and Nielsen, M. Ø.: (2019), Asymptotic theory and wild bootstrap inference with clustered errors, *Journal of Econometrics* **212**, 393–412.

Driscoll, J. C. and Kraay, A. C.: (1998), Consistent covariance matrix estimation with spatially dependent panel data, *The Review of Economics and Statistics* **80**, 549–560.

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

Grenander, U. and Rosenblatt, M.: (1953), Statistical spectral analysis of time series arising from stationary stochastic processes, *Annals of Mathematical Statistics* **24**, 537–558.

Hansen, B. E. and Lee, S.: (2019), Asymptotic theory for clustered samples, *Journal of Econometrics* **210**, 268–290.

Hansen, C. B.: (2007), Asymptotic properties of a robust variance matrix estimator for panel data when T is large, *Journal of Econometrics* **141**(2), 597–620.

Ibragimov, R. and Müller, U. K.: (2010), t -statistic based correlation and heterogeneity robust inference, *Journal of Business and Economic Statistics* **28**(4), 453–468.

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

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

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

Newey, W. K. and West, K. D.: (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Phillips, P. C.: (2005), HAC estimation by automated regression, *Econometric Theory* **21**(1), 116–142.

Phillips, P. C. and Durlauf, S. N.: (1986), Multiple time series regression with integrated processes, *Review of Economic Studies* **53**, 473–95.

Sun, Y.: (2013), A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator, *The Econometrics Journal* **16**(1), 1–26.

Sun, Y.: (2014), 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. and Jin, S.: (2008), Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* **76**, 175–194.

Vogelsang, T. J.: (2012), Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects, *Journal of Econometrics* **166**, 303–319.

White, H. (ed.): (2001), *Asymptotic theory for econometricians*, Academic Press. Revised Edition.

Table 1: CHAC: Empirical Null Rejections Using Asymptotic Critical Values, Bartlett Kernel

ρ	$\frac{M_G}{G}$	$G \to \infty$				Fixed G			
		values of G				values of G			
		6	12	30	60	6	12	30	60
0	0.17	0.071	0.056	0.049	0.048	0.049	0.050	0.049	0.049
	0.50	0.072	0.056	0.051	0.050	0.048	0.050	0.051	0.051
	0.83	0.067	0.052	0.046	0.048	0.050	0.050	0.048	0.050
	1.00	0.067	0.052	0.048	0.047	0.050	0.049	0.048	0.048
0.5	0.17	0.092	0.074	0.075	0.075	0.062	0.069	0.075	0.077
	0.50	0.083	0.070	0.070	0.069	0.058	0.065	0.068	0.069
	0.83	0.079	0.067	0.067	0.067	0.057	0.065	0.069	0.070
	1.00	0.080	0.068	0.068	0.068	0.057	0.065	0.069	0.070
0.8	0.17	0.158	0.151	0.153	0.153	0.113	0.141	0.153	0.155
	0.50	0.122	0.115	0.115	0.115	0.089	0.107	0.114	0.115
	0.83	0.118	0.113	0.112	0.112	0.094	0.109	0.114	0.116
	1.00	0.119	0.114	0.114	0.114	0.094	0.110	0.115	0.116

CHAC: Empirical Null Rejections Using Asymptotic Critical Values, Bartlett Kernel (cont'd)

ρ	$\frac{M_G}{G}$	$G \to \infty$						Fixed G					
		values of G						values of G					
		3	6	12	15	30	60	3	6	12	15	30	60
0	0.33	0.135	0.071	0.055	0.052	0.048	0.048	0.050	0.049	0.050	0.049	0.047	0.048
	0.67	0.130	0.069	0.053	0.051	0.049	0.049	0.048	0.049	0.050	0.048	0.049	0.051
	1.00	0.132	0.067	0.052	0.049	0.048	0.047	0.048	0.050	0.049	0.048	0.048	0.048
0.5	0.33	0.145	0.086	0.069	0.070	0.068	0.068	0.054	0.060	0.064	0.066	0.068	0.068
	0.67	0.141	0.083	0.068	0.068	0.066	0.066	0.052	0.058	0.064	0.066	0.067	0.068
	1.00	0.142	0.080	0.068	0.068	0.068	0.068	0.052	0.057	0.065	0.067	0.069	0.070
0.8	0.33	0.171	0.125	0.120	0.120	0.119	0.119	0.064	0.093	0.113	0.114	0.118	0.120
	0.67	0.163	0.119	0.113	0.113	0.114	0.113	0.063	0.091	0.108	0.110	0.114	0.116
	1.00	0.166	0.119	0.114	0.114	0.114	0.114	0.063	0.094	0.110	0.112	0.115	0.116

Note: Table 1 reports empirical null rejection rates for the Bartlett kernel CHAC approach based on simulated asymptotic critical values with $b = M_G/G$ fixed. The left panel contains rejection rates for $G \to \infty$ with n_G-fixed case and the right panel contains rejection rates for $n_G \to \infty$ with G-fixed.

Table 2: CHAC: Empirical Null Rejections Using Bootstrap Critical Values, Bartlett Kernel

		G block bootstrap				i.i.d. bootstrap			
ρ	$\frac{M_G}{G}$	values of G				values of G			
		6	12	30	60	6	12	30	60
0	0.17	0.036	0.043	0.049	0.051	0.049	0.050	0.051	0.051
	0.50	0.038	0.043	0.049	0.051	0.049	0.051	0.051	0.051
	0.83	0.037	0.043	0.048	0.050	0.048	0.050	0.049	0.050
	1.00	0.037	0.044	0.049	0.050	0.048	0.050	0.050	0.050
0.5	0.17	0.044	0.062	0.074	0.079	0.063	0.071	0.077	0.079
	0.50	0.042	0.059	0.068	0.070	0.059	0.065	0.069	0.070
	0.83	0.041	0.057	0.068	0.070	0.060	0.065	0.070	0.070
	1.00	0.041	0.057	0.068	0.070	0.060	0.064	0.069	0.070
0.8	0.17	0.075	0.125	0.153	0.158	0.116	0.144	0.156	0.158
	0.50	0.065	0.096	0.113	0.117	0.090	0.109	0.116	0.117
	0.83	0.066	0.097	0.112	0.116	0.094	0.108	0.115	0.116
	1.00	0.066	0.099	0.113	0.116	0.094	0.110	0.115	0.116

CHAC: Empirical Null Rejections Using Bootstrap Critical Values, Bartlett Kernel (cont'd)

		G block bootstrap						i.i.d. bootstrap					
ρ	$\frac{M_G}{G}$	values of G						values of G					
		3	6	12	15	30	60	3	6	12	15	30	60
0	0.33	0.031	0.036	0.045	0.044	0.048	0.051	0.052	0.049	0.050	0.051	0.050	0.051
	0.67	0.032	0.037	0.043	0.045	0.047	0.050	0.049	0.049	0.050	0.049	0.050	0.050
	1.00	0.032	0.037	0.044	0.045	0.049	0.050	0.049	0.048	0.050	0.050	0.050	0.050
0.5	0.33	0.032	0.043	0.058	0.061	0.069	0.070	0.055	0.060	0.066	0.067	0.070	0.070
	0.67	0.030	0.043	0.059	0.062	0.066	0.068	0.054	0.060	0.063	0.066	0.068	0.068
	1.00	0.030	0.041	0.057	0.062	0.068	0.070	0.054	0.060	0.064	0.067	0.069	0.070
0.8	0.33	0.030	0.065	0.102	0.107	0.118	0.121	0.065	0.094	0.114	0.115	0.120	0.121
	0.67	0.030	0.067	0.096	0.104	0.112	0.116	0.064	0.092	0.108	0.111	0.115	0.116
	1.00	0.030	0.066	0.099	0.104	0.113	0.116	0.064	0.094	0.110	0.111	0.115	0.116

Note: Table 2 reports empirical null rejection rates for the Bartlett kernel CHAC approach based on the overlapping n_G block bootstrap (left panel) and the i.i.d. bootstrap (right panel) critical values. The nominal level is 5% and $T = 60$.

Table 3: CHAC: Empirical Null Rejections Using Bootstrap Critical Value, QS Kernel

ρ	$\frac{M_G}{G}$	G block bootstrap, values of G: 6	12	30	60	i.i.d. bootstrap, values of G: 6	12	30	60
0	0.17	0.036	0.045	0.050	0.052	0.050	0.050	0.052	0.052
	0.50	0.044	0.046	0.052	0.051	0.053	0.050	0.052	0.051
	0.83	0.045	0.046	0.050	0.050	0.051	0.049	0.049	0.050
	1.00	0.045	0.046	0.048	0.049	0.051	0.049	0.048	0.049
0.5	0.17	0.043	0.057	0.058	0.059	0.063	0.062	0.060	0.059
	0.50	0.046	0.052	0.055	0.055	0.056	0.055	0.056	0.055
	0.83	0.046	0.051	0.052	0.052	0.055	0.053	0.052	0.052
	1.00	0.046	0.051	0.053	0.053	0.054	0.053	0.052	0.053
0.8	0.17	0.073	0.097	0.103	0.104	0.112	0.113	0.105	0.104
	0.50	0.054	0.062	0.065	0.067	0.069	0.066	0.066	0.067
	0.83	0.051	0.057	0.061	0.062	0.064	0.061	0.062	0.062
	1.00	0.052	0.059	0.060	0.062	0.063	0.061	0.061	0.062

CHAC: Empirical Null Rejections Using Bootstrap Critical Values, QS Kernel (cont'd)

ρ	$\frac{M_G}{G}$	G block bootstrap, values of G: 3	6	12	15	30	60	i.i.d. bootstrap, values of G: 3	6	12	15	30	60
0	0.33	0.031	0.040	0.045	0.049	0.051	0.051	0.052	0.051	0.051	0.052	0.053	0.051
	0.67	0.030	0.045	0.046	0.049	0.049	0.050	0.047	0.052	0.050	0.052	0.050	0.050
	1.00	0.029	0.045	0.046	0.047	0.048	0.049	0.047	0.051	0.049	0.049	0.048	0.049
0.5	0.33	0.032	0.043	0.052	0.053	0.054	0.056	0.055	0.057	0.055	0.057	0.056	0.056
	0.67	0.028	0.045	0.051	0.052	0.053	0.053	0.053	0.054	0.054	0.053	0.052	0.053
	1.00	0.027	0.046	0.051	0.051	0.053	0.053	0.050	0.054	0.053	0.054	0.052	0.053
0.8	0.33	0.029	0.058	0.069	0.070	0.072	0.074	0.064	0.079	0.075	0.075	0.074	0.074
	0.67	0.028	0.052	0.058	0.060	0.063	0.062	0.065	0.064	0.062	0.063	0.062	0.062
	1.00	0.027	0.052	0.059	0.060	0.060	0.062	0.061	0.063	0.061	0.062	0.061	0.062

Note: Table 3 reports empirical null rejection rates for the QS kernel CHAC approach based on the overlapping n_G block bootstrap (left panel) and the i.i.d. bootstrap (right panel) critical values. The nominal level is 5% and $T = 60$.

Table 4: CEWC: Empirical Null Rejections Using t_B Critical Values

ρ	B	values of G										
		2	3	4	5	6	10	12	15	20	30	60
0	1	0.050	0.049	0.052	0.050	0.053	0.052	0.051	0.053	0.051	0.053	0.053
	2		0.050	0.049	0.048	0.050	0.051	0.050	0.051	0.048	0.050	0.049
	3			0.051	0.051	0.049	0.053	0.050	0.051	0.050	0.050	0.050
	4				0.051	0.052	0.051	0.049	0.049	0.051	0.050	0.050
	5					0.049	0.050	0.049	0.050	0.050	0.048	0.050
	6						0.050	0.051	0.052	0.049	0.051	0.050
0.5	1	0.049	0.051	0.050	0.053	0.052	0.051	0.048	0.051	0.050	0.050	0.050
	2		0.053	0.053	0.052	0.051	0.052	0.052	0.050	0.051	0.051	0.051
	3			0.055	0.058	0.056	0.056	0.054	0.055	0.055	0.054	0.054
	4				0.061	0.061	0.058	0.058	0.056	0.057	0.056	0.055
	5					0.062	0.061	0.060	0.058	0.058	0.057	0.055
	6						0.065	0.064	0.062	0.060	0.060	0.058
0.8	1	0.054	0.055	0.051	0.054	0.051	0.051	0.051	0.052	0.052	0.051	0.051
	2		0.064	0.064	0.063	0.063	0.060	0.059	0.058	0.059	0.058	0.058
	3			0.080	0.078	0.079	0.076	0.074	0.074	0.073	0.073	0.072
	4				0.092	0.096	0.089	0.087	0.087	0.086	0.085	0.085
	5					0.114	0.107	0.103	0.101	0.098	0.097	0.096
	6						0.125	0.121	0.118	0.117	0.114	0.113

Note: Table 4 reports empirical null rejection rates for the CEWC approach. The nominal level is 5% and $T = 60$.

Table 5: CHAC: Average Size-adjusted Power for $\beta \in (0, 5]$

ρ	$\frac{M_G}{G}$	values of G: 6	12	30	60
		Bartlett Kernel			
0	0.17	0.938	0.942	0.944	0.944
	0.50	0.930	0.934	0.935	0.936
	0.83	0.928	0.933	0.934	0.934
	1.00	0.928	0.933	0.935	0.935
0.5	0.17	0.874	0.884	0.886	0.886
	0.50	0.858	0.865	0.867	0.867
	0.83	0.856	0.863	0.863	0.865
	1.00	0.856	0.863	0.865	0.865
0.8	0.17	0.695	0.711	0.710	0.710
	0.50	0.638	0.649	0.650	0.649
	0.83	0.632	0.642	0.644	0.645
	1.00	0.632	0.643	0.644	0.645
		QS Kernel			
0	0.17	0.937	0.939	0.939	0.939
	0.50	0.907	0.911	0.910	0.910
	0.83	0.889	0.893	0.891	0.892
	1.00	0.880	0.885	0.886	0.887
0.5	0.17	0.874	0.878	0.879	0.879
	0.50	0.815	0.819	0.820	0.820
	0.83	0.776	0.783	0.786	0.786
	1.00	0.764	0.767	0.770	0.770
0.8	0.17	0.689	0.696	0.697	0.697
	0.50	0.555	0.567	0.570	0.570
	0.83	0.500	0.511	0.513	0.513
	1.00	0.475	0.483	0.486	0.485

ρ	$\frac{M_G}{G}$	values of G: 3	6	12	15	30	60
		Bartlett Kernel					
0	0.33	0.901	0.932	0.937	0.938	0.939	0.940
	0.67	0.898	0.928	0.933	0.934	0.934	0.934
	1.00	0.898	0.928	0.933	0.934	0.935	0.935
0.5	0.33	0.804	0.865	0.872	0.875	0.876	0.875
	0.67	0.797	0.855	0.863	0.864	0.863	0.863
	1.00	0.797	0.856	0.863	0.863	0.865	0.865
0.8	0.33	0.546	0.660	0.672	0.674	0.673	0.673
	0.67	0.514	0.630	0.642	0.644	0.645	0.644
	1.00	0.514	0.632	0.643	0.646	0.644	0.645
		QS Kernel					
0	0.33	0.900	0.923	0.925	0.924	0.924	0.924
	0.67	0.889	0.897	0.900	0.899	0.900	0.900
	1.00	0.873	0.880	0.885	0.886	0.886	0.887
0.5	0.33	0.803	0.848	0.849	0.850	0.850	0.850
	0.67	0.780	0.795	0.798	0.801	0.800	0.800
	1.00	0.756	0.764	0.767	0.765	0.770	0.770
0.8	0.33	0.544	0.618	0.627	0.625	0.627	0.626
	0.67	0.493	0.530	0.537	0.532	0.536	0.536
	1.00	0.447	0.475	0.483	0.485	0.486	0.485

Note: Table 5 reports average size adjusted power for the Bartlett and QS kernels CHAC approach. The nominal level is 5% and $T = 60$. The alternative hypothesis is $\beta \in (0, 5]$.

Table 6: CEWC: Average Size-adjusted Power for $\beta \in (0, 5]$

ρ	B	values of G										
		2	3	4	5	6	10	12	15	20	30	60
0	1	0.739	0.745	0.724	0.736	0.719	0.726	0.732	0.727	0.733	0.724	0.715
	2		0.901	0.902	0.903	0.902	0.901	0.901	0.901	0.902	0.901	0.903
	3			0.924	0.924	0.925	0.922	0.924	0.924	0.924	0.924	0.924
	4				0.932	0.932	0.932	0.933	0.933	0.932	0.933	0.932
	5					0.938	0.937	0.937	0.937	0.937	0.938	0.937
	6						0.940	0.939	0.939	0.940	0.939	0.939
0.5	1	0.529	0.524	0.525	0.507	0.517	0.521	0.536	0.522	0.523	0.521	0.526
	2		0.804	0.805	0.808	0.808	0.807	0.807	0.810	0.808	0.808	0.807
	3			0.852	0.848	0.853	0.850	0.852	0.850	0.850	0.850	0.849
	4				0.866	0.865	0.866	0.866	0.867	0.865	0.865	0.865
	5					0.874	0.876	0.876	0.876	0.875	0.877	0.876
	6						0.881	0.880	0.880	0.881	0.881	0.881
0.8	1	0.247	0.253	0.267	0.256	0.265	0.263	0.263	0.258	0.261	0.263	0.262
	2		0.546	0.546	0.544	0.549	0.545	0.550	0.548	0.549	0.550	0.550
	3			0.641	0.636	0.637	0.638	0.640	0.639	0.639	0.640	0.640
	4				0.670	0.674	0.677	0.677	0.678	0.677	0.678	0.678
	5					0.695	0.696	0.698	0.699	0.699	0.698	0.698
	6						0.710	0.711	0.714	0.714	0.715	0.714

Note: Table 6 reports average size adjusted power for the CEWC approach. The nominal level is 5% and $T = 60$. The alternative hypothesis is $\beta \in (0, 5]$.

Table 7: CHAC: Empirical Null Rejections Using Bootstrap Critical Values, Weekends Missing ($T = 60$ out of 84) , Bartlett Kernel

		overlapping G block bootstrap critical value										i.i.d. bootstrap critical value									
ρ	M_G	values of G										values of G									
		2	3	4	5	6	10	12	15	30	60	2	3	4	5	6	10	12	15	30	60
0.5	1	0.011	0.031	0.031	0.040	0.036	0.058	0.049	0.077	0.132	0.208	0.045	0.053	0.049	0.058	0.052	0.070	0.058	0.086	0.134	0.208
	2	0.011	0.032	0.032	0.038	0.039	0.054	0.048	0.060	0.088	0.137	0.045	0.053	0.050	0.057	0.052	0.063	0.053	0.065	0.089	0.137
	3		0.032	0.032	0.040	0.040	0.053	0.047	0.056	0.076	0.107		0.053	0.050	0.057	0.054	0.061	0.054	0.063	0.076	0.107
	4			0.032	0.041	0.038	0.053	0.048	0.056	0.068	0.092			0.050	0.058	0.054	0.061	0.054	0.062	0.071	0.092
	5				0.041	0.039	0.053	0.049	0.057	0.064	0.083				0.058	0.053	0.059	0.056	0.063	0.067	0.083
	6					0.039	0.053	0.049	0.058	0.062	0.077					0.053	0.061	0.057	0.064	0.065	0.077
	10						0.054	0.048	0.058	0.062	0.067						0.060	0.056	0.062	0.065	0.067
0.8	1	0.010	0.029	0.034	0.047	0.053	0.113	0.122	0.183	0.317	0.455	0.051	0.062	0.061	0.079	0.080	0.137	0.139	0.195	0.322	0.455
	2	0.010	0.028	0.036	0.045	0.050	0.083	0.084	0.115	0.203	0.329	0.051	0.061	0.063	0.071	0.071	0.097	0.099	0.125	0.207	0.329
	3		0.028	0.035	0.045	0.050	0.076	0.076	0.095	0.154	0.256		0.061	0.063	0.071	0.070	0.087	0.087	0.104	0.158	0.256
	4			0.035	0.045	0.051	0.073	0.073	0.086	0.129	0.209			0.063	0.072	0.072	0.085	0.082	0.093	0.133	0.209
	5				0.045	0.051	0.075	0.073	0.084	0.116	0.180				0.072	0.073	0.086	0.081	0.091	0.117	0.180
	6					0.051	0.075	0.072	0.085	0.107	0.160					0.073	0.087	0.081	0.092	0.108	0.160
	10						0.075	0.075	0.085	0.092	0.118						0.088	0.083	0.090	0.094	0.118

Table 8: CHAC: Empirical Null Rejections Using Bootstrap Critical Values, Weekends Missing ($T = 60$ out of 84), QS Kernel

		overlapping G block bootstrap critical value										i.i.d. bootstrap critical value									
ρ	M_G	values of G										values of G									
		2	3	4	5	6	10	12	15	30	60	2	3	4	5	6	10	12	15	30	60
0.5	1	0.011	0.032	0.032	0.040	0.036	0.056	0.048	0.073	0.119	0.186	0.045	0.053	0.049	0.058	0.053	0.068	0.057	0.080	0.120	0.186
	2	0.011	0.031	0.034	0.041	0.039	0.051	0.048	0.052	0.074	0.112	0.045	0.054	0.050	0.056	0.051	0.059	0.052	0.058	0.075	0.112
	3		0.030	0.037	0.042	0.041	0.050	0.049	0.052	0.060	0.083		0.054	0.050	0.055	0.050	0.057	0.054	0.055	0.062	0.083
	4			0.038	0.043	0.042	0.049	0.049	0.051	0.053	0.070			0.052	0.053	0.052	0.055	0.054	0.055	0.055	0.070
	5				0.044	0.044	0.051	0.047	0.052	0.052	0.064				0.054	0.053	0.057	0.052	0.055	0.054	0.064
	6					0.044	0.053	0.049	0.051	0.052	0.059					0.052	0.057	0.054	0.054	0.054	0.059
	10						0.054	0.052	0.053	0.053	0.054						0.056	0.054	0.054	0.055	0.054
0.8	1	0.010	0.029	0.035	0.046	0.051	0.104	0.112	0.165	0.283	0.417	0.051	0.062	0.062	0.080	0.079	0.128	0.128	0.176	0.288	0.417
	2	0.010	0.028	0.039	0.044	0.048	0.067	0.070	0.090	0.163	0.280	0.051	0.061	0.059	0.065	0.066	0.080	0.080	0.097	0.166	0.280
	3		0.028	0.042	0.047	0.047	0.058	0.060	0.069	0.114	0.202		0.062	0.061	0.062	0.061	0.067	0.067	0.076	0.117	0.202
	4			0.042	0.049	0.051	0.054	0.056	0.062	0.090	0.161			0.063	0.061	0.063	0.062	0.064	0.067	0.091	0.161
	5				0.050	0.051	0.055	0.056	0.060	0.078	0.134				0.062	0.063	0.062	0.061	0.066	0.080	0.134
	6					0.051	0.057	0.056	0.058	0.071	0.113					0.062	0.063	0.061	0.063	0.074	0.113
	10						0.059	0.059	0.058	0.061	0.079						0.064	0.063	0.062	0.064	0.079

Note: Tables 7 and 8 report empirical null rejection rates for the weekends missing case for the CHAC approach. The rejection rates are computed based on the overlapping n_G moving block bootstrap and i.i.d. bootstrap critical values. The nominal level is 5%.

Table 9: CHAC: Average Size-adjusted Power for $\beta \in (0, 5]$, Weekends Missing ($T = 60$ out of 84)

ρ	M	values of G									
		2	3	4	5	6	10	12	15	30	60
							Barlett Kernel				
0.5	1	0.587	0.821	0.868	0.881	0.888	0.899	0.902	0.903	0.907	0.908
	2	0.587	0.811	0.856	0.868	0.879	0.892	0.898	0.900	0.905	0.907
	3		0.811	0.855	0.862	0.871	0.886	0.893	0.896	0.903	0.906
	4			0.855	0.864	0.868	0.883	0.888	0.891	0.900	0.905
	5				0.864	0.870	0.879	0.883	0.887	0.899	0.903
	6					0.870	0.875	0.880	0.883	0.896	0.903
	10						0.875	0.878	0.877	0.887	0.899
0.8	1	0.283	0.598	0.699	0.728	0.743	0.767	0.775	0.777	0.782	0.785
	2	0.283	0.566	0.664	0.697	0.715	0.750	0.760	0.765	0.777	0.783
	3		0.566	0.661	0.677	0.697	0.729	0.743	0.753	0.772	0.779
	4			0.661	0.676	0.690	0.716	0.730	0.736	0.766	0.776
	5				0.676	0.692	0.705	0.721	0.726	0.759	0.774
	6					0.692	0.700	0.714	0.717	0.754	0.773
	10						0.701	0.705	0.701	0.726	0.760
							QS Kernel				
0.5	1	0.587	0.820	0.867	0.880	0.887	0.898	0.902	0.903	0.906	0.908
	2	0.587	0.791	0.834	0.848	0.865	0.884	0.892	0.897	0.903	0.907
	3		0.762	0.804	0.819	0.838	0.868	0.877	0.887	0.900	0.905
	4			0.774	0.799	0.816	0.851	0.864	0.875	0.896	0.903
	5				0.781	0.798	0.836	0.850	0.864	0.891	0.902
	6					0.786	0.819	0.836	0.853	0.886	0.900
	10						0.780	0.798	0.814	0.863	0.892
0.8	1	0.283	0.595	0.694	0.725	0.742	0.763	0.771	0.774	0.780	0.785
	2	0.283	0.528	0.605	0.652	0.682	0.732	0.747	0.758	0.774	0.781
	3		0.483	0.532	0.581	0.622	0.699	0.716	0.735	0.769	0.777
	4			0.496	0.541	0.565	0.662	0.688	0.710	0.758	0.774
	5				0.508	0.531	0.624	0.656	0.686	0.748	0.773
	6					0.511	0.594	0.628	0.663	0.736	0.769
	10						0.509	0.541	0.572	0.687	0.748

Note: Table 9 reports average size adjusted power for the weekends missing case for the Bartlett and QS kernels CHAC approach. The nominal level is 5%, and the alternative hypothesis is $\beta \in (0, 5]$.

Table 10: CEWC: Empirical Null Rejections Using t_B Critical Values, Weekends Missing ($T = 60$ out of 84)

ρ	B	values of G										
		2	3	4	5	6	10	12	15	20	30	60
0.5	1	0.043	0.050	0.052	0.049	0.048	0.048	0.048	0.049	0.049	0.047	0.047
	2		0.052	0.051	0.053	0.050	0.052	0.050	0.050	0.052	0.050	0.050
	3			0.049	0.053	0.051	0.054	0.052	0.052	0.053	0.053	0.053
	4				0.056	0.052	0.057	0.053	0.054	0.054	0.054	0.053
	5					0.052	0.059	0.052	0.054	0.057	0.054	0.054
	6						0.059	0.054	0.053	0.055	0.055	0.053
0.8	1	0.052	0.050	0.053	0.057	0.053	0.053	0.053	0.053	0.053	0.052	0.052
	2		0.059	0.058	0.057	0.058	0.056	0.054	0.056	0.055	0.055	0.055
	3			0.061	0.065	0.061	0.061	0.059	0.060	0.059	0.059	0.059
	4				0.081	0.071	0.072	0.067	0.069	0.068	0.067	0.067
	5					0.080	0.082	0.077	0.079	0.077	0.076	0.075
	6						0.096	0.087	0.088	0.086	0.085	0.083

Note: Table 10 reports empirical null rejection rates for the weekends missing case for the CEWC approach. The nominal level is 5%.

Table 11: CEWC: Average Size-adjusted Power for $\beta \in (0, 5]$, Weekends Missing ($T = 60$ out of 84)

ρ	B	values of G										
		2	3	4	5	6	10	12	15	20	30	60
0.5	1	0.587	0.554	0.532	0.564	0.572	0.569	0.569	0.561	0.567	0.582	0.582
	2		0.821	0.823	0.822	0.826	0.822	0.826	0.827	0.822	0.824	0.826
	3			0.868	0.867	0.866	0.866	0.864	0.864	0.864	0.864	0.863
	4				0.881	0.880	0.879	0.879	0.881	0.879	0.879	0.879
	5					0.888	0.886	0.888	0.888	0.886	0.887	0.886
	6						0.892	0.892	0.894	0.893	0.892	0.892
0.8	1	0.283	0.300	0.288	0.259	0.285	0.282	0.286	0.282	0.283	0.283	0.283
	2		0.598	0.595	0.603	0.599	0.597	0.600	0.602	0.603	0.602	0.603
	3			0.699	0.695	0.696	0.697	0.699	0.698	0.697	0.697	0.699
	4				0.728	0.726	0.728	0.730	0.727	0.730	0.729	0.729
	5					0.743	0.746	0.744	0.747	0.744	0.745	0.745
	6						0.754	0.755	0.756	0.756	0.755	0.755

Note: Table 11 reports average size adjusted power for the CEWC approach. The nominal level is 5%, and the alternative hypothesis is $\beta \in (0, 5]$.

Appendix

In this appendix we provide proofs for Theorems 1 and 2. Theorem 1 provides asymptotic results for the $G \to \infty$ with n_G fixed case. The proof closely follows proofs in the existing literature (Sun (2013) and Vogelsang (2012)). Here we provide key arguments for completeness.

Proof of Theorem 1(a): Under Assumption A, the following result is straightforward:

$$\sqrt{G}\left(\widehat{\beta}-\beta\right)=\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}G^{-1/2}\sum_{g=1}^{G}\overline{v}_g \Rightarrow Q_c^{-1}\Lambda_c\mathcal{W}_k(1). \tag{3}$$

□

Proof of Theorem 1(b): When the kernel function satisfies relevant conditions that the kernel function is symmetric, piecewise smooth with $\mathcal{K}(0)=1$ and $\int_0^\infty \mathcal{K}(x)xdx<\infty$, the kernel function $\mathcal{K}_b(r,s)=\mathcal{K}((r-s)/b)$ on $[0,1]\times[0,1]$ can be expanded by Mercer's Theorem as $\mathcal{K}_b(r,s)=\sum_{n=1}^{\infty}\nu_n f_n(r)f_n(s)$, where ν_n is the eigenvalue of the kernel and $f_n(s)$ is the corresponding eigenfunction. Then, under Assumption A, the following holds with b fixed (See Sun (2014) for details):

$$\widehat{\Omega}^{CHAC}\Rightarrow \Lambda_c\int_0^1\int_0^1\mathcal{K}_b^*(r,s)d\mathcal{W}_k(r)\mathcal{W}_k(s)'\Lambda_c'. \tag{4}$$

Here, $\mathcal{K}_b^*(r,s)=\mathcal{K}\left((r-s)/b\right)-\int_0^1\mathcal{K}\left((r-\tau)/b\right)d\tau-\int_0^1\mathcal{K}\left((t-s)/b\right)dt+\int_0^1\int_0^1\mathcal{K}\left((t-\tau)/b\right)dtd\tau$. Then under H_0,

$$\begin{aligned}
W_{CHAC}&=\sqrt{G}\left[R\widehat{\beta}-r\right]'\left[R\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}\widehat{\Omega}^{CHAC}\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}R'\right]^{-1}\sqrt{G}\left[R\widehat{\beta}-r\right]\\
&\Rightarrow \mathcal{W}_k(1)'\Lambda_c'Q_c^{-1}R'\left[RQ_c^{-1}\Lambda_c\int_0^1\int_0^1\mathcal{K}_b^*(r,s)d\mathcal{W}_k(r)\mathcal{W}_k(s)'\Lambda_c'Q_c^{-1}R'\right]RQ_c^{-1}\Lambda_c\mathcal{W}_k(1)\\
&=\mathcal{W}_m(1)'\int_0^1\int_0^1\mathcal{K}_b^*(r,s)d\mathcal{W}_m(r)d\mathcal{W}_m(s)'\mathcal{W}_m(1).
\end{aligned}$$

The weak convergence ($\Rightarrow$) result is straightforward from (3) and (4). In case of $m=1$,

$$t_{CHAC}\Rightarrow\frac{\mathcal{W}_1(1)}{\sqrt{\int_0^1\int_0^1\mathcal{K}_b^*(r,s)d\mathcal{W}_1(r)d\mathcal{W}_1(s)}}.$$

□

Proof of Theorem 1(c): Under Assumption A, the relevant LLN for S_g^{xx} and the multivariate CLT for $\overline{v}_g$ are satisfied. Furthermore, the cosine basis functions $\phi_j(r)=\sqrt{2}cos\left(r\pi j\right)$ are orthonormal with $\int_0^1\phi_j(r)dr=0$. Therefore, the calculations in Sun (2013) apply. First, note that

$$\begin{aligned}
\widehat{\Lambda}_j=\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\phi_j\left(\frac{g-0.5}{G}\right)\widehat{\overline{v}}_g &\Rightarrow \Lambda_c\int_0^1\phi_j(r)\left(d\mathcal{W}_k(r)-dr\mathcal{W}_k(1)\right)\\
&=\Lambda_c\int_0^1\phi_j(r)d\mathcal{W}_k(r)\quad\because\int_0^1\phi_j(r)dr=0\\
&\overset{d}{=}\Lambda_c\xi_j^{(k)},\quad \xi_j^{(k)}=\int_0^1\phi_j(r)d\mathcal{W}_k(r)\overset{i.i.d.}{\sim}N(0,I_k).
\end{aligned}$$

It follows that

$$\widehat{\Omega}_j=\widehat{\Lambda}_j\widehat{\Lambda}_j'\Rightarrow\Lambda_c\xi_j^{(k)}\xi_j^{(k)\prime}\Lambda_c',$$

which implies

$$\widehat{\widehat{\Omega}}^{CEWC}=\frac{1}{B}\sum_{j=1}^{B}\widehat{\widehat{\Omega}}_j\Rightarrow\Lambda_c\frac{1}{B}\sum_{j=1}^{B}\xi_j^{(k)}\sum_{j=1}^{B}\xi_j^{(k)\prime}\Lambda_c^{\prime}.$$

Here $\xi_j^{(k)}$ are *i.i.d.* $N(0,I_k)$ distributed. Hence, by definition, $\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)\prime}$ is Wishart distribution with B degrees of freedom and covariance matrix I_k: $\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)\prime}\overset{d}{=}\mathbb{W}_k\left(B,I_k\right)$. Then under H_0,

$$\begin{aligned}W_{CEWC}&=\sqrt{G}\left(R\widehat{\beta}-r\right)'\left[R\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}\widehat{\widehat{\Omega}}^{CEWC}\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}R'\right]^{-1}\sqrt{G}\left(R\widehat{\beta}-r\right)\\&\Rightarrow\mathcal{W}_k(1)'\Lambda_c'Q_c^{-1}R'\left[RQ_c^{-1}\Lambda_c\frac{1}{B}\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)\prime}\Lambda_c'Q_c^{-1}R'\right]^{-1}RQ_c^{-1}\Lambda_c\mathcal{W}_k(1)\\&=\mathcal{W}_m(1)'\left[\frac{1}{B}\sum_{j=1}^{B}\xi_j^{(m)}\xi_j^{(m)\prime}\right]^{-1}\mathcal{W}_m(1)\overset{d}{=}T^2_{m,B}.\end{aligned}$$

Because $\phi_j(r)$ are orthonormal basis functions, $\mathcal{W}_m(1)$ and $\{\xi_j^{(m)}\}$ are independent. Then, by definition, the limiting distribution of W_{CEWC} is Hotelling's T-squared distribution with dimensionality parameter m and B degrees of freedom, $T_2(m,B)$. Using the relationship between the Hotelling's T-squared distribution and the F distribution, it follows that

$$F_{CEWC}=\frac{B-m+1}{mB}W_{CEWC}\Rightarrow F_{m,B-m+1},$$

where $F_{m,B-m+1}$ is the F distribution with degrees of freedom $(m,B-m+1)$. When $m=1$,

$$t_{CEWC}\Rightarrow t_B,$$

which is the student t distribution with degree of freedom B. □

Next, we provide a proof for Theorem 2 which gives asymptotic results for the $n_G\to\infty$, G-fixed case.

Proof of Theorem 2(a): When G is fixed, Assumption B2 implies

$$\frac{1}{T}S_g^{xx}=\frac{1}{T}\sum_{t=(g-1)n_G+1}^{gn_G}x_tx_t'\Rightarrow\frac{g}{G}Q-\frac{g-1}{G}Q=\frac{1}{G}Q,\tag{5}$$

and Assumption B3 implies

$$T^{-1/2}\overline{v}_g=\frac{1}{T}\sum_{t=(g-1)n_G+1}^{gn_G}v_g\Rightarrow\Lambda\left(\mathcal{W}_k\left(\frac{g}{G}\right)-\mathcal{W}_k\left(\frac{g-1}{G}\right)\right).\tag{6}$$

Hence,

$$\begin{aligned}\sqrt{T}\left(\widehat{\beta}-\beta\right)&=\left(\frac{1}{T}\sum_{g=1}^{G}S_g^{xx}\right)^{-1}T^{-\frac{1}{2}}\sum_{g=1}^{G}\overline{v}_g\\&\Rightarrow\left(\sum_{g=1}^{G}\frac{1}{G}Q\right)^{-1}\Lambda\sum_{g=1}^{G}\left(\mathcal{W}_k\left(\frac{g}{G}\right)-\mathcal{W}_k\left(\frac{g-1}{G}\right)\right)\\&=Q^{-1}\Lambda\mathcal{W}_k\left(1\right).\end{aligned}$$

□

Proof of Theorem 2(b): Recall that Assumption B states that the LLN and FCLT are satisfied for the unclustered series v_t. Hence,

$$
\begin{aligned}
T^{-\frac{1}{2}} \sum_{t=1}^{[rT]} \widehat{v}_t &= T^{-\frac{1}{2}} \sum_{j=1}^{[rT]} v_t - \frac{1}{T} \sum_{j=1}^{[rT]} x_t x_t' \sqrt{T} \left(\widehat{\beta} - \beta\right) \\
&\Rightarrow \Lambda \mathcal{W}_k(r) - r \Lambda \mathcal{W}_k(1) \\
&= \Lambda \widetilde{\mathcal{W}}(r),
\end{aligned}
$$

where $\widetilde{\mathcal{W}}(r)$ is a Brownian bridge. Next, note that

$$
\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq t \leq n_{\mathrm{G}} j\right] = g \quad \text{for} \quad t \in \left[n_{\mathrm{G}}(g-1) + 1 \leq t \leq n_{\mathrm{G}} g\right].
$$

Hence we can rewrite the CHAC estimator as

$$
\frac{G}{T} \widehat{\Omega}^{CHAC} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \mathcal{K} \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq t \leq n_{\mathrm{G}} j\right] - \sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq \tau \leq n_{\mathrm{G}} j\right]}{bG} \right) \widehat{v}_t \widehat{v}_\tau'.
$$

Expanding the kernel function $\mathcal{K}_b(r,s) = \mathcal{K}((r-s)/b)$ on $[0,1] \times [0,1]$ by Mercer's Theorem as $\mathcal{K}_b(r,s) = \sum_{n=1}^{\infty} \nu_n f_n(r) f_n(s)$, where ν_n is the eigenvalue of the kernel and $f_n(s)$ is the corresponding eigenfunction (See Sun (2014) for details) gives

$$
\begin{aligned}
&\frac{G}{T} \widehat{\Omega}^{CHAC} \\
=& \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \sum_{i=1}^{\infty} \lambda_i f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq t \leq n_{\mathrm{G}} j\right]}{bG} \right) f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq \tau \leq n_{\mathrm{G}} j\right]}{bG} \right) \widehat{v}_t \widehat{v}_\tau' \\
=& \sum_{i=1}^{\infty} \lambda_i \left[\frac{1}{T} \sum_{t=1}^{T} f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq t \leq n_{\mathrm{G}} j\right]}{bG} \right) \sqrt{T} \widehat{v}_t \right] \left[\frac{1}{T} \sum_{\tau=1}^{T} f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[n_{\mathrm{G}}(j-1) + 1 \leq \tau \leq n_{\mathrm{G}} j\right]}{bG} \right) \sqrt{T} \widehat{v}_\tau \right] \\
&\Rightarrow \sum_{i=1}^{\infty} \lambda_i \int_0^1 f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq r \leq \frac{j}{G}\right]}{bG} \right) \Lambda d\widetilde{\mathcal{W}}(r) \int_0^1 f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq s \leq \frac{j}{G}\right]}{bG} \right) d\widetilde{\mathcal{W}}(s)' \Lambda' \\
&= \Lambda \int_0^1 \int_0^1 \sum_{i=1}^{\infty} \lambda_i f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq r \leq \frac{j}{G}\right]}{bG} \right) f_i \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq s \leq \frac{j}{G}\right]}{bG} \right) d\widetilde{\mathcal{W}}(r) d\widetilde{\mathcal{W}}(s)' \Lambda' \\
&= \Lambda \int_0^1 \int_0^1 \mathcal{K} \left(\frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq r \leq \frac{j}{G}\right]}{bG} - \frac{\sum_{j=1}^{G} j \mathbb{1}\left[\frac{(j-1)+1}{G} \leq s \leq \frac{j}{G}\right]}{bG} \right) d\widetilde{\mathcal{W}}(r) d\widetilde{\mathcal{W}}(s)' \Lambda' \\
&\equiv \Lambda P_k(G,b) \Lambda'.
\end{aligned}
$$

Then under the null hypothesis H_0, the W_{CHAC} statistic follows the limiting distribution as defined below:

$$
\begin{aligned}
W_{CHAC} &= \sqrt{T} \left(R\widehat{\beta} - r \right)' \left[R \widehat{V}_{CHAC} R' \right]^{-1} \sqrt{T} \left(R\widehat{\beta} - r \right) \\
&\Rightarrow \left[RQ^{-1} \Lambda \mathcal{W}_k(1) \right]' \left[RQ^{-1} \Lambda P_k(G,b) \Lambda'^{-1} R \right]^{-1} RQ^{-1} \Lambda \mathcal{W}_k(1) \\
&= \mathcal{W}_m(1)' \left[P_m(G,b) \right]^{-1} \mathcal{W}_m(1).
\end{aligned}
$$

When $m = 1$,

$$
t_{CHAC} \Rightarrow \frac{\mathcal{W}_1(1)}{P_1(G,b)}.
$$

An alternative approach to obtaining a limiting result for the variance estimator follows Kiefer and Vogelsang (2005). First note that

$$\begin{aligned} T^{-1/2}\widehat{S}_g &= T^{-\frac{1}{2}}\sum_{j=1}^{g}\widehat{v}_j = T^{-\frac{1}{2}}\sum_{j=1}^{g}\overline{v}_j - \frac{1}{T}\sum_{j=1}^{g}S_j^{xx}\sqrt{T}\left(\widehat{\beta}-\beta\right) \\ &\Rightarrow \Lambda\mathcal{W}_k\left(\frac{g}{G}\right) - \frac{g}{G}QQ^{-1}\Lambda\mathcal{W}_k(1) \\ &= \Lambda\left(\mathcal{W}_k\left(\frac{g}{G}\right) - \frac{g}{G}\Lambda\mathcal{W}_k(1)\right) \equiv \Lambda\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right). \end{aligned} \quad (7)$$

This result is straightforward given (5) and (6). Using summation by parts

$$\begin{aligned} \frac{G}{T}\widehat{\Omega}^{CHAC} &= \sum_{g=1}^{G-1}\sum_{h=1}^{G-1}T^{-\frac{1}{2}}\widehat{S}_g\left[2\mathcal{K}\left(\frac{|g-h|}{M_G}\right)-\mathcal{K}\left(\frac{|g-h+1|}{M_G}\right)-\mathcal{K}\left(\frac{|g-h-1|}{M_G}\right)\right]T^{-\frac{1}{2}}\widehat{S}_h' \\ &\Rightarrow \Lambda\left[\sum_{g=1}^{G-1}\sum_{h=1}^{G-1}\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)\left(2\mathcal{K}\left(\frac{|g-h|}{M_G}\right)-\mathcal{K}\left(\frac{|g-h+1|}{M_G}\right)-\mathcal{K}\left(\frac{|g-h-1|}{M_G}\right)\right)\ddot{\mathcal{W}}_k\left(\frac{h}{G}\right)'\right]\Lambda'. \end{aligned}$$

For the Bartlett kernel, the specific result is

$$\begin{aligned} \frac{G}{T}\widehat{\Omega}^{CHAC} &= \sum_{g=1}^{G-1}\sum_{h=1}^{G-1}T^{-\frac{1}{2}}\widehat{S}_g\left[2\mathcal{K}\left(\frac{|g-h|}{M_G}\right)-k\left(\frac{|g-h+1|}{M_G}\right)-\mathcal{K}\left(\frac{|g-h-1|}{M_G}\right)\right]T^{-\frac{1}{2}}\widehat{S}_h' \\ &= \frac{2}{M_G}\sum_{g=1}^{G-1}T^{-\frac{1}{2}}\widehat{S}_gT^{-\frac{1}{2}}\widehat{S}_g' - \frac{1}{M_G}\sum_{g=1}^{G-M_G-1}\left(T^{-\frac{1}{2}}\widehat{S}_gT^{-\frac{1}{2}}\widehat{S}_{g+M_G}'^{-\frac{1}{2}}\widehat{S}_{g+M_G}T^{-\frac{1}{2}}\widehat{S}_g'\right) \\ &\Rightarrow \Lambda\left[\frac{2}{M_G}\sum_{g=1}^{G-1}\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)' - \frac{1}{M_G}\sum_{g=1}^{G-M_G-1}\left(\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)\ddot{\mathcal{W}}_k\left(\frac{g+M_G}{G}\right)' + \ddot{\mathcal{W}}_k\left(\frac{g+M_G}{G}\right)\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)'\right)\right]\Lambda'. \end{aligned}$$

□

Proof of Theorem 2(c): Straightforward calculations give

$$\begin{aligned} \left(\frac{T}{G}\right)^{-1/2}\Lambda_j &= \sum_{g=1}^{G}\sqrt{2}cos\left(\frac{g-0.5}{G}\pi j\right)T^{-1/2}\widehat{v}_g \\ &\Rightarrow \Lambda\sum_{g=1}^{G}\sqrt{2}cos\left(\frac{g-0.5}{G}\pi j\right)\left(\mathcal{W}_k\left(\frac{g}{G}\right)-\mathcal{W}_k\left(\frac{g-1}{G}\right)-\frac{1}{G}\mathcal{W}_k(1)\right) && (8) \\ &= \Lambda\sum_{g=1}^{G}\sqrt{2}cos\left(\frac{g-0.5}{G}\pi j\right)\left(\mathcal{W}_k\left(\frac{g}{G}\right)-\mathcal{W}_k\left(\frac{g-1}{G}\right)\right) && (9) \\ &\overset{d}{=} \Lambda\sum_{g=1}^{G}\sqrt{2}cos\left(\frac{g-0.5}{G}\pi j\right)Z_g, \quad Z_g \overset{i.i.d.}{\sim} N\left(0,\frac{1}{G}I_k\right) \\ &\overset{d}{=} \Lambda\xi_j, \quad \xi_j \overset{i.i.d.}{\sim} N(0,I_k). && (10) \end{aligned}$$

The weak convergence ($\Rightarrow$) in (8) is obvious using (7):

$$T^{-1/2}\widehat{v}_g \Rightarrow \Lambda\left[\ddot{\mathcal{W}}_k\left(\frac{g}{G}\right)-\ddot{\mathcal{W}}_k\left(\frac{g-1}{G}\right)\right] = \Lambda\left[\mathcal{W}_k\left(\frac{g}{G}\right)-\mathcal{W}_k\left(\frac{g-1}{G}\right)-\frac{1}{G}\mathcal{W}_k(1)\right].$$

The equality in (9) is also straightforward because $\sum_{g=1}^{G}\cos\left[\left(\frac{(g-0.5)\pi j}{G}\right)\right] = 0$ for $j \in \{1,\ldots,B\}$, $B \leq G$, as shown below:

$$\sum_{g=1}^{G}\cos\left[\left(\frac{(g-0.5)\pi j}{G}\right)\right] = Re\left\{\sum_{g=1}^{G} e^{i\frac{(g-0.5)\pi j}{G}}\right\} = Re\left\{\frac{1-e^{i\pi j}}{1-e^{\frac{i\pi j}{G}}}\times e^{\frac{0.5 i\pi j}{G}}\right\}$$
$$= Re\left\{\frac{\left(1-e^{i\pi j}\right)\left(e^{\frac{i\pi j}{2G}}-e^{\frac{-i\pi j}{2G}}\right)}{2\left(1-cos\frac{\pi j}{G}\right)}\right\}$$
$$= \frac{\sin(\pi j)\sin\left(\frac{\pi j}{2G}\right)}{1-\cos\left(\frac{\pi j}{G}\right)} = 0. \tag{11}$$

Finally, the last equivalence in distribution $\left(\overset{d}{=}\right)$ in (10) holds from the following two equalities. First, for $j \in \{1, \ldots, G-1\}$,

$$\sum_{g=1}^{G}\frac{1}{G}2cos\left(\frac{g-0.5}{G}\pi j\right)^2 = \sum_{g=1}^{G}\frac{1}{G}\left\{1+cos\left(\frac{g-0.5}{G}2\pi j\right)\right\}$$
$$= 1+\frac{1}{G}Re\left\{\sum_{g=1}^{G} e^{\frac{i(g-0.5)2\pi j}{G}}\right\}$$
$$= 1+\frac{1}{G}Re\left\{\frac{1-e^{i2\pi j}}{1-e^{\frac{i2\pi j}{G}}}\times e^{\frac{i\pi j}{G}}\right\}$$
$$= 1+\frac{1}{G}Re\left\{\frac{\left(1-e^{i2\pi j}\right)\left(e^{\frac{i\pi j}{G}}-e^{\frac{-i\pi j}{G}}\right)}{2\left(1-\cos\left(\frac{2\pi j}{G}\right)\right)}\right\}$$
$$= 1+\frac{1}{G}\frac{\sin(2\pi j)\sin\left(\frac{\pi j}{G}\right)}{1-\cos\left(\frac{2\pi j}{G}\right)} = 1.$$

Second, $\{\xi_j\}$ is a sequence of independent random variables because for $j \neq k$, $j+k < 2G$,

$$\sum_{g=1}^{G} cos\left(\frac{g-0.5}{G}\pi j\right)cos\left(\frac{g-0.5}{G}\pi k\right) = \frac{1}{2}\sum_{g=1}^{G}\left(cos\left(\frac{g-0.5}{G}\pi(j-k)\right)+cos\left(\frac{g-0.5}{G}\pi(j+k)\right)\right)$$
$$= \frac{1}{2}\left(\frac{sin(\pi(j-k))}{sin\left(\frac{\pi(j-k)}{2G}\right)}+\frac{sin(\pi(j+k))}{sin\left(\frac{\pi(j+k)}{2G}\right)}\right) = 0.$$

From (10),

$$\frac{G}{T}\widehat{\Lambda}_j\widehat{\Lambda}_j' \Rightarrow \Lambda\xi_j^{(k)}\xi_j^{(k)'}\Lambda'$$

and the asymptotic limit of the CEWC estimator is

$$\frac{G}{T}\widehat{\Omega}^{CEWC} = \frac{G}{T}\frac{1}{B}\sum_{j=1}^{B}\widehat{\Lambda}_j\widehat{\Lambda}_j' \Rightarrow \Lambda\frac{1}{B}\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)'}\Lambda'.$$

By definition, $\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)'}$ is a Wishart distribution: $\sum_{j=1}^{B}\xi_j^{(k)}\xi_j^{(k)'} \overset{d}{=} \mathbb{W}_k(I_k, B)$. The limits F_{CEWC} and t_{CEWC} easily follow using similar arguments as in the proof of Theorem 1(c).

□