A FULLY FLEXIBLE CHANGEPOINT TEST FOR REGRESSION MODELS WITH STATIONARY ERRORS

Michael W. Robbins

RAND Corporation

Abstract: Temporal discontinuities in time series present one of the classic problems of time series analysis. Such discontinuities are often analyzed through the detection of changes at specific times in the parameters governing a regression model that is fit to the series. The regression framework examined here contains three classes of predictors: functional form, seasonal, and stochastic. Regression errors are allowed to observe a general stationary structure. Methods are proposed that provide the analyst full flexibility in selecting which set of regression parameters are allowed to change under the alternative hypothesis. Several mathematical complications arise in the development of such procedures; manners of addressing these complications are illustrated. A simulation study illustrates the efficacy of the proposed methodology—a test statistic based on residuals from an ARMA model is shown to perform most favorably. The methods are applied to the carbon dioxide time series measured at Mauna Loa Observatory, where a shift in the seasonal variations is detected (in addition to a known shift in trend), and to a series of monthly temperatures at Barrow, Alaska, where only a shift in trend is found.

Key words and phrases: Asymptotic theory; Changepoints; Time series analysis

1. Introduction

The detection of a changepoint (or structural break) within an ordered sequence of data is one of the classical problems of statistical analysis. Changepoint methods for regression models have proliferated in recent decades (e.g., MacNeill, 1978; Hansen, 2000; Aue *et al.*, 2006, 2008; Gallagher *et al.*, 2013). A related problem in changepoint diagnostics involves the incorporation of autocorrelation (see, for example, Bai, 1993; Antoch *et al.*, 1997; Yu, 2007; Robbins *et al.*, 2011a, among many others). The goal here is to develop a comprehensive and flexible changepoint test for models of a general regression structure with stationary error sequences. Specifically, we consider detection of temporal shifts in the coefficients of a regression model where the outcome series is allowed to depend on three classes of predictors: 1) terms governing the trend (where the trend may be, for example, constant, linear, quadratic, etc.); 2) seasonal terms (i.e., which control periodic oscillations from the trend function); and 3) stochastic covariates used to explain error in the outcome.

To briefly review foundational work that is particularly pertinent to the stated objective: Gombay (2010) and Aue *et al.* (2012) present diagnostic methods for changepoints in regression models while enabling autocorrelated regression errors. Specifically, Aue *et al.* (2012) examine a regression structure with predictor terms that have a general functional form. However, they model autocorrelation through the use of a Bartlett-based variance expression and approximate the large sample distribution of their test statistics through extreme value expressions. As a result, their asymptotic approximations do not perform well on finite samples; they use bootstrapping techniques to address this issue.

Robbins *et al.* (2016) extend the framework of Aue *et al.* (2012) to a more general model that is designed to incorporate seasonality and covariate information in addition to functional trend. Further, they fit an autoregressive moving average (ARMA) model to the regression residuals and develop a changepoint test statistic that is based off of the resulting ARMA fit. However, their method is limited in that only the underlying functional trend is allowed to change. Detection of changepoints in the seasonal structure and/or covariate/outcome relationships within a regression model is an untapped problem. (Note that Aue *et al.* 2012 discuss changepoints in a seasonal cycle while using harmonic functions with a fixed number of oscillations to capture seasonality. However, asymptotic methods that mandate such a representation are not necessarily congenial with real data; i.e., they do not enable the period of the seasonal cycle to remain constant as the total sample size increases.)

The detection of changes in the seasonal structure of a time series or in the temporal relationship between a predictor and the resulting outcome is certainly of practical relevance. Consider the following climate examples. First, several authors (e.g., Buermann *et al.*, 2007; Zeng *et al.*, 2014) have posited that the amplitude of the seasonal oscillations in atmospheric carbon dioxide (CO_2) measurements is increasing over time. Further, researchers have found evidence that warming of surface temperatures in polar climates is greater in winter seasons than in summer seasons (Lu & Cai, 2009; Screen & Simmonds, 2010). Lastly, Elsner *et al.* (2001) find that the magnitude of the well established dependence between El Niño and hurricane frequency has declined over time. .

Furthermore, most existing methods (e.g., Aue *et al.*, 2012) mandate that all regression coefficients change simultaneously. Thus, we expand upon existing tests for changes in trend and by developing tests for changes in the seasonal structure and outcome/covariate relationships separately. Then, we illustrate that these three classes of tests are asymptotically independent which enables them to packaged as a single omnibus test that can detect discontinuities within any pre-selected subset of coefficients in the general regression model described earlier. Throughout, regression errors are allowed to contain serial correlation. Flexible procedures based on ordinary least squares (OLS) residuals and ARMA residuals are developed. Several mathematical complications arise in efforts to extend existing changepoint methods; innovative approaches are required to overcome these obstacles.

The methods presented here are designed for the at-most-one-changepoint (AMOC) alternative. Nonetheless, we illustrate that the proposed method can be used to first detect a shift in trend and then separately test for a change in seasonal oscillations while incorporating the discontinuity in trend (if found); disentanglement of changes in trend from changes in seasonality proves prudent within the data examples presented here.

The article proceeds in the following manner. Section 2 provides mathematical context and in the process paraphrases technical details of relevant extant methodology, and Section 3 outlines the foundations for our general method by addressing the setting of independent and identically distributed (IID) regression errors. In Section 4, procedures that encapsulate autocorrelation in regression errors are developed. In Section 5, the finite sample performance of the proposed methods through simulation is examined, whereas Section 6 is used to present an application of the developed techniques to two datasets: 1) carbon dioxide levels measured at the Mauna Loa Observatory in Hawaii, and 2) average temperatures in Barrow, Alaska. These applications illustrate the importance of flexibility in comprehensive changepoint detection methods.

2. Technical Preliminaries

Let $\{Y_t\}$ for $t = 1, \ldots, n$ denote the response sequence where n is the sample size. The null hypothesis model for the t^{th} time point is

$$Y_t = \widetilde{\boldsymbol{\alpha}}'\widetilde{\mathbf{x}}_t + \widetilde{\boldsymbol{\beta}}'\widetilde{\mathbf{s}}_t + \widetilde{\boldsymbol{\gamma}}'\widetilde{\mathbf{v}}_t + \epsilon_t, \tag{1}$$

where $\widetilde{\mathbf{x}}_t$ is a vector of deterministic design points that control any long-term temporal trend, $\widetilde{\mathbf{s}}_t$ is a vector of terms that determine any seasonal cycle, and $\widetilde{\mathbf{v}}_t$ is a vector of stochastic covariates. These vectors have length p_x, p_s and p_v, respectively. Additionally, $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\gamma}}$ are vectors of regression coefficients, and $\{\epsilon_t\}$ is a stationary mean zero error sequence. All sequences indexed by t are defined for $t \in (1, \ldots, n)$. A functional form is imposed on $\{\widetilde{\mathbf{x}}_t\}$. That is, set $\widetilde{\mathbf{x}}_t = (f_1(t/n), \ldots, f_{p_x}(t/n))'$, where for each $j = 1, \ldots, p_x$, $f_j(z)$ is a continuous function for $z \in (0, 1)$. Commonly, these functions will encompass an intercept term by setting $f_1(z) = 1$. The functions are evaluated at times scaled to the unit interval for mathematical convenience. In the case of polynomial trend, for example, one can safely define predictors using t^j instead of $(t/n)^j$ (Aue *et al.*, 2012; Robbins *et al.*, 2016). Further, $\{\widetilde{\mathbf{s}}_t\}$ contains periodic deterministic design points with known period T so that $\widetilde{\mathbf{s}}_{t+T} = \widetilde{\mathbf{s}}_t$ for all t. It is also assumed without loss of generality (as long as the model contains an intercept term) that $\sum_{t=1}^{T} \widetilde{\mathbf{s}}_t = \mathbf{0}$. As a result of this imposition, the predictor terms $\mathbf{s}_t$ are interpreted as governing seasonal oscillations from the overall trend function, and the parameters $\widetilde{\boldsymbol{\beta}}$ control the magnitude of these oscillations. Finally, $\{\widetilde{\mathbf{v}}_t\}$ contains stationary terms that without loss of generality satisfy $\mathrm{E}[\widetilde{\mathbf{v}}_t] = \mathbf{0}$ and have finite second moments. Note that the first q_x terms of $\widetilde{\boldsymbol{\alpha}}$, where $q_x < p_x$, will be allowed to shift at the changepoint time; q_s and q_v are defined similarly.

Define $K \subset (0, 1)$ so that $K = \{z \in (0, 1) : k = \lfloor nz \rfloor$ is an admissible changepoint time$\}$, with $\lfloor \cdot \rfloor$ indicating the floor function. We use $K = \{z : \ell \le z \le h\}$, where $0 < \ell < h < 1$. Restricting this set of admissible changepoints away from the boundaries of $[0, 1]$ is done to ensure that the test statistics defined in the forthcoming discourse are finite asymptotically.

The formal assumptions imposed here are outlined in Appendix A of the supplement. Note (A.1) in Assumption 4, wherein the regression errors $\{\epsilon_t\}$ are written via a causal expression in terms of a sequence $\{Z_t\}$ of white noise innovations that have variance σ^2 (this implies stationarity and facilitates autocorrelation in the error sequence). Furthermore, the variability that is attributable to autocorrelation can often be encapsulated by monitoring the quantity

$$\tau^2 := \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}\left(\sum_{t=1}^{n} \epsilon_t\right). \tag{2}$$

A process that observes (A.1) has short memory in that $\tau^2 < \infty$.

We consider an alternative hypothesis that allows pre-selected subsets of the regression coefficients $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\gamma}}$ to shift at a single unknown time c that satisfies $1 \leq c < n$. To enable complete flexibility in the choice of which parameters change, $\widetilde{\mathbf{x}}_t$ is decomposed into two subvectors: $\mathbf{x}_t$, which has length q_x, and $\mathbf{x}_t^*$ of length $p_x - q_x$. That is, we write $\widetilde{\mathbf{x}}_t = (\mathbf{x}_t', (\mathbf{x}_t^*)')'$. Similarly, we set $\widetilde{\mathbf{s}}_t = (\mathbf{s}_t', (\mathbf{s}_t^*)')'$, and $\widetilde{\mathbf{v}}_t = (\mathbf{v}_t', (\mathbf{v}_t^*)')'$, where $\mathbf{s}_t$ and $\mathbf{v}_t$ have length q_s and q_v, respectively. The alternative hypothesis model for $\{Y_t\}$ is

$$\begin{aligned} Y_t &= (\boldsymbol{\alpha} + \boldsymbol{\delta}_{\mathrm{x},t})'\mathbf{x}_t + (\boldsymbol{\beta} + \boldsymbol{\delta}_{\mathrm{s},t})'\mathbf{s}_t + (\boldsymbol{\gamma} + \boldsymbol{\delta}_{\mathrm{v},t})'\mathbf{v}_t \\ &\quad + (\boldsymbol{\alpha}^*)'\mathbf{x}_t^* + (\boldsymbol{\beta}^*)'\mathbf{s}_t^* + (\boldsymbol{\gamma}^*)'\mathbf{v}_t^* + \epsilon_t, \end{aligned}$$

where the vectors of regression coefficients have been decomposed in a similar manner. That is, we set $\widetilde{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}', (\boldsymbol{\alpha}^*)')'$, $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}', (\boldsymbol{\beta}^*)')'$ and $\widetilde{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}', (\boldsymbol{\gamma}^*)')'$. The expressions $\boldsymbol{\delta}_{\mathrm{x},t}$, $\boldsymbol{\delta}_{\mathrm{s},t}$ and $\boldsymbol{\delta}_{\mathrm{v},t}$ quantify the magnitude of the structural break. Specifically,

$$\boldsymbol{\delta}_{\mathrm{x},t} = \begin{cases} \mathbf{0}, & t \leq c, \\ \boldsymbol{\Delta}_x, & t > c, \end{cases} \quad \boldsymbol{\delta}_{\mathrm{s},t} = \begin{cases} \mathbf{0}, & t \leq c, \\ \boldsymbol{\Delta}_s, & t > c, \end{cases} \quad \text{and} \quad \boldsymbol{\delta}_{\mathrm{v},t} = \begin{cases} \mathbf{0}, & t \leq c, \\ \boldsymbol{\Delta}_v, & t > c, \end{cases}$$

which are vectors of dimension q_x, q_s and q_v, respectively. The double subscript (e.g., x, t) seen in the above definitions is a notational structure that is used frequently throughout this article; the first subscript (e.g., x) simply refers to a set of predictor variables (out of the functional form, seasonal, or stochastic predictor sets) and the second subscript (e.g., t for $t \in (1, \ldots, n)$) is a time-varying index. Defining $\boldsymbol{\Delta} = ((\boldsymbol{\Delta}_x)', (\boldsymbol{\Delta}_s)', (\boldsymbol{\Delta}_v)')'$, we test

$$\mathcal{H}_0 : \boldsymbol{\Delta} = \mathbf{0} \quad \text{against} \quad \mathcal{H}_1 : \boldsymbol{\Delta} \neq \mathbf{0}. \tag{3}$$

Since the actual changepoint time, c, is assumed to be unknown, the methodology developed here involves examining processes of test statistics that are evaluated for each element in the set of admissible changepoint times. We define order of probability notation that indicates convergence rates for such processes when considered across all k that satisfy $k/n \in K$. Given a random process $\{\mathbf{X}_n(k)\}$ for $k/n \in K$ and for some sequence a_n , we write

$$\mathbf{X}_n(k) = o_p(a_n, k) \quad \text{when} \quad \max_{\frac{k}{n} \in K} ||\mathbf{X}_n(k)||_\infty = o_p(a_n), \tag{4}$$

and

$$\mathbf{X}_n(k) = \mathcal{O}_p(a_n, k) \quad \text{when} \quad \max_{\frac{k}{n} \in K} ||\mathbf{X}_n(k)||_\infty = \mathcal{O}_p(a_n),$$

as $n \to \infty$. Similar notation for deterministic processes is defined via $o(a_n, k)$ and $\mathcal{O}(a_n, k)$.

3. Fully Flexible Changepoint Tests

In this section, tests for the hypotheses in (3) are developed while assuming IID regression errors. The resultant procedure is extended to settings involving general stationary errors later. In line with earlier work (e.g., Robbins *et al.*, 2016), we use a Wald-type statistic based on an estimator for $\boldsymbol{\Delta}$ which is then expressed as weighted sum of OLS residuals.

Letting $\widehat{\boldsymbol{\Delta}}_k$ denote the OLS estimate of $\boldsymbol{\Delta}$ (which disregards autocorrelation within the errors) and assuming that a changepoint occurs at time k with $k/n \in K$, the Wald statistic used to test for the presence of a change at time k is $F_k := \widehat{\boldsymbol{\Delta}}_k' \mathrm{Var}(\widehat{\boldsymbol{\Delta}}_k)^{-1} \widehat{\boldsymbol{\Delta}}_k$. Note that it is assumed throughout that all processes indexed by k (where k typically indicates a hypothetical changepoint time) have k that satisfies $k/n \in K$. Letting $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ denote OLS estimators of $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\gamma}}$ calculated under $\mathcal{H}_0$, the OLS residuals are defined as

$$\hat{\epsilon}_t = Y_t - (\widehat{\boldsymbol{\alpha}}' \widetilde{\mathbf{x}}_t + \widehat{\boldsymbol{\beta}}' \widetilde{\mathbf{s}}_t + \widehat{\boldsymbol{\gamma}}' \widetilde{\mathbf{v}}_t). \tag{5}$$

Defining $\mathbf{m}_t = (\mathbf{x}_t', \mathbf{s}_t', \mathbf{v}_t')'$ and $\mathbf{N}_k = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{m}_t$, it holds that $F_k = \mathbf{N}_k' \mathrm{Var}(\mathbf{N}_k)^{-1} \mathbf{N}_k$. In practice, we use

$$\widehat{F}_k = \frac{\mathbf{N}_k' \mathbf{C}_k^{-1} \mathbf{N}_k}{\hat{\tau}^2}, \tag{6}$$

to test for a changepoint at time k, where $\mathbf{C}_k$ is a matrix satisfying $\mathrm{Var}(\mathbf{N}_k) = \tau^2 \mathbf{C}_k$ conditional upon $\{\widetilde{\mathbf{v}}_t\}$ whenever $\{\epsilon_t\}$ is white noise with variance τ^2. The specific form of $\mathbf{C}_k$ is given in (A.5) from Appendix F in the supplement. Discussion of choices for $\hat{\tau}^2$, which estimates τ^2, is postponed until Section 4, wherein the white noise assumption is relaxed.

To detect a changepoint at an unknown time, consider the maximally selected statistic

$$\widehat{F} = \max_{\frac{k}{n} \in K} \widehat{F}_k. \tag{7}$$

For this statistic and others defined later, the estimated changepoint time $\hat{c}$ is the argument k that maximizes $\widehat{F}_k$ (although statistical confirmation of the presence of a changepoint, as opposed to properties of its estimator, is our focus here). The large-sample behavior of $\widehat{F}$ is formally stated in Theorem 1 below; however, we first provide further discourse to help illustrate the components of its limiting process.

The process $\{\mathbf{N}_k\}$, which drives $\widehat{F}$, is itself underpinned by three separate processes:

$$\mathbf{N}_{\mathrm{x},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{x}_t, \quad \mathbf{N}_{\mathrm{s},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{s}_t, \quad \text{and} \quad \mathbf{N}_{\mathrm{v},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{v}_t. \tag{8}$$

for $1 \leq k \leq n$. It would appear evident given the orthogonal nature of the regression design that these three processes are pairwise asymptotically uncorrelated; however, this is shown formally within the proof of Theorem 1 provided in the supplement (Appendix F). Thus, for large n, $\widehat{F}_k$ may be dissected into three uncorrelated components:

$$\widehat{F}_k \approx \widehat{F}_{\mathrm{x},k} + \widehat{F}_{\mathrm{s},k} + \widehat{F}_{\mathrm{v},k}$$

where $\widehat{F}_{\mathrm{x},k} = \mathbf{N}'_{\mathrm{x},k} \mathrm{Var}(\mathbf{N}_{\mathrm{x},k})^{-1} \mathbf{N}_{\mathrm{x},k}$ and $\widehat{F}_{\mathrm{s},k}$ and $\widehat{F}_{\mathrm{v},k}$ are defined in an analogous manner. The variance terms in these expressions can be approximated via equations of the form $\widehat{\mathrm{Var}}(\mathbf{N}_{\mathrm{s},k}) = \hat{\tau}^2 \mathbf{C}_{\mathrm{s},k}$, where $\mathbf{C}_{\mathrm{s},k}$ is defined in (A.10) from Appendix F of the supplement.

The process $\{\widehat{F}_{\mathrm{x},k}\}$ was described by Robbins *et al.* (2016) since it may be used to detect a shift in only the trend function. Those results are rehashed briefly here. Letting $\Rightarrow$ denote weak convergence in $D[0,1]$, the space of right-continuous functions with left-hand limits,

$$\widehat{F}_{\mathrm{x},\lfloor nz \rfloor} \Rightarrow \widetilde{B}_1(z), \quad \text{for } z \in K, \text{ where} \quad \widetilde{B}_1(z) = \mathbf{\Lambda}(z)' \mathbf{\Omega}(z)^{-1} \mathbf{\Lambda}(z). \tag{9}$$

See Appendix B of the supplement for details regarding the convergence and terms in (9).

The processes $\{\widehat{F}_{\mathrm{s},k}\}$ and $\{\widehat{F}_{\mathrm{v},k}\}$ have not been considered in earlier work. They behave as the square of scaled multidimensional Brownian bridges for large sample sizes. Although detailed justification for this claim is provided in the proof of Theorem 1, key concepts that yield the claim are sketched below. First, note that

$$(n\tau^2)^{-1}\mathrm{Var}(\mathbf{N}_{\mathrm{s},\lfloor nz\rfloor}) \Rightarrow z(1-z)\mathbf{D}_T, \tag{10}$$

and

$$(n\tau^2)^{-1}\mathrm{Var}(\mathbf{N}_{\mathrm{v},\lfloor nz\rfloor}) \Rightarrow z(1-z)\mathbf{\Sigma}_v, \tag{11}$$

as $n \to \infty$ for $z \in K$, where $\mathbf{D}_T = \sum_{j=1}^{T} \mathbf{s}_j\mathbf{s}_j'/T$ and $\mathbf{\Sigma}_v = \mathrm{Var}(\mathbf{v}_1)$. In addition,

$$\mathbf{N}_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t\epsilon_t - \frac{k}{n}\sum_{t=1}^{n} \mathbf{s}_t\epsilon_t + o_p(\sqrt{n}, k), \tag{12}$$

and

$$\mathbf{N}_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t\epsilon_t - \frac{k}{n}\sum_{t=1}^{n} \mathbf{v}_t\epsilon_t + o_p(\sqrt{n}, k), \tag{13}$$

where the order of probability notation introduced in (4) has been employed. Letting $\{\mathbf{W}_d(z)\}_{z\in[0,1]}$ denote a d-dimensional Wiener process, (10)-(13) yield that

$$(n\tau^2\mathbf{D}_T)^{-1/2}\mathbf{N}_{\mathrm{s},\lfloor nz\rfloor} \Rightarrow \mathbf{W}_{q_s}(z) - z\mathbf{W}_{q_s}(1), \tag{14}$$

and

$$(n\tau^2\mathbf{\Sigma}_v)^{-1/2}\mathbf{N}_{\mathrm{v},\lfloor nz\rfloor} \Rightarrow \mathbf{W}_{q_v}(z) - z\mathbf{W}_{q_v}(1), \tag{15}$$

The asymptotic behavior of $\widehat{F}$ is stated within the following theorem.

Theorem 1. *Assume that the null hypothesis is true, i.e., that the data $\{Y_t\}$ obey the model in (1), and that Assumptions 1–3 in Appendix A hold. Further, assume that the errors $\{\epsilon_t\}$*

are IID with variance τ^2*. If* $\widehat{F}$ *is calculated in accordance with (7),*

$$\widehat{F} \xrightarrow{\mathcal{D}} \sup_{z \in K} \widetilde{B}(z)$$

as $n \to \infty$*, where*

$$\widetilde{B}(z) = \widetilde{B}_1(z) + \widetilde{B}_2(z). \tag{16}$$

The stochastic process $\{\widetilde{B}_1(z)\}$*, which is defined in (9), is independent of* $\{\widetilde{B}_2(z)\}$ *where*

$$\widetilde{B}_2(z) = \frac{\mathbf{B}_d(z)'\mathbf{B}_d(z)}{z(1-z)}, \tag{17}$$

for $z \in K$ *with* $d = q_s + q_v$*. Further,* $\{\mathbf{B}_d(z) := \mathbf{W}_d(z) - z\mathbf{W}_d(1)\}$ *denotes a* d*-dimensional set of independent Brownian bridges each defined for* $z \in [0, 1]$*.*

Note that the limit process $\{\widetilde{B}(z)\}$ is not influenced by the characteristics of $\{\widetilde{\mathbf{s}}_t\}$ and $\{\widetilde{\mathbf{v}}_t\}$ (aside from their dimensionality) but depends on the specific functions that underpin $\{\widetilde{\mathbf{x}}_t\}$. Thus, if $q_x = 0$ or if $p_x = 1$ with $f_1(z) \propto 1$ for all $z \in [0, 1]$, it holds that $\widetilde{B}(z)$ simplifies to a sum of $q_x + q_s + q_v$ scaled and squared Brownian bridges (see Csörgő & Horváth 1997 and Robbins *et al.* 2011b for closed-form approximations of the supremum of this process).

Consider also that the procedure outlined here can be used to test for changes in coefficients governing $\mathbf{s}_t$ and $\mathbf{v}_t$ while incorporating a known change in trend. Let $f_{i+1}(z) = \mathbf{1}_{\{t>c^*\}} f_i(z)$ for odd i, where $\mathbf{1}_{\{A\}}$ is the indicator of event A and $c^*/n \in (0, 1)$. In order to satisfy Assumption 2 from Appendix A of the supplement with this model, one may impose $q_x = 0$. In Section 6, it proves prudent to disentangle changes in trend from changes in seasonality in such a manner.

4. Tests under Autocorrelated Errors

Here, we consider extension of the test statistics illustrated in the previous section to settings where regression errors contain autocorrelation. That is, the regression errors $\{\epsilon_t\}$ are allowed to obey the general stationary structure outlined in Section 2. The $\widehat{F}$ statistic introduced above has power to detect discontinuities in the presence of autocorrelated errors—we must discern manners of adjustment so that type I error rates can be controlled.

Robbins *et al.* (2016) argue that $\{\widehat{F}_{\mathrm{x},k}\}$ has the same limit process in circumstances where regression errors contain autocorrelation as is observed when regression errors are white noise if the estimate of the marginal error variance is replaced with a consistent estimator of the long-run variance term τ^2 from (2). Further, their calculations incorporate the following Bartlett-based estimator of τ^2 in the event that regression errors contain a stationary autocorrelation structure:

$$\hat{\tau}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2 + 2\sum_{j=1}^{q_n}\left(1-\frac{j}{q_n+1}\right)\frac{1}{n-j}\sum_{t=1}^{n-j}\hat{\epsilon}_t\hat{\epsilon}_{t+j}, \tag{18}$$

where q_n is a bandwidth that diverges to infinity as $n\to\infty$, but the divergence is slow enough to ensure that $|\hat{\tau}^2-\tau^2| = o_p(1)$. Setting $\widehat{\mathrm{Var}}(\mathbf{N}_{\mathrm{x},k}) = \hat{\tau}^2\mathbf{C}_{\mathrm{x},k}$, where $\mathbf{C}_{\mathrm{x},k}$ is defined in (A.9) in Appendix F of the supplement, it holds that $\widehat{F}_{\mathrm{x},\lfloor nz\rfloor} \Rightarrow \widetilde{B}_1(z)$ for $z\in K$, where $\widehat{F}_{\mathrm{x},k} = \mathbf{N}'_{\mathrm{x},k}\widehat{\mathrm{Var}}(\mathbf{N}_{\mathrm{x},k})^{-1}\mathbf{N}_{\mathrm{x},k}$ and where $\widetilde{B}_1(z)$ is defined in (9).

We use related techniques to adjust $\widehat{F}_{\mathrm{s},k}$ and $\widehat{F}_{\mathrm{v},k}$ for autocorrelation. The process $\{\epsilon_t\mathbf{s}_t\}$ is not necessarily stationary, nor is the long-run variance of its partial sums sequence assured to be proportional to τ (i.e., it does not obey a version of Lemma A.1 of Robbins *et al.*, 2016, for general $\mathbf{s}_t$). Estimation of the long-run variance of $\{\mathbf{N}_{\mathrm{s},k}\}$ requires intra-period aggregation.

Specifically, let $\boldsymbol{e}_i = \sum_{t=T(i-1)+1}^{iT} \mathbf{s}_t \epsilon_t$ and $\hat{\boldsymbol{e}}_i = \sum_{t=T(i-1)+1}^{iT} \mathbf{s}_t \hat{\epsilon}_t$ for $i \in (1, \ldots, m)$ where $m = \lfloor n/T \rfloor$. Note that the sequence $\{\boldsymbol{e}_i\}$ is stationary—therefore, the variance of $\{\mathbf{N}_{\mathrm{s},k}\}$ may be approximated through incorporation of the term $\boldsymbol{\tau}_s := \lim_{m\to\infty} \frac{1}{Tm} \mathrm{Var}\left[\sum_{i=1}^{m} \boldsymbol{e}_i \boldsymbol{e}_i'\right]$. A consistent estimate of $\boldsymbol{\tau}_s$ is

$$\widehat{\boldsymbol{\tau}}_s = \frac{1}{Tm} \sum_{i=1}^{m} \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i' + \sum_{j=1}^{q_m} \left(1 - \frac{j}{q_m + 1}\right) \frac{1}{T(m-j)} \sum_{i=1}^{m-j} \left(\hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_{i+j}' + \hat{\boldsymbol{e}}_{i+j} \hat{\boldsymbol{e}}_i'\right), \tag{19}$$

In light of (12), it follows that $(n\widehat{\boldsymbol{\tau}}_s)^{-1/2} \mathbf{N}_{\mathrm{s},\lfloor nz \rfloor} \Rightarrow \mathbf{W}_{q_s}(z) - z\mathbf{W}_{q_s}(1)$.

To approximate the long-run variance of $\{\mathbf{N}_{\mathrm{v},k}\}$, use

$$\widehat{\boldsymbol{\tau}}_v = \frac{1}{n} \sum_{t=1}^{n} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' + \sum_{j=1}^{q_n} \left(1 - \frac{j}{q_n + 1}\right) \frac{1}{n-j} \sum_{t=1}^{n-j} \left(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t+j}' + \hat{\boldsymbol{\varepsilon}}_{t+j} \hat{\boldsymbol{\varepsilon}}_{t,}' \right), \tag{20}$$

where $\hat{\boldsymbol{\varepsilon}}_t = \hat{\epsilon}_t \widetilde{\mathbf{v}}_t$. Akin to (13), we see $(n\widehat{\boldsymbol{\tau}}_v)^{-1/2} \mathbf{N}_{\mathrm{v},\lfloor nz \rfloor} \Rightarrow \mathbf{W}_{q_v}(z) - z\mathbf{W}_{q_v}(1)$. Therefore, setting

$$\widehat{F}_k^* = \hat{\tau}^{-2} \mathbf{N}_{\mathrm{x},k}' \mathbf{C}_{\mathrm{x},k}^{-1} \mathbf{N}_{\mathrm{x},k} + \mathbf{N}_{\mathrm{s},k}' \widehat{\boldsymbol{\tau}}_s^{-1} \mathbf{N}_{\mathrm{s},k} + \mathbf{N}_{\mathrm{v},k}' \widehat{\boldsymbol{\tau}}_v^{-1} \mathbf{N}_{\mathrm{v},k}, \tag{21}$$

the following result is evident.

Theorem 2. *Assume that the conditions of Theorem 1 hold with the exception that the error sequence $\{\epsilon_t\}$ follows the general stationary structure outlined in Section 2. Further, assume that $\widehat{F}_k^*$ is in accordance with (21). It holds that*

$$\widehat{F}^* := \max_{\frac{k}{n} \in K} \widehat{F}_k^* \xrightarrow{\mathcal{D}} \sup_{z \in K} \widetilde{B}(z)$$

as $n \to \infty$, where $\widetilde{B}(z)$ is defined in (16).

Note that there are alternatives to the Bartlett-based method of estimation of the long-run variance terms (e.g., τ^2, $\boldsymbol{\tau}_s$ and $\boldsymbol{\tau}_v$) described above. For example, one may consider estimation of these terms through a spectral density (Andrews & Monahan, 1992) or through data-dependent bandwidths (Newey & West, 1994). However, previous work (e.g., Robbins *et al.*, 2011a) shows that in changepoint settings, tests that require such variance terms have performance issues in finite samples that extend to circumstances where the variance term is assumed known. We expect similar results to hold for our methods.

4.1 Tests Based on ARMA Residuals

An advantage of $\widehat{F}_k^*$ from (21) is that it does not impose a parametric model for the error sequence $\{\epsilon_t\}$. However, convergence of this statistic can be quite slow; this problem is exacerbated in the presence of strong autocorrelation. A solution to this issue, in line with the suggestions of several authors (e.g., Bai, 1993; Robbins *et al.*, 2011a, 2016), is to construct changepoint statistics using residuals frpm an autoregressive moving average (ARMA) model instead of OLS residuals. This effectively reduces the statistic to its white noise components; the resulting procedure is more stable across a wide array of autocorrelation structures.

In the remainder of this section, assume that the IID innovations sequence $\{Z_t\}$ generates the error sequence $\{\epsilon_t\}$ via ARMA formulation:

$$\epsilon_t - \phi_1 \epsilon_{t-1} - \cdots - \phi_{p_{\mathrm{ar}}} \epsilon_{t-p_{\mathrm{ar}}} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_{q_{\mathrm{ma}}} Z_{t-q_{\mathrm{ma}}}, \qquad t \in \mathbb{Z}, \tag{22}$$

where p_{ar} and q_{ma} are the ARMA orders; if the coefficients in the above define a stationary model, this formulation obeys (A.1) from Appendix A of the supplement and enables estimation of the innovations sequence. Specifically, the sequence of ARMA residuals $\{\hat{Z}_t\}$ is calculated to satisfy the recursion

$$\hat{Z}_t = \hat{\epsilon}_t - \hat{\hat{\phi}}_1 \hat{\epsilon}_{t-1} - \cdots - \hat{\hat{\phi}}_{p_{\mathrm{ar}}} \hat{\epsilon}_{t-p_{\mathrm{ar}}} - \hat{\hat{\theta}}_1 \hat{Z}_{t-1} - \cdots - \hat{\hat{\theta}}_{q_{\mathrm{ma}}} \hat{Z}_{t-q_{\mathrm{ma}}},$$

for all t. In the above, the $\hat{\hat{\phi}}_i$ and $\hat{\hat{\theta}}_j$ are consistent (under $\mathcal{H}_0$) estimators of ARMA parameters. Robbins *et al.* (2016) consider ARMA-based approahces for the trend component. They define $\mathbf{R}_{\mathrm{x},k} = \sum_{t=1}^{k} \mathbf{x}_t \hat{Z}_t$ and illustrate that

$$\frac{\mathbf{R}_{\mathrm{x},k}}{\hat{\sigma}\sqrt{n}} - \frac{\mathbf{N}_{\mathrm{x},k}}{\hat{\tau}\sqrt{n}} = o_p(1, k), \tag{23}$$

where $\hat{\sigma}^2 = \sum_{t=1}^{n} \hat{Z}_t^2/n$ estimates the white noise variance and where $\hat{\tau}^2$ was defined in (18), $\mathbf{N}_{\mathrm{x},k}$ is calculated in the same manner as in (8), and the $o_p(a_n, k)$ notation is as defined in (4). Then, they define an ARMA residuals-based analogue to $\widehat{F}_{\mathrm{x},k}$ via

$$\widehat{L}_{\mathrm{x},k} = \mathbf{R}'_{\mathrm{x},k} \mathbf{C}^{-1}_{\mathrm{x},k} \mathbf{R}_{\mathrm{x},k} / \hat{\sigma}^2.$$

It follows that the processes $\{\widehat{F}_{\mathrm{x},k}\}$ and $\{\widehat{L}_{\mathrm{x},k}\}$ are asymptotically equivalent, which implies that their maximally selected analogues have the same limit distribution.

Derivation of statistics based on ARMA residuals for seasonal and covariate components requires different approaches. To elaborate, note that the result in (23) is based off the fact that for a continuous function $f(\cdot)$ and for large n, $f((t+1)/n) \approx f(t/n)$—no such relationship holds for the sequences $\{\mathbf{s}_t\}$ or $\{\mathbf{v}_t\}$. That is, letting

$$\mathbf{R}_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t \hat{Z}_t \quad \text{and} \quad \mathbf{R}_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t \hat{Z}_t,$$

there is no expression akin to that of (23) that connects $\mathbf{R}_{\mathrm{s},k}$ to $\mathbf{N}_{\mathrm{s},k}$ or $\mathbf{R}_{\mathrm{v},k}$ to $\mathbf{N}_{\mathrm{v},k}$ in general settings. Further efforts to extract the asymptotic behavior of $\{\mathbf{R}_{\mathrm{s},k}\}$ and $\{\mathbf{R}_{\mathrm{v},k}\}$ directly do not bear fruit. To explain, Bai (1993) establishes an asymptotic equivalence between a partial sums sequence of ARMA residuals and an analogous partial sums sequence defined using the true ARMA errors; however, $\{\mathbf{R}_{\mathrm{s},k}\}$ and $\{\mathbf{R}_{\mathrm{v},k}\}$ do not yield similar results. I.e.,

$$\mathbf{R}_{\mathrm{s},k} - \sum_{t=1}^{k} \mathbf{s}_t Z_t = \mathcal{O}_p(\sqrt{n}, k) \quad \text{and} \quad \mathbf{R}_{\mathrm{v},k} - \sum_{t=1}^{k} \mathbf{v}_t Z_t = \mathcal{O}_p(\sqrt{n}, k),$$

and faster rates of convergence do not hold in general. Instead, we examine the processes

$$\mathbf{R}^*_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t \hat{Z}_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{s}_t \hat{Z}_t, \quad \text{and} \quad \mathbf{R}^*_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t \hat{Z}_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{v}_t \hat{Z}_t. \tag{24}$$

The limit behavior of these quantities (under both of $\mathcal{H}_0$ and $\mathcal{H}_1$) is established in the following lemma that is proven in the appendix.

Lemma 1. *Assume that the conditions for Theorem 1 hold with the exception that $\{\epsilon_t\}$ obeys the ARMA formulation in (22). Letting*

$$\mathbf{U}_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t Z_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{s}_t Z_t \quad \textit{and} \quad \mathbf{U}_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t Z_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{v}_t Z_t,$$

it holds that

$$\frac{\mathbf{R}^*_{\mathrm{s},k}}{\sqrt{n}} - \frac{\mathbf{U}_{\mathrm{s},k}}{\sqrt{n}} = o_p(1, k), \quad \textit{and} \quad \frac{\mathbf{R}^*_{\mathrm{v},k}}{\sqrt{n}} - \frac{\mathbf{U}_{\mathrm{v},k}}{\sqrt{n}} = o_p(1, k),$$

The large sample behavior of $\{\mathbf{U}_{\mathrm{s},k}\}$ and $\{\mathbf{U}_{\mathrm{v},k}\}$ follows from the case of IID regression errors as considered earlier. Specifically, the fact that (12) yields (14) similarly implies

$$(\hat{\sigma}^2 n\mathbf{D}_T)^{-1/2}\mathbf{U}_{\mathrm{s},\lfloor nz\rfloor} \Rightarrow \mathbf{W}_{q_s}(z) - z\mathbf{W}_{q_s}(1).$$

Likewise,

$$(\hat{\sigma}^2 n\widehat{\boldsymbol{\Sigma}}_v)^{-1/2}\mathbf{U}_{\mathrm{v},\lfloor nz\rfloor} \Rightarrow \mathbf{W}_{q_v}(z) - z\mathbf{W}_{q_v}(1).$$

Therefore, defining statistics for a change at time k via

$$\widehat{L}^*_{\mathrm{s},k} = \frac{(\mathbf{R}^*_{\mathrm{s},k})'\mathbf{D}_T^{-1}\mathbf{R}^*_{\mathrm{s},k}}{\hat{\sigma}^2 k(1-\frac{k}{n})} \quad \text{and} \quad \widehat{L}^*_{\mathrm{v},k} = \frac{(\mathbf{R}^*_{\mathrm{v},k})'\widehat{\boldsymbol{\Sigma}}_v^{-1}\mathbf{R}^*_{\mathrm{v},k}}{\hat{\sigma}^2 k(1-\frac{k}{n})}.$$

it follows that

$$\widehat{L}^*_{\mathrm{s},\lfloor nz\rfloor} \Rightarrow \frac{\mathbf{B}_{q_s}(z)'\mathbf{B}_{q_s}(z)}{z(1-z)} \quad \text{and} \quad \widehat{L}^*_{\mathrm{v},\lfloor nz\rfloor} \Rightarrow \frac{\mathbf{B}_{q_v}(z)'\mathbf{B}_{q_v}(z)}{z(1-z)}, \tag{25}$$

for $z \in K$, where $\{\mathbf{B}_d(z)\}$ is a d-dimensional Brownian bridge. Further, an omnibus test is defined via

$$\widehat{L}_k = \widehat{L}_{\mathrm{x},k} + \widehat{L}^*_{\mathrm{s},k} + \widehat{L}^*_{\mathrm{v},k}, \tag{26}$$

the limit distribution distribution of the maximally selected version of which is stated as follows.

Theorem 3. *Assume that the conditions of Lemma 1 hold and that $\{\widehat{L}_k\}$ is calculated in accordance with (26). Then, it holds that*

$$\widehat{L} := \max_{\frac{k}{n}\in K} \widehat{L}_k \xrightarrow{\mathcal{D}} \sup_{z\in K} \widetilde{B}(z)$$

as $n \to \infty$*, where* $\{\widetilde{B}(z)\}$ *is defined in (16).*

A proof of Theorem 3 is given in the supplement (Appendix F). Therein, it is illustrated that the processes $\{\mathbf{R}_{\mathrm{x},k}\}$, $\{\mathbf{R}^*_{\mathrm{s},k}\}$ and $\{\mathbf{R}^*_{\mathrm{v},k}\}$ are asymptotically uncorrelated. Otherwise, the theorem is a direct consequence of (23), Lemma 1 and (25). Arguments regarding the power of $\widehat{L}$ are given next.

The quantities $\widehat{F}$ and $\widehat{F}^*$ are derived from Wald-based expressions, therefore it can be presumed that these statistics will have power to detect changepoints. Further, existing theory (e.g., Bai, 1997) establishes consistency of a changepoint estimator (e.g., $\hat{c} = \arg\max_k \widehat{F}^*_k$) found using these statistics. However, we use the following result, which is also proven in the appendix, to demonstrate that $\widehat{L}_{\mathrm{x},k}$ has asymptotic power of one in the event that $\mathbf{\Delta}_x \neq \mathbf{0}$ (and likewise for $\widehat{L}^*_{\mathrm{s},k}$ and $\widehat{L}^*_{\mathrm{v},k}$). The corollary also shows that the respective changepoint estimators consistently estimate the changepoint time when written as a proportion of the sample size.

Corollary 1. *Assume that conditions of Theorem 3 hold; however, we relax the assumption that* $\mathcal{H}_0$ *is true (and therefore permit* $\mathcal{H}_1$ *to hold). It follows that*

$$\lim_{n\to\infty} P\left(\max_{\frac{k}{n}\in K} \widehat{L}_{\mathrm{x},k} > c_\alpha\right) = 1 \quad \textit{and} \quad n^{-1} \arg\max_k \widehat{L}_{\mathrm{x},k} \xrightarrow{\mathcal{P}} \kappa \quad \textit{if } \mathbf{\Delta}_x \neq 0$$

for any constant c_α *when* $c/n \to \kappa$ *where* $\xrightarrow{\mathcal{P}}$ *denotes convergence in probability as* $n \to \infty$*. Likewise,*

$$\lim_{n\to\infty} P\left(\max_{\frac{k}{n}\in K} \widehat{L}^*_{\mathrm{s},k} > c_\alpha\right) = 1 \quad \textit{and} \quad n^{-1} \arg\max_k \widehat{L}^*_{\mathrm{s},k} \xrightarrow{\mathcal{P}} \kappa \quad \textit{if } \mathbf{\Delta}_s \neq 0,$$

with

$$\lim_{n\to\infty} P\left(\max_{\frac{k}{n}\in K} \widehat{L}^*_{\mathrm{v},k} > c_\alpha\right) = 1 \quad \textit{and} \quad n^{-1}\arg\max_k \widehat{L}^*_{\mathrm{v},k} \xrightarrow{\mathcal{P}} \kappa \quad \textit{if } \boldsymbol{\Delta}_v \neq 0.$$

A direct consequence of Corollary 1 is that $\widehat{L}$ has asymptotic power of one and that $n^{-1}\arg\max_k \widehat{L}_k \xrightarrow{\mathcal{P}} \kappa$ when $\boldsymbol{\Delta} = (\boldsymbol{\Delta}'_x, \boldsymbol{\Delta}'_s, \boldsymbol{\Delta}'_v)' \neq \mathbf{0}$.

The supplemental materials provide additional theoretical results. Appendix C illustrates simplifications of the ARMA residuals-based statistic that are observed if $\mathbf{s}_t$ follows some commonly used expressions, including harmonic terms and categorical representations of the seasonal fluctuations. Also, Appendix D considers the more general circumstance where $\{\widetilde{\mathbf{v}}_t\}$ has a nonstationary mean structure.

5. Simulations

In this section, simulated data are used to study the efficacy of the proposed methods on samples of finite size. Although there are a myriad of ways that to generate data under the general regression models studied here, we focus on the following baseline model. Specifically, the vector of responses, $(Y_1, \ldots, Y_n)'$, is generated using a null hypothesis model of

$$Y_t = \alpha_1 + \alpha_2\left(\frac{t}{n}\right) + \beta_1 \cos\left(\frac{2\pi t}{12}\right) + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \gamma_1 v_{1,t} + \gamma_2 v_{2,t} + \epsilon_t,$$

for $t = 1, \ldots, n$, where we fix $n = 1000$. This model contains a linear trend, two harmonic terms that govern periodicity ($T = 12$), and a pair of stochastic covariates, $\{v_{1,t}\}$ and $\{v_{2,t}\}$.

The stochastic covariates are generated with $v_{j,t} = \zeta_{1,j} + \zeta_{2,j}(t/n) + u_{j,t}$, when $u_{j,t} = \Phi_j u_{j,t-1} + W_{j,t}$, for $j = 1, 2$, where Φ_1 and Φ_2 are autoregressive coefficients (we set $\Phi_1 = 0.5$ and $\Phi_2 = -0.2$) and $\{W_{1,t}\}$ and $\{W_{2,t}\}$ are (Gaussian) white noise processes. Lastly, the regression errors $\{\epsilon_t\}$ are generated using an ARMA(1,1) model satisfying $\epsilon_t - \phi\epsilon_{t-1} =$

$Z_t + \theta Z_{t-1}$, where $\{Z_t\}$ is also Gaussian white noise. All white noise processes have unit variance, although we set $\text{Cor}(W_{1,t}, W_{2,t}) = 0.3$ and $\text{Cor}(W_{j,t}, Z_t) = 0$ for $j = 1, 2$ (and no cross correlation exists at nonzero lags in these processes). The terms ϕ and θ will be varied throughout, and we fix $\zeta_{1,1} = \zeta_{1,2} = 0$ with $\zeta_{2,1} = 1$ and $\zeta_{2,2} = -0.5$.

The regression coefficients that define the mean function of Y_t are determined as follows. For $t \leq c$, where c is the changepoint time, we set $\alpha_1 = \alpha_2 = 1$, whereas for $t > c$, we use $\alpha_1 = \alpha_2 = 1 + \delta_x$. Likewise, for $t \leq c$, we use $\beta_1 = \beta_2 = 1$ and $\gamma_1 = \gamma_2 = 1$, with $\beta_1 = \beta_2 = 1 + \delta_s$ and $\gamma_1 = \gamma_2 = 1 + \delta_v$ for $t > c$. Note that $\boldsymbol{\Delta} = (\delta_x, \delta_x, \delta_s, \delta_s, \delta_v, \delta_v)'$ where δ_x, δ_s, and δ_v are treated as bandwidth parameters used to vary the magnitude of a change. Throughout, we use $c = n/2 = 500$.

We will study the empirical performance of the $\widehat{F}^*$ test of Theorem 2 and the $\widehat{L}$ test of Theorem 3. Specifically, we estimate size and power for a test of $\mathcal{H}_0 : \boldsymbol{\Delta} = \mathbf{0}$ against three separate alternative hypotheses:

- H1a – All regression coefficients are allowed to change at an unknown time c.
- H1c – Only the vector of coefficients that govern the periodicity, i.e., $(\beta_1, \beta_2)'$, is allowed to change at unknown time c.
- H1d – Only the vector of coefficients that govern the relationship with the covariates, i.e., $(\gamma_1, \gamma_2)'$, is allowed to change at unknown time c.

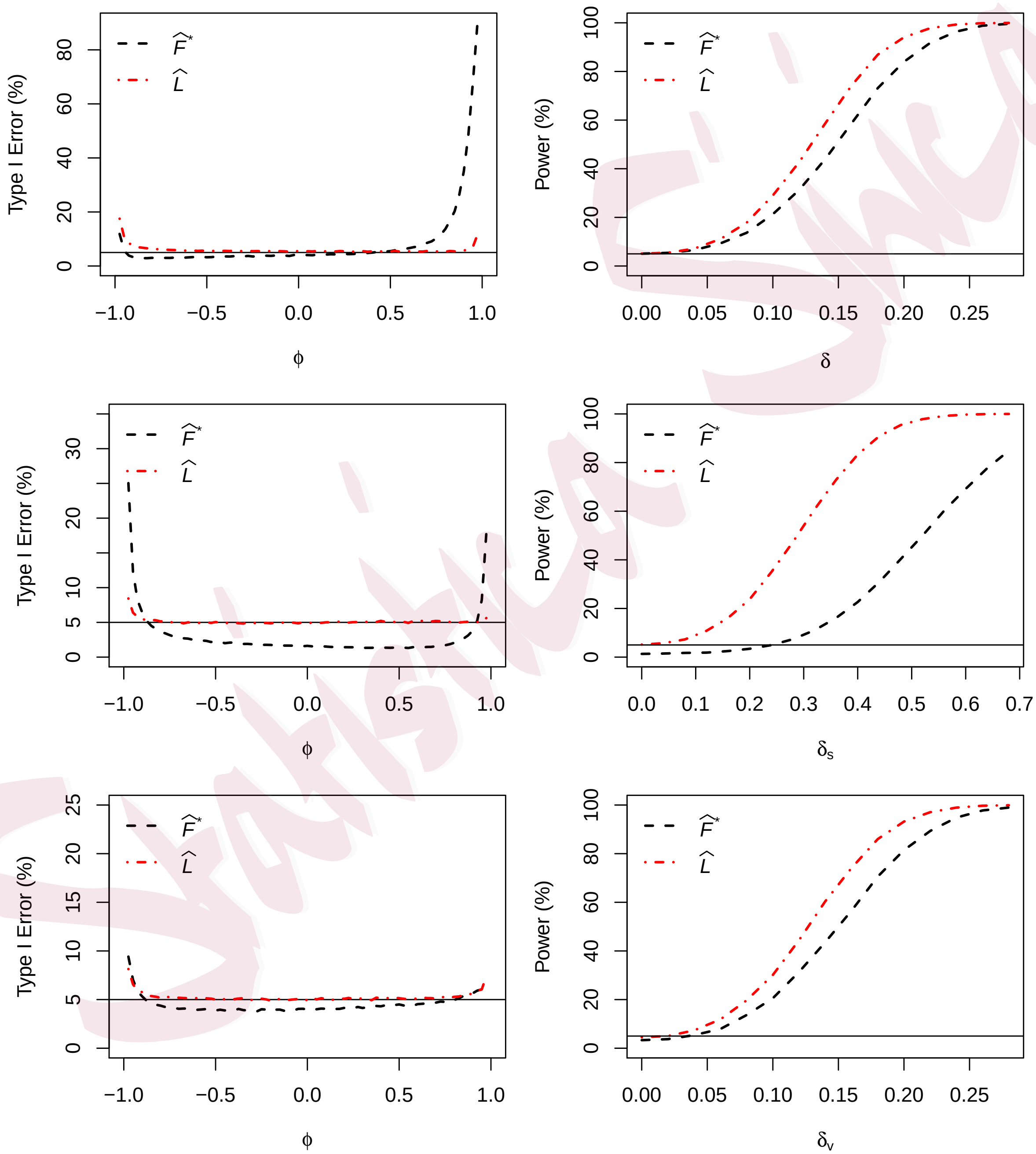
We test H1c using $\widehat{F}_k^* = \widehat{F}_{\mathrm{s},k}^*$ and $\widehat{L}_k = \widehat{L}_{\mathrm{s},k}$, whereas H1d is tested using $\widehat{F}_k^* = \widehat{F}_{\mathrm{v},k}^*$ and $\widehat{L}_k = \widehat{L}_{\mathrm{v},k}$. The setting where only the coefficients that govern the trend function are allowed to change (which would be alternative H1b here), was evaluated by Robbins *et al.* (2016) and has been studied in great detail by several other authors under less general regression models;

therefore, it is not included here (although it is considered within the data applications in Section 6). For the $\widehat{F}^*$ test, we use $q_n = \lfloor n^{1/3} \rfloor$ when calculating the expressions seen in (18) and (20), and thus $q_m = \lfloor (n/12)^{1/3} \rfloor$ when calculating the term in (19). For all tests, the set of admissible changepoint times is set as $\{k : 0.05n \leq k \leq 0.95n\}$. This selection is in line with earlier work (e.g., Robbins *et al.*, 2011a,b; Gallagher *et al.*, 2013), although alternatives should be considered if *a priori* knowledge suggests a wider or narrower bound (note that wider bounds will yield higher critical values for the test statistics).

As noted towards the end Section 3, in certain circumstances, closed form approximations for the limit process $\{\widetilde{B}(z)\}$ exist. Otherwise, critical values for the $\widehat{F}^*$ and $\widehat{L}$ statistics need to be derived by simulating realizations of the limit process. Throughout the remainder of the article, critical values of our test statistics are derived via simulation by simulating n discrete time points for each realization of $\{\widetilde{B}(z)\}$. We base critical values on 1,000,000 independently generated realizations of this process.

To begin, we study the empirical type I error (i.e., it is imposed that $\mathbf{\Delta} = \mathbf{0}$) of the tests. The size of the tests is most sensitive to the choice of the parameters that govern the autocorrelation within the regression errors, so results are provided for various choices of ϕ while fixing $\theta = 0$ (which implies $\{\epsilon_t\}$ follows an AR(1)). It is assumed that the correct ARMA order is known (although ϕ is estimated), and trends in the stochastic covariates are filtered prior to applying the methods. Findings are illustrated in Figure 1 (left column). The results uniformly indicate that the $\widehat{L}$ test is preferred to the $\widehat{F}^*$ test. The $\widehat{L}$ statistic gives well-controlled type I error across wide ranges of ϕ, which is not the case for the $\widehat{F}^*$ statistic. The $\widehat{F}^*$ test performs particularly poorly when considering alternative H1c; specifically, it is overly conservative. The process $\{\widehat{F}^*_{\mathrm{s},k}\}$ observes slow convergence due to difficulties in

Figure 1: Simulated type I error (left column) and power (right column) rates for tests based on the $\widehat{F}^*$ and $\widehat{L}$ statistics at a significance level of 0.05 with $n = 1000$. Results are shown for tests of three different alternative hypotheses: H1a (top row); H1c (middle row); H1d (bottom row). Type I error rates are shown as a function of the AR(1) parameter ϕ. Power is given for various choices of δ_x, δ_s, and δ_v (with $\phi = 0.5$), where $\delta_x = \delta_s = \delta_v = \delta$ when examining alternative H1a. Type I errors are based on 100,000 independently simulated datasets for each value of ϕ, whereas power is based on 25,000 datasets for each value of $\delta/\delta_s/\delta_v$



estimating the quantity in (19)—this quality is estimated using $\lfloor 1000/12 \rfloor = 83$ observations. For larger n, the empirical type I error is closer to the nominal value for this test.

Next, the power of these tests is examined. Figure 1 (right column) displays the power of the $\widehat{F}^*$ and $\widehat{L}$ statistics for tests of the alternative hypotheses mentioned above. When alternative H1a is examined, various values of $\delta = \delta_x = \delta_s = \delta_v$ are employed, whereas when alternatives H1c and H1d are considered, we vary the choice of δ_s and δ_v, respectively. These analyses use $\phi = 0.5$ and $\theta = 0$ throughout (where the correct ARMA order is again assumed to be known). The findings show that both tests have power to detect changepoints under each alternative examined, although the $\widehat{L}$ test shows power superior to that of $\widehat{F}^*$ (varying ϕ does not alter this conclusion).

Simulations for misspecified ARMA models are shown in the supplement (Appendix E).

6. Data applications

6.1 Mauna Loa CO_2

To examine the performance of the proposed methods on real data, the methodology is first applied to a time series of average atmospheric carbon dioxide (CO_2) levels measured monthly (in parts per million by volume) at the Mauna Loa Observatory on the island of Hawaii from March 1958 to June 2015 ($n = 688$). The series exhibits a marked increasing trend (which is often thought to be the byproduct of anthropogenic carbon emissions), and a multitude of authors have examined this series in search of structural breaks (e.g., Lund & Reeves, 2002; Beaulieu *et al.*, 2012; Robbins *et al.*, 2016)—these endeavors frequently yield evidence of a shift in the underlying trend structure that coincides with the eruption of Mount Pinatubo in June 1991. The CO_2 time series also observes pronounced periodicity. Seasonality in CO_2 levels is thought to be a direct consequence of vegetation growth; therefore, the periodic structure of the series has been heavily scrutinized within the climate literature. Many authors have claimed that the amplitude of the series has increased over time (Bacastow *et al.*, 1985; Buermann *et al.*, 2007; Zeng *et al.*, 2014).

Changes in the seasonal pattern of the CO_2 data have a variety of posited causes, including: a) increased prominence of droughts purportedly brought on by global warming; b) increased overall levels of vegetation, brought on by the higher overall CO_2 levels; and c) increased prominence of agriculture to accommodate a growing human population (Zeng *et al.*, 2014). Despite the wealth of literature on the issue, there does not appear to be a consensus as to the cause of the changes in the seasonal CO_2 fluctuations, nor have such changes been previously verified with rigorous statistical tools. This example illustrates the prudence of methods that examine changes in trend and seasonal structure separately since these aspects of the CO_2 data sequence are underpinned by different environmental processes.

The following null hypothesis model, which enables a quadratic trend function, is fit to the Mauna Loa CO_2 series. Let Y_t indicate CO_2 level at month t and assume

$$Y_t = \alpha_1 + \alpha_2\left(\frac{t}{n}\right) + \alpha_3\left(\frac{t}{n}\right)^2 + \sum_{j=1}^{4}\left[\beta_{1,j}\cos\left(\frac{2\pi jt}{12}\right) + \beta_{2,j}\sin\left(\frac{2\pi jt}{12}\right)\right] + \gamma_1 \mathrm{ENSO}_{t-12} + \epsilon_t, \tag{27}$$

where ENSO_t denotes the El Niño/Southern Oscillation index at month t (this index is used with a time lag of one year to improve predictive power). First, two separate alternatives to this model are considered:

- H1a – All regression coefficients are allowed to change at an unknown time c.
- H1b – Only elements in the vector of terms that govern the trend sequence, i.e., $(\alpha_1, \alpha_2, \alpha_3)$, are allowed to change at time c.

To test against alternative H1a, we use the OLS residuals-based statistic $\widehat{F}^*$ from Theorem 2 and the ARMA residuals-based statistic $\widehat{L}$ from Theorem 3. To test alternative H1c, we set $\widehat{F}_k^* = \widehat{F}_{x,k}^*$ and $\widehat{L}_k = \widehat{L}_{x,k}$ to account for the fact that only the trend function may change.

In order to study changes in the seasonal and covariate coefficients separately, the following revision of (27) is considered:

$$\begin{aligned} Y_t &= \alpha_1 + \alpha_2\left(\frac{t}{n}\right) + \alpha_3\left(\frac{t}{n}\right)^2 + \alpha_4\mathbf{1}_{\{t>c^*\}} + \alpha_5\left(\frac{t}{n}\right)\mathbf{1}_{\{t>c^*\}} + \alpha_6\left(\frac{t}{n}\right)^2\mathbf{1}_{\{t>c^*\}} \\ &\quad + \sum_{j=1}^{4}\left[\beta_{1,j}\cos\left(\frac{2\pi jt}{12}\right) + \beta_{2,j}\sin\left(\frac{2\pi jt}{12}\right)\right] + \gamma_1\mathrm{ENSO}_{t-12}, \end{aligned} \tag{28}$$

where $\mathbf{1}_{\{A\}}$ is the indicator of event A and where c^* is a known time that satisfies $1 \leq c^* < n$. The revised null hypothesis model is designed to incorporate a known shift in trend. Therefore, in the results shown in this section, c^* is the changepoint time estimated under a test of the hypothesis in H1b. Note that in order to satisfy Assumption 2 in Appendix A, changes in α_4, α_5, or α_6 cannot be allowed. We test (28) against alternative hypotheses described below:

- H1c* – Only elements in $(\beta_{1,1}, \beta_{2,1}, \ldots, \beta_{1,4}, \beta_{2,4})'$, the vector of terms that govern the periodicity, are allowed to change at time c.

- H1d* – Only γ_1 is allowed to change at time c.

These hypotheses may be tested using $\widehat{F}^*$ and $\widehat{L}^*$, where we set $\widehat{F}_k^* = \widehat{F}_{s,k}^*$ and $\widehat{L}_k = \widehat{L}_{s,k}$ for H1c* and $\widehat{F}_k^* = \widehat{F}_{v,k}^*$ and $\widehat{L}_k = \widehat{L}_{v,k}$ for H1d*. Under the notation of Theorem 1, the limit distribution of these statistics observes $\widetilde{B}(z) = \widetilde{B}_2(z)$ where $d = q_s$ for H1c* and $d = q_v$ for H1d*. Throughout this section, p-values are approximated using 1,000,000 independently simulated realizations of $\sup_{z\in K}\widetilde{B}(z)$—each realization is calculated using n discrete time points from the process $\{\widetilde{B}(z)\}$. As in Section 5, this section imposes that an admissible

Table 1: Results for application of the methodology to the Mauna Loa CO_2 time series, where ρ denotes a bandwidth parameter. When the $\widehat{F}^*$ test is used, we set $q_n = \rho$ when calculating the expressions seen in (18) and (20) and we set $q_m = \rho/12^{1/3}$ (rounded to the nearest integer) when calculating the term in (19); likewise, for the $\widehat{L}$ test, we set $p_{\text{ar}} = \rho$ and $q_{\text{ma}} = 0$ (i.e., an AR(ρ) is used). For hypotheses H1c* and H1d*, $c^* = 400$ is used.

	Alternative Model H1a				Alternative Model H1b				Alternative Model H1c*				Alternative Model H1d*			
	$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test	
ρ	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.
2	402	0.000	229	0.000	402	0.000	400	0.002	188	0.053	216	0.000	463	0.510	584	0.557
4	368	0.000	216	0.000	402	0.000	400	0.012	188	0.174	216	0.000	463	0.694	584	0.551
8	368	0.000	229	0.000	402	0.000	400	0.021	188	0.359	216	0.000	463	0.818	584	0.476
12	368	0.000	229	0.000	402	0.000	400	0.011	644	0.116	216	0.000	463	0.836	584	0.392
16	379	0.000	226	0.000	402	0.000	400	0.060	644	0.000	217	0.000	463	0.832	624	0.423
24	368	0.000	229	0.001	402	0.003	400	0.033	650	0.002	222	0.000	463	0.867	624	0.355

changepoint time k satisfies $0.05n \leq k \leq 0.95n$.

Table 1 shows results for application of the $\widehat{F}^*$ test of Theorem 2 and the $\widehat{L}$ test of Theorem 3 to the CO_2 data across each of the four alternative hypotheses mentioned above. Further, the table illustrates results for various choices of a bandwidth parameter ρ (see the description in the caption to the table), which governs the selection of terms like q_n, q_m, p_{ar} and q_{ma}. We prefer to use $\rho = 12$ when applying the $\widehat{L}$ statistic (as this value of p_{ar} ensures that there the ARMA residuals are devoid of autocorrelation) and $\rho = 8$ when using the $\widehat{F}^*$ statistic (which is consistent with the rule-of-thumb $q_n = n^{1/3}$). The value of $\widehat{F}^*$ is much more sensitive to the choice of bandwidth parameter than is the value of $\widehat{L}$; for this reason, and in accordance with the results of Section 5, the $\widehat{L}$ test is our preferred method. The shift in trend of CO_2 levels that occurs in June 1991 ($\hat{c} = 400$) and was detected by other authors is confirmed by the test of hypothesis H1b. However, shifts in trend are not the focus here (refer to the aforementioned references for discussion of the trend behavior in this dataset).

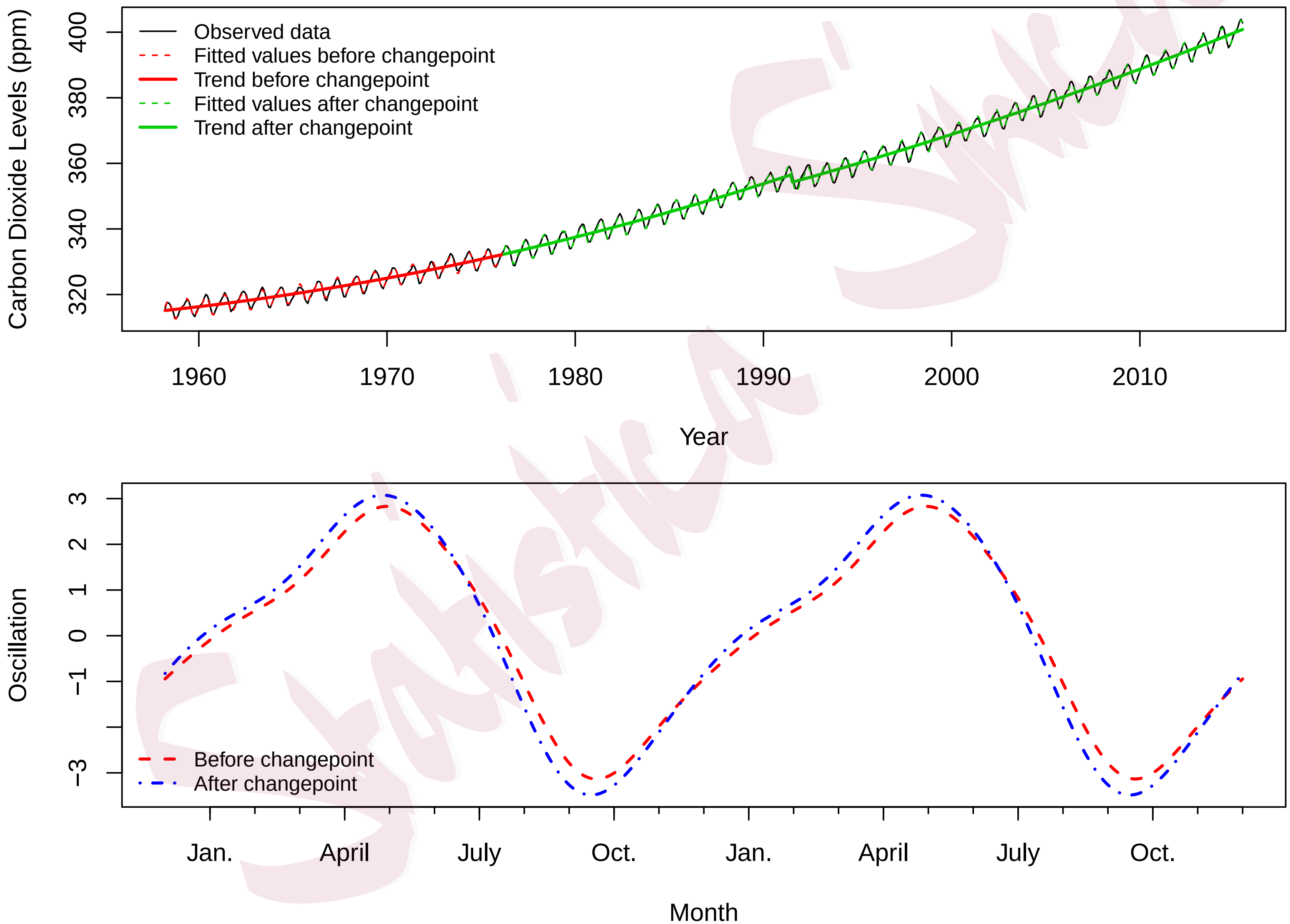
When all regression coefficients are allowed to shift (hypothesis H1a), we see a strongly significant changepoint that occurs in late 1976 (as estimated by the $\widehat{L}$ test), which "overrides" the discontinuity in trend that is often attributed to Mount Pinatubo. The test of the hypothesis H1c* indicates a strongly significant change in the parameters governing the seasonal pattern that also occurs in 1976. In short, it appears that the shift in the seasonal structure is given precedence when using the omnibus test of H1a—this fact is likely an artifact of the underlying mathematical model (there are more parameters that govern the seasonal behavior than those that govern the trend function).

Figure 2 includes a plot of the predicted CO_2 series overlaid on the observed data values; however, a change in seasonal behavior is not evident from visual investigation of the series. To offer further analysis, Figure 2 also shows the expected oscillations (i.e., departures) from the trend function for the CO_2 data. The pattern of oscillations observed before and after the changepoint are shown, where it is assumed that a change in the seasonal pattern occurs at time $\hat{c} = 216$, as estimated under alternative H1c*. The changepoint indicated by alternative H1c* indeed resulted in a subtle shift in the seasonal behavior. Specifically, the amplitude of the oscillations is 5.97 ppm prior to the changepoint and 6.56 ppm afterwards (which represents a 10% increase in amplitude). This finding is in line with the report of Zeng *et al.* (2014), who observe a 15% long term increase in the seasonal amplitude. OLS parameter estimates under the null and alternative for this example and the ensuing one are provided in Appendix G of the supplement.

6.2 Temperatures at Barrow, AK

We next apply the proposed methodology to a series of average monthly temperatures measured (in degrees Celsius) at the Wiley Post–Will Rogers Memorial Airport in Barrow, Alaska from October 1920 to June 2015 ($n = 1137$). At 515 kilometers north of the Arctic Circle, Barrow is the northernmost city in the United States. Due to obvious implications for melting of ice sheets, patterns of warming temperatures in polar climates such as Barrow

Figure 2: The fitted CO_2 values (with underlying trend function) as calculated using the model in alternative hypothesis H1c* (top) and a plot of the expected oscillations, i.e., seasonal variations from the trend, as estimated before and after the changepoint that was found under alternative H1c* (bottom)



have been heavily scrutinized in the scientific literature—it is frequently observed that polar regions have endured greater rates of warming than other regions (e.g., Holland & Bitz, 2003; Serreze & Francis, 2006). Differential rates of warming by season are also of interest. Using historical and simulated data, several researchers (e.g., Lu & Cai, 2009; Screen & Simmonds, 2010; Manabe *et al.*, 2011; Bintanja & Van der Linden, 2013) stipulate warming is occurring at higher rates during winter seasons than in summer seasons within polar regions.

Since a linear trend is frequently used to model temperature data, the null hypothesis model that will be fit to the Barrow temperatures is

$$Y_t = \alpha_1 + \alpha_2 \left(\frac{t}{n}\right) + \sum_{j=1}^{11} \beta_j s_{j,t} + \gamma_1 \text{ENSO}_{t-4} + \epsilon_t,$$

where $s_{j,t}$ obeys (A.4) from the supplemental materials with $T = 12$ and where ENSO_t again denotes the El Niño/Southern Oscillation index. The validity of this null hypothesis model is tested against alternative models indicated by corresponding versions of H1a (all coefficients are allowed to shift) and H1b (only the trend is allowed to shift). Given the presence of a shift in trend at some time c^*, we also test for a changepoint in only the seasonal coefficients $(\beta_1, \ldots, \beta_{11})$ (i.e., alternative H1c*) and for a change in only γ_1 (i.e., alternative H1d*).

Results are shown in Table 2. The omnibus changepoint test of alternative H1a indicates a statistically significant discontinuity at time $\hat{c} = 914$ (November 1996); whereas, when we search for a change only in trend (alternative H1b), the changepoint, though still statistically significant, is estimated to occur at $\hat{c} = 513$ (June 1963). When testing for a change in only the seasonal variations (alternative H1c*), the estimated changepoint time is again $\hat{c} = 914$. However, the change in seasonal variation is not statistically significant ($p = 0.265$). These

Table 2: Results for application of the methodology to the Barrow temperature time series, where ρ denotes a bandwidth parameter. See the caption to Table 1 for a description of ρ. Under hypotheses H1c* and H1d*, $c^* = 513$ is used.

	Alternative Model H1a				Alternative Model H1b				Alternative Model H1c*				Alternative Model H1d*			
	$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test		$\widehat{F}^*$ Test		$\widehat{L}$ Test	
ρ	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.	$\hat{c}$	p-val.
2	895	0.000	819	0.000	520	0.000	513	0.000	914	0.582	914	0.250	692	0.890	692	0.967
4	895	0.000	914	0.001	520	0.000	513	0.000	60	0.775	914	0.268	692	0.913	692	0.965
8	895	0.001	914	0.004	520	0.000	513	0.001	60	0.850	914	0.259	692	0.932	692	0.978
12	896	0.006	914	0.024	520	0.000	513	0.009	60	0.702	914	0.270	692	0.937	692	0.976
16	896	0.022	914	0.046	520	0.000	513	0.043	60	0.250	914	0.246	692	0.930	692	0.962
24	60	0.005	914	0.149	520	0.002	375	0.117	60	0.003	914	0.337	692	0.923	692	0.923

observations indicate that the omnibus statistic is dominated by the piece that is estimating the change in seasonality. Note that there are two parameters that control the trend, whereas there are 11 parameters that control seasonality. Therefore, the addition of the $\widehat{L}_x$ (which finds a significant changepoint) to the $\widehat{L}_s$ statistic (which finds a non-significant changepoint) is enough to result in an omnibus statistic that estimates a significant changepoint at the time given by the $\widehat{L}_s$ statistic. This example further emphasizes the need to test for changes in trend and seasonal variation separately.

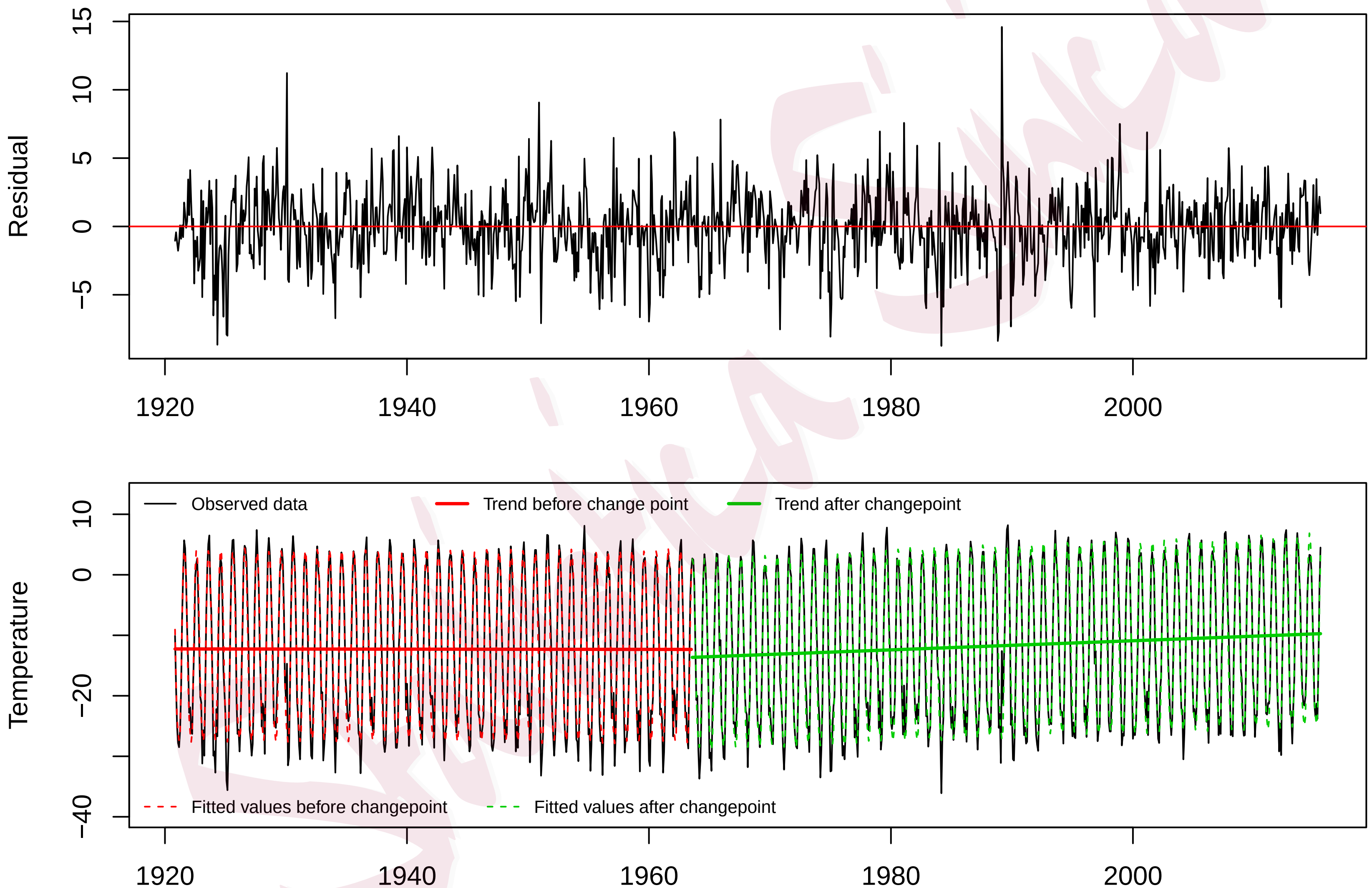
Figure 3 illustrates the OLS residuals that are estimated under alternative model H1a. The autocorrelation observed in these residuals is small (the lag-1 correlation is 0.27 under the null hypothesis); an AR(2) sufficiently captures the residual autocorrelation. Figure 3 also shows the temperature series with fitted values and trend when the change in trend has been included. The estimated rate of temperature increase under the null hypothesis is 2.19 °C per century; whereas, when a change in trend is allowed to occur at time $\hat{c} = 513$, it is estimated that temperature decreased at a rate of 0.19 °C per century prior to June 1963 and has increased at a rate of 7.54 °C per century since. This result is in line with findings given by other authors (e.g., Bloomfield, 1992; Jones & Moberg, 2003; Jones *et al.*, 2011) who observe relative stability in global temperatures from the mid 1940's to the mid

1970s followed since by an extended period of warming. However, the magnitude of warming that we observe over the last few decades in Barrow is substantially greater than the global rate of 2.06 °C per century from 1977-2001 as estimated by Jones & Moberg (2003)—this observation is in accordance the theory of amplified warming in the Arctic (Holland & Bitz, 2003). However, our study does not confirm previous literature that stipulates that the amount of amplification varies by season (Screen & Simmonds, 2010).

7. Discussion As was observed in prior work (Robbins *et al.*, 2011a, 2016), we saw herein that tests based on ARMA residuals ($\widehat{L}$) outperform those that fail to exploit the error structure ($\widehat{F}^*$). This is due in large part to difficulties with estimating a long-run variance term (τ^2) in finite samples. In addition, convergence of a partial sums sequence of independent terms ($\{Z_t\}$) is known to be quicker that that of autocorrelated variates ($\{\epsilon_t\}$).

The methods introduced here were developed for the AMOC setting. Although we illustrated that these methods can be used to detect shifts that occur at separate times in different components of the regression model, some discussion of the multiple changepoint setting is warranted. Segmentation (see, for example Menne & Williams Jr., 2009; Robbins *et al.*, 2011b) is often used to detect multiple changepoints with AMOC methods. This process works well so long as the shifts occur in a manner that yields a monotonic increasing or decreasing outcome. In other situations, such as those involving wavelets, it is usually necessary to consider procedures developed specifically for the multiple changepoint setting (e.g., Bai & Perron, 1998; Cho & Fryzlewicz, 2011; Yau & Zhao, 2016; Horvath *et al.*, 2017). Extension of such methods to the general framework considered here is left for future work. Furthermore, note that local alternatives (e.g. Andrews, 1993) were not discussed here; this is also left for future work.

Figure 3: The fitted temperature values (measured in °C) for the Barrow, AK series with underlying trend function after Alternative Model H1a has been fit (top) and the resulting OLS residuals (bottom)



An important innovation provided by the methodology introduced in herein is the ability to test for changes in a specific coefficient (or set of coefficients) of a large regression model. This enables the analyst to run a variety of tests to assess regression coefficients separately (as was done in Sections 5 and 6). To hedge against false-positive changepoint detections, the omnibus test (i.e., hypothesis H1a) should be considered prior to evaluating a variety of coefficients separately. Regardless, if a substantial number of tests are being applied within the same dataset, the analyst should consider corrections for multiple testing such as those that control the false discovery rate (Benjamini & Hochberg, 1995).

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* , 821–856.

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

Antoch, J., Hušková, M. and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference* **60**, 291–310.

Aue, A., Horváth, L. and Hušková, M. (2012). Segmenting mean-nonstationary time series via trending regressions. *Journal of Econometrics* **168**, 367–381.

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal* **9**, 373–403.

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2008). Testing for changes in polynomial regression. *Bernoulli* **14**, 637–660.

Bacastow, R., Keeling, C. and Whorf, T. (1985). Seasonal amplitude increase in atmospheric CO_2 concentration at mauna loa, hawaii, 1959–1982. *Journal of Geophysical Research: Atmospheres (1984–2012)* **90**, 10529–10540.

Bai, J. (1993). On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis* **14**, 247–260.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics* **79**, 551–563.

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

Beaulieu, C., Chen, J. and Sarmiento, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **370**, 1228–1249.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* , 289–300.

Bintanja, R. and Van der Linden, E. (2013). The changing seasonal climate in the arctic. *Scientific reports* **3**.

Bloomfield, P. (1992). Trends in global temperature. *Climatic Change* **21**, 1–16.

Buermann, W., Lintner, B. R., Koven, C. D., Angert, A., Pinzon, J. E., Tucker, C. J. and Fung, I. Y. (2007). The changing carbon cycle at mauna loa observatory. *Proceedings of the National Academy of Sciences* **104**, 4249–4254.

Cho, H. and Fryzlewicz, P. (2011). Multiscale interpretation of taut string estimation and its connection to unbalanced haar wavelets. *Statistics and Computing* **21**, 671–681.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* John Wiley & Sons Ltd.

Elsner, J. B., Bossak, B. H. and Niu, X.-F. (2001). Secular changes to the enso-us hurricane relationship. *Geophysical Research Letters* **28**, 4123–4126.

Gallagher, C., Lund, R. and Robbins, M. (2013). Changepoint detection in climate time series with long-term trends. *Journal of Climate* **26**, 4994–5006.

Gombay, E. (2010). Change detection in linear regression with time series errors. *Canadian Journal of Statistics* **38**, 65–79.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics* **97**, 93–115.

Holland, M. and Bitz, C. (2003). Polar amplification of climate change in coupled models. *Climate Dynamics* **21**, 221–232.

Horvath, L., Pouliot, W. and Wang, S. (2017). Detecting at-most-m changes in linear regression models. *Journal of Time Series Analysis* **38**, 552–590.

Jones, P., Wigley, T. and Wright, P. (2011). Global temperature variations between 1861 and 1984. *The Warming Papers* , 208.

Jones, P. D. and Moberg, A. (2003). Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of Climate* **16**, 206–223.

Lu, J. and Cai, M. (2009). Seasonality of polar surface warming amplification in climate simulations. *Geophysical Research Letters* **36**.

Lund, R. B. and Reeves, J. (2002). Detection of undocumented changepoints — a revision of the two-phase regression model. *Journal of Climate* **17**, 2547–2554.

MacNeill, I. (1978). Limit processes for sequences of partial sums of regression residuals. *The Annals of Probability* **6**, 695–698.

Manabe, S., Ploshay, J. and Lau, N.-C. (2011). Seasonal variation of surface temperature change during the last several decades. *Journal of Climate* **24**, 3817–3821.

Menne, M. J. and Williams Jr., C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal* *of Climate* **22**, 1700–1717.

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

Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011a). Mean shift testing in correlated data. *Journal of Time Series Analysis* **32**, 498–511.

Robbins, M., Lund, R., Gallagher, C. and Lu, Q. (2011b). Changepoints in the north atlantic tropical cyclone record. *Journal of the American Statistical Association* **106**, 89–99.

Robbins, M. W., Gallagher, C. M. and Lund, R. B. (2016). A general regression changepoint test for time series data. *Journal of the American Statistical Association* **111**, 670–683.

Screen, J. A. and Simmonds, I. (2010). The central role of diminishing sea ice in recent arctic temperature amplification. *Nature* **464**, 1334–1337.

Serreze, M. C. and Francis, J. A. (2006). The arctic amplification debate. *Climatic Change* **76**, 241–264.

Yau, C. Y. and Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**, 895–916.

Yu, H. (2007). High moment partial sum processes of residuals in arma models and their applications. *Journal of Time Series Analysis* **28**, 72–91.

Zeng, N., Zhao, F., Collatz, G. J., Kalnay, E., Salawitch, R. J., West, T. O. and Guanter, L. (2014). Agricultural green revolution as a driver of increasing atmospheric CO_2 seasonal amplitude. *Nature* **515**, 394–397.

RAND Corporation, Pittsburgh, PA 15213 U.S.A.

E-mail: (mrobbins@rand.org); Phone (1-412-683-2300 x4914)

SUPPLEMENT: A FULLY FLEXIBLE CHANGEPOINT TEST FOR REGRESSION MODELS WITH STATIONARY ERRORS

Michael W. Robbins

RAND Corporation

Appendix A: Technical Assumptions

The following assumptions, which are largely borrowed from work such as Robbins *et al.* (2016), are imposed on the predictor sequences $\{\widetilde{\mathbf{x}}_t\}$, $\{\widetilde{\mathbf{s}}_t\}$, and, $\{\widetilde{\mathbf{v}}_t\}$ and error sequence $\{\epsilon_t\}$:

Assumption 1. *The sequence* $\{\widetilde{\mathbf{v}}_t\}$ *satisfies a functional central limit theorem.*

Assumption 2. *If* $q_x > 0$*, the functions* f_1 *through* f_{q_x} *are continuous and differentiable over the set* K *of admissible changepoints. It is also imposed that* $f_j^2 > 0$ *over the set* K *for* $j = 1, \ldots, p_x$.

Assumption 3. *Let* $\widetilde{\boldsymbol{\chi}}_t = (\widetilde{\mathbf{x}}_t', \widetilde{\mathbf{s}}_t', \widetilde{\mathbf{v}}_t')'$. *The matrix* $n^{-1}\sum_{t=1}^{n} \widetilde{\boldsymbol{\chi}}_t \widetilde{\boldsymbol{\chi}}_t'$ *is invertible for each* $n \geq p_x + p_s + p_v$ *with probability 1 in that it has a probability limit with a minimum eigenvalue that is bounded away from zero.*

Assumption 4. *The regression errors* $\{\epsilon_t\}$ *are independent of the process* $\{\widetilde{\mathbf{v}}_t\}$ *and satisfy*

$$\epsilon_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{A.1}$$

where $\{Z_t\}$ *is a sequence of mean zero independent and identically distributed (IID) random innovations that have a variance denoted by* σ^2 *and a finite* $(2+\eta)^{\text{th}}$ *moment for some* $\eta > 0$. *Also, the causal coefficients* $\{\psi_j\}$ *have a geometrically decaying structure that obeys*

$|\psi_j| \leq \omega r^{-j}$ *for all* $j \geq 0$ *and some finite* ω *and* $r > 1$.

Appendix B: The Limit of $\mathbf{N}_{\mathrm{x},k}$

Define strict functional form versions of $\widetilde{\mathbf{x}}_t$ and $\mathbf{x}_t$ as $\widetilde{\boldsymbol{f}}(z) = (f_1(z), \ldots, f_{p_x}(z))'$ and $\boldsymbol{f}(z) = (f_1(z), \ldots, f_{q_x}(z))'$, respectively, for $z \in [0,1]$. Also, let

$$\mathbf{G}(z) = \int_0^z \boldsymbol{f}(u)\boldsymbol{f}(u)'du, \qquad \mathbf{G}^*(z) = \int_0^z \widetilde{\boldsymbol{f}}(u)\boldsymbol{f}(u)'du$$

$$\text{and} \qquad \widetilde{\mathbf{G}}(z) = \int_0^z \widetilde{\boldsymbol{f}}(u)\widetilde{\boldsymbol{f}}(u)'du.$$

Likewise, let

$$\boldsymbol{\Gamma}(z) = \int_0^z \boldsymbol{f}(u)dW(u) \quad \text{and} \quad \widetilde{\boldsymbol{\Gamma}}(z) = \int_0^z \widetilde{\boldsymbol{f}}(u)dW(u),$$

where $\{W(z)\}_{z\in[0,1]}$ is a Wiener process. Define

$$\boldsymbol{\Omega}(z) = \mathbf{G}(z) - \mathbf{G}^*(z)'\widetilde{\mathbf{G}}(1)^{-1}\mathbf{G}^*(z) \quad \text{and} \quad \boldsymbol{\Lambda}(z) = \boldsymbol{\Gamma}(z) - \mathbf{G}^*(z)'\widetilde{\mathbf{G}}(1)^{-1}\widetilde{\boldsymbol{\Gamma}}(1).$$

Robbins *et al.* (2016) prove that

$$(n\tau^2)^{-1}\widehat{\mathrm{Var}}(\mathbf{N}_{\mathrm{x},\lfloor nz\rfloor}) \Rightarrow \boldsymbol{\Omega}(z) \quad \text{and} \quad (n\tau^2)^{-1/2}\mathbf{N}_{\mathrm{x},\lfloor nz\rfloor} \Rightarrow \boldsymbol{\Lambda}(z), \tag{A.2}$$

as $n \to \infty$ for $z \in K$. The result in (9) follows directly from the above.

Appendix C: Specific Representations for $\mathbf{s}_t$

The form of the ARMA residuals-based statistic $\widehat{L}^*_{\mathrm{s},k}$ can be simplified if seasonal component $\mathbf{s}_t$ obeys one of a pair of commonly used representations. First, consider that $\mathbf{s}_t$ takes the harmonic form

$$\mathbf{s}_t = (\mathbf{s}'_{j_1,t}, \ldots, \mathbf{s}'_{j_\rho,t})' \quad \text{where} \quad \mathbf{s}_{j,t} = (\cos(2\pi jt/T), \sin(2\pi jt/T))' \tag{A.3}$$

for $j \in (j_1, \ldots, j_\rho) \subseteq (1, \ldots, T/2)$ and $\rho \leq p_s/2$. Further, assume that any terms contained within $\mathbf{s}^*_t$ are among those in $(\mathbf{s}'_{1,t}, \ldots, \mathbf{s}'_{T/2,t})'$ which are not in $\mathbf{s}_t$. If $\{\mathbf{s}_t\}$ and $\{\mathbf{s}^*_t\}$ follow this representation, the matrix $\mathbf{D}_T$ has a diagonal form; specifically, $\mathbf{D}_T = \mathbf{I}_{q_s}/2$ where $\mathbf{I}_d$ is an identity matrix in d dimensions. Consider a second situation where seasonality is modeled exhaustively (i.e., each season is allocated its own mean term through the use of dummy variables). In this case, it holds that $q_s = T - 1$ so long as $\mathbf{x}_t$ contains an intercept term. Further, write $\mathbf{s}_t = (s_{1,t}, \ldots, s_{T-1,t})'$ and let

$$s_{j,t} = \begin{cases} 1 - T^{-1}, & \text{if } (t-j)/T \text{ an integer,} \\ -T^{-1}, & \text{otherwise.} \end{cases} \tag{A.4}$$

This equation defines an indicator variable that has been centered so as to satisfy the requirement that $\sum_{t=1}^{T} \mathbf{s}_t = \mathbf{0}$. Since $\mathbf{s}_t$ exhaustively models the periodicity, $\mathbf{s}^*_t$ is empty. In this case, $\mathbf{D}_T = \mathbf{I}_{q_s} - T^{-1}\mathbf{J}$, where $\mathbf{J}$ is a $q_s \times q_s$ matrix of ones. Under either of the above formulations for $\mathbf{s}_t$, the quantity $\widehat{L}_k$ can be further simplified by replacing $\mathbf{R}^*_{\mathrm{s},k}$ with the process $\mathbf{R}_{\mathrm{s},k}$ due to the following (a proof of which is provided in the supplement.).

Corollary A.1. *Given the conditions of Theorem 3, assume that $\{\mathbf{s}_t\}$ obeys (A.3) or (A.4). Let $\mathbf{R}_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t \hat{Z}_t$ and*

$$\widehat{L}_{\mathrm{s},k} = \frac{\mathbf{R}'_{\mathrm{s},k}(\mathbf{D}_T)^{-1}\mathbf{R}_{\mathrm{s},k}}{\hat{\sigma}^2 k(1 - \frac{k}{n})}.$$

It follows that

$$\widehat{L}_{\mathrm{s},k} - \widehat{L}^*_{\mathrm{s},k} = o_p(1, k).$$

Appendix D: Nonstationary Stochastic Covariates

In the main text, we assumed that $\{\widetilde{\mathbf{v}}_t\}$ is stationary with zero mean. Now, we generalize to circumstances where $\{\widetilde{\mathbf{v}}_t\}$ has nonzero mean; specifically, consider that $\{\widetilde{\mathbf{v}}_t\}$ is generated via

$$\widetilde{\mathbf{v}}_t = \boldsymbol{\xi}'\boldsymbol{a}_t + \widetilde{\mathbf{u}}_t,$$

where $\{\widetilde{\mathbf{u}}_t\}$ (which is decomposed as $\widetilde{\mathbf{u}}_t = (\mathbf{u}_t', (\mathbf{u}_t^*)')'$ in the same manner as the other regressor vectors introduced in Sections 1 and 2) is stationary with zero mean, $\{\boldsymbol{a}_t\}$ is a vector of known deterministic design points, and $\boldsymbol{\xi}$ is a matrix of constants.

Assume that predictor sequence given by $\{\boldsymbol{a}_t\}$ is contained within the predictors in $\{(\mathbf{x}_t', \mathbf{s}_t')'\}$. It follows that the OLS residuals take on the same values when $\{\widetilde{\mathbf{v}}_t\}$ is used as a predictor as they do when $\{\widetilde{\mathbf{u}}_t\}$ is used in its place when fitting the regression (this is the case for residuals calculated under both the null and alternative hypotheses). Therefore, the sequences $\{\widehat{F}_k\}$ and $\{\widehat{F}_k^*\}$, defined in (6) and (21), respectively, are unchanged if $\{\widetilde{\mathbf{u}}_t\}$ were used in place of $\{\widetilde{\mathbf{v}}_t\}$, and the limit laws given in Theorems 1 and 2 still hold.

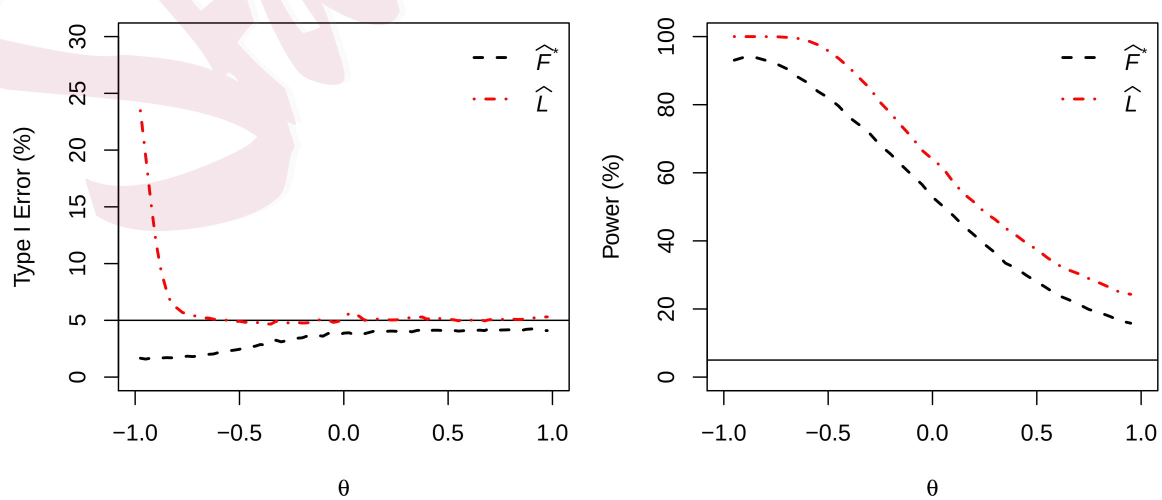
However, one must filter a nonstationary mean sequence out of $\{\mathbf{v}_t\}$ prior to calculating $\{\mathbf{R}^*_{\mathrm{v},k}\}$, as defined in (24). Let $\{\widehat{\mathbf{u}}_t\}$ denote residuals from a regression of $\{\mathbf{v}_t\}$ on $\{\boldsymbol{a}_t\}$ and let $\widehat{\boldsymbol{\xi}}_q$ denote a $\sqrt{n}$-consistent estimator of $\boldsymbol{\xi}_q$, where $\boldsymbol{\xi}_q$ gives the first q_v rows of $\boldsymbol{\xi}$. Define

$$\mathbf{R}^\dagger_{\mathrm{v},k} = \sum_{t=1}^{k} \widehat{\mathbf{u}}_t \hat{Z}_t - \frac{k}{n}\sum_{t=1}^{n} \widehat{\mathbf{u}}_t \hat{Z}_t.$$

If $\mathbf{R}^*_{\mathrm{v},k}$ is replaced with $\mathbf{R}^\dagger_{\mathrm{v},k}$ in the calculation of $\widehat{L}$, the convergence illustrated in Theorem 3 will hold. If $\boldsymbol{a}_t$ contains terms exogenous to $(\mathbf{x}_t', \mathbf{s}_t')'$, we recommend homogenizing $\{\widetilde{\mathbf{v}}_t\}$ prior fitting any regressions (and therefore prior to calculating test statistics). In this case, the limit theory outlined above applies; formal proof of these claims is omitted for brevity.

Appendix E: Additional Simulations The motivation for use of the $\widehat{F}^*$ statistic is that it does not impose a parametric model on the error structure. Therefore, we examine the performance of the $\widehat{L}$ statistic when the serial correlation in $\{\epsilon_t\}$ is not correctly modeled. Specifically, we generate $\{\epsilon_t\}$ using various values of θ while fixing $\phi = 0$ (this implies the errors are sampled from a MA(1) model). Then, when the $\widehat{L}$ statistic is calculated, an AR(p_{ar}) model, with p_{ar} selected using the AIC criterion, is fit to the regression residuals. The size of the $\widehat{F}^*$ and the misspecified $\widehat{L}$ tests are approximated under alternative H1a, and then the power for this alternative is calculated while fixing $\delta_x = \delta_s = \delta_v = 0.107$. Results are shown

Figure A.1: Simulated size (left) and power (right) of the $\widehat{F}^*$ and *misspecified* $\widehat{L}$ tests for a nominal significance level of 0.05 when alternative model H1a is considered and when the error sequence $\{\epsilon_t\}$ is generated from an MA(1) model with parameter θ with $n = 1000$. Results are shown for various choices of θ, where $\delta_x = \delta_s = \delta_v = 0.107$ for power comparisons. Results for size are based on 100,000 independently simulated datasets for each value of θ, whereas 25,000 datasets are generated for power calculations



in Figure A.1. The findings indicate that the $\widehat{L}$ statistic still outperforms the $\widehat{F}^*$ statistic (with regards to both size and power), even when the error model is incorrectly specified. It is expected that power for both tests will decrease as θ increases (Robbins *et al.*, 2011a).

Appendix F: Proofs

Theorem 1. As is stipulated by the conditions of Theorem 1, this proof assumes IID regression errors (i.e., $\epsilon_t = Z_t$ and thus $\tau^2 = \sigma^2$). To begin, let

$$\mathbf{X}_t = (\mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{0}, \ldots, \mathbf{0})' \quad \text{and} \quad \widetilde{\mathbf{X}}_t = (\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_t, \mathbf{0}, \ldots, \mathbf{0})',$$

which are matrices of dimension $n \times q_x$ and $n \times p_x$, respectively, the last $n-t$ rows of which contain zeros. Similarly, define

$$\mathbf{S}_t = (\mathbf{s}_1, \ldots, \mathbf{s}_t, \mathbf{0}, \ldots, \mathbf{0})' \quad \text{and} \quad \widetilde{\mathbf{S}}_t = (\widetilde{\mathbf{s}}_1, \ldots, \widetilde{\mathbf{s}}_t, \mathbf{0}, \ldots, \mathbf{0})',$$

and

$$\mathbf{V}_t = (\mathbf{v}_1, \ldots, \mathbf{v}_t, \mathbf{0}, \ldots, \mathbf{0})' \quad \text{and} \quad \widetilde{\mathbf{V}}_t = (\widetilde{\mathbf{v}}_1, \ldots, \widetilde{\mathbf{v}}_t, \mathbf{0}, \ldots, \mathbf{0})'.$$

Further, let $\mathbf{M}_t = (\mathbf{X}_t, \mathbf{S}_t, \mathbf{V}_t)$ and $\widetilde{\mathbf{M}}_t = (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{S}}_t, \widetilde{\mathbf{V}}_t)$. Note that $\mathbf{M}_n$ is the full design matrix under $\mathcal{H}_0$. The null hypothesis OLS estimator of $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_x', \boldsymbol{\Delta}_s', \boldsymbol{\Delta}_v')'$, when a changepoint is assumed to occur at time k with $k/n \in K$, is $\widehat{\boldsymbol{\Delta}}_k = -\mathbf{C}_k^{-1}\mathbf{N}_k$ where

$$\mathbf{C}_k = \mathbf{M}_k'(\mathbf{I} - \mathbf{P}_n)\mathbf{M}_k, \tag{A.5}$$

and

$$\mathbf{N}_k = \mathbf{M}_k'(\mathbf{I} - \mathbf{P}_n)\mathbf{Y}, \tag{A.6}$$

with $\mathbf{Y} = (Y_1, \ldots, Y_n)'$. In the above, $\mathbf{P}_n = \widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1}\widetilde{\mathbf{M}}_n'$ is the projection matrix under the null hypothesis. Conditional on $\widetilde{\mathbf{V}}_n$, it holds that $\mathrm{Var}(\widehat{\boldsymbol{\Delta}}_k) = \tau^2\mathbf{C}_k^{-1}$.

Note that

$$n^{-1}\mathbf{M}_k'\mathbf{M}_k = \begin{pmatrix} n^{-1}\mathbf{X}_k'\mathbf{X}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n^{-1}\mathbf{S}_k'\mathbf{S}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n^{-1}\mathbf{V}_k'\mathbf{V}_k \end{pmatrix} + o_p(1,k).$$

Lemmas A.1 and A.2 of Robbins *et al.* (2016) are used to show that the off-diagonal blocks of the above matrix are zero asymptotically. Similarly,

$$(n^{-1}\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1} = \begin{pmatrix} (n^{-1}\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (n^{-1}\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (n^{-1}\widetilde{\mathbf{V}}_n'\widetilde{\mathbf{V}}_n)^{-1} \end{pmatrix} + \mathcal{O}_p(n^{-1/2}).$$

Likewise, we now see

$$\begin{aligned} n^{-1}\mathbf{M}_k'\widetilde{\mathbf{M}}_n &= n^{-1}\begin{pmatrix} \mathbf{X}_k'\widetilde{\mathbf{X}}_n & \mathbf{X}_k'\widetilde{\mathbf{S}}_n & \mathbf{X}_k'\widetilde{\mathbf{V}}_n \\ \mathbf{S}_k'\widetilde{\mathbf{X}}_n & \mathbf{S}_k'\widetilde{\mathbf{S}}_n & \mathbf{S}_k'\widetilde{\mathbf{V}}_n \\ \mathbf{V}_k'\widetilde{\mathbf{X}}_n & \mathbf{V}_k'\widetilde{\mathbf{S}}_n & \mathbf{V}_k'\widetilde{\mathbf{V}}_n \end{pmatrix} \\ &= \begin{pmatrix} n^{-1}\mathbf{X}_k'\widetilde{\mathbf{X}}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n^{-1}\mathbf{S}_k'\widetilde{\mathbf{S}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n^{-1}\mathbf{V}_k'\widetilde{\mathbf{V}}_n \end{pmatrix} + o_p(1,k). \end{aligned}$$

Continuing, we see

$$\mathbf{M}_k'\widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1} = \begin{pmatrix} \mathbf{X}_k'\widetilde{\mathbf{X}}_n(\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_k'\widetilde{\mathbf{S}}_n(\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_k'\widetilde{\mathbf{V}}_n(\widetilde{\mathbf{V}}_n'\widetilde{\mathbf{V}}_n)^{-1} \end{pmatrix} + o_p(1,k). \quad \text{(A.7)}$$

Hence,

$$n^{-1}\mathbf{C}_k = \begin{pmatrix} n^{-1}\mathbf{C}_{\mathrm{x},k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n^{-1}\mathbf{C}_{\mathrm{s},k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n^{-1}\mathbf{C}_{\mathrm{v},k} \end{pmatrix} + o_p(1,k), \quad \text{(A.8)}$$

where

$$\mathbf{C}_{\mathrm{x},k} = \mathbf{X}_k'\mathbf{X}_k - \mathbf{X}_k'\widetilde{\mathbf{X}}_n(\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n)^{-1}\widetilde{\mathbf{X}}_n'\mathbf{X}_k, \quad \text{(A.9)}$$

with

$$\mathbf{C}_{\mathrm{s},k} = \mathbf{S}_k'\mathbf{S}_k - \mathbf{S}_k'\widetilde{\mathbf{S}}_n(\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n)^{-1}\widetilde{\mathbf{S}}_n'\mathbf{S}_k, \quad \text{(A.10)}$$

and

$$\mathbf{C}_{\mathrm{v},k} = \mathbf{V}_k'\mathbf{V}_k - \mathbf{V}_k'\widetilde{\mathbf{V}}_n(\widetilde{\mathbf{V}}_n'\widetilde{\mathbf{V}}_n)^{-1}\widetilde{\mathbf{V}}_n'\mathbf{V}_k,$$

for $\mathbf{C}_k$ is defined in (A.5).

Shifting the focus to the process $\{\mathbf{N}_k\}$, we first note that

$$\mathbf{N}_k = \begin{pmatrix} \mathbf{N}_{\mathrm{x},k} \\ \mathbf{N}_{\mathrm{s},k} \\ \mathbf{N}_{\mathrm{v},k} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_k'\boldsymbol{\epsilon} - \mathbf{X}_k'\widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1}\widetilde{\mathbf{M}}_n'\boldsymbol{\epsilon} \\ \mathbf{S}_k'\boldsymbol{\epsilon} - \mathbf{S}_k'\widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1}\widetilde{\mathbf{M}}_n'\boldsymbol{\epsilon} \\ \mathbf{V}_k'\boldsymbol{\epsilon} - \mathbf{V}_k'\widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1}\widetilde{\mathbf{M}}_n'\boldsymbol{\epsilon} \end{pmatrix},$$

for $\mathbf{N}_k$ as defined in (A.6) and for $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. Recall that $\mathbf{N}_{\mathrm{x},k}$, $\mathbf{N}_{\mathrm{s},k}$, and $\mathbf{N}_{\mathrm{v},k}$ were defined in (8). Using $\mathrm{Var}(\mathbf{N}_k) = \tau^2\mathbf{C}_k$ and (A.8), it holds that these three processes are asymptotically uncorrelated. Applying the result of (A.7), we see

$$n^{-1/2}\begin{pmatrix} \mathbf{N}_{\mathrm{x},k} \\ \mathbf{N}_{\mathrm{s},k} \\ \mathbf{N}_{\mathrm{v},k} \end{pmatrix} = n^{-1/2}\begin{pmatrix} \mathbf{X}_k'\boldsymbol{\epsilon} - \mathbf{X}_k'\widetilde{\mathbf{X}}_n(\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n)^{-1}\widetilde{\mathbf{X}}_n'\boldsymbol{\epsilon} \\ \mathbf{S}_k'\boldsymbol{\epsilon} - \mathbf{S}_k'\widetilde{\mathbf{S}}_n(\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n)^{-1}\widetilde{\mathbf{S}}_n'\boldsymbol{\epsilon} \\ \mathbf{V}_k'\boldsymbol{\epsilon} - \mathbf{V}_k'\widetilde{\mathbf{V}}_n(\widetilde{\mathbf{V}}_n'\widetilde{\mathbf{V}}_n)^{-1}\widetilde{\mathbf{V}}_n'\boldsymbol{\epsilon} \end{pmatrix} + o_p(1,k). \tag{A.11}$$

The processes $\{\mathbf{C}_{\mathrm{x},k}\}$ and $\{\mathbf{N}_{\mathrm{x},k}\}$ were studied in the proof of Lemma 2.1 in Robbins *et al.* (2016); the focus now turns to $\{\mathbf{C}_{\mathrm{s},k}\}$ and $\{\mathbf{N}_{\mathrm{s},k}\}$.

Let $\mathbf{D}_T = \sum_{j=1}^T \mathbf{s}_j\mathbf{s}_j'/T$ with $\mathbf{D}_T^* = \sum_{j=1}^T \widetilde{\mathbf{s}}_j\mathbf{s}_j'/T$ and $\widetilde{\mathbf{D}}_T = \sum_{j=1}^T \widetilde{\mathbf{s}}_j\widetilde{\mathbf{s}}_j'/T$. Consequentially,

$$n^{-1}\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n = \widetilde{\mathbf{D}}_T + \mathcal{O}(n^{-1}),$$

and

$$n^{-1}\mathbf{S}_k'\widetilde{\mathbf{S}}_n = (k/n)\mathbf{D}_T^* + \mathcal{O}(n^{-1},k), \quad \text{with} \quad n^{-1}\mathbf{S}_k'\mathbf{S}_k = (k/n)\mathbf{D}_T + \mathcal{O}(n^{-1},k).$$

To derive (10), note that

$$\begin{aligned} n^{-1}\mathbf{C}_{\mathrm{s},\lfloor nz\rfloor} &\Rightarrow z\mathbf{D}_T - z^2(\mathbf{D}_T^*)'(\widetilde{\mathbf{D}}_T)^{-1}\mathbf{D}_T^*, \\ &= z\mathbf{D}_T - z^2\begin{pmatrix} \mathbf{I}_{p_s} & \mathbf{0} \end{pmatrix}\mathbf{D}_T^*, \\ &= z(1-z)\mathbf{D}_T. \end{aligned}$$

The second line follows from the fact that $\mathbf{D}_T^*$ equals the first q_s columns of $\widetilde{\mathbf{D}}_T$, and similarly the third line uses the observation that the first q_s rows of $\mathbf{D}_T^*$ equal $\mathbf{D}_T$. Likewise,

$$\begin{aligned}
\mathbf{N}_{\mathrm{s},k} &= \mathbf{S}_k'\boldsymbol{\epsilon} - (k/n)(\widetilde{\mathbf{D}}_T^*)'(\widetilde{\mathbf{D}}_T)^{-1}\widetilde{\mathbf{S}}_n'\boldsymbol{\epsilon} + o_p(\sqrt{n}, k) \\
&= \mathbf{S}_k'\boldsymbol{\epsilon} - (k/n)\begin{pmatrix} \mathbf{I}_{q_s} & \mathbf{0} \end{pmatrix}^{-1}\widetilde{\mathbf{S}}_n'\boldsymbol{\epsilon} + o_p(\sqrt{n}, k) \\
&= \mathbf{S}_k'\boldsymbol{\epsilon} - (k/n)\mathbf{S}_n'\boldsymbol{\epsilon} + o_p(\sqrt{n}, k),
\end{aligned}$$

which illustrates (12).

Recall that $\boldsymbol{e}_i = \sum_{t=T(i-1)+1}^{iT} \mathbf{s}_t\epsilon_t$ and note that $\mathbf{S}_k'\boldsymbol{\epsilon} = \sum_{i=1}^{m^*} \boldsymbol{e}_i$, where it is assumed that $n = Tm$ and $k = Tm^*$. Note further that $\mathrm{Var}(\boldsymbol{e}_i) = \tau^2 T\mathbf{D}_T$. It follows that

$$(T/n)^{1/2}\mathbf{N}_{\mathrm{s},k} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m^*}\boldsymbol{e}_i - \frac{k}{n}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\boldsymbol{e}_i\right) + o_p(1, k),$$

This formula, in combination with (10), yields (14).

Similar approaches are taken to extract the limit behavior of $\mathbf{C}_{\mathrm{v},k}$ and $\mathbf{N}_{\mathrm{v},k}$. Define $\widetilde{\boldsymbol{\Sigma}}_v = \mathrm{E}\left[\widetilde{\mathbf{v}}_t\widetilde{\mathbf{v}}_t'\right]$, let $\boldsymbol{\Sigma}_v^*$ denote the first q_v columns of $\widetilde{\boldsymbol{\Sigma}}_v$ and let $\boldsymbol{\Sigma}_v$ denote the first q_v rows of $\boldsymbol{\Sigma}_v^*$. Furthermore,

$$n^{-1}\widetilde{\mathbf{V}}_n'\widetilde{\mathbf{V}}_n \to \widetilde{\boldsymbol{\Sigma}}_v, \quad n^{-1}\mathbf{V}_{\lfloor nz\rfloor}'\widetilde{\mathbf{V}}_n \Rightarrow z\boldsymbol{\Sigma}_v^*, \quad \text{and} \quad n^{-1}\mathbf{V}_{\lfloor nz\rfloor}'\mathbf{V}_{\lfloor nz\rfloor} \Rightarrow z\boldsymbol{\Sigma}_v.$$

Formulas (11) and (13) are derived using arguments akin to those that provide (10) and (12). Specifically,

$$n^{-1}\mathbf{C}_{\mathrm{v},\lfloor nz\rfloor} \Rightarrow z(1-z)\boldsymbol{\Sigma}_v$$

and

$$\mathbf{N}_{\mathrm{v},k} = \mathbf{V}_k'\boldsymbol{\epsilon} - (k/n)\mathbf{V}_n'\boldsymbol{\epsilon} + o_p(\sqrt{n}, k).$$

Note that the sequence $\{\mathbf{v}_t\epsilon_t\}$ is devoid of autocorrelation and observes $\mathrm{Var}(\mathbf{v}_t\epsilon_t) = \tau^2\mathbf{\Sigma}_v$. Therefore, the identity in (15) is now evident.

From (A.8) and (A.11), it follows that

$$\widehat{F}_k = \frac{\mathbf{N}'_{\mathrm{x},k}\mathbf{C}^{-1}_{\mathrm{x},k}\mathbf{N}_{\mathrm{x},k}}{\hat{\tau}^2} + \frac{\mathbf{N}'_{\mathrm{s},k}(\mathbf{D}_T)^{-1}\mathbf{N}_{\mathrm{s},k}}{\hat{\tau}^2 k(1-\frac{k}{n})} + \frac{\mathbf{N}'_{\mathrm{v},k}(\widehat{\mathbf{\Sigma}}_v)^{-1}\mathbf{N}_{\mathrm{v},k}}{\hat{\tau}^2 k(1-\frac{k}{n})} + o_p(1,k),$$

where $\widehat{F}_k$ is defined in (6). The limit behavior of the term involving $\mathbf{N}_{\mathrm{x},k}$ follows from (A.2), and the limit behavior of the term involving $\mathbf{N}_{\mathrm{s},k}$ follows from (10) and (12). Likewise, the limit distribution of the term involving $\mathbf{N}_{\mathrm{v},k}$ follows from (11) and (13). The block-diagonal form of $\mathrm{Var}(\mathbf{N}_k) = \tau^2\mathbf{C}_k$ as $n \to \infty$, which is evident in (A.8), implies pairwise asymptotic independence of $\mathbf{N}_{\mathrm{x},k}$, $\mathbf{N}_{\mathrm{s},k}$ and $\mathbf{N}_{\mathrm{v},k}$. To establish (asymptotic) process independence, calculations similar to those which yield the form of $\mathbf{C}_k$ can be used to establish that $\mathbf{N}_{\mathrm{x},k}$ and $\mathbf{N}_{\mathrm{s},k'}$, for example, are asymptotically uncorrelated for $k \neq k'$. □

Lemma 1. Let $\{\mathbf{b}_t\}$ and $\{\widetilde{\mathbf{b}}_t\}$ be sequences of vectors that satisfy

$$\sum_{t=1}^{k}\mathbf{b}_t\hat{\epsilon}_{t-i} = \sum_{t=1}^{k}\mathbf{b}_t\epsilon_{t-i} - \sum_{t=1}^{k}\mathbf{b}_{t+i}\widetilde{\mathbf{b}}_t\left(\sum_{t=1}^{n}\widetilde{\mathbf{b}}_t\widetilde{\mathbf{b}}'_t\right)^{-1}\sum_{t=1}^{n}\widetilde{\mathbf{b}}_t\epsilon_t + o_p(\sqrt{n},k), \tag{A.12}$$

where $\{\hat{\epsilon}_t\}$ is the sequence of OLS residuals generated using (5) and where $\{\epsilon_t\}$ is the sequence of regression errors generated from the white noise ARMA errors $\{Z_t\}$ in accordance with (22). Assume further that

$$\frac{1}{n}\sum_{t=1}^{\lfloor nz\rfloor}\mathbf{b}_{t+i}\widetilde{\mathbf{b}}'_t \Rightarrow z\mathbf{\Gamma}_b(i) \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{b}}_t\widetilde{\mathbf{b}}'_t \to \widetilde{\mathbf{\Gamma}}_b(0) \tag{A.13}$$

for some sequence of deterministic matrices $\{\mathbf{\Gamma}_b(i)\}$ with arbitrary $i \geq 0$ and for some matrix

$\widetilde{\mathbf{\Gamma}}_b(0)$. It follows that

$$\begin{aligned}\sum_{t=1}^{k}\mathbf{b}_t\hat{\epsilon}_{t-i}-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{\epsilon}_{t-i} &= \sum_{t=1}^{k}\mathbf{b}_t\epsilon_{t-i}-(k/n)\mathbf{\Gamma}_b(i)[\widetilde{\mathbf{\Gamma}}_b(0)]^{-1}\sum_{t=1}^{n}\widetilde{\mathbf{b}}_t\epsilon_t\\ &\quad -\frac{k}{n}\left(\sum_{t=1}^{n}\mathbf{b}_t\epsilon_{t-i}-\mathbf{\Gamma}_b(i)[\widetilde{\mathbf{\Gamma}}_b(0)]^{-1}\sum_{t=1}^{n}\widetilde{\mathbf{b}}_t\epsilon_t\right)+o_p(\sqrt{n},k)\\ &= \sum_{t=1}^{k}\mathbf{b}_t\epsilon_{t-i}-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\epsilon_{t-i}+o_p(\sqrt{n},k). \end{aligned} \tag{A.14}$$

Let the sequence $\{\pi_j\}_{j=0}^{\infty}$ denote the coefficients from the expansion of $(1-\phi_1 z-\cdots-\phi_p z^p)/(1+\theta_1 z+\cdots+\theta_q z^q)$, and let $\{\hat{\pi}_j\}_{j=0}^{\infty}$ represent the versions of these coefficients when calculated using the $\hat{\phi}_j$ and $\hat{\theta}_j$. Hence,

$$Z_t=\sum_{j=0}^{\infty}\pi_j\epsilon_{t-j} \quad \text{and} \quad \hat{Z}_t=\sum_{j=0}^{\infty}\hat{\pi}_j\hat{\epsilon}_{t-j}, \tag{A.15}$$

and

$$\begin{aligned}\sum_{t=1}^{k}\mathbf{b}_t\hat{Z}_t-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{Z}_t &= \sum_{j=0}^{\infty}\hat{\pi}_j\left(\sum_{t=1}^{k}\mathbf{b}_t\hat{\epsilon}_{t-j}-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{\epsilon}_{t-j}\right)\\ &= \sum_{j=0}^{\infty}\pi_j\left(\sum_{t=1}^{k}\mathbf{b}_t\epsilon_{t-j}-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\epsilon_{t-j}\right)+o_p(\sqrt{n},k). \end{aligned} \tag{A.16}$$

The last line in the above uses (A.14) and the facts that the elements of $\{\pi_t\}$ decay at an exponential rate while $\{\hat{\pi}_t\}$ converges to $\{\pi_t\}$ at an even quicker rate. It follows that

$$\sum_{t=1}^{k}\mathbf{b}_t\hat{Z}_t-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{Z}_t=\sum_{t=1}^{k}\mathbf{b}_tZ_t-\frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_tZ_t+o_p(\sqrt{n},k).$$

Note that (A.7) and, therefore, (A.11) hold in the event that $\mathbf{M}_k$ is substituted with an analogous version that has the row of $\mathbf{M}_k$ that corresponds to $(\mathbf{x}_t',\mathbf{s}_t',\mathbf{v}_t')$ replaced with

$(\mathbf{x}'_{t+i}, \mathbf{s}'_{t+i}, \mathbf{v}'_{t+i})$ for $t = 1, \ldots, k$ and for $i \geq 0$. Using this observation as well as the fact that $\{n^{-1/2} \sum_{t=1}^{k} \mathbf{b}_t \hat{\epsilon}_{t-i}\}$ and $\{n^{-1/2} \sum_{t=1}^{k} \mathbf{b}_{t+i} \hat{\epsilon}_t\}$ are asymptotically equivalent, we see that $\{\mathbf{s}_t\}$ and $\{\widetilde{\mathbf{s}}_t\}$ obey (A.12), as do $\{\mathbf{v}_t\}$ and $\{\widetilde{\mathbf{v}}_t\}$. Lastly, it holds that these predictor sequences obey (A.13), which yields the lemma's main result. □

Theorem 3. It is first illustrated that the processes $\{\mathbf{R}_{\mathrm{x},k}\}$, $\{\mathbf{R}_{\mathrm{s},k}\}$ and $\{\mathbf{R}_{\mathrm{v},k}\}$ are asymptotically uncorrelated. In light of (23), (A.11), and Lemma 1, it is sufficient to show that the following three processes are asymptotically uncorrelated:

$$\left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \mathbf{x}_t \epsilon_t \right\}, \quad \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \mathbf{s}_t Z_t \right\}, \quad \text{and} \quad \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \mathbf{v}_t Z_t \right\},$$

for $1 \leq k \leq n$, where $\{\epsilon_t\}$ is generated from $\{Z_t\}$ via the causal representation in (A.1). Calculations show that $\sum_{t=1}^{k} \mathbf{x}_t \epsilon_t = \sum_{t=1}^{k} \mathbf{y}_t Z_t$ where $\mathbf{y}_t = \sum_{j=t}^{k} \psi_{j-t} \mathbf{x}_j$.

The proof of Theorem 1 illustrates that the latter two processes of the three processes above are uncorrelated in large samples. Further calculations show that the remaining pairwise covariances of these processes are diven by $n^{-1} \sum_{t=1}^{k} \mathbf{y}_t \mathbf{s}'_t$ and $n^{-1} \sum_{t=1}^{k} \mathbf{y}_t \mathbf{v}'_t$. Furthermore,

$$\sum_{t=1}^{k} \mathbf{y}_t \mathbf{s}'_t = \sum_{j=0}^{k-1} \psi_j \sum_{t=1}^{k} \mathbf{x}_{t+j} \mathbf{s}'_t \quad \text{and} \quad \sum_{t=1}^{k} \mathbf{y}_t \mathbf{v}'_t = \sum_{j=0}^{k-1} \psi_j \sum_{t=1}^{k} \mathbf{x}_{t+j} \mathbf{v}'_t.$$

Using the geometrically decaying structure of $\{\psi_j\}$ in addition to Lemmas A.1 and A.2 of Robbins *et al.* (2016), it holds that $n^{-1} \sum_{t=1}^{k} \mathbf{y}_t \mathbf{s}'_t = \mathcal{O}(n^{-1}, k)$ and $n^{-1} \sum_{t=1}^{k} \mathbf{y}_t \mathbf{v}'_t = \mathcal{O}_p(n^{-1/2}, k)$. This, in combination with (23), Lemma 1 and (25), illustrates the result in the theorem. □

Corollary 1. Let $\mathbf{b}_t$ denote one of $\mathbf{x}_t$, $\mathbf{s}_t$ or $\mathbf{v}_t$ and correspondingly let $\mathbf{\Delta}_b$ denote either $\mathbf{\Delta}_x$,

$\boldsymbol{\Delta}_s$ or $\boldsymbol{\Delta}_v$. In the event that $\boldsymbol{\Delta}_b \neq \mathbf{0}$ and that the changepoint occurs at time c, we see

$$\begin{aligned}\sum_{t=1}^{k} \mathbf{b}_t \hat{\epsilon}_{t-i} &= \left[\sum_{t=1}^{\min\{k,c\}} \mathbf{b}_{t+i}\mathbf{b}_t' - \sum_{t=1}^{k} \mathbf{b}_{t+i}\widetilde{\mathbf{b}}_t \left(\sum_{t=1}^{n} \widetilde{\mathbf{b}}_t\widetilde{\mathbf{b}}_t'\right)^{-1} \sum_{t=1}^{c} \mathbf{b}_t\mathbf{b}_t'\right] \boldsymbol{\Delta}_b \\ &\quad + \sum_{t=1}^{k} \mathbf{b}_t \epsilon_{t-i} - \sum_{t=1}^{k} \mathbf{b}_{t+i}\widetilde{\mathbf{b}}_t \left(\sum_{t=1}^{n} \widetilde{\mathbf{b}}_t\widetilde{\mathbf{b}}_t'\right)^{-1} \sum_{t=1}^{n} \widetilde{\mathbf{b}}_t \epsilon_t + o_p(\sqrt{n}, k), \quad \text{(A.17)}\end{aligned}$$

which expands upon (A.12).

Our focus now turns to $\mathbf{b}_t = \mathbf{x}_t$. Let

$$\mathcal{A}_k = \sum_{i=0}^{p_{\rm ar}} \hat{\phi}_i^* \sum_{t=1}^{k} f\left(\frac{t}{n}\right) \hat{\epsilon}_t - \sum_{j=0}^{q_{\rm ma}} \hat{\theta}_j^* \sum_{t=1}^{k} f\left(\frac{t}{n}\right) \hat{Z}_t,$$

where $f(t/n)$ denotes an arbitrary element of $\mathbf{x}_t$ and where $\hat{\phi}_i^* = -\hat{\phi}_i$ and $\hat{\theta}_j^* = \hat{\theta}_j$ for $i = 1, \ldots, p_{\rm ar}$ and $j = 1, \ldots, q_{\rm ma}$ with $\hat{\phi}_0^* = \hat{\theta}_0^* = 1$. Following the proof of Lemma 2.2 of Robbins *et al.* (2016), it holds that

$$\mathcal{A}_k + \frac{1}{n} \sum_{i=0}^{p_{\rm ar}} i\hat{\phi}_i^* \sum_{t=1}^{k} \dot{f}(\xi_{ti}) \hat{\epsilon}_t - \frac{1}{n} \sum_{j=0}^{q_{\rm ma}} j\hat{\theta}_j^* \sum_{t=1}^{k} \dot{f}(\xi_{tj}) \hat{Z}_t = o_p(n^{1/\nu}, k),$$

for some $\nu \geq 2$ where $\cdot f(z)$ is the first derivative of $f(z)$. Let

$$\mathcal{B}_c(k) = \left[\sum_{t=1}^{\min\{k,c\}} \mathbf{x}_t\mathbf{x}_t' - \sum_{t=1}^{k} \mathbf{x}_t\widetilde{\mathbf{x}}_t \left(\sum_{t=1}^{n} \widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t'\right)^{-1} \sum_{t=1}^{c} \mathbf{x}_t\mathbf{x}_t'\right] \boldsymbol{\Delta}_x,$$

and note that $n^{-1}\mathcal{B}_c(k) = \mathcal{O}_p(1, k)$. Using this and calculations that illustrate (A.17), we can show that $n^{-1}\sum_{t=1}^{k} \dot{f}(\xi_{ti})\hat{\epsilon}_t = \mathcal{O}_p(1, k)$ when $\boldsymbol{\Delta}_x \neq \mathbf{0}$. Similarly, $n^{-1}\sum_{t=1}^{k} \dot{f}(\xi_{tj})\hat{Z}_t = \mathcal{O}_p(1, k)$, which follows from the application of (A.15). Consequentially,

$$\mathcal{A}_k = o_p(n^{1/\nu}, k). \quad \text{(A.18)}$$

Using (A.17), (A.18), and the fact that $n^{-1}\mathcal{B}_c(k) = \mathcal{O}_p(1,k)$, we see that $n^{-1/2}\mathbf{R}_{\mathrm{x},k}$ diverges at rate of $n^{1/2}$ if $\boldsymbol{\Delta}_x \neq \mathbf{0}$, which proves that $\lim_{n\to\infty} P(\widehat{L}_{\mathrm{x},k} > c_\alpha) = 1$ for any constant c_α.

To illustrate consistency of $\arg\max_k \widehat{L}_{\mathrm{x},k}$ as an estimator of the changepoint time, note that $(\mathcal{B}_c(k))'\mathbf{C}_{\mathrm{x},k}^{-1}\mathcal{B}_c(k)$ is maximized when $k = c$ by Lemma A.2 of Bai (1997). This, (A.18), and Lemma A.4 of Bai (1997) show that $\arg\max_k \widehat{L}_{\mathrm{x},k} \xrightarrow{\mathcal{P}} \kappa$, where $c/n \to \kappa$.

Next, we focus on the case where $\mathbf{b}_t = \mathbf{s}_t$ or $\mathbf{b}_t = \mathbf{v}_t$. Formula (A.17) and calculations akin to those which provide (A.14) imply

$$\begin{aligned}\sum_{t=1}^{k}\mathbf{b}_t\hat{\epsilon}_{t-i} - \frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{\epsilon}_{t-i} \;&=\; n\min\left\{\frac{k}{n},\frac{c}{n}\right\}\left(1-\max\left\{\frac{k}{n},\frac{c}{n}\right\}\right)\boldsymbol{\Gamma}_b(i)\boldsymbol{\Delta}_b \\ &\quad \sum_{t=1}^{k}\mathbf{b}_t\epsilon_{t-i} - \frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\epsilon_{t-i} + o_p(\sqrt{n},k).\end{aligned}$$

So long as $\{\hat{\pi}_j\}$ are reasonable approximations under the alternative hypothesis, we mimic (A.16) to yield

$$\begin{aligned}\sum_{t=1}^{k}\mathbf{b}_t\hat{Z}_t - \frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t\hat{Z}_t \;&=\; n\min\left\{\frac{k}{n},\frac{c}{n}\right\}\left(1-\max\left\{\frac{k}{n},\frac{c}{n}\right\}\right)\sum_{j=0}^{\infty}\pi_j\boldsymbol{\Gamma}_b(j)\boldsymbol{\Delta}_b \\ &\quad + \sum_{t=1}^{k}\mathbf{b}_t Z_t - \frac{k}{n}\sum_{t=1}^{n}\mathbf{b}_t Z_t + o_p(\sqrt{n},k).\end{aligned}$$

Therefore, $n^{-1/2}\mathbf{R}^*_{\mathrm{s},k}$ diverges at rate of $n^{1/2}$ if $\boldsymbol{\Delta}_s \neq \mathbf{0}$. Note that $\min\{k,c\}(n-\max\{k,c\})$ is maximized when $k = c$. Furthermore, from Lemma A.4 of Bai (1997), it holds that $\arg\max_k \widehat{L}_{\mathrm{s},k} \xrightarrow{\mathcal{P}} \kappa$, where $c/n \to \kappa$. Analogous results hold for $\mathbf{R}^*_{\mathrm{v},k}$ and $\arg\max_k \widehat{L}_{\mathrm{v},k}$.

□

Corollary A.1. We first assume that the seasonal terms obey the harmonic representation in (A.3). Basic trigonometric identities can be used to establish that $\mathbf{D}_T = \mathbf{I}_{q_s}/2$. This result in combination with the observation that $\sum_{t=1}^{n} \mathbf{s}_t \hat{Z}_t = \mathcal{O}_p(1)$ establishes the finding of Corollary A.1. To illustrate the latter formula, we establish that $\sum_{t=1}^{n} \mathbf{s}_t \hat{\epsilon}_{t-i} = \mathcal{O}_p(1)$ for all $i \geq 0$; the invertibility expansions used in (A.16) can then be applied in order to show that $\sum_{t=1}^{n} \mathbf{s}_t \hat{Z}_t = \mathcal{O}_p(1)$.

Define

$$\mathbf{H}_{j,i} = \begin{pmatrix} \cos(2\pi ji/T) & -\sin(2\pi ji/T) \\ \sin(2\pi ji/T) & \cos(2\pi ji/T) \end{pmatrix},$$

and let $\mathbf{H}_i$ denote a block-diagonal matrix of dimension $p_s \times p_s$ written as

$$\mathbf{H}_i = \begin{pmatrix} \mathbf{H}_{j_1,i} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{j_2,i} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{j_J,i} \end{pmatrix}.$$

The matrix $\mathbf{H}_i$ can be used to establish a recursion between the $\mathbf{s}_t$. Specifically,

$$\mathbf{H}_i \mathbf{s}_t = \mathbf{s}_{t+i}. \tag{A.19}$$

Further, note that $\mathbf{H}_i = 2\sum_{t=1}^{T} \mathbf{s}_{t+i}\mathbf{s}_t'/T$. Define

$$\begin{aligned} \mathbf{N}_{s,n}(i) := \sum_{t=1}^{n} \mathbf{s}_{t+i}\hat{\epsilon}_t &= \mathbf{S}_{n,i}'\boldsymbol{\epsilon} - \mathbf{S}_{n,i}'\widetilde{\mathbf{M}}_n(\widetilde{\mathbf{M}}_n'\widetilde{\mathbf{M}}_n)^{-1}\widetilde{\mathbf{M}}_n\boldsymbol{\epsilon} \\ &= \mathbf{S}_{n,i}'\boldsymbol{\epsilon} - \mathbf{S}_{n,i}'\widetilde{\mathbf{S}}_n(\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n)^{-1}\widetilde{\mathbf{S}}_n\boldsymbol{\epsilon} + \mathcal{O}_p(1), \end{aligned}$$

where $\mathbf{S}_{n,i} = (\mathbf{s}_{1+i}, \ldots, \mathbf{s}_{n+i})'$, for $t = 1, \ldots, n$.

It follows that

$$\widetilde{\mathbf{S}}_n'\widetilde{\mathbf{S}}_n/n = \mathbf{I}_{p_s}/2 + \mathcal{O}(n^{-1}), \quad \text{and} \quad \mathbf{S}_{n,i}'\widetilde{\mathbf{S}}_n/n = \begin{pmatrix} \mathbf{H}_i/2 & \mathbf{0} \end{pmatrix} + \mathcal{O}(n^{-1}).$$

Continuing,

$$\begin{aligned} \mathbf{N}_{\mathrm{s},n}(i) &= \sum_{t=1}^{n} \mathbf{s}_{t+i}\epsilon_t - \mathbf{H}_i \sum_{t=1}^{n} \mathbf{s}_t\epsilon_t + \mathcal{O}_p(1) \\ &= \mathbf{0} + \mathcal{O}_p(1), \end{aligned} \tag{A.20}$$

which is derived using the recursion in (A.19). The above also implies that $\sum_{t=1}^{n} \mathbf{s}_t\hat{\epsilon}_{t-i} = \mathcal{O}_p(1)$.

Next, assume that $\mathbf{s}_t$ satisfies the formulation in (A.4). Then, it holds that

$$\mathbf{S}_{n,i}'\mathbf{S}_n(\mathbf{S}_n'\mathbf{S}_n)^{-1} = \mathbf{H}_i^* + \mathcal{O}(n^{-1}),$$

where $\mathbf{H}_i^*$ is defined as follows. First,

$$\mathbf{H}_1^* = \begin{pmatrix} -\mathbf{1}' & -1 \\ \mathbf{I}_{T-2} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{1}$ is a vector of ones. Next, $\mathbf{H}_2^*$ is defined by permuting the rows of $\mathbf{H}_1$ so that the bottom row of $\mathbf{H}_2^*$ equals the top row of $\mathbf{H}_1^*$ and the remaining rows of $\mathbf{H}_1^*$ are each shifted down. In general, define $\mathbf{H}_i^*$ by permuting the rows of $\mathbf{H}_{i-1}^*$ in a similar manner. A recursion analogous to (A.19) can be established: $\mathbf{H}_i^*\mathbf{s}_t = \mathbf{s}_{t+i}$. Therefore, (A.20) is again satisfied, and the proof is completed. □

Appendix G: Parameter Estimates for Data Examples

OLS parameter estimates under both the null and alternative are provided in Table A.1 for the Mauna Loa data example and in Table A.2 for the Barrow, AK data example. (Notation in the tables is in line with the notation provided in the article. That is, α's govern trend, β's govern seasonality, γ's govern covariates, and Δ's quantify changes. For example, $\hat{\Delta}_{s,1,1}$ is the post-changepoint change in the parameter $\hat{\beta}_{1,1}$.)

Supplement References

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics* **79**, 551–563.

Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011). Mean shift testing in correlated data. *Journal of Time Series Analysis* **32**, 498–511.

Robbins, M. W., Gallagher, C. M. and Lund, R. B. (2016). A general regression changepoint test for time series data. *Journal of the American Statistical Association* **111**, 670–683.

Table A.1: Parameter estimates for the Mauna Loa data analysis

	Coef.	Estimate
H0	$\hat{\alpha}_1$	315.086
	$\hat{\alpha}_2$	34.833
	$\hat{\alpha}_3$	62.580
	$\hat{\alpha}_4$	1.763
	$\hat{\alpha}_5$	1.664
	$\hat{\alpha}_6$	-15.103
	$\hat{\beta}_{1,1}$	2.552
	$\hat{\beta}_{2,1}$	-0.653
	$\hat{\beta}_{1,2}$	0.022
	$\hat{\beta}_{2,2}$	-0.057
	$\hat{\beta}_{1,3}$	1.120
	$\hat{\beta}_{2,3}$	-0.413
	$\hat{\beta}_{1,4}$	-0.086
	$\hat{\beta}_{2,4}$	0.044
	$\hat{\gamma}_1$	-0.015
H1c*	$\hat{\alpha}_1$	315.088
	$\hat{\alpha}_2$	34.819
	$\hat{\alpha}_3$	62.586
	$\hat{\alpha}_4$	1.793
	$\hat{\alpha}_5$	1.628
	$\hat{\alpha}_6$	-15.098
	$\hat{\beta}_{1,1}$	2.424
	$\hat{\beta}_{2,1}$	-0.614
	$\hat{\beta}_{1,2}$	0.018
	$\hat{\beta}_{2,2}$	-0.049
	$\hat{\beta}_{1,3}$	0.890
	$\hat{\beta}_{2,3}$	-0.335
	$\hat{\beta}_{1,4}$	-0.096
	$\hat{\beta}_{2,4}$	0.086
	$\hat{\gamma}_1$	-0.014
	$\widehat{\Delta}_{s,1,1}$	0.186
	$\widehat{\Delta}_{s,2,1}$	-0.058
	$\widehat{\Delta}_{s,1,2}$	0.006
	$\widehat{\Delta}_{s,2,2}$	-0.012
	$\widehat{\Delta}_{s,1,3}$	0.336
	$\widehat{\Delta}_{s,2,3}$	-0.113
	$\widehat{\Delta}_{s,1,4}$	0.014
	$\widehat{\Delta}_{s,2,4}$	-0.061

Table A.2: Parameter estimates for the Barrow, AK data analysis

	Coef.	Estimate
H0	$\hat{\alpha}_1$	-13.003
	$\hat{\alpha}_2$	2.072
	$\hat{\beta}_1$	-8.151
	$\hat{\beta}_2$	-17.113
	$\hat{\beta}_3$	-22.994
	$\hat{\beta}_4$	-25.149
	$\hat{\beta}_5$	-26.663
	$\hat{\beta}_6$	-25.111
	$\hat{\beta}_7$	-17.225
	$\hat{\beta}_8$	-6.285
	$\hat{\beta}_9$	1.796
	$\hat{\beta}_{10}$	4.885
	$\hat{\beta}_{11}$	4.095
	$\hat{\gamma}_1$	-0.017
H1b	$\hat{\alpha}_1$	-12.263
	$\hat{\alpha}_2$	-0.169
	$\hat{\beta}_1$	-7.731
	$\hat{\beta}_2$	-16.982
	$\hat{\beta}_3$	-23.250
	$\hat{\beta}_4$	-25.322
	$\hat{\beta}_5$	-27.109
	$\hat{\beta}_6$	-25.276
	$\hat{\beta}_7$	-17.045
	$\hat{\beta}_8$	-6.445
	$\hat{\beta}_9$	1.806
	$\hat{\beta}_{10}$	5.030
	$\hat{\beta}_{11}$	4.340
	$\hat{\gamma}_1$	-0.017
	$\widehat{\Delta}_{x,1}$	-4.616
	$\widehat{\Delta}_{x,2}$	7.312
	$\widehat{\Delta}_{s,1}$	-0.787
	$\widehat{\Delta}_{s,2}$	-0.259
	$\widehat{\Delta}_{s,3}$	0.442
	$\widehat{\Delta}_{s,4}$	0.302
	$\widehat{\Delta}_{s,5}$	0.788
	$\widehat{\Delta}_{s,6}$	0.282
	$\widehat{\Delta}_{s,7}$	-0.354
	$\widehat{\Delta}_{s,8}$	0.266
	$\widehat{\Delta}_{s,9}$	-0.050
	$\widehat{\Delta}_{s,10}$	-0.256
	$\widehat{\Delta}_{s,11}$	-0.443
	$\widehat{\Delta}_{v,1}$	-0.007