Statistica Sinica **30** (2020), 1657-1683
doi:https://doi.org/10.5705/ss.202018.0275

A FULLY FLEXIBLE CHANGEPOINT TEST FOR REGRESSION MODELS WITH STATIONARY ERRORS

Michael W. Robbins

RAND Corporation

Abstract: Temporal discontinuities in time series represent one of the classic problems of time series. Such discontinuities are often analyzed by detecting changes at specific times in the parameters governing a regression model 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 with full flexibility in selecting which set of regression parameters are allowed to change under the alternative hypothesis. Here, we also examine several mathematical complications that arise in the development of such procedures. A simulation study illustrates the efficacy of the proposed methodology, where a test statistic based on the residuals from an ARMA model is shown to perform most favorably. The methods are applied to a 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, Lund and Robbins (2013)). A related problem in changepoint diagnostics is how to incorporate autocorrelation (e.g., see Bai (1993); Antoch, Hušková and Prášková (1997); Yu (2007); Robbins et al. (2011a), etc.). 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 in which the outcome series is allowed to depend on three classes of predictors: 1) the terms governing the trend (where the trend may be, for example, constant, linear, quadratic, etc.); 2) seasonal terms (i.e., that control periodic oscillations from the trend function); and 3) stochastic covariates, which we use to explain the error in the outcome.

In related works, Gombay (2010) and Aue, Horváth and Hušková (2012) present diagnostic methods for changepoints in regression models, while enabling autocorrelated regression errors. Specifically, Aue, Horváth and Hušková (2012) examine a regression structure with predictor terms that have a general functional form. However, they model autocorrelation using a Bartlett-based variance expression and approximate the large-sample distribution of their test statistics using extreme value expressions. As a result, their asymptotic approximations do not perform well on finite samples, where they employ bootstrapping techniques to address this issue.

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

Detecting 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. Furthermore, researchers have found evidence that the warming of surface temperatures in polar climates is greater in winter seasons than it is in summer seasons (Lu and Cai (2009); Screen and Simmonds (2010)). Lastly, Elsner, Bossak and Niu (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, Horváth and Hušková (2012)) mandate that all regression coefficients change simultaneously. Thus, we expand upon existing tests for changes in a trend by developing separate tests for changes in the seasonal structure and outcome/covariate relationships. 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 preselected 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 show 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 any discontinuity in the trend. Disentangling 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 the mathematical context, including a discussion of technical details of relevant extant methodologies. Section 3 outlines the foundations for our general method by addressing the setting of independent and identically distributed (i.i.d.) regression errors. In Section 4, procedures that encapsulate the autocorrelation in regression errors are developed. In Section 5, we examine the finite-sample performance of the proposed methods using simulations. Section 6 presents an application of the developed techniques to two data sets: 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 a response sequence, where n is the sample size. The null hypothesis model for the tth time point is

$$Y_t = \widetilde{\boldsymbol{\alpha}}'\widetilde{\mathbf{x}}_t + \widetilde{\boldsymbol{\beta}}'\widetilde{\mathbf{s}}_t + \widetilde{\boldsymbol{\gamma}}'\widetilde{\mathbf{v}}_t + \epsilon_t, \tag{2.1}$$

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. In addition, $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ 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 a polynomial trend, for example, one can safely define predictors using t^j instead of $(t/n)^j$ (Aue, Horváth and Hušková (2012); Robbins, Gallagher and Lund (2016)). Furthermore, $\{\widetilde{\mathbf{s}}_t\}$ contains periodic deterministic design points with known period T, such 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 $q_x < p_x$, are allowed to shift at the changepoint time; q_s and q_v are defined similarly.

Define $K \subset (0, 1)$, such 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 $K = \{z : \ell \leq z \leq h\}$, where $0 < \ell < h < 1$. Restricting this set of admissible changepoints away from the boundaries of $[0, 1]$ ensures that the test statistics defined in the forthcoming discourse are asymptotically finite.

The formal assumptions imposed here are outlined in Appendix A of the Supplementary Material. Note that in (A.1) in Assumption 4, wherein the regression errors $\{\epsilon_t\}$ are written as 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 attributable to autocorrelation can often be encapsulated by monitoring the quantity

$$\tau^2 := \lim_{n \to \infty} \frac{1}{n} \mathrm{Var}\left(\sum_{t=1}^{n} \epsilon_t \right). \tag{2.2}$$

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

We consider an alternative hypothesis that allows preselected subsets of the regression coefficients $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$, and $\widetilde{\boldsymbol{\gamma}}$ to shift at a single unknown time c, for $1 \leq$

$c < n$. To enable complete flexibility in the choice of which parameters change, $\widetilde{\mathbf{x}}_t$ is decomposed into two sub-vectors: $\mathbf{x}_t$, 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

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

where the vectors of the 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,

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

which are vectors of dimension q_x, q_s, and q_v, respectively. The double subscript (e.g., x, t) in the above definitions is a notational structure used frequently throughout this paper. The first subscript (e.g., x) simply refers to a set of predictor variables (from the functional form, seasonal, or stochastic predictor sets), and the second (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

$$\mathcal{H}_0 : \boldsymbol{\Delta} = \mathbf{0} \quad \text{against} \quad \mathcal{H}_1 : \boldsymbol{\Delta} \neq \mathbf{0}. \tag{2.3}$$

Because 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 the 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

$$\mathbf{X}_n(k) = o_p(a_n, k) \quad \text{when} \quad \max_{k/n \in K} ||\mathbf{X}_n(k)||_\infty = o_p(a_n), \tag{2.4}$$

and

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

as $n \to \infty$. Similar notation for deterministic processes is defined using $o(a_n, k)$

and $\mathcal{O}(a_n, k)$.

3. Fully Flexible Changepoint Tests

In this section, tests for the hypotheses in (2.3) are developed, assuming i.i.d regression errors. Later, we extend the resultant procedure to settings involving general stationary errors later. In line with earlier work (e.g., Robbins, Gallagher and Lund (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 $F_k := \widehat{\boldsymbol{\Delta}}_k' \text{Var}(\widehat{\boldsymbol{\Delta}}_k)^{-1} \widehat{\boldsymbol{\Delta}}_k$. 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 $\widetilde{\boldsymbol{\gamma}}$, respectively, calculated under $\mathcal{H}_0$, the OLS residuals are defined as

$$\hat{\epsilon}_t = Y_t - (\widehat{\boldsymbol{\alpha}}'\widetilde{\mathbf{x}}_t + \widehat{\boldsymbol{\beta}}'\widetilde{\mathbf{s}}_t + \widehat{\boldsymbol{\gamma}}'\widetilde{\mathbf{v}}_t). \tag{3.1}$$

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 $F_k = \mathbf{N}_k' \text{Var}(\mathbf{N}_k)^{-1} \mathbf{N}_k$. In practice, we use

$$\widehat{F}_k = \frac{\mathbf{N}_k' \mathbf{C}_k^{-1} \mathbf{N}_k}{\hat{\tau}^2} \tag{3.2}$$

to test for a changepoint at time k, where $\mathbf{C}_k$ is a matrix satisfying $\text{Var}(\mathbf{N}_k) = \tau^2 \mathbf{C}_k$, 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) in Appendix F of the Supplementary Material. A 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

$$\widehat{F} = \max_{\frac{k}{n} \in K} \widehat{F}_k. \tag{3.3}$$

For this statistic, and others defined later, the estimated changepoint time $\hat{c}$ is the argument k that maximizes $\widehat{F}_k$ (although we focus on a statistical confirmation of the presence of a changepoint, as opposed to the properties of its estimator). The large-sample behavior of $\widehat{F}$ is stated formally 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:

$$\mathbf{N}_{\mathrm{x},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{x}_t, \quad \mathbf{N}_{\mathrm{s},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{s}_t, \quad \text{and} \quad \mathbf{N}_{\mathrm{v},k} = \sum_{t=1}^{k} \hat{\epsilon}_t \mathbf{v}_t, \tag{3.4}$$

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

The process $\{\widehat{F}_{\mathrm{x},k}\}$ was used by Robbins, Gallagher and Lund (2016) to detect a shift in the trend function only. We briefly repeat the results here. Letting $\Rightarrow$ denote weak convergence in $D[0,1]$, the space of right-continuous functions with left-hand limits,

$$\widehat{F}_{\mathrm{x},\lfloor nz \rfloor} \Rightarrow \widetilde{B}_1(z), \quad \text{for } z \in K, \text{ where} \quad \widetilde{B}_1(z) = \mathbf{\Lambda}(z)'\mathbf{\Omega}(z)^{-1}\mathbf{\Lambda}(z). \tag{3.5}$$

See Appendix B of the Supplementary Material for details on the convergence and terms in (3.5).

The processes $\{\widehat{F}_{\mathrm{s},k}\}$ and $\{\widehat{F}_{\mathrm{v},k}\}$ have not been considered previously and behave as the square of scaled multidimensional Brownian bridges for large sample sizes. Although a 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

$$(n\tau^2)^{-1}\mathrm{Var}(\mathbf{N}_{\mathrm{s},\lfloor nz \rfloor}) \Rightarrow z(1-z)\mathbf{D}_T, \tag{3.6}$$

and

$$(n\tau^2)^{-1}\mathrm{Var}(\mathbf{N}_{\mathrm{v},\lfloor nz \rfloor}) \Rightarrow z(1-z)\mathbf{\Sigma}_v, \tag{3.7}$$

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

$$\mathbf{N}_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t \epsilon_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{s}_t \epsilon_t + o_p(\sqrt{n}, k), \tag{3.8}$$

and

$$\mathbf{N}_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t \epsilon_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{v}_t \epsilon_t + o_p(\sqrt{n}, k), \tag{3.9}$$

where the order-of-probability notation introduced in (2.4) is employed. Letting $\{\mathbf{W}_d(z)\}_{z \in [0,1]}$ denote a d-dimensional Wiener process, (3.6)–(3.9) yield that

$$(n\tau^2 \mathbf{D}_T)^{-1/2} \mathbf{N}_{\mathrm{s},\lfloor nz \rfloor} \Rightarrow \mathbf{W}_{q_s}(z) - z\mathbf{W}_{q_s}(1), \tag{3.10}$$

and

$$(n\tau^2 \boldsymbol{\Sigma}_v)^{-1/2} \mathbf{N}_{\mathrm{v},\lfloor nz \rfloor} \Rightarrow \mathbf{W}_{q_v}(z) - z\mathbf{W}_{q_v}(1). \tag{3.11}$$

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* (2.1)*, and that Assumptions* 1–3 *in Appendix A of the Supplementary Material hold. Furthermore, assume that the errors* $\{\epsilon_t\}$ *are i.i.d. with variance* τ^2*. If* $\widehat{F}$ *is calculated from* (3.3)*,*

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

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

$$\widetilde{B}(z) = \widetilde{B}_1(z) + \widetilde{B}_2(z). \tag{3.12}$$

The stochastic process $\{\widetilde{B}_1(z)\}$*, which is defined in* (3.5)*, is independent of* $\{\widetilde{B}_2(z)\}$*, where*

$$\widetilde{B}_2(z) = \frac{\mathbf{B}_d(z)' \mathbf{B}_d(z)}{z(1-z)}, \tag{3.13}$$

for $z \in K$*, with* $d = q_s + q_v$*. In addition,* $\{\mathbf{B}_d(z) := \mathbf{W}_d(z) - z\mathbf{W}_d(1)\}$ *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 does depend 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 and Horváth (1997) and Robbins et al. (2011b) for closed-form approximations of the supremum of this process).

Note 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 Supplementary Material 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 extend 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 the power to detect discontinuities in the presence of autocorrelated errors. Thus, we need to identify manners of adjustment so that type-I error rates can be controlled.

Robbins, Gallagher and Lund (2016) argue that $\{\widehat{F}_{\mathrm{x},k}\}$ has the same limit process in circumstances where regression errors contain autocorrelation. This 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 (2.2). Furthermore, their calculations incorporate the following Bartlett-based estimator of τ^2 in the event that regression errors contain a stationary autocorrelation structure:

$$\hat{\tau}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2 + 2\sum_{j=1}^{q_n}\left(1 - \frac{j}{q_n+1}\right)\frac{1}{n-j}\sum_{t=1}^{n-j}\hat{\epsilon}_t\hat{\epsilon}_{t+j}, \tag{4.1}$$

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 Supplementary Material, 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 $\widetilde{B}_1(z)$ is defined in (3.5).

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, Gallagher and Lund (2016) for general $\mathbf{s}_t$). An 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 by incorporating 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

$$\widehat{\boldsymbol{\tau}}_s = \frac{1}{Tm} \sum_{i=1}^{m} \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i' + \sum_{j=1}^{q_m} \left(1 - \frac{j}{q_m + 1}\right) \frac{1}{T(m-j)} \sum_{i=1}^{m-j} \left(\hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_{i+j}' + \hat{\boldsymbol{e}}_{i+j} \hat{\boldsymbol{e}}_i'\right). \quad (4.2)$$

In light of (3.8), 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

$$\widehat{\boldsymbol{\tau}}_v = \frac{1}{n} \sum_{t=1}^{n} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' + \sum_{j=1}^{q_n} \left(1 - \frac{j}{q_n + 1}\right) \frac{1}{n-j} \sum_{t=1}^{n-j} \left(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t+j}' + \hat{\boldsymbol{\varepsilon}}_{t+j} \hat{\boldsymbol{\varepsilon}}_t',\right), \quad (4.3)$$

where $\hat{\boldsymbol{\varepsilon}}_t = \hat{\epsilon}_t \widetilde{\mathbf{v}}_t$. Similarly to (3.9), 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

$$\widehat{F}_k^* = \hat{\tau}^{-2} \mathbf{N}_{\mathrm{x},k}' \mathbf{C}_{\mathrm{x},k}^{-1} \mathbf{N}_{\mathrm{x},k} + \mathbf{N}_{\mathrm{s},k}' \widehat{\boldsymbol{\tau}}_s^{-1} \mathbf{N}_{\mathrm{s},k} + \mathbf{N}_{\mathrm{v},k}' \widehat{\boldsymbol{\tau}}_v^{-1} \mathbf{N}_{\mathrm{v},k}, \quad (4.4)$$

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. *Furthermore, assume that* $\widehat{F}_k^*$ *is in accordance with* (4.4). *Then it holds that*

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

as $n \to \infty$, *where* $\widetilde{B}(z)$ *is defined in* (3.12).

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, we can estimate the terms using a spectral density (Andrews and Monahan (1992)) or data-dependent bandwidths (Newey and 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 (4.4) is that it does not impose a parametric model on 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); Robbins, Gallagher and Lund (2016)), is to construct changepoint statistics using residuals from an autoregressive moving average (ARMA) model, instead of using 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 i.i.d. innovations sequence $\{Z_t\}$ generates the error sequence $\{\epsilon_t\}$ via the ARMA formulation:

$$\epsilon_t - \phi_1\epsilon_{t-1} - \cdots - \phi_{p_{\rm ar}}\epsilon_{t-p_{\rm ar}} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_{q_{\rm ma}} Z_{t-q_{\rm ma}}, \qquad t \in \mathbb{Z}, \quad (4.5)$$

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

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

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

$$\frac{\mathbf{R}_{{\rm x},k}}{\hat{\sigma}\sqrt{n}} - \frac{\mathbf{N}_{{\rm x},k}}{\hat{\tau}\sqrt{n}} = o_p(1,k), \quad (4.6)$$

where $\hat{\sigma}^2 = \sum_{t=1}^n \hat{Z}_t^2/n$ estimates the white noise variance, $\hat{\tau}^2$ is defined in (4.1), $\mathbf{N}_{{\rm x},k}$ is calculated in the same manner as in (3.4), and the $o_p(a_n,k)$ notation is as defined in (2.4). Then, they define an ARMA residuals-based analogue to $\widehat{F}_{{\rm x},k}$ via

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

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

Deriving statistics based on ARMA residuals for seasonal and covariate components requires different approaches. Note that the result in (4.6) is based on 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 (4.6) 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. That is,

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

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

$$\mathbf{R}^*_{\mathrm{s},k} = \sum_{t=1}^{k} \mathbf{s}_t \hat{Z}_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{s}_t \hat{Z}_t, \quad \text{and} \quad \mathbf{R}^*_{\mathrm{v},k} = \sum_{t=1}^{k} \mathbf{v}_t \hat{Z}_t - \frac{k}{n} \sum_{t=1}^{n} \mathbf{v}_t \hat{Z}_t. \tag{4.7}$$

The limit behavior of these quantities (under both of $\mathcal{H}_0$ and $\mathcal{H}_1$) is established in the following lemma; see Appendix F of the Supplementary Material for the proof.

Lemma 1. *Assume that the conditions for Theorem* 1 *hold, with the exception that* $\{\epsilon_t\}$ *obeys the ARMA formulation in* (4.5). *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 i.i.d regression errors considered earlier. Specifically, the fact that (3.8) yields (3.10) implies

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

Similarly,

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

Therefore, defining statistics for a change at time k via

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

it follows that

$$\widehat{L}^*_{\mathrm{s},\lfloor nz \rfloor} \Rightarrow \frac{\mathbf{B}_{q_s}(z)' \mathbf{B}_{q_s}(z)}{z(1-z)} \quad \text{and} \quad \widehat{L}^*_{\mathrm{v},\lfloor nz \rfloor} \Rightarrow \frac{\mathbf{B}_{q_v}(z)' \mathbf{B}_{q_v}(z)}{z(1-z)}, \tag{4.8}$$

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

$$\widehat{L}_k = \widehat{L}_{\mathrm{x},k} + \widehat{L}^*_{\mathrm{s},k} + \widehat{L}^*_{\mathrm{v},k}. \tag{4.9}$$

The limit distribution of the maximally selected version of the above test is stated as follows.

Theorem 3. *Assume that the conditions of Lemma* 1 *hold and that* $\{\widehat{L}_k\}$ *is calculated in accordance with* (4.9)*. 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* (3.12)*.*

A proof of Theorem 3 is given in the Supplementary Material (Appendix F), showing 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 (4.6), Lemma 1, and (4.8). Arguments related to 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 sufficient power to detect changepoints. Furthermore, existing theory (e.g., Bai (1997)) establishes the 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 Appendix F of the Supplementary Material, to demonstrate that $\widehat{L}_{\mathrm{x},k}$ has asymptotic power of one in the event that $\boldsymbol{\Delta}_x \neq \mathbf{0}$ (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 the 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*

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

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

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

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 Supplementary Material provides 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. 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 many ways 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 $n = 1,000$. 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 from $v_{j,t} = \zeta_{1,j} + \zeta_{2,j}(t/n) + u_{j,t}$, 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 $\mathrm{Cor}(W_{1,t}, W_{2,t}) = 0.3$ and $\mathrm{Cor}(W_{j,t}, Z_t) = 0$ 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 $\zeta_{1,1} = \zeta_{1,2} = 0$, $\zeta_{2,1} = 1$, 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 $\alpha_1 = \alpha_2 = 1 + \delta_x$. Similarly, 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 examine the empirical performance of the $\widehat{F}^*$ test of Theorem 2 and the $\widehat{L}$ test of Theorem 3. Specifically, we estimate the size and power of 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, Gallagher and Lund (2016) and has been studied in 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 (4.1) and (4.3), and thus $q_m = \lfloor (n/12)^{1/3} \rfloor$ when calculating the term in (4.2). 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 works (e.g., Robbins et al. (2011a,b); Gallagher, Lund and Robbins (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 earlier, 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 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; thus, 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. The 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 estimating the quantity in (4.2)—this quality is estimated using $\lfloor 1,000/12 \rfloor = 83$ 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 sufficient 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 Supplementary Material (Appendix E).

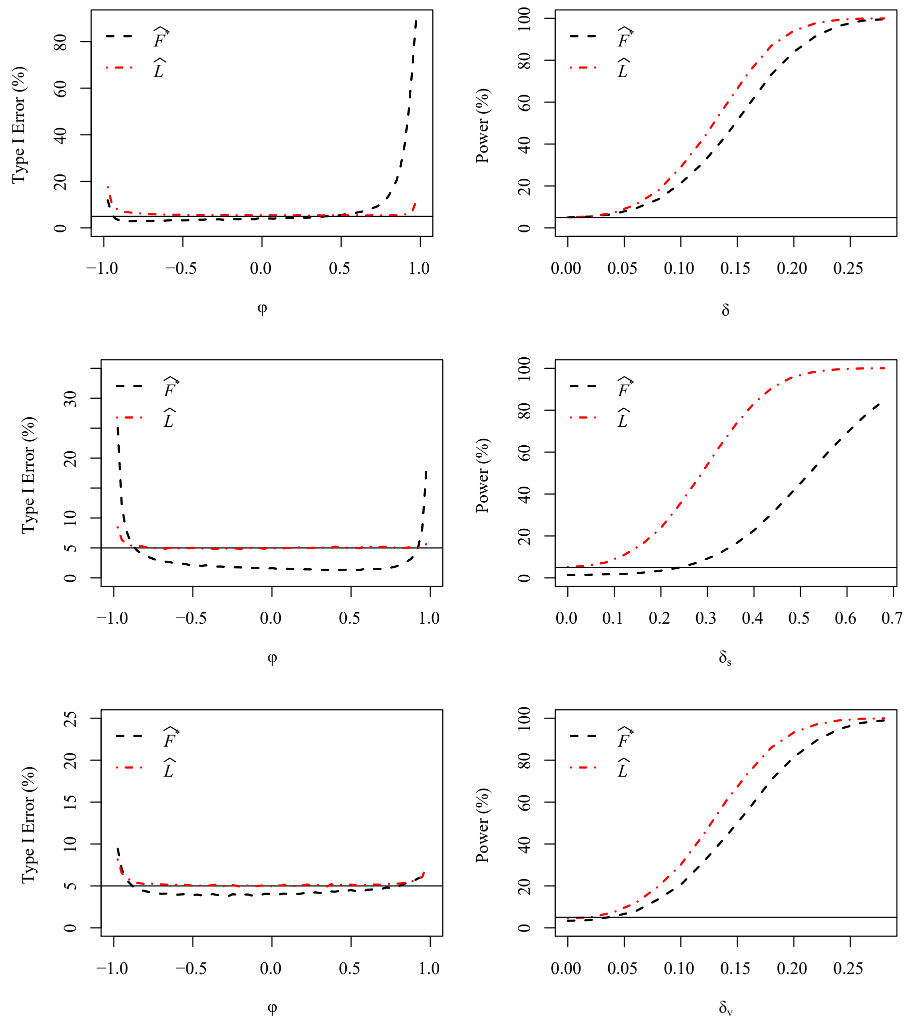

Figure 1. Simulated type-I error (left column) and power (right column) rates for tests based on the $\widehat{F}^*$ and $\widehat{L}$ statistics at a significance level of 0.05 with $n = 1,000$. Results are shown for tests of three alternative hypotheses: H1a (top row); H1c (middle row); H1d (bottom 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 data sets for each value of ϕ, whereas power is based on 25,000 data sets for each value of $\delta/\delta_s/\delta_v$.

6. Data Applications

6.1. Mauna Loa CO_2

To examine the performance of the proposed methods using 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), which many authors have examined in search of structural breaks (e.g., Lund and Reeves (2002); Beaulieu, Chen and Sarmiento (2012); Robbins, Gallagher and Lund (2016)). Such 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 examined closely within the climate literature. Many authors have claimed that the amplitude of the series has increased over time (e.g., Bacastow, Keeling and Whorf (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 on the causes of the changes in the seasonal CO_2 fluctuations, nor have such changes been verified using rigorous statistical tools. This example illustrates the prudence of methods that examine changes in trend and seasonal structures separately, because 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 the CO_2 level at month t, and assume

$$\begin{aligned} Y_t =& \alpha_1 + \alpha_2 \left(\frac{t}{n}\right) + \alpha_3 \left(\frac{t}{n}\right)^2 \\ &+ \sum_{j=1}^{4} \left[\beta_{1,j} \cos\left(\frac{2\pi jt}{12}\right) + \beta_{2,j} \sin\left(\frac{2\pi jt}{12}\right)\right] + \gamma_1 \text{ENSO}_{t-12} + \epsilon_t, \end{aligned} \tag{6.1}$$

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 $\widehat{F}_k^* = \widehat{F}_{\mathrm{x},k}^*$ and $\widehat{L}_k = \widehat{L}_{\mathrm{x},k}$ 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 (6.1) is considered:

$$Y_t = \alpha_1 + \alpha_2\left(\frac{t}{n}\right) + \alpha_3\left(\frac{t}{n}\right)^2 + \alpha_4\mathbf{1}_{\{t>c^*\}} + \alpha_5\left(\frac{t}{n}\right)\mathbf{1}_{\{t>c^*\}} + \alpha_6\left(\frac{t}{n}\right)^2\mathbf{1}_{\{t>c^*\}} + \sum_{j=1}^{4}\left[\beta_{1,j}\cos\left(\frac{2\pi jt}{12}\right) + \beta_{2,j}\sin\left(\frac{2\pi jt}{12}\right)\right] + \gamma_1\mathrm{ENSO}_{t-12}, \tag{6.2}$$

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 the test of the hypothesis in H1b. Note that in order to satisfy Assumption 2 in Appendix A of the Supplementary Material, changes in α_4, α_5, or α_6 cannot be allowed. We test (6.2) against the 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 $\widehat{F}_k^* = \widehat{F}_{\mathrm{s},k}^*$ and $\widehat{L}_k = \widehat{L}_{\mathrm{s},k}$ for H1c*, and $\widehat{F}_k^* = \widehat{F}_{\mathrm{v},k}^*$ and $\widehat{L}_k = \widehat{L}_{\mathrm{v},k}$ 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 the condition that an admissible changepoint time k satisfies $0.05n \leq k \leq 0.95n$.

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 (4.1) and (4.3), and we set $q_m = \rho/12^{1/3}$ (rounded to the nearest integer) when calculating the term in (4.2); similarly, for the $\widehat{L}$ test, we set $p_{\rm ar} = \rho$ and $q_{\rm ma} = 0$ (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

Table 1 shows the results for the 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. The table also shows the results for various choices of the bandwidth parameter ρ (see the description in the caption to the table) that governs the selection of terms such as q_n, q_m, $p_{\rm ar}$, and $q_{\rm ma}$. We prefer to use $\rho = 12$ when applying the $\widehat{L}$ statistic (as this value of $p_{\rm ar}$ ensures that there the ARMA residuals that 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 the bandwidth parameter than is the value of $\widehat{L}$; thus, and in accordance with the results of Section 5, the $\widehat{L}$ test is our preferred method. The shift in the 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 a discussion of the trend behavior in this data set).

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 the trend that is often attributed to Mount Pinatubo. The test of 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).

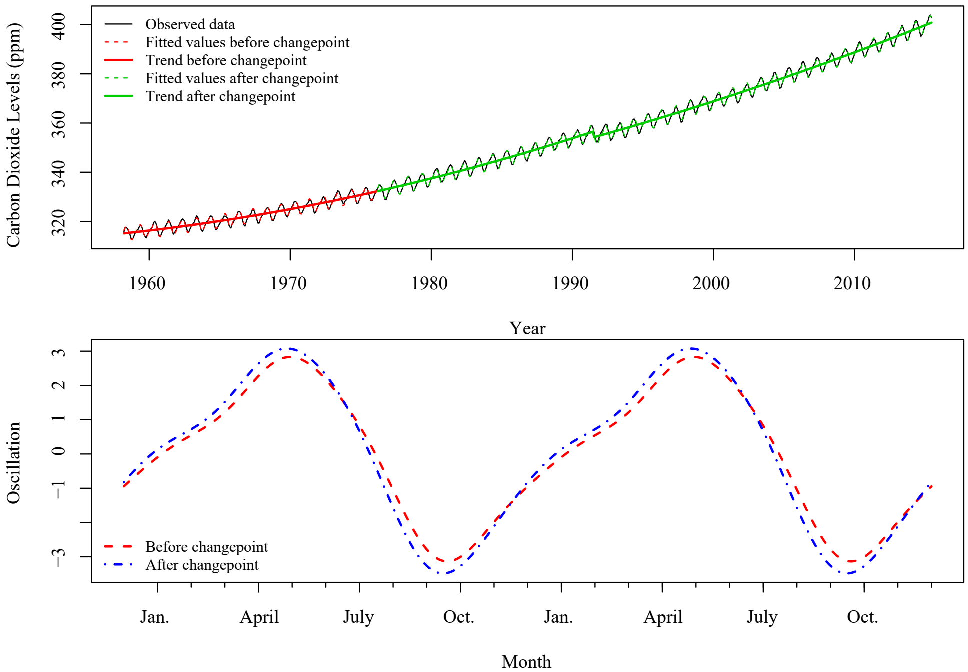

Figure 2. The fitted CO_2 values (with underlying trend function) calculated using the model in alternative hypothesis H1c* (top), and a plot of the expected oscillations, i.e., seasonal variations from the trend, estimated before and after the changepoint found under alternative H1c* (bottom).

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 a 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 (and the following) example are provided in Appendix G of the Supplementary Material.

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 ($n = 1,137$). At 515 km north of the Arctic Circle, Barrow is the northernmost city in the United States. Owing to obvious implications of melting of ice sheets, patterns of warming temperatures in polar climates such as Barrow have been heavily researched in the scientific literature. Such studies frequently observe that polar regions have endured greater rates of warming than other regions have (e.g., Holland and Bitz (2003); Serreze and Francis (2006)). Differential rates of warming by season are also of interest. Using historical and simulated data, several researchers (e.g., Lu and Cai (2009); Screen and Simmonds (2010); Manabe, Ploshay and Lau (2011); Bintanja and Van der Linden (2013)) stipulate that warming is occurring at higher rates during winter seasons than during summer seasons within polar regions.

Because a linear trend is frequently used to model temperature data, the null hypothesis model fitted to the Barrow data is

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

where $s_{j,t}$ obeys (A.4) from the Supplementary 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 (onl 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 γ_1 only (i.e., alternative H1d*).

The 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). However, when we search for a change in trend only (alternative H1b), the changepoint, though still statistically significant, is estimated to occur at $\hat{c} = 513$ (June 1963). When testing for a change in the seasonal variation only (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 observations indicate that the omnibus statistic is dominated by the component that is estimating the change in seasonality. Note that there are two

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

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 nonsignificant 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 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 is 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 the 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 the findings of other authors (e.g., Bloomfield (1992); Jones and Moberg (2003); Jones, Wigley and Wright (2011)), who observe relative stability in global temperatures from the mid-1940's to the mid-1970s followed by an extended period of warming. However, the magnitude of the 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 and Moberg (2003). This observation supports the theory of amplified warming in the Arctic (Holland and Bitz (2003)). However, our study does not confirm the finding of previous work that stipulates that the amount of amplification varies by season (Screen and Simmonds (2010)).

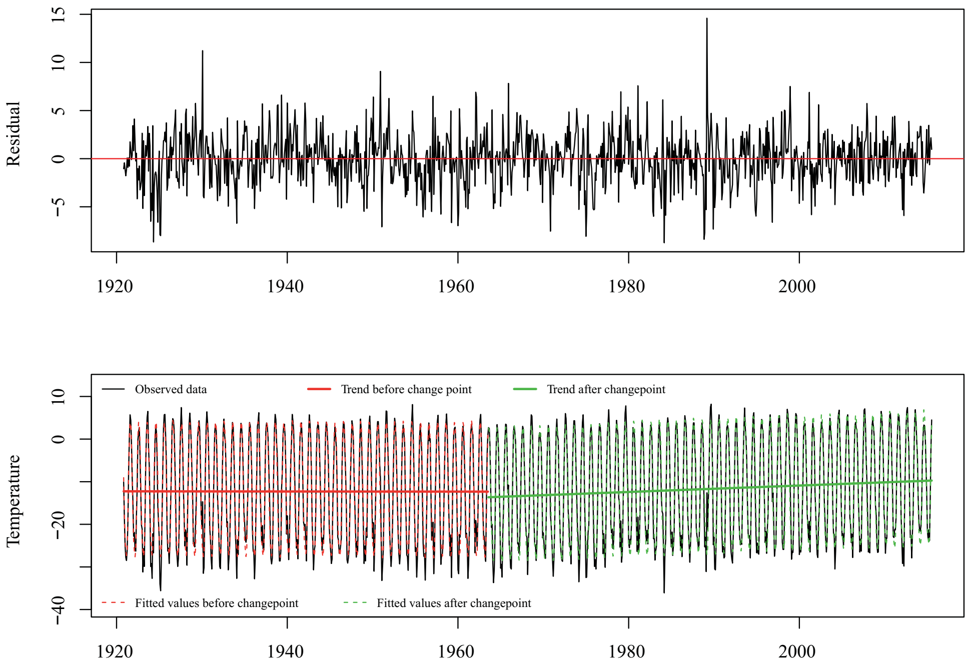

Figure 3. The fitted temperature values (measured in °C) for the Barrow, AK, series with an underlying trend function after alternative model H1a has been fit (top), and the resulting OLS residuals (bottom).

7. Discussion

As observed in prior works (Robbins et al. (2011a); Robbins, Gallagher and Lund (2016)), 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, the convergence of a partial sums sequence of independent terms ($\{Z_t\}$) is known to be quicker than that of autocorrelated variates ($\{\epsilon_t\}$).

The methods introduced here were developed for the AMOC setting. Although we have shown 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 (e.g., see Menne and Williams Jr. (2009); Robbins et al. (2011b)) is often used to detect multiple changepoints with AMOC methods. This process works well, as 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 a multiple changepoint setting (e.g., Bai and Perron (1998); Cho and Fryzlewicz (2011); Yau and Zhao (2016); Horvath, Pouliot and Wang (2017)). Extension of such methods to the general framework considered here is left for future work. Furthmore, note that local alternatives (e.g., Andrews (1993)) are not discussed here; this is also left for future work.

An important innovation provided by the methodology introduced in this paper is the ability to test for changes in a specific coefficient (or set of coefficients) of a large regression model. This enables an analyst to run a variety of tests to assess regression coefficients separately (as 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 coefficients separately. However, if a substantial number of tests are being applied to the same data set, the analyst should consider corrections for multiple testing, such as those that control the false discovery rate (Benjamini and Hochberg (1995)).

Supplementary Material

The online Supplementary Material contains miscellaneous additional content, including technical details and proofs.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821–856.

Andrews, D. W. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 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* **66**, 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 (Statistical Methodology)* , 289–300.

Bintanja, R. and Van der Linden, E. (2013). The changing seasonal climate in the Arctic. *Scientific Reports* **3**, 1556.

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. M. and Bitz, C. M. (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. In *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. B. (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. W., Lund, R. B., Gallagher, C. M. and Lu, Q. 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, 4570 Fifth Avenue #600 Pittsburgh, PA 15213, USA.

E-mail: mrobbins@rand.org

(Received January 2018; accepted October 2018)