BACKWARD CUSUM FOR TESTING AND MONITORING STRUCTURAL CHANGE WITH AN APPLICATION TO COVID-19 PANDEMIC DATA

SVEN OTTO
Institute of Finance and Statistics, University of Bonn

JÖRG BREITUNG
Institute of Econometrics and Statistics, University of Cologne

It is well known that the conventional cumulative sum (CUSUM) test suffers from low power and large detection delay. In order to improve the power of the test, we propose two alternative statistics. The backward CUSUM detector considers the recursive residuals in reverse chronological order, whereas the stacked backward CUSUM detector sequentially cumulates a triangular array of backwardly cumulated residuals. A multivariate invariance principle for partial sums of recursive residuals is given, and the limiting distributions of the test statistics are derived under local alternatives. In the retrospective context, the local power of the tests is shown to be substantially higher than that of the conventional CUSUM test if a break occurs in the middle or at the end of the sample. When applied to monitoring schemes, the detection delay of the stacked backward CUSUM is found to be much shorter than that of the conventional monitoring CUSUM procedure. Furthermore, we propose an estimator of the break date based on the backward CUSUM detector and show that in monitoring exercises this estimator tends to outperform the usual maximum likelihood estimator. Finally, an application of the methodology to COVID-19 data is presented.

1. INTRODUCTION

Cumulative sums have become a standard statistical tool for testing and monitoring structural changes in time series models. The CUSUM test was introduced by Brown, Durbin, and Evans (1975) as a test for structural breaks in the coefficients of a linear regression model $y_t = x_t' \beta_t + u_t$ with time index t, where β_t denotes the coefficient vector, x_t is the vector of regressor variables and u_t is a zero mean error term. Under the null hypothesis, there is no structural change in β_t, while,
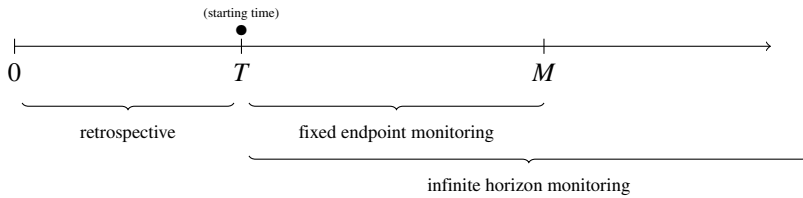
659

FIGURE 1. Retrospective testing and monitoring.

under the alternative hypothesis, the coefficient vector changes at unknown time $T^* \leq T$.

Sequential tests, such as the CUSUM test, consist of a detector statistic and a critical boundary function. The CUSUM detector sequentially cumulates standardized one-step ahead forecast errors, which are also referred to as recursive residuals. The detector is evaluated for each time point within the testing period, and, if its path crosses the boundary function at least once, the null hypothesis is rejected. If the endpoint of the sample is fixed and the test is applied once to the full sample by comparing the path of the detector with the boundary function, the test is called a retrospective test (henceforth: *R-test*). A variety of *R-tests* have been proposed in the literature (for recent reviews, see Robbins et al., 2011; Aue and Horváth, 2013; Casini and Perron, 2019).

Since the seminal work of Chu, Stinchcombe, and White (1996), increasing interest has been focused on monitoring structural stability in real time. Sequential monitoring procedures (henceforth: *M-tests*) consist of a detector statistic and a boundary function that are evaluated for periods beyond some historical time span. The monitoring time span with $t > T$ can either have a fixed endpoint $M < \infty$ or an infinite horizon. In the fixed endpoint setting, the monitoring period starts at $T + 1$ and ends at M, while the boundary function depends on the ratio $m = M/T$. In case of an infinite horizon, the monitoring time span does not need to be specified before the monitoring procedure starts. These two monitoring schemes are also referred to as closed-end and open-end procedures (see Kirch and Kamgaing, 2015). The null hypothesis of no structural change is rejected whenever the path of the detector crosses some critical boundary function for the first time. Monitoring procedures for a fixed end point were proposed in Leisch, Hornik, and Kuan (2000), Zeileis et al. (2005), Wied and Galeano (2013), and Dette and Gösmann (2020), whereas Chu et al. (1996), Horváth et al. (2004), Aue et al. (2006), Fremdt (2015), and Gösmann, Kley, and Dette (2021) considered an infinite monitoring horizon. In recent years, *M-tests* have become popular as tests for speculative bubbles in financial markets (e.g., Phillips, Wu, and Yu, 2011; Homm and Breitung, 2012; Astill et al., 2018).

A well-known drawback of the conventional CUSUM *R-test* is its low power, whereas the CUSUM *M-test* may exhibit large detection delays. This is due to the

fact that the pre-break recursive residuals are uninformative, as their expectation is equal to zero up to the break date, while the recursive residuals have a nonzero expectation after the break. Hence, the cumulative sums of the recursive residuals contain a large number of uninformative residuals that only add noise to the statistic. In contrast, if one cumulates the recursive residuals backwardly from the end of the sample to the beginning, the cumulative sum collects the informative residuals first, and the likelihood of exceeding the critical boundary will typically be larger than when cumulating residuals from the beginning onwards. In this paper, we show that backward CUSUM test procedures may indeed have a much higher power and lower detection delays than the conventional CUSUM *R*- and *M-tests*. Phillips and Shi (2018) proposed a reverse sample scheme for the PSY procedure that is used for detecting crises (bubble collapses). The main difference with our approach is that their regression is performed in reverse order, while our approach estimates the model in the original time but reverses the order of the (recursive) residuals.

Another way of motivating the backward CUSUM testing approach is to consider the simplest possible situation, where, under the null hypothesis, it is assumed that the process is generated as $y_t = \mu + u_t$, with μ and $\sigma^2 = Var(u_t)$ assumed to be known. To test the hypothesis that the mean changes at T^*, we introduce the dummy variable D_t^*, which is unity for $t \geq T^*$ and zero elsewhere. The uniformly most powerful test statistic is the *t*-statistic for the hypothesis $\delta = 0$ in the regression $y_t - \mu = \delta D_t^* + u_t$, which is given by $\sigma^{-1}(T - T^* + 1)^{-1/2} \sum_{t=T^*}^{T} (y_t - \mu)$. If μ is unknown, we may replace it by the full sample mean $\overline{y}$, resulting in the backward cumulative sum of the ordinary least squares (OLS) residuals from period T through T^*. If T^* is unknown, the test statistic is computed for all possible values of T^*, whereas the starting point T of the backward cumulative sum remains constant. Since the sum of the OLS residuals is zero, it follows that the test is equivalent to a test based on the forward cumulative sum of the OLS residuals. In contrast, if we replace μ with the recursive mean $\overline{\mu}_{t-1} = (t-1)^{-1} \sum_{i=1}^{t-1} y_t$, we obtain a test statistic based on the backward cumulative sum of the recursive residuals (henceforth: backward CUSUM). In this case, however, the test is different from a test based on the forward cumulative sum of the recursive residuals (henceforth: forward CUSUM). This is due to the fact that the sum of the recursive residuals is an unrestricted random variable. Accordingly, the two versions of the test may have quite different properties. In particular, it turns out that the backward CUSUM is much more powerful than the standard forward CUSUM at the end of the sample. Accordingly, this version of the CUSUM test procedure is better suited for the purpose of real-time monitoring, where it is crucial to be powerful at the end of the sample.

An additional problem of the conventional CUSUM test is that it has no power against alternatives that do not affect the unconditional mean of y_t (see Krämer, Ploberger, and Alt, 1988). For both retrospective testing and monitoring, we propose a multivariate sequential statistic in the fashion of the score-based cumulative sum statistic of Hansen (1992) and the tests by Jiang and Kurozumi

(2019). The maximum vector entry of the multivariate statistic yields a detector and a sequential test that has power against a much larger class of structural breaks than when using conventional CUSUM detectors.

We also suggest a new estimator for the break date based on backwardly cumulated recursive residuals. This estimator outperforms the conventional estimator constructed by the sum of squared residuals whenever the break occurs close to the end of the sample, which is the relevant scenario for online monitoring.

This paper is organized as follows. In Section 2, the limiting distribution of the multivariate CUSUM process is derived under both the null hypothesis and local alternatives. Section 3 introduces the backward CUSUM and the stacked backward CUSUM tests for both retrospective testing and monitoring. While the backward CUSUM is only defined for $t \leq T$ and can thus be implemented only for retrospective testing, the stacked backward CUSUM cumulates recursive residuals backwardly in a triangular scheme and is therefore suitable for real-time monitoring. The local powers of the tests are compared in Section 4. In the retrospective setting, the powers of the backward CUSUM and the stacked backward CUSUM tests are substantially higher than that of the conventional forward CUSUM test if a single break occurs after one third of the sample size. In the case of monitoring, the detection delay of the stacked backward CUSUM under local alternatives is shown to be much lower than that of the monitoring CUSUM detector by Chu et al. (1996). In Section 5, we present a strong invariance principle for the multivariate CUSUM process and propose an infinite horizon monitoring procedure. Section 6 considers the estimation of the break date based on backwardly cumulated recursive residuals. We present an estimator, which is more accurate than the conventional maximum likelihood estimator if the break is located at the end of the sample. Section 7 presents Monte Carlo simulation results, in Section 8, we provide a real-data example on monitoring SARS-CoV-2 infections during the COVID-19 pandemic, and Section 9 concludes.

Throughout the paper, we use the following notation: $\|a\| = \max_{i=1,\ldots,k} |a_i|$ denotes the maximum norm and $\|A\|_M = \max_{i=1,\ldots,k} \sum_{j=1}^{l} |A_{i,j}|$ denotes the maximum absolute row sum norm, where $a \in \mathbb{R}^k$, and $A \in \mathbb{R}^{k \times l}$. We use $\overset{p}{\longrightarrow}$ to denote convergence in probability as $T \to \infty$, $\overset{d}{\longrightarrow}$ for convergence in distribution, and $\overset{d}{=}$ to indicate that two random variables have the same distribution. The space of right continuous functions with left limits (càdlàg) on $[0, m]$, where $0 < m < \infty$, is denoted as $D([0, m])$, and its k-fold product space is $D([0, m])^k = D([0, m]) \times \cdots \times D([0, m])$. The space is equipped with the Skorokhod metric (see Billingsley, 1999), and the symbol "$\Rightarrow$" denotes weak convergence with respect to this metric.

2. THE MULTIVARIATE CUSUM PROCESS

We consider the multiple linear regression model

$$y_t = x_t'\beta_t + u_t, \quad t \in \mathbb{N}, \tag{1}$$

where y_t is the dependent variable, and $x_t = (1, x_{t2}, \ldots, x_{tk})'$ is the vector of regressor variables including a constant. The $k \times 1$ vector of regression coefficients β_t depends on the time index t, and u_t is an error term. The time point T divides the time horizon into the retrospective time period $t \leq T$ and the monitoring period $t > T$. We impose the following assumptions on the regressors and the error term.

Assumption 1.

(a) The errors satisfy $E[u_t] = 0$, $E[u_t^2] = \sigma^2 > 0$, and $E[|u_t|^8] < \infty$ for all t.
(b) The regressors satisfy $E[\|x_t\|^8] < \infty$ for all t, and the sample covariance matrices $\widehat{C}_t = t^{-1} \sum_{j=1}^{t} x_j x_j'$ are uniformly positive definite for all $t > k$ with $\mathrm{plim}_{T \to \infty} \widehat{C}_T = C$.
(c) There exists a positive definite Ω such that $\mathrm{plim}_{T \to \infty} T^{-1} (\sum_{t=1}^{T} x_t u_t)(\sum_{t=1}^{T} x_t u_t)' = \Omega$.

Model (1) allows for conditionally heteroskedastic errors and local nonstationary regressors, provided that a global long-run covariance matrix Ω exists. The regressors can contain lagged dependent variables such as in autoregressive distributed lag models. We focus on models with a correctly specified dynamic structure and uncorrelated errors.

Assumption 2. The error process u_t is a martingale difference sequence with respect to $\mathcal{F}_t$, the σ-algebra generated by $\{(x_{i+1}', u_i)', \ i \leq t\}$.

Following Brown et al. (1975) this assumption rules out autocorrelated error processes. In practice, this may require a dynamic specification with a suitable lag distribution of the variables. In Remark 1, we show that autocorrelated errors can be accommodated by replacing the ordinary covariance matrix by a (consistent estimate of the) long-run covariance matrix. Since the estimation of long-run covariances can lead to finite sample size distortions (see e.g., Casini, 2021), Assumption 2 is a common and convenient assumption in practice. The expression of the global covariance matrix simplifies to $\Omega = \sigma^2 C$ under Assumption 2.

Recursive residuals for linear regression models were introduced by Brown et al. (1975) as standardized one-step ahead forecast errors, and are defined as

$$w_t = \frac{y_t - x_t' \widehat{\beta}_{t-1}}{(1 + x_t'(\sum_{i=1}^{t-1} x_i x_i')^{-1} x_t)^{1/2}}, \quad t \geq k+1,$$

and $w_t = 0$ for $t = 1, \ldots, k$, where $\widehat{\beta}_{t-1} = (\sum_{i=1}^{t-1} x_i x_i')^{-1} \sum_{i=1}^{t-1} x_i y_i$. Using recursive residuals instead of ordinary OLS residuals as in Ploberger and Krämer (1992) has a number of advantages. First, the recursive residuals behave exactly as under the null hypothesis until the parameters change, whereas a structural break affects all OLS residuals in a different manner. Second, under Assumptions 1 and 2, the recursive residuals form a martingale difference sequence regardless of the estimation error in the recursive residuals. By contrast, the OLS residuals are (slightly) autocorrelated, which only disappears if the sample size gets large.

The conventional CUSUM detector is given by $S_{t,T} = \widehat{\sigma}_T^{-1} T^{-1/2} \sum_{i=1}^{t} w_i$, where $\widehat{\sigma}_T^2$ denotes the sample variance of $\{w_{k+1}, \ldots, w_T\}$. Under the null hypothesis $H_0 : \beta_t = \beta_0$ for all t, the univariate CUSUM process obeys the functional central limit theorem $S_{\lfloor rT \rfloor, T} \Rightarrow W(r)$, where $W(r)$ is a standard Brownian motion (see Sen, 1982). The univariate CUSUM *R-test* of Brown et al. (1975) rejects the null hypothesis if the path of $|S_{t,T}|$ exceeds the linear critical boundary function $b_t = \lambda_\alpha d_{\text{lin}}(t/T)$ for at least one time index $t = 1, \ldots, T$, where

$$d_{\text{lin}}(r) = 1 + 2r. \tag{2}$$

The critical value λ_α is the $(1 - \alpha)$ quantile of $\sup_{0 \le r \le 1} |W(r)|/d_{\text{lin}}(r)$ and determines the significance level α, which accounts for the multiplicity issue of the sequential test procedure. In the monitoring context, Chu et al. (1996) considered the radical type boundary function $b_{\text{rad}}(r) = r^{1/2}(\log(r) - \log(\alpha^2))^{1/2}$, which is derived from the boundary crossing probability for a Brownian motion (see Robbins and Siegmund, 1970). The conventional univariate CUSUM *M-test* rejects the null hypothesis if the detector statistic $|S_{t,T} - S_{T,T}|$ exceeds $b_t = b_{\text{rad}}(t/T)$ for some $t > T$.

A weakness of univariate CUSUM tests is that they focus on breaks in the intercept. Ploberger and Krämer (1990) studied local alternatives of the form $\beta_t = \beta_0 + T^{-1/2} g(t/T)$, where $g : \mathbb{R} \to \mathbb{R}^k$ is piecewise constant and bounded. The authors showed that $S_{\lfloor rT \rfloor, T} \Rightarrow W(r) + \pi' h(r)$, where $\pi = e_1' C$, $e_1 = (1, 0, \ldots, 0)'$, and

$$h(r) = \frac{1}{\sigma} \int_0^r g(z)\, dz - \frac{1}{\sigma} \int_0^r \int_0^z \frac{1}{z} g(v)\, dv\, dz. \tag{3}$$

Consequently, univariate CUSUM tests have no power if $g(r)$ is orthogonal to π. To sidestep this difficulty, we follow Jiang and Kurozumi (2019) and consider the multivariate statistic

$$Q_T(r) = \frac{1}{\widehat{\sigma}_T \sqrt{T}} \widehat{C}_T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} x_t w_t.$$

Under Assumption 1, the multivariate series $x_t u_t$ obeys a multivariate functional central limit theorem (see Phillips and Durlauf, 1986), which also applies to the multivariate CUSUM process of recursive residuals.

THEOREM 1. *Let Assumptions 1 and 2 hold true. If $\beta_t = \beta_0$ for all t, then*

$$Q_T(r) \Rightarrow W^{(k)}(r), \quad r \in [0, m], \tag{4}$$

for any $m < \infty$, as $T \to \infty$, where $W^{(k)}(r)$ is a k-dimensional standard Brownian motion. If $\beta_t = \beta_0 + T^{-1/2} g(t/T)$, where $g(r)$ is piecewise constant and bounded, then

$$Q_T(r) \Rightarrow W^{(k)}(r) + C^{1/2} h(r), \quad r \in [0, m].$$

This is an extension of the results in Jiang and Kurozumi (2019), who considered slightly stronger assumptions and no local alternatives. Note that the function $g(r)$ is constant if and only if β_t is constant. If $\beta_t = \beta_0$ for all t, we have $h(r) = 0$. By contrast, under a local alternative with a nonconstant break function $g(r)$, it follows that $h(r)$ is nonzero, and, consequently, $C^{1/2}h(r)$ is nonzero, since $C^{1/2}$ is positive definite. Hence, sequential tests that are based on $Q_T(r)$ have power against a larger class of alternatives than the tests of Brown et al. (1975) and Chu et al. (1996).

Therefore, we consider *R*- and *M-tests* that are based on the multivariate detector $Q_{t,T} = Q_T(t/T)$. Note that $Q_{t,T} = S_{t,T}$ if there is only an intercept in the model. The multivariate forward CUSUM *R-test* is defined by the following rule: the null hypothesis is rejected if the path of $\|Q_{t,T}\|$ exceeds the boundary function $b_t = \lambda_\alpha d(t/T)$ for at least one index $t = 1, \ldots, T$. Equivalently, we can express this sequential test as a one-shot test, where H_0 is rejected if the maximum statistic $\mathcal{Q}_T = \max_{t=1,\ldots,T} \|Q_{t,T}\|/d(t/T)$ exceeds the critical value λ_α, which is the $(1-\alpha)$ quantile of its limiting null distribution.

Assumption 3. The boundary function is of the form $b(r) = \lambda_\alpha d(r)$, where $d(r)$ is continuous. There exists $\epsilon > 0$ such that $d(r) > \epsilon$ for all $r \geq 0$.

By Theorem 1 and the continuous mapping theorem it follows that

$$\mathcal{Q}_T \xrightarrow{d} \sup_{r \in (0,1)} \frac{\|W^{(k)}(r)\|}{d(r)}$$

under the null hypothesis. The multivariate forward CUSUM *M-test* with fixed endpoint $M = \lfloor mT \rfloor$ rejects H_0 if the path of $\|Q_{t,T} - Q_{T,T}\|$ exceeds the boundary function $b_t = \lambda_\alpha d((t-T)/T)$ for at least one index $t = T+1, \ldots, \lfloor mT \rfloor$, where $1 < m < \infty$. The corresponding maximum statistic is $\mathcal{Q}_{T,m} = \max_{t=T+1,\ldots,\lfloor mT \rfloor} \|Q_{t,T} - Q_{T,T}\|/d((t-T)/T)$, where, under H_0,

$$\mathcal{Q}_{T,m} \xrightarrow{d} \sup_{r \in (0,m-1)} \frac{\|W^{(k)}(r)\|}{d(r)}.$$

Remark 1. If the dynamics of the model are not specified correctly, the errors may be autocorrelated and Assumption 2 does not apply. In this case, the limiting distribution differs from that in (4) and depends on the global long-run covariance matrix Ω. Under additional strong mixing assumptions, the process $x_t u_t$ obeys the multivariate functional central limit theorem $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} x_t u_t \Rightarrow \Omega^{1/2} W^{(k)}(r)$ (see Wooldridge and White, 1988). To obtain the same limiting distribution as in Theorem 1, we may consider the modified multivariate CUSUM detector

$$\widetilde{Q}_T(r) = \frac{1}{\sqrt{T}} \widehat{\Omega}_T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} x_t w_t, \quad r \in [0,m],$$

where $\widehat{\Omega}_T$ is some consistent estimator for Ω. Suitable choices are the long-run covariance estimators of Newey and West (1987) and Andrews (1991). An

alternative is the double kernel HAC estimator by Casini (2021), which performs well in the presence of locally stationary regressors. In the Appendix, we show that if Assumption 1 holds and if there exists $\kappa \geq 8$ such that $\sup_t E[\|x_t\|^\kappa] < \infty$, $\sup_t E[|u_t|^\kappa] < \infty$, and $(x_t, u_t)'$ is strong mixing of size $-\kappa/(\kappa - 6)$, then, under the null hypothesis, $\widetilde{Q}_T(r) \Rightarrow W^{(k)}(r)$, as $T \to \infty$. Therefore, all R- and M-tests can also be constructed based on the modified detector $\widetilde{Q}_{t,T} = \widetilde{Q}_T(t/T)$. A second approach to deal with possible autocorrelation was proposed by Robbins et al. (2011). Their two-step adjustment approach first employs an ARMA model in order to obtain the prewhitened residuals which in turn replaces the original residuals in the detector. In our case the ARIMA prewhitening may be performed in a recursive fashion. As shown by Robbins et al. (2011) the resulting detector possesses similar asymptotic properties as the original detector (apart from a scaling factor that depends on the long-run variance).

Remark 2. In practice, partial or one-sided tests can be beneficial in terms of a more powerful test if one is interested in breaks in certain coefficients or directions. For testing the partial hypothesis $H_0 : H'\beta_t = H'\beta_0$, where H is a $k \times l$ matrix with full column rank, we consider the partial CUSUM process $Q_T^*(t/T) = Q_{t,T}^* = \widehat{\sigma}_T^{-1} T^{-1/2} (H'C_T H)^{-1/2} H' \sum_{j=1}^{t} x_j w_j$. All R- and M-tests can be defined with respect to $Q_{t,T}^*$, where $Q_{t,T}^*(r) \Rightarrow W^{(l)}(r)$, under H_0 and the conditions of Theorem 1. The R-test by Brown et al. (1975) and the M-test by Chu et al. (1996) are partial structural break tests for which the matrix H coincides with the first unit vector. In case of one-sided tests, e.g., $H_1 : H'\beta_t > H'\beta_0$, the maximum norm can be replaced by the simple maximum, so that H_0 is rejected if $p(H'Q_{t,T})$ exceeds the respective boundary function, where $p(x) = \max_{i=1,\ldots,l} x_i, x \in \mathbb{R}^l$.

3. BACKWARD CUSUM R- AND M-TESTS

3.1. Backward CUSUM R-Test

An alternative approach is to cumulate the recursive residuals in reversed order. Suppose there is a single break in β_t at time $t = T^*$. Then, $\{w_t, \ t < T^*\}$ are the residuals from the prebreak period, and $\{w_t, \ t \geq T^*\}$ are those from the postbreak period. As the prebreak recursive residuals are not affected by a violation of the null hypothesis, they do not provide useful information about a subsequent break. Accordingly, the partial sum process $T^{-1/2} \sum_{j=1}^{t} w_j$ behaves like a pure random walk for $t < T^*$ and cumulating those residuals brings nothing but noise to the detector statistic. In contrast, the postbreak residuals have nonzero mean and reveal relevant information about a possible break. In order to focus on the postbreak residuals, we therefore consider backwardly cumulated partial sums of the form $T^{-1/2} \sum_{j=0}^{t-1} w_{T-j}$. We define the retrospective backward CUSUM detector as

$$BQ_{t,T} = Q_T(1) - Q_T\left(\frac{t-1}{T}\right) = \frac{1}{\widehat{\sigma}_T \sqrt{T}} C_T^{-1/2} \sum_{j=t}^{T} x_j w_j \quad (t = 1, \ldots, T).$$

TABLE 1. Asymptotic critical values for $\mathcal{Q}_T$ and $\mathcal{BQ}_T$.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\alpha=10\%$	0.848	0.944	0.996	1.031	1.058	1.080	1.097	1.112	1.125	1.138
$\alpha=5\%$	0.947	1.034	1.082	1.115	1.141	1.161	1.177	1.190	1.203	1.214
$\alpha=1\%$	1.144	1.219	1.258	1.283	1.303	1.324	1.343	1.357	1.368	1.381

Note: Simulated $(1-\alpha)$ quantiles of $\sup_{r\in(0,1)}\|W^{(k)}(r)\|/(1+2r)$ based on 100,000 Monte Carlo replications are reported. The Wiener process is approximated on a grid of 50,000 equidistant points.

The null hypothesis is rejected if $\|B\mathcal{Q}_{t,T}\|$ exceeds the boundary $b_t = \lambda_\alpha d((T-t-1)/T)$ for at least one time index t. The maximum statistic is given by

$$\mathcal{BQ}_T = \max_{t=1,\dots,T} \frac{\|B\mathcal{Q}_{t,T}\|}{d(\frac{T-t+1}{T})},$$

and, under the local alternatives defined in Theorem 1, the continuous mapping theorem implies

$$\mathcal{BQ}_T \xrightarrow{d} \sup_{r\in(0,1)} \frac{\|W^{(k)}(1)+C^{1/2}h(1)-W^{(k)}(r)-C^{1/2}h(r)\|}{d(1-r)}$$

$$\stackrel{d}{=} \sup_{r\in(0,1)} \frac{\|W^{(k)}(r)+C^{1/2}(h(1)-h(1-r))\|}{d(r)}. \tag{5}$$

Hence, the limiting distribution of $\mathcal{BQ}_T$ under H_0 coincides with that of $\mathcal{Q}_T$. Simulated asymptotic critical values under the linear boundary (2) are presented in Table 1. Under local alternatives, the limiting distributions of $\mathcal{BQ}_T$ and $\mathcal{Q}_T$ differ. A simple illustrative example of the detector paths together with the linear boundary (2) of Brown et al. (1975) are depicted in Figure 2, in which two processes with $k=1$ and a single break in the mean at 3/4 and 1/4 of the sample are simulated.

Unlike the forward CUSUM detector, the backward CUSUM detector is not measurable with respect to the filtration of available information at time t and is therefore not suitable for a monitoring procedure. The path of $\|B\mathcal{Q}_{t,T}\|$ cannot be monitored in real-time, as it is only defined for $t \leq T$ with fixed endpoint T. To obtain a feasible *M-test* in practice, we resort to a triangular backward inspection scheme of recursive residuals, which is discussed below.

3.2. Stacked Backward CUSUM R-Test

Let $\mathcal{BQ}_T(t) = \max_{s=1,\dots,t} \|\mathcal{Q}_T(t/T) - \mathcal{Q}_T((s-1)/T)\|/d((t-s+1)/T)$ be the backward CUSUM maximum statistic with endpoint t. The idea of the stacked backward CUSUM scheme is to compute this statistic sequentially for each t,
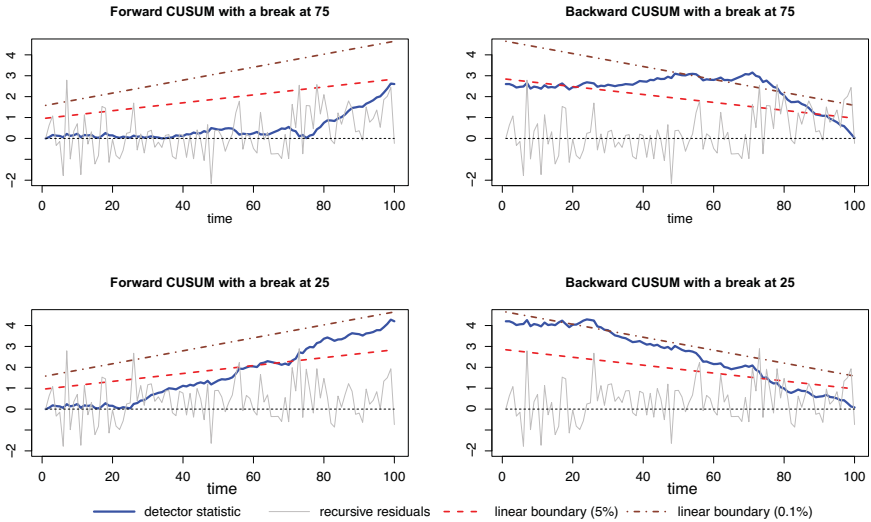
FIGURE 2. Illustrative example for the backward CUSUM with a break in the mean.

Note: The process $y_t = \mu_t + u_t$, $t = 1, \dots, T$, is simulated for $T = 100$ with $\mu_t = 0$ for $t < \tau^* T$, $\mu_t = 1$ for $t \geq \tau^* T$, and $u_t \sim NID(0,1)$, where $\tau^* = 0.75$ in the upper panels and $\tau^* = 0.25$ in the lower panels. The bold solid line paths are the trajectories of $\|Q_{t,T}\|$ and $\|BQ_{t,T}\|$, where the detectors are univariate such that the norm is just the absolute value. In the background, the recursive residuals are plotted. The dashed and dash-dotted lines correspond to the linear boundary $d_{\text{lin}}(r)$ with significance levels $\alpha = 5\%$ and $\alpha = 0.1\%$, respectively.

yielding $\mathcal{BQ}_T(1), \dots, \mathcal{BQ}_T(T)$. The corresponding maximum statistic $\mathcal{SBQ}_T$ is the maximum among this sequence of backward CUSUM statistics. An important feature is that this sequence is measurable with respect to the filtration of information at time t, so that $\mathcal{BQ}_T(t)$ is itself a sequential statistic. Stacking all backward CUSUM statistics on one another leads to a triangular array structure given by

$$SBQ_{s,t,T} = Q_T\left(\tfrac{t}{T}\right) - Q_T\left(\tfrac{s-1}{T}\right) = \frac{1}{\widehat{\sigma}_T \sqrt{T}} C_T^{-1/2} \sum_{j=s}^{t} x_j w_j \quad (t \in \mathbb{N},\ s = 1, \dots, t),$$

(6)

which is denoted as the stacked backward CUSUM detector. We reject H_0 if $\|SBQ_{s,t,T}\|$ exceeds the triangular boundary $b_{s,t} = b(\tfrac{t}{T}, \tfrac{s-1}{T})$ for some $t = 1, \dots, T$ and $s = 1, \dots, t$.

Assumption 4. The triangular boundary function is of the form $b(r,s) = \lambda_\alpha d(r,s)$, where $d(r,s)$ is continuous. There exists $\epsilon > 0$ such that $d(r,s) > \epsilon$ for all $0 \leq s \leq r$.

The stacked backward CUSUM *R-test* can be equivalently expressed in terms of a maximum statistic. H_0 is rejected if the double maximum statistic

$$\mathcal{SBQ}_T = \max_{t=1,\dots,T} \mathcal{BQ}_T(t) = \max_{t=1,\dots,T} \max_{s=1,\dots,t} \frac{\|SBQ_{s,t,T}\|}{d(\frac{t}{T}, \frac{s-1}{T})},$$

exceeds λ_α. Under the local alternatives defined in Theorem 1, it follows that

$$\mathcal{SBQ}_T \xrightarrow{d} \sup_{r \in (0,1)} \sup_{s \in (0,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s) + C^{1/2}[h(r) - h(s)]\|}{d(r,s)}.$$

3.3. Stacked Backward CUSUM M-Test

Since the triangular detector (6) is measurable with respect to the information set at time t, it can be monitored online across all time points $t > T$. The null hypothesis is rejected if $\|SBQ_{s,t,T}\|$ exceeds $b_{s,t} = b(\frac{t}{T}, \frac{s-1}{T})$ at least once for some $t \geq T+1$ and $s = T+1, \dots, t$. The *M-test* maximum statistic with a fixed horizon $m < \infty$ is given by

$$\mathcal{SBQ}_{T,m} = \max_{t=T+1,\dots,\lfloor mT \rfloor} \max_{s=T+1,\dots,t} \frac{\|SBQ_{s,t,T}\|}{d(\frac{t}{T}, \frac{s-1}{T})},$$

where, analogously to (5),

$$\mathcal{SBQ}_{T,m} \xrightarrow{d} \sup_{r \in (1,m)} \sup_{s \in (1,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s) + C^{1/2}[h(r) - h(s)]\|}{d(r,s)}$$

$$\overset{d}{=} \sup_{r \in (0,m-1)} \sup_{s \in (0,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s) + C^{1/2}[h(r+1) - h(s+1)]\|}{d(r,s)}.$$

Simulated critical values for the stacked backward CUSUM *R*- and *M*-tests under the linear triangular boundary

$$d_{\text{sbq}}(r,s) = 1 + 2(r-s) \tag{7}$$

are presented in Table 2.

4. LOCAL POWER

In order to illustrate the advantages of the backward CUSUM tests, we consider the simple local break model $\beta_t = \beta_0 + T^{-1/2}g(t/T)$ with $g(r) = c1_{\{r \geq \tau^*\}}$, where $c \in \mathbb{R}^k$, and τ^* denotes the break location. From (3) it follows that

$$h(r) = c\sigma^{-1}\left(\int_{\tau^*}^{r} dz - \int_0^r \int_{\tau^*}^z \frac{1}{z} dv\, dz\right) = c\sigma^{-1}\tau^*(\ln(r) - \ln(\tau^*))1_{\{r \geq \tau^*\}}, \tag{8}$$

TABLE 2. Asymptotic critical values for $\mathcal{SBQ}_{T,m}$ (fixed m).

|	$k=1$			$k=2$			$k=3$			$k=4$		
α	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$m=1.2$	0.780	0.859	1.023	0.857	0.932	1.082	0.900	0.973	1.121	0.930	1.002	1.147
$m=1.4$	0.944	1.030	1.208	1.026	1.107	1.270	1.073	1.153	1.316	1.107	1.183	1.345
$m=1.6$	1.024	1.114	1.290	1.109	1.189	1.356	1.156	1.235	1.398	1.190	1.266	1.428
$m=1.8$	1.077	1.166	1.341	1.161	1.241	1.406	1.207	1.285	1.446	1.241	1.318	1.476
$m=2$	1.116	1.202	1.374	1.195	1.274	1.438	1.243	1.319	1.479	1.275	1.351	1.506
$m=4$	1.268	1.346	1.510	1.342	1.414	1.567	1.386	1.455	1.600	1.415	1.483	1.625
$m=10$	1.392	1.462	1.610	1.460	1.527	1.665	1.499	1.564	1.695	1.526	1.589	1.722
	$k=5$			$k=6$			$k=7$			$k=8$		
α	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$m=1.2$	0.953	1.021	1.167	0.971	1.038	1.182	0.986	1.052	1.194	0.999	1.065	1.205
$m=1.4$	1.131	1.206	1.363	1.151	1.225	1.378	1.167	1.240	1.390	1.180	1.253	1.402
$m=1.6$	1.214	1.290	1.446	1.235	1.310	1.461	1.251	1.324	1.473	1.264	1.337	1.486
$m=1.8$	1.265	1.340	1.493	1.285	1.360	1.512	1.301	1.374	1.525	1.314	1.387	1.538
$m=2$	1.299	1.374	1.529	1.318	1.392	1.544	1.334	1.407	1.555	1.347	1.419	1.565
$m=4$	1.436	1.504	1.644	1.453	1.522	1.659	1.469	1.536	1.673	1.482	1.548	1.683
$m=10$	1.546	1.608	1.739	1.563	1.624	1.755	1.576	1.638	1.765	1.587	1.649	1.774

Note: Simulated $(1-\alpha)$ quantiles of $\sup_{r\in(0,m-1)}\sup_{s\in(0,r)}\|W^{(k)}(r)-W^{(k)}(s)\|/(1+2(r-s))$ based on 100,000 Monte Carlo replications are reported. The Wiener process is approximated on a grid of 50,000 equidistant points. The critical values for $\mathcal{SBQ}_T$ (R-test) coincide with those for $\mathcal{SBQ}_{T,m}$ (M-test) with $m=2$.

and, under the linear boundaries (2) and (7), the R-tests satisfy

$$\mathcal{Q}_T \xrightarrow{d} \sup_{r\in(0,1)} \frac{\|W(r)+h(r)\|}{1+2r}, \quad \mathcal{BQ}_T \xrightarrow{d} \sup_{r\in(0,1)} \frac{\|W(r)+h(1)-h(1-r)\|}{1+2r},$$

$$\mathcal{SBQ}_T \xrightarrow{d} \sup_{r\in(0,1)}\sup_{s\in(0,r)} \frac{\|W(r)-W(s)+h(r)-h(s)\|}{1+2(r-s)}. \tag{9}$$

Asymptotic local power curves from the limiting distributions in (9) for the case $k=1$ are presented in Figure 3. The $(2,3)$-element of the panel of figures shows that for a fixed break size the backward CUSUM and the stacked backward CUSUM outperform the forward CUSUM if a single break τ^* is located after 15% of the sample size. If the break date τ^* tends to the end of the sample, the power gain of $\mathcal{BQ}_T$ and $\mathcal{SBQ}_T$ increases substantially.

For the M-test statistics with fixed endpoint $m=2$, the limiting distributions of $\mathcal{Q}_{T,m}$ and $\mathcal{SBQ}_{T,m}$ for a break at $\tau^* \in (1,2)$ coincide with those for the R-tests presented in (9) for a break at $\tau^* \in (0,1)$. Hence, the power of $\mathcal{SBQ}_{T,m}$
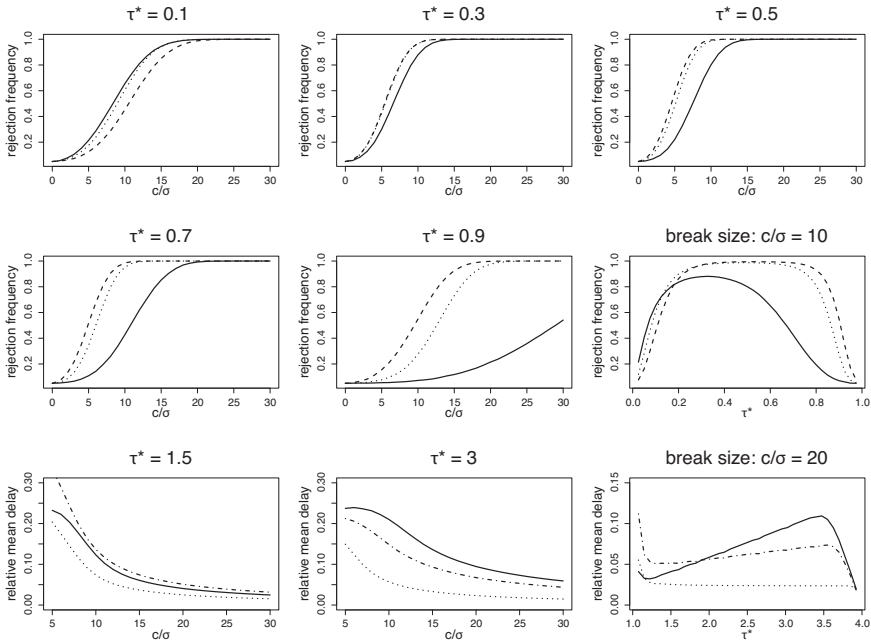
FIGURE 3. Asymptotic local power curves.

Note: The upper six panels show simulated asymptotic local power curves for the *R-tests* $\mathcal{Q}_T$ (solid), $\mathcal{BQ}_T$ (dashed) and $\mathcal{SBQ}_T$ (dotted) from equation (9). The bottom three panels show simulated asymptotic mean local delays for the *M-tests* $\mathcal{Q}_{T,m}$ (solid) and $\mathcal{SBQ}_{T,m}$ (dotted) from equation (10), and the test by Chu et al. (1996) (dash-dotted), where $m = 4$. The Brownian motions in the limiting distributions are approximated on a grid of 1,000 equidistant points and the rates are obtained from 100,000 Monte Carlo repetitions using size-adjusted 5% critical values.

is higher than that of $\mathcal{Q}_{T,m}$ if breaks are located after 15% of the premonitoring sample.

Another important performance measure for *M-tests* is the delay between the actual break and the detection time point. Aue and Horváth (2004) and Aue, Horváth, and Reimherr (2009) derived the asymptotic distribution of the detection stopping link time of CUSUM *M*-tests that are based on OLS residuals. For $\mathcal{Q}_{T,m}$ and $\mathcal{SBQ}_{T,m}$, the detection stopping times are given by

$$\mathcal{D}_{T,m,Q} = \min\left\{t \in \{T+1,\ldots,\lfloor mT\rfloor\}\,\middle|\, \frac{\|Q_{t,T} - Q_{T,T}\|}{d(\frac{t-T}{T})\lambda_{\alpha,m,Q}} \geq 1\right\},$$

$$\mathcal{D}_{T,m,SBQ} = \min\left\{t \in \{T+1,\ldots,\lfloor mT\rfloor\}\,\middle|\, \max_{T<s\leq t}\frac{\|SBQ_{s,t,T}\|}{d(\frac{t}{T},\frac{s-1}{T})\lambda_{\alpha,m,SBQ}} \geq 1\right\},$$

where $\lambda_{\alpha,m,Q}$ and $\lambda_{\alpha,m,SBQ}$ are the corresponding critical values. Under the same setting as in (9), the relative detection stopping times satisfy

$$\frac{\mathcal{D}_{T,m,Q}}{T} \xrightarrow{d} \min\left\{r \in (0, m-1) \,\middle|\, \frac{\|W(r) + h(r+1) - h(r)\|}{(1+2r)\lambda_{\alpha,m,Q}} \geq 1\right\}, \qquad (10)$$

$$\frac{\mathcal{D}_{T,m,SBQ}}{T} \xrightarrow{d} \min\left\{r \in (0, m-1) \,\middle|\, \sup_{s \in (0,r)} \frac{\|W(r) - W(s) + h(r+1) - h(s+1)\|}{(1+2(r-s))\lambda_{\alpha,m,Q}} \geq 1\right\},$$

as $T \to \infty$, where the limiting relative stopping times are denoted as $\tau_{D,Q}$ and $\tau_{D,SBQ}$, respectively. The asymptotic mean local delays $E(\tau_D \mid \tau^* \leq \tau_D \leq m) - \tau^*$, where $\tau_D \in \{\tau_{D,Q}, \tau_{D,SBQ}\}$, are presented in the bottom panels of Figure 3 for $m = 4$ and different break locations. The asymptotic mean local delay of $\mathcal{SBQ}_{T,m}$ is much lower than that of $\mathcal{Q}_{T,m}$. Moreover, the asymptotic mean local delay of $\mathcal{SBQ}_{T,m}$ slowly decreases in τ^* and is much lower than that of $\mathcal{Q}_{T,m}$, except for early breaks.

Remark 3. While, for one-shot tests, the critical value determines the type I error, sequential testing involves two degrees of freedom. Besides the test size, which is controlled asymptotically by an appropriately chosen value for λ_α, the shape of the boundary determines the distribution of potential relative crossing time points r. As already noted by Brown et al. (1975), the forward CUSUM with the linear boundary (2) puts more weight on detecting breaks that occur early in the sample (c.f. Figure 3). In Figure 4, we present the distributions of the first boundary crossing under the null hypothesis, which is also referred to as the "distribution of the size" (see Anatolyev and Kosenok, 2018). The results indicate that the size is skewed for the forward and backward CUSUM tests and almost evenly distributed for stacked backward CUSUM tests, which is due to the weighting scheme of the linear triangular boundary function (7). There is no consensus on which distribution should be preferred, as whether one wishes to put more weight on particular regions of time points of rejection depends on the particular application. However, Zeileis et al. (2005) and Anatolyev and Kosenok (2018) argue that if no further information is available, one might prefer a uniform distribution to a skewed one.

5. INFINITE HORIZON MONITORING

The functional central limit theorem given by Theorem 1 is not suitable for analyzing the asymptotic behavior of an infinite horizon monitoring statistic, since the variance of $Q_T(r)$ is unbounded as $r \to \infty$, and $\sup_{r \geq 1} \|Q_T(r) - W^{(k)}(r)\|$ might not converge in general. Instead, we need an almost sure invariance principle, which is specified by the following high level condition:

Assumption 5. There exists a k-dimensional Brownian motion $W^{(k)}(t)$ such that

$$\sum_{j=1}^{t} x_j u_j = \Omega^{1/2} W^{(k)}(t) + o(t^{1/2}), \qquad \text{(a.s.)}.$$
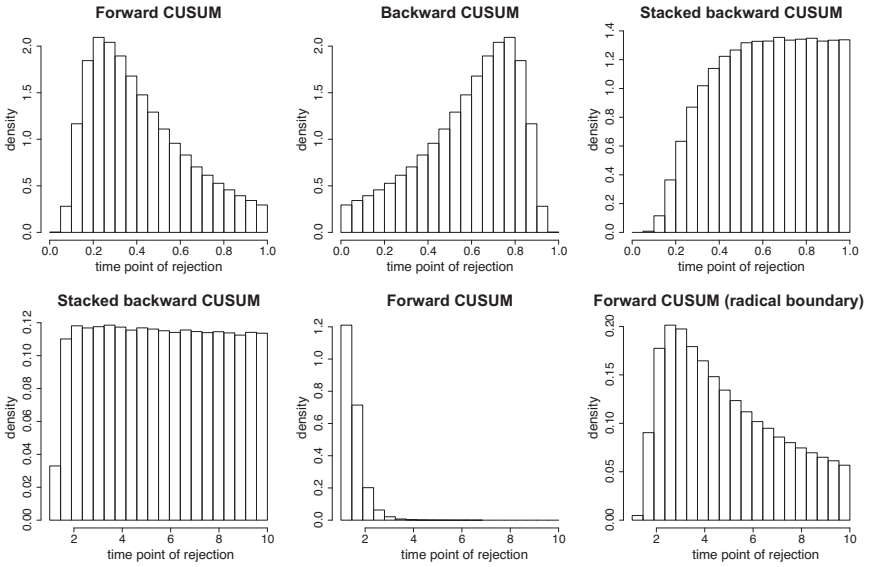
FIGURE 4. Size distributions of the retrospective and monitoring detectors.
Note: The frequencies of the location of the first boundary exceedance under the null hypothesis are shown for a significance level of 5% for the model with $k = 1$. The frequencies are based on random draws under the limiting null distribution of the maximum statistics. The retrospective cases are considered for the upper three histograms and the fixed endpoint monitoring case with $m = 10$ for the lower three. The linear boundaries (2) and (7) are considered in the first five plots and the radical boundary by Chu et al. (1996) is used in the last plot.

Almost sure invariance principles were first studied by Strassen (1967), who verified Assumption 5 under the additional assumption that $x_t u_t$ is a stationary and ergodic martingale difference sequence. Optimal rates were first derived by Komlós, Major, and Tusnády (1975). Aue and Horváth (2004) and Aue et al. (2009) present examples where Assumption 5 is satisfied, which include martingale difference sequences and linear processes with GARCH-type innovations under mild regularity conditions. For more general dependent processes under suitable regularity conditions, Assumption 5 was shown by Wu et al. (2007) and Berkes, Liu, and Wu (2014) with respect to a physical dependence measure (see also Berkes et al., 2011).

THEOREM 2. *Let Assumptions 1, 2, and 5 hold true and let $\beta_t = \beta_0$ for all $t \in \mathbb{N}$. There exists a k-dimensional standard Brownian motion $W^{(k)}(r)$, such that, as $T \to \infty$,*

$$\sup_{r>1} r^{-1/2} \|Q_T(r) - W^{(k)}(r)\| = o_P(1).$$

This result is the key tool to establish the limiting distributions of infinite horizon monitoring statistics under H_0 and indicates the need of further restrictions on the boundary function. In the Appendix, we also show that this result remains valid if we replace $Q_T(r)$ by the autocorrelation robust statistic $\widetilde{Q}_T(r)$ under additional mixing assumptions. The infinite horizon forward CUSUM and stacked backward CUSUM maximum statistics are defined as

$$
\mathcal{Q}_{T,\infty} = \max_{t \geq T+1} \frac{\|Q_T(\frac{t}{T}) - Q_T(1)\|}{d(\frac{t-T}{T})},
$$

$$
\mathcal{SBQ}_{T,\infty} = \max_{t \geq T+1} \max_{s=T+1,\ldots,t} \frac{\|Q_T(\frac{t}{T}) - Q_T(\frac{s-1}{T})\|}{d(\frac{t}{T}, \frac{s-1}{T})}.
$$

Analogously to the fixed horizon case, H_0 is rejected if the statistic exceeds the $(1 - \alpha)$ quantile of its limiting null distribution. Since a maximum over a noncompact set can be unbounded, we need further restrictions on the boundary functions for infinite horizon monitoring.

Assumption 6. The boundary functions that are defined in Assumptions 3 and 4 satisfy $\sup_{r>1} \sqrt{r}/d(r-1) < \infty$ and $\sup_{r>1} \sup_{s \in (1,r)} \sqrt{r}/d(r,s) < \infty$.

Under this assumption we show the following theorem:

THEOREM 3. *Let $\beta_t = \beta_0$ for all t, and let Assumptions 1–6 hold true. Then, as $T \to \infty$,*

$$
\mathcal{Q}_{T,\infty} \xrightarrow{d} \sup_{r \in (0,1)} \frac{\|B^{(k)}(r)\|}{(1-r)d(\frac{r}{1-r})},
$$

$$
\mathcal{SBQ}_{T,\infty} \xrightarrow{d} \sup_{r \in (0,1)} \sup_{s \in (0,r)} \frac{\|(1-s)B^{(k)}(r) - (1-r)B^{(k)}(s)\|}{(1-r)(1-s)d(\frac{1}{1-r}, \frac{1}{1-s})},
$$

where $B^{(k)}(r)$ is a k-dimensional standard Brownian bridge.

The linear boundary $d_{\text{lin}}(r)$ (see equation (2)) satisfies Assumption 6, whereas for the linear triangular boundary $d_{\text{sbq}}(r,s)$ (see equation (7)) it is not satisfied. Instead, the boundary must be at least of order $\sqrt{r}$ uniformly among all s, which motivates the alternative boundary

$$
d_{\text{inf}}(r,s) = \sqrt{r}(1 + 2(r-s)), \quad 0 \leq s \leq r.
$$

Simulated critical values for the *M-tests* under the boundaries $d_{\text{lin}}(r)$ and $d_{\text{inf}}(r,s)$ are presented in Table 3.

6. ESTIMATION OF THE BREAKPOINT LOCATION

As soon as the testing procedure has indicated a structural instability in the coefficient vector, the next step is to locate the break point. In the single break

TABLE 3. Asymptotic critical values for infinite horizon M-tests.

	Stacked backward CUSUM					Forward CUSUM				
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\alpha = 10\%$	0.911	0.974	1.010	1.035	1.054	0.864	0.956	1.006	1.040	1.066
$\alpha = 5\%$	0.976	1.036	1.071	1.094	1.113	0.958	1.044	1.090	1.121	1.146
$\alpha = 1\%$	1.113	1.169	1.199	1.219	1.236	1.148	1.222	1.261	1.289	1.308

Note: Simulated $(1-\alpha)$ quantiles of the limiting distributions presented in Theorem 3 based on 100,000 Monte Carlo replications are reported. For $\mathcal{SBQ}_{T,\infty}$ we use boundary $d_{\inf}(r,s)$, and for $\mathcal{Q}_{T,\infty}$ the boundary $d_{\lin}(r)$ is implemented. The Brownian bridge is approximated on a grid of 50,000 equidistant points.

model with $\beta_t = \beta_0 + \delta 1_{\{t \geq T^*\}}$, where $\delta \neq 0$, Horváth (1995) suggested to estimate the relative break date $\tau^* = T^*/T$ by the relative time index for which the likelihood ratio statistic is maximized. As an asymptotically equivalent estimator, Bai (1997) proposed the maximum likelihood estimator

$$\widehat{\tau}_{ML}^{\mathrm{ret}} = T^{-1} \cdot \underset{t=1,\ldots,T}{\operatorname{argmin}} \{R_1(t) + R_2(t)\}, \tag{11}$$

where $R_1(t)$ is the OLS residual sum of squares when using observations until time point t and $R_2(t)$ is the OLS residual sum of squares when using observations from time $t+1$ onwards. In case of monitoring, Chu et al. (1996) considered

$$\widehat{\tau}_{ML}^{\mathrm{mon}} = T^{-1} \cdot \underset{t=T+1,\ldots,T_d}{\operatorname{argmin}} \{R_1(t) + R_2(t)\} \tag{12}$$

to estimate $\tau_{\mathrm{mon}}^* = T^*/T_d$, where T_d denotes the detection time point, which is the stopping time at which the detector statistic exceeds the boundary function for the first time. The maximum likelihood estimator is very accurate if the breakpoint is located in the middle of the sample. However, by construction, the true breakpoint T^* tends to be close to the stopping time T_d, and $R_2(T^*)$ is computed from very few observations, which may lead to a large finite sample estimation error for the maximum likelihood estimator. A theoretical explanation for this effect is given in Casini and Perron (2021a, 2021b), where the finite-sample distribution of the least squares estimator is investigated using a continuous record asymptotic framework.

To bypass this problem, we use backwardly cumulated recursive residuals to estimate the relative break location. In the single break model, $\|BQ_{\lfloor rT \rfloor,T}\|$ is asymptotically proportional to $\|h(1) - h(r)\|$, which is constant in the prebreak period and decreases to zero in the postbreak period. When scaled by its asymptotic standard deviation, the detector is asymptotically proportional to $\|h(1) - h(r)\|/\sqrt{1-r}$, which in turn (see equation (8)) is proportional to

$$\left(-\ln(\tau^*)1_{\{r<\tau^*\}} - \ln(r)1_{\{r \geq \tau^*\}}\right)/\sqrt{1-r},$$

where the maximum is attained at $r = \tau^*$. Accordingly, we consider

$$\widehat{\tau}_{\text{ret}} = \frac{1}{T} \cdot \operatorname*{argmax}_{t=1,\dots,T} \frac{\|BQ_{t,T}\|}{\sqrt{T-t+1}}, \qquad \widehat{\tau}_{\text{mon}} = \frac{1}{T} \cdot \operatorname*{argmax}_{t=T+1,\dots,T_d} \frac{\|BQ_{t,T_d}\|}{\sqrt{T_d+t-1}}. \tag{13}$$

THEOREM 4. *Let $\beta_t = \beta_0 + \delta 1_{\{t/T \geq \tau^*\}}$, where $\delta \neq 0$, and let Assumption 1 hold true. If $\tau^* \in (0,1]$, then $\widehat{\tau}_{\text{ret}} \overset{p}{\longrightarrow} \tau^*$, as $T \to \infty$; if $\tau^* \in (1, T_d/T]$, then $\widehat{\tau}_{\text{mon}} \overset{p}{\longrightarrow} \tau^*$, as $T \to \infty$.*

This result implies that the breakpoint estimators (13) are consistent, as $T \to \infty$.

7. FINITE SAMPLE PERFORMANCE

We illustrate the finite sample performance of the *R-tests* and *M-tests* for the models

$$y_t = \gamma_t + u_t, \tag{model I}$$
$$y_t = 1 + \gamma_t z_t + u_t, \tag{model II}$$
$$y_t = \gamma_t + 0.5 y_{t-1} + u_t, \quad (t = 1, \dots, T), \tag{model III}$$

where $\gamma_t = 0.8 \cdot 1_{\{t/T \geq \tau^*\}}$, u_t and e_t are independent and $NID(0,1)$, and $z_t = (1 - 0.5L)e_t$, where L is the lag operator. For model I and model II we consider the full structural break tests, and for model III partial break tests with $H = (1,0)$ are considered (see Remark 2).

7.1. Retrospective Tests (R-Tests)

In Table 4, the empirical sizes and powers of the retrospective tests are compared with that of the sup-Wald test of Andrews (1993). First, we observe that $\mathcal{BQ}_T$ and $\mathcal{SBQ}_T$ outperform $\mathcal{Q}_T$, except for the case $\tau^* = 0.1$. Second, while $\mathcal{Q}_T$ has much lower power than the sup-Wald test, the reversed order cumulation structure in $\mathcal{BQ}_T$ and $\mathcal{SBQ}_T$ seems to compensate for the weakness of $\mathcal{Q}_T$. Andrews (1993) showed that the sup-Wald test is weakly optimal in the sense that, in the case of a single structural break, its asymptotic local power curve approaches the power curve from the infeasible point optimal maximum likelihood test, as the significance level tends to zero. Within the framework of the considered models, $\mathcal{BQ}_T$ performs similarly well as the sup-Wald test and thus has comparably good power properties as the weakly optimal test. In contrast to $\mathcal{SBQ}_T$, the the sup-Wald test is not suitable for monitoring since its statistic is not measurable with respect to the filtration of information at time t.

7.2. Monitoring Procedures (M-Tests)

Fixed endpoint *M-tests* are particularly useful when the monitoring period is short. *M-tests* with infinite horizon can be used for long monitoring periods of arbitrary length. To evaluate the performance of the *M*-tests for finite samples, we simulate

TABLE 4. Empirical sizes and powers of the R-tests.

	Model I				Model II				Model III			
	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	$\mathcal{Q}_T$	supW	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	$\mathcal{Q}_T$	supW	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	$\mathcal{Q}_T$	supW
Size	4.3	3.4	4.2	4.6	4.4	4.1	4.1	4.8	4.9	4.2	5.3	5.8
τ^*						Power						
0.1	54.4	67.4	73.6	51.8	51.2	63.0	68.2	53.0	24.3	26.5	40.5	36.1
0.4	99.8	99.5	92.3	99.8	99.9	99.7	93.2	99.9	98.2	94.6	66.9	98.7
0.6	99.9	99.4	72.5	99.8	99.8	99.3	72.9	99.9	99.4	96.8	47.0	98.8
0.9	56.4	25.2	5.8	46.9	52.3	33.2	6.0	49.8	49.8	19.9	6.6	34.7

Note: Rejection rates of the retrospective tests are reported for a significance level of 5% and a sample size of $T = 200$. The results are obtained from 100,000 Monte Carlo repetitions under the linear boundaries $d_{\mathrm{lin}}(r)$ and $d_{\mathrm{sbq}}(r,s)$. The sup-Wald test by Andrews (1993) with trimming parameter 0.15 is denoted as supW.

TABLE 5. Empirical sizes and mean detection delays of the fixed endpoint M-tests.

	Model I		Model II		Model III	
	$\mathcal{SBQ}$	$\mathcal{Q}$	$\mathcal{SBQ}$	$\mathcal{Q}$	$\mathcal{SBQ}$	$\mathcal{Q}$
Size	4.0	4.6	6.4	5.6	4.4	5.0
τ^*			Absolute mean detection delay			
1.1	27.4	27.3	26.9	27.1	29.1	28.8
1.2	26.6	31.9	26.2	31.1	27.9	34.6
1.3	26.1	36.5	25.9	35.3	27.2	40.2
1.4	25.9	40.9	25.7	39.3	26.8	45.1
1.5	25.8	44.7	25.4	42.2	26.4	47.7

Note: The empirical sizes are reported in percentage points. Starting in row 4, the empirical mean detection delays are presented for different breakpoints τ^*. Results are obtained for a significance level of 5%, a pre-monitoring sample size of $T = 200$, a monitoring period of $m = 2$, boundaries $d_{\mathrm{lin}}(r)$ and $d_{\mathrm{sbq}}(r,s)$, and from 100,000 Monte Carlo repetitions. Critical values from Tables 1 and 2 are implemented.

the same models as for the R-tests for time points $t = 1, \ldots, mT$, where we specify $m = 2$ for the fixed endpoint tests and $m = 20$ for the infinite horizon tests. The results in Table 5 show that the mean delay for $\mathcal{SBQ}_{T,m}$ is much lower than that of $\mathcal{Q}_{T,m}$ and is almost constant across the breakpoint locations.

For infinite horizon monitoring, $\mathcal{SBQ}_{T,\infty}$ performs similarly well compared to conventional tests (see Table 6). The detection delay of $\mathcal{Q}_{T,\infty}$ is much higher than that of $\mathcal{SBQ}_{T,\infty}$, and the gap increases further with increasing τ^*. Compared to the tests of Chu et al. (1996) and Fremdt (2015), we find a similar picture. Note that the two alternative tests have no power in model II and are therefore omitted for this case.

TABLE 6. Empirical sizes and mean detection delays of the infinite horizon M-tests.

	Model I				Model II		Model III			
	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	FR	$\mathcal{SBQ}$	$\mathcal{Q}$	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	FR
Size	3.8	4.8	2.2	4.3	6.4	5.8	6.3	5.2	2.2	6.1
τ^*				Absolute mean detection delay						
1.5	27.6	46.4	70.2	53.6	27.3	44.2	29.7	52.8	84.1	51.8
2	33.8	69.5	80.3	76.7	32.9	65.6	36.7	82.2	94.3	74.8
4	57.2	162.2	111.9	169.8	52.9	150.8	64.4	200.9	126.3	167.7
6	81.1	254.9	136.9	263.6	72.3	235.4	93.9	319.6	151.3	261.6

Note: The empirical sizes are reported in percentage points. Starting in row 4, the empirical mean detection delays are presented for different breakpoints τ^*. Results are obtained for a significance level of 5%, a pre-monitoring sample size of $T = 200$, a monitoring period of $m = 20$, boundaries $d_{\text{lin}}(r)$ and $d_{\text{inf}}(r,s)$, and from 100,000 Monte Carlo repetitions. Critical values from Table 3 are implemented. The univariate infinite horizon M-tests by Chu et al. (1996) and Fremdt (2015) with boundary parameter 0.25 are denoted as CSW and FR, respectively.

TABLE 7. Bias and RMSE of breakpoint estimators.

	$T = 100$				$T = 200$			
	Bias		RMSE		Bias		RMSE	
τ^*	BQ	ML	BQ	ML	BQ	ML	BQ	ML
0.5	−0.03	0.01	0.14	0.11	−0.02	0.01	0.08	0.05
0.65	−0.03	0.00	0.14	0.12	−0.02	0.00	0.08	0.05
0.8	−0.03	−0.04	0.15	0.18	−0.01	−0.01	0.09	0.08
0.85	−0.04	−0.07	0.17	0.22	−0.02	−0.02	0.10	0.11
0.9	−0.06	−0.13	0.19	0.30	−0.03	−0.04	0.12	0.17
0.95	−0.10	−0.25	0.24	0.43	−0.05	−0.14	0.17	0.32
0.97	−0.13	−0.33	0.28	0.50	−0.08	−0.24	0.22	0.42
0.99	−0.20	−0.44	0.35	0.58	−0.15	−0.40	0.30	0.56

Note: The bias and root mean squared error (RMSE) for the break date estimators (11) and (13) are reported based on 100,000 Monte Carlo repetitions, where model (model I) is simulated for $t = 1, \ldots, T$. BQ denotes the backward CUSUM estimator (13), and ML denotes the maximum likelihood estimator (11).

7.3. Breakpoint Estimators

To compare the breakpoint estimator in equation (13) with its maximum likelihood benchmark in (11) and (12), we present Monte Carlo simulation results in Table 7. If the break τ^* is located after 85% of the sample, the estimator based on backwardly cumulated recursive residuals has a much lower bias and root mean
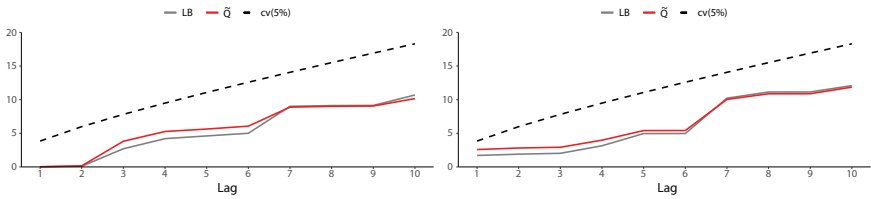
FIGURE 5. Ljung-Box and robust Q statistics of the residuals.

Note: The cumulative Ljung-Box (LB) and robust Q-statistics ($\widetilde{Q}$) of Dalla, Giraitis, and Phillips (2020) are plotted for the residuals of model (14). The left plot shows the statistics using the sample from the first premonitoring training period (April 10 until May 21) and the right plot for the second premonitoring training period (July 20 until August 30). The dashed line indicates the 5% critical values. The plots are created using the R-package *testcorr* provided by Dalla et al. (2020).

squared error than the maximum likelihood estimator, which is due to the fact that the postbreak entails only few observations.

8. EMPIRICAL APPLICATION TO COVID-19 INFECTIONS

We consider the time series y_t of daily new cases of SARS-CoV-2 infections in the United States during the first months of the COVID-19 pandemic. The starts of the first three waves of infections are observed in March, June, and September 2020. To monitor for a second and third waves of infections, we consider premonitoring training samples of 6 weeks ($T = 46$) starting at the times after the first and second peak. The training sample periods are given by April 10 until May 21, and July 20 until August 30. To account for a seasonal unit root in y_t we consider the seasonally differenced series $\widetilde{y}_t = y_t - y_{t-7}$. Weekly differences are used to account for the seasonalities resulting from the weekly reporting pattern of COVID-19 data with lower numbers on weekends. We estimate the dynamic model

$$\widetilde{y}_t = \phi_0 + \phi_1 \widetilde{y}_{t-2} + \phi_2 \widetilde{y}_{t-7} + u_t = x_t' \beta + u_t, \tag{14}$$

where $x_t = (1, \widetilde{y}_{t-1}, \widetilde{y}_{t-7})'$, and $\beta = (\phi_0, \phi_1, \phi_2)'$. The parameters for lags 2 and 7 are the only significant autoregressive parameters.

Both the Ljung-Box and the robust Q-statistic of Dalla et al. (2020) do not indicate any significant autocorrelation in the residuals for the pre-break training periods (see Figure 5). We consider the infinite-horizon stacked backward CUSUM statistic for a break in the intercept and the infinite horizon forward CUSUM of Chu et al. (1996). We are interested in detecting positive changes in the intercept ϕ_0 and apply one-sided infinite horizon monitoring statistics with a significance level of 5%. The critical values for the partial right-sided tests are given by those of the full test with $\alpha = 0.1$ and $k = 1$ (see Remark 2). We consider the infinite-horizon stacked backward CUSUM statistic for a break in the intercept and the infinite horizon forward CUSUM of Chu et al. (1996). To compare the detector statistics,
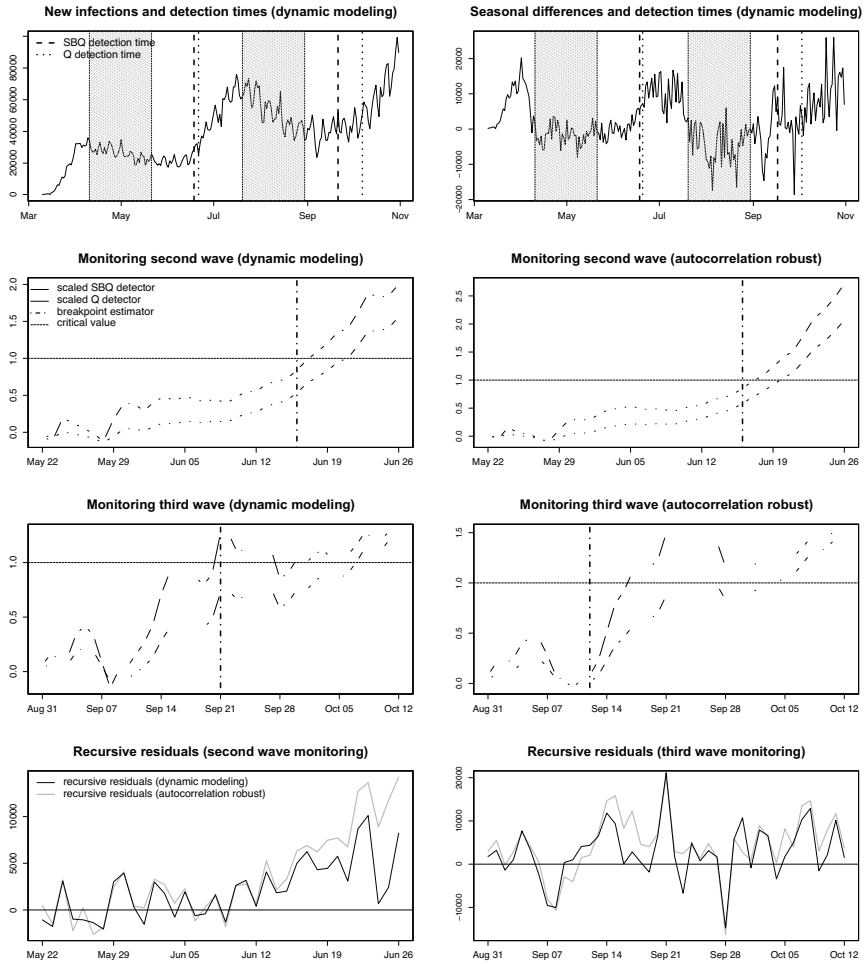
FIGURE 6. Monitoring daily new COVID-19 infections in the US.

Note: The results of the *M-tests* for the dynamic model (left column) and the autocorrelation robust *M-tests* (right column) are presented. The plots show the new infections (first row) and the seasonally differenced infections (second row). The third and fourth row present the standardized recursive residuals (solid), the scaled stacked backward CUSUM detector (dashed), and the scaled forward CUSUM detector (dotted). The shaded areas represent the pre-monitoring training periods, the dashed vertical lines represent the detection time point of the stacked backward CUSUM, the dotted vertical lines are the detection times of the forward CUSUM, and the vertical dash-dotted lines are the backward CUSUM breakpoint estimators given by equation (13). The horizontal solid line represents the critical boundary of the scaled detectors.

we scale them by their boundaries and critical values, so that H_0 is rejected in favor of a positive change in ϕ_0 if the detector exceeds unity, respectively. An alternative

to the dynamic modeling in (14) is to apply autocorrelation robust *M-tests* to the model $\widetilde{y}_t = \phi_0 + u_t$, where we replace $\widehat{\sigma}_T$ with the long-run variance estimator by Newey and West (1987) (see Remark 1).

The results are presented in Figure 6. Both monitoring procedures find an indication for a rise in SARS-CoV-2 infections in the US at the end of June and the end of September. The stacked backward CUSUM detects the breaks much earlier and becomes significant between 2 and 16 days before the forward CUSUM becomes significant. This confirms our theoretical analysis and shows that precious time can be saved by applying the backward monitoring scheme.

9. CONCLUSION

In this paper, we propose two alternatives to the conventional CUSUM detectors by Brown et al. (1975) and Chu et al. (1996). It has been demonstrated that cumulating the recursive residuals backwardly results in much higher power than using forwardly cumulated recursive residuals, in particular if the break is located at the end of the sample. Accordingly, the backward scheme is especially attractive for online monitoring. To this end, the stacked triangular array of backwardly cumulated recursive residuals is employed and we find that this approach yields a much lower detection delay than the monitoring procedure by Chu et al. (1996). Due to the multivariate nature of our tests, they also have power against structural breaks that do not affect the unconditional mean of the dependent variable. We also propose a new break date estimator which outperforms conventional estimators if the break is located at the end of the sample.

APPENDIX

A. Technical Proofs

A.1. *Auxiliary Lemmas.*   We first present some auxiliary lemmas which we require for the main proofs.

LEMMA 1. *Let Assumptions 1 and 2 hold true. Then, for any fixed $m < \infty$, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} x_t u_t \Rightarrow \sigma C^{1/2} W^{(k)}(r), \quad r \in [0, m].$$

Proof. Note that $M^{-1/2} \sum_{t=1}^{\lfloor sM \rfloor} x_t u_t \Rightarrow \Omega^{1/2} W^{(k)}(s), s \in [0,1]$, as $M \to \infty$, by Theorem 7.19 in White (2001). Then, on the space $D([0,m])^k$, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} x_t u_t = \frac{\sqrt{m}}{\sqrt{M}} \sum_{t=1}^{\lfloor (r/m)M \rfloor} x_t u_t \Rightarrow \sqrt{m} \Omega^{1/2} W^{(k)}(r/m) \stackrel{d}{=} \Omega^{1/2} W^{(k)}(r), \quad r \in [0, m]. \quad \textbf{(15)}$$

$\square$

Assumption 7. There exists $\kappa \geq 8$ such that $\sup_t E[\|x_t\|^\kappa] < \infty$, $\sup_t E[|u_t|^\kappa] < \infty$, and $(x_t, u_t)'$ is strong mixing of size $-\kappa/(\kappa - 6)$.

LEMMA 2. *Let Assumptions 1 and 7 hold true. Then, for any fixed $m < \infty$, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} x_t u_t \Rightarrow \Omega^{1/2} W^{(k)}(r), \quad r \in [0, m].$$

Proof. We have $M^{-1/2} \sum_{t=1}^{\lfloor sM \rfloor} x_t u_t \Rightarrow \Omega^{1/2} W^{(k)}(s)$, $s \in [0, 1]$, as $M \to \infty$, by Corollary 4.2 in Wooldridge and White (1988). The result follows by analogous arguments as in (15). $\qquad \square$

LEMMA 3. *Let $\beta_t = \beta_0$ for all $t \in \mathbb{N}$, and let Assumption 1 hold true. Moreover, let either Assumption 2 or 7 hold true. Then, for any $m < \infty$, as $T \to \infty$,*

$$\max_{1 \leq t \leq mT} \frac{1}{\sqrt{T}} \left\| Y_t - \sum_{j=1}^{t-1} \frac{1}{j} Y_j - X_t \right\| = o_P(1),$$

where $X_t = \sum_{j=1}^t x_j w_j$ and $Y_t = \sum_{j=1}^t x_j u_j$.

Proof. Let $f_j = (1 + (j-1)^{-1} x_j' \widehat{C}_{j-1}^{-1} x_j)^{-1/2} 1_{\{j > k\}}$. Since $\widehat{\beta}_j = \beta_0 + j^{-1} \widehat{C}_j^{-1} Y_j$, we can represent the recursive residuals as

$$w_j = f_j(y_j - x_j' \widehat{\beta}_{j-1}) = f_j(u_j - (j-1)^{-1} x_j' \widehat{C}_{j-1}^{-1} Y_{j-1}).$$

The multivariate cumulative sum can be written as

$$X_t = \sum_{j=1}^t x_j u_j f_j - \sum_{j=1}^{t-1} \frac{f_{j+1}}{j} x_{j+1} x_{j+1}' \widehat{C}_j^{-1} Y_j,$$

and we have

$$T^{-1/2} \left(Y_t - \sum_{j=1}^{t-1} \frac{1}{j} Y_j - X_t \right) = T^{-1/2}(A_t + B_t),$$

where

$$A_t = \sum_{j=1}^t x_j u_j (1 - f_j), \quad B_t = \sum_{j=1}^{t-1} (f_{j+1} x_{j+1} x_{j+1}' \widehat{C}_j^{-1} - I_k) \frac{1}{j} Y_j. \tag{16}$$

It remains to show that $\max_{1 \leq t \leq mT} T^{-1/2} \|A_t\| = o_P(1)$ and $\max_{1 \leq t \leq mT} T^{-1/2} \|B_t\| = o_P(1)$. For the first part, note that, for $j > k$,

$$0 \leq \sqrt{j-1}(1 - f_j) = \frac{\sqrt{j-1}\sqrt{1 + \frac{1}{j-1} x_j' \widehat{C}_{j-1}^{-1} x_j} - 1}{\sqrt{1 + \frac{1}{j-1} x_j' \widehat{C}_{j-1}^{-1} x_j}} \leq \sqrt{x_j' \widehat{C}_{j-1}^{-1} x_j}, \tag{17}$$

which implies that $\sqrt{j}(1-f_j) = O_P(1)$, as $j \to \infty$. Since $\widehat{C}_j$ is uniformly positive definite, there exists a uniformly minimal eigenvalue $\lambda_{min} > 0$, which is defined as the infimum of all eigenvalues of the matrices $\{\widehat{C}_j\}_{j>k}$. Any Rayleigh quotient of C_{j-1}^{-1} is bounded above by $\lambda_{min}^{-1} < \infty$. Therefore, for any j and $\delta > 0$,

$$E[|\sqrt{j}(1-f_j)|^\delta] \leq E[|x_j'\widehat{C}_{j-1}^{-1}x_j|^{\delta/2}] \leq \lambda_{min}^{-\delta/2} E[|x_j'x_j|^{\delta/2}] \leq \lambda_{min}^{-\delta/2} E[\|x_j\|^\delta].$$

Hence, by the Hölder inequality,

$$E\big[\|x_j u_j \sqrt{j}(1-f_j)\|^{2+\epsilon}\big] \leq \big(E[\|x_j\|^{6+3\epsilon}]E[|u_j|^{6+3\epsilon}]E[|\sqrt{j}(1-f_j)|^{6+3\epsilon}]\big)^{1/3} < \infty, \quad \textbf{(18)}$$

for any $0 < \epsilon \leq (\kappa - 6)/3$. Then,

$$Var\Big[\max_{1 \leq t \leq mT} T^{-1/2}\|A_t\|\Big] \leq \frac{1}{T} Var\Big[\sum_{j=1}^{\lfloor mT \rfloor} \|x_j u_j (1-f_j)\|\Big] = A_{T,1}^* + A_{T,2}^*, \quad \textbf{(19)}$$

where

$$A_{T,1}^* = \frac{1}{T}\sum_{j=1}^{\lfloor mT \rfloor} Var\big[\|x_j u_j (1-f_j)\|\big] \leq \frac{1}{T}\sum_{j=1}^{\lfloor mT \rfloor} \frac{1}{j}E\big[\|x_j u_j \sqrt{j}(1-f_j)\|^2\big], \quad \textbf{(20)}$$

which is $O(\ln(mT)/T)$ since the harmonic series satisfies $\sum_{j=1}^{\lfloor mT \rfloor} j^{-1} = O(\ln(mT))$, and

$$A_{T,2}^* = \frac{2}{T}\sum_{\tau=1}^{\lfloor mT \rfloor - 1}\sum_{j=1}^{\lfloor mT \rfloor - \tau} Cov\big[\|x_j u_j (1-f_j)\|, \|x_{j+\tau} u_{j+\tau}(1-f_{j+\tau})\|\big],$$

which is zero under Assumption 2. Under Assumption 7, we apply Corollary 14.3 of Davidson (1994). Let $\alpha(\tau)$ be the α-mixing sequence of (x_t', u_t), and consider $0 < \epsilon \leq (\kappa - 6)/3$. Then, by (18), there exists a constant $K < \infty$, such that

$$A_{T,2}^* \leq \frac{12}{T}\sum_{\tau=1}^{\lfloor mT \rfloor - 1}\sum_{j=1}^{\lfloor mT \rfloor - \tau} \big(E[\|x_j u_j (1-f_j)\|^{2+\epsilon}]E[\|x_{j+\tau} u_{j+\tau}(1-f_{j+\tau})\|^{2+\epsilon}]\big)^{1/(2+\epsilon)}$$

$$\alpha(\tau)^{1-\frac{2}{2+\epsilon}}$$

$$\leq \frac{12K}{T}\sum_{\tau=1}^{\infty}\alpha(\tau)^{1-\frac{2}{2+\epsilon}}\sum_{j=1}^{\lfloor mT \rfloor}\frac{1}{j} = O(\ln(mT)/T) = o(1), \quad \textbf{(21)}$$

since $1 - 2/(2+\epsilon) \leq (\kappa - 6)/\kappa$. Consequently, $\max_{1 \leq t \leq mT}\|A_{t,T}\| = o_P(1)$ by Chebyshev's inequality. For the second term, we consider the decomposition $B_t = B_{t,1} + B_{t,2}$, where

$$B_{t,1} = \sum_{j=1}^{t-1}\frac{1}{j}x_{j+1}x_{j+1}'(f_{j+1}\widehat{C}_j^{-1} - C^{-1})Y_j, \quad B_{t,2} = \sum_{j=1}^{t-1}\frac{1}{j}(x_{j+1}x_{j+1}'C^{-1} - I_k)Y_j.$$

Note that, by (17), $\sqrt{j}(f_{j+1}^{-1}\widehat{C}_j - \widehat{C}_j) = O_P(1)$, as $j \to \infty$. Thus, by Lemmas 1 and 2, for any $s > 0$, $\text{plim}_{T \to \infty} x_{\lfloor sT \rfloor + 1} x_{\lfloor sT \rfloor + 1}' (f_{\lfloor sT \rfloor + 1}\widehat{C}_{\lfloor sT \rfloor}^{-1} - C^{-1})T^{-1/2}Y_{\lfloor sT \rfloor} = 0$. Consequently,

as $T \to \infty$,

$$
\max_{1 \le t \le mT} T^{-1/2} \|B_{t,1}\| = \sup_{r \in [0,m]} \int_0^r x_{\lfloor sT \rfloor + 1} x'_{\lfloor sT \rfloor + 1} (f_{\lfloor sT \rfloor + 1} \widehat{C}_{\lfloor sT \rfloor}^{-1} - C^{-1}) T^{-1/2} Y_{\lfloor sT \rfloor} \, ds
$$
$$
+ o_P(1),
$$

which is $o_P(1)$ by the continuous mapping theorem. For $B_{t,2}$, we apply Abel's formula of summation by parts, which is given by

$$
\sum_{j=1}^t A_j b_j = \sum_{j=1}^t A_j b_t + \sum_{j=1}^{t-1} \sum_{i=1}^j A_i (b_j - b_{j+1}), \qquad A_j \in \mathbb{R}^{k \times k}, \quad b_j \in \mathbb{R}^k, \quad t \in \mathbb{N}. \tag{22}
$$

By setting $A_j = x_{j+1} x'_{j+1} C^{-1} - I_k$ and $b_j = j^{-1} T^{-1/2} Y_j$ we get

$$
T^{-1/2} B_{t,2} = \sum_{j=1}^{t-1} A_j b_{t-1} + \sum_{j=1}^{t-2} \sum_{i=1}^j A_i (b_j - b_{j+1}) = B^*_{t,T,1} + B^*_{t,T,2} + B^*_{t,T,3},
$$

where

$$
B^*_{t,T,1} = (\widetilde{C}_t C^{-1} - I_k) T^{-1/2} Y_{t-1}, \quad B^*_{t,T,2} = \sum_{j=1}^{t-2} \frac{j-1}{j(j+1)} (\widetilde{C}_j C^{-1} - I_k) T^{-1/2} Y_j,
$$

$B^*_{t,T,3} = -\sum_{j=1}^{t-2} \frac{j-1}{j+1} (\widetilde{C}_j C^{-1} - I_k) T^{-1/2} x_{j+1} u_{j+1}$, and $\widetilde{C}_t = (t-1)^{-1} \sum_{j=2}^t x_j x'_j$. Lemmas 1 and 2, the continuous mapping theorem, and the fact that $\operatorname{plim}_{t \to \infty} \widetilde{C}_t = C$ imply

$$
\max_{1 \le t \le mT} \|B^*_{t,T,1}\| = \sup_{r \in [0,m]} \|(\widetilde{C}_{\lfloor rT \rfloor} C^{-1} - I_k) T^{-1/2} Y_{\lfloor rT \rfloor}\| = o_P(1),
$$

and

$$
\max_{1 \le t \le mT} \|B^*_{t,T,2}\| = \sup_{r \in [0,m]} \int_0^r \|(\widetilde{C}_{\lfloor rT \rfloor} C^{-1} - I_k) T^{-1/2} Y_{\lfloor rT \rfloor}\| \, dr = o_P(1).
$$

For the last term, note that $(\widetilde{C}_j C^{-1} - I_k) T^{-1/2} x_{j+1} u_{j+1}$ is either a martingale difference sequence or strong mixing. Then, by the fact that we have bounded eighth moments, and by Theorems 24.3 and 24.6 in Davidson (1994), we have $\max_{1 \le t \le mT} \|B^*_{t,T,3}\| = o_P(1)$, and the assertion is shown. □

LEMMA 4. *Let Assumptions 1 and 5 hold true, and let $\beta_t = \beta_0$ for all $t \in \mathbb{N}$. Moreover, let either Assumption 2 or 7 hold true. Then, as $T \to \infty$,*

$$
\sup_{t > T} \frac{1}{\sqrt{t}} \left\| Y_t - \sum_{j=1}^{t-1} \frac{1}{j} Y_j - X_t \right\| = o_P(1),
$$

where $X_t = \sum_{j=1}^t x_j w_j$ and $Y_t = \sum_{j=1}^t x_j u_j$.

Proof. Analogously to the proof of Lemma 3, it remains to show that

$$\sup_{t>T} \frac{1}{\sqrt{t}} \left\| Y_t - \sum_{j=1}^{t-1} \frac{1}{j} Y_j - X_t \right\| = \sup_{t>T} \frac{1}{\sqrt{t}} \|A_t + B_t\| = o_P(1),$$

where A_t and B_t are defined in (16), which satisfy $\sup_{t \leq T} T^{-1/2} \|A_t + B_t\|$. Thus, for any $n \in \mathbb{N}$,

$$\sup_{T<t\leq nT} t^{-1/2} \|A_t + B_t\| \leq \sup_{T<t\leq nT} T^{-1/2} \|A_t + B_t\| \leq \sup_{t\leq nT} T^{-1/2} \|A_t + B_t\| = o_P(1),$$

$$\plim_{t\to\infty} t^{-1/2} \|A_t + B_t\| \leq \plim_{t\to\infty} \sup_{l\leq t} t^{-1/2} \|A_l + B_l\| = 0,$$

and the assertion follows. $\qquad\square$

LEMMA 5. *Let Assumption 1 hold true, let $\beta_t = \beta_0$ for all t, and let $m < \infty$. Under Assumption 2,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT\rfloor} x_t w_t \Rightarrow \sigma C^{1/2} W^{(k)}(r), \quad r \in [0,m],$$

and, under Assumption 7,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT\rfloor} x_t w_t \Rightarrow \Omega^{1/2} W^{(k)}(r), \quad r \in [0,m],$$

where $W^{(k)}(r)$ is a k-dimensional standard Brownian motion.

Proof. Lemma 3 implies that

$$\sup_{r\in[0,m]} \frac{1}{\sqrt{T}} \left\| Y_{\lfloor rT\rfloor} - \int_0^r z^{-1} Y_{\lfloor zT\rfloor} \, \mathrm{d}z - X_{\lfloor rT\rfloor} \right\| = o_P(1), \tag{23}$$

where $X_t = \sum_{j=1}^t x_j w_j$ and $Y_t = \sum_{j=1}^t x_j u_j$. Under Assumption 2, Lemma 1, the continuous mapping theorem, and (23) imply

$$\frac{1}{\sqrt{T}} X_{\lfloor rT\rfloor} \Rightarrow \sigma C^{1/2} \left(W^{(k)}(r) - \int_0^r z^{-1} W^{(k)}(z) \, \mathrm{d}z \right), \quad r \in [0,m].$$

Analogously, under Assumption 7,

$$\frac{1}{\sqrt{T}} X_{\lfloor rT\rfloor} \Rightarrow \Omega^{1/2} \left(W^{(k)}(r) - \int_0^r z^{-1} W^{(k)}(z) \, \mathrm{d}z \right), \quad r \in [0,m],$$

by Lemma 2. It remains to show that, for any $r \geq 0$,

$$W^{(k)}(r) - \int_0^r z^{-1} W^{(k)}(z) \, \mathrm{d}z \stackrel{d}{=} W^{(k)}(r). \tag{24}$$

Let $W_j(r)$ and $B_j(r)$ be the jth component of $W^{(k)}(r)$ and $B^{(k)}(r)$, respectively. We show the identities for each $j = 1, \ldots, k$, separately. Using Cauchy–Schwarz and Jensen's inequalities, we obtain $\int_0^r z^{-1} E[|W_j(z)|] \, \mathrm{d}z < \infty$ as well as $\int_0^r z^{-1} E[|W_j(r) W_j(z)|] \, \mathrm{d}z < \infty$, which

justifies the application of Fubini's theorem in the subsequent steps. Since both $W_j(r)$ and $F(W_j(r)) = W_j(r) - \int_0^r z^{-1} W_j(z)\,dz$ are Gaussian with zero mean, it remains to show that their covariance functions coincide. Let w.l.o.g. $r \leq s$. Then,

$$E[F(W_j(r))F(W_j(s))] - E[W_j(r)W_j(s)]$$

$$= \int_0^r \int_0^s \frac{E[W_j(z_1)W_j(z_2)]}{z_1 z_2}\,dz_2\,dz_1 - \int_0^s \frac{E[W_j(r)W_j(z_2)]}{z_2}\,dz_2 - \int_0^r \frac{E[W_j(s)W_j(z_1)]}{z_1}\,dz_1$$

$$= (2r + r\ln(s) - r\ln(r)) - (r + r\ln(s) - r\ln(r)) - r = 0,$$

and the assertion follows. $\qquad\square$

LEMMA 6. *Let h be a $\mathbb{R}^k$-valued function of bounded variation, and let $\{A_t\}_{t\in\mathbb{N}}$ be a sequence of random $(k \times k)$ matrices with $\sup_{r\in[0,m]} \|T^{-1} \sum_{t=1}^{\lfloor rT \rfloor}(A_t - A)\|_M = o_P(1)$, where $m < \infty$. Then, as $T \to \infty$,*

$$\sup_{r\in[0,m]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor}(A_t - A)h(\tfrac{t}{T}) \right\| = o_P(1).$$

Proof. By the application of Abel's formula of summation by parts, which is given in (22), it follows that

$$\sum_{t=1}^{\lfloor rT \rfloor}(A_t - A)h(\tfrac{t}{T}) = \sum_{t=1}^{\lfloor rT \rfloor}(A_t - A)h(\tfrac{\lfloor rT \rfloor}{T}) + \sum_{t=1}^{\lfloor rT \rfloor-1}\sum_{j=1}^{t}(A_j - A)(h(\tfrac{t}{T}) - h(\tfrac{t+1}{T})).$$

The fact that $h(r)$ is of bounded variation yields $\sup_{r\in[0,m]}\|h(r)\| = O(1)$ as well as $\sup_{r\in[0,m]}\|\sum_{t=1}^{\lfloor rT \rfloor-1}(h(\tfrac{t}{T}) - h(\tfrac{t+1}{T}))t/T\| = O(1)$. Consequently,

$$\sup_{r\in[0,m]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor}(A_t - A)h(\tfrac{\lfloor rT \rfloor}{T}) \right\| \leq \sup_{r\in[0,m]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor}(A_t - A) \right\|_M \left\| h(\tfrac{\lfloor rT \rfloor}{T}) \right\| = o_P(1)$$

and

$$\sup_{r\in[0,m]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor-1}\sum_{j=1}^{t}(A_j - A)(h(\tfrac{t}{T}) - h(\tfrac{t+1}{T})) \right\|$$

$$\leq \sup_{r\in[0,m]} \sum_{t=1}^{\lfloor rT \rfloor-1} \frac{t}{T} \left\| \frac{1}{t} \sum_{j=1}^{t}(A_j - A) \right\|_M \left\| h(\tfrac{t}{T}) - h(\tfrac{t+1}{T}) \right\| = o_P(1).$$

Then, by the triangle inequality, the assertion follows. $\qquad\square$

A.2. *Main Proofs.*

Proof of Theorem 1. Consider the auxiliary sequence $y_t^* = x_t'\beta_0 + u_t$, which coincide with y_t if $\beta_t = \beta_0$. Moreover, define $\widehat{\beta}_{t-1}^* = (\sum_{j=1}^{t-1} x_j x_j')^{-1} \sum_{j=1}^{t-1} x_j y_j^*$ and $f_t = (1 + (t-1)^{-1} x_t' \widehat{C}_{t-1}^{-1} x_t)^{-1/2} 1_{\{t>k\}}$. Then, $w_t^* = f_t(y_t^* - x_t'\widehat{\beta}_{t-1}^*)$ are recursive

residuals from a regression without any structural break in the coefficients. If $\beta_t = \beta_0 + T^{-1/2}g(t/T)$, we have $y_t = x_t'\beta_t + u_t = y_t^* + T^{-1/2}x_t'g(t/T)$, and $\widehat{\beta}_{t-1} = \widehat{\beta}_{t-1}^* + T^{-1/2}(t-1)^{-1}\widehat{C}_{t-1}^{-1}\sum_{j=1}^{t-1}x_jx_j'g(j/T)$, which implies that

$$w_t = w_t^* + f_t T^{-1/2}x_t'g(t/T) - f_t T^{-1/2}(t-1)^{-1}\widehat{C}_{t-1}^{-1}\sum_{j=1}^{t-1}x_jx_j'g(j/T).$$

We decompose $T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor}x_t w_t = \mathcal{S}_{1,T}(r) + \mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r)$, where

$$\mathcal{S}_{1,T}(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor}x_t w_t^*, \quad \mathcal{S}_{2,T}(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}f_t x_t x_t'g(\tfrac{t}{T}),$$

$$\mathcal{S}_{3,T}(r) = -\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}f_t(t-1)^{-1}x_t x_t'\widehat{C}_{t-1}^{-1}\sum_{j=1}^{t-1}x_jx_j'g(\tfrac{j}{T}).$$

Lemma 5 yields $\mathcal{S}_{1,T}(r) \Rightarrow \Omega^{1/2}W^{(k)}(r)$, where $\Omega^{1/2} = \sigma C^{1/2}$ under Assumption 2. Analogously to (19)–(21), we have

$$E\left[\sup_{r\in[0,m]}\left\|\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}(f_t-1)x_t x_t'\right\|_M^2\right] = o(1).$$

Moreover, $\sup_{r\in[0,m]}\|\widehat{C}_{\lfloor rT \rfloor} - C\|_M = o_P(1)$ due to Assumption 1(b), and, consequently,

$$\sup_{r\in[0,m]}\left\|\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}f_t x_t x_t' - C\right\|_M = o_P(1). \tag{25}$$

Since $g(r)$ is piecewise constant and therefore of bounded variation, Lemma 6 yields

$$\sup_{r\in[0,m]}\left\|\mathcal{S}_2(r) - \int_0^r Cg(s)\,ds\right\| = \sup_{r\in[0,m]}\left\|\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}(f_t x_t x_t' - C)g(\tfrac{t}{T})\right\| = o_P(1).$$

For the third term, let

$$p_{1,t} = \widehat{C}_t^{-1}\frac{1}{t}\sum_{j=1}^{t}x_jx_j'g(\tfrac{j}{T}), \quad p_{2,t} = \widehat{C}_t^{-1}\frac{1}{t}\sum_{j=1}^{t}Cg(\tfrac{j}{T}), \quad p_{3,t} = \frac{1}{t}\sum_{j=1}^{t}g(\tfrac{j}{T}).$$

From Assumption 1(b) and Lemma 6, $\|p_{1,T} - p_{3,T}\| \le \|p_{1,T} - p_{2,T}\| + \|p_{2,T} - p_{3,T}\| = o_P(1)$. Consequently,

$$\sup_{r\in[0,m]}\left\|\mathcal{S}_{3,T}(r) + \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}f_t x_t x_t' p_{3,t-1}\right\| \le \int_0^m \|f_{\lfloor sT \rfloor}x_{\lfloor sT \rfloor}x_{\lfloor sT \rfloor}'(p_{3,\lfloor sT \rfloor} - p_{1,\lfloor sT \rfloor})\,ds\|$$

$$= o_P(1).$$

Since $p_3(r) = p_{3,\lfloor rT \rfloor}$ is a partial sum of a piecewise constant function, it is of bounded variation, and, together with (25), Lemma 6 yields

$$\sup_{r\in[0,m]}\left\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(f_t^{-1}x_t x_t' - C)p_{3,t-1}\right\| = o_P(1),$$

which implies that $\sup_{r\in[0,m]}\|\mathcal{S}_{3,T}(r) + \frac{1}{T}C\sum_{t=1}^{\lfloor rT\rfloor}p_{3,t-1}\| = o_P(1)$, and

$$\sup_{r\in[0,m]}\left\|\mathcal{S}_{3,T}(r) + \int_0^r\int_0^s\frac{1}{s}Cg(v)\,dv\,ds\right\| = o_P(1).$$

Consequently, $\sup_{r\in[0,m]}\|\mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r) - \sigma Ch(t/T)\| = o_P(1)$, and, by Slutsky's theorem,

$$\mathcal{S}_{1,T}(r) + \mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r) \Rightarrow \sigma C^{1/2}W^{(k)}(r) + \sigma Ch(r). \tag{26}$$

Finally,

$$Q_T(r) = \widehat{\sigma}_T^{-1}C_T^{-1/2}(\mathcal{S}_{1,T}(r) + \mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r)) \Rightarrow W^{(k)}(r) + C^{1/2}h(r),$$

since $\widehat{\sigma}_T^2$ is consistent for σ^2 (see Krämer et al., 1988).

Note that, if we replace Assumption 2 by Assumption 7, the corresponding limiting result for the autocorrelation robust statistic $\widetilde{Q}_T(r)$ under H_0 follows analogously. Moreover, under the local alternatives we have

$$\mathcal{S}_{1,T}(r) + \mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r) \Rightarrow \Omega^{1/2}W^{(k)}(r) + \sigma Ch(r),$$

in equation (26) so that $\widetilde{Q}_T(r) \Rightarrow W^{(k)}(r) + \sigma\Omega^{-1/2}Ch(r)$. $\quad\square$

Proof of Theorem 2. By Assumption 5 there exists a k-dimensional standard Brownian motion $W^{(k)}(t)$ such that

$$\sup_{t>T}t^{-1/2}\|Y_t - \Omega^{1/2}W^{(k)}(t)\| = o_P(1),$$

and there exists a random variable ξ and some $\epsilon > 0$ such that $\|Y_t - \Omega^{1/2}W^{(k)}(t)\| \le \xi t^{1/2-\epsilon}$, which implies that

$$\sup_{t>T}t^{-1/2}\left\|\sum_{j=1}^{t-1}j^{-1}(Y_j - \Omega^{1/2}W^{(k)}(j))\right\| \le \sup_{t>T}t^{-\epsilon}\xi\sum_{j=1}^{t}j^{-1} = o_P(1).$$

Therefore,

$$\sup_{t>T}t^{-1/2}\left\|Y_t - \sum_{j=1}^{t-1}j^{-1}Y_j - \Omega^{1/2}\left(W^{(k)}(t) - \sum_{j=1}^{t-1}j^{-1}W^{(k)}(j)\right)\right\| = o_P(1).$$

By Lemma 4, from the fact that $T^{-1/2}W^{(k)}(t) \stackrel{d}{=} W^{(k)}(t/T)$, and from (24), it follows that there exists another k-dimensional standard Brownian motion $\widetilde{W}^{(k)}(t)$, such that

$$\sup_{r>1}r^{-1/2}\|T^{-1/2}X_{\lfloor rT\rfloor} - \Omega^{1/2}\widetilde{W}^{(k)}(r)\| = o_P(1),$$

and the assertion follows, since $\Omega = \sigma^2 C$ under Assumption 2, and since $\widehat{\sigma}_T^2$, $\widehat{C}_T$, and $\widehat{\Omega}_T$ are consistent estimators for their population counterparts.   □

Proof of Theorem 3. By the triangle inequality, we have

$$\|Q_T(r) - Q_T(s)\| - \|W^{(k)}(r) - W^{(k)}(s)\| \leq \|Q_T(r) - W^{(k)}(r)\| + \|Q_T(s) - W^{(k)}(s)\|$$

for any r and s. Thus,

$$\sup_{r>1} \frac{\|Q_T(r) - Q_T(1)\|}{d(r-1)} - \sup_{r>1} \frac{\|W^{(k)}(r) - W^{(k)}(1)\|}{d(r-1)}$$

$$\leq \sup_{r>1} \frac{\|Q_T(r) - W^{(k)}(r)\|}{d(r-1)} + \sup_{r>1} \frac{\|Q_T(1) - W^{(k)}(1)\|}{d(r-1)}$$

$$\leq \sup_{r>1} \left(\frac{\|Q_T(r) - W^{(k)}(r)\|}{\sqrt{r}} \frac{\sqrt{r}}{d(r-1)} \right) + \sup_{r>1} \left(\frac{\|Q_T(1) - W^{(k)}(1)\|}{\sqrt{r}} \frac{\sqrt{r}}{d(r-1)} \right)$$

$$\leq 2 \left(\sup_{r>1} \frac{\|Q_T(r) - W^{(k)}(r)\|}{\sqrt{r}} \right) \left(\sup_{r>1} \frac{\sqrt{r}}{d(r-1)} \right) = o_P(1)$$

for some k-dimensional standard Brownian motion $W^{(k)}(r)$, which implies that

$$\mathcal{Q}_{T,\infty} = \sup_{r \in (1,\infty)} \frac{\|Q_T(r) - Q_T(1)\|}{d(r-1)} \xrightarrow{d} \sup_{r \in (1,\infty)} \frac{\|W^{(k)}(r) - W^{(k)}(1)\|}{d(r-1)}.$$

We transform the expression into a supremum over the unit interval. A k-dimensional Brownian motion $W^{(k)}(r)$ and a k-dimensional Brownian Bridge $B^{(k)}(r)$ have the distributional relation $B^{(k)}(r) \overset{d}{=} (1-r)W^{(k)}(r/(1-r))$, which can be verified by comparing their covariance functions. Consider the bijective function $f : (0,1) \to (0,\infty)$ given by $f(\eta) = \eta/(1-\eta)$. The limiting distribution of $\mathcal{Q}_{T,m}$ can be rearranged as

$$\sup_{r \in (1,\infty)} \frac{\|W^{(k)}(r) - W^{(k)}(1)\|}{d(r-1)} \overset{d}{=} \sup_{r \in (0,\infty)} \frac{\|W^{(k)}(r)\|}{d(r)}$$

$$= \sup_{\eta \in (0,1)} \frac{\|W^{(k)}(f(\eta))\|}{d(f(\eta))} \overset{d}{=} \sup_{\eta \in (0,1)} \frac{\|B^{(k)}(\eta)\|}{(1-\eta)d\left(\frac{\eta}{1-\eta}\right)}.$$

For the second result, we analogously have

$$\sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|Q_T(r) - Q_T(s)\|}{d(r,s)} - \sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r,s)}$$

$$\leq \sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|Q_T(r) - W^{(k)}(r)\|}{d(r,s)} + \sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|Q_T(s) - W^{(k)}(s)\|}{d(r,s)}$$

$$\leq 2 \left(\sup_{r \in (1,\infty)} \frac{\|Q_T(r) - W^{(k)}(r)\|}{\sqrt{r}} \right) \left(\sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\sqrt{r}}{d(r,s)} \right) = o_P(1)$$

for some k-dimensional standard Brownian motion $W^{(k)}(r)$. Then,

$$\mathcal{SBQ}_{T,\infty} = \sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|Q_T(r) - Q_T(s)\|}{d(r,s)} \xrightarrow{d} \sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r,s)}.$$

Consider again the bijective function f from above, which analogously implies that

$$\sup_{r \in (1,\infty)} \sup_{s \in (1,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r,s)} \overset{d}{=} \sup_{r \in (0,\infty)} \sup_{s \in (0,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r+1,s+1)}$$

$$= \sup_{\eta \in (0,1)} \sup_{s \in (0,f(\eta))} \frac{\|W^{(k)}(f(\eta)) - W^{(k)}(s)\|}{d(f(\eta)+1, s+1)} = \sup_{\eta \in (0,1)} \sup_{\zeta \in (0,\eta)} \frac{\|W^{(k)}(f(\eta)) - W^{(k)}(f(\zeta))\|}{d(f(\eta)+1, f(\zeta)+1)}$$

$$\overset{d}{=} \sup_{\eta \in (0,1)} \sup_{\zeta \in (0,\eta)} \frac{\|B^{(k)}(\eta)/(1-\eta) - B^{(k)}(\zeta)/(1-\zeta)\|}{d((1-\eta)^{-1}, (1-\zeta)^{-1})}$$

$$= \sup_{\eta \in (0,1)} \sup_{\zeta \in (0,r)} \frac{\|(1-\zeta)B^{(k)}(\eta) - (1-\eta)B^{(k)}(\zeta)\|}{(1-\eta)(1-\zeta)d((1-\eta)^{-1}, (1-\zeta)^{-1})}. \qquad \square$$

Proof of Theorem 4. Adopting the notation of the local break in Theorem 1, we have $\beta_t + T^{-1/2}g(t/T)$ with $g(t/T) = T^{1/2}\delta 1_{\{t \geq T^*\}}$. Note that

$$\int_0^r 1_{\{z \geq \tau^*\}}\, dz - \int_0^r \int_0^z \frac{1}{z} 1_{\{v \geq \tau^*\}}\, dv\, dz = \int_0^r \frac{\tau^*}{z} 1_{\{z \geq \tau^*\}}\, dz = \tau^* \big(\ln(r) - \ln(\tau^*)\big) 1_{\{r \geq \tau^*\}}.$$

By equation (3) and Theorem 1,

$$Q_T(r) - T^{1/2}\sigma^{-1}\tau^* C^{1/2}\delta \big(\ln(r) - \ln(\tau^*)\big) 1_{\{r \geq \tau^*\}} \Rightarrow W^{(k)}(r), \quad r \in [0,m].$$

Then, by the continuous mapping theorem,

$$\widehat{\tau}_{\mathrm{ret}} = \frac{1}{T}\left(\underset{1 \leq t \leq T}{\mathrm{argmax}}\, \frac{\|Q_T(1) - Q_T(\frac{t+1}{T})\|}{\sqrt{T-t+1}}\right) = \underset{r \in (0,1)}{\mathrm{argsup}}\, \frac{\|Q_T(1) - Q_T(r)\|}{\sqrt{T}\sqrt{1-r}}$$

$$= \underset{r \in (0,1)}{\mathrm{argsup}}\, \frac{\big(\ln(1) - \ln(\tau^*)\big) 1_{\{1 \geq \tau^*\}} - \big(\ln(r) - \ln(\tau^*)\big) 1_{\{r \geq \tau^*\}}}{\sqrt{1-r}} + o_P(1)$$

$$= \tau^* + o_P(1),$$

since $-\ln(\tau^*)/\sqrt{1-r}$ is strictly increasing for $r \in (0,\tau^*)$ and $-\ln(r)/\sqrt{1-r}$ is strictly decreasing for $r \in [\tau^*, 1)$. Analogously, if $\tau^* \in (1, \tau_d]$,

$$\widehat{\tau}_{\mathrm{mon}} = \frac{1}{T}\left(\underset{T < t \leq T_d}{\mathrm{argmax}}\, \frac{\|Q_{T_d}(1) - Q_{T_d}(\frac{t+1}{T_d})\|}{\sqrt{T_d - t + 1}}\right) = \underset{r \in (0,1)}{\mathrm{argsup}}\, \frac{-\ln(r) 1_{\{r \geq \tau^*\}} - \ln(\tau^*) 1_{\{r < \tau^*\}}}{\sqrt{\tau_d - r}}$$

$$\qquad + o_P(1)$$

$$= \tau^* + o_P(1),$$

where $\tau_d = T_d/T$. $\qquad \square$

REFERENCES

Anatolyev, S. & G. Kosenok (2018) Sequential testing with uniformly distributed size. *Journal of Time Series Econometrics* 10, 1941–1928.

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

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

Astill, S., D.I. Harvey, S.J. Leybourne, R. Sollis, & A. Robert Taylor (2018) Real-time monitoring for explosive financial bubbles. *Journal of Time Series Analysis* 39, 863–891.

Aue, A. & L. Horváth (2004) Delay time in sequential detection of change. *Statistics & Probability Letters* 67, 221–231.

Aue, A. & L. Horváth (2013) Structural breaks in time series: Structural breaks in time series. *Journal of Time Series Analysis* 34, 1–16.

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

Aue, A., L. Horváth, & M.L. Reimherr (2009) Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics* 149, 174–190.

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

Berkes, I., S. Hörmann, J. Schauer, et al. (2011) Split invariance principles for stationary processes. *Annals of Probability* 39, 2441–2473.

Berkes, I., W. Liu, & W.B. Wu (2014) Komlós–Major–Tusnády approximation under dependence. *Annals of Probability* 42, 794–817.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edition. Wiley.

Brown, R.L., J. Durbin, & J.M. Evans (1975) Techniques for testing the Constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B* 37, 149–192.

Casini, A. (2021) Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Preprint, arXiv:2103.02981.

Casini, A. and P. Perron (2019) Structural breaks in time series. In *Oxford Research Encyclopedia of Economics and Finance*.

Casini, A. and P. Perron (2021a) Continuous record asymptotics for structural change models. Preprint, arXiv:1803.10881.

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

Chu, C.-S.J., M. Stinchcombe, & H. White (1996) Monitoring structural change. *Econometrica* 64, 1045–1065.

Dalla, V., L. Giraitis, & P.C.B. Phillips (2020) Robust tests for white noise and cross-correlation. *Econometric Theory*, forthcoming. DOI:10.1017/S0266466620000341.

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

Dette, H. & J. Gösmann (2020) A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association* 115, 1361–1377.

Fremdt, S. (2015) Page's sequential procedure for change-point detection in time series regression. *Statistics* 49, 128–155.

Gösmann, J., T. Kley, & H. Dette (2021) A new approach for open-end sequential change point monitoring. *Journal of Time Series Analysis* 42, 63–84.

Hansen, B.E. (1992) Testing for parameter instability in linear models. *Journal of Policy Modeling* 14, 517–533.

Homm, U. & J. Breitung (2012) Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics* 10(1), 198–231.

Horváth, L. (1995) Detecting changes in linear regressions. *Statistics* 26, 189–208.

Horváth, L., M. Hušková, P. Kokoszka, & J. Steinebach (2004) Monitoring changes in linear models. *Journal of Statistical Planning and Inference* 126, 225–251.

Jiang, P. & E. Kurozumi (2019) Power properties of the modified CUSUM tests. *Communications in Statistics - Theory and Methods* 48, 2962–2981.

Kirch, C. & J.T. Kamgaing (2015) On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference* 161, 25–49.

Komlós, J., P. Major, & G. Tusnády (1975) An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 32, 111–131.

Krämer, W., W. Ploberger, & R. Alt (1988) Testing for structural change in dynamic models. *Econometrica* 56, 1355–1369.

Leisch, F., K. Hornik, & C.-M. Kuan (2000) Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* 16, 835–854.

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

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B. & S.-P. Shi (2018) Financial bubble implosion and reverse regression. *Econometric Theory* 34(4), 705–753.

Phillips, P.C.B., Y. Wu, & J. Yu (2011) Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review* 52, 201–226.

Ploberger, W. & W. Krämer (1990) The local power of the CUSUM and CUSUM of squares tests. *Econometric Theory* 6, 335–347.

Ploberger, W. & W. Krämer (1992) The CUSUM test with OLS residuals. *Econometrica* 60, 271–285.

Robbins, H. & D. Siegmund (1970) Boundary crossing probabilities for the wiener process and sample sums. *Annals of Mathematical Statistics* 41, 1410–1429.

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

Sen, P.K. (1982) Invariance principles for recursive residuals. *Annals of Statistics* 10, 307–312.

Strassen, V. (1967) Almost sure behavior of sums of independent random variables and martingales. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 2, 315–343.

White, H. (2001). *Asymptotic Theory for Econometricians*, revised edition. Academic Press.

Wied, D. & P. Galeano (2013) Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference* 143, 186–196.

Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

Wu, W.B., et al. (2007) Strong invariance principles for dependent random variables. *Annals of Probability* 35, 2294–2320.

Zeileis, A., F. Leisch, C. Kleiber, & K. Hornik (2005) Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* 20, 99–121.