Backward CUSUM for Testing and Monitoring Structural Change

Sven Otto
University of Bonn, Institute for Finance and Statistics,
sven.otto@uni-bonn.de

Jörg Breitung*
Institute of Econometrics and Statistics, University of Cologne,
breitung@statistik.uni-koeln.de

July 10, 2020

Abstract

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 backward cumulated residuals. The existing invariance principle for partial sums of recursive residuals is extended to a multivariate version, and the limiting distributions of the test statistics are derived under suitable sequences of alternatives. In the retrospective context, the local power of the tests is shown to be substantially higher than that for 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 shown 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 ourperform the usual maximum likelihood estimator. Finally, an application to Covid-19 data is presented.

Keywords: Sequential tests; Recursive residuals; Open-end monitoring; Local delay; Breakpoint estimation.

*Corresponding author: Jörg Breitung, Institute of Econometrics and Statistics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany. Mail: breitung@statistik.uni-koeln.de

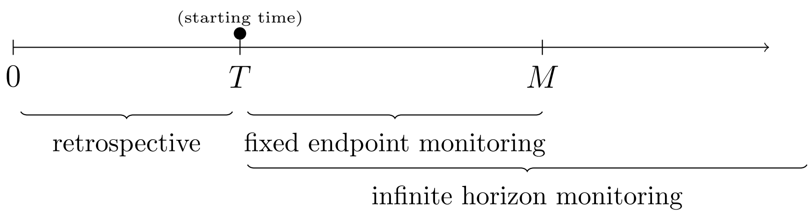
1 Introduction

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 et al. (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, 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. A variety of retrospective structural break tests have been proposed in the literature (for reviews, see Perron, 2006; Aue and Horváth, 2013).

Since the seminal work of Chu et al. (1996), increasing interest has been focused on monitoring structural stability in real time. Sequential monitoring procedures 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 endpoint were proposed in Leisch et al. (2000), Zeileis et al. (2005), Wied and Galeano (2013), and Dette and Gösmann (2019), whereas Chu et al. (1996), Horváth et al. (2004), Aue et al. (2006), Fremdt (2015), and Gösmann et al. (2019) considered an infinite monitoring horizon.

Figure 1: Retrospective testing and monitoring



A drawback of the conventional retrospective CUSUM test is its low power, whereas the conventional monitoring CUSUM procedure exhibits 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 non-zero 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 backwards 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 such backward CUSUM tests may indeed have a much higher power and lower detection delay than the conventional forward CUSUM tests.

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 uniform most powerful test statistic is the t-statistic of 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 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 et al., 1988). We extend the existing invariance principle for recursive residuals to a multivariate version and consider a vector-valued CUSUM process. For both retrospective testing and monitoring, we propose a vector-valued sequential statistic in the fashion of the score-based cumulative sum statistic of Hansen (1992). 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 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 on-line 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 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 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 backward 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 simulated critical values and Monte Carlo simulation results, in Section 8 we provide a real-data example on monitoring Covid-19 infections, and Section 9 concludes.

2 The multivariate CUSUM process

We consider the multiple linear regression model

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + u_t, \quad t \in \mathbb{N},$$

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. *The regressors x_t are stationary and ergodic with $E(x_t x_t') = C$, where C is positive definite. The error process u_t is a stationary martingale difference sequence with respect to $\mathcal{F}_t$, the σ-algebra generated by $\{(x_{i+1}', u_i)',\ i \leq t\}$, where $E(u_t^2 \mid \mathcal{F}_{t-1}) = \sigma^2 > 0$. Further, there exists $\kappa > 2$ such that $E(|x_{tj}|^\kappa) < \infty$ and $E(|u_t|^\kappa) < \infty$, for all $t \in \mathbb{N}$ and $j = 2, \ldots, k$.*

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 $\widehat{\beta}_{t-1} = (\sum_{i=1}^{t-1} x_i x_i')^{-1} \sum_{i=1}^{t-1} x_i y_i$. For testing against structural changes in the retrospective time period, the conventional univariate CUSUM statistic 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 $\{w_{k+1}, \ldots, w_T\}$ (see Brown et al., 1975; Krämer et al., 1988). Under the null hypothesis $H_0 : \beta_t = \beta_0$ for all t, the underlying process $S_{\lfloor rT \rfloor,T}$ converges weakly and uniformly to a standard Brownian motion $W(r)$, $r \in [0,1]$ (see Sen, 1982). The null hypothesis is rejected 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

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

The critical value λ_α is the $(1-\alpha)$ quantile of $\sup_{0 \leq r \leq 1} |W(r)|/d_{\text{lin}}(r)$ and determines the significance level α of the sequential test. 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 null hypothesis is rejected, if the detector statistic $|S_{t,T} - S_{T,T}|$ exceeds $b_t = b_{\text{rad}}(t/T)$ for some $t > T$. 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. Let $\pi = \text{plim}_{T\to\infty}(\overline{x}_1, \ldots, \overline{x}_k)'$ be the mean regressor, where $\overline{x}_j$ is the sample mean of the j-th component of the regressors, and let

$$h(r) = \frac{1}{\sigma} \int_0^r g(z)\,\mathrm{d}z - \frac{1}{\sigma} \int_0^r \int_0^z \frac{1}{z} g(v)\,\mathrm{d}v\,\mathrm{d}z. \tag{2}$$

The authors showed that $S_{\lfloor rT \rfloor,T}$ converges weakly and uniformly to $W(r) + \pi' h(r)$, which implies that the conventional univariate CUSUM tests have no power if the break vector

$g(r)$ is orthogonal to π. To sidestep this difficulty, we consider the multivariate statistic

$$Q_T(r) = \frac{1}{\widehat{\sigma}_T\sqrt{T}} C_T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} x_t w_t, \quad r \geq 0,$$

where $C_T = T^{-1}\sum_{t=1}^T x_t x_t'$. Let $m < \infty$. For $r \in [0, m]$ the process $Q_T(r)$ is an element of the k-fold product space $D([0,m])^k = D([0,m]) \times \ldots \times D([0,m])$, where $D([0,m])$ is in the space of càdlàg functions on $[0,m]$. The space is equipped with the Skorokhod metric (see Billingsley, 1999), and the symbol "$\Rightarrow$" denotes weak convergence with respect to this metric.

Theorem 1. *Let $g(r)$ be piecewise constant and bounded, and let $\beta_t = \beta_0 + T^{-1/2}g(t/T)$ for all $t \in \mathbb{N}$. Then, under Assumption 1 and for any fixed and positive $m < \infty$,*

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

as $T \to \infty$, where $W^{(k)}(r)$ is a k-dimensional standard Brownian motion.

Note that $g(r)$ is constant if and only if β_t is constant. If $\beta_t = \beta_0$ for all $t \in \mathbb{N}$, we have $C^{1/2}h(r) = 0$. By contrast, under a local alternative with a non-constant break function $g(r)$, it follows that $h(r)$ is non-zero, and, consequently, $C^{1/2}h(r)$ is non-zero, since $C^{1/2}$ is positive definite. Therefore, 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).

As an extension of the univariate CUSUM detector $S_{t,T}$ we consider the multivariate CUSUM detector $Q_{t,T} = Q_T(t/T)$. Let $\|x\| = \max_{i=1,\ldots,k} x_i$, $x \in R^k$, denote the maximum norm. The general retrospective forward CUSUM test rejects H_0 if the path of $\|Q_{t,T}\|$ exceeds the boundary function $b_t = b(t/T)$ for at least one index $t = 1, \ldots, T$. The general forward CUSUM for fixed endpoint monitoring rejects H_0 if the path of $\|Q_{t,T} - Q_{T,T}\|$ exceeds the boundary function $b_t = b((t-T)/T)$ for at least one index $t = T+1, \ldots, \lfloor mT \rfloor$, where $1 < m < \infty$. We make the following assumption on the general boundary function:

Assumption 2. *The boundary is of the form $b(r) = \lambda_\alpha d(r)$, where λ_α denotes the critical value for significance level α, and $d(r)$ is continuous and strictly increasing with $d(0) > 0$.*

A sequential tests can be equivalently expressed as a one-shot test, where H_0 is rejected if the corresponding maximum statistic exceeds the critical value λ_α, which is the $(1-\alpha)$ quantile of its limiting distribution under H_0. Under Assumption 2 and the conditions of Theorem 1, the continuous mapping theorem yields

$$\mathcal{Q}_T = \max_{t=1,\ldots,T} \|Q_{t,T}\|/d(t/T) \to \sup_{0\leq r\leq 1} \|W^{(k)}(r) + C^{1/2}h(r)\|/d(r),$$

$$\mathcal{Q}_{T,m} = \max_{t=T+1,\ldots,\lfloor mT\rfloor} \frac{\|Q_{t,T} - Q_{T,T}\|}{d((t-T)/T)} \to \sup_{0\leq r\leq m-1} \frac{\|W^{(k)}(r) + C^{1/2}(h(r+1) - h(1))\|}{d(r)},$$

in distribution, as $T \to \infty$.

3 Backward CUSUM tests

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 pre-break period, and $\{w_t,\ t \geq T^*\}$ are those from the post-break period. The pre-break residuals do not contain any information about the break and have mean zero. The partial sum process $T^{-1/2}\sum_{j=1}^{t} w_j$ has a random walk behaviour for the pre-break period $t < T^*$, and cumulating those residuals brings nothing but noise to the detector statistic. In contrast, the post-break residuals have nonzero mean and reveal relevant information about a possible break. In order to focus on the post-break residuals, we 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

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

The null hypothesis is rejected if $\|BQ_{t,T}\|$ exceeds the boundary $b_t = b((T-t-1)/T)$ for at least one time index t.

Theorem 2. *Let $g(r)$ be piecewise constant and bounded, and let $\beta_t = \beta_0 + T^{-1/2}g(t/T)$ for all $t \in \mathbb{N}$. Then, under Assumptions 1 and 2,*

$$\mathcal{BQ}_T = \max_{t=1,\ldots,T} \|BQ_{t,T}\|/d\big(\tfrac{T-t+1}{T}\big) \to \sup_{0\leq r\leq 1} \|W^{(k)}(r) + C^{1/2}(h(1) - h(1-r))\|/d(r)$$

in distribution, as $T \to \infty$.

Figure 2: Illustrative example for the backward CUSUM with a break in the mean

Note: The process $y_t = \mu_t + u_t$, $t = 1, \ldots, T$, is simulated for $T = 100$ with $\mu_t = 0$ for $t < 75$, $\mu_t = 1$ for $t \geq 75$, and i.i.d. standard normal innovations u_t. 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 lines correspond to the linear boundary (1) with significance level $\alpha = 5\%$ and critical value $\lambda_\alpha = 0.948$.

The limiting distributions of $\mathcal{Q}_T$ and $\mathcal{BQ}_T$ coincide under H_0 and differ under the alternative. A simple illustrative example of the detector paths together with the linear boundary of Brown et al. (1975) are depicted in Figure 2, in which a process with $k = 1$ and a single break in the mean at 3/4 of the sample is 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 $\|BQ_{t,T}\|$ is only defined for $t \leq T$, as its endpoint T is fixed.

To combine the advantages of $BQ_{t,T}$ with the measurability properties of $Q_{t,T}$, we resort to an inspection scheme, which goes back to Page (1954) and involves a triangular array of residuals. Let $\mathcal{BQ}_T(t) = \max_{s=1,\ldots,t} \|Q_T(t/T) - Q_T((s-1)/T)\|/d((t-s+1)/T)$ be the backward CUSUM statistic with endpoint t. The idea is to compute this statistic sequentially for each time point, yielding $\mathcal{BQ}_T(1), \ldots, \mathcal{BQ}_T(T)$. The stacked backward CUSUM statistic is the maximum among this sequence of backward CUSUM statistics. An important feature of this sequence is that it 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 triangular array structure. The stacked backward CUSUM detector is defined as

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

We reject H_0 if $\|SBQ_{s,t,T}\|$ exceeds the two-dimensional boundary $b_{s,t} = b((t-s+1)/T)$ for some $t = 1, \dots, T$ and $s = 1, \dots, t$. Equivalently, H_0 is rejected if the double maximum statistic

$$\mathcal{SBQ}_T = \max_{t=1,\dots,T} \mathcal{BQ}_T(t) = \max_{t=1,\dots,T} \max_{s=1,\dots,t} \|SBQ_{s,t,T}\| / d\left(\tfrac{t-s+1}{T}\right)$$

exceeds λ_α. The triangular detector can also be monitored on-line across all the time points $t > T$. The null hypothesis is rejected if $\|SBQ_{s,t,T}\|$ exceeds $b_{s,t} = b((t-s+1)/T)$ at least once for some $t \geq T+1$ and $s = T+1, \dots, t$. Analogously to the retrospective case, let

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

be the maximum statistic for fixed endpoint monitoring.

Theorem 3. *Let $g(r)$ be piecewise constant and bounded, let $\beta_t = \beta_0 + T^{-1/2} g(t/T)$ for all $t \in \mathbb{N}$, and let $m < \infty$ be a positive constant. Then, under Assumptions 1 and 2,*

$$\mathcal{SBQ}_T \to \sup_{0 \leq r \leq 1} \sup_{0 \leq s \leq r} \|W^{(k)}(r) - W^{(k)}(s) + C^{1/2}[h(r) - h(s)]\| / d(r-s)$$

$$\mathcal{SBQ}_{T,m} \to \sup_{0 \leq r \leq m-1} \sup_{0 \leq s \leq r} \|W^{(k)}(r) - W^{(k)}(s) + C^{1/2}[h(r+1) - h(s+1)]\| / d(r-s)$$

in distribution, as $T \to \infty$.

Remark 1. In practice one is often interested in breaks in certain coefficients or directions. Partial or one-sided tests can be beneficial in terms of a more powerful test. For testing the partial hypothesis $H_0 : H'\beta_t = H'\beta_0$, where H is an orthonormal $k \times l$ matrix, consider the partial CUSUM process $Q_T^*(t/T) = Q_{t,T}^* = \widehat{\sigma}_T^{-1} T^{-1/2} (H' C_T H)^{-1/2} H' \sum_{j=1}^{t} x_j w_j$. Then, H_0 is rejected if $\|Q_{t,T}^* - Q_{s,T}^*\|$ (partial SBQ test) or $\|Q_{T,T}^* - Q_{t,T}^*\|$ (partial BQ test) exceeds the respective boundary. Note that the critical values for the l-dimensional CUSUM test from Tables 1 and 2 can be applied for the partial test, since $Q_T^*(r) \Rightarrow W^{(l)}(r)$ under H_0. When testing against the one-sided alternative $H'\beta_t > H'\beta_0$, we reject if $p(H'BQ_{t,T})$ or $p(H'SBQ_{s,t,T})$ exceeds the respective boundary function, where $p(x) = \max_{i=1,\dots,l} x_i$,

$x \in \mathbb{R}^l$. The critical value coincides with the $(1-2\alpha)$ quantile of the respective l-dimensional limiting process. For instance, when testing for positive breaks in the intercept with $\alpha = 5\%$, the retrospective stacked backward CUSUM rejects if $p(H'SBQ_{s,t,T})$ exceeds $1.112 \cdot d((t-s+1)/T)$.

4 Local power

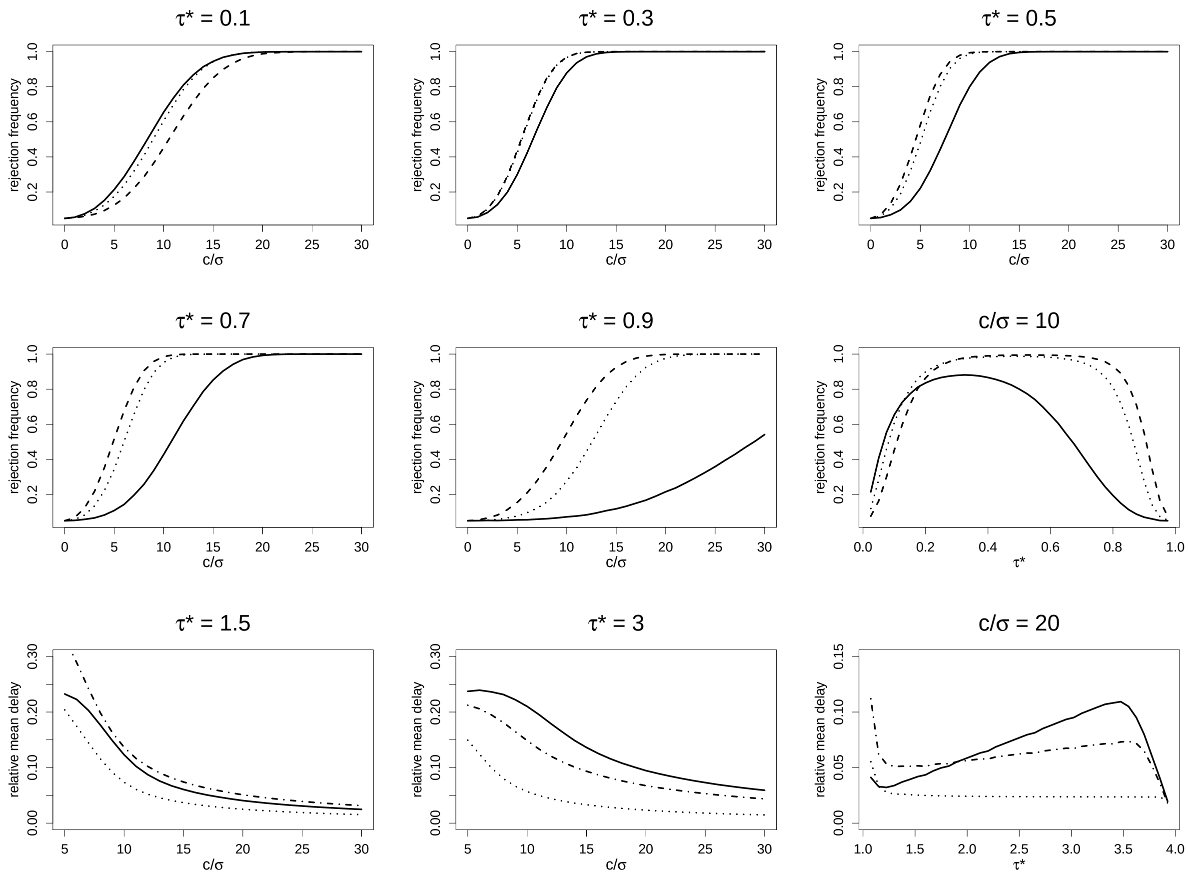
In order to illustrate the advantages of the backward CUSUM tests, we consider the simple local break model $\beta_t = \beta_0 + T^{-1/2} g(t/T)$ with $g(r) = c1_{\{r \geq \tau^*\}}$, where $c \in \mathbb{R}^k$, and τ^* denotes the break location. From (2) it follows that

$$h(r) = c\sigma^{-1}\Big(\int_{\tau^*}^{r} \mathrm{d}z - \int_0^r \int_{\tau^*}^{z} \frac{1}{z} \, \mathrm{d}v \, \mathrm{d}z \Big) = c\sigma^{-1}\tau^*(\ln(r) - \ln(\tau^*))1_{\{r \geq \tau^*\}}. \tag{3}$$

Simulated asymptotic local power curves of $\mathcal{Q}_T$, $\mathcal{BQ}_T$, and $\mathcal{SBQ}_T$ under the limiting distribution for the case $k = 1$ are presented in Figure 3, where the linear boundary (1) is implemented. The plots show that for a single break that is located after 15% of the sample size, the backward CUSUM and the stacked backward CUSUM clearly outperform the forward CUSUM in terms of power. The local power curves of $\mathcal{Q}_{T,m}$ and $\mathcal{SBQ}_{T,m}$ for a break at $\tau^* \in (1, 2)$ with fixed endpoint $m = 2$ coincide with those of $\mathcal{Q}_T$ and $\mathcal{SBQ}_T$ in the upper six panels, except that τ^* is shifted by 1 to the right. Hence, the power of $\mathcal{SBQ}_{T,m}$ is higher than that of $\mathcal{Q}_{T,m}$ if $\tau^* \geq 1.15$.

The more important performance measure for monitoring detectors is the delay between the actual break and the detection time point, since every fixed nontrivial alternative will be detected if the monitoring horizon is long enough. Let T_d be the stopping time of the time point of the first boundary crossing, and let $E(T_d/T \mid \tau^* \leq T_d/T \leq m) - \tau^*$ be the mean local relative delay. The bottom panels of Figure 3 present the simulated mean local relative delay curves for the fixed endpoint $m = 4$ for $\mathcal{SBQ}_{T,4}$ with the linear boundary and for $\mathcal{Q}_{T,4}$ with both the linear and the radical boundary. The mean local relative delay of $\mathcal{SBQ}_{T,4}$ is much lower than that of $\mathcal{Q}_{T,4}$. Furthermore, the mean local relative delay is constant across break locations with $\tau^* > 1.15$.

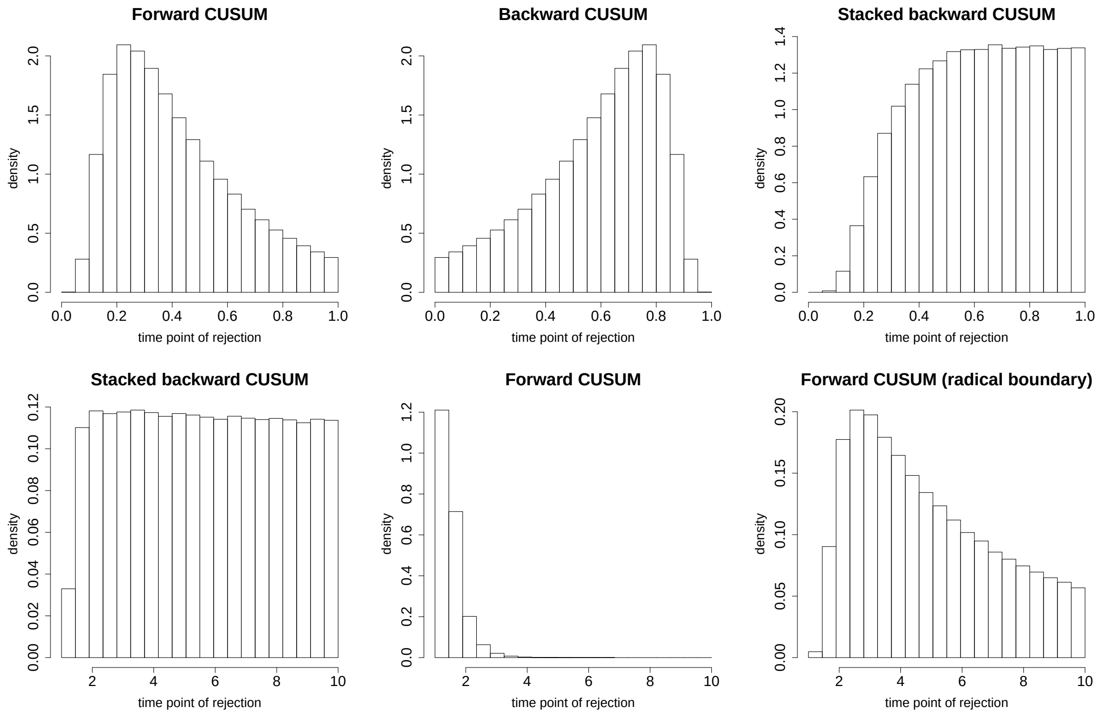
Figure 3: Asymptotic local power curves



Note: The plots show simulated asymptotic local power curves for retrospective tests (upper six panels) and mean local relative delay curves for fixed endpoint monitoring with $m = 4$ (bottom three panels) under a single break in the mean ($k = 1$) for $\mathcal{Q}_T$ and $\mathcal{Q}_{T,4}$ (solid), $\mathcal{BQ}_T$ (dashed) and $\mathcal{SBQ}_T$ and $\mathcal{SBQ}_{T,4}$ (dotted), under the linear boundary (1), as well as the test by Chu et al. (1996) (dash-dotted). Brownian motions are approximated on a grid of 1,000 equidistant points and rejection rates are obtained from 100,000 Monte Carlo repetitions using size-adjusted 5% critical values.

Remark 2. 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 (1) 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 for the stacked backward CUSUM the size is much more evenly distributed than that for the forward CUSUM, which is right-skewed. There is no consensus on which distribution

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 is considered for the upper three histograms and the fixed endpoint monitoring case with $m = 10$ for the lower three. The linear boundary (1) is considered in the first five plots and the radical boundary by Chu et al. (1996) is used in the last plot.

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. However, in the context of infinite horizon monitoring the size can never be uniformly distributed.

5 Infinite horizon monitoring

The functional central limit theorem given by Theorem 1 is not suitable for analysing the asymptotic behaviour 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. For an i.i.d. random process v_t, $t \in \mathbb{N}$, with $E[v_1] = 0$, $E[v_1^2] = \sigma^2$, and $E[v_1^\kappa] < \infty$, where $\kappa > 2$, Komlós et al. (1975) showed that there exists a standard Brownian motion $W(r)$, such that

$\sigma^{-1}\sum_{t=1}^{T} v_t = W(T) + o(T^{1/\kappa})$, a.s., as $T \to \infty$. This almost sure invariance principle was employed by Horváth (1995) to derive the limiting distribution of the infinite horizon statistic $\sup_{t>T} |S_{t,T} - S_{T,T}|/d(t/T)$ for an appropriate boundary function $d(r)$. Wu et al. (2007) and Berkes et al. (2014) extended this invariance principle to more general classes of dependent random processes, which can be used to formulate the following stochastic approximation result:

Theorem 4. *Let $\beta_t = \beta_0$ for all $t \in \mathbb{N}$. Then, under Assumption 1, there exists a k-dimensional standard Brownian motion $W^{(k)}(r)$, such that*

$$\sup_{r\geq 1} r^{-1/2} \|Q_T(r) - W^{(k)}(r)\| \to 0,$$

in probability, as $T \to \infty$.

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.

Theorem 5. *Let $\beta_t = \beta_0$ for all $t \in \mathbb{N}$, and let $\sup_{r\geq 0}(r+1)^{1/2}/d(r) < \infty$. Then, under Assumptions 1 and 2,*

$$\mathcal{Q}_{T,\infty} = \max_{t\geq T+1} \|Q_T(t/T) - Q_T(1)\|/d(\tfrac{t-1}{T}) \to \sup_{0\leq r\leq 1} \|B^{(k)}(r)\|/((1-r)d(\tfrac{r}{1-r})),$$

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

$$\to \sup_{0\leq r\leq 1} \sup_{0\leq s\leq r} \|(1-s)B^{(k)}(r) - (1-r)B^{(k)}(s)\|/((1-r)(1-s)d(\tfrac{r-s}{1-r-s+rs})),$$

in distribution, as $T \to \infty$, where $B^{(k)}(r)$ is a k-dimensional standard Brownian bridge.

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 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

$$\widehat{\tau}_{ML}^{\text{ret}} = T^{-1} \cdot \underset{t=1,\dots,T}{\operatorname{argmin}} \{R_1(t) + R_2(t)\}, \tag{4}$$

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

$$\widehat{\tau}_{ML}^{\text{mon}} = T^{-1} \cdot \underset{t=T+1,\dots,T_d}{\operatorname{argmin}} \{R_1(t) + R_2(t)\} \tag{5}$$

to estimate $\tau^*_{\text{mon}} = T^*/T_d$, 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.

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 pre-break period and decreases to zero in the post-break period. When scaled by its asymptotic standard deviation, the detector is asymptotically proportional to $\|h(1) - h(r)\|/\sqrt{1-r}$, which is in turn proportional to

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

where the maximum is attained at $r = \tau^*$ (see equation (3)). Accordingly, we consider

$$\widehat{\tau}_{\text{ret}} = \frac{1}{T} \cdot \underset{t=1,\dots,T}{\operatorname{argmax}} \frac{\|BQ_{t,T}\|}{\sqrt{T-t+1}}, \qquad \widehat{\tau}_{\text{mon}} = \frac{1}{T} \cdot \underset{t=T+1,\dots,T_d}{\operatorname{argmax}} \frac{\|BQ_{t,T_d}\|}{\sqrt{T-d+t-1}}. \tag{6}$$

Theorem 6. *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 $\widehat{\tau}_{ret} \to \tau^*$, in probability, as $T \to \infty$; if $\tau^* \in (1, T_d/T]$, then $\widehat{\tau}_{mon} \to \tau^*$, in probability, as $T \to \infty$.*

This result implies that the breakpoint estimators (6) are consistent, as $T \to \infty$.

Table 1: Asymptotic critical values for $\mathcal{BQ}_T$

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$\alpha = 1\%$	1.143	1.219	1.260	1.287	1.307	1.323	1.338	1.349
$\alpha = 5\%$	0.945	1.032	1.081	1.114	1.139	1.160	1.176	1.189
$\alpha = 10\%$	0.847	0.941	0.993	1.029	1.056	1.077	1.095	1.110

Note: Asymptotic critical values are reported for $\mathcal{BQ}_T$, $d(r) = 1 + 2r$, based on 100,000 Monte Carlo replications, where the Wiener process is approximated on a grid of 10,000 equidistant points.

Table 2: Asymptotic critical values for $\mathcal{SBQ}_{T,m}$

	$k=1$			$k=2$			$k=3$			$k=4$		
$m \setminus \alpha$	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.2	0.782	0.859	1.024	0.859	0.935	1.092	0.902	0.975	1.129	0.932	1.003	1.152
1.4	0.941	1.030	1.208	1.028	1.111	1.277	1.076	1.156	1.320	1.108	1.185	1.345
1.6	1.026	1.113	1.292	1.111	1.192	1.365	1.158	1.238	1.406	1.189	1.269	1.432
1.8	1.077	1.162	1.344	1.161	1.244	1.411	1.208	1.286	1.452	1.240	1.317	1.476
2	1.113	1.198	1.374	1.196	1.277	1.442	1.244	1.321	1.481	1.275	1.350	1.506
4	1.262	1.339	1.500	1.336	1.410	1.564	1.378	1.450	1.599	1.407	1.478	1.621
10	1.367	1.440	1.588	1.437	1.503	1.644	1.475	1.540	1.677	1.500	1.565	1.703
∞	1.450	1.514	1.648	1.512	1.573	1.703	1.547	1.606	1.736	1.570	1.629	1.760
	$k=5$			$k=6$			$k=7$			$k=8$		
$m \setminus \alpha$	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.2	0.954	1.023	1.170	0.972	1.041	1.186	0.987	1.054	1.198	1.000	1.065	1.206
1.4	1.133	1.208	1.366	1.152	1.225	1.381	1.167	1.241	1.396	1.181	1.253	1.409
1.6	1.214	1.293	1.452	1.235	1.311	1.466	1.251	1.325	1.477	1.265	1.339	1.488
1.8	1.265	1.340	1.496	1.283	1.357	1.511	1.300	1.372	1.525	1.315	1.385	1.537
2	1.299	1.372	1.526	1.317	1.388	1.541	1.333	1.404	1.556	1.347	1.418	1.566
4	1.429	1.497	1.638	1.446	1.513	1.651	1.461	1.527	1.665	1.473	1.539	1.679
10	1.520	1.584	1.718	1.536	1.599	1.732	1.551	1.612	1.744	1.562	1.623	1.752
∞	1.589	1.647	1.775	1.604	1.661	1.788	1.617	1.673	1.799	1.627	1.683	1.807

Note: Asymptotic critical values are reported for $\mathcal{SBQ}_{T,m}$, $d(r) = 1 + 2r$, based on 100,000 Monte Carlo replications, where the Wiener process is approximated on a grid of 10,000 equidistant points.

Table 3: Empirical sizes of the retrospective tests

	$T = 100$			$T = 200$			$T = 500$		
	$\mathcal{Q}_T$	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	$\mathcal{Q}_T$	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	$\mathcal{Q}_T$	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$
Model I	3.9	4.2	2.9	4.2	4.3	3.5	4.6	4.6	4.1
Model II	3.8	4.4	3.3	4.2	4.6	3.6	4.5	4.8	4.1

Note: Empirical sizes of the retrospective tests with $d(r) = 1 + 2r$ are presented in percentage points. The results are based on 100,000 Monte Carlo repetitions, using the 5% critical values from Tables 1 and 2

Table 4: Size-adjusted powers of the retrospective tests

	Model I				Model II			
	$\mathcal{Q}_T$	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	supW	$\mathcal{Q}_T$	$\mathcal{BQ}_T$	$\mathcal{SBQ}_T$	supW
$\tau^* = 0.1$	46.7	27.8	40.8	26.3	33.8	19.7	27.2	21.2
$\tau^* = 0.2$	63.6	64.6	71.2	74.1	48.9	49.2	53.9	59.5
$\tau^* = 0.3$	67.2	83.9	84.0	87.0	52.6	71.4	69.8	75.3
$\tau^* = 0.4$	63.9	91.4	88.9	91.6	48.9	82.5	77.0	82.1
$\tau^* = 0.5$	53.9	93.6	89.5	92.6	39.9	86.0	78.1	84.2
$\tau^* = 0.6$	39.5	93.3	86.8	91.6	27.8	84.4	73.2	82.3
$\tau^* = 0.7$	23.1	88.9	77.1	86.9	16.4	76.4	60.7	75.3
$\tau^* = 0.8$	10.5	67.7	49.1	68.5	8.1	52.2	36.7	54.1
$\tau^* = 0.9$	5.4	24.6	11.4	21.3	5.1	21.3	12.4	17.8

Note: Size-adjusted powers of the retrospective tests with $d(r) = 1 + 2r$ are reported for a significance level of 5% and a sample size of $T = 100$, based on 100,000 Monte Carlo repetitions. The sup-Wald test by Andrews (1993) with trimming parameter 0.15 is denoted as supW.

Table 5: Empirical sizes of the infinite horizon monitoring detectors

	Model I						Model II					
	$T = 100$			$T = 500$			$T = 100$		$T = 200$		$T = 500$	
horizon	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	$\mathcal{SBQ}$	$\mathcal{Q}$	$\mathcal{SBQ}$	$\mathcal{Q}$	$\mathcal{SBQ}$	$\mathcal{Q}$
$m = 1.5$	0.1	2.8	0.0	0.0	3.0	0.0	0.2	3.8	0.1	3.3	0.1	3.2
$m = 2$	0.3	4.3	0.1	0.2	4.4	0.1	0.7	5.6	0.4	4.9	0.3	4.7
$m = 4$	1.0	4.9	1.0	0.8	4.8	0.8	2.7	6.2	1.6	5.4	1.3	5.2
$m = 10$	3.2	4.9	2.4	2.7	4.8	2.0	8.0	6.2	4.9	5.4	4.1	5.2

Note: Empirical sizes of the monitoring detectors using infinite horizon 5% critical values are reported based 100,000 Monte Carlo repetitions. While $\mathcal{SBQ}$ corresponds the infinite horizon stacked backward CUSUM with linear boundary (1) and $\mathcal{Q}$ is the infinite horizon multivariate forward CUSUM with linear boundary (1), the univariate infinite horizon monitoring by Chu et al. (1996) is denoted as CSW.

Table 6: Empirical mean detection delays of the monitoring detectors

	Model I			Model II			Model I			Model II	
breakpoint	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	$\mathcal{SBQ}$	$\mathcal{Q}$	breakpoint	$\mathcal{SBQ}$	$\mathcal{Q}$	CSW	$\mathcal{SBQ}$	$\mathcal{Q}$
$\tau^* = 1.5$	37.7	39.5	52.3	52.5	49.9	$\tau^* = 3$	32.9	99.4	69.8	44.3	124.8
$\tau^* = 2$	35.0	59.6	58.7	48.0	74.7	$\tau^* = 5$	31.6	179.0	88.2	41.2	223.0
$\tau^* = 2.5$	33.6	79.6	64.6	45.8	99.9	$\tau^* = 10$	30.6	377.1	123.2	39.6	461.9

Note: Size-adjusted empirical mean detection delays ($\alpha = 5\%$) of the monitoring detectors, with $T = 100$ and $m = 20$ are reported based on 100,000 Monte Carlo repetitions. While $\mathcal{SBQ}$ corresponds the infinite horizon stacked backward CUSUM with linear boundary (1) and $\mathcal{Q}$ is the infinite horizon multivariate forward CUSUM with linear boundary (1), the univariate infinite horizon monitoring by Chu et al. (1996) is denoted as CSW.

Table 7: Bias and MSE of breakpoint estimators

	$T = 100$				$T = 200$			
	Bias		MSE		Bias		MSE	
τ^*	BQ	ML	BQ	ML	BQ	ML	BQ	ML
0.5	-0.03	0.01	0.02	0.01	-0.02	0.01	0.01	0.00
0.65	-0.03	0.00	0.02	0.01	-0.02	0.00	0.01	0.00
0.8	-0.03	-0.04	0.02	0.03	-0.01	-0.01	0.01	0.01
0.85	-0.04	-0.07	0.03	0.05	-0.02	-0.02	0.01	0.01
0.9	-0.06	-0.13	0.04	0.09	-0.03	-0.04	0.02	0.03
0.95	-0.10	-0.25	0.06	0.18	-0.05	-0.14	0.03	0.11
0.97	-0.13	-0.33	0.08	0.25	-0.08	-0.24	0.05	0.18
0.99	-0.20	-0.44	0.13	0.33	-0.15	-0.40	0.09	0.31

Note: The bias and mean squared error (MSE) for the breakdate estimators (4) and (6) are reported based on 100,000 Monte Carlo repetitions, where model (Model I) is simulated for $t = 1, \dots, T$. BQ denotes the backward CUSUM estimator (6), and ML denotes the maximum likelihood estimator (4).

7 Finite sample performance

Tables 1 and 2 present critical values for the retrospective and monitoring detectors using the linear boundary (1). The critical values for $\mathcal{SBQ}_T$ coincide with those for $\mathcal{SBQ}_{T,2}$. Empirical sizes in Table 3 indicate that the retrospective tests have only minor size distortions in finite samples. The empirical powers of the retrospective tests are compared with that of the sup-Wald test of Andrews (1993), where the trimming parameter is $r_0 = 0.15$. The sup-Wald test has weak optimality properties 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 asymptotically, as the significance level tends to zero. Note that 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.

We illustrate the finite sample performance for the models,

$$y_t = \mu_t + u_t, \quad \mu_t = 2 + 0.8 \cdot 1_{\{t/T \geq \tau^*\}}, \qquad \text{(Model I)}$$

$$y_t = \mu_t + \gamma_t x_t + u_t, \quad \mu_t = 2, \quad \gamma_t = 1 + 0.8 \cdot 1_{\{t/T \geq \tau^*\}} \quad (t = 1, \ldots, T), \qquad \text{(Model II)}$$

where u_t and x_t are i.i.d. standard normal. Table 4 presents the size-adjusted power results. 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 the backward CUSUM scheme seems to compensate for this weakness of the forward CUSUM test. The backward CUSUM performs equally well than the sup-Wald test, which is remarkable since, as discussed previously, the latter test has weak optimality properties. Finally, while the sup-Wald statistic and the backward CUSUM detector are not suitable for monitoring, $\mathcal{SBQ}_T$ is much more powerful than $\mathcal{Q}_T$, and its detector statistic is therefore well suited for real-time monitoring.

In order to evaluate the finite sample performances of the monitoring detectors, we consider the same models for the time points $t = T+1, \ldots, \lfloor mT \rfloor$. We simulate the series up to different fixed endpoints m, while the critical values for the case $m = \infty$ are implemented. For $\mathcal{Q}_{T,\infty}$ with the linear boundary, the 5% critical values are given by 0.957 for $k = 1$ and 1.044 for $k = 2$. Table 5 presents the empirical sizes. The tests are undersized by construction, as not all of the size is used up to the time point mT. For $k \geq 2$, we observe some size distortions for small sample sizes. The results in Table 6 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.

To compare the breakpoint estimator in equation (6) with its maximum likelihood benchmark in (4) and (5), 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 mean squared error than the maximum likelihood estimator, which is due to the fact that the post-break sample consists of too few observations for an accurate maximum likelihood estimation.

8 Real-data example

We consider the daily time series of Covid-19 new infections for the US. In order to control for a second wave of infections we estimate the following model for the time after the first peak (April 10):

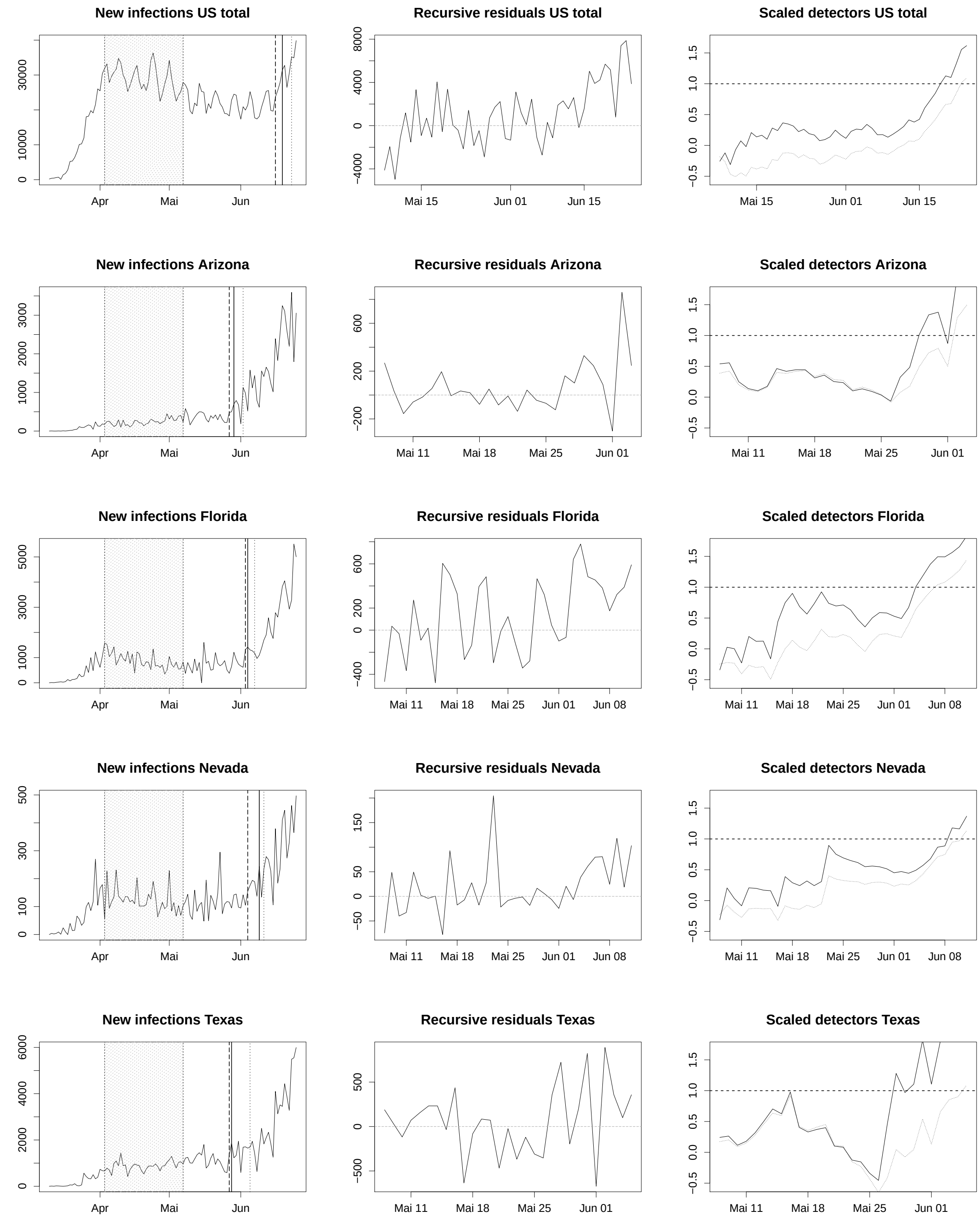
$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_7 y_{t-7} + u_t = x_t'\beta + u_t \tag{7}$$

with $x_t = (1, y_{t-1}, y_{t-7})'$ and $\beta = (\phi_0, \phi_1, \phi_7)'$. The parameters ϕ_1 and ϕ_7 control for the observed persistence and seasonality in the series. According to the Ljung-Box statistics with 8 lags, no serial correlation is left in the residuals for the first $T = 28$ observations (April 10 – May 8), which is used as a pre-monitoring training sample.

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%, which implies $\alpha = 0.1$ (see Remark 1). We consider the univariate one-sided forward CUSUM statistic $T^{-1/2}\widehat{\sigma}_T^{-1}\sum_{j=T+1}^{t} w_t$, which rejects if its path exceeds the boundary of Chu et al. (1996) for $\alpha = 0.1$, and the one-sided stacked backward CUSUM statistic for a break in the intercept, which is $T^{-1/2}\widehat{\sigma}_T^{-1}\max_{s=T+1,\dots,t}(1+2(t-s)/T)\sum_{j=s}^{t} w_t$ and rejects if its path exceeds 1.45. Figure 5 presents the detectors, which are scaled by their boundary.

Both monitoring procedures find an indication for a rise in Covid-19 infections in the US at the end of June. According to the panels in the first column of Figure 5, also for Arizona, Florida, Nevada, and Texas the tests indicate a second wave of infection. The stacked backward CUSUM detects the break much earlier and becomes significant between 2 and 8 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.

Figure 5: Monitoring daily new Covid-19 infections in the US



Note: The plots in the first column presents daily new Covid-19 infections, where the shaded areas represent the pre-monitoring training period, the solid vertical line represents the detection time point of the stacked backward CUSUM, the dashed line is the backward CUSUM breakpoint estimator given by equation (6), and the dotted vertical line is the detection time point of the forward CUSUM. The detectors in the plots of the third column are scaled by their boundary function, where the solid line represents the one-sided stacked backward CUSUM and the dotted line is the one-sided forward CUSUM of Chu et al. (1996). The horizontal dashed line represents the critical boundary of the scaled detectors.

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 result 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 on-line 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.

Acknwoledgements

We are thankful to Holger Dette, Josua Gösmann and Dominik Wied for very helpful comments and suggestions. Furthermore, the usage of the CHEOPS HPC cluster for parallel computing is gratefully acknowledged.

Supporting Information

An accompanying `R`-package for all methods presented in this article is available online at `https://github.com/ottosven/backCUSUM`.

A Appendix: Technical proofs

A.1 Auxiliary lemmas

We first present some auxiliary lemmas which we require for the proofs of the main theorems.

Lemma 1. *Under Assumption 1, there exists a k-dimensional standard Brownian motion $W^{(k)}(r)$, such that the following statements hold true: (a) for any fixed $m < \infty$, as $T \to \infty$,*

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

where "$\Rightarrow$" denotes weak convergence on the k-fold càdlàg space $D([0, m])^k$ with respect to the Skorokhod metric; (b)

$$\lim_{t \to \infty} \frac{\| \sum_{j=1}^{t} x_j u_j - \sigma C^{1/2} W^{(k)}(t) \|}{\sqrt{t}} = 0 \quad (a.s.).$$

Proof. For (a), note that a direct consequence of the functional central limit theorem for multiple time series on the space $D([0, 1])^k$ given by Theorem 2.1 in Phillips and Durlauf (1986) is that $M^{-1/2} \sum_{t=1}^{\lfloor sM \rfloor} x_t u_t \Rightarrow \sigma C^{1/2} W^{(k)}(s)$, $s \in [0, 1]$, as $M \to \infty$ (see also Lemma 3 in Krämer et al. 1988). Then, on the space $D([0, m])^k$,

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

To show (b), note that $\{x_t u_t\}_{t \in \mathbb{N}}$ is a stationary and ergodic martingale difference sequence with $E(x_t u_t) = 0$ and $E[(x_t u_t)(x_t u_t)'] = \sigma^2 C$. We apply the strong invariance principle given by Theorem 3 in Wu et al. (2007). Then,

$$\lim_{t \to \infty} \frac{\|\sigma^{-1} C^{-1/2} \sum_{j=1}^{t} x_j u_j - W^{(k)}(t)\|}{t^{1/q} \sqrt{\ln(t)} (\ln(\ln(t)))^{1/4}} < \infty, \quad \text{(a.s.)},$$

where $q = \min\{\kappa, 4\}$ (see also Strassen 1967), and the assertion follows from the fact that

$$\lim_{t \to \infty} t^{1/q} \sqrt{\ln(t)} (\ln(\ln(t)))^{1/4} / \sqrt{t} = 0$$

□

Lemma 2. *Let* $\{(x_t, u_t)\}_{t\in\mathbb{N}}$ *satisfy Assumption 1, let* $\beta_t = \beta_0$ *for all* $t \in \mathbb{N}$*, and let* $m < \infty$*. Let* $X_t = \sum_{j=1}^{t} x_j w_j$*,* $Y_t = \sum_{j=1}^{t} x_j u_j$*, and* $Z_t = \sum_{j=1}^{t-1}\sum_{i=1}^{j} j^{-1} x_i u_i$*. Then, as* $T \to \infty$*,*

$$\sup_{1\le t\le mT} \frac{\|X_t - (Y_t - Z_t)\|}{\sqrt{T}} = o_P(1), \quad \textit{and} \quad \sup_{T<t<\infty} \frac{\|X_t - (Y_t - Z_t)\|}{\sqrt{t}} = o_P(1).$$

Proof. First, note that $w_t = 0$ for $t \le k$. For $t > k$ let $f_t = (1 + (t-1)^{-1} x_t' C_{t-1}^{-1} x_t)^{1/2}$ be the denominator of w_t. Then,

$$f_t w_t = y_t - x_t'\widehat{\beta}_{t-1} = u_t - x_t'\Big(\sum_{j=1}^{t-1} x_j x_j'\Big)^{-1}\Big(\sum_{j=1}^{t-1} x_j u_j\Big) = u_t - x_t' C_{t-1}^{-1}\Big(\frac{1}{t-1}\sum_{j=1}^{t-1} x_j u_j\Big).$$

Furthermore, let $\widetilde{Y}_t = \sum_{j=k+1}^{t} f_j^{-1} x_j u_j$, and $\widetilde{Z}_t = \sum_{j=k}^{t-1}\sum_{i=1}^{j} j^{-1} f_{j-1}^{-1} x_{j+1} x_{j+1}' C_j^{-1} x_i u_i$. Then, $X_t = \sum_{j=k+1}^{t} f_j^{-1} x_j (u_j - (j-1)^{-1} x_j' C_{j-1}^{-1} \sum_{i=1}^{j-1} x_i u_i) = \widetilde{Y}_t - \widetilde{Z}_t$. Hence, it remains to show, that

$$\sup_{1\le t\le mT} \frac{\|\widetilde{Y}_t - Y_t\|}{\sqrt{T}} = o_P(1), \quad \text{and} \quad \sup_{T<t<\infty} \frac{\|\widetilde{Y}_t - Y_t\|}{\sqrt{t}} = o_P(1), \tag{8}$$

and that

$$\sup_{1\le t\le mT} \frac{\|\widetilde{Z}_t - Z_t\|}{\sqrt{T}} = o_P(1), \quad \text{and} \quad \sup_{T<t<\infty} \frac{\|\widetilde{Z}_t - Z_t\|}{\sqrt{t}} = o_P(1). \tag{9}$$

To show (8) and (9), we apply Abel's formula of summation by parts, which is given by

$$\sum_{t=1}^{n} A_t b_t = \sum_{t=1}^{n} A_t b_n + \sum_{t=1}^{n-1}\sum_{j=1}^{t} A_j (b_t - b_{t+1}), \qquad A_t \in \mathbb{R}^{k\times k}, \quad b_t \in \mathbb{R}^k, \quad n \in \mathbb{N}. \tag{10}$$

Let $\psi_T = \sqrt{T}((f_T - 1)1_{\{T>k\}} - 1_{\{T\le k\}})$, which is $O_P(1)$, since $\sqrt{T}(f_T - 1) = O_P(1)$, as $T \to \infty$, and let $a_t = t^{-1/2}\sum_{j=1}^{t} \psi_j x_j u_j$, where $\|a_T\| = O_P(1)$. Furthermore, note that $j^{-1/2} - (j+1)^{-1/2} < j^{-3/2}$. Then,

$$\widetilde{Y}_t - Y_t = \sum_{j=1}^{t} (\psi_j x_j u_j) j^{-1/2} = a_t + \sum_{j=1}^{t-1}\Big(a_j j^{1/2}\big[j^{-1/2} - (j+1)^{-1/2}\big]\Big) < a_t + \sum_{j=1}^{t-1}\frac{1}{j} a_j,$$

which implies that

$$\sup_{1\le t\le mT} \frac{\|\widetilde{Y}_t - Y_t\|}{\sqrt{T}} < \sup_{1\le t\le mT}\Big(\frac{\|a_t\|}{\sqrt{T}} + \frac{m}{T^{1/4}}\sum_{j=1}^{t-1}\frac{\|a_j\|}{j^{5/4}}\Big) = o_P(1),$$

and

$$\sup_{T<t<\infty} \frac{\|\widetilde{Y}_t - Y_t\|}{\sqrt{t}} < \sup_{T<t<\infty} \Big(\frac{\|a_t\|}{\sqrt{T}} + \frac{1}{T^{1/4}} \sum_{j=1}^{t-1} \frac{\|a_j\|}{j^{5/4}} \Big) = o_P(1).$$

To show (9), let $Z_t^* = \sum_{j=1}^{t-1} \sum_{i=1}^{j} j^{-1} x_{j+1} x_{j+1}' C^{-1} x_i u_i$, $\widetilde{A}_j = f_{j-1}^{-1} C_j^{-1} 1_{\{j \geq k\}} - C^{-1}$, and $\widetilde{a}_j = j^{-1/2} \sum_{i=1}^{j} x_{j+1} x_{j+1}' \widetilde{A}_j x_i u_i$, such that $\widetilde{Z}_t - Z_t^* = \sum_{j=1}^{t-1} j^{-1/2} \widetilde{a}_j$. Since $\{x_t\}_{t\in\mathbb{N}}$ is ergodic, we have $\|\widetilde{A}_T\|_M = o_P(1)$, as $T \to \infty$, where $\|\cdot\|_M$ denotes the matrix norm induced by $\|\cdot\|$, and $\|\widetilde{a}_T\| = o_P(1)$. Moreover, there exists some $\epsilon > 0$ and some random variable ξ, such that $\|\widetilde{a}_j\| \leq j^{-\epsilon}\xi$. Thus,

$$\sup_{1\leq t\leq mT} \frac{\|\widetilde{Z}_t - Z_t^*\|}{\sqrt{T}} \leq \frac{m\xi}{T^\epsilon} \sum_{j=1}^{\infty} \frac{1}{j^{1+\epsilon}} = o_P(1), \quad \sup_{T<t<\infty} \frac{\|\widetilde{Z}_t - Z_t^*\|}{\sqrt{t}} \leq \frac{\xi}{T^\epsilon} \sum_{j=1}^{\infty} \frac{1}{j^{1+\epsilon}} = o_P(1).$$

Finally, with $A_j^* = x_{j+1} x_{j+1}' C^{-1} - I_K$ and $b_t^* = t^{-1} \sum_{j=1}^{t} x_j u_j$, (10) yields

$$\begin{aligned} Z_t^* - Z_t &= \sum_{j=1}^{t-1} A_j^* b_j^* = \sum_{j=1}^{t-1} A_j^* b_{t-1}^* + \sum_{j=1}^{t-2} \sum_{i=1}^{j} A_i^* \big[b_j^* - b_{j+1}^* \big] \\ &= (t-1) B_{t-1}^* b_{t-1}^* + \sum_{j=1}^{t-2} j B_j^* \Big[\frac{1}{j+1} b_{j+1}^* + \frac{1}{j} x_{j+1} u_{j+1} \Big], \end{aligned}$$

where $B_t^* = t^{-1} \sum_{j=1}^{t} A_j^*$. Since $\|B_T^*\|_M = o_P(1)$ and $\|b_T^*\| = O_P(T^{-1/2})$, there exists some $\gamma > 0$ and some random variable ζ, such that $\|B_t^* b_t^*\| \leq t^{-1/2-\gamma}\zeta$, $\|B_t^* b_{t+1}^*\| \leq t^{-1/2-\gamma}\zeta$, and $\|\sum_{j=1}^{t} B_j^* x_{j+1} u_{j+1}\| \leq t^{1/2-\gamma}\zeta$, which yields

$$\begin{aligned} \|Z_t^* - Z_t\| &\leq \zeta \Big[(t-1) t^{-1/2-\gamma} + \sum_{j=1}^{t-2} \frac{j^{1/2-\gamma}}{j+1} + (t-2)^{1/2-\gamma} \Big] \\ &\leq \zeta \Big[2t^{1/2-\gamma} + t^{1/2-\gamma/2} \sum_{j=1}^{t-2} \frac{1}{j^{1+\gamma/2}} \Big] \leq \zeta K t^{1/2-\gamma/2} \end{aligned}$$

for some constant $K < \infty$. Consequently,

$$\sup_{1\leq t\leq mT} \frac{\|Z_t^* - Z_t\|}{\sqrt{T}} = o_P(1), \quad \text{and} \quad \sup_{T<t<\infty} \frac{\|Z_t^* - Z_t\|}{\sqrt{t}} = o_P(1),$$

and (9) follows by the triangle inequality. □

Lemma 3. *Let $W^{(k)}(r)$ be a k-dimensional standard Brownian motion and let $B^{(k)}(r)$ be a k-dimensional standard Brownian bridge. Then,* *(a)* $W^{(k)}(r) - \int_0^r z^{-1} W^{(k)}(z)\,\mathrm{d}z \overset{d}{=} W^{(k)}(r)$, *for* $r \geq 0$,

(b) $W^{(k)}(r/(1-r)) \overset{d}{=} B^{(k)}(r)/(1-r)$, *for* $r \in (0,1)$.

Proof. Let $W_j(r)$ and $B_j(r)$ be the j-th 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 $F(W_j(r)) = W_j(r) - \int_0^r z^{-1} W_j(z)\,\mathrm{d}z$ are Gaussian with zero mean, it remains to show that their covariance functions coincide. Let w.l.o.g. $r \leq s$. Then,

$$
\begin{aligned}
&E[F(W_j(r))F(W_j(s))] - E[W_j(r)W_j(s)] \\
&= \int_0^r \int_0^s \frac{E[W_j(z_1)W_j(z_2)]}{z_1 z_2}\,\mathrm{d}z_2\,\mathrm{d}z_1 - \int_0^s \frac{E[W_j(r)W_j(z_2)]}{z_2}\,\mathrm{d}z_2 - \int_0^r \frac{E[W_j(s)W_j(z_1)]}{z_1}\,\mathrm{d}z_1 \\
&= (2r + r\ln(s) - r\ln(r)) - (r + r\ln(s) - r\ln(r)) - r = 0,
\end{aligned}
$$

and (a) has been shown. The second result follows from the fact that both processes are Gaussian with zero mean and

$$
E\left[\frac{B_j(r)}{1-r}\frac{B_j(s)}{1-s}\right] = \frac{\min\{r(1-s), s(1-r)\}}{(1-r)(1-s)} = \min\left\{\frac{r}{1-r}, \frac{s}{1-s}\right\} = E\big[W_j(\tfrac{r}{1-r})W_j(\tfrac{s}{1-s})\big].
$$

□

Lemma 4. *Let* $\{(x_t, u_t)\}_{t\in\mathbb{N}}$ *satisfy Assumption 1, let* $\beta_t = \beta_0$ *for all* $t \in \mathbb{N}$*, and let* $m \in (0,\infty)$*. Then, as* $T \to \infty$,

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

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

Proof. From Lemma 2, we have $\sup_{r\in[0,m]} T^{-1/2}\|X_{\lfloor rT \rfloor} - (Y_{\lfloor rT \rfloor} - Z_{\lfloor rT \rfloor})\| = o_P(1)$. Let $F(Y_{\lfloor rT \rfloor}) = Y_{\lfloor rT \rfloor} - \int_0^r z^{-1} Y_{\lfloor zT \rfloor}\,\mathrm{d}z$. Then, $\lim_{T\to\infty} \|(Y_{\lfloor rT \rfloor} - Z_{\lfloor rT \rfloor}) - F(Y_{\lfloor rT \rfloor}))\| = 0$, and $\sup_{r\in[0,m]} \|T^{-1/2}X_{\lfloor rT \rfloor} - F(T^{-1/2}Y_{\lfloor rT \rfloor})\| = o_P(1)$. Lemma 1(a) and the continuous mapping theorem imply $F(T^{-1/2}Y_{\lfloor rT \rfloor}) \Rightarrow F(\sigma C^{-1/2}W^{(k)}(r)) = \sigma C^{-1/2}F(W^{(k)}(r))$. Furthermore, from Lemma 3, it follows that $F(W^{(k)}(r)) \overset{d}{=} W^{(k)}(r)$. Consequently, $T^{-1/2}X_{\lfloor rT \rfloor} \Rightarrow \sigma C^{1/2}W^{(k)}(r)$. □

Lemma 5. *Let $\|\cdot\|_M$ be the induced matrix norm of $\|\cdot\|$. 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 $m\in(0,\infty)$. Then, as $T\to\infty$,*

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

Proof. By the application of Abel's formula of summation by parts, which is given in (10), 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

$$\sup_{r\in[0,m]}\|h(r)\| = O(1), \qquad \sup_{r\in[0,m]}\Big\|\sum_{t=1}^{\lfloor rT\rfloor-1}\frac{t}{T}(h(\tfrac{t}{T})-h(\tfrac{t+1}{T}))\Big\| = O(1).$$

Consequently,

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

and

$$\begin{aligned}&\sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t}(A_j-A)(h(\tfrac{t}{T})-h(\tfrac{t+1}{T}))\Big\|\\ &\le \sup_{r\in[0,m]}\sum_{t=1}^{\lfloor rT\rfloor-1}\frac{t}{T}\Big\|\frac{1}{t}\sum_{j=1}^{t}(A_j-A)\Big\|_M\Big\|h(\tfrac{t}{T})-h(\tfrac{t+1}{T})\Big\| = o_P(1).\end{aligned}$$

Then, by the triangle inequality, the assertion follows. □

A.2 Proof of Theorem 1

Let $w_t^* = f_t^{-1}(y_t^* - x_t'\widehat{\beta}_{t-1}^*)$, which are recursive residuals from a regression without any structural break, where $f_t = (1+(t-1)^{-1}x_t'C_{t-1}^{-1}x_t)^{1/2}$,

$$y_t^* = x_t'\beta_0 + u_t, \quad \text{and} \quad \widehat{\beta}_{t-1}^* = \Big(\sum_{j=1}^{t-1}x_jx_j'\Big)^{-1}\Big(\sum_{j=1}^{t-1}x_jy_j^*\Big).$$

Then, $y_t = x_t'\beta_t + u_t = y_t^* + T^{-1/2}x_t'g(t/T)$, and

$$\widehat{\beta}_{t-1} = \widehat{\beta}_{t-1}^* + \frac{1}{\sqrt{T}(t-1)}C_{t-1}^{-1}\sum_{j=1}^{t-1} x_j x_j' g(j/T).$$

Furthermore, $w_t = w_t^* + f_t^{-1}T^{-1/2}x_t'g(t/T) - f_t^{-1}T^{-1/2}(t-1)^{-1}C_{t-1}^{-1}\sum_{j=1}^{t-1} x_j x_j' g(j/T)$. We can decompose the partial sum process as $T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor} x_t w_t = \mathcal{S}_{1,T}(r) + \mathcal{S}_{2,T}(r) + \mathcal{S}_{3,T}(r)$, where

$$\mathcal{S}_{1,T}(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} x_t w_t^*, \quad \mathcal{S}_{2,T}(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} f_t^{-1} x_t x_t' g(\tfrac{t}{T}), \tag{11}$$

$$\mathcal{S}_{3,T}(r) = -\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} \frac{1}{f_t(t-1)} x_t x_t' C_{t-1}^{-1} \sum_{j=1}^{t-1} x_j x_j' g(\tfrac{j}{T}). \tag{12}$$

Let $\|\cdot\|_M$ be the induced matrix norm of $\|\cdot\|$. Lemma 4 yields $\mathcal{S}_{1,T}(r) \Rightarrow \sigma C^{1/2} W^{(k)}(r)$. For the second term, note that, from Assumption 1 and the fact that $\sqrt{T}(f_T^{-1} - 1) = O_P(1)$, it follows that

$$\sup_{r\in[0,m]} \Big\| \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} x_t x_t' - C) \Big\|_M = o_P(1). \tag{13}$$

Since $g(r)$ is piecewise constant and therefore of bounded variation, Lemma 5 yields

$$\sup_{r\in[0,m]} \Big\| \mathcal{S}_2(r) - \int_0^r Cg(s)\,\mathrm{d}s \Big\| = \sup_{r\in[0,m]} \Big\| \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} x_t x_t' - C) g(\tfrac{t}{T}) \Big\| = o_P(1). \tag{14}$$

For the third term, let

$$p_1(r) = \frac{1}{\lfloor rT \rfloor} C_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} x_j x_j' g(\tfrac{j}{T}), \quad p_2(r) = \frac{1}{\lfloor rT \rfloor} C_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} C g(\tfrac{j}{T}),$$

$$p_3(r) = \frac{1}{\lfloor rT \rfloor} \sum_{j=1}^{\lfloor rT \rfloor} g(\tfrac{j}{T}).$$

From Assumption 1, it follows that $\sup_{r\in[0,m]} \|p_2(r) - p_3(r)\|_M = o_P(1)$. Furthermore, from Lemma 5 and from the fact that $\sup_{r\in[0,m]} \| \frac{1}{\lfloor rT \rfloor}\sum_{t=1}^{\lfloor rT \rfloor}(x_t x_t' - C)\|_M = o_P(1)$, it follows that $\sup_{r\in[0,m]} \|p_1(r) - p_2(r)\| = o_P(1)$. Thus, $\sup_{r\in[0,m]} \|p_1(r) - p_3(r)\| = o_P(1)$. Consequently,

$$\begin{aligned} &\sup_{r\in[0,m]} \Big\| \mathcal{S}_{3,T}(r) + \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} f_t^{-1} x_t x_t' h_3(\tfrac{t-1}{T}) \Big\| \\ &\leq \sup_{r\in[0,m]} \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} \|f_t^{-1} x_t x_t'\|_M \|p_1(\tfrac{t-1}{T}) - p_3(\tfrac{t-1}{T})\|, \end{aligned} \tag{15}$$

which is $o_P(1)$. Since p_3 is a partial sum of a piecewise constant function, it is of bounded variation, and, together with (13), we can apply Lemma 5. Then,

$$\sup_{r\in[0,m]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} x_t x_t' - C) p_3(\tfrac{t-1}{T}) \right\| = o_P(1),$$

which yields

$$\begin{aligned}
&\sup_{r\in[0,m]} \left\| \mathcal{S}_{3,T}(r) + \int_0^r \int_0^s \frac{1}{s} C g(v) \, \mathrm{d}v \, \mathrm{d}s \right\| \\
&= \sup_{r\in[0,m]} \left\| \mathcal{S}_{3,T}(r) + \frac{1}{T} C \sum_{t=1}^{\lfloor rT \rfloor} p_3(\tfrac{t-1}{T}) \right\| + o_P(1) = o_P(1).
\end{aligned}$$

Finally, Slutsky's theorem implies that $\mathcal{S}_{1,T}(r)+\mathcal{S}_{2,T}(r)+\mathcal{S}_{3,T}(r) \Rightarrow \sigma C^{1/2} W^{(k)}(r) + \sigma C h(r)$, which yields

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

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

A.3 Proof of Theorem 2

Theorem 1 and the continuous mapping theorem imply that

$$\begin{aligned}
\mathcal{BQ}_T &= \sup_{r\in(0,1)} \frac{\|Q_T(1) - Q_T(r)\|}{d(1-r)} \\
&\xrightarrow{d} \sup_{r\in(0,1)} \frac{\|W^{(k)}(1) + C^{1/2} h(1) - W^{(k)}(r) - C^{1/2} h(r)\|}{d(1-r)} \\
&\stackrel{d}{=} \sup_{r\in(0,1)} \frac{\|W^{(k)}(r) + C^{1/2}(h(1) - h(1-r))\|}{d(r)}.
\end{aligned}$$

A.4 Proof of Theorem 3

Analogously to the proof of Theorem 2, Theorem 1 and the continuous mapping theorem imply that

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

and

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

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

A.5 Proof of Theorem 4

Lemma 2 yields

$$\sup_{t\geq T} \frac{\| \sum_{j=1}^t x_j w_j - \sum_{j=1}^t (x_j u_j - j^{-1} \sum_{i=1}^j x_i u_i)\|}{\sqrt{t}} = o_P(1).$$

Let $W^{(k)}(r)$ be the k-dimensional standard Brownian motion given by Lemma 1(b). Then,

$$A_T = \sup_{t\geq T} \frac{\| \sum_{j=1}^t x_j u_j - \sigma C^{1/2} W^{(k)}(t)\|}{\sqrt{t}} = o_P(1),$$

Furthermore, $\| \sum_{j=1}^t x_t u_t - W^{(k)}(t)\| \leq \xi t^{1/2-\epsilon}$, for some $\epsilon > 0$ and some random variable ξ, for all $t \in \mathbb{N}$. It follows that

$$\sup_{t\geq T} \frac{\|(\sum_{j=1}^t x_j u_j - j^{-1} \sum_{i=1}^j x_i u_i) - \sigma C^{1/2}(W^{(k)}(t) - \sum_{j=1}^t j^{-1} W^{(k)}(j))\|}{\sqrt{t}}$$

$$\leq A_T + \sup_{t\geq T} \sum_{j=1}^t \frac{\| \sum_{i=1}^j x_i u_i - W^{(k)}(j)\|}{j\sqrt{t}} \leq A_T + \xi \cdot \left(\sup_{t\geq T} \sum_{j=1}^t \frac{j^{1/2-\epsilon}}{j\sqrt{t}} \right) = o_P(1),$$

since

$$\sup_{t\geq T} \sum_{j=1}^t \frac{j^{1/2-\epsilon}}{j\sqrt{t}} \leq \sup_{t\geq T} \sum_{j=1}^t \frac{1}{j^{1+\epsilon} T^\epsilon} \leq \frac{1}{T^\epsilon} \sum_{j=1}^\infty \frac{1}{j^{1+\epsilon}} = o_P(1).$$

Consequently,

$$\sup_{t\geq T} \frac{\| \sum_{j=1}^t x_j w_j - \sigma C^{-1/2}(W^{(k)}(t) - \sum_{j=1}^t j^{-1} W^{(k)}(j))\|}{\sqrt{t}} = o_P(1).$$

From the fact that $T^{-1/2} W^{(k)}(t) \stackrel{d}{=} W^{(k)}(t/T)$ it follows that there exists some k-dimensional standard Brownian motion $W_1^{(k)}(t)$, such that

$$\sup_{r\geq 1} \frac{\|T^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} x_j w_j - \sigma C^{-1/2}(W_1^{(k)}(r) - \sum_{j=1}^{\lfloor rT \rfloor} j^{-1} W_1^{(k)}(j/T))}{\sqrt{t}} = o_P(1).$$

Moreover, from Lemma 3 and $\lim_{T\to\infty}\sum_{j=1}^{\lfloor rT\rfloor} j^{-1}W_1^{(k)}(j/T) = \int_0^r z^{-1}W_1^{(k)}(z)\,\mathrm{d}z$, there exists some k-dimensional standard Brownian motion $W_2^{(k)}(t)$, such that

$$\sup_{r\geq 1}\frac{\|T^{-1/2}\sum_{j=1}^{\lfloor rT\rfloor} x_j w_j - \sigma C^{1/2}W_2^{(k)}(r)\|}{\sqrt{r}} = o_P(1),$$

and, therefore,

$$\sup_{r\geq 1}\frac{\|\sigma^{-1}C^{-1/2}T^{-1/2}\sum_{j=1}^{\lfloor rT\rfloor} x_j w_j - W_2^{(k)}(r)\|}{\sqrt{r}} = o_P(1).$$

Since $\hat{\sigma}$ is consistent for σ (see Krämer et al. 1988) and $\{x_t\}_{t\in\mathbb{N}}$ is ergodic, we have

$$\|\hat{\sigma}^{-1}C_T^{-1/2} - \sigma^{-1}C^{-1/2}\|_M = o_P(1),$$

where $\|\cdot\|_M$ denotes the matrix norm induced by $\|\cdot\|$. Consequently,

$$\sup_{r\geq 1}\frac{\|Q_T(r) - W_2^{(k)}(r)\|}{\sqrt{r}} = o_P(1).$$

A.6 Proof of Theorem 5

For the first result, Theorem 4 and Assumption 2 imply

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

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

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

We transform the supremum to a supremum over a subset of the unit interval. Consider the bijective function $g:(0,1)\to(0,\infty)$ that is given by $g(\eta)=\eta/(1-\eta)$. Furthermore, note that $W^{(k)}(g(\eta)) \stackrel{d}{=} B^{(k)}(\eta)/(1-\eta)$, which follows from Lemma 3. Consequently,

$$\begin{aligned}
&\sup_{r\in(1,\infty)} \frac{\|W^{(k)}(r) - W^{(k)}(1)\|}{d(r-1)} \stackrel{d}{=} \sup_{r\in(0,\infty)} \frac{\|W^{(k)}(r)\|}{d(r)} \\
&= \sup_{\eta\in(0,1)} \frac{\|W^{(k)}(g(\eta))\|}{d(g(\eta))} \stackrel{d}{=} \sup_{\eta\in(0,1)} \frac{\|B^{(k)}(\eta)\|}{(1-\eta)d\left(\frac{\eta}{1-\eta}\right)}.
\end{aligned}$$

For the second result, Theorem 2 and Assumption 2 imply

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

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 from above. With Lemma 3(b), we have

$$\begin{aligned}
&\sup_{r\in(1,\infty)} \sup_{s\in(1,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r-s)} \stackrel{d}{=} \sup_{r\in(0,\infty)} \sup_{s\in(0,r)} \frac{\|W^{(k)}(r) - W^{(k)}(s)\|}{d(r-s)} \\
&= \sup_{\eta\in(0,1)} \sup_{s\in(0,g(\eta))} \frac{\|W^{(k)}(g(\eta)) - W^{(k)}(s)\|}{d(g(\eta)-s)} \\
&= \sup_{\eta\in(0,1)} \sup_{\zeta\in(0,\eta)} \frac{\|W^{(k)}(g(\eta)) - W^{(k)}(g(\zeta))\|}{d(g(\eta)-g(\zeta))} \\
&\stackrel{d}{=} \sup_{\eta\in(0,1)} \sup_{\zeta\in(0,\eta)} \frac{\|B^{(k)}(\eta)/(1-\eta) - W^{(k)}(\zeta)/(1-\zeta)\|}{d\left(\frac{\eta}{1-\eta} - \frac{\zeta}{1-\zeta}\right)} \\
&= \sup_{\eta\in(0,1)} \sup_{\zeta\in(0,r)} \frac{\|(1-\zeta)B^{(k)}(\eta) - (1-\eta)B^{(k)}(\zeta)\|}{(1-\eta)(1-\zeta)d\left(\frac{\eta-\zeta}{(1-\eta)(1-\zeta)}\right)}.
\end{aligned}$$

A.7 Proof of Theorem 6

Adopting the notation of the local break in Theorem 1, we have $\beta_t = \beta_0 + T^{-1/2}g(t/T)$ with $g(t/T) = T^{1/2}\delta 1_{\{t\geq T^*\}}$. Unlike in Theorem 1, the alternative does not converge to the null as the sample size grows. Following equations (11)–(15), we have

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

where $\sup_{r\in[0,1]}\|T^{-1/2}\mathcal{S}_{1,T}(r)\| = o_P(1)$, and

$$\sup_{r\in[0,1]}\left\|\mathcal{S}_{2,T}(r)+\mathcal{S}_{3,T}(r) - C\Big(\int_0^r g^*(z)\,\mathrm{d}z - \int_0^r\int_0^z \frac{1}{z}g^*(v)\,\mathrm{d}v\,\mathrm{d}z\Big)\right\| = o_P(1),$$

where $g^*(r) = \delta 1_{\{r\geq\tau^*\}}$. Note that

$$\begin{aligned}\int_0^r g^*(z)\,\mathrm{d}z - \int_0^r\int_0^z \frac{1}{z}g^*(v)\,\mathrm{d}v\,\mathrm{d}z &= \delta\int_0^r\Big(1_{\{s\geq\tau^*\}} - \int_0^s\frac{1}{s}1_{\{v\geq\tau^*\}}\Big)\,\mathrm{d}s\\ = \delta\int_{\tau^*}^r\Big(1-\frac{s-\tau^*}{s}\Big)\,\mathrm{d}s &= \delta\int_{\tau^*}^r\frac{1}{s}\,\mathrm{d}s = \tau^*\delta\big(\ln(r)-\ln(\tau^*)\big)1_{\{r\geq\tau^*\}},\end{aligned}$$

which implies that $\sigma T^{-1/2}Q_T(r)\Rightarrow \tau^* C^{1/2}\delta\big(\ln(r)-\ln(\tau^*)\big)1_{\{r\geq\tau^*\}}$. Then,

$$\begin{aligned}\widehat{\tau}_{\text{ret}} &= \frac{1}{T}\cdot\underset{1\leq t\leq T}{\operatorname{argmax}}\left\|\frac{\widehat{\sigma}_T\sqrt{T}}{\sqrt{T-t+1}}\big(Q_T(1)-Q_T(\tfrac{t+1}{T})\big)\right\|,\\ \widehat{\tau}_{\text{mon}} &= \frac{1}{T}\cdot\underset{T<t\leq T_d}{\operatorname{argmax}}\left\|\frac{\widehat{\sigma}_T\sqrt{T_d}}{\sqrt{T_d-t+1}}\big(Q_{T_d}(1)-Q_{T_d}(\tfrac{t+1}{T_d})\big)\right\|,\end{aligned}$$

and $\sup_{r\in[0,1]}\|Q_{T_d}(r)-Q_T(r\tau_d)\| = o_P(1)$, where $\tau_d = T_d/T$. If $r\in[\tau^*,1)$, the continuous mapping theorem yields

$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{\tau}_{\text{ret}} &= \underset{0<r<1}{\operatorname{argsup}}\frac{1}{\sqrt{1-r}}\Big(\big(\ln(1)-\ln(\tau^*)\big)1_{\{1\geq\tau^*\}} - \big(\ln(r)-\ln(\tau^*)\big)1_{\{r\geq\tau^*\}}\Big)\\ &= \underset{0<r<1}{\operatorname{argsup}}\frac{1}{\sqrt{1-r}}\big(-\ln(r)1_{\{r\geq\tau^*\}} - \ln(\tau^*)1_{\{r<\tau^*\}}\big) = \tau^*,\end{aligned}$$

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 $\tau^*\in(1,\tau_d]$,

$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{\tau}_{\text{mon}} &= \underset{1<r<\tau_d}{\operatorname{argsup}}\frac{1}{\sqrt{\tau_d-r}}\Big(\big(\ln(\tau_d)-\ln(\tau^*)\big)1_{\{\tau_d\geq\tau^*\}} - \big(\ln(r)-\ln(\tau^*)\big)1_{\{r\geq\tau^*\}}\Big)\\ &= \underset{1<r<\tau_d}{\operatorname{argsup}}\frac{1}{\sqrt{\tau_d-r}}\big(\ln(\tau_d)-\ln(r)1_{\{r\geq\tau^*\}} - \ln(\tau^*)1_{\{r<\tau^*\}}\big) = \tau^*.\end{aligned}$$

References

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

Andrews, D. W. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 61(4):821–856.

Aue, A. and Horváth, L. (2013). Structural Breaks in Time Series: Structural Breaks in Time Series. *Journal of Time Series Analysis*, 34(1):1–16.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal*, 9(3):373–403.

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

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

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Probability and Statistics Section. Wiley, New York, 2nd ed edition.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for Testing the Constancy of Regression Relationships Over Time. *Journal of the Royal Statistical Society. Series B*, 37(2):149–192.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64:1045–65.

Dette, H. and Gösmann, J. (2019). A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, 0:1–17.

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

Gösmann, J., Kley, T., and Dette, H. (2019). A new approach for open-end sequential change point monitoring. *https://arxiv.org/abs/1906.03225*.

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

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

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

Kirch, C. and Kamgaing, J. T. (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., Major, P., and Tusnády, G. (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., Ploberger, W., and Alt, R. (1988). Testing for Structural Change in Dynamic Models. *Econometrica*, 56(6):1355–1369.

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

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1):100–115.

Perron, P. (2006). Dealing with Structural Breaks. In *Palgrave Handbook of Econometrics*, volume 1 of *1*, pages 278–352.

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

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

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

Sen, P. K. (1982). Invariance principles for recursive residuals. *The 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.

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

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

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