Backward CUSUM for Testing and Monitoring Structural Change*

Sven Otto** University of Cologne
Jörg Breitung University of Cologne

May 21, 2019

Abstract

It is well known that the conventional CUSUM test suffers from low power and large detection delay. We therefore propose two alternative detector statistics. The backward CUSUM detector sequentially cumulates the recursive residuals in reverse chronological order, whereas the stacked backward CUSUM detector considers a triangular array of backward cumulated residuals. Accordingly, the stacked backward CUSUM detector can be monitored on-line, while the backward CUSUM detector is only suitable for retrospective testing. We derive the limiting distributions of the maximum statistics under suitable sequences of alternatives. The distributions are obtained for retrospective testing, fixed endpoint monitoring, and infinite horizon monitoring. In the retrospective testing context, the local power of the tests is shown to be substantially higher than for the conventional CUSUM test if a single break occurs after one third of the sample size. 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.

Keywords: structural breaks, recursive residuals, sequential tests, change-point detection, local power, local delay

*We are thankful to Holger Dette, Josua Gösmann, Dominik Wied, David Hendry, Juan Carlos Escanciano, and Miguel A. Delgado for very helpful comments and suggestions. Further, we would like to thank the participants of the RMSE meeting 2018 in Vallendar, the econometrics research seminar at the UC3M in Madrid, and the DAGStat Conference 2019 in Munich.

**Corresponding author: Sven Otto, University of Cologne, Institute of Econometrics and Statistics, Albertus-Magnus-Platz, 50923 Köln, Germany. Tel.: +49-221-470-6186. Mail: sven.otto@statistik.uni-koeln.de.

1 Introduction

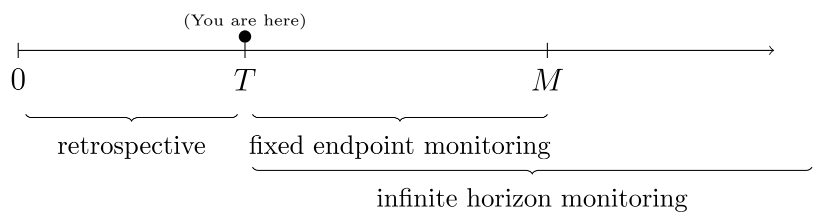
Cumulative sums have become a standard statistical tool for testing and monitoring structural changes in time series models. The CUSUM test was introduced by Brown et al. (1975) as a structural break test for the coefficient vector in the linear regression model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + u_t$ with time index t. Under the null hypothesis, there is no structural change, such that $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $1 \leq t \leq T$, while, under the alternative hypothesis, the coefficient vector changes at unknown time T^*, where $1 < T^* \leq T$.

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. Krämer et al. (1988) investigated the CUSUM test of Brown et al. (1975) under a more general setting. The MOSUM tests by Bauer and Hackl (1978) and Chu et al. (1995) are based on a moving time window of fixed length. A CUSUM test statistic that cumulates OLS residuals was proposed by Ploberger and Krämer (1992), and Ploberger et al. (1989) presented a fluctuation test based on a sequence of OLS estimates. Kuan and Hornik (1995) studied generalized fluctuation tests. Andrews et al. (1993) and Andrews and Ploberger (1994) proposed a sup-Wald test, and the tests by Nyblom (1989) and Hansen (1992) consider maximum likelihood scores instead of residuals.

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 $\{1, 2, \ldots, T\}$. It is assumed that there is no structural change within the historical time period. The monitoring time span with $t > T$ can either have a fixed endpoint $M < \infty$ or an infinite horizon (see Figure 1). In case of a fixed endpoint, the monitoring period starts at $T+1$ and stops at M, such that $T+1 \leq t \leq M < \infty$. In case of an infinite horizon, the monitoring period has infinite length, such that $M \to \infty$. The null hypothesis of no structural change is rejected whenever the path of the detector crosses some critical boundary function for the first time. Leisch et al. (2000) and Zeileis et al. (2005) proposed CUSUM-based monitoring procedures for a fixed endpoint, whereas Chu et al. (1996), Horváth et al. (2004), and Aue et al. (2006) 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 typically 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 $y_t = \beta + u_t$, with β and $\sigma^2 = Var(u_t)$ assumed to be known. We are interested in testing the hypothesis, that at some time period T^*, the mean changes to some unknown value $\beta^* > 0$. To test this hypothesis, we introduce the dummy variable D_t^*, which is unity for $t \geq T^*$ and zero elsewhere. For this one-sided testing problem, there exists a uniform most powerful (UMP) test statistic, which is the t-statistic of the hypothesis $\delta = 0$ in the regression $(y_t - \beta) = \delta D_t^* + u_t$:

$$\tau_{T^*} = \frac{1}{\sigma\sqrt{T - T^* - 1}} \sum_{t=T^*}^{T} (y_t - \beta).$$

If β 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^*. Note that if T^* is unknown, the test statistic is computed for all values of T^*, whereas the starting point of the backward cumulative sum T 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 β 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 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 approach 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 be powerful at the end of the sample.

Furthermore, the conventional CUSUM test has no power against alternatives that do not affect the unconditional mean of y_t. In order to obtain tests that have power against breaks of this kind, we extend the existing invariance principle for recursive residuals to a multivariate version and consider a vector-valued CUSUM process instead of the univariate CUSUM detector. For both retrospective testing and monitoring, we propose a vector-valued sequential statistic in the fashion of the score-based cumulative sum statistic of Nyblom (1989) and Hansen (1992). The application of a vector norm then yields a detector and a sequential test, that has power against a much larger class of structural breaks.

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 forward CUSUM, 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. In Section 4, the local powers of the tests are compared. 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). Furthermore, simulated critical values as well as Monte Carlo simulation results are presented. Finally, Section 5 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 $\boldsymbol{x}_t = (1, x_{t2}, \ldots, x_{tk})'$ is the vector of regressor variables, and y_t is the dependent variable. The $k \times 1$ vector of regression coefficients $\boldsymbol{\beta}_t$ depends on the time index t, and u_t is an error term. Let $\{(y_t, \boldsymbol{x}_t')', \ 1 \leq t \leq T\}$ be the set of historical observations, such that the time point T divides the time horizon into the retrospective time period $1 \leq t \leq T$ and the monitoring period $t > T$. We impose the following assumptions on the regressors and the error term, which are common in the literature on CUSUM statistics and also include the case of lagged dependent variables (see e.g. Krämer et al. 1988):

Assumption 1. *(a) Let $\boldsymbol{C}_T = T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'$ be the empirical covariance matrix, and let $\| \cdot \|_M$ denote some matrix norm. Then,* $\operatorname{plim}_{T \to \infty} \|\boldsymbol{C}_T - \boldsymbol{C}\|_M = 0$*, where* $\boldsymbol{C}$

is a positive definite $k \times k$ matrix. Furthermore, there exists some $\delta > 0$ such that $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E|x_{tj}|^{2+\delta} < \infty$ for all $j = 2, \ldots, k$.

(b) Let $\mathcal{F}_t$ be the σ-algebra generated by $\{(\boldsymbol{x}_{i+1}', u_i)',\ i \leq t\}$. The error process $\{u_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t$, where $E[u_t|\mathcal{F}_{t-1}] = 0$ and $E[u_t^2|\mathcal{F}_{t-1}] = \sigma^2$ with $0 < \sigma^2 < \infty$.

Recursive residuals for linear regression models were introduced by Brown et al. (1975) as standardized one-step ahead forecast errors. Let $\widehat{\boldsymbol{\beta}}_{t-1} = \big(\sum_{i=1}^{t-1} \boldsymbol{x}_i \boldsymbol{x}_i'\big)^{-1}\big(\sum_{i=1}^{t-1} \boldsymbol{x}_i y_i\big)$ be the OLS estimator at time $t-1$. The recursive residuals are given by

$$w_t = \frac{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_{t-1}}{\sqrt{1 + \boldsymbol{x}_t'(\sum_{i=1}^{t-1} \boldsymbol{x}_i \boldsymbol{x}_i')^{-1}\boldsymbol{x}_t}}, \quad t \geq k+1,$$

and $w_t = 0$ for $t = 1, \ldots, k$.

For testing against structural changes in the regression coefficient vector, Brown et al. (1975) introduced the sequential statistic $Q_t = (\widehat{\sigma}^2 T)^{-1/2} \sum_{j=1}^{t} w_j$ for $t = 1, \ldots, T$, where $\widehat{\sigma}^2$ is a consistent estimator for σ^2. In the monitoring context, Chu et al. (1996) considered the detector statistic $Q_t - Q_T$ for $t > T$. The limiting behavior of the underlying empirical process has been thoroughly analyzed in the literature. Under $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, Sen (1982) showed that $Q_{\lfloor rT \rfloor} = (\widehat{\sigma}^2 T)^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} w_j$ converges weakly and uniformly to a standard Brownian motion $W(r)$. Ploberger and Krämer (1990) studied local alternatives of the form $H_1 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0 + T^{-1/2}\mathbf{g}(t/T)$, where $\mathbf{g}(r)$ is piecewise constant and bounded. Let $\boldsymbol{\mu} = \lim_{T\to\infty}(\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_k)'$ be the mean regressor, where $\overline{\boldsymbol{x}}_j$ is the sample mean of the j-th component of the regressors, and let

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

The authors showed that $Q_{\lfloor rT \rfloor}$ converges weakly and uniformly to $W(r) + \boldsymbol{\mu}'\mathbf{h}(r)$. As noted by Krämer et al. (1988), if the break vector $\mathbf{g}(r)$ is orthogonal to $\boldsymbol{\mu}$, the limiting distributions under H_0 and H_1 coincide. Hence, if a break in the coefficient vector does not affect the unconditional mean of y_t, then the CUSUM tests of Brown et al. (1975) and Chu et al. (1996) have no power against such an alternative.

Accordingly, we consider a multivariate cumulative sum process of recursive residuals, which is defined as

$$\mathbf{Q}_T(r) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t, \qquad r \in [0, \infty).$$

Following Krämer et al. (1988), the consistent estimator $\widehat{\sigma}^2 = (T-k-1)^{-1}\sum_{j=1}^{T}(w_j - \overline{w})^2$ is considered. Note that $\mathbf{Q}_T(r)$ is a vector of piecewise constant processes, where each component is in the space $D([0,\infty))$ of càdlàg functions on the domain $[0,\infty)$. Hence, $\mathbf{Q}_T(r)$ is an element of the k-fold product space $D([0,\infty))^k = D([0,\infty)) \times \ldots \times D([0,\infty))$. The space is equipped with a suitable metric, and the symbol "$\Rightarrow$" denotes weak convergence of the associated probability measure. A metric that induces a separable topology on $D([0,\infty))^k$ can be defined by a sequence of Skorokhod metrics on each compact interval $[0,n]$, where $n \in \mathbb{N}$ (see Billingsley 1999, p. 168 and p. 244). The result presented below summarizes the limiting behavior of $\mathbf{Q}_T(r)$ for both the retrospective and the monitoring time period under both H_0 and H_1:

Theorem 1. *Let $\{\boldsymbol{x}_t, u_t\}_{t\in\mathbb{N}}$ satisfy Assumption 1, let $\mathbf{g}(r)$ be piecewise constant and bounded, and let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0 + T^{-1/2}\mathbf{g}(t/T)$ for all $t \in \mathbb{N}$. Then, as $T \to \infty$,*

$$\mathbf{Q}_T(r) \Rightarrow \mathbf{W}(r) + \boldsymbol{C}^{1/2}\mathbf{h}(r),$$

where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions and where $\mathbf{h}(r)$ is defined as in (1).

Note that the function $\mathbf{g}(r)$ is constant if and only if $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$. Under H_0, we then obtain $\boldsymbol{C}^{1/2}\mathbf{h}(r) = \mathbf{0}$, and thus $\mathbf{Q}_T(r) \Rightarrow \mathbf{W}(r)$. By contrast, under a local alternative with a non-constant break function $\mathbf{g}(r)$, it follows that $\mathbf{h}(r)$ is non-zero, and, consequently, $\boldsymbol{C}^{1/2}\mathbf{h}(r)$ is non-zero, since $\boldsymbol{C}^{1/2}$ is positive definite. The limiting distributions of $\mathbf{Q}_T(r)$ under both H_0 and H_1 thus coincide only for the trivial case where $\mathbf{g}(r)$ is constant. Therefore, tests that are based on $\mathbf{Q}_T(r)$ have power against a larger class of alternatives than the tests of Brown et al. (1975) and Chu et al. (1996).

3 CUSUM detectors

In this section, we consider sequential tests for both retrospective testing and monitoring that are based on the multivariate CUSUM processes $\mathbf{Q}_T(r)$. The null hypothesis of no structural change in the regression coefficient vector is formulated as $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathcal{T}$, where the testing period is given by

$$\mathcal{T} = \begin{cases} \{1, 2, \ldots T-1, T\} & \text{in the retrospective context,} \\ \{T+1, T+2, \ldots, M-1, M\} & \text{in the fixed endpoint monitoring context,} \\ \{T+1, T+2, \ldots\} & \text{in the infinite horizon monitoring context.} \end{cases}$$

In the monitoring context, the non-contamination assumption $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ is imposed for the historical time period $t = 1, \ldots, T$, and the monitoring period needs to be predefined. The monitoring time span could have either a fixed endpoint $M < \infty$ with $M > T$ or an infinite horizon such that $M \to \infty$. The sequential tests consist of a detector statistic and a critical boundary function, in which 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. While the forward CUSUM detectors for retrospective testing and monitoring are discussed in Subsection 3.1, we introduce the backward CUSUM detector in Subsection 3.2 and the stacked backward CUSUM detectors in Subsection 3.3. Throughout this section, we assume that the boundary function is of the form $b(r) = \lambda_\alpha \cdot d(r)$, where λ_α denotes the critical value, which depends on the significance level α of the test, and $d(r)$ is a continuous and positive function that is bounded away from zero for all $r \geq 0$.

3.1 Forward CUSUM

Following the univariate CUSUM test by Brown et al. (1975), we consider the multivariate retrospective CUSUM detector

$$\mathbf{Q}_{t,T} = \mathbf{Q}_T\left(\tfrac{t}{T}\right) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=1}^{t} \boldsymbol{x}_t w_j, \quad 1 \leq t \leq T.$$

The detector is inspired by the score-based cumulative sum statistic proposed by Hansen (1992), where OLS residuals are considered. While Hansen (1992) proposed averaging the entries of the vector-valued cumulative sum, we consider the maximum vector entry. Let $\|\boldsymbol{a}\| = \max_{i=1,\ldots,k} |a_i|$ be the maximum norm, where $\boldsymbol{a} = (a_1, \ldots, a_k)' \in \mathbb{R}^k$. The null hypothesis is rejected if the path of $\|\mathbf{Q}_{t,T}\|$ exceeds the critical boundary function $b_t = \lambda_\alpha \cdot d\big(t/T\big)$ for at least some time index within the retrospective testing period. The critical value λ_α determines the significance level α such that

$$\lim_{T\to\infty} P\Big(\|\mathbf{Q}_{t,T}\| \geq \lambda_\alpha \cdot d\big(\tfrac{t}{T}\big) \text{ for at least one index } t = 1, \ldots, T \,\Big|H_0\Big) = \alpha.$$

Let $\boldsymbol{\mathcal{M}}_Q^{\text{ret}} = \max_{1\leq t\leq T} \|\mathbf{Q}_{t,T}\|/d\big(t/T\big)$ denote the maximum statistic representation of the CUSUM detector. The above condition can be equivalently expressed as

$$\lim_{T\to\infty} P(\boldsymbol{\mathcal{M}}_Q^{\text{ret}} \geq \lambda_\alpha | H_0) = \alpha.$$

Hence, λ_α is the $(1-\alpha)$ quantile of the limiting null distribution of $\boldsymbol{\mathcal{M}}_Q^{\text{ret}}$, and Theorem 1 together with the continuous mapping theorem yields

$$\boldsymbol{\mathcal{M}}_Q^{\text{ret}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\mathbf{W}(r)\|}{d(r)}$$

under H_0, as $T \to \infty$. Note that $\boldsymbol{\mathcal{M}}_Q^{\text{ret}}$ together with the critical value λ_α defines a one-shot test that is equivalent to the sequential CUSUM test.

For real-time monitoring, we follow Chu et al. (1996) and define the multivariate retrospective CUSUM detector as

$$\mathbf{Q}_{t,T}^{\text{mon}} = \mathbf{Q}_T\big(\tfrac{t}{T}\big) - \mathbf{Q}_T(1) = \frac{1}{\widehat{\sigma}\sqrt{T}} \sum_{j=T+1}^{t} \boldsymbol{x}_t w_j, \quad t > T,$$

and H_0 is rejected if its maximum norm $\|\mathbf{Q}_{t,T}^{\text{mon}}\|$ exceeds the boundary $b_t = \lambda_\alpha \cdot d\big(t/T - 1\big)$ at least once for $t > T$. Let $M = \lfloor mT \rfloor$, such that $1 < m < \infty$ in the fixed endpoint monitoring context and $m \to \infty$ in the infinite horizon context. The corresponding maximum statistics are given by $\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} = \max_{T<t\leq\lfloor mT\rfloor} \|\mathbf{Q}_{t,T}^{\text{mon}}\|/d\big(t/T - 1\big)$ and $\boldsymbol{\mathcal{M}}_{Q,\infty}^{\text{mon}} = \max_{T<t<\infty} \|\mathbf{Q}_{t,T}^{\text{mon}}\|/d\big(t/T - 1\big)$, respectively. From Theorem 1 and the continuous mapping theorem, it follows that

$$\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} \xrightarrow{\mathcal{D}} \sup_{r\in(1,m)} \frac{\|\mathbf{W}(r) - \mathbf{W}(1)\|}{d(r-1)} \overset{\mathcal{D}}{=} \sup_{r\in(0,m-1)} \frac{\|\mathbf{W}(r)\|}{d(r)},$$
$$\boldsymbol{\mathcal{M}}_{Q,\infty}^{\text{mon}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,\infty)} \frac{\|\mathbf{W}(r)\|}{d(r)}.$$

In order to obtain a limiting distribution that includes a supremum over a set of finite length, we consider the bijective function $g : (0, (m-1)/m) \to (0, m-1)$ that is given by $g(\eta) = \eta/(1-\eta)$. Furthermore, note that $\mathbf{W}(g(\eta)) \overset{\mathcal{D}}{=} \mathbf{B}(\eta)/(1-\eta)$, where $\mathbf{B}(r)$ is a vector of k independent standard Brownian bridges. This follows from the fact that both $\mathbf{W}(g(\eta))$ and $\mathbf{B}(\eta)/(1-\eta)$ are Gaussian with mean zero and have the same covariance function. Consequently,

$$\sup_{r\in(0,m-1)} \frac{\|\mathbf{W}(r)\|}{d(r)} = \sup_{\eta\in(0,\frac{m-1}{m})} \frac{\|\mathbf{W}(g(\eta))\|}{d(g(\eta))} \overset{\mathcal{D}}{=} \sup_{\eta\in(0,\frac{m-1}{m})} \frac{\|\mathbf{B}(\eta)\|}{(1-\eta)d\left(\frac{\eta}{1-\eta}\right)}.$$

Hence, under H_0, the maximum statistics for fixed endpoint monitoring and infinite horizon monitoring satisfy

$$\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,\frac{m-1}{m})} \frac{\|\boldsymbol{B}(r)\|}{(1-r)d\left(\frac{r}{1-r}\right)},$$
$$\boldsymbol{\mathcal{M}}_{Q,\infty}^{\text{mon}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\boldsymbol{B}(r)\|}{(1-r)d\left(\frac{r}{1-r}\right)},$$

as $T \to \infty$, which follows from Theorem 1 and the continuous mapping theorem.

While, for one-shot tests, the critical value determines the type I error, for sequential tests, the critical boundary 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 the first boundary crossing under the null hypothesis, which is also referred to as the "distribution of the size" (see Anatolyev and Kosenok 2018). Brown et al. (1975) suggested the linear boundary function

$$b(r) = \lambda_\alpha(1+2r), \tag{2}$$

which is our main benchmark. In this case, the retrospective maximum statistic satisfies

$$\max_{1\le t\le T} \frac{\|\mathbf{Q}_{t,T}\|}{1+2\left(\frac{t}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\mathbf{W}(r)\|}{1+2r}$$

under H_0, as $T \to \infty$, whereas, for the monitoring maximum statistics, we obtain

$$\max_{T+1\le t\le \lfloor mT\rfloor} \frac{\|\mathbf{Q}_{t,T}\|}{1+2\left(\frac{t}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,\frac{m-1}{m})} \frac{\|\boldsymbol{B}(r)\|}{1+r},$$
$$\max_{T+1\le t<\infty} \frac{\|\mathbf{Q}_{t,T}\|}{1+2\left(\frac{t}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\boldsymbol{B}(r)\|}{1+r}.$$

The linear boundary is widely applied in practice, but, as already noted by Brown et al. (1975), the crossing probabilities cannot be constant for all potential relative crossing time points r. The authors argued that it is more natural to consider a boundary that is proportional to the standard deviation of the limiting process. Such a boundary is given by the radical function $b(r) = \lambda_\alpha \sqrt{r}$. As noted by Zeileis (2004), if there is a single break in the middle or at the end of the retrospective sample, there is no power gain using the radical boundary when compared to the linear boundary. Only in cases where a break occurs at the beginning of the sample, some increased power may be observed. Another problem associated with the radical boundary is that it is not bounded away from zero. In order to obtain critical values and avoid size distortions, some trimming at the beginning of the sample in the fashion of the sup-Wald test by Andrews et al. (1993) is necessary. For infinite horizon monitoring, Chu et al. (1996) also considered a boundary function of radical type, which is given by

$$b(r) = \sqrt{(r+1)\ln\left(\tfrac{r+1}{\alpha^2}\right)}. \tag{3}$$

The boundary is based on a result on boundary crossing probabilities for the path of Brownian motions. Robbins and Siegmund (1970) showed that

$$P\Big(|W(r)| \geq \sqrt{(r+1)\ln\left(\tfrac{r+1}{\alpha^2}\right)} \text{ for some } r \geq 0\Big) = \alpha,$$

and the univariate monitoring CUSUM detector together with the radical boundary by Chu et al. (1996) thus yields a sequential test that has size α, as $m \to \infty$. Anatolyev and Kosenok (2018) derived a theoretical boundary that yields a uniformly distributed size. However, their boundary has no closed form solution and is only valid for the univariate retrospective and fixed endpoint monitoring cases. Furthermore, simulations, which are omitted here, indicate that their approximative boundary does indeed yield a uniform size distribution, but that the CUSUM test performs uniformly worse in terms of power compared to the test when using the linear boundary of Brown et al. (1975). Note that in the context of infinite horizon monitoring the size cannot be uniformly distributed.

3.2 Backward CUSUM

An alternative approach is to cumulate the recursive residuals in reversed order. Suppose there is a single break point at time 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 behavior for the pre-break period $t < T^*$, and cumulating those residuals brings nothing but noise to the detector statistic. In contrast, under a structural break, 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

$$\mathbf{BQ}_{t,T} = \mathbf{Q}_T(1) - \mathbf{Q}_T\big(\tfrac{t-1}{T}\big) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=t}^{T} \boldsymbol{x}_t w_j,$$

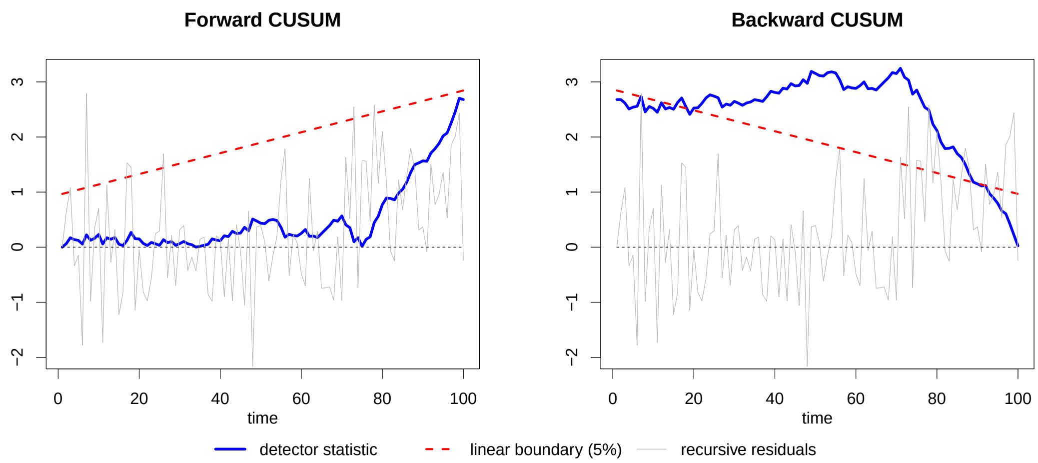
where $1 \leq t \leq T$. The null hypothesis is rejected if the path of $\|\mathbf{BQ}_{t,T}\|$ exceeds the boundary $b_t = \lambda_\alpha \cdot d\big(1 - (t+1)/T\big)$ for at least some time index t. From Theorem 1 and the continuous mapping theorem, it follows that the maximum statistic satisfies

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}} = \max_{1 \leq t \leq T} \frac{\|\mathbf{BQ}_{t,T}\|}{d\big(\frac{T-t+1}{T}\big)} \xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \frac{\|\mathbf{W}(1) - \mathbf{W}(r)\|}{d(1-r)} \overset{\mathcal{D}}{=} \sup_{r \in (0,1)} \frac{\|\mathbf{W}(r)\|}{d(r)}$$

under H_0, as $T \to \infty$. Using the same boundary as for the retrospective CUSUM, the limiting null distributions of their maximum statistics coincide. 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 $\|\mathbf{BQ}_{t,T}\|$ is only defined for $t \leq T$, as its endpoint T is fixed.

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. Since $k = 1$, the detectors are univariate, and the vector norm is simply the absolute value. The bold solid line paths are the trajectories of $|\mathbf{Q}_{t,T}|$ and $|\mathbf{BQ}_{t,T}|$. In the background, the recursive residuals are plotted. The dotted lines shows the linear boundary (2) with $\alpha = 5\%$ and $\lambda_\alpha = 0.948$.

3.3 Stacked backward CUSUM

In order to combine the advantages of the backward CUSUM with the measurability properties of the forward CUSUM for monitoring, we propose the stacked backward CUSUM detector. Let

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t) = \max_{1 \leq s \leq t} \frac{\|\mathbf{Q}_T\big(\frac{t}{T}\big) - \mathbf{Q}_T\big(\frac{s-1}{T}\big)\|}{d\big(\frac{T-t+1}{T}\big)}$$

be the backward CUSUM statistic with endpoint t. The idea is to compute this statistic sequentially for each time point $t = 1, \ldots, T$, yielding $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(1)$, $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(2), \ldots, \boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(T)$. 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 and $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t)$ can thus be adapted for real-time monitoring.

The stacked backward CUSUM detector can be defined as

$$\mathbf{SBQ}_{s,t,T} = \mathbf{Q}_T\big(\tfrac{t}{T}\big) - \mathbf{Q}_T\big(\tfrac{s-1}{T}\big) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=s}^{t} \boldsymbol{x}_t w_j, \quad 1 \leq s \leq t < \infty.$$

Since the upper and the lower summation index of $\mathbf{SBQ}_{s,t,T}$ are both flexible with $s \leq t$, this induces a triangular scheme. H_0 is rejected if $\|\mathbf{SBQ}_{s,t,T}\|$ exceeds $b_{s,t} = \lambda_\alpha \cdot d\big((t-s+1)/T\big)$

for some s and t with $1 \leq s \leq t \leq T$, or, equivalently, if

$$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}} = \max_{1 \leq t \leq T} \boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t) = \max_{1 \leq t \leq T} \max_{1 \leq s \leq t} \frac{\|\mathbf{SBQ}_{s,t,T}\|}{d\left(\frac{t-s+1}{T}\right)}$$

exceeds λ_α. Under H_0, we then obtain

$$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}} \xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \sup_{s \in (0,r)} \frac{\|\mathbf{W}(r) - \mathbf{W}(s)\|}{d(r-s)},$$

as $T \to \infty$, which follows from Theorem 1 and the continuous mapping theorem. The backward CUSUM maximum statistic $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t)$ is itself a sequential statistic, and stacking all those maximum statistics on one another leads to an additional maximum and a double supremum in the limiting distribution. The stacked backward CUSUM uses the recursive residuals in a multiple way such that the set over which the maximum is taken has many more elements than the forward CUSUM and the backward CUSUM. For $t = 1$ only w_1 is cumulated, for $t = 2$ the residuals w_2 and w_1 are cumulated, for $t = 3$ we consider w_3, w_2, and w_1, and so forth.

A similar procedure was proposed by Dette and Gösmann (2019) in the context of likelihood ratio (LR) tests for change point detection. Their detector is given by the maximum of a triangular array of LR statistics, which also leads to a double maximum statistic.

The triangular detector can also be monitored on-line across all the time points $t > T$. The null hypothesis is rejected if $\|\mathbf{SBQ}_{s,t,T}\|$ exceeds the boundary $b_{s,t} = \lambda_\alpha \cdot d\big((t-s+1)/T\big)$ at least once for some indices s and t with $T < s \leq t$. Analogously to the retrospective case, let

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{mon}}(t) = \max_{T < s \leq t} \frac{\|\mathbf{Q}_T\left(\frac{t}{T}\right) - \mathbf{Q}_T\left(\frac{s-1}{T}\right)\|}{d\left(\frac{t-s+1}{T}\right)}$$

be the sequence of backward CUSUM maximum statistics for $t > T$. Its maximum statistic for fixed endpoint monitoring satisfies

$$\boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}} = \max_{T < t \leq \lfloor mT \rfloor} \boldsymbol{\mathcal{M}}_{BQ}^{\text{mon}}(t) \xrightarrow{\mathcal{D}} \sup_{r \in (0,m-1)} \sup_{s \in (0,r)} \frac{\|\mathbf{W}(r) - \mathbf{W}(s)\|}{d(r-s)}$$

under H_0, as $T \to \infty$, which follows from Theorem 1 and the continuous mapping theorem. The limiting distribution can also be formulated with respect to Brownian bridge processes.

Analogously to the Forward CUSUM, let the function $g : (0, (m-1)/m) \to (0, m-1)$ be given by $g(\eta) = \eta/(1-\eta)$. Then,

$$\begin{aligned}
&\sup_{r\in(0,m-1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r) - \mathbf{W}(s)\|}{d(r-s)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{s\in(0,g(\eta))} \frac{\|\mathbf{W}(g(\eta)) - \mathbf{W}(s)\|}{d(g(\eta)-s)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,\eta)} \frac{\|\mathbf{W}(g(\eta)) - \mathbf{W}(g(\zeta))\|}{d(g(\eta)-g(\zeta))} \\
&\overset{\mathcal{D}}{=} \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,\eta)} \frac{\|\mathbf{B}(\eta)/(1-\eta) - \mathbf{W}(\zeta)/(1-\zeta)\|}{d\big(\frac{\eta}{1-\eta} - \frac{\zeta}{1-\zeta}\big)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,r)} \frac{\|(1-\zeta)\mathbf{B}(\eta) - (1-\eta)\mathbf{B}(\zeta)\|}{(1-\eta)(1-\zeta)d\big(\frac{\eta-\zeta}{(1-\eta)(1-\zeta)}\big)}.
\end{aligned}$$

Hence, in the infinite horizon monitoring context, we obtain

$$\boldsymbol{\mathcal{M}}_{SBQ,\infty}^{\text{mon}} = \max_{T<t<\infty} \boldsymbol{\mathcal{M}}_{BQ}^{\text{mon}}(t) \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r) - (1-r)\mathbf{B}(s)\|}{(1-r)(1-s)d\big(\frac{r-s}{(1-r)(1-s)}\big)}$$

under H_0, as $T \to \infty$, and λ_α is equal to its $(1-\alpha)$ quantile. For the linear boundary of Brown et al. (1975), it follows that

$$\max_{T+1\leq t<\infty} \max_{T\leq s\leq t-1} \frac{\|\mathbf{SBQ}_{s,t,T}^{\text{mon}}\|}{1+2(\frac{t-s}{T})} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r) - (1-r)\mathbf{B}(s)\|}{(1-r)(1-s)+2(r-s)}$$

under H_0, as $T \to \infty$.

4 Simulations

In this section, we compare both the asymptotic and finite sample properties of the tests. While in Subsection 4.1 local asymptotic power an local asymptitoc mean delay curves are simulated, we present simulation results on the finite sample size and power in Subsection 4.2. Furthermore, asymptotic critical values for the tests are provided.

4.1 Local asymptotic power and delay

In order to illustrate the advantages of the backward CUSUM and the stacked backward CUSUM tests, we consider the simple model $y_t = \beta_t + u_t$ with a local break in the mean.

Let the mean be given by $\beta_t = \beta^0 + T^{-1/2} g(t/T)$, where $g(r)$ is a piecewise constant and bounded function. Note that in this case the multivariate CUSUM process coincides with the univariate CUSUM process $Q_{\lfloor rT \rfloor}$. Furthermore, note that the covariance matrix $\boldsymbol{C}$ is equal to unity, and the vector norm for $k = 1$ is simply the absolute value. Theorem 1 yields $Q_T(r) \Rightarrow W(r) + h(r)$, where

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

and together with the continuous mapping theorem, it follows that

$$\begin{aligned}
\boldsymbol{\mathcal{M}}_Q^{\text{ret}} &\xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \frac{|W(r) + h(r)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}} &\xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \frac{|W(r) + h(1) - h(1-r)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}} &\xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \sup_{s \in (0,r)} \frac{|W(r) - W(s) + h(r) - h(s)|}{d(r-s)},
\end{aligned}$$

as $T \to \infty$. While, under H_0, the limiting distributions for the retrospective forward CUSUM and the retrospective backward CUSUM coincide, they differ from each other under the alternative. The maximum statistics in the fixed endpoint monitoring case satisfy

$$\begin{aligned}
\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} &\xrightarrow{\mathcal{D}} \sup_{r \in (0,m-1)} \frac{|W(r) + h(r+1) - h(1)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}} &\xrightarrow{\mathcal{D}} \sup_{r \in (0,m-1)} \sup_{s \in (0,r)} \frac{|W(r) - W(s) + h(r+1) - h(s+1)|}{d(r-s)},
\end{aligned}$$

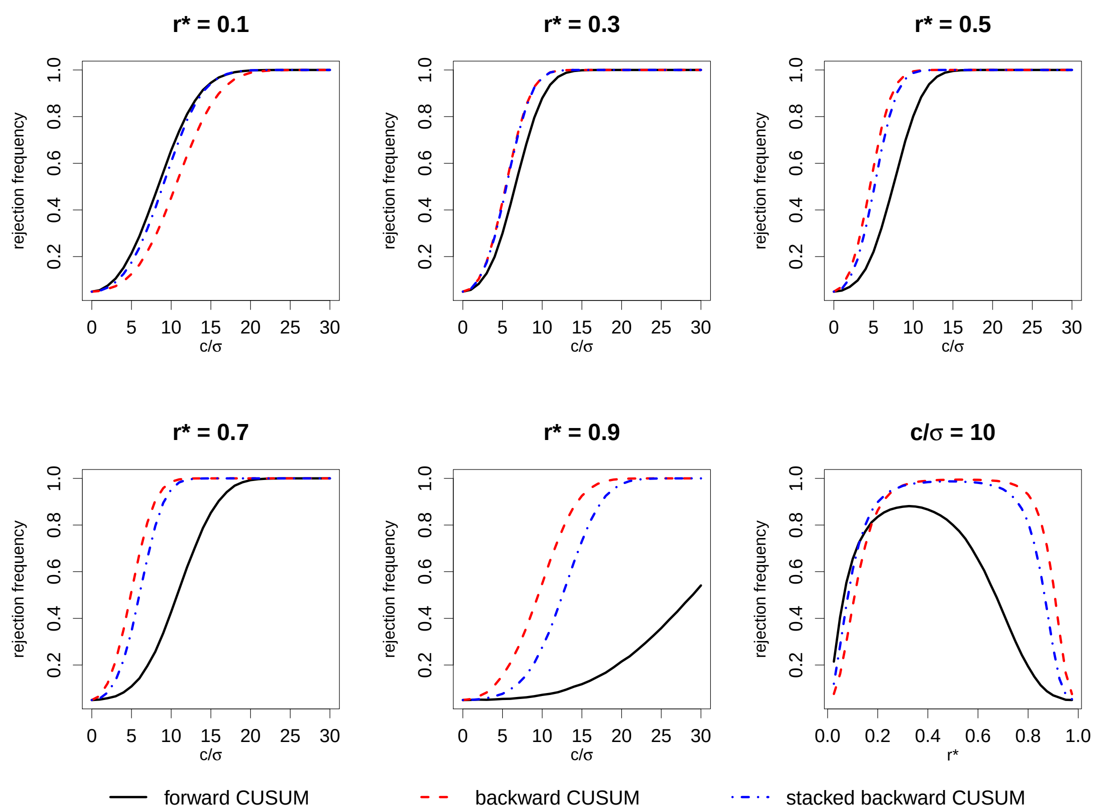
as $T \to \infty$.

Generally, none of the tests can be shown to be uniformly more powerful in comparison to the other tests. However, we can compare the tests under particular alternatives. We consider a single break in the mean, where the break function is given by $g(r) = c \cdot 1_{\{r \geq r^*\}}$ and r^* denotes the break location. Then,

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

Simulated asymptotic local power curves under the limiting distribution at a 5% significance level are presented in Figure 3 for the retrospective case. The Brownian motions are approximated on a grid of 1,000 equidistant points, and the linear boundary $d(r) = 1+2r$ is implemented. The size-adjusted rejection rates are obtained from 100,000 Monte Carlo repetitions for different break locations. 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 backward CUSUM performs best for $r^* > 0.3$, while the stacked backward CUSUM outperforms the other two tests if the break is located at around 1/5 of the sample size.

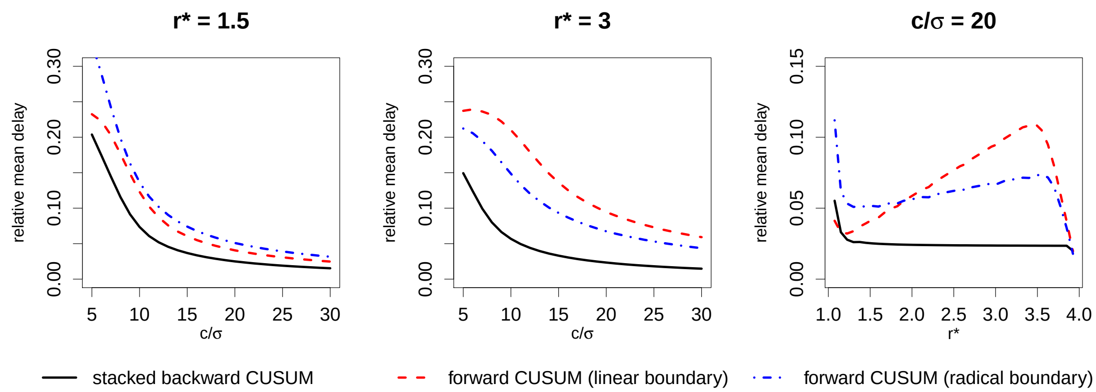
Figure 3: Asymptotic local power curves for retrospective testing



Note: The plots show simulated local power curves. While, for the plots at the top and the first two plots at the bottom, the break location is fixed with $r^* \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and local break sizes c/σ are shown on the x-axis, for the last plot, the local break size is fixed with $c/\sigma = 10$, and the breakpoint locations r^* are given on the x-axis. The linear boundary (2) is implemented for a significance level of $\alpha = 5\%$.

For the monitoring case with fixed endpoint $m = 2$, the local power curves of the forward CUSUM test and the stacked backward CUSUM test have exactly the same shape as in the retrospective case. The monitoring local power curve for a break at $r^* \in (1, 2)$ then coincides with the corresponding retrospective curve in Figure 3 with a single break at $r^* - 1$. Hence, the power of the stacked backward CUSUM is always higher than that of the forward CUSUM if $r^* \geq 1.15$. Because every fixed nontrivial alternative will be detected at some time point in the infinite monitoring context, the delay between the actual break and the detection time point is a much more important performance measure for monitoring detectors than the power itself. Let τ be the stopping time of the time point of the first boundary crossing, and let the mean local relative delay be given by $E\big[\tau/T | r^* \leq \tau/T \leq m\big] - r^*$.

Figure 4: Asymptotic local mean delay curves for monitoring ($m = 4$)

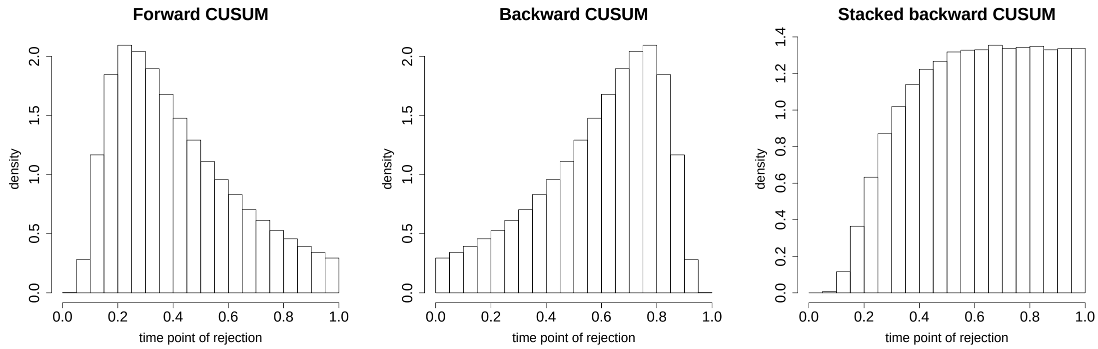


Note: The plots show simulated local mean delay curves, where the relative mean delays are given on the y-axis. While, for the first two plots, the break locations are fixed with $r^* \in \{1.5, 3\}$ and local break sizes c/σ are given on the x-axis, for the last plot, the local break size is fixed with $c/\sigma = 20$, and the breakpoint locations r^* are given on the x-axis. The linear boundary (2) is considered for $\alpha = 5\%$.

Figure 4 presents the simulated mean local relative delay curves for the fixed endpoint $m = 4$ for $\boldsymbol{\mathcal{M}}_{SBQ,4}^{\text{mon}}$ with the linear boundary, for $\boldsymbol{\mathcal{M}}_{Q,4}^{\text{mon}}$ with the linear boundary, and for $\boldsymbol{\mathcal{M}}_{Q,4}^{\text{mon}}$ with the radical boundary by Chu et al. (1996). The mean local relative delay of the stacked backward CUSUM is much lower than that of the forward CUSUM. Furthermore, the mean local relative delay is constant across different break locations, with the exception of breaks that are located at $r^* < 1.15$.

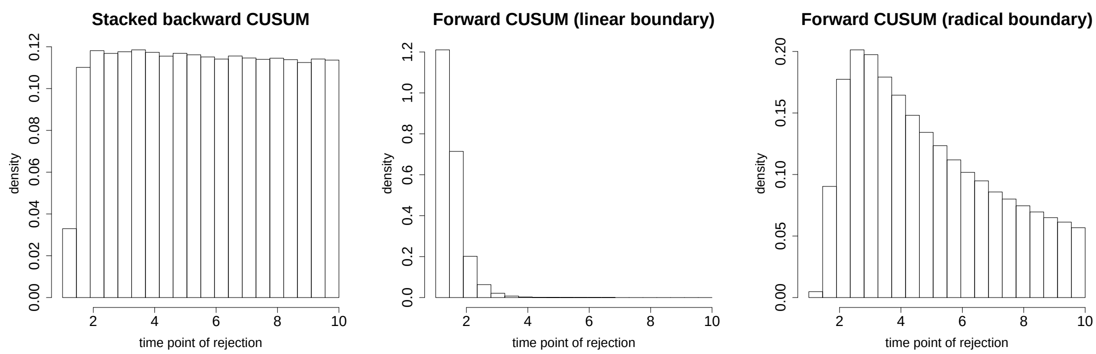
Furthermore, we compare the asymptotic distributions of the size, which is the distribution of the time point of the first boundary crossing under H_0. Figure 5 presents histograms of the asymptotic size distributions for retrospective testing under the linear boundary. For the forward CUSUM, the highest rejection rates under H_0 are obtained at relative locations between 0.15 and 0.4 of the sample. For the backward CUSUM, the

Figure 5: Size distributions of the retrospective detectors



Note: The plots show the frequencies of the location of the first boundary exceedance under the null hypothesis. The frequencies are based on random draws under the limiting distribution of the maximum statistics of the forward CUSUM, the backward CUSUM, and the stacked backward CUSUM detector using the linear boundary in (2) with a significance level of 5% under a model with $k = 1$.

Figure 6: Size distributions of the monitoring detectors ($m = 10$)



Note: The plots show the frequencies of the location of the first boundary exceedance under the null hypothesis. The frequencies are based on random draws under the limiting distribution of the monitoring maximum statistics with $m = 10$. The stacked backward CUSUM detector using the linear boundary, the forward CUSUM detector using the linear boundary, and the forward CUSUM detector using the radical boundary by Chu et al. (1996) are considered at a significance level of 5% under a model with $k = 1$.

picture is mirror-inverted, such that most weight is put on rejections at relative locations between 0.6 and 0.85. The distribution for the CUSUM is right-skewed, whereas, for the backward CUSUM, it is left-skewed. For the stacked backward CUSUM, the distribution is much closer to a uniform distribution, although it is slightly left-skewed. Note that the size distributions provide information about the location of false rejections, but, when comparing Figure 3 with Figure 5, it is reasonable to assume that this is also related to the distribution of the power across different time points. There is no consensus on which distribution should be preferred, as whether one wishes to put more weight on particular

Table 1: Asymptotic critical values for the retrospective tests under the linear boundary

	$\boldsymbol{\mathcal{M}}_{Q}^{\text{ret}}$ and $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}$					$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}}$				
k	20%	10%	5%	2.5%	1%	20%	10%	5%	2.5%	1%
1	0.734	0.847	0.945	1.034	1.143	1.018	1.113	1.198	1.278	1.374
2	0.839	0.941	1.032	1.115	1.219	1.107	1.196	1.277	1.352	1.442
3	0.895	0.993	1.081	1.163	1.260	1.156	1.244	1.321	1.392	1.481
4	0.933	1.029	1.114	1.192	1.287	1.190	1.275	1.350	1.419	1.506
5	0.962	1.056	1.139	1.216	1.307	1.216	1.299	1.372	1.441	1.526
6	0.985	1.077	1.160	1.235	1.323	1.237	1.317	1.388	1.457	1.541
7	1.005	1.095	1.176	1.249	1.338	1.253	1.333	1.404	1.471	1.556
8	1.021	1.110	1.189	1.261	1.349	1.268	1.347	1.418	1.483	1.566

Note: Critical values λ_α are reported for the linear boundary in (2) from 100,000 Monte Carlo repetitions. The Gaussian processes in the limiting distributions are simulated on a grid of 10,000 equidistant points.

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. Figure 6 presents the distributions of the size for the fixed monitoring horizon with $m = 10$. The distribution for the stacked backward CUSUM is much closer to a uniform distribution compared to those of the forward CUSUM variants.

4.2 Critical values and finite sample performance

Table 1 presents critical values for the retrospective case using the linear boundary, while the empirical size results for a significance level of 5% are shown in Table 2. The 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 et al. (1993). The sup-Wald statistic is given by

$$\max_{r \in [r_0, 1-r_0]} T \cdot \frac{S_0 - S_1(r) - S_2(r)}{r(1-r)},$$

where S_0 is the OLS residual sum of squares using observations $1, \ldots, T$, $S_1(r)$ is the OLS residual sum of squares using observations $1, \ldots, \lfloor rT \rfloor$, and $S_2(r)$ is the OLS residual sum of squares using observations $\lfloor rT \rfloor + 1, \ldots, T$. The parameter r_0 defines the lower and

Table 2: Empirical sizes of the retrospective tests

	$k=1$			$k=2$			$k=3$			$k=4$		
T	100	200	500	100	200	500	100	200	500	100	200	500
$\boldsymbol{\mathcal{M}}_{Q}^{\text{ret}}$	3.8	4.2	4.6	4.0	4.4	4.5	4.0	4.4	4.5	4.1	4.3	4.5
$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}$	4.1	4.2	4.6	4.8	4.7	4.6	5.4	4.9	4.6	6.0	5.3	4.7
$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}}$	2.8	3.5	4.2	3.9	4.0	4.2	4.7	4.5	4.2	5.7	4.9	4.4

Note: Simulated rejection rates under H_0 are presented in percentage points. The values are obtained from 100,000 Monte Carlo repetitions using the critical values from Table 1 at a significance level of 5% for the linear boundary (2). The cases $k = 1, \dots, 4$ represent the models $y_t = \beta_1 + u_t$, $y_t = \beta_1 + \beta_2 x_{t2} + u_t$, $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, and $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, respectively, where x_{t2}, x_{t3}, x_{t4}, and u_t are simulated independently as standard normal random variables for all $t = 1, \dots, T$.

upper trimming parameters. In the subsequent simulations, we consider $r_0 = 0.15$, which is the default setting suggested by Andrews et al. (1993). The limiting distribution is given by $\sup_{r \in [r_0, 1-r_0]} \boldsymbol{B}(r)'\boldsymbol{B}(r)/(r(1-r))$, and critical values for different values of r_0 and k are tabulated in Andrews et al. (1993). The author showed that the sup-Wald test has weak optimality properties in the sense that, in the case of a single structural break, its 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 statistic is not suitable for monitoring, since its numerator statistic $T(S_0 - S_1(t/T) - S_2(t/T))$ is not measurable with respect to the filtration of information at time t.

We illustrate the finite sample performance for a simple model with $k = 1$ and a break in the mean, which is given by

$$y_t = \mu_t + u_t, \quad \mu_t = 2 + 0.8 \cdot 1_{\{\frac{t}{T} \geq r^*\}}, \quad u_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{4}$$

and for a univariate linear regression model with a break in the slope coefficient, which is given by

$$y_t = \mu_t + \beta_t x_t + u_t, \quad \mu_t = 2, \quad \beta_t = 1 + 0.8 \cdot 1_{\{\frac{t}{T} \geq r^*\}}, \quad x_t, u_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{5}$$

where $t = 1, \dots, T$. Table 3 presents the size-adjusted power results.

First, we observe that the backward CUSUM and the stacked backward CUSUM outperform the forward CUSUM, except for the case $r^* = 0.1$. Second, while the forward CUSUM test has much lower power than the sup-Wald test, the reversed order cumulation

Table 3: Size-adjusted powers of the retrospective tests for $T = 100$ and $\alpha = 5\%$

	Model (4) ($k = 1$)				Model (5) ($k = 2$)			
	$\mathcal{M}_Q^{\text{ret}}$	$\mathcal{M}_{BQ}^{\text{ret}}$	$\mathcal{M}_{SBQ}^{\text{ret}}$	supW	$\mathcal{M}_Q^{\text{ret}}$	$\mathcal{M}_{BQ}^{\text{ret}}$	$\mathcal{M}_{SBQ}^{\text{ret}}$	supW
$r^* = 0.1$	46.9	28.3	40.7	26.3	32.5	19.0	25.9	21.5
$r^* = 0.2$	63.5	65.0	71.2	73.9	47.2	47.4	51.7	59.3
$r^* = 0.3$	67.1	84.0	83.9	86.8	50.8	70.3	68.1	75.3
$r^* = 0.4$	63.5	91.5	88.7	91.4	47.1	81.9	75.9	82.3
$r^* = 0.5$	54.0	93.8	89.4	92.5	38.2	85.7	77.0	84.3
$r^* = 0.6$	39.4	93.3	86.6	91.4	26.6	84.1	72.0	82.2
$r^* = 0.7$	23.4	89.0	77.0	86.9	15.6	75.5	58.9	75.3
$r^* = 0.8$	11.0	74.2	51.6	74.1	8.2	56.0	37.0	59.5
$r^* = 0.9$	5.5	31.4	12.9	26.2	5.1	24.6	13.3	21.4

Note: Simulated size-adjusted rejection rates under models (4) and (5) are presented in percentage points for a significance level of 5%, where supW denotes the sup-Wald test with $r_0 = 0.15$. The values are obtained from 100,000 Monte Carlo repetitions for a sample size of $T = 100$, while the linear boundary (2) is implemented.

structure in the backward CUSUM 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, the stacked backward CUSUM test is much more powerful than the forward CUSUM test, and its detector statistic is therefore well suited for real-time monitoring.

For the monitoring case, the critical values for the stacked backward CUSUM are shown in Table 4. For the forward CUSUM with the linear boundary (2), the simulated 5% critical values for $m = \infty$ are given by 0.957 for $k = 1$ and 1.044 for $k = 2$.

In order to evaluate the finite sample performances of the monitoring detectors, we consider models (4) and (5) for the time points $t = T + 1, \ldots, \lfloor mT \rfloor$. We simulate the series up to the fixed endpoints $m \in \{1.5, 2, 4, 10\}$, while the critical values for the case $m = \infty$ are implemented. Table 5 presents the size results. Note, that the tests are undersized by construction, as not all of the size is used up to the time point $\lfloor mT \rfloor$. For $k \geq 2$, we observe some size distortions for small sample sizes. The results in Table 6 show that the mean delay for the stacked backward CUSUM is much lower than that of the

Table 4: Asymptotic critical values for $\boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}}$ under the linear boundary

	$k=1$			$k=2$			$k=3$			$k=4$		
m	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
3	1.211	1.293	1.462	1.291	1.366	1.524	1.334	1.407	1.558	1.363	1.436	1.582
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
6	1.316	1.390	1.544	1.387	1.460	1.606	1.428	1.496	1.638	1.456	1.522	1.660
8	1.346	1.419	1.569	1.417	1.486	1.629	1.456	1.522	1.661	1.483	1.548	1.686
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	1.612	1.745	1.570	1.629	1.760
	$k=5$			$k=6$			$k=7$			$k=8$		
m	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
3	1.386	1.457	1.601	1.404	1.472	1.615	1.420	1.487	1.629	1.433	1.500	1.640
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
6	1.476	1.541	1.680	1.492	1.557	1.696	1.507	1.571	1.709	1.519	1.583	1.718
8	1.504	1.566	1.708	1.519	1.582	1.718	1.533	1.596	1.728	1.545	1.607	1.739
10	1.521	1.582	1.713	1.536	1.599	1.724	1.551	1.612	1.744	1.562	1.623	1.752
∞	1.588	1.650	1.777	1.604	1.661	1.788	1.617	1.673	1.799	1.630	1.683	1.812

Note: Critical values λ_α are reported for the linear boundary (2) from 10,000 Monte Carlo repetitions. The Gaussian processes in the limiting distributions are simulated on a grid of 10,000 equidistant points.

Table 5: Empirical sizes of the infinite horizon monitoring detectors

	$k=1$						$k=2$					
	$T=100$			$T=500$			$T=100$		$T=200$		$T=500$	
horizon	SBQ	Q	CSW	SBQ	Q	CSW	SBQ	Q	SBQ	Q	SBQ	Q
$m=1.5$	0.1	2.8	0.0	0.1	3.0	0.0	0.5	4.5	0.2	3.7	0.1	3.2
$m=2$	0.2	4.2	0.1	0.2	4.4	0.1	1.4	6.6	0.7	5.5	0.4	4.8
$m=4$	1.0	4.7	0.9	0.9	4.8	0.8	4.8	7.3	2.5	6.0	1.4	5.2
$m=6$	1.7	4.7	1.6	1.4	4.8	1.4	7.7	7.4	4.1	6.0	2.3	5.2
$m=8$	2.4	4.7	2.0	2.0	4.8	1.8	10.3	7.4	5.7	6.0	3.3	5.2
$m=10$	3.1	4.7	2.3	2.7	4.8	2.0	12.7	7.4	7.2	6.0	4.3	5.2

Note: Simulated rejection rates under H_0 are presented in percentage points. The linear boundary (2) is implemented, while critical values for $\alpha = 5\%$ and $m = \infty$ are considered. The values are obtained from 100,000 random draws of the models $y_t = \beta_1 + u_t$ and $y_t = \beta_1 + \beta_2 x_{t2} + u_t$ for $t = 1, \ldots, \lfloor mT \rfloor$, where x_{t2} and u_t are i.i.d. and standard normal. While SBQ and Q denote the tests $\boldsymbol{\mathcal{M}}^{mon}_{SBQ,\infty}$ and $\boldsymbol{\mathcal{M}}^{mon}_{Q,\infty}$, respectively, the univariate test by Chu et al. (1996) using the radical boundary (3) is denoted by CSW.

forward CUSUM and is almost constant across the breakpoint locations.

5 Conclusion

Two alternatives to the conventional CUSUM detectors by Brown et al. (1975) and Chu et al. (1996) have been proposed. It has been demonstrated that a detector that backwardly cumulates recursive residuals yields much higher power than when using forwardly cumulated recursive residuals when the break is located in the middle or at the end of the sample. Furthermore, the stacked triangular array of backwardly cumulated recursive residuals can be applied for monitoring and yields a much lower detection delay than that of the monitoring procedure by Chu et al. (1996). Due to the multivariate nature of the tests, we also have power against structural breaks that do not affect the unconditional mean of the dependent variable.

Table 6: Empirical mean detection delays of the monitoring detectors

	Model (4)			Model (5)	
	SBQ	Q	CSW	SBQ	Q
$r^* = 1.5$	41.4	39.4	53.6	62.2	50.4
$r^* = 2$	38.4	59.4	60.1	57.7	77.0
$r^* = 2.5$	36.9	79.2	65.8	54.6	103.4
$r^* = 3$	36.0	99.1	71.1	52.4	129.6
$r^* = 5$	34.5	178.0	89.4	48.1	233.6
$r^* = 10$	33.5	374.6	124.2	45.7	487.8

Note: The empirical mean detection delays are obtained from 100,000 Monte Carlo repetitions using size-adjusted critical values for a significance level of 5%, where models (4) and (5) are simulated for $t = 1, \ldots, \lfloor mT \rfloor$ with $T = 100$ and $m = 20$. While SBQ and Q correspond to the tests $\boldsymbol{\mathcal{M}}^{\text{mon}}_{SBQ,\infty}$ and $\boldsymbol{\mathcal{M}}^{\text{mon}}_{Q,\infty}$ with the linear boundary (2), the univariate test by Chu et al. (1996) with the radical boundary (3) is denoted by CSW.

Appendix: Proofs

We present some auxiliary lemmas that are required to prove Theorem 1.

Lemma 1. *Let $\{\boldsymbol{x}_t, u_t\}_{t\in\mathbb{N}}$ satisfy Assumption 1, and let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t u_t \Rightarrow \sigma \boldsymbol{C}^{-1/2} \mathbf{W}(r), \quad r \in [0, \infty),$$

where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions.

Proof. The result is shown in Phillips and Durlauf (1986) for the space $D([0,1])^k$, and its extension to the space $D([0,\infty))^k$ is discussed in Leisch et al. (2000). The Skorokhod metric for the product space $D([0,\infty))^k$ is defined in Billingsley (1999), p. 168 and p. 244. □

Lemma 2. *Let $\{\boldsymbol{x}_t, u_t\}_{t\in\mathbb{N}}$ satisfy Assumption 1, and let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$. Then, as $T \to \infty$,*

$$\sup_{r\in[0,\infty)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Big(\boldsymbol{x}_t u_t - \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{x}_j u_j \Big) \right\|_V = o_P(1),$$

where $\| \cdot \|_V$ denotes some vector norm on $\mathbb{R}^k$.

Proof. Let $f_t = \sqrt{1 + \boldsymbol{x}_t'(\sum_{i=1}^{t-1} \boldsymbol{x}_i \boldsymbol{x}_i')^{-1} \boldsymbol{x}_t}$, which yields $f_t w_t = 0$ for $t \leq k$, and

$$f_t w_t = y_t - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_{t-1} = u_t - \boldsymbol{x}_t' \Big(\sum_{j=1}^{t-1} \boldsymbol{x}_j \boldsymbol{x}_j' \Big)^{-1} \Big(\sum_{j=1}^{t-1} \boldsymbol{x}_j u_j \Big) = u_t - \boldsymbol{x}_t' \boldsymbol{C}_{t-1}^{-1} \Big(\frac{1}{t-1} \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j \Big),$$

for $t \geq k+1$. Consequently, $w_t = f_t^{-1} u_t - f_t^{-1} \boldsymbol{x}_t' \boldsymbol{C}_{t-1}^{-1}((t-1)^{-1} \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j)$. We decompose

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Big(\boldsymbol{x}_t u_t - \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{x}_j u_j \Big) = \boldsymbol{Z}_{1,T}(r) + \boldsymbol{Z}_{2,T}(r) + \boldsymbol{Z}_{3,T}(r),$$

where

$$\boldsymbol{Z}_{1,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \boldsymbol{x}_t u_t \Big(\frac{1 - f_t}{f_t} \Big),$$

$$\boldsymbol{Z}_{2,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \frac{1}{t} \boldsymbol{x}_t \boldsymbol{x}_t' \Big(\boldsymbol{C}^{-1} - \frac{t}{f_t(t-1)} \boldsymbol{C}_{t-1}^{-1} \Big) \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j,$$

$$\boldsymbol{Z}_{3,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \Big(\frac{1}{t} \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \boldsymbol{x}_t u_t \Big) - \frac{1}{\sqrt{T}} \sum_{t=1}^{k} \Big(\boldsymbol{x}_t u_t + \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{x}_j u_j \Big).$$

It remains to show that the norm of these random vectors converges to zero in probability. First, note that, from Assumption 1(a), it follows that $f_T \xrightarrow{p} 1$. Together with Lemma 1, we obtain

$$\sup_{r \in [0,\infty)} E[\|\boldsymbol{Z}_{1,T}(r)\|_V^2] = o(1), \qquad \sup_{r \in [0,\infty)} E[\|\boldsymbol{Z}_{3,T}(r)\|_V^2] = o(1),$$

which yields $\sup_{r \in [0,\infty)} \|\boldsymbol{Z}_{1,T}(r) + \boldsymbol{Z}_{3,T}(r)\|_V = o_P(1)$. The remaining term can be expressed as $\boldsymbol{Z}_{2,T}(r) = \int_0^r \boldsymbol{z}_{2,T}(s) \, \mathrm{d}s + o_P(1)$, where

$$\boldsymbol{z}_{2,T}(s) = \frac{1}{s} \boldsymbol{x}_{\lfloor sT \rfloor} \boldsymbol{x}_{\lfloor sT \rfloor}' \Big(\frac{1}{f_{\lfloor sT \rfloor}} \boldsymbol{C}_{\lfloor sT \rfloor}^{-1} - \boldsymbol{C}^{-1} \Big) \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor sT \rfloor} \boldsymbol{x}_j u_j$$

for $s \geq k/T$, and $\boldsymbol{z}_{2,T}(s) = \boldsymbol{0}$ for $s < k/T$. Assumption 1(a) and the continuous mapping theorem imply that $\sup_{s \in [0,\infty)} \|f_{\lfloor sT \rfloor}^{-1} \boldsymbol{C}_{\lfloor sT \rfloor}^{-1} - \boldsymbol{C}^{-1}\|_V = o_P(1)$. Lemma 1 yields $\sup_{s \in [0,\infty)} \|\boldsymbol{z}_{2,T}(s)\|_V = o_P(1)$, and, by the continuous mapping theorem, it follows that $\sup_{r \in [0,\infty)} \|\boldsymbol{Z}_{2,T}(r)\|_V = o_P(1)$. □

Lemma 3. *Let $\{\boldsymbol{x}_t, u_t\}_{t\in\mathbb{N}}$ satisfy Assumption 1. Then, as $T \to \infty$,*

$$\sup_{r\in[0,\infty)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{x}_j u_j - \int_0^r \frac{1}{s} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor sT \rfloor} \boldsymbol{x}_j u_j \right) \mathrm{d}s \right\|_V = o_P(1),$$

where $\|\cdot\|_V$ denotes some vector norm on $\mathbb{R}^k$.

Proof. Let $\boldsymbol{A}_t = \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1}$ and $\boldsymbol{b}_t = t^{-1} \sum_{j=1}^{t} \boldsymbol{x}_j u_j$. Note that for any $r \geq 1/T$, Abel's formula of summation by parts yields

$$\sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{A}_t \boldsymbol{b}_t = \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{A}_t \boldsymbol{b}_{\lfloor rT \rfloor} + \sum_{t=1}^{\lfloor rT \rfloor - 1} \sum_{j=1}^{t} \boldsymbol{A}_j (\boldsymbol{b}_t - \boldsymbol{b}_{t+1}) = \boldsymbol{Z}_{4,T}(r) + \boldsymbol{Z}_{5,T}(r) + \boldsymbol{Z}_{6,T}(r),$$

where we consider the decomposition given by

$$\begin{aligned}
\boldsymbol{Z}_{4,T}(r) &= \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{A}_t \boldsymbol{b}_{\lfloor rT \rfloor} - \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j u_j = (\boldsymbol{C}_{\lfloor rT \rfloor} \boldsymbol{C}^{-1} - \boldsymbol{I}) \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j u_j \\
\boldsymbol{Z}_{5,T}(r) &= \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j u_j - \sum_{t=1}^{\lfloor rT \rfloor - 1} \sum_{j=1}^{t} \boldsymbol{A}_j (\boldsymbol{b}_{t+1} - \tfrac{t}{t+1} \boldsymbol{b}_t) = \boldsymbol{x}_1 u_1 + \sum_{t=2}^{\lfloor rT \rfloor} \left(\boldsymbol{I}_k - \tfrac{t-1}{t} \boldsymbol{C}_{t-1} \boldsymbol{C}^{-1} \right) \boldsymbol{x}_t u_t \\
\boldsymbol{Z}_{6,T}(r) &= \sum_{t=1}^{\lfloor rT \rfloor - 1} \sum_{j=1}^{t} \boldsymbol{A}_j (\boldsymbol{b}_t - \tfrac{t}{t+1} \boldsymbol{b}_t) = \sum_{t=1}^{\lfloor rT \rfloor - 1} \tfrac{1}{t+1} \boldsymbol{C}_t \boldsymbol{C}^{-1} \sum_{j=1}^{t} \boldsymbol{x}_j u_j
\end{aligned}$$

Assumption 1(a) and Lemma 1 yield $\sup_{r\in[0,\infty)} \|\boldsymbol{Z}_{4,T}(r) + \boldsymbol{Z}_{5,T}(r)\|_V = o_P(\sqrt{T})$. Furthermore, Assumption 1(a) implies that

$$\sup_{r\in[0,\infty)} \left\| \frac{1}{\sqrt{T}} \boldsymbol{Z}_{6,T}(r) - \int_0^r \frac{1}{s} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor sT \rfloor} \boldsymbol{x}_j u_j \right) \mathrm{d}s \right\|_V = o_P(1),$$

and the assertion follows. □

Lemma 4. *Let $W(r)$ be a standard Brownian motion. Furthermore, for $r \geq 0$, let $F(W(r)) = W(r) - \int_0^r z^{-1} W(z) \,\mathrm{d}z$. Then $F(W(r)) \stackrel{\mathcal{D}}{=} W(r)$.*

Proof. Note that, by the Cauchy-Schwarz inequality and Jensen's inequality, we obtain $\int_0^r z^{-1} E[|W(z)|] \,\mathrm{d}z < \infty$ as well as $\int_0^r z^{-1} E[|W(r)W(z)|] \,\mathrm{d}z < \infty$, which justifies the application of Fubini's theorem in the subsequent steps. Since both $W(r)$ and $F(W(r))$ are Gaussian, it remains to show that their covariance functions coincide. First, note that

$E[F(W(r))] = E[W(r)] = 0$. Furthermore, let w.l.o.g. $r \leq s$. Then, the assertion follows from

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

□

Lemma 5. *Let $\{\boldsymbol{x}_t, u_t\}_{t\in\mathbb{N}}$ satisfy Assumption 1, and let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t \Rightarrow \sigma \boldsymbol{C}^{-1/2} \mathbf{W}(r), \quad r \in [0, \infty),$$

where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions.

Proof. Let $\mathbf{X}_T(r) = T^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j u_j$, and let $\mathbf{Y}_T(r) = T^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j w_j$. From Lemmas 2 and 3, it follows that $\sup_{r\in[0,\infty)} \|\mathbf{Y}_T(r) - F(\mathbf{X}_T(r))\|_V = o_P(1)$. Therefore, the Skorokhod metric of $\mathbf{X}_T(r)$ and $\mathbf{Y}_T(r)$ tends to zero in probability, and they thus have the same limiting probability measure. Lemma 1 and the continuous mapping theorem imply that $F(\mathbf{X}_T(r)) \Rightarrow F(\sigma \boldsymbol{C}^{-1/2}\mathbf{W}(r)) = \sigma \boldsymbol{C}^{-1/2} F(\mathbf{W}(r))$. Furthermore, from Lemma 4, it follows that $F(\mathbf{W}(r)) \overset{\mathcal{D}}{=} \mathbf{W}(r)$. Consequently, $\mathbf{Y}_T(r) \Rightarrow \sigma \boldsymbol{C}^{-1/2}\mathbf{W}(r)$. □

Lemma 6. *Let $\|\cdot\|_V$ be some vector norm on $\mathbb{R}^k$, and let $\|\cdot\|_M$ be the induced matrix norm. Let $\mathbf{h}$ be a $\mathbb{R}^k$-valued function of bounded variation, and let $\{\boldsymbol{A}_t\}_{t\in\mathbb{N}}$ be a sequence of random $(k \times k)$ matrices with $\sup_{r\in[0,\infty)} \|\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}(\boldsymbol{A}_t - \boldsymbol{A})\|_M = o_P(1)$. Then, as $T \to \infty$,*

$$\sup_{r\in[0,\infty)} \Big\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (\boldsymbol{A}_t - \boldsymbol{A}) \mathbf{h}(\tfrac{t}{T}) \Big\|_V = o_P(1).$$

Proof. By the application of Abel's formula of summation by parts it follows that

$$\sum_{t=1}^{\lfloor rT \rfloor} (\boldsymbol{A}_t - \boldsymbol{A}) \mathbf{h}(\tfrac{t}{T}) = \sum_{t=1}^{\lfloor rT \rfloor} (\boldsymbol{A}_t - \boldsymbol{A}) \mathbf{h}(\tfrac{\lfloor rT \rfloor}{T}) + \sum_{t=1}^{\lfloor rT \rfloor - 1} \sum_{j=1}^{t} (\boldsymbol{A}_j - \boldsymbol{A})(\mathbf{h}(\tfrac{t}{T}) - \mathbf{h}(\tfrac{t+1}{T})).$$

The fact that $\mathbf{h}(r)$ is of bounded variation yields

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

Consequently,

$$\sup_{r\in[0,\infty)} \Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t - \boldsymbol{A})\mathbf{h}(\tfrac{\lfloor rT\rfloor}{T})\Big\|_V \leq \sup_{r\in[0,\infty)} \Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t - \boldsymbol{A})\Big\|_M \Big\|\mathbf{h}(\tfrac{\lfloor rT\rfloor}{T})\Big\|_V = o_P(1)$$

and

$$\begin{aligned}
&\sup_{r\in[0,\infty)} \Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor -1}\sum_{j=1}^{t}(\boldsymbol{A}_j - \boldsymbol{A})(\mathbf{h}(\tfrac{t}{T}) - \mathbf{h}(\tfrac{t+1}{T}))\Big\|_V \\
&\leq \sup_{r\in[0,\infty)} \sum_{t=1}^{\lfloor rT\rfloor -1}\frac{t}{T}\Big\|\frac{1}{t}\sum_{j=1}^{t}(\boldsymbol{A}_j - \boldsymbol{A})\Big\|_M \Big\|\mathbf{h}(\tfrac{t}{T}) - \mathbf{h}(\tfrac{t+1}{T})\Big\|_V = o_P(1).
\end{aligned}$$

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

Proof of Theorem 1

Let $w_t^* = f_t^{-1}(y_t^* - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_{t-1}^*)$ be the recursive residuals from a regression without any structural break, where $f_t = (1 + (t-1)^{-1}\boldsymbol{x}_t'\boldsymbol{C}_{t-1}^{-1}\boldsymbol{x}_t)^{1/2}$,

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta}^0 + u_t, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{t-1}^* = \Big(\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\Big)^{-1}\Big(\sum_{j=1}^{t-1}\boldsymbol{x}_j y_j^*\Big)$$

Then, $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + u_t = y_t^* + T^{-1/2}\boldsymbol{x}_t'\mathbf{g}(t/T)$, and

$$\widehat{\boldsymbol{\beta}}_{t-1} = \widehat{\boldsymbol{\beta}}_{t-1}^* + \frac{1}{\sqrt{T}(t-1)}\boldsymbol{C}_{t-1}^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\mathbf{g}(j/T).$$

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

$$\begin{aligned}
\mathbf{S}_{1,T}(r) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{x}_t w_t^*, \quad \mathbf{S}_{2,T}(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor} f_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\mathbf{g}(\tfrac{t}{T}), \\
\mathbf{S}_{3,T}(r) &= -\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}\frac{1}{f_t(t-1)}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{C}_{t-1}^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\mathbf{g}(\tfrac{j}{T}).
\end{aligned}$$

Let $\|\cdot\|_V$ be some vector norm on $\mathbb{R}^k$, and let $\|\cdot\|_M$ be the induced matrix norm. Theorem 5 yields $\mathbf{S}_{1,T}(r) \Rightarrow \sigma \boldsymbol{C}^{1/2}\mathbf{W}(r)$. For the second term, note that, from Assumption 1(a), it follows that

$$\sup_{r\in[0,\infty)} \Big\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C}) \Big\|_M = o_P(1). \tag{6}$$

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

$$\sup_{r\in[0,\infty)} \Big\| \mathbf{S}_2(r) - \int_0^r \boldsymbol{C}\mathbf{g}(s)\,\mathrm{d}s \Big\|_V = \sup_{r\in[0,\infty)} \Big\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C})\mathbf{g}(\tfrac{t}{T}) \Big\|_V = o_P(1).$$

For the third term, let

$$\mathbf{p}_1(r) = \frac{1}{\lfloor rT \rfloor} \boldsymbol{C}_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j \boldsymbol{x}_j' \mathbf{g}(\tfrac{j}{T}), \quad \mathbf{p}_2(r) = \frac{1}{\lfloor rT \rfloor} \boldsymbol{C}_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{C}\mathbf{g}(\tfrac{j}{T}),$$

$$\mathbf{p}_3(r) = \frac{1}{\lfloor rT \rfloor} \sum_{j=1}^{\lfloor rT \rfloor} \mathbf{g}(\tfrac{j}{T}).$$

From Assumption 1(a), it follows that $\sup_{r\in[0,\infty)} \|\mathbf{p}_2(r) - \mathbf{p}_3(r)\|_M = o_P(1)$. Furthermore, from Lemma 6 and from the fact that $\sup_{r\in[0,\infty)} \| \frac{1}{\lfloor rT \rfloor} \sum_{t=1}^{\lfloor rT \rfloor} (\boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C}) \|_M = o_P(1)$, it follows that $\sup_{r\in[0,\infty)} \|\mathbf{p}_1(r) - \mathbf{p}_2(r)\|_V = o_P(1)$. Thus, $\sup_{r\in[0,\infty)} \|\mathbf{p}_1(r) - \mathbf{p}_3(r)\|_V = o_P(1)$. Consequently,

$$\begin{aligned} &\sup_{r\in[0,\infty)} \Big\| \mathbf{S}_{3,T}(r) + \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' \mathbf{h}_3(\tfrac{t-1}{T}) \Big\|_V \\ &\leq \sup_{r\in[0,\infty)} \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} \|f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t'\|_M \|\mathbf{p}_1(\tfrac{t-1}{T}) - \mathbf{p}_3(\tfrac{t-1}{T})\|_V, \end{aligned}$$

which is $o_P(1)$. Since $\mathbf{p}_3$ is a partial sum of a piecewise constant function, it is of bounded variation, and, together with (6), we can apply Lemma 6. Then,

$$\sup_{r\in[0,\infty)} \Big\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C}) \mathbf{p}_3(\tfrac{t-1}{T}) \Big\| = o_P(1),$$

which yields

$$\begin{aligned} &\sup_{r\in[0,\infty)} \Big\| \mathbf{S}_{3,T}(r) + \int_0^r \int_0^s \frac{1}{s} \boldsymbol{C}\mathbf{g}(v)\,\mathrm{d}v\,\mathrm{d}s \Big\|_V \\ &= \sup_{r\in[0,\infty)} \Big\| \mathbf{S}_{3,T}(r) + \frac{1}{T}\boldsymbol{C} \sum_{t=1}^{\lfloor rT \rfloor} \mathbf{p}_3(\tfrac{t-1}{T}) \Big\|_V + o_P(1) = o_P(1). \end{aligned}$$

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

$$\mathbf{Q}_T(r) = \widehat{\sigma}^{-1}\boldsymbol{C}_T^{-1/2}(\mathbf{S}_{1,T}(r) + \mathbf{S}_{2,T}(r) + \mathbf{S}_{3,T}(r)) \Rightarrow \mathbf{W}(r) + \boldsymbol{C}^{1/2}\mathbf{h}(r),$$

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

References

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

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

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414.

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

Bauer, P. and Hackl, P. (1978). The use of mosums for quality control. *Technometrics*, 20:431–436.

Billingsley, P. (1999). *Convergence of probability measures, 2nd Edition.* John Wiley & Sons.

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*, 37:149–192.

Chu, C.-S. J., Hornik, K., and Kaun, C.-M. (1995). Mosum tests for parameter constancy. *Biometrika*, 82:603–617.

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*, forthcoming.

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

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

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

Kuan, C.-M. and Hornik, K. (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*, 14:135–161.

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

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84:223–230.

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.

Ploberger, W. and Krämer, W. (1992). The cusum test with ols residuals. *Econometrica*, 60:271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40:307–318.

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.

Zeileis, A. (2004). Alternative boundaries for cusum tests. *Statistical Papers*, 45:123–131.

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