FACULTY WORKING PAPER 92-0164

Mosum Tests for Parameter Constancy

Chia-Shang J. Chu
Department of Economics
University of Southern California

Kurt Hornik
Institut für Statistik und
Wahrscheinlichkeitstheorie
Technische Universität Wien

Chung-Ming Kuan
Department of Economics
University of Illinois at Urbana-Champaign

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 92-0164

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 1992

Mosum Tests for Parameter Constancy

Chia-Shang J. Chu
Department of Economics

Kurt Hornik
Institut für Statistik und Wahrscheinlichkeitstheorie

Chung-Ming Kuan
Department of Economics

MOSUM TESTS FOR PARAMETER CONSTANCY

Chia-Shang J. Chu
Department of Economics
University of Southern California

Kurt Hornik
Institut für Statistik und Wahrscheinlichkeitstheorie
Technische Universität Wien

Chung-Ming Kuan
Department of Economics
University of Illinois at Urbana-Champaign

This Draft: Sept. 22, 1992

Abstract

In this paper, the MOSUM tests with recursive and OLS residuals are considered, and their asymptotic null distributions are characterized analytically. We show that the limiting processes of moving sums of recursive residuals is the increments of a standard Wiener process. For a particular choice of the size of moving sums, a formula representing the probability of this limiting process crossing constant boundaries is derived, from which correct asymptotic critical values are calculated and tabulated. We also show that the limiting process of moving sums of OLS residuals is the increments of a Brownian bridge whose boundary-crossing probability is known in literature. We then prove that the MOSUM tests are consistent and have non-trivial local power under a general class of alternatives. Our simulation further indicates that the propsoed MOSUM tests can complement other CUSUM-type of tests when there is a single structural change and have power advantage when there are certain double structural changes.

JEL Classification Number: 211

Keywords: CUSUM, MOSUM, Recursive Residuals, OLS Residuals, Boundary-Crossing Probability, Brownian Bridge, Wiener Process, Structural Change.

1 Introduction

For testing the null hypothesis that regression coefficients are constant over time, Brown, Durbin & Evans (1975) (hereafter BDE) introduced the CUSUM test based on cumulated sums of recursive residuals. Ploberger & Krämer (1992), motivated by the intuition that OLS residuals can better approximate the true disturbances under the null hypothesis, propose a CUSUM test based on cumulated sums of OLS residuals. (In what follows, the CUSUM tests with recursive and OLS residuals will be referred to as the BDE- and OLS-CUSUM tests, respectively.) While the BDE-CUSUM test is one of the most commonly used tests for parameter constancy in applications, Bauer & Hackl (1978) find that cumulated sums (of recursive residuals) may not be very sensitive to certain types of parameter changes.

Another shortcoming of the CUSUM test seems to have been overlooked. Recall that the limiting distribution of the BDE-CUSUM test is determined by the probability of a standard Wiener process crossing a pair of linear boundaries. Intuitively, root-t boundaries should be used to account for the growing variance of the limiting Wiener process, but for convenience BDE use a pair of straight lines tangent to the desired curved boundaries as an approximation. This results in a loss of test power asymptotically because the diameter of the acceptance region is larger than it should be. Also, the asymptotic null distribution of the OLS-CUSUM test is determined by the probability of a Brownian bridge crossing a pair of constant boundaries. This also results in biased power performance because the variance of a Brownian bridge is not a constant.

In this paper we consider the MOSUM test based on *moving* sums of residuals. That is, we fix the number of residuals in each moving sum and let these sums move accross the whole sample. This is intuitively appealing because moving sums retain only recent information by gradually discarding the residual in the distant past, hence may be more sensitive to parameter variation. In particular, we study the MOSUM tests with recursive and OLS residuals. The former is introduced in Bauer & Hackl (1978), but its (conservative) critical values are obtained by incorrectly ignoring correlations of moving sums; the latter is new. (In what follows, these two MOSUM tests will be referred to as the BH- and OLS-MOSUM tests, respectively.) We first show that the limiting processes of moving sums of recursive and OLS residuals are, respectively, the increments of a standard

Wiener process and the increments of a Brownian bridge. As both limiting processes have constant variance, it is legitimate to consider the probability of these limiting processes crossing a pair of constant boundaries. The MOSUM tests thus avoid the aforementioned drawback of the CUSUM test.

We also characterize the asymptotic null distributions of the proposed MOSUM tests for a particular choice of the size of moving sums. We derive a formula representing the boundary-crossing probability of the limiting BH-MOSUM process, which, to the best of our knowledge, is a *new* result and yields correct asymptotic critical values. The boundary-crossing probability of the limiting OLS-MSOUM process is obtained from a result in Chu, Hornik, & Kuan (1992). Concerning the power performance, we show that the MOSUM tests are consistent and have non-trivial local power against a wide class of alternatives. Our simulation indicates that the BH-MOSUM test is comparable with the BDE- and OLS-CUSUM tests when there is a single structural change, although none of them uniformly dominates the other. Under the alternative of double structural changes, we also find that the proposed MOSUM tests outperform both CUSUM tests quite significantly.

This paper proceeds as follows. We introduce the MOSUM tests in section 2 and derive their asymptotic null distribution in section 3. We analyze the power performance of the MOSUM tests and report simulation results in section 4. Section 5 concludes the paper. All proofs are deferred to the Appendix.

2 The MOSUM Tests

Given the regression model

$$y_i = x_i'\beta_i + \epsilon_i, \qquad i = 1, 2, \ldots, T,$$

where β_i is a $k \times 1$ vector, we are interested in the null hypothesis that $\beta_i = \beta_0$ for all i. Following Krämer, Ploberger, & Alt (1988) (hereafter KPA) and Ploberger & Krämer (1992) (hereafter PK), we assume that, in addition to other technical conditions, $\{\epsilon_i\}$ is a martingale difference sequence with respect to some sequence of σ-algebras $\{\mathcal{F}_i\}$ with $E(\epsilon_i^2|\mathcal{F}_{i-1}) = \sigma^2$ and that x_i is $\mathcal{F}_{i-1}$-measurable and obeys the weak laws of large numbers:

$$\frac{1}{T}\sum_{i=1}^{T} x_i \quad \to^p \quad \mu, \tag{1}$$

$$\frac{1}{T}\sum_{i=1}^{T} x_i x_i' \quad \rightarrow^p \quad R, \tag{2}$$

where $\rightarrow^p$ stands for convergence in probability, μ and R are $k \times 1$ and $k \times k$ non-stochastic matrices, respectively. These conditions are not the weakest possible but are general enough to cover many interesting cases, including dynamic models; see KPA and PK for more details. In what follows these conditions are always assumed but will not be mentioned explicitly.

Let $X_n = [x_1\, x_2 \cdots x_n]'$ and $Y_n = [y_1\, y_2 \cdots y_n]'$ be the data matrices of dimensions $n \times k$ and $n \times 1$, respectively. Then $\hat{\beta}_n = (X_n' X_n)^{-1} X_n' Y_n$ is the OLS estimator at time n, and $\hat{\epsilon}_n = y_n - x_n' \hat{\beta}_{n-1}$ is the prediction error. Recursive residuals are defined as

$$e_n = \frac{\hat{\epsilon}_n}{\sqrt{1 + x_n'(X_{n-1}' X_{n-1})^{-1} x_n}}, \qquad n = k+1, \cdots, T.$$

For $s = k+1, \ldots, T$, the s-th *cumulated sum* of recursive residuals is

$$\frac{1}{\hat{\sigma}\sqrt{T-k}} \sum_{n=k+1}^{s} e_n,$$

where $\hat{\sigma}$ is a consistent estimator of σ. Let

$$Q_T^r(t) = \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} e_n = \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{y_n - x_n' \hat{\beta}_{n-1}}{v_n} \tag{3}$$

be the corresponding empirical BDE-CUSUM process in $D([0,1])$, the space of functions that are right continuous with left-hand limits on $[0,1]$, where for notational convenience we write $\tau = T - k$ and $v_n = (1 + x_n'(X_{n-1}' X_{n-1})^{-1} x_n)^{1/2}$, and $[\tau t]$ is the integer part of τt. Throughout this paper, we shall use the superscript "r" to signify the processes and statistics that are based on recursive residuals. KPA show that under the null hypothesis, $Q_T^r \Rightarrow W$, where $\Rightarrow$ stands for weak convergence (of the associated probability measures) and W is a standard Wiener process, cf. Sen (1982).

We first consider moving sums of recursive residuals, i.e., the sums of $[\tau h]$ recursive residuals moving across the whole sample, where h $(0 < h < 1)$ denotes the proportion of the residuals used to construct each moving sum. For $j = 0, \cdots, \tau - [\tau h]$, the j-th *moving sum* is

$$\frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+j+1}^{k+j+[\tau h]} e_n$$

$$= \frac{1}{\hat{\sigma}\sqrt{\tau}} \left(\sum_{n=k+1}^{k+j+[\tau h]} e_n - \sum_{n=k+1}^{k+j} e_n \right)$$
$$= Q_T^\tau\left(\frac{j+[\tau h]}{\tau}\right) - Q_T^\tau\left(\frac{j}{\tau}\right).$$

Let

$$S_{T,h}^\tau(t) = \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+[N_\tau t]+1}^{k+[N_\tau t]+[\tau h]} e_n = Q_T^\tau\left(\frac{[N_\tau t]+[\tau h]}{\tau}\right) - Q_T^\tau\left(\frac{[N_\tau t]}{\tau}\right) \tag{4}$$

be the corresponding empirical MOSUM process on $[0, 1-h]$, where $N_\tau = (\tau-[\tau h])/(1-h)$. The BH-MOSUM test statistic introduced in Bauer & Hackl (1978) is

$$MS_{T,h}^\tau = \max_{0 \leq j \leq \tau-[\tau h]} \frac{1}{\hat{\sigma}\sqrt{\tau}} \left| \sum_{n=k+j+1}^{k+j+[\tau h]} e_n \right| = \max_{0\leq t\leq 1-h} |S_{T,h}^\tau(t)|. \tag{5}$$

We then have

Theorem 2.1 *Under the null hypothesis,*

$$S_{T,h}^\tau \Rightarrow S_h^\tau,$$

where for $0 \leq t \leq 1-h$, $S_h^\tau(t) = W(t+h) - W(t)$. *In particular,*

$$MS_{T,h}^\tau \Rightarrow \max_{0\leq t\leq 1-h} |S_h^\tau(t)|.$$

This result says that the empirical BH-MOSUM process converges in distribution to the increments of a Wiener process. We observe that for $0 \leq s \leq t$,

$$\begin{aligned} &\text{cov}(S_h^\tau(t), S_h^\tau(s)) \\ &= \text{cov}(W(t+h), W(s+h)) - \text{cov}(W(t+h), W(s)) \\ &\qquad - \text{cov}(W(t), W(s+h)) + \text{cov}(W(t), W(s)) \\ &= s+h-s-\min(t, s+h)+s \\ &= \max(h+(s-t), 0). \end{aligned}$$

The covariance function of S_h^τ is thus $\max(h-|t-s|, 0)$ so that the variance is h. The asymptotic critical values of the MOSUM test are then determined by the two-sided boundary-crossing probabilities of S_h^τ:

$$\lim_{T\to\infty} \mathbb{P}\{MS_{T,h}^\tau > b^*\} = \mathbb{P}\{|S_h^\tau(t)| > b^* \text{ for some } 0 \leq t \leq 1-h\}.$$

Let $\hat{\epsilon}_{T,n} = y_n - x_n'\hat{\beta}_T$ be the OLS residual. Note that there are T OLS residuals, but there are only $T-k$ recursive residuals. For $s = 1,\ldots,T$, the s-th cumulated sum of OLS residuals is

$$\frac{1}{\hat{\sigma}\sqrt{T}}\sum_{n=1}^{s}\hat{\epsilon}_{T,n}.$$

and the corresponding empirical OLS-CUSUM process in $D([0,1])$ is

$$Q_T^o(t) = \frac{1}{\hat{\sigma}\sqrt{T}}\sum_{n=1}^{[Tt]}\hat{\epsilon}_{T,n}. \tag{6}$$

We also use the superscript "o" to signify the processes and statistics that are based on OLS residuals. PK show that under the null hypothesis, $Q_T^o \Rightarrow W^0$, where W^0 is a Brownian bridge, provided that x contains a constant term. For $j = 0,\cdots,T-[Th]$, the j-th moving sum of OLS residuals is

$$\begin{aligned}
&\frac{1}{\hat{\sigma}\sqrt{T}}\sum_{n=j+1}^{j+[Th]}\hat{\epsilon}_{T,n} \\
&= \frac{1}{\hat{\sigma}\sqrt{T}}\left(\sum_{n=1}^{j+[Th]}\hat{\epsilon}_{T,n} - \sum_{n=1}^{j}\hat{\epsilon}_{T,n}\right) \\
&= Q_T^o\left(\frac{j+[Th]}{T}\right) - Q_T^o\left(\frac{j}{T}\right).
\end{aligned}$$

Similar to (4), let

$$S_{T,h}^o(t) = \frac{1}{\hat{\sigma}\sqrt{T}}\sum_{n=[N_Tt]+1}^{[N_Tt]+[Th]}\hat{\epsilon}_{T,h} = Q_T^o\left(\frac{[N_Tt]+[Th]}{T}\right) - Q_T^o\left(\frac{[N_Tt]}{T}\right) \tag{7}$$

be the empirical OLS-MOSUM process on $[0,1-h]$, where $N_T = (T-[Th])/(1-h)$, and let the OLS-MOSUM statistic be defined as

$$MS_{T,h}^o = \max_{0\le j\le T-[Th]}\frac{1}{\hat{\sigma}\sqrt{T}}\left|\sum_{n=j+1}^{j+[Th]}\hat{\epsilon}_{T,n}\right| = \max_{0\le t\le 1-h}|S_{T,h}^o(t)|. \tag{8}$$

The result below is analogous to Theorem 2.1.

Theorem 2.2 *Under the null hypothesis,*

$$S_{T,h}^o \Rightarrow S_h^o,$$

where for $0 \le t \le 1-h$, $S_h^o(t) = W^0(t+h) - W^0(t)$. *In particular,*

$$MS_{T,h}^o \Rightarrow \max_{0 \le t \le 1-h} |S_h^o(t)|.$$

We note that the limiting process S_h^o is the same as that of the empirical moving-estimates process studied in Chu, Hornik, & Kuan (1992); in particular, its covariance function is $h(1-h) - \min(h, |t-s|)$, and the variance is $h(1-h)$. Again, the asymptotic critical values of the OLS-MOSUM test are determined by the two-sided boundary-crossing probability of S_h^o:

$$\lim_{T \to \infty} \mathbb{P}\{MS_{T,h}^o > b^*\} = \mathbb{P}\{|S_h^o(t)| > b^* \text{ for some } 0 \le t \le 1-h\}.$$

The MOSUM test differs from the CUSUM test in the following respects. First, each moving sum uses the most recent $[\tau h]$ ($[Th]$) residuals, whereas cumulated sums incorporate more and more residuals. Intuitively, moving sums should be more sensitive to parameter changes because they contain only recent information. A similar intuition has been confirmed in Chu, Hornik, & Kuan (1992), where a moving-estimates test is shown to be more sensitive to double structural changes than the recursive-estimates-based fluctuation test of Sen (1980) and Ploberger, Krämer, & Kontrus (1989). Second, as the variances of the limiting BH-MOSUM and OLS-MOSUM processes are h and $h(1-h)$, respectively, it suffices to consider the probability of S_h^r (S_h^o) crossing a pair of constant boundaries. On the other hand, the desired boundaries for the BDE-CUSUM test should be of the form $\pm\lambda\sqrt{t}$, where λ is a constant depending on the size of the test, to account for the growing variance of the Wiener process (BDE, p. 153). However, BDE use linear boundaries tangent to the desired curved boundaries as an approximation to derive asymptotic critical values. This approximation causes loss of test power when the change point of parameter is away from the center of sample. Also, the limiting OLS-CUSUM process is a Brownian bridge whose variance is $t(1-t)$. Thus, the OLS-CUSUM test with critical values derived from the probability of a Brownian bridge crossing constant boundaries will also result in biased power performance. The MOSUM tests discussed above obviously do not suffer this drawback.

An interesting and important issue arising in the MOSUM test is of course the choice of h, the size of moving sums. If h is large, each moving sum includes "too many" residuals, and only a few moving sums are available to detect possible parameter changes. Hence, moving sums with large h are not very sensitive to parameter variation. If h is small, the limiting process S_h^τ (S_h^o) is not a good approximation of moving sums in finite samples, and sample variation of each moving sum is likely to be large. Therefore, the leading choice is $h = 1/2$ so that the number of residuals in each moving sum equals $[\tau/2]$ ($[T/2]$). We note that the MOSUM tests with $h < 1/2$ and h not "too" small are also of interest; their asymptotic distributions appear to be difficult to calculate, though.

3 Asymptotic Null Distributions

We first derive the asymptotic null distributions of the BH-MOSUM test.

Lemma 3.1 *Let $\Delta W(t) = W(t+1) - W(t)$. Then*

$$\max_{0 \le t \le 1-h} \frac{1}{\sqrt{h}} |S_h^\tau(t)| =^d \max_{0 \le t \le 1/h - 1} |\Delta W(t)|,$$

where $=^d$ stands for equality in distribution.

Let $b = b^*/\sqrt{h}$, then by Theorem 2.1 and Lemma 3.1,

$$\begin{aligned} \lim_{T\to\infty} \mathbb{P}\{MS_{T,h}^\tau > b^*\} &= \lim_{T\to\infty} \mathbb{P}\left\{\frac{1}{\sqrt{h}} MS_{T,h}^\tau > b\right\} \\ &= \mathbb{P}\left\{\max_{0 \le t \le 1/h-1} |\Delta W(t)| > b\right\}. \end{aligned} \tag{9}$$

In what follows, we compute this probability for $h = 1/2$ such that $1/h - 1 = 1$.

For $f \in C[0,1]$, the space of all continuous functions on $[0,1]$, define

$$m(f) = \min\nolimits_{0 \le t \le 1} f(t), \qquad M(f) = \max\nolimits_{0 \le t \le 1} f(t).$$

We also let Φ and ϕ denote the distribution and density functions of the standard normal random variable, respectively. We first consider the probability that ΔW stays within two constant boundaries a and b on $[0,1]$ conditional on $\Delta W(0) = x$:

$$\begin{aligned} &\mathbb{P}\{a \le \Delta W(t) \le b \text{ for all } 0 \le t \le 1 | \Delta W(0) = x\} \\ &= \mathbb{P}\{a \le m(\Delta W) \le M(\Delta W) \le b | \Delta W(0) = x\}. \end{aligned}$$

Let $\mu_{\Delta W|x}$, μ_Z, and μ_W be the measures on $C[0,1]$ induced by ΔW conditional on $\Delta W(0) = x$, $Z = x + \sqrt{2}W$, and the Wiener process, respectively.

Lemma 3.2 (Shepp (1966)) *The measures $\mu_{\Delta W|x}$ and μ_Z are equivalent, and the Radon-Nikodym derivative of $\mu_{\Delta W|x}$ with respect to μ_Z is*

$$\frac{\mu_{\Delta W|x}}{d\mu_Z}(f) = \sqrt{2}e^{x^2/2}e^{-(x+f(1))^2/4}.$$

Observe that under μ_Z, the functions $g(t) = (f(t) - x)/\sqrt{2}$ are distributed according to μ_W. Then by Lemma 3.2 and the Radon-Nikodym theorem,

$$\begin{aligned}
&\mathbb{P}\{a \le m(\Delta W) \le M(\Delta W) \le b|\Delta W(0) = x\} \\
&= \int_{a\le m(f)\le M(f)\le b} d\mu_{\Delta W|x}(f) \\
&= \int_{a\le m(f)\le M(f)\le b} \frac{d\mu_{\Delta W|x}}{d\mu_Z}(f)\, d\mu_Z(f) \\
&= \int_{(a-x)/\sqrt{2}\le m(g)\le M(g)\le (b-x)/\sqrt{2}} \sqrt{2}e^{x^2/2}e^{-(2x+\sqrt{2}g(1))^2/4}\, d\mu_W(g) \\
&= \int_{(a-x)/\sqrt{2}}^{(b-x)/\sqrt{2}} \frac{1}{\phi(x)\sqrt{\pi}} e^{-(2x+\sqrt{2}u)^2/4} \\
&\qquad d\mathbb{P}\left\{\frac{a-x}{\sqrt{2}} \le m(W) \le M(W) \le \frac{b-x}{\sqrt{2}},\ W(1) \le u\right\}.
\end{aligned} \tag{10}$$

By (11.10) of Billingsley (1968),

$$\begin{aligned}
&d\mathbb{P}\left\{\alpha \le m(W) \le M(W) \le \beta,\ W(1) \le u\right\}/du \\
&= \sum_{k=-\infty}^{\infty} \phi(u + 2k(\beta-\alpha)) - \sum_{k=-\infty}^{\infty} \phi(u - 2\beta - 2k(\beta-\alpha)).
\end{aligned} \tag{11}$$

From (10) and (11), routine calculation shows that

Lemma 3.3 *For $a \le b$,*

$$\begin{aligned}
&\mathbb{P}\{a \le \Delta W(t) \le b \text{ for all } 0 \le t \le 1|\Delta W(0) = x\} \\
&= \frac{1}{\phi(x)} \sum_{k=-\infty}^{\infty} \Big[\phi(x - k(b-a))(\Phi(b + k(b-a)) - \Phi(a + k(b-a))) \\
&\qquad - \Phi(x - k(b-a))(\phi(b + k(b-a)) - \phi(a + k(b-a)))\Big];
\end{aligned} \tag{12}$$

in particular, when $a = -b$,

$$\begin{aligned}
&\mathbb{P}\{|\Delta W(t)| \le b \text{ for all } 0 \le t \le 1|\Delta W(0) = x\} \\
&= \frac{1}{\phi(x)} \sum_{k=-\infty}^{\infty} \Big[\phi(x - 2kb)(\Phi((2k+1)b) - \Phi((2k-1)b)) \\
&\qquad - \Phi(x - 2kb)(\phi((2k+1)b) - \phi((2k-1)b))\Big].
\end{aligned} \tag{13}$$

Remark: For $a \to -\infty$, all terms $\phi(y \pm k(b-a))$ tend to zero unless $k = 0$. It follows from Lemma 3.3 that

$$\begin{aligned} &\mathbb{P}\{\Delta W(t) \le b \text{ for all } 0 \le t \le 1 | \Delta W(0) = x\} \\ &= \phi(x)^{-1}\Big(\phi(x)\Phi(b) - \phi(b)\Phi(x)\Big) \\ &= \Phi(b) - \frac{\phi(x)}{\phi(b)}\,\Phi(x). \end{aligned} \tag{14}$$

This is the same as the result calculated from (17.8) of Shepp (1966).

Because $\Delta W(0) = W(1) - W(0)$ is a standard normal random variable, the unconditional probability that ΔW stays within the constant boundaries $-b$ and b on $[0,1]$ is thus

$$\begin{aligned} &\mathbb{P}\{|\Delta W(t)| \le b \text{ for all } 0 \le t \le 1\} \\ &= \int_{-b}^{b} \mathbb{P}\{|\Delta W(t)| \le b \text{ for all } 0 \le t \le 1 | \Delta W(0) = x\}\phi(x)\,dx \\ &= \sum_{k=-\infty}^{\infty} \Bigg((\Phi((2k+1)b) - \Phi((2k-1)b)) \int_{-b}^{b} \phi(x - 2kb)\,dx \\ &\qquad + (\phi((2k+1)b) - \phi((2k-1)b)) \int_{-b}^{b} \Phi(x + 2kb)\,dx \Bigg). \end{aligned}$$

Note that we have changed the sign of the second term in (13) by replacing k by $-k$. As clearly

$$\int_{-b}^{b} \phi(x - 2kb)\,dx = \int_{(2k-1)b}^{(2k+1)b} \phi(u)\,du = \Phi((2k+1)b) - \Phi((2k-1)b)$$

and

$$\begin{aligned} &\int_{-b}^{b} \Phi(x + 2kb)\,dx \\ &= \int_{(2k-1)b}^{(2k+1)b} \Phi(u)\,du \\ &= (u\Phi(u) + \phi(u))\Big|_{(2k-1)b}^{(2k+1)b} \\ &= (2k+1)b\Phi((2k+1)b) - (2k-1)b\Phi((2k-1)b) + \phi((2k+1)b) - \phi((2k-1)b), \end{aligned}$$

we thus obtain

Theorem 3.4 *For $b > 0$,*

$$\mathbb{P}\{|\Delta W(t)| \leq b \text{ for all } 0 \leq t \leq 1\} = \sum_{k=-\infty}^{\infty} (A_k + B_k + C_k), \tag{15}$$

where

$$\begin{aligned} A_k &= (\Phi((2k+1)b) - \Phi((2k-1)b))^2, \\ B_k &= (\phi((2k+1)b) - \phi((2k-1)b)) \\ &\qquad \times ((2k+1)b\Phi((2k+1)b) - (2k-1)b\Phi((2k-1)b)), \\ C_k &= (\phi((2k+1)b) - \phi((2k-1)b))^2. \end{aligned}$$

Concerrning the boundary-crossing probability of the limiting OLS-MOSUM process, Chu, Hornik, & Kuan (1992) show that

$$\mathbb{P}\{|W^0(t+\tfrac{1}{2}) - W^0(t)| \leq b^* \text{ for all } 0 \leq t \leq \tfrac{1}{2}\} = 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-k^2\pi^2/8(b^*)^2}. \tag{16}$$

Combining (9), (15), and (16) we immediately get

Corollary 3.5 *Under the null hypothesis, the BH-MOSUM test has the limiting distribution*

$$\lim_{T\to\infty} \mathbb{P}\{MS^r_{T,1/2} \leq b^*\} = \sum_{k=-\infty}^{\infty} (A_k + B_k + C_k),$$

where A_k, B_k, and C_k are defined in Theorem 3.4; the OLS-MOSUM test has the limiting distribution

$$\lim_{T\to\infty} \mathbb{P}\{MS^o_{T,1/2} \leq b^*\} = 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-k^2\pi^2/8(b^*)^2}.$$

Given different probabilities, the asymptotic critical values of the BH-MOSUM test are $b^* = b/\sqrt{2}$, where b's are computed from (15). The asymptotic critical values of the OLS-MOSUM test are calculated directly from (16). Some of these values are summarized in Table 1.

[Table 1 About Here]

Remarks:

(1) Qualls & Watanabe (1972) study the asymptotic distribution of $\sup_{0\leq t\leq T} X(t)$, where $X(t)$ is a Gaussian process with covariance function

$$1 - |t-s|^\alpha H(|t-s|) + o(|t-s|^\alpha H(|t-s|)),$$

$0 < \alpha \leq 2$, and H is a function slowly varying at zero. Their result, however, does not give a readily usable representation of the distribution.

(2) In the Appendix we show that

$$\mathbb{P}\{\Delta W(t) \leq b \text{ for all } 0 \leq t \leq 1\} = \Phi(b)^2 - \phi(b)(b\Phi(b) + \phi(b)), \tag{17}$$

which is the asymptotic distribution of the one-sided BH-MOSUM test. The one-sided MOSUM test becomes relevant when one is concerned only whether the parameter has changed in a particular direction. For probabilities 0.90, 0.95, 0.975, and 0.99, (17) yields one-sided critical values ($b/\sqrt{2}$) 1.57596, 1.80413, 2.00372, and 2.23765, respectively. Some of the one-sided critical values of the OLS-MOSUM test can be found in Chu, Hornik, & Kuan (1992).

4 Test Performance and Simulation

Consider a sequence of alternatives

$$\beta_i^T = \beta_0 + T^{-c} g(i/T), \tag{18}$$

where $g : [0,1] \to \mathbb{R}^k$ is a vector-valued, non-constant function of bounded variation on $[0,1]$. Note that (18) represents a general class of alternatives including a global alternative ($c = 0$), a sequence of local alternatives ($c = 1/2$), and a sequence of non-local alternatives ($0 < c < 1/2$). Here, g may e.g. be a step or continuous function to characterize various types of parameter changes.

Under the maintained assumptions, it can be shown that if

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T}(y_i - x_i'\hat{\beta}_T)^2 \tag{19}$$

is the standard OLS estimator of σ^2, then $\hat{\sigma}^2 \longrightarrow^p \sigma_c^2$, where

$$\sigma_c^2 = \begin{cases} \sigma^2, & 0 < c \leq 1/2, \\ \sigma^2 + \int_0^1 \Big(g(u) - \int_0^1 g(v)\,dv\Big) R \Big(g(u) - \int_0^1 g(v)\,dv\Big)'\,du, & c = 0, \end{cases} \tag{20}$$

and R is defined in (2). It is evident from (20) that $\hat{\sigma}^2$ is not consistent under the global alternative but is consistent otherwise. We obtain the following result for the BH-MOSUM test.

Theorem 4.1 *Under the alternative* (18), *if* $0 \leq c < 1/2$,

$$T^{c-1/2} MS^{\tau}_{T,h} \longrightarrow^p \sigma_c^{-1} \max_{0 \leq t \leq 1-h} |\mu' J^{\tau}_h g(t)|,$$

where μ *is defined in* (1) *and*

$$J^{\tau}_h g(t) = \int_t^{t+h} \left(g(u) - \frac{1}{u} \int_0^u g(v)\,dv \right) du;$$

if $c = 1/2$,

$$MS^{\tau}_{T,h} \Rightarrow \max_{0 \leq t \leq 1-h} \left| S^{\tau}_h(t) + \sigma^{-1} \mu' J^{\tau}_h g(t) \right|.$$

Similarly, the following result holds for the OLS-MOSUM test.

Theorem 4.2 *Under the alternative* (18), *if* $0 \leq c < 1/2$,

$$T^{c-1/2} MS^{o}_{T,h} \longrightarrow^p \sigma_c^{-1} \max_{0 \leq t \leq 1-h} |\mu' J^{o}_h g(t)|,$$

where

$$J^{o}_h g(t) = \int_t^{t+h} g(u)\,du - h \int_0^1 g(u)\,du;$$

if $c = 1/2$,

$$MS^{o}_{T,h} \Rightarrow \max_{0 \leq t \leq 1-h} |S^{o}_h(t) + \sigma^{-1} \mu' J^{o}_h g(t)|.$$

Remarks:

(1) Under the alternative (18) with $0 \leq c < 1/2$, if $\mu' J^{\tau}_h g(t) \neq 0$ and $\mu' J^{o}_h g(t) \neq 0$ for some t, then $MS^{\tau}_{T,h}$ and $MS^{o}_{T,h}$ must grow at the rate $T^{1/2-c}$ so that the two MOSUM tests are consistent for such sequences of alternatives. In particular, as noted in Chu, Hornik, & Kuan (1992), if $\mu' g$ is not periodic with period h, then $\mu' J^{o}_h g(t)$ does not vanish identically. On the other hand,

- if $\mu' g = 0$, i.e., μ is orthogonal to g, then $\mu' J_h^r g(t) = 0$ for all t;
- if $\mu' g = 0$ or if $\mu' g$ is periodic with period h and $1/h$ is an integer, then $\mu' J_h^o g(t) = 0$ for all t.

Under these situations, consistency of the MOSUM tests is not ensured.

(2) Under local alternatives (i.e., $c = 1/2$), if μ is orthogonal to g, then the weak limits of the BH- and OLS-MOSUM tests reduce to their corresponding weak limits under the null hypothesis. Thus, the MOSUM tests have only trivial local power when the mean of regressors is orthogonal to parameter changes; in particular, they should *not* be applied when the regressors have mean zero. Note that the BDE- and OLS-CUSUM tests suffer the same local problem, as shown in KPA and PK. Similarly, if g is periodic with period h and $1/h$ is an integer, the OLS-MOSUM test again has trivial local power. This local power deficiency is the same as that of the moving-estimates test (Chu, Hornik, & Kuan (1992)).

(3) If in a model the mean of regressors is a zero vector, one should subtract a nonzero constant from the regressors, which helps to circumvent the local power problem and affects only the estimates of the constant term in the model.

Our simulation uses the location model

$$y_i = \beta_i + \epsilon_i, \tag{21}$$

where ϵ_i generated from i.i.d. $N(0,1)$. The results of size simulation are summarized in Table 2. It can be seen that all the tests are conservative, in the sense that the type I errors are less than nominal sizes. We also observe that the BDE-CUSUM test has the least size distortion and that the OLS-MOSUM test has the greatest size distortion in all cases considered.

[Table 2 About Here]

Consider the alternative of a single structural change:

$$\beta_i = \begin{cases} \beta_0, & i = 1, \cdots, [T\lambda], \\ \beta_0 + \Delta, & i = [T\lambda] + 1, \cdots, T, \end{cases} \tag{22}$$

where $\Delta \neq 0$ represents the jump in the parameter and λ characterizes the change point. The result below follows from Theorems 4.1 and 4.2.

Corollary 4.3 *Under the alternative* (22), *we have*

$$\frac{1}{\sqrt{T}} MS^{r}_{T,h} \quad \to^{p} \quad \lambda(\log(\min(\lambda+h,1)) - \log\lambda)\frac{|\Delta|}{\sigma_0}, \tag{23}$$

$$\frac{1}{\sqrt{T}} MS^{o}_{T,h} \quad \to^{p} \quad \max\Big(\min(\lambda(1-h), h(1-\lambda)), \min((1-h)(1-\lambda), h\lambda)\Big)\frac{|\Delta|}{\sigma_0}, \tag{24}$$

where σ_0 *is defined in* (20).

Without loss of generality, we set $\Delta = \sigma_0 = 1$. If $h = 1/2$, then

$$\frac{1}{\sqrt{T}} MS^{r}_{T,1/2} \to^{p} \begin{cases} \lambda(\log(\lambda+1/2) - \log\lambda), & \lambda \leq 1/2, \\ \lambda|\log\lambda|, & \lambda \geq 1/2. \end{cases} \tag{25}$$

This limit, as a function of λ, reaches its maximum at $\lambda = 1/2$. Note, however, that the power performance of the BH-MOSUM test is *not* symmetric about 1/2. Also, when $h = 1/2$, we have

$$\frac{1}{\sqrt{T}} MS^{o}_{T,1/2} \to^{p} \frac{1}{2}\min(\lambda, 1-\lambda). \tag{26}$$

The above limit is just a linear interpolation of the points (0,0), (1/2,1/4), and (1,0), which is the same as that of the moving-estimates test. It is clear that the power performance of the OLS-MOSUM test with $h = 1/2$ is symmetric about $\lambda = 1/2$.

For the BDE-CUSUM statistic, it can be shown that

$$\frac{1}{\sqrt{T}} \max_{0\leq t\leq 1} |Q^{r}_{T}(t)| \to^{p} \max_{0\leq t\leq 1} \max(\lambda\log(t/\lambda), 0) = \lambda|\log\lambda|, \tag{27}$$

cf. equation (33) of PK. This limit reaches its maximum at $\lambda = 1/e \approx 0.368$. It is also readily seen that

$$\frac{1}{\sqrt{T}} \max_{0\leq t\leq 1} |Q^{o}_{T}(t)| \to^{p} \max_{0\leq t\leq 1} \max(t(1-\lambda), (1-t)\lambda) = \lambda(1-\lambda), \tag{28}$$

cf. equation (34) of PK. The limit of (28) is clearly symmetric in λ. It can also be seen that the limit in (26) is no greater than the limit in (28) and that both limits have the same maximum 1/4 at $\lambda = 1/2$. As the critical values of the OLS-MOSUM test are greater than those of the OLS-CUSUM test, the latter rejects whenever the former rejects. Therefore, the OLS-CUSUM test dominates the OLS-MOSUM test for all possible change points.

The results of power simulation under a single structural change with the break point $\lambda = 0.1, 0.2, \ldots, 0.9$ and parameter jumps $\Delta = 0.4, 0.6$ are summarized in Table 3. These results basically agree with the qualitative findings of the analytic results above. We also observe the following.

1. The BDE-CUSUM test performs the best when the break point is at 0.1, the BH-MOSUM test performs the best when the break point is at 0.2, and the OLS-CUSUM test outperforms the other tests for other break points.

2. When the break point occurs at and after 0.3, the BDE-CUSUM test is dominated by the other three tests quite significantly.

3. The OLS-CUSUM test beats the OLS-MOSUM test for all possible change points.

4. Owing to the asymmetric performance of the BH-MOSUM test, the BH-MOSUM test outperforms the OLS-MOSUM test when the break point occurs before 0.5 but is beaten by the OLS-MOSUM test when the break point occurs after 0.5,

We conclude from these simulation results that, under a single structural change, the OLS-CUSUM (BDE-CUSUM) test is the most (least) favorable except when the break point occurs at the beginning of the sample. Also, the BH-MOSUM test performs quite well when the break point occurs at and before 0.5.

[Table 3 About Here]

Under the alternative of double structural changes:

$$\beta_i = \begin{cases} \beta_0, & i = 1, \cdots, [T\lambda_1], \\ \beta_0 + \Delta_1, & i = [T\lambda_1] + 1, \cdots, [T\lambda_2], \\ \beta_0 + \Delta_2, & i = [T\lambda_2] + 1, \cdots, T, \end{cases} \tag{29}$$

where $\Delta_1 \neq \Delta_2$ and $\Delta_1 \neq 0$, we may also derive analytic results from Corollary 4.3, analogous to (25) and (26). However, as we are unable to compute the size-corrected power function, we do not pursue these technical details. Our simulation is based on $\Delta_1 = 0.4, 0.6$ and $\Delta_2 = 0$. We fix the first break point at 0.3 and consider the second break point $\lambda_2 = 0.4, 0.5, \ldots, 0.9$. These results are collected in Table 4 from which we can see that the following ordering holds for all models under consideration:

$$\text{OLS-MOSUM} \succ \text{BH-MOSUM} \succ \text{OLS-CUSUM} \succ \text{BDE-CUSUM}.$$

where $\succ$ is used to denote "performs better than". In fact, both the BH- and OLS-MOSUM tests dominate the two CUSUM tests quite significantly. Note also that the OLS-MOSUM test has the greatest power when the duration of parameter changes equals the size of moving sum, i.e., $\lambda_2 - \lambda_1 = 0.5$.

[Table 4 About Here]

Instead of postulating that the parameter jumps suddenly from one regime to another as in (22), we consider the case that there is a smooth transition between two regimes:

$$\beta_i = \begin{cases} \beta_0, & i = 1, \cdots, [T\lambda_1], \\ \beta_0 + \Delta_i, & i = [T\lambda_1] + 1, \cdots, [T\lambda_2], \\ \beta_0 + \Delta, & i = [T\lambda_2] + 1, \cdots, T, \end{cases} \tag{30}$$

where

$$\Delta_i = \Delta \left(\frac{i - [T\lambda_1]}{[T\lambda_2] - [T\lambda_1]} \right).$$

That is, during the transition period between $[T\lambda_1]$ and and $[T\lambda_2]$, the parameter β_i increases linearly from β_0 to a new level $\beta_0 + \Delta$. We simulate the case where $\Delta = 0.4$, 0.6 and transition periods 0.3–0.5, 0.3–0.7, and 0.3–0.9. These results are summarized in Table 5. We find that the OLS-CUSUM (BDE-CUSUM) test is the most (least) favorable in the models considered, while the BH-MOSUM test also performs very well. These conclusions are similar to those under a single structural change.

[Table 5 About Here]

We further consider the case that the parameter first has a sudden jump and then gradually returns to its original level. The model is

$$\beta_i = \begin{cases} \beta_0, & i = 1, \cdots, [T\lambda_1], \\ \beta_0 + \Delta_i, & i = [T\lambda_1] + 1, \cdots, [T\lambda_2], \\ \beta_0, & i = [T\lambda_2] + 1, \cdots, T, \end{cases} \tag{31}$$

where

$$\Delta_i = \Delta \left(\frac{[T\lambda_2] - i + 1}{[T\lambda_2] - [T\lambda_1]} \right),$$

This model may be more realistic than the model of double structural changes (29), because an economic relationship may change after a sudden shock but will gradually adjust back to the original level. In this simulation, Δ is taken to be 0.4 and 0.6, and the change periods are again 0.3–0.5, 0.3–0.7, and 0.3–0.9. These results are summarized in Table 6.

The conclusions here are similar to those under double structural changes. It can be seen that the OLS-MOSUM (BDE-CUSUM) test is the most (least) favorable and that the BH-MOSUM test performs better than the OLS-CUSUM test when the change period is long.

[Table 6 About Here]

Harvey (1975) proposes a different variance estimate for the CUSUM test:

$$\tilde{\sigma}^2 = \frac{1}{T-k} \sum_{i=k+1}^{T} (e_i - \bar{e})^2,$$

where $\bar{e}$ is the average of e_i. We have found that the general conclusions of power performance are not altered if this variance estimate is used.

5 Summary

In this paper, two MOSUM tests for parameter constancy are proposed, and their asymptotic null distributions are characterized analytically. In particular, a formula representing the boundary-crossing probability of the increments of a standard Wiener process is derived. Our analytic result shows that both the MOSUM tests are consistent and have non-trivial local power against a general class of alternatives. Our simulation results indicate that the BDE-CUSUM test performs quite poorly under various alternatives and that the OLS-CUSUM test usually outperforms other tests when the parameter basically obeys two regimes, whether the transition between these two regimes is abrupt or smooth. We also find that the BH-MOSUM test can complement other CUSUM tests if a single structural change is present, and the OLS-MOSUM test dominates the other tests when the parameter first changes to a new level and then returns to the original level. However, we have only obtained the boundary-crossing probability for the MOSUM process such that the size of moving sums equals one half of the entire sample, i.e., $h = 1/2$. The general case where $h < 1/2$ is currently under investigation.

Appendix

Chu, Hornik, & Kuan (1992) prove the following lemma.

Lemma A *Let X_T be a sequence of random processes in $D([0,1])^k$ converging in distribution (with respect to the Skorohod topology) to a random process X in $C([0,1])^k$ (i.e., the limiting process has continuous paths). Further, let $0 < h_T < 1$ be a sequence converging to $0 < h < 1$, and let $\kappa_T : [0, 1-h] \to [0, 1-h_T]$ be a sequence of maps such that $\sup_{0\le t\le 1-h} |\kappa_T(t) - t|$ tends to zero. Then, if Z_T is the random process on $D([0,1-h])^k$ given by*

$$Z_T(t) = X_T(\kappa_T(t) + h_T) - X_T(\kappa_T(t)),$$

we have $Z_T \Rightarrow Z$, where for $0 \le t \le 1-h$, $Z(t) = X(t+h) - X(t)$.

Proof of Theorem 2.1. We apply Lemma A by setting $X_T = Q_T^\tau$, $X = W$, $\kappa_T(t) = [N_\tau t]/\tau$ and $h_T = [\tau h]/\tau$. Clearly, $[\tau h]/\tau \to h$, and

$$\left|\frac{[N_\tau t]}{\tau} - t\right| \le \frac{1}{\tau}$$

so that $\sup_{0\le t\le 1-h} |\kappa_T(t) - t| \to 0$ as $T \to \infty$. It follows from Lemma A that

$$\begin{aligned} &\left(Q_T^\tau\left(\frac{[N_\tau t] + [\tau h]}{\tau}\right) - Q_T^\tau\left(\frac{[N_\tau t]}{\tau}\right), 0 \le t \le 1-h\right) \\ &\Rightarrow \quad (W(t+h) - W(t),\ 0 \le t \le 1-h). \end{aligned}$$

This proves the first assertion. In light of (5), the second assertion follows immediately from the continuous mapping theorem. □

Proof of Theorem 2.2. The assertions also follow from Lemma A and the continuous mapping theorem, as in the proof of Theorem 2.1. □

Proof of Lemma 3.1: Observe that $\tilde{W}_h(t) = W(th)/\sqrt{h}$ is again a Wiener process, hence, as

$$\frac{1}{\sqrt{h}} S_h^\tau(t) = \frac{W(t+h) - W(t)}{\sqrt{h}} = \tilde{W}_h(t/h + 1) - \tilde{W}_h(t/h),$$

we obtain by writing $u = t/h$ that

$$\begin{aligned} \max_{0\le t\le 1-h} |S_h^\tau(t)| \quad &= \quad \max_{0\le u\le (1-h)/h} |\tilde{W}_h(u+1) - \tilde{W}_h(u)| \\ &=^d \quad \max_{0\le u\le 1/h-1} |W(u+1) - W(u)|. \quad \square \end{aligned}$$

Proof of Lemma 3.2: Clearly, ΔW is zero-mean Gaussian process with continuous paths. Also, for $0 \leq s \leq t \leq 1$,

$$
\begin{aligned}
&\mathrm{cov}(\Delta W(t), \Delta W(s)) \\
&= \mathrm{cov}(W(t+1), W(s+1)) - \mathrm{cov}(W(t+1), W(s)) \\
&\qquad -\mathrm{cov}(W(t), W(s+1)) + \mathrm{cov}(W(t), W(s)) \\
&= s + 1 - s - t + s \\
&= 1 + (s - t).
\end{aligned}
$$

Thus, ΔW has linear covariance $1 - |t - s|$. The lemma now follows immediately from Shepp (1966, pp. 345–347). □

Proof of Lemma 3.3: In view of (10) and (11), we must compute

$$\frac{1}{\sqrt{\pi}} \int_\alpha^\beta e^{-(2x+\sqrt{2}u)^2/4} \phi(u + 2k(\beta - \alpha))\, du$$

and

$$\frac{1}{\sqrt{\pi}} \int_\alpha^\beta e^{-(2x+\sqrt{2}u)^2/4} \phi(u - 2\beta - 2k(\beta - \alpha))\, du,$$

for $\alpha = (a - x)/\sqrt{2}$ and $\beta = (b - x)/\sqrt{2}$. It is straightforward to see that

$$
\begin{aligned}
&\int_\alpha^\beta e^{-(2x+\sqrt{2}u)^2/4} \phi(u + \sqrt{2}v)\, du \\
&= \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-(2x+\sqrt{2}u)^2/4} e^{-(u+\sqrt{2}v)^2/2}\, du \\
&= \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-(\sqrt{2}u+x+v)^2/2}\, e^{-(x-v)^2/2}\, du \\
&= \frac{1}{\sqrt{2}} e^{-(x-v)^2/2} \int_{\sqrt{2}\alpha+x+v}^{\sqrt{2}\beta+x+v} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \frac{1}{\sqrt{2}} e^{-(x-v)^2/2} \Big(\Phi(\sqrt{2}\beta + x + v) - \Phi(\sqrt{2}\alpha + x + v) \Big).
\end{aligned}
$$

Setting $\alpha = (a - x)/\sqrt{2}$, $\beta = (b - x)/\sqrt{2}$, and $v = \sqrt{2}k(\beta - \alpha) = k(b - a)$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{\pi}} \int_\alpha^\beta e^{-(2x+\sqrt{2}u)^2/4} \phi(u + 2k(\beta - \alpha))\, du \\
&= \frac{1}{\sqrt{2\pi}} e^{-(x-k(b-a))^2/2} \Big(\Phi(b + k(b - a)) - \Phi(a + k(b - a)) \Big) \\
&= \phi(x - k(b - a)) \Big(\Phi(b + k(b - a)) - \Phi(a + k(b - a)) \Big),
\end{aligned}
$$

and similarly for $v = \sqrt{2}(-\beta - k(\beta - \alpha)) = -(b - x + k(b - a))$,

$$\begin{aligned}
&\frac{1}{\sqrt{\pi}} \int_{\alpha}^{\beta} e^{-(2x+\sqrt{2}u)^2/4}\, \phi(u - 2\beta - 2k(\beta - \alpha))\, du \\
&= \frac{1}{\sqrt{2\pi}} e^{-(b+k(b-a))^2/2} \Big(\Phi(x - k(b-a)) - \Phi(x - (k+1)(b-a)) \Big) \\
&= \phi(b + k(b-a)) \Big(\Phi(x - k(b-a)) - \Phi(x - (k+1)(b-a)) \Big).
\end{aligned}$$

The first assertion (12) now follows by combining these results and rearranging terms. The second assertion follows immediately form (12) by substituting $-b$ for a. □

Proof of Theorem 3.4: Obvious from the text. □

Proof of Corollary 3.5: The assertions follow from Theorem 3.4 and (16). □

Proof of Equation (17): We have from (14)

$$\begin{aligned}
&\mathbb{P}\{\Delta W(t) \le b \text{ for all } 0 \le t \le 1\} \\
&= \int_{-\infty}^{b} \left(\Phi(b) - \frac{\phi(b)}{\phi(x)} \Phi(b) \right) \phi(x)\, dx \\
&= \Phi(b)^2 - \phi(b) \int_{-\infty}^{b} \Phi(x)\, dx \\
&= \Phi(b)^2 - \phi(b) \left((x\Phi(x) + \phi(x) \Big|_{-\infty}^{b} \right) \\
&= \Phi(b)^2 - \phi(b)(b\Phi(b) + \phi(b)). \quad \square
\end{aligned}$$

Proof of Theorem 4.1: The proof of this theorem relies heavily on the results in the proof of Theorem 2 of KPA. If $c = 1/2$, we follow KPA to write

$$\begin{aligned}
y_i^T &= x_i^{T\prime} \beta_i^T + \epsilon_i \\
&= x_i^{T\prime} \beta_0 + \frac{1}{\sqrt{T}} x_i^{T\prime} g(i/T) + \epsilon_i.
\end{aligned}$$

Because x may contain lagged y, hence x_i also depends on T under local alternatives. Under the alternative (18), denote the OLS estimator at time n as

$$\tilde{\beta}_n^T = (X_n^{T\prime} X_n^T)^{-1} X_n^{T\prime} Y_n^T.$$

and recursive residuals as $e_n^T = (y_n^T - x_n^{T\prime} \tilde{\beta}_{n-1}^T)/v_n^T$. The cumulated sums of recursive residuals under local alternatives can be written as

$$\frac{1}{\hat{\sigma}\sqrt{T}} \sum_{n=k+1}^{k+[\tau t]} \frac{y_n^T - x_n^{T\prime} \tilde{\beta}_{n-1}^T}{v_n^T}$$

$$
\begin{aligned}
&= \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{x_n^T{}'\beta_0 + T^{-1/2} x_n^T{}' g(n/T) + \epsilon_n - x_n^T{}'\tilde{\beta}_{n-1}^T}{v_n^T} \\
&= \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{\epsilon_n - x_n'(\hat{\beta}_{n-1} - \beta_0)}{v_n^T} \\
&\quad + \frac{1}{\hat{\sigma}\sqrt{T\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{x_n^T{}' g(n/T)}{v_n^T} \\
&\quad + \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{x_n'(\hat{\beta}_{n-1} - \beta_0) - x_n^T{}'(\tilde{\beta}_{n-1}^T - \beta_0)}{v_n^T}. \qquad (32)
\end{aligned}
$$

The first term of (32) is just $Q_T^\tau(t)$ under the null hypothesis, hence converges weakly to $W(t)$. KPA show that the second term of (32) converges weakly to

$$
\frac{1}{\sigma}\mu' \int_0^t g(u)\, du,
$$

and the third terms behaves like

$$
-\frac{1}{\hat{\sigma}\sqrt{T\tau}} \sum_{n=k+1}^{k+[\tau t]} x_n' \frac{T}{n-1} \left(\frac{X_{n-1}' X_{n-1}}{n-1} \right)^{-1} \frac{1}{v_n^T} \frac{1}{T} \sum_{i=1}^{n-1} x_i x_i' g(i/T) \qquad (33)
$$

which converges weakly to

$$
\frac{-1}{\sigma}\mu' \int_0^t \left(\frac{1}{u} \int_0^u g(v)\, dv \right) du.
$$

In view of (4), we can apply the above limits to get

$$
\begin{aligned}
S_{T,h}^\tau(t) &= \frac{1}{\hat{\sigma}\sqrt{\tau}} \left(\sum_{n=k+1}^{k+[N_\tau t]+[\tau h]} e_n^T - \sum_{n=k+1}^{k+[\tau t]} e_n^T \right) \\
&\Rightarrow S_h^\tau(t) + \frac{1}{\sigma}\mu' \int_t^{t+h} \left(g(u) - \frac{1}{u} \int_0^u g(v)\, dv \right) du.
\end{aligned}
$$

This proves the second assertion. If $0 \leq c < 1/2$,

$$
y_i^T = x_i^T{}'\beta_0 + T^{-c} x_i^T{}' g(i/T) + \epsilon_i,
$$

and

$$
\begin{aligned}
& T^{c-1/2} \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{y_n^T - x_n^T{}'\tilde{\beta}_{n-1}^T}{v_n^T} \\
&= T^{c-1/2} \frac{1}{\hat{\sigma}\sqrt{\tau}} \sum_{n=k+1}^{k+[\tau t]} \frac{x_n^T{}'\beta_0 + T^{-c} x_n^T{}' g(n/T) + \epsilon_n - x_n^T{}'\tilde{\beta}_{n-1}^T}{v_n^T}
\end{aligned}
$$

$$= T^{c-1/2}\frac{1}{\hat\sigma\sqrt{\tau}}\sum_{n=k+1}^{k+[\tau t]}\frac{\epsilon_n - x_n'(\hat\beta_{n-1}-\beta_0)}{v_n^T}$$
$$+\frac{1}{\hat\sigma\sqrt{T\tau}}\sum_{n=k+1}^{k+[\tau t]}\frac{x_n^{T\prime} g(n/T)}{v_n^T}$$
$$+T^{c-1/2}\frac{1}{\hat\sigma\sqrt{\tau}}\sum_{n=k+1}^{k+[\tau t]}\frac{x_n'(\hat\beta_{n-1}-\beta_0)-x_n^{T\prime}(\tilde\beta_{n-1}^T-\beta_0)}{v_n^T}. \tag{34}$$

Owing to the presence of $T^{c-1/2}$ and $c < 1/2$, the first term of (34) above converges to zero; the probability limit of the second term is again

$$\frac{1}{\sigma}\mu'\int_0^t g(u)\,du.$$

Note that the third term in (34) differs from the third term in (32) by a factor $T^{c-1/2}$; but in view of equations (55) and (56) of KPA, they have the same limit

$$\frac{-1}{\sigma}\mu'\int_0^t\left(\frac{1}{u}\int_0^u g(v)\,dv\right)\,du.$$

Hence, for $0 \le c < 1/2$, the probability limit of $T^{c-1/2}MS_{T,h}^\tau$ is the same as the deterministic part of the limit under local laternatives. This gives the first assertion and completes the proof. □

Proof of Theorem 4.2: We use the same notations as in the preceding proof. By Theorem 2 of PK, we have under local alternatives,

$$\frac{1}{\hat\sigma\sqrt{T}}\sum_{n=1}^{[Tt]}(y_n^T - x_n^T\tilde\beta_T^T) \Rightarrow W^0(t) + \sigma^{-1}\mu' J_h^o g(t).$$

Hence, the assertions can be proved along the the same line as the proof of Theorem 4.1. We omit the details. □

Proof of Corollary 4.3: By Theorem 4.1,

$$\frac{1}{\sqrt{T}}MS_{T,h}^\tau \to^p \frac{1}{\sigma_0}\max_{0\le t\le 1-h}|J_h^\tau g(t)|.$$

Now

$$\begin{aligned} J_h^\tau g(t) &= \int_t^{t+h}\left(g(u)-\frac{1}{u}\int_0^u g(v)\,dv\right)du \\ &= \begin{cases} 0, & t \le \lambda - h, \\ \lambda\log((t+h)/\lambda)\Delta, & \lambda - h \le t \le \lambda, \\ \lambda\log((t+h)/t)\Delta, & t \ge \lambda, \end{cases}\end{aligned}$$

attains the maximum of its absolute value over $[0, 1-h]$ at $t = \min(\lambda, 1-h)$ with value $\lambda \log(\min(\lambda + h, 1)/\lambda)|\Delta|$. This proves the first assertion. By Theorem 4.2,

$$\frac{1}{\sqrt{T}} MS^o_{T,h} \to^p \sigma_0^{-1} \max_{0 \le t \le 1-h} |J^o_h g(t)|.$$

We observe that $\int_t^{t+h} g(u)\,du$ is non-decreasing in t and

$$\int_t^{t+h} g(u)\,du = \begin{cases} \max(h-\lambda,\, 0)\Delta, & t = 0, \\ \min(1-\lambda,\, h)\Delta, & t = 1-h. \end{cases}$$

Consequently,

$$\begin{aligned} |J^o_h g(t)| &= \left| \int_t^{t+h} g(u)\,du - h \int_0^1 g(u)\,du \right| \\ &= \max\left(|[\max(h-\lambda,\, 0) - h(1-\lambda)]\Delta|,\ |[\min(1-\lambda,\, h) - h(1-\lambda)]\Delta|\right) \\ &= \max\left(\min(\lambda(1-h),\, h(1-\lambda)),\ \min((1-\lambda)(1-h),\, h\lambda)\right)|\Delta|. \end{aligned}$$

This establishes the second assertion. □

References

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

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Brown, R. L., J. Durbin, & J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society*, Series B, **37**, 149–163.

Chu, C.-S. J., K. Hornik, & C.-M. Kuan (1992). A moving-estimates test for parameter stability and its boundary-crossing probability, BEBR Working Paper 92-0148, College of Commerce, University of Illinois, Urbana-Champaign.

Harvey, A. (1975). Comment on the paper by Brown, Durbin, & Evans. *Journal of the Royal Statistical Society*, Series B, **37**, 179–180.

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

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

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

Qualls, C., & H. Watanabe (1972). Asymptotic properties of Gaussian processes, *Annals of Mathematical Statistics*, **43**, 580–596.

Sen, P. K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **52**, 203–218.

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

Shepp, L. (1966). Radon-Nikodym derivatives of Gaussian process, *Annals of Mathematical Statistics*, **37**, 312–354.

Table 1: The Asymptotic Critical Values b^* of the MOSUM Tests.

MOSUM Tests	Probabilities					
	0.80	0.85	0.90	0.95	0.975	0.99
BH	1.57368	1.67357	1.80345	2.00350	2.18316	2.39798
OLS	1.21803	1.28636	1.37506	1.51151	1.63408	1.78082

Note: The critical values of the MOSUM tests are solved using *Mathematica*. For the BH-MSOUM test, we use $k = -5, \cdots, 5$ in the summation of (14), even though we notice that the effective terms are $k = -1, 0, 1$. For the OLS-MOSUM test, we use 5 terms in the summation of (15).

Table 2: Size Simulation of the MOSUM and CUSUM Tests.

Tests	$\alpha = 5\%$			$\alpha = 10\%$		
	T=100	T=200	T=300	T=100	T=200	T=300
BDE-CUSUM	4.04	4.14	4.51	8.37	8.67	8.94
OLS-CUSUM	3.26	3.83	4.25	7.57	8.58	8.56
BH-MOSUM	3.64	3.74	4.33	7.71	8.66	8.49
OLS-MOSUM	3.02	3.58	4.00	6.72	7.84	7.96

Note: The numbers in the table are empirical sizes (in percentages). Observations are generated from i.i.d. $N(2, 1)$; The number of replications is 10,000.

Table 3: Power Simulation of the Model (22): A Single Structural Change.

λ	Parameter Change: $\Delta = 0.4$				Parameter Change: $\Delta = 0.6$			
	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM
0.1	23.28	14.36	20.52	12.56	36.92	21.52	32.88	15.84
	36.88	19.76	31.56	16.04	63.48	35.28	55.56	23.48
0.2	28.52	27.96	32.76	20.48	49.80	51.36	57.92	33.88
	48.96	46.96	54.44	33.60	78.76	81.56	84.56	58.40
0.3	28.16	43.40	41.68	31.04	52.56	73.36	72.92	52.76
	49.00	70.24	67.96	51.56	80.24	95.96	94.24	81.92
0.4	25.76	50.80	47.40	39.72	50.60	82.96	78.96	68.60
	44.12	77.92	72.00	64.96	80.76	97.84	97.12	93.56
0.5	21.80	53.96	46.76	44.40	42.04	85.24	79.76	77.48
	39.84	81.68	75.00	73.60	71.64	98.88	97.68	97.40
0.6	17.16	52.52	35.92	39.20	30.68	82.36	65.08	68.56
	32.20	79.28	61.28	66.00	58.00	98.04	91.64	93.92
0.7	13.72	42.56	25.36	31.20	20.56	74.92	46.96	54.28
	20.00	67.44	42.68	50.72	39.80	95.64	73.48	81.76
0.8	11.08	28.40	16.24	20.28	12.24	52.20	24.44	32.92
	14.20	47.40	22.16	32.92	21.28	81.72	41.88	56.76
0.9	09.04	14.88	10.64	13.00	10.40	24.20	13.32	17.56
	10.20	19.68	11.16	16.60	11.76	37.76	15.84	24.92

Note: The first and second numbers in each cell are empirical power (in percentages) of the samples 100 and 200, respectively, based on empirical critical values at 10% level. The other tables use the same convention. For all power simulations, $\beta_0 = 2$, ϵ_i i.i.d. $N(0,1)$, and the number of replications is 2500.

Table 4: Power Simulation of the Model (29): Double Structural Changes with $\lambda_1 = 0.3$.

λ_2	Parameter Changes: $\Delta_1 = 0.4$, $\Delta_2 = 0$				Parameter Changes: $\Delta_1 = 0.6$ $\Delta_2 = 0$			
	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM
0.4	11.12	11.64	11.92	12.84	13.64	16.12	15.08	17.04
	13.04	15.60	14.36	17.84	18.64	21.00	20.64	23.08
0.5	14.88	16.32	19.32	22.20	20.96	26.00	31.00	31.96
	21.32	23.72	26.80	31.52	35.40	43.52	50.72	55.40
0.6	16.32	21.12	23.32	31.36	29.20	36.92	42.72	53.80
	29.60	34.52	39.00	51.40	53.92	63.60	69.80	82.36
0.7	21.68	23.60	31.04	39.64	37.68	44.52	55.84	70.08
	36.80	40.80	51.60	66.76	64.48	73.36	83.44	93.56
0.8	24.28	25.80	38.80	45.64	41.04	45.96	66.36	76.84
	40.04	41.72	59.88	71.40	74.20	79.24	91.60	96.60
0.9	25.80	29.88	40.12	39.20	47.04	52.92	67.04	68.04
	47.44	53.60	66.92	66.36	77.92	84.32	92.64	92.96

Table 5: Power Simulation of the Model (30): Smooth Transition between Two Regimes.

Change Period	Parameter Changes from 2 to 2.4				Parameter Changes from 2 to 2.6			
	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM
0.3–0.5	25.92	49.64	44.84	36.36	47.44	80.44	76.28	65.56
	44.96	76.32	71.68	62.68	80.80	97.20	96.24	91.84
0.3–0.7	22.40	48.48	40.36	35.72	37.12	77.36	67.56	61.84
	37.44	72.28	64.08	59.36	70.24	96.04	93.32	89.92
0.3–0.9	16.24	39.84	29.96	26.32	27.88	65.76	52.56	49.00
	25.96	62.68	48.28	45.12	53.64	91.08	82.72	77.52

Table 6: Power Simulation of the Model (31): Parameter has a Sudden Jump and Then Declines Gradually.

Change Period	Initial Parameter Change: $\Delta = 0.4$				Initial Parameter Change: $\Delta = 0.6$			
	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM	BDE-CUSUM	OLS-CUSUM	BH-MOSUM	OLS-MOSUM
0.3–0.5	10.52	12.40	12.56	12.56	13.00	15.04	14.84	16.88
	12.56	15.20	14.40	16.24	17.36	21.08	21.72	25.12
0.3–0.7	13.24	14.92	16.80	17.16	17.28	20.72	21.80	27.88
	17.16	19.84	22.28	27.00	30.24	33.96	38.16	50.12
0.3–0.9	15.28	15.88	20.20	22.24	24.32	23.76	32.60	40.32
	23.36	23.00	29.52	35.92	39.60	38.96	53.64	66.08