Sequential testing with uniformly distributed size

Stanislav Anatolyev*
CERGE-EI and New Economic School

Grigory Kosenok
New Economic School

September 2017

Abstract

Sequential procedures for the testing for structural stability do not provide enough guidance on the shape of boundaries that are used to decide on acceptance or rejection, requiring only that the overall size of the test is asymptotically controlled. We introduce and motivate a reasonable criterion for the shape of boundaries which requires that the test size be uniformly distributed over the testing period. Under this criterion, we numerically construct boundaries for the most popular sequential tests that are characterized by a test statistic behaving asymptotically either as a Wiener process or Brownian bridge. We handle this problem both in the context of retrospecting a historical sample and in the context of monitoring newly arriving data. We tabulate the boundaries by fitting them to certain flexible yet parsimonious functional forms. Interesting patterns emerge in an illustrative application of sequential tests to the Phillips curve model.

Key words: Structural stability; sequential tests; CUSUM; retrospection; monitoring; boundaries; asymptotic size.

*Corresponding author. Address: Stanislav Anatolyev, CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic; e-mail stanislav.anatolyev@cerge-ei.cz. We are grateful to the Editor Javier Hidalgo and two anonymous referees for useful suggestions that significantly improved the paper. We have benefited from discussions with Helmut Lütkepohl, Franco Peracchi and Vladimir Spokoiny. Our thanks also go to numerous seminar audiences, especially at the Bar-Ilan University, CIREQ, Humboldt Universität zu Berlin, EIEF, EUI, Koç University, Monash University, QUT, Università di Bologna, University of Haifa, University of Helsinki, Universität Zürich, as well as the audience at the 70th European Meeting of the Econometric Society in Lisbon.

1 Introduction

From mid-seventies onward, one can encounter applications of sequential testing tools in applied econometric and statistical work. Sequential testing methods are usually used in the context of testing for the structural stability of coefficients in a regression, although not necessarily. The CUSUM and CUSUM of squares tests introduced in Brown, Durbin, and Evans (1975) belong to this class and can be found in many textbooks, including those of introductory level.

Consider a linear regression framework of testing for structural stability. Of interest for us is the stability of the regression relationship

$$y_\tau = x'_\tau \beta_\tau + u_\tau \tag{1}$$

over time indexed by τ. Formally, the stability of the regression relationship (1) is formulated as the hypothesis $H_0 : \beta_\tau = \beta$ for all τ, where β is unknown. From this point, one may take a number of approaches to test this null. One approach starts from the formulation of a specific alternative hypothesis that assumes a particular type of non-stability of coefficients (e.g., a single structural break), and proceeds by constructing a test designed specifically for this alternative. Such test is expected to also have power against alternatives other than the one to which it is tailored. Typically, this specific alternative assumes one or several abrupt changes in coefficients at specific dates. This leads to a usual decision rule when a scalar test statistic is compared to a critical value. A radically different approach, *sequential* or *recursive testing*, avoids specifying any particular alternative to construct a *sequential statistic*, which is a sequence of the same statistic computed over (usually) an expanding time window. The decision rule involves a comparison of a *trajectory* or *path* (i.e. a sequence of values) of this sequential statistic to a *boundary* (i.e. a sequence of separate critical values for each time period). The typical outcome of a sequential test is 'do not reject' if the entire trajectory stays below the boundary, and 'reject' if it crosses the boundary at least once.[1]

It is worth mentioning the relationship of sequential testing to multiple testing, i.e. testing a finite number of hypotheses using the same dataset. A sequential test can be viewed as a collection of many identical one-shot tests of the same null hypothesis of stability on

[1]Some existing tests are in fact a mixture of the two extreme approaches. For example, the test for a single structural break of Andrews (1993) has a structure of a classical test, but can be interpreted as sequential tests with a particular boundary (see section 2.2).

(usually) an expanding dataset. As a result, the one-shot test statistics (i.e. those specific for each τ) are highly dependent, and accounting for their multiplicity and dependence drives critical values up from their one-shot values in a systematic way, according to the asymptotic theory as the number of periods and hence the numerosity of one-shot tests tends to infinity. In contrast, under multiple testing, the number of the hypotheses being tested is limited and asymptotically fixed, and the null hypotheses are typically different and may be quite heterogenous.[2]

More formally, the sequential statistic Q_τ, also called a *detector*, of a sequential test is computed on various subintervals indexed by τ, typically on $[1+k, \tau]$, where k is dimensionality of β. Denote the corresponding boundary by b_τ. The decision rule is: reject H_0 if the path of Q_τ hits the boundary b_τ at least once, otherwise do not reject. The requirement for b_τ to be a valid boundary is that the test size α be asymptotically controlled. In case H_0 is rejected, as a by-product, the researcher gets an idea about the timing of the structural instability; note, however, that such speculations do not formally belong to an outcome of the testing procedure.

Before formulating our objectives and contribution, let us distinguish two environments where sequential testing is used. The first one is classical, which we call *retrospection*, when one tests for structural stability in a given historical sample, i.e. for $\tau = k+1, ..., T$. Most of the sequential testing tools are developed for this retrospective context; see references in Table 1 of Section 2. However, starting from Chu, Stinchcombe, and White (1996), researchers became interested in implementing sequential testing in the *monitoring* context by using data arriving in real time, i.e. for $\tau = T+1, T+2, ...$, conditional that H_0 holds for the historical interval; see references in Table 1 of Section 2. In this paper, we handle both retrospection and monitoring situations, placing more weight on the latter, as it is more natural for sequential testing and because it poses more interesting challenges.

The critical issue in sequential testing is the choice of which boundary to use. For the sake of simplicity, consider one-sided testing when rejection occurs for large positive values of a statistic. The only formal requirement imposed on the boundary is that the test size be controlled, which leaves many degrees of freedom as far as the boundary shape is concerned.[3]

[2]See Romano, Shaikh and Wolf (2010) for examples of multiple tests and related techniques.

[3]This slackness in the decision rule of sequential tests leads to an unpleasant phenomenon of different conclusions reached by different researchers and possibility of manipulating the testing outcome.

While in their original paper Brown, Durbin, and Evans (1975) derived linear boundaries for retrospective CUSUM tests, the choice of a linear shape is arbitrary. For example, Inclán and Tiao (1994) and Anatolyev (2008) used horizontal retrospective boundaries. In the monitoring context, Chu, Stinchcombe, and White (1996) derived so-called parabolic monitoring boundaries for some tests, where "parabolic" is an informal term indicating that the shape of such boundaries is close to a root of the time index. Later, Zeileis, Leisch, Kleiber, and Hornik (2005) criticized the parabolic shape and suggested linear monitoring boundaries instead. Indeed, one may suggest many legitimate boundaries with different shapes, since fixing the asymptotic test size, which is just one number, is insufficient to pin down the shape of a boundary. There is no consensus in the literature on which shape is more reasonable, although, clearly, the shape of boundaries strongly affects testing outcomes. The arguments that are typically given in favor of one shape or in criticism of another are twofold. The first argument is that some boundaries, such as horizontal retrospective and parabolic monitoring ones, can be derived analytically as functions of size in a closed form. The second argument is that with some boundaries, such as linear monitoring ones, the test size tends to be distributed more evenly over time than with others, even though one has to employ simulations to deduce parameters of their shape.

In this paper, we suggest a reasonable criterion that allows one to fix the shape of a boundary. This criterion requires that the prescribed asymptotic test size be *uniformly distributed* over the retrospective or monitoring interval. In other words, the likelihood of rejecting true stability in any specific time period, given that it is not rejected yet, does not depend on the time period. Such requirement leads to a fair and dynamically consistent testing procedure, as the test size is equally allocated to equal-sized time subintervals. For example, under structural stability, the type I error of rejecting stability during the first half of the historical sample should be equal to that during the second half. As we already know (see the previous paragraph), the monitoring literature tends to favor a more even distribution of size over time (in particular, linear boundaries are motivated as advantageous to parabolic ones). Arguably, in various circumstances, other distributions of size may also make perfect sense, in particular those with time discounting, possibly motivated by a desire to reject the truth earlier, if reject at all. In such cases, corresponding boundaries can be

constructed using the guidelines and algorithms we provide here.[4]

The next thing we do in this paper is a derivation of boundaries under the criterion of uniform size distribution, using the integral equations for determination of first passage probabilities (Durbin, 1971). While this famous result has been heavily used in the statistical literature to derive the distribution of size given a boundary, we switch the input and output and instead derive the boundary given a distribution of size. This task turns out to be much more challenging due to the nature of the integral equation and is further complicated by the presence of singularities. To attain our goal, we use numerical methods of integration and solving equations that also take account of the singularities. This technology has been previously applied by the authors in a related problem of finding critical values for the Andrews (1993) test (Anatolyev and Kosenok, 2011).

We construct the boundaries for the two large classes of tests most often encountered in the sequential testing literature. These classes are characterized by the processes that are asymptotic analogs of a detector: the Wiener process and Brownian bridge. We consider each of the two cases separately, managing both one- and two-sided testing. As mentioned before, in doing this, we handle both retrospection and monitoring situations. When the detector is asymptotically a Wiener process, it turns out that one "baseline" retrospective boundary derived for a particular value of size can be exploited in other situations (i.e. for other values of size and any finite monitoring horizons) by using a scaling transformation. That is, different boundaries are "homothetic" to each other, possibly after a rightward shift. The case where the detector is asymptotically a Brownian bridge is more complex. Here, in contrast, boundaries are specific for the value of size in the retrospection context, as well as for the monitoring horizon in the monitoring context.

Because the boundaries are computed numerically, we provide a tabulated version of the boundaries. We handle this by fitting the computed boundaries to parsimonious but flexible parametric functional forms, paying special attention to the boundary curvature near the points 0 and 1. The quality of fit is very high, so that the computed and parameterized boundaries are practically indistinguishable. We demonstrate via simulations that the

[4]A chosen shape of boundaries, motivated by a particular distribution of size, will of course affect the power of the test. However, according to the paradigm of statistical testing, power comes into play after the test size is controlled. In the case of a classical test with a scalar test statistic, control over the size means fixing its critical value(s). In the context of a sequential test with a functional test statistic, control over the size means controlling its entire distribution, which implies fixing the entire boundary.

parameterized boundaries do possess the property of distributing the size uniformly. We also conduct a simulation experiment that compares hitting rates and average delays across different tests and different boundaries.

Finally, we apply sequential testing tools to the Phillips curve model using US monthly data. We perform a few testing experiments, both retrospective and monitoring, using different boundaries and different testing intervals. The application illustrates interesting patterns that one can encounter in practice.

The rest of the paper is organized as follows. Section 2 gives some technical details on sequential testing. In Section 3, we describe the method of obtaining the boundaries with a particular distribution of test size, show the results of constructing the "uniform" boundaries, report asymptotic simulation results on the distribution of size, and compare hitting rates and average delays. In Section 4, we illustrate implementation of both our and alternative procedures in an empirical application. Finally, Section 5 concludes the paper. More technical material is relegated to the Appendix, while the Online Appendix at `is.gd/unisize` contains technical details about boundary parameterization.

2 Sequential testing: some details

2.1 Asymptotics

A sequential test has the following elements: a detector Q_τ, a boundary b_τ with the property

$$\Pr\{Q_\tau < b_\tau \ \forall \ \tau \in \mathcal{T} | H_0\} = 1 - \alpha$$

in case testing is one-sided, or

$$\Pr\{|Q_\tau| < b_\tau \ \forall \ \tau \in \mathcal{T} | H_0\} = 1 - \alpha$$

in case testing is two-sided,[5] where

$$\mathcal{T} = \begin{cases} \{k+1, k+2, ..., T-1, T\} & \text{in the retrospective context,} \\ \{T+1, T+2, ..., KT-1, KT\} & \text{in the monitoring context,} \end{cases}$$

and K is a finite monitoring horizon (see below on why K has to be finite), and, finally, a decision rule prescribing to reject structural stability if the detector hits the boundary at least once.

[5]It is conventional that two-sided boundaries are symmetric in the sense that the lower boundary is $-b_\tau$ when the upper boundary is b_τ for detector Q_τ that is asymptotically a symmetric gaussian process.

Usually, the test size can be controlled only asymptotically. As $T \to \infty$, we have

$$Q_{\lfloor rT \rfloor} \Rightarrow Q(r)$$

on $\mathcal{R}$ and

$$b_{\lfloor rT \rfloor} \to b(r),$$

where $Q(r)$ is the limiting continuous time process for the detector,

$$\mathcal{R} = \begin{cases} [0,1] & \text{in the retrospective context,} \\ [1,K] & \text{in the monitoring context,} \end{cases}$$

and $b(r)$ is a deterministic asymptotic boundary. Asymptotic control of the size means that

$$\Pr\{Q(r) < b(r) \ \forall\, r \in \mathcal{R} | H_0\} = 1 - \alpha$$

when testing is one-sided or

$$\Pr\{-b(r) < Q(r) < b(r) \ \forall\, r \in \mathcal{R} | H_0\} = 1 - \alpha$$

when testing is two-sided.

While $b(r)$ is arbitrary subject to the size control requirement, the asymptotic process $Q(r)$ depends on the detector used in testing. Table 1 contains a list of sequential tests drawn from the literature, with corresponding limiting distributions. Although there are more complicated limits, most often $Q(r)$ is one of the following two processes: the Wiener process W:

$$Q(r) = \begin{cases} W(r),\ r \in [0,1] & \text{for retrospection,} \\ W(r-1),\ r \in [1,K] & \text{for monitoring,} \end{cases}$$

or the Brownian bridge B:

$$Q(r) = \begin{cases} B(r),\ r \in [0,1] & \text{for retrospection,} \\ B(r),\ r \in [1,K] & \text{for monitoring,} \end{cases}$$

or, possibly, an absolute value thereof.[6] Usually, a detector is a normalized cumulative sum computed in an expanding window, possibly contrasted with a similar measure on the whole historical interval. Without contrasting, the asymptotic process is likely the Wiener process; with no contrasting, the asymptotic process is likely the Brownian bridge.

[6]Note that for the process B which is tied down at $r = 1$, the argument of asymptotic process is r irrespective of whether it is retrospection or monitoring, while for the untied process W the argument is $r - 1$ in case of monitoring. We focus on cases the asymptotic process starts off from the non-random value (typically, zero). The "uniform" boundaries in cases when the starting point is random would have a shape starting off from infinity because otherwise there would be a positive probability mass of the process concentrated above the boundary, which is incompatible with uniform size distribution over a continuum.

Test characterization	Sources	Limiting distribution $Q(r)$
Retrospective CUSUM	Brown, Durbin & Evans (1975), Ploberger, Krämer & Alt (1988)	$W(r),\ r \in [0,1]$
Sequential t-statistic	Anatolyev (2008)	$W(r),\ r \in [0,1]$
Monitoring CUSUM	Chu, Stinchcombe & White (1996)	$W(r-1),\ r \in [1,K]$
Monitoring based on prediction errors	Aue, Horváth, Hušková & Kokoszka (2006)	$\lvert W(r)\rvert,\ r \in [0,1]$
Monitoring with endogeneity	Kurozumi (2017)	$\lvert W(r)\rvert,\ r \in [0,1]$
Monitoring multivariate CUSUM	Groen, Kapetanios & Price (2013)	$\max_j \lvert W_p^{(j)}(r)\rvert,\ r \in [1,K]$
Retrospective CUSUM of squares	Brown, Durbin & Evans (1975), Deng & Perron (2008)	$B(r),\ r \in [0,1]$
Detection of variance change	Inclán & Tiao (1994)	$B(r),\ r \in [0,1]$
Retrospective OLS-based CUSUM	Ploberger & Krämer (1992)	$B(r),\ r \in [0,1]$
Fluctuation test	Ploberger, Krämer & Kontrus (1989), Kuan & Hornik (1995)	$B(r),\ r \in [0,1]$
Monitoring fluctuation test	Leisch, Hornik and Kuan (2000)	$B(r),\ r \in [1,K]$
Monitoring OLS-based CUSUM	Zeileis, Leisch, Kleiber & Hornik (2005)	$B(r),\ r \in [1,K]$
Monitoring volatility dynamics	Andreou and Ghysels (2002, 2006)	$B(r),\ r \in [0,1],\ r \in [1,K]$
MOSUM-type tests	Chu, Hornik & Kuan (1995), Leisch, Hornik & Kuan (2000)	$B(r) - B(r-h),\ r \in [h,K],\ h \in (0,1)$
Recursive predictability tests	Inoue and Rossi (2005)	$W_p(r)'W_p(r),\ r \in [0,1]$
Andrews stability sup-test	Anatolyev and Kosenok (2011)	$B_p(r)'B_p(r),\ r \in [\pi, 1-\pi]$
Predictive test for stability	Ghysels, Guay & Hall (1997)	$B_p(r)'B_p(r) + W_{q-p}(r)'W_{q-p}(r),\ r \in [0,1]$

Note: $W_p(r)$ is a p-variate Wiener process, $W_p^{(j)}(r)$ is its j^{th} element, $B_p(r)$ is a p-variate Brownian bridge.

Note: K is a monitoring horizon which is equal to ∞ in most cases.

Table 1: Examples of sequential tests with their limiting processes.

2.2 Boundaries

As explained above, in practice one uses a boundary from a small set of possibilities suggested in the literature. Let us list those suggestions that are documented in the literature, in the case of two sided testing.

When the asymptotic process is Wiener process W, and the context is retrospective, one has a choice between a linear boundary

$$b(r) = \lambda(2r + 1)$$

derived in Brown, Durbin and Evans (1975) for the CUSUM test, where $\lambda = 0.948$ for $\alpha = 5\%$, and a horizontal boundary from Anatolyev (2008):

$$b(r) = \mu,$$

where $\mu = 2.241$ for $\alpha = 5\%$. In an attempt to make the distribution of size relatively even, Zeileis (2004) suggests an ad hoc boundary

$$b(r) = \nu\sqrt{r}$$

where $\nu = 3.15$ for $\alpha = 5\%$, motivated by its proportionality to the standard deviation of the Wiener process. In the monitoring context, the parabolic boundary derived in Chu, Stinchcombe and White (1996) is

$$b(r) = \left(r \log \frac{r}{\alpha^2}\right)^{1/2}.$$

This is an exact formula presuming that the monitoring horizon is infinite.

When the asymptotic process is Brownian bridge B and the context is retrospective, the most widespread boundary is horizontal (e.g., Brown, Durbin and Evans, 1975; Inclán and Tiao, 1994)

$$b(r) = \mu,$$

where $\mu = 1.358$ for $\alpha = 5\%$, which is implicit in the usually used functional $\sup_{r\in[0,1]}$. An alternative choice is again suggested in Zeileis (2004):

$$b(r) = \nu\sqrt{r(1-r)},$$

where $\nu = 3.37$ for $\alpha = 5\%$, which is proportionate to the standard deviation of the Brownian bridge. The same boundary shape lies in the construction of the Andrews (1993) test, but

the test period is restricted to be $[\pi, 1-\pi]$ with $0 < \pi < \frac{1}{2}$. In the monitoring context, the leading choice is the (slightly curved) parabolic boundary derived in Chu, Stinchcombe and White (1996)

$$b(r) = \left(r(r-1) \left(a^2 + \log \frac{r}{r-1} \right) \right)^{1/2},$$

where a^2 depends only on α. The monitoring horizon is presumed infinite. A linear boundary

$$b(r) = \lambda r$$

was suggested in Zeileis, Leisch, Kleiber and Hornik (2005). The authors give critical values for λ for integer values of the monitoring horizon K from 2 to 10.

As stated in the Introduction, we aim at constructing the "uniform" boundaries, i.e. such that the size is uniformly distributed over the relevant testing horizon. More formally, when testing is one-sided, we want to find the retrospective boundary $b_\alpha^R(r)$ such that for all $s \in [0,1]$

$$\Pr\left\{ Q(r) < b_\alpha^R(r) \ \forall\, r \in [0,s] \,|H_0 \right\} = 1 - \alpha s$$

or the monitoring boundary $b_\alpha^M(r)$ such that for all $s \in [1,K]$

$$\Pr\left\{ Q(r) < b_\alpha^M(r) \ \forall\, r \in [1,s] \,|H_0 \right\} = 1 - \alpha \frac{s-1}{K-1}.$$

Similarly the "uniform" boundaries are defined when testing is two-sided.

Remark 1. Of course, uniform distribution over a monitoring period is possible only if the monitoring horizon K is finite.[7] In the monitoring literature, the monitoring horizon is typically infinite, which is an approximation for "very long" monitoring and is convenient for analytic work (for example, the parabolic boundaries are specific for ever-lasting monitoring and derived from certain statistical properties of the Wiener process, see Robbins and Siegmund, 1970). However, an infinite horizon is implausible in practice, and may be an inadequate approximation for "very long" monitoring, in cases when most of the size is "consumed" only after an implausibly long period of monitoring is elapsed (see an example below). Interestingly, some published simulation studies verify properties of tests relying on finite monitoring horizons, even though the boundaries are derived for the infinite horizon (e.g., Chu, Stinchcombe and White, 1996; Leisch, Hornik and Kuan, 2000).[8]

[7] If K was infinite, it would mean that the size accumulated with constant rate would also be infinite, which is impossible by the definition of the total test size.

[8] It is usually verified that the size actually used does not exceed the total size, and that the rate of its accummulation makes it unlikely that it will ever exceed the total size.

Remark 2. Though lying outside the sequential testing paradigm, interesting questions are what actions a user of the model is supposed to take when a rejection of stability occurs during the monitoring period or when the monitoring period elapses before a rejection is made. In the former case, arguably, the user reformulates the model, possibly adapting to a new prevailing DGP or expanding it to incorporate a change in the DGP. In the latter case, if still there is a need to keep the same model (or reformulated one as the given horizon was chosen for a reason) running, a new monitoring test may be launched, with a new size and horizon specific to the next purpose of its use.[9] Alternatively, one may prolongate the previous testing procedure by increasing its horizon and the total size accordingly, with continuous extension of the uniform boundary.

Figures 1a, 1b, 1c and 1d present distributions of size[10] in the four situations with the Wiener process described at the beginning of this subsection. One can easily see that in all cases the distribution of size is far from even. In particular, because the linear, horizontal and parabolic boundaries do not start off from zero, the chances of crossing it near the beginning of the testing period are very slim. Obviously, the "uniform" boundaries have to take off from zero, with an infinite slope. The Zeileis (2004) boundary does start off from zero and have an infinite slope, but, among other things, the curvature at zero is too high. Special attention deserves Figure 1d with parabolic monitoring boundaries corresponding to an infinite monitoring horizon. One can see that a significant portion of size corresponds to the period beyond $K = 10$; in fact, only about 2% out of 5% are used before $10T$ time periods elapsed (about 3% before $30T$ periods, and about 4% before $100T$ periods) . Suppose that the data are quarterly covering 25 years, so the historical interval has length $T = 100$. This means that even in 250 years even half of the prescribed size will not be used, and in a plausible exercise only a tiny fraction of it will. Hence, in practice the actual size in a plausible procedure is likely to have little to do with the nominal size, when the monitoring horizon is assumed infinite.

Figures 2a and 2b show a couple of situations for the Brownian bridge. Analogously, with the horizontal boundaries most of crossings are concentrated in the middle of the unit

[9]Determining a monitoring horizon is akin to determining a size of a test; both are fully at the discretion of the tester and are driven by her attitude towards risk.

[10]In these simulations, one million trajectories of an appropriate asymptotic process are generated. Each trajectory corresponding to the Wiener process is approximated by a relevant portion of a suitably normalized sum of 100,000 standard normals.

interval. With the Zeileis (2004) boundary (the suggestion closest to what should implement the idea of the "uniform" boundary), most of crossings lie near the endpoints of the unit interval.

From these figures one can see that if a boundary starts off too high, crossings near zero are very rare. Obviously, the "uniform" boundary has to start from zero at zero. On the other hand, it should start off steeply enough so that not the whole crossing mass is concentrated at zero. Two theorems below formalize these observations by stating conditions that rule out a possibility of uniform distribution of size. Let us denote

$$\ell(r) = \sqrt{2r \ln(-\ln r)}.$$

This is a familiar "knife-edge" boundary for the Brownian motion that figures in the "zero-time" law of iterated logarithms (LIL, see, e.g., Karatzas and Shreve, 1988, theorem 9.23i):

$$\overline{\lim_{r \downarrow 0}} \frac{W(r)}{\ell(r)} = 1. \tag{2}$$

Consider a smooth boundary $b(r)$. The first impossibility result covers the boundaries that start off not from (or, if applicable, arrive not at) zero. The second impossibility result rejects the boundaries that do start off from zero, but with an insufficiently high rate.

Theorem 1 *Uniform distribution of size is impossible using boundary* $b(r)$ *if*

(i) $b(0) > 0$ *when* $Q(r) = W(r)$, $r \in [0,1]$,

(ii) $b(0) > 0$ *or* $b(1) > 0$ *when* $Q(r) = B(r)$, $r \in [0,1]$,

(iii) $b(1) > 0$ *when* $Q(r) = W(r)$ *or* $Q(r) = B(r)$, $r \in [1,K]$.

Theorem 2 *Uniform distribution of size is impossible using boundary* $b(r)$ *if*

(i) $Q(r) = W(r)$, $r \in [0,1]$, $b(0) = 0$ *and*

$$\lim_{r \downarrow 0} \frac{b(r)}{\ell(r)} < 1. \tag{3}$$

(ii) $Q(r) = B(r)$, $r \in [0,1]$, $b(0) = b(1) = 0$ *and*

$$\lim_{r \downarrow 0} \frac{b(r)}{\ell(r)} < 1 \tag{4}$$

or

$$\lim_{r \uparrow 1} \frac{b(r)}{\ell(1-r)} < 1. \tag{5}$$

(iii) $Q(r) = W(r)$ *or* $Q(r) = B(r)$, $r \in [1, K]$, $b(1) = 0$ *and*

$$\lim_{r\downarrow 1} \frac{b(r)}{\ell(r-1)} < 1. \tag{6}$$

According to Theorem 1, the linear and horizontal boundaries in the retrospective context $b(r) = \lambda(2r+1)$ and $b(r) = \mu$ do not satisfy the necessary requirements, and so does not the linear boundary $b(r) = \lambda r$ in the monitoring context. According to Theorem 2, the retrospective Zeileis boundaries $b_R(r) = \nu\sqrt{r}$ when $Q(r) = W(r)$ and $b_M(r) = \nu\sqrt{r(1-r)}$ when $Q(r) = B(r)$ do not satisfy the necessary requirements:

$$\begin{aligned}
\lim_{r\downarrow 0} \frac{b_R(r)}{\ell(r)} &= \frac{\nu}{\sqrt{2}} \lim_{r\downarrow 0} \frac{1}{\sqrt{\ln(-\ln r)}} = 0, \\
\lim_{r\downarrow 0} \frac{b_M(r)}{\ell(r)} &= \frac{\nu}{\sqrt{2}} \lim_{r\downarrow 0} \sqrt{\frac{1-r}{\ln(-\ln r)}} = 0, \\
\lim_{r\uparrow 1} \frac{b_M(r)}{\ell(1-r)} &= \frac{\nu}{\sqrt{2}} \lim_{r\uparrow 1} \sqrt{\frac{r}{\ln(-\ln(1-r))}} = 0.
\end{aligned}$$

The retrospective parabolic boundary

$$b(r) = \left(r \ln \frac{r}{\alpha^2}\right)^{1/2}$$

does satisfy the necessary condition of Theorem 1 as $b(1) = 0$, but it does not satisfy that of Theorem 2:

$$\lim_{r\downarrow 1} \frac{b_M(r)}{\ell(r-1)} = \left(\lim_{r\downarrow 1} \frac{\ln r}{\ln(-\ln(r-1))}\right)^{1/2} = 0.$$

The monitoring parabolic boundary, however, $b(r) = \left(r(r-1)\left(a^2 + \ln r/(r-1)\right)\right)^{1/2}$ does satisfy both necessary conditions:

$$b_M(1) = \left(\lim_{r\downarrow 1} r(r-1)\left(a^2 + \ln \frac{r}{r-1}\right)\right)^{1/2} = \left(\lim_{s\to+\infty} \frac{a^2 + \ln(1+s)}{s}\right)^{1/2} = 0,$$

where the change of variable $s = 1/(r-1)$ is employed, and

$$\lim_{r\downarrow 1} \frac{b_M(r)}{\ell(r-1)} = \left(\frac{1}{2} \lim_{s\downarrow 0} \frac{a^2 + \ln(1+s^{-1})}{\ln(-\ln s)}\right)^{1/2} = +\infty,$$

where the change of variable $s = r - 1$ is employed.

3 Construction of "uniform" boundaries

3.1 Determination of boundaries

The integral equation relating the boundaries to first passage probabilities was derived in Durbin (1971). Subsequently, this technique was intensively used in the statistical literature

(in particular, numerous articles in subsequent issues of the *Journal of Applied Probability*) to derive the distribution of crossing probabilities for boundaries of various shape. Here, we "reverse" the usual procedure and derive the boundary for a particular (namely, uniform) distribution of crossing probabilities over the relevant interval. We show the technique in the retrospective context; the monitoring situation is handled similarly.

Denote by $p_r(y)$ the unconditional density of $Q(r)$, and by $p_{r|s}(y|x)$ the conditional density of $Q(r)$ given that $Q(s)$ took the value x. The exact forms of $p_r(y)$ and $p_{r|s}(y|x)$ will be specified later when we move on to concrete processes for $Q(r)$. Note that the process $Q(r)$ is assumed to be Markovian in a sense: the transition density $p_{r|s}(y|x)$ is supposed to depend on value x taken at s but not on the rest of the trajectory preceding s. This is true for our processes of interest, $W(r)$ and $B(r)$, as well as for, e.g., squared Bessel process (as in Inoue and Rossi, 2005) and squared Bessel Bridge (as in Anatolyev and Kosenok, 2011), but may not hold for other processes, e.g., the one at the very bottom of Table 1.

Let $\Psi(r)$ be a one-sided boundary on $[0,1]$ such that the distribution of size is $\alpha(r)$, $r \in [0,1]$. According to Durbin (1971, sec.2), it is implicitly defined by the integral equation

$$p_r(\Psi(r)) = \int_0^r p_{r|s}(\Psi(r)|\Psi(s))\, d\alpha(s) \tag{7}$$

that should hold for all $r \in [0,1]$. Intuitively, the meaning of the equality in (7) is the following: the unconditional density of $Q(r)$ at the boundary $\Psi(r)$ can be alternatively obtained via the law of total probability by counting, along the boundary from 0 to r, the total measure for those trajectories that pass through $\Psi(r)$ for the first time.

When $\Psi(r)$ is the upper (positive) part of the symmetric two-sided boundary on $[0,1]$, according to Durbin (1971, sec.4) it is implicitly defined by the integral equation

$$p_r(\Psi(r)) = \frac{1}{2}\int_0^r p_{r|s}(\Psi(r)|\Psi(s))\, d\alpha(s) + \frac{1}{2}\int_0^r p_{r|s}(\Psi(r)|-\Psi(s))\, d\alpha(s) \tag{8}$$

that should hold for all $r \in [0,1]$.[11] Now at the right hand one counts the total measure from 0 to r along both positive and negative parts of the boundary.

Suppose we need the size α to be uniformly distributed over $[0,1]$. Then we set

$$d\alpha(s) = \alpha ds.$$

[11] For gaussian processes and symmetric boundaries the first passage density is the same when evaluated at both upper and lower boundaries.

For our two gaussian processes the integral equations (7) or (8) belong to the class of nonlinear Volterra equations of the second kind with weak singularity of Abel type (e.g., Brunner and van der Houwen, 1986). Singularities occur when s is near r because $p_{r|s}\left(\Psi(r)|\Psi(s)\right) \to \infty$ as $s \to r$ from the left. Of course, there is no hope for an analytical solution, so we use numerical methods of integration and solving equations. More exactly, we, moving from $r = 0$ to $r = 1$, construct a piecewise linear boundary, at each step determining the slope of a current linear segment by using the bisection method in equating the left and right sides of (7) or (8), each time computing integrals at the right-side of (7) or (8) using the trapezoidal rule and analytically derived asymptotic solutions near singularity points. The details of the algorithm are described in Appendix B. The trapezoidal rule is utilized for the following reasons. First, the rate of convergence is close to higher order approximations requiring much more complicated programming. Second, it is the best method given that our functions are at most twice continuously differentiable. Third, Diogo, Lima and Rebelo (2005) prove that the trapezoid method has the property of uniform convergence for this type of integral equations.

Some details about parameters of the numerical algorithm follow; Appendix B contains full details. There are 500 knots in the piecewise linear boundary that approximates the smooth one. These knots are not uniformly distributed on $[0, 1]$, but rather the closer to zero, the more dense they are. In the numerical integration, approximately 10,000 gridpoints are uniformly distributed over the domain of integration.

3.2 "Uniform" boundaries for Wiener process

When the asymptotic process $Q(r)$ is Wiener process $Q(r) = W(r)$, we know that for $s < r$,

$$\begin{aligned} Q(r) &\sim N(0, r), \\ Q(r)|Q(s) &\sim N(Q(s), r - s). \end{aligned}$$

Therefore, the densities entering (7) and (8) are

$$\begin{aligned} p_r(y) &= \frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{y^2}{2r}\right), \\ p_{r|s}(y|x) &= \frac{1}{\sqrt{2\pi(r-s)}} \exp\left(-\frac{(y-x)^2}{2(r-s)}\right). \end{aligned}$$

It turns out that we only need to derive two (one for one-sided testing, another for two-sided testing) what we call *baseline boundaries* corresponding to the maximum needed test

size A. It turns out that both retrospective and monitoring boundaries corresponding to any size $\alpha \leq A$ can be easily obtained from the baseline boundaries by using an appropriate transformation. Suppose we have derived a baseline boundary $\Psi(r)$ over $[0,1]$ corresponding to the maximum needed test size A. However, one is interested in a retrospective boundary $b_\alpha^R(r)$ corresponding to the test size $\alpha \leq A$, or a monitoring boundary $b_\alpha^M(r)$ over the period $[1,K]$ corresponding to the test size $\alpha < A$. The transformations

$$b_\alpha^R(r) = \sqrt{\zeta_\alpha}\Psi\left(\zeta_\alpha^{-1} r\right) \tag{9}$$

and

$$b_\alpha^M(r) = \sqrt{\xi_{\alpha,K}}\Psi\left(\xi_{\alpha,K}^{-1}(r-1)\right), \tag{10}$$

where

$$\zeta_\alpha \equiv \frac{A}{\alpha}, \quad \xi_{\alpha,K} \equiv \frac{A}{\alpha}(K-1), \tag{11}$$

accomplish this job, which can be easily seen from the integral equation (7) or (8). Intuitively, as the size is distributed uniformly over $[0,1]$, the baseline boundary $\Psi(r)$ accumulates exactly α over the segment $[0,\zeta_\alpha^{-1}]$. This portion of $\Psi(r)$ should be extended over the retrospective or monitoring interval and then scaled to restore the target size.

To report the baseline boundary so that it can be used in practice, we fit to it a flexible, but parsimonious parametric function, and report the parameters of this function. The parameterization has to fit well both for r distant from zero and for r close to zero, as the behavior of the boundary near zero is critical for ensuring the correct crossing intensity rate. Therefore, we use a product of different functions of r, each function being designed to capture one of these sharply distinct dependencies. The function, call it $b_{(0,1]}(r)$, covering dependence on the bulk of $(0,1]$ is, after taking a log, a polynomial function of r. The function, call it $b_{\downarrow 0}(r)$, covering dependence in the vicinity of zero is derived from the asymptoric behavior of integral equations (7) or (8) in the neighborhood of $r = 0$. To summarize, the baseline boundary is parameterized by the following functional form:

$$\Psi(r) = b_{\downarrow 0}(r)\, b_{(0,1]}(r), \tag{12}$$

where

$$b_{(0,1]}(r) = \exp\left(\sum_{j=1}^{J} \psi_j r^j\right),$$

and the form of $b_{\downarrow 0}(r)$ can be found in the Online Appendix, section 1.[12] This form turns out to be sufficiently flexible even for low values of J, and quite convenient, as, after taking logs, $\ln \Psi(r) - \ln b_{\downarrow 0}(r)$ is a linear form in powers, up to J^{th}, in r, which allows us to easily estimate the coefficients ψ_j, $j = 1, ..., J$, by ordinary least squares; see the Online Appendix for details. The coefficients for the choice $J = 3$ and $A = 10\%$ are tabulated below.

	One-sided	Two-sided
ψ_1	0.18795	0.10834
ψ_2	−0.19150	−0.10585
ψ_3	0.21882	0.06987

As mentioned in the Introduction, the degree of fit turns out to be very high even with $J = 3$: the regression (in logs) R^2 is about 99.99%, and the computed and parameterized boundaries are practically indistinguishable, both visually and in terms of maximal discrepancy.

According to the transformations (9)–(10), an arbitrary target boundary $b_\alpha(r)$ corresponding to target size α may be obtained as

$$b_\alpha^R(r) = b_{\downarrow 0}(dr) \exp\left(\sum_{j=0}^{J} c_j r^j\right)$$

in the retrospective case and

$$b_\alpha^M(r) = b_{\downarrow 0}(d(r-1)) \exp\left(\sum_{j=0}^{J} c_j (r-1)^j\right)$$

in the monitoring case, where the coefficients c_j, $j = 0, ..., J$ and d are functions of ψ_j, $j = 1, ..., J$ and ζ_α or $\xi_{\alpha,K}$ defined in (11). In particular, in the case $J = 3$, we have the following correspondences:

$$c_0 = \frac{1}{2}\ln \zeta_\alpha, \quad c_1 = \zeta_\alpha^{-1}\psi_1, \quad c_2 = \zeta_\alpha^{-2}\psi_2, \quad c_3 = \zeta_\alpha^{-3}\psi_3, \quad d = \zeta_\alpha^{-1}$$

in the retrospective case, and

$$c_0 = \frac{1}{2}\ln \xi_{\alpha,K}, \quad c_1 = \xi_{\alpha,K}^{-1}\psi_1, \quad c_2 = \xi_{\alpha,K}^{-2}\psi_2, \quad c_3 = \xi_{\alpha,K}^{-3}\psi_3, \quad d = \xi_{\alpha,K}^{-1}$$

in the monitoring case.

Figures 3a and 3b depict the retrospective and monitoring, respectively, "uniform" boundaries corresponding to the three conventional levels of size.

[12] The function $b_{\downarrow 0}(r)$ solves a certain first order differential equation derived from the asymptotic behavior of the integral equation (7) or (8) as $r \downarrow 0$.

3.3 "Uniform" boundaries for Brownian Bridge

In the case of Brownian bridge, in contrast to the case of Wiener process, the retrospective and monitoring boundaries have to be handled separately. This is due to the property of Brownian bridge of being tied down at $r = 1$. When the asymptotic process $Q(r)$ is the Brownian bridge process $Q(r) = B(r) = W(r) - rW(1)$, we can derive that when $s < r \leq 1$,

$$\begin{aligned} Q(r) &\sim N(0, r(1-r)), \\ Q(r)|Q(s) &\sim N\left(\frac{1-r}{1-s}Q(s), \frac{1-r}{1-s}(r-s)\right). \end{aligned}$$

Therefore, the densities entering (7) and (8) are

$$\begin{aligned} p_r(y) &= \frac{1}{\sqrt{2\pi r(1-r)}} \exp\left(-\frac{y^2}{2r(1-r)}\right), \\ p_{r|s}(y|x) &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{(1-s)}{(r-s)(1-r)}} \exp\left(-\frac{((1-s)y-(1-r)x)^2}{2(r-s)(1-r)(1-s)}\right). \end{aligned}$$

It turns out that now, in contrast to the case of Wiener process, we cannot obtain a retrospective boundary corresponding to some value of α from a boundary corresponding to a different value of α. In other words, there does not exist a baseline boundary that would be able to generate a whole family of size-specific boundaries. This is, of course, due to the property of the Brownian bridge of being tied down at $r = 1$.

Hence, to report the family of boundaries in the case of retrospection, we fit the whole family to a parametric function not only of r, but also of α. For fixed α, we use a functional form similar to (12), which takes into account tiedness to zero both at $r = 0$ and at $r = 1$:

$$b_\alpha^R(r,\alpha) = b_{\downarrow 0}(r,\alpha)\, b_{(0,1)}(r,\alpha)\, b_{\uparrow 1}(r,\alpha), \tag{13}$$

where

$$b_{(0,1)}(r,\alpha) = \exp\left(\sum_{j=0}^{J} \psi_j(\alpha)\, r^j\right), \tag{14}$$

and the functions $b_{\downarrow 0}(r,\alpha)$ and $b_{\uparrow 1}(r,\alpha)$ are described in the Online Appendix, section 2.[13] The coefficients $\psi_j(\alpha)$ in (14) are parameterized as cubic functions of α:

$$\psi_j(\alpha) = \psi_j^{(0)} + \psi_j^{(1)}\alpha + \psi_j^{(2)}\alpha^2 + \psi_j^{(3)}\alpha^3.$$

[13] The functions $b_{\downarrow 0}(r,\alpha)$ and $b_{\uparrow 1}(r,\alpha)$ solve certain first order and second order differential equations, respectively, derived from the asymptotic behavior of the integral equation (7) or (8) as $r \downarrow 0$ or as $r \uparrow 1$.

The coefficients for the choice $J = 3$ are given below. The size α is unitless, i.e., for example, $\alpha = 0.05$.[14]

	One-sided				Two-sided			
i	0	1	2	3	0	1	2	3
$\psi_0^{(i)}$	−1.1113	13.511	−113.26	536.9	−1.1857	10.706	−94.85	403.2
$\psi_1^{(i)}$	−0.0258	−2.398	23.67	−316.9	−0.0262	−1.023	4.62	−36.1
$\psi_2^{(i)}$	−0.0417	−2.419	−12.47	303.7	−0.0254	−2.306	13.83	−41.3
$\psi_3^{(i)}$	0.0340	2.692	2.14	−89.4	0.0244	2.081	−11.10	43.3

For a particular value of α, a researcher may find the parameterization of $b_\alpha^R(r)$ using the tabulated coefficients, and use this $b_\alpha^R(r)$ as a retrospective boundary on $[0, 1]$.

When $r > s \geq 1$,

$$\begin{aligned} Q(r) &\sim N(0, r(r-1)), \\ Q(r)|Q(s) &\sim N\left(\frac{r}{s}Q(s), \frac{r}{s}(r-s)\right). \end{aligned}$$

Therefore, the densities entering (7) and (8) are

$$\begin{aligned} p_r(y) &= \frac{1}{\sqrt{2\pi r(r-1)}} \exp\left(-\frac{y^2}{2r(r-1)}\right), \\ p_{r|s}(y|x) &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{s}{(r-s)r}} \exp\left(-\frac{(sy-rx)^2}{2(r-s)rs}\right). \end{aligned}$$

As could be expected, we cannot obtain a monitoring boundary from a retrospective one. This is again due to the property of the Brownian bridge of being tied down at $r = 1$.

Compared to the case of retrospection, here we have, along with α, an additional parameter K, that determines the boundary. Fortunately, a particular boundary can be characterized by a single combination of α and K, namely the "crossing intensity"

$$\gamma = \frac{\alpha}{K-1}.$$

This property can be easily confirmed by analyzing the integral equation (7) or (8).

Thus, we can fit the whole family of monitoring boundaries to a parametric function of r and γ. For fixed γ, we use a familiar functional form, which takes into account tiedness to zero at $r = 1$:

$$b_\alpha^M(r) = b_{\downarrow 1}(r, \gamma)\, b_{(1,K]}(r, \gamma). \tag{15}$$

[14]Because each coefficient is multiplied by α to a certain power, we reduce the number of reported digits after the decimal point as the power increases.

where

$$b_{(1,K]}(r,\gamma) = \exp\left(\sum_{j=0}^{J} \psi_j(\gamma)(r-1)^j\right), \tag{16}$$

and the function $b_{\downarrow 1}(r,\gamma)$ are described in the Online Appendix, section 2.[15] The coefficients $\psi_j(\gamma)$ in (16) are parameterized as functions of γ in the same way:

$$\psi_j(\gamma) = \psi_j^{(0)} + \psi_j^{(1)}\gamma + \psi_j^{(2)}\gamma^2 + \psi_j^{(3)}\gamma^3.$$

The coefficients for the choice $J = 3$ are given below. The size γ is unitless, i.e., for example, $\gamma = 0.05/(5-1) = 0.0125$.[16]

	One-sided				Two-sided			
i	0	1	2	3	0	1	2	3
$\psi_0^{(i)}$	0.0682	−2.017	18.58	−51.1	0.0680	−2.079	19.60	−55.8
$\psi_1^{(i)}$	0.2338	3.308	9.08	−153.5	0.2375	3.348	5.60	−141.7
$\psi_2^{(i)}$	−0.0229	−0.809	−38.97	129.1	−0.0234	−0.778	−36.62	132.5
$\psi_3^{(i)}$	0.0010	0.032	3.93	113.2	0.0010	0.031	3.79	97.7

Having particular values of α and K, a researcher can compute γ, find the parameterization of $b_\alpha^M(r)$ using the tabulated coefficients, and use this $b_\alpha^M(r)$ as a monitoring boundary on $[1,K]$.

Figures 3c and 3d depict the retrospective and monitoring, respectively, "uniform" boundaries corresponding to the three conventional levels of size. All boundaries except the Brownian bridge retrospective boundaries, as expected, start off from zero with an infinite derivative, and are increasing throughout entire intervals. The Brownian bridge retrospective boundaries (see Figure 3c) have an inverted U-shape and come to zero at the end of the retrospective interval, also with an infinite derivative. Their shape is similar to that of the Zeileis boundaries (see Figure 2b), but they are asymmetric (for example, the maximum is reached at 0.48 rather than at 0.50 when $\alpha = 5\%$); they are also steeper at the beginning and end of the interval, in the sense that their ratio diverges to $+\infty$ as $r \downarrow 0$ or $r \uparrow 1$.

The relative positioning of boundaries of different type can be observed in Figures 5a–5d, see Section 4.

[15] The function $b_{\downarrow 1}(r,\gamma)$ solves a first order differential equation similar to that for $b_{\downarrow 0}(r,\alpha)$, derived from the asymptotic behavior of the integral equation (7) or (8) as $r \downarrow 1$.

[16] Again, because each coefficient is multiplied by γ to a certain power, we reduce the number of reported digits after the decimal point as the power increases.

3.4 Distribution of size with "uniform" boundaries

Figures 4a and 4b present the results of asymptotic simulations in a retrospective context and one-testing with the Wiener process and the Brownian bridge. It is clear that with the constructed boundaries the size is indeed distributed uniformly across the interval $[0, 1]$.

All imperfections in these distributions are due to insufficient accuracy of approximations during numerically solving an integral equation, or to insufficient flexibility of a parameterization, or to insufficient number of simulation repetitions. All three sources can be potentially driven to nullity if desired, although at a non-negligible expense.

3.5 Detection of instability

While the primary concern of this article is the distribution of size, it is still interesting what kind of power properties are exibited by tests with the same total size but different distributions thereof. We perform a simple simulation experiment with an abrupt structural break that occurs at various moments during the monitoring interval, namely, at T (i.e. at the very beginning), at $\frac{1}{4}T$, at $\frac{1}{2}T$ (i.e. halfway), and at $\frac{3}{4}T$. The simulated model contains only a zero constant which then shifts to take value c, where c is either $\frac{1}{5}\sigma$ or σ, and σ is the error standard deviation which is normalized to unity. We set $T = 1000$ and $K = 2$ (i.e. monitoring is meant to continue for 1000 time periods). We explore the classical CUSUM test with uniform and parabolic boundaries and the OLS-based CUSUM with uniform, parabolic and linear boundaries (see Table 1 and subsection 2.2). The overall size for the tests is set at the 5% level. Table 2 reports hitting rates (we count only those hits that occur after the moment of the break) and average delays (i.e., the average number of time periods passed from the period when the break occurs to the period after the break in which the instability was detected).

When the structural shift is large, it is detected with a high probability irrespective of which boundary is used, even when the break is late in the monitoring interval. When the shift is small, the hitting rate gets smaller, and the later the break date is the smaller is the hitting rate. The parabolic boundaries suffer most, presumably because they are constructed for an infinite monitoring horizon. The CUSUM test equipped with the "uniform" boundaries seems to be most powerful among the five.

The average delay figures reflect shapes of corresponding boundaries. If the structural

moment of break	CUSUM		OLS-based CUSUM		
	uniform	parabolic	uniform	parabolic	linear
mean shift is equal to $\frac{1}{5}$ of error standard deviation					
T	0.99 (278)	0.84 (597)	0.99 (276)	0.97 (278)	0.99 (341)
$\frac{1}{4}T$	0.91 (364)	0.58 (530)	0.86 (389)	0.70 (417)	0.89 (384)
$\frac{1}{2}T$	0.61 (314)	0.20 (386)	0.46 (321)	0.26 (335)	0.53 (313)
$\frac{3}{4}T$	0.15 (171)	0.02 (191)	0.10 (164)	0.03 (173)	0.13 (162)
mean shift is equal to error standard deviation					
T	1.00 (16)	1.00 (85)	1.00 (16)	1.00 (13)	1.00 (53)
$\frac{1}{4}T$	0.99 (55)	1.00 (97)	0.99 (61)	0.98 (63)	1.00 (67)
$\frac{1}{2}T$	0.99 (68)	1.00 (105)	0.98 (81)	0.99 (91)	0.99 (79)
$\frac{3}{4}T$	0.96 (77)	1.00 (114)	0.96 (97)	0.97 (117)	0.97 (92)

Note: for each test and boundary, documented are two figures: hitting rate and (average delay).

Table 2: Hitting rates and average delays.

shift occurs early in the monitoring interval, it can be faster captured by the tests with boundaries that start off from zero – both tests coupled with the "uniform" boundaries, and the OLS-based CUSUM test coupled with the parabolic boundaries. If the break occurs late in the monitoring period, the linear boundary for the OLS-based CUSUM test catches up, while the CUSUM parabolic boundary is much harder to reach, again because they are constructed for an infinite monitoring horizon. The "uniform" boundaries (especially those designed for the Wiener process) tend to provide smaller and more robust average delays overall.

4 Empirical illustration

In this Section we illustrate sequential testing tools using an empirical application. We perform a few testing experiments, both retrospective and monitoring, using different detectors, different boundaries and different testing intervals. The purpose of this exercise is to show interesting patterns that one may encounter in practice rather than to contrast the merits of different boundaries or detectors.

We use the Phillips curve model analogous to one of applications in Bai and Perron (2003). However, we use monthly US data instead of annual UK data,[17] for the sake of

[17]See also an application to monthly Japanese data in Kurozumi (2017).

larger sample sizes.[18] The Phillips curve equation we estimate is

$$E\left[\Delta w_t | \pi_{t-1}, \pi_{t-2}, ..., u_t, u_{t-1}, ...\right] = \gamma_1 + \gamma_2 \pi_{t-1} + \gamma_3 \pi_{t-2} + \gamma_4 u_t + \gamma_5 u_{t-1} + \gamma_6 u_{t-2},$$

where w_t is nominal log wage, π_t is inflation (difference of log nominal price index), u_t is unemployment. Here, in contrast to Bai and Perron (2003), we allow one more lag for inflation and unemployment because of the higher data frequency.

We carry out four experiments, two retrospective and two monitoring, in each case carrying out two-sided testing at the 5% significance level. We use two detectors: the CUSUM and OLS-based CUSUM. Recall that the CUSUM detector asymptotically behaves as a Wiener process, and the OLS-based CUSUM detector – as a Brownian bridge. The results are presented in Figures 5a through 5d which show the (absolute values of) detectors (in ragged bold) and various boundaries (the "uniform" boundaries are in solid bold).

We start from retrospective testing with a historical interval spread from 1965:02 till 1970:11 (70 observations) which is presumably stable: the Andrews (1993) stability test with the truncation parameter $\pi = 0.20$ does not reject stability even at the 10% significance. Figure 5a attests that according to sequential testing there is no evidence of structural instability either. Both detectors uniformly lie below all corresponding boundaries: the "uniform" (solid), horizontal (short dashes), Zeileis (long dashes) and, in the case of CUSUM, linear (dots and dashes).

Next we carry out two monitoring experiments, one with a shorter horizon, one with a longer horizon. Figure 5b shows the results for $K = 2$ (so that monitoring starts from 1970:12 and a researcher commits to continue it till 1976:09), and Figure 5c – for $K = 5$ (so that monitoring starts from 1970:12 and continues till 1984:03). In the case of CUSUM monitoring, we check the "uniform" (solid) and "parabolic" (short dashes) boundaries; in the case of OLS-based CUSUM, we check the "uniform" (solid), "parabolic" (short dashes) and linear (long dashes) boundaries.

In the case of the shorter monitoring interval, the CUSUM detector passes through the "uniform" boundary in 14 time periods (i.e. on 1972:01). Note that the instability is not detected by the parabolic boundary at all, the main reason being that it is relying on the infinite horizon (which is too different from the actual horizon) and thus starts off too high

[18]The (seasonally adjusted) data are taken from the FRED database of the Federal Reserve Board at `fred.stlouisfed.org` (series AHETPI, CPIAUCNS, UNRATE).

(in other words, only a small portion of 5% size is "utilized" when $K = 2$). The OLS-based CUSUM detector touches on both the "uniform" and parabolic boundaries even faster, in 6 time periods (i.e. on 1971:05), and the linear boundary more than twice as late, in 13 time periods (i.e. on 1971:12).

In the case of the longer monitoring interval, the CUSUM detector again hits the "uniform" boundary for the first time in 14 time periods (i.e. on 1972:01). Now the instability is detected by the parabolic boundary too, but very much later, after 190 time periods pass (i.e. only on 1986:09). Thus, the parabolic CUSUM boundaries, even when they detect an instability, may make an impression of a very late break when in fact it is very early (cf. subsection 3.5). The OLS-based CUSUM detector touches on the parabolic boundaries first, in 6 time periods (i.e. on 1971:05), the "uniform" boundary in 9 time periods (i.e. on 1971:08), and the linear boundary in 14 time periods (i.e. on 1972:01).

Finally, we repeat a retrospection experiment on a longer historical interval, which now presumably includes structural instability evidenced by the previous monitoring tests. We set the end of the historical interval now to 1984:11 (so that the interval contains 237 observations). The results are presented on Figure 5d. Interestingly, no boundary detects instabilities when the CUSUM detector is used. In the case of the OLS-based detector, however, the structural instability is sensed by the "uniform" (solid) boundary in the 208th time period (i.e. on 1982:06) and by the Zeileis (long dashes) boundary a bit later, in the 217th time period (i.e. on 1983:03). Both time periods are quite late compared to when the structural instability must in fact have taken place, but note an important fact that the horizontal boundary (short dashes), which is implicit in the most popular sup functional, is very far at all from detecting this instability.

5 Concluding remarks

We have numerically derived boundaries for major classes of sequential (CUSUM-type) tests, both retrospective and monitoring, both one-sided and two-sided, such that the overall test size is uniformly distributed over the testing (historical or monitoring) interval. We have reported these boundaries as tables of coefficients of fitted parsimonious but flexible parametric forms. We have also provided asymptotic simulation evidence that these (parametric) boundaries do an excellent job in distributing test size uniformly.

The two major classes of sequential tests considered are based on asymptotical behavior of the Wiener process or Brownian bridge. Nevertheless, in the literature one can encounter, although more rarely, other asymptotic processes for sequential detectors, which are usually functions of the two processes above; see Table 1. Some of these (having the transitional density in a convenient analytical form) can be handled similarly to the technique we have proposed. MOSUM-type tests are more problematic because of the presence of an additional parameter. The problems just described may constitute an agenda for future work.

References

[1] Anatolyev, S. (2008) Nonparametric retrospection and monitoring of predictability of financial returns. *Journal of Business & Economic Statistics* 27, 149–160.

[2] Anatolyev, S. and G. Kosenok (2011) Another numerical method of finding critical values for the Andrews stability test. *Econometric Theory* 28, 239–246.

[3] Andreou, E. and E. Ghysels (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600.

[4] Andreou, E. and E. Ghysels (2006) Monitoring disruptions in financial markets. *Journal of Econometrics* 135, 77–124.

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

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

[7] Bai, J. and P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

[8] Bai, J. and P. Perron (2003) Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18, 1–22.

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

[10] Brunner, H. and P.J. van der Houwen (1986) The numerical solution of Volterra equations. CWI Monographs, North–Holland, Amsterdam.

[11] Chu, C.S.J., K. Hornik, and C.M. Kuan (1995) The moving-estimates test for parameter stability. *Econometric Theory* 11, 669–720.

[12] Chu, C.S.J., M. Stinchcombe, and H. White (1996) Monitoring structural change. *Econometrica* 64, 1045–1065.

[13] Deng, A. and P. Perron (2008) The limit distribution of the Cusum of squares test under general mixing conditions. *Econometric Theory* 24, 809–822.

[14] Diogo, N., P. Lima and M. Rebelo (2005) Computational methods for a nonlinear Volterra integral equation. *Proceedings of HERCMA* 2005, 100–107.

[15] Durbin, J. (1971) Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov–Smirnov test. *Journal of Applied Probability* 8, 431–453.

[16] Ghysels, E., A. Guay and A. Hall (1997) Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209–233.

[17] Groen, J., G. Kapetanios and S. Price (2013) Multivariate methods for monitoring structural change. *Journal of Applied Econometrics* 28, 250–274.

[18] Inclán, C. and G.C. Tiao (1994) Use of cumulatrive sums of squares for retrospective detection of changes in variance. *Journal of American Statistical Association* 89, 913–923.

[19] Inoue, A. and B. Rossi (2005) Recursive predictability tests for real time data. *Journal of Business and Economic Statistics* 23, 336–345.

[20] Karatzas, I. and S.E. Shreve (1988) *Brownian motion and stochastic calculus*, New York: Springer-Verlag.

[21] Kurozumi, E. (2017) Monitoring parameter constancy with endogenous regressors. *Journal of Time Series Analysis* 38, 791–805.

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

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

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

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

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

[27] Romano, J.P., A.M. Shaikh and M. Wolf (2010) Multiple testing. In: *The New Palgrave Dictionary of Economics*. (eds. S.N. Durlauf and L.E. Blume). Palgrave Macmillan.

[28] Zeileis A (2004) Alternative boundaries for CUSUM tests. *Statistical Papers* 45, 123–131.

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

A Appendix: proofs

Lemma 1 *Parts (ii) and (iii) of Theorems 1 and 2 follow from part (i).*

Proof Observe that

$$\overline{\lim_{r\downarrow 0}}\frac{B(r)}{\ell(r)}=\overline{\lim_{r\downarrow 0}}\frac{W(r)-rW(1)}{\ell(r)}=\overline{\lim_{r\downarrow 0}}\frac{W(r)}{\ell(r)}-W(1)\lim_{r\downarrow 0}\sqrt{\frac{r}{2\ln\left(-\ln r\right)}}=\overline{\lim_{r\downarrow 0}}\frac{W(r)}{\ell(r)},$$

therefore the same LIL holds for the Brownian bridge at zero. Also, observe that

$$\overline{\lim_{r\uparrow 1}}\frac{B(r)}{\ell(1-r)}=\overline{\lim_{s\downarrow 0}}\frac{B(1-s)}{\ell(s)}=\overline{\lim_{s\downarrow 0}}\frac{B(s)}{\ell(s)},$$

where the change of variable $s=1-r$ and equality in distribution of $B(r)$ and $B(1-r)$ are employed. Therefore the same LIL holds for the Brownian bridge at unity. Thus, all steps of the proof of part (i) can be carried over to the case (ii).

Next, observe that, because for $r>1$

$$B(r)\overset{d}{=}(r-1)\,W\left(1+\frac{1}{r-1}\right),$$

we have

$$\overline{\lim_{r\downarrow 1}}\frac{B(r)}{\ell(r-1)}=\overline{\lim_{s\to+\infty}}\frac{s^{-1}W\left(1+s\right)}{\ell(s^{-1})}=\overline{\lim_{s\to+\infty}}\frac{W\left(s\right)}{\sqrt{2s\ln\left(\ln s\right)}},$$

where the change of variable $s=1/(r-1)$ is employed. This is unity by the "infinite-time" LIL (Karatzas and Shreve, 1988, theorem 9.23iii). Thus, all steps of the proof of part (i) can be carried over to the case (iii). □

Proof of Theorem 1 (i) For sufficiently small $\varepsilon>0$ there is $0<\bar{r}<1$ such that $b(r)>\ell(r)+\varepsilon$ for all $r\in[0,\bar{r}]$, and by the LIL (2) the equality

$$\Pr\left\{W\left(r\right)/\ell(r)>1+\varepsilon\text{ for some }r\in[0,\bar{r}]\right\}=0$$

holds. Then

$$\Pr\left\{W\left(r\right)>b(r)\text{ for some }r\in[0,\bar{r}]\right\}\leq\Pr\left\{W\left(r\right)/\ell(r)>1+\varepsilon\text{ for some }r\in[0,\bar{r}]\right\}=0,$$

i.e. $W\left(r\right)$ never hits $b(r)$ on $[0,\bar{r}]$ almost surely, which makes uniform distribution of finite size impossible. For parts (ii) and (iii), apply Lemma 1. □

Proof of Theorem 2 (b) From (3) and positivity of $b(r)$ and $\ell(r)$ it follows that for sufficiently small $\varepsilon > 0$ there exists $\bar{r}$ such that $b(r)/\ell(r) < 1 - \varepsilon$ for all $r \in (0, \bar{r}]$, and by the LIT (2),

$$\Pr\{W(r)/\ell(r) \geq 1 - \varepsilon \text{ for some } r \in (0, \bar{r}]\} = 1.$$

These two statements imply that

$$\Pr\{W(r) \geq b(r) \text{ for some } r \in (0, \bar{r}]\} = 1,$$

i.e. $W(r)$ hits $b(r)$ on $(0, \bar{r}]$ almost surely, which makes uniform distribution of finite size impossible. For parts (ii) and (iii), apply Lemma 1. □

B Appendix: Numerical solution of integral equation

The integral equation has the following generic form:

$$\varphi(r, b(r)) = \alpha \int_0^r f(s, r, b(\cdot))\, ds, \tag{1}$$

where s and r are such that $0 \leq s \leq r \leq K$. We are looking for a numerical solution for the boundary $b(r)$ on the interval $[0, K]$. We set a nonuniform grid with $I + 1$ gridpoints that gets finer as one approached the zero boundary value. Denote this grid by $\{r_i\}_{i=0}^I$, where $r_0 = 0$ and $r_I = K$. The approximate solution is a piecewise linear function on the grid with the corresponding values $\{b_i\}_{i=0}^I$, where $b_0 = 0$, and the other values are evaluated sequentially as we describe next.[19]

For r_1, the integral equation has the form

$$\varphi(r_1, b_1) = \alpha \int_0^{r_1} f(s, r_1, b(\cdot))\, ds. \tag{2}$$

In order to find the value of b_1, we create a uniform subgrid $\{s_j\}_{j=0}^M$ on the interval $[0, r_1]$, where $s_0 = 0$ and $s_M = r_1$.[20] We replace the function $b(s)$ inside the integral sign by a linear function on the interval $[r_0, r_1]$ of the form

$$b(s) = \frac{r_1 - s}{r_1 - r_0} b_0 + \frac{s - r_0}{r_1 - r_0} b_1, \ s \in [r_0, r_1].$$

[19] Our algorithm is based on the method described in Diogo et. al. (2005) who prove a uniform convergence to the solution as the number of grid points grows to infinity.

[20] In computations, we set $M = 100$.

Based on this function we calculate the values of $f\{s, r_1, b(\cdot)\}$ at the subgrid points $\{s_j\}_{j=0}^{j=M-1}$. Denote these values by $\{f_j\}_{j=0}^{j=M-1}$. Based on them we can evaluate the integral in (2) by the trapezoidal rule. However, we do not yet have a value of f_M because at r_1 there is a singularity of Abel type: $f(s, r_1, b(\cdot)) \approx D/\sqrt{r_1 - s} \overset{s \to r_1}{\to} \infty$. Still, we can introduce such 'fake' value of f_M that the trapezoidal rule yields a correct approximate value of the integral on the subinterval $[s_{M-1}, s_M = r_1]$. Approximately, the corresponding integral is

$$\int_{s_{M-1}}^{s_M} f(s, r_1, b(\cdot))\, ds \approx \int_{s_{M-1}}^{s_M} \frac{D}{\sqrt{s_M - s}} ds = 2D\sqrt{s_M - s_{M-1}} \approx 2(s_M - s_{M-1}) f_{M-1},$$

where the last approximate equality follows from the approximate equality $f_{M-1} \approx D/\sqrt{s_M - s_{M-1}}$. In the case of a direct application of the trapezoidal rule we would use the formula

$$\int_{s_{M-1}}^{s_M} f(s, r_1, b(\cdot))\, ds \approx (s_M - s_{M-1}) \frac{f_{M-1} + f_M}{2}.$$

Setting $f_M = 3f_{M-1}$ forces these formulas to coincide. To summarize, by applying the trapezoidal integration rule to the sequence $\{f_j\}_{j=0}^{M}$ with $f_M = 3f_{M-1}$ we can approximate the integral on the interval $[0, r_1]$ as

$$\int_0^{r_1} f(s, r_1, b(\cdot))\, ds \approx \Delta s \left(\sum_{j=1}^{M-1} f_j + \frac{f_0 + f_M}{2} \right)$$

where $\Delta s = s_M - s_{M-1}$ is the size of a subgrid.

At last, by varying the value of b_1 we can reach the following equality:

$$\varphi(r_1, b_1) = \alpha \Delta s \left(\sum_{j=1}^{M-1} f_j + \frac{f_0 + f_M}{2} \right),$$

which is going to be our approximate solution for b_1. We applied bisection method to find the solution of the above equality.

After having evaluated b_1 we switch to the interval $[r_1, r_2]$, and in a similar way we find the value b_2, although note that this requires evaluation of the integral over two intervals: $[r_0, r_1]$ and $[r_1, r_2]$. Then, after having evaluated b_2 we find b_3, and so on until we evaluate the last value of the boundary b_I.

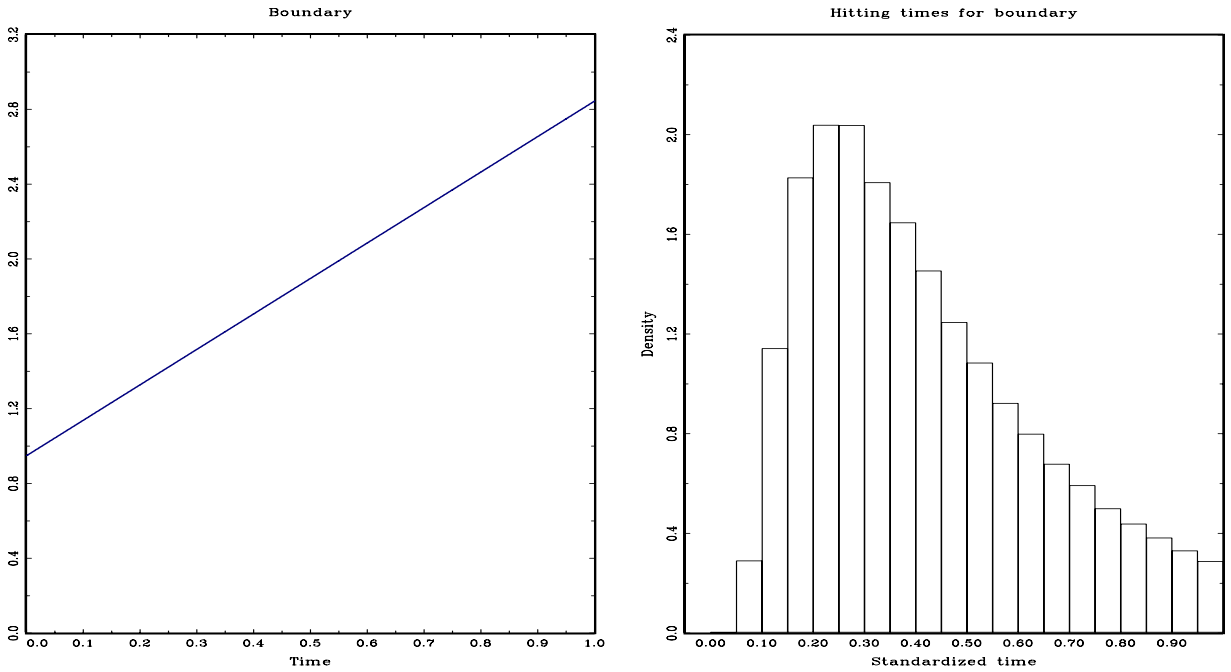

Figure 1a. Two-sided linear boundary for $W(r)$ on $[0,1]$ with distribution of size $\alpha = 5\%$

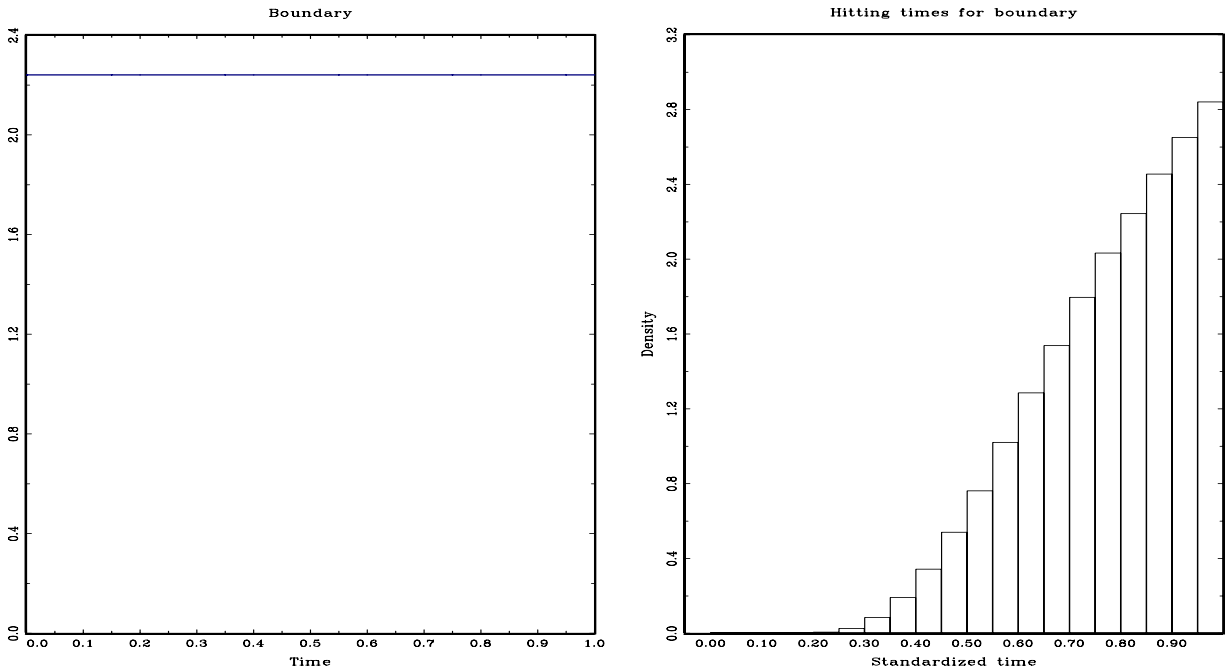

Figure 1b. Two-sided horizontal boundary for $W(r)$ on $[0,1]$ with distribution of size $\alpha = 5\%$

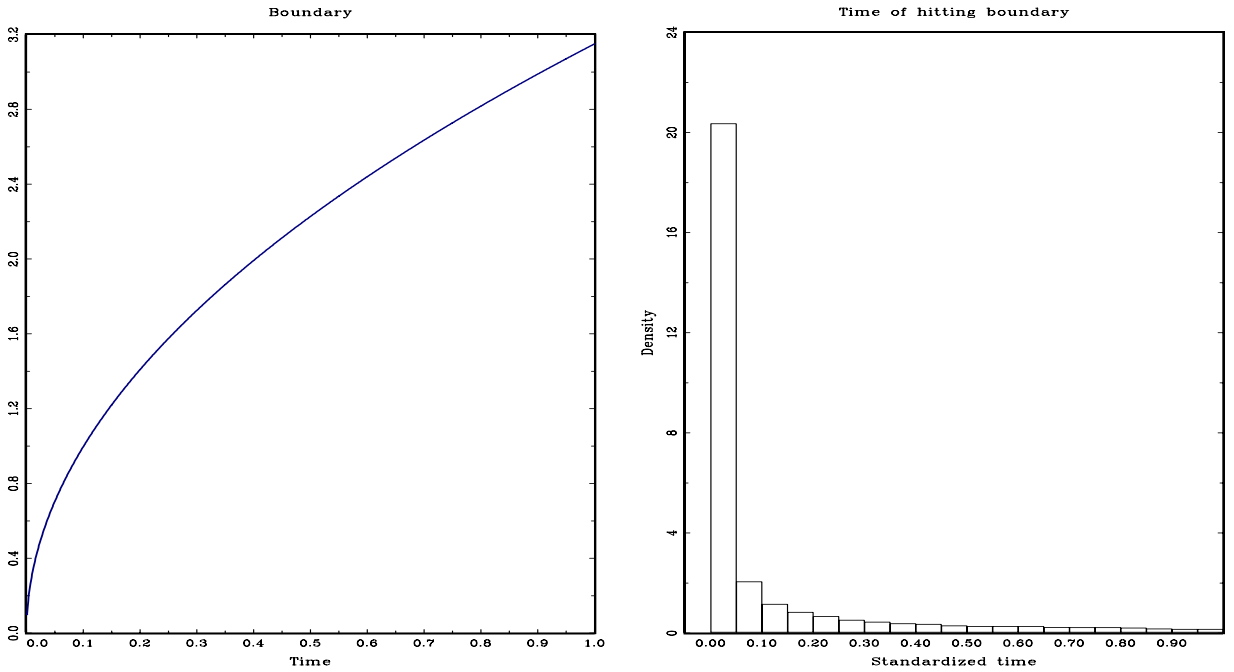

Figure 1c. Two-sided Zeileis boundary for $W(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

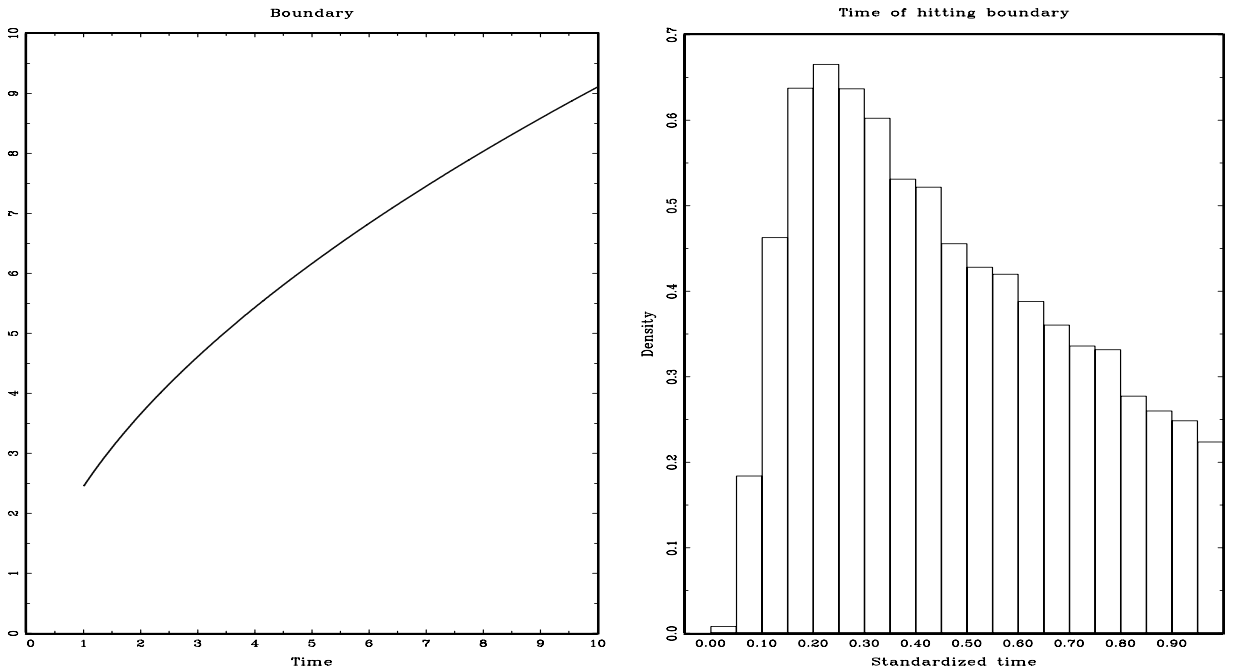

Figure 1d. Two-sided parabolic boundary for $W(r-1)$ on $[1, \infty)$ with distribution of size $\alpha = 5\%$

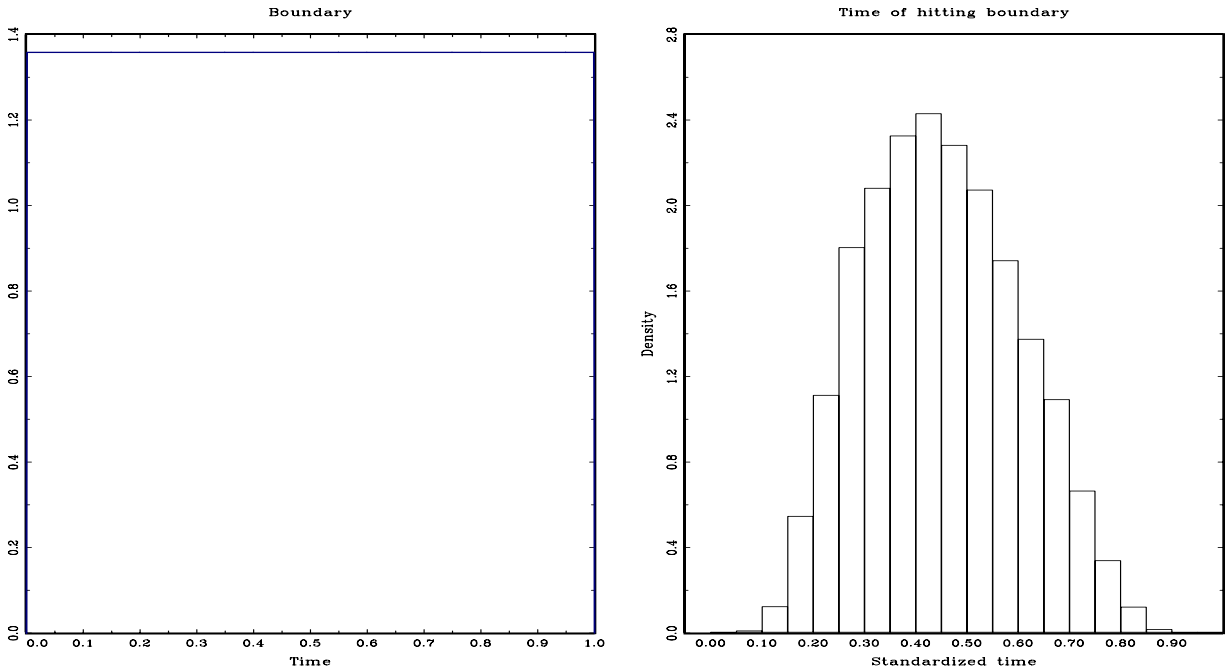

Figure 2a. Two-sided horizontal boundary for $B(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

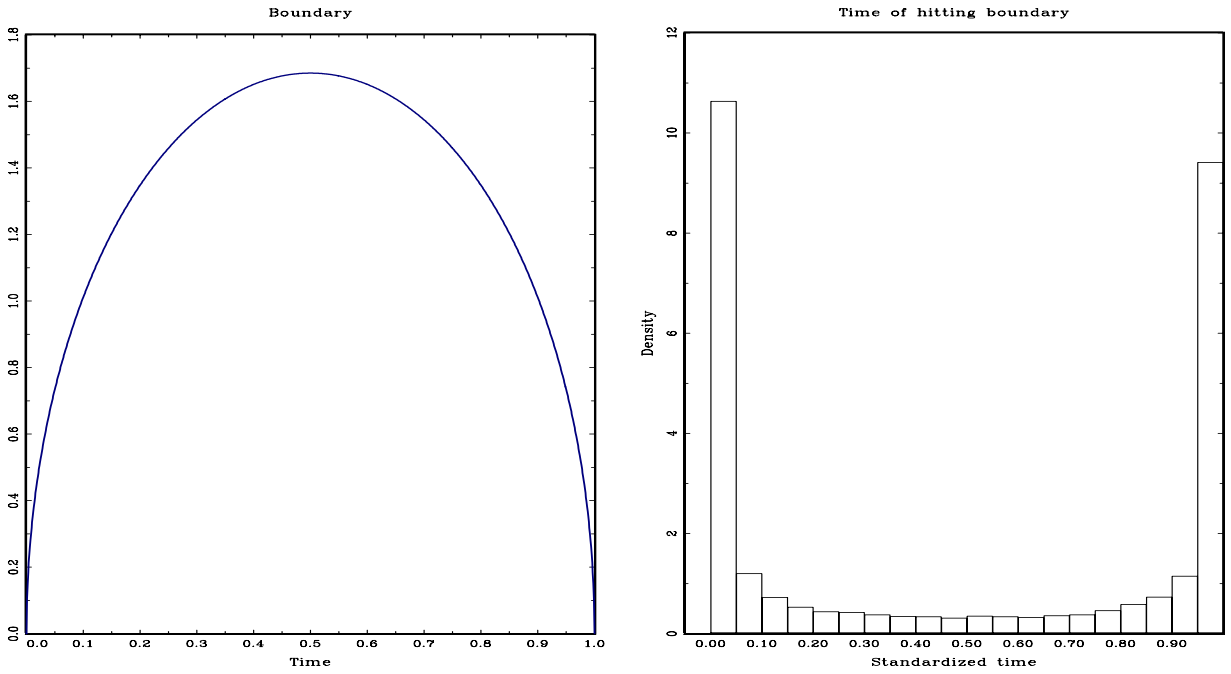

Figure 2b. Two-sided Zeileis boundary for $B(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

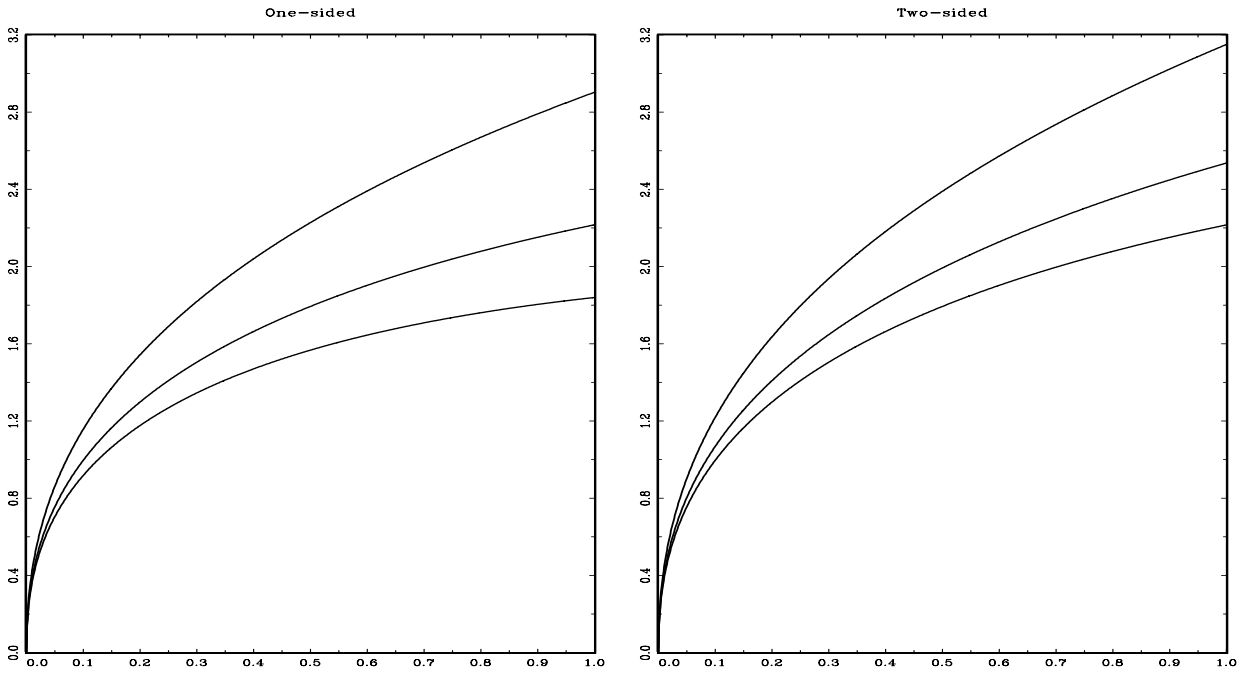

Figure 3a. Shapes of boundaries for $W(r)$ on $[0,1]$ with uniform distribution of sizes 1%, 5% and 10%

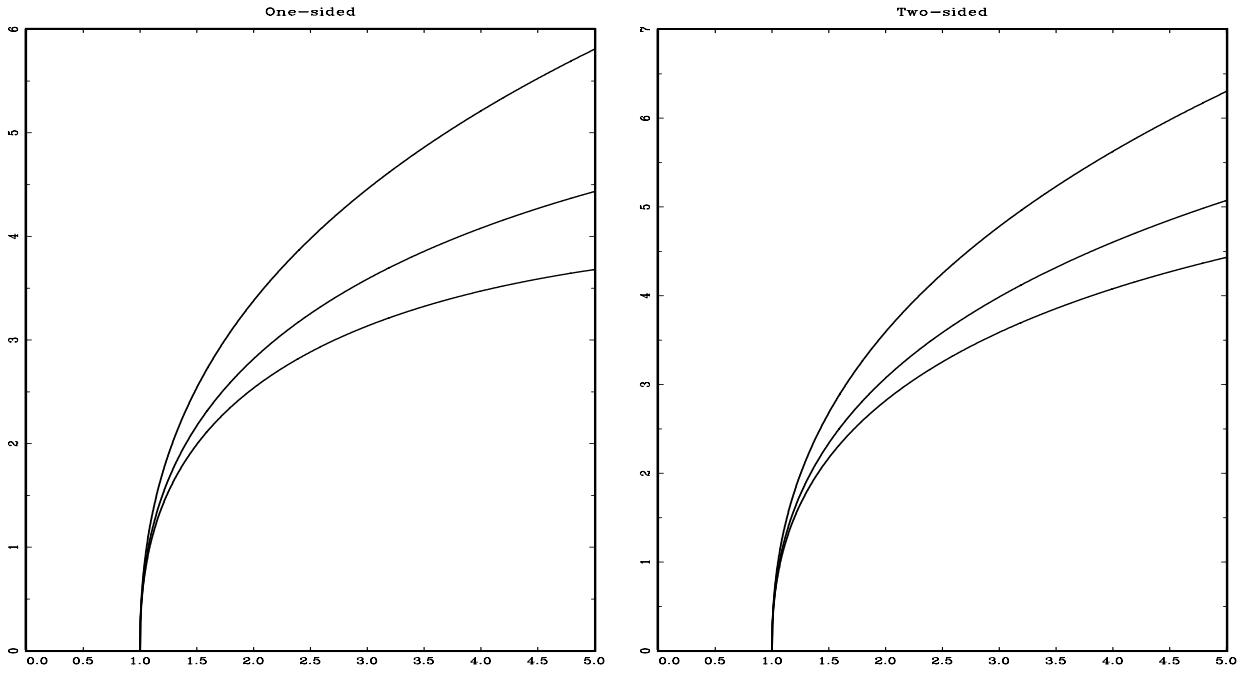

Figure 3b. Shapes of boundaries for $W(r-1)$ on $[1,5]$ with uniform distribution of sizes 1%, 5% and 10%

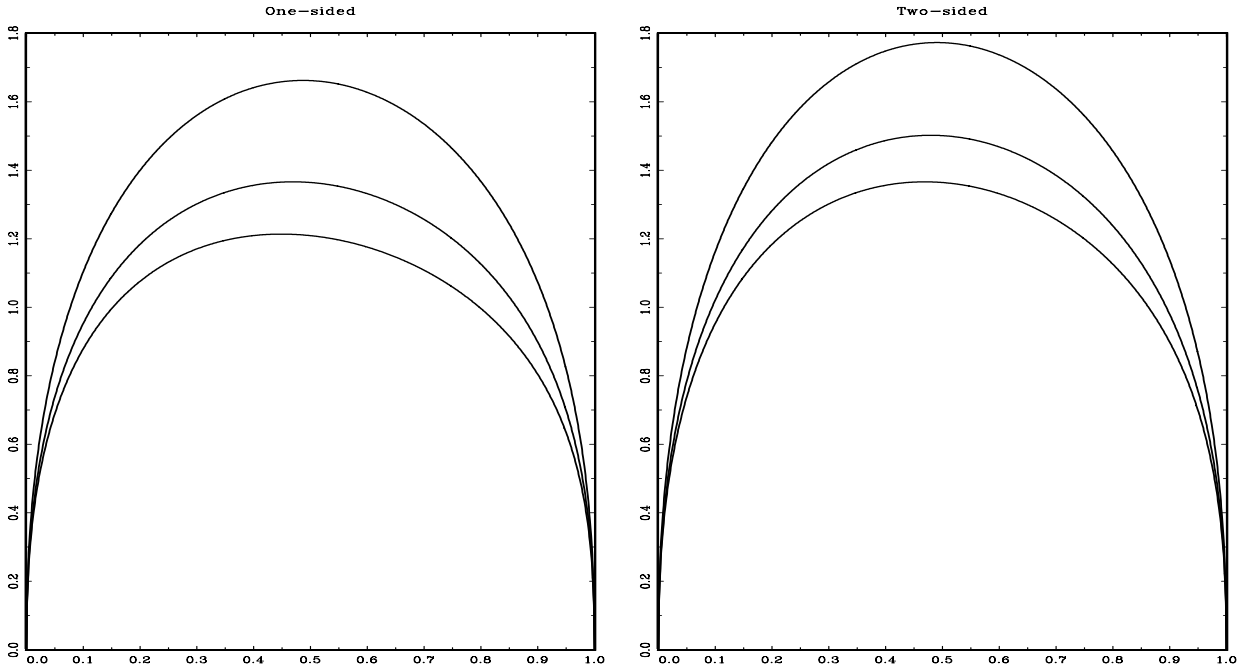

Figure 3c. Shapes of boundaries for $B(r)$ on $[0,1]$ with uniform distribution of sizes 1%, 5% and 10%

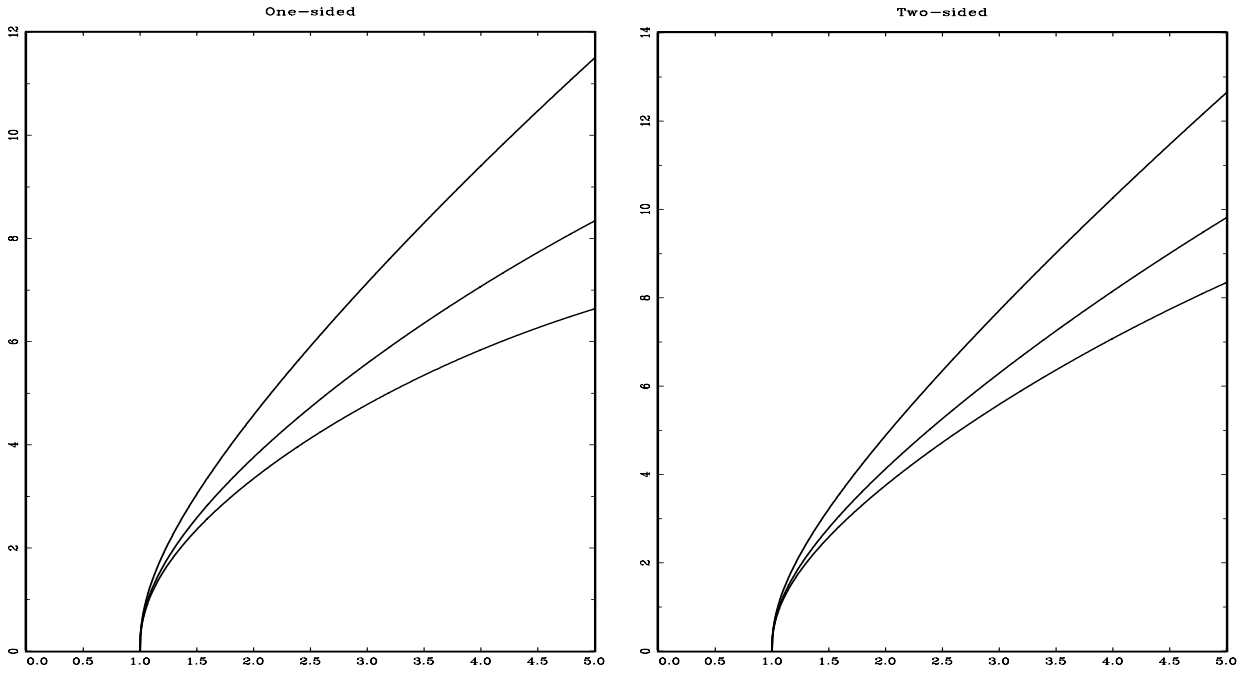

Figure 3d. Shapes of boundaries for $B(r)$ on $[1,5]$ with uniform distribution of sizes 1%, 5% and 10%

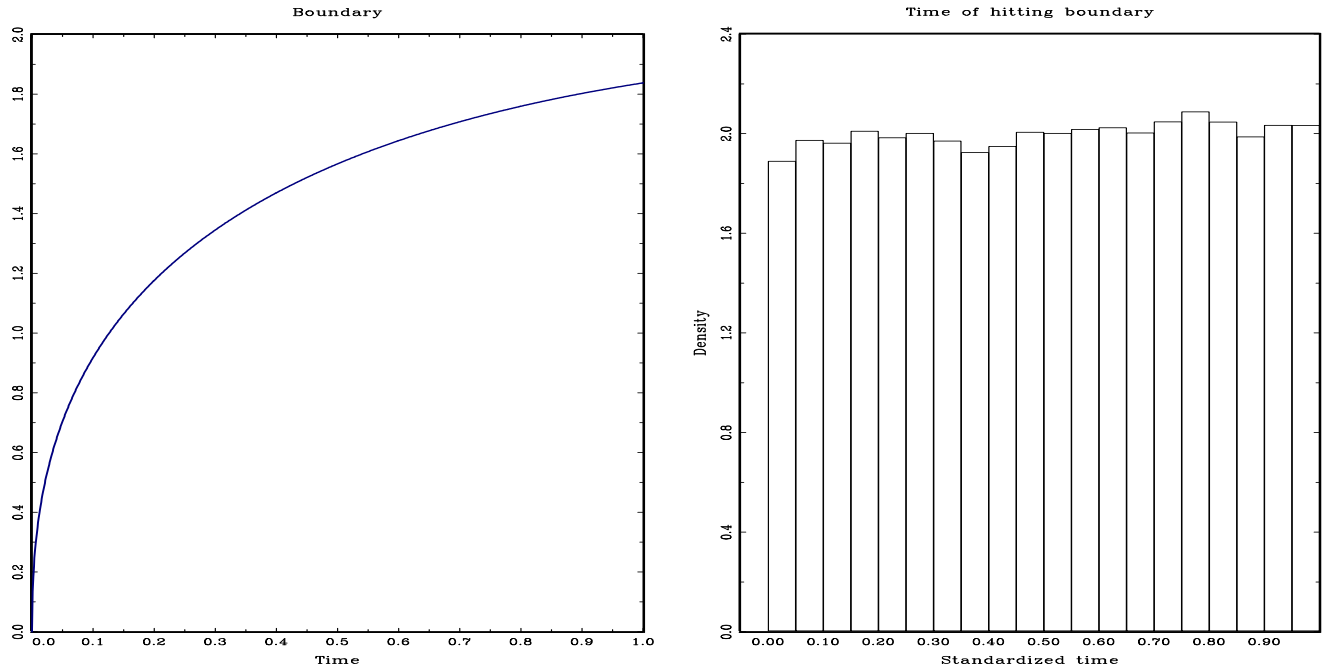

Figure 4a. One-sided boundary for $W(r)$ on $[0,1]$ with uniform distribution of size $\alpha = 5\%$

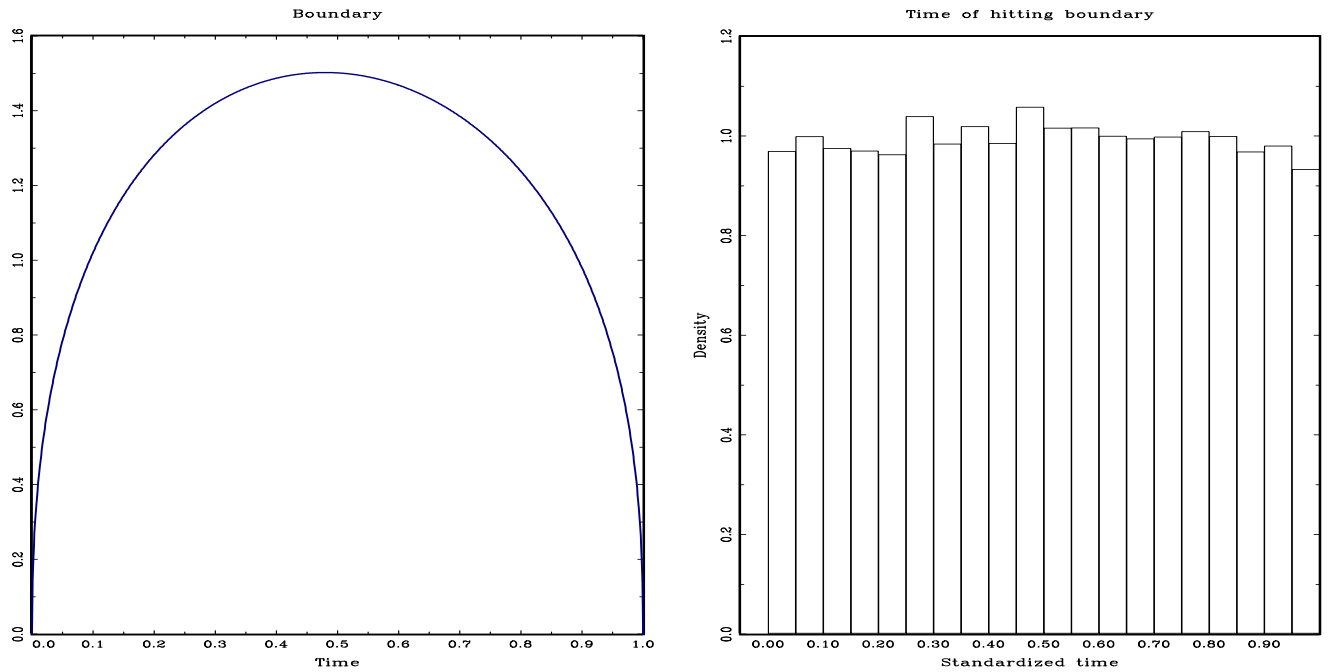

Figure 4b. One-sided boundary for $B(r)$ on $[0,1]$ with uniform distribution of size $\alpha = 5\%$

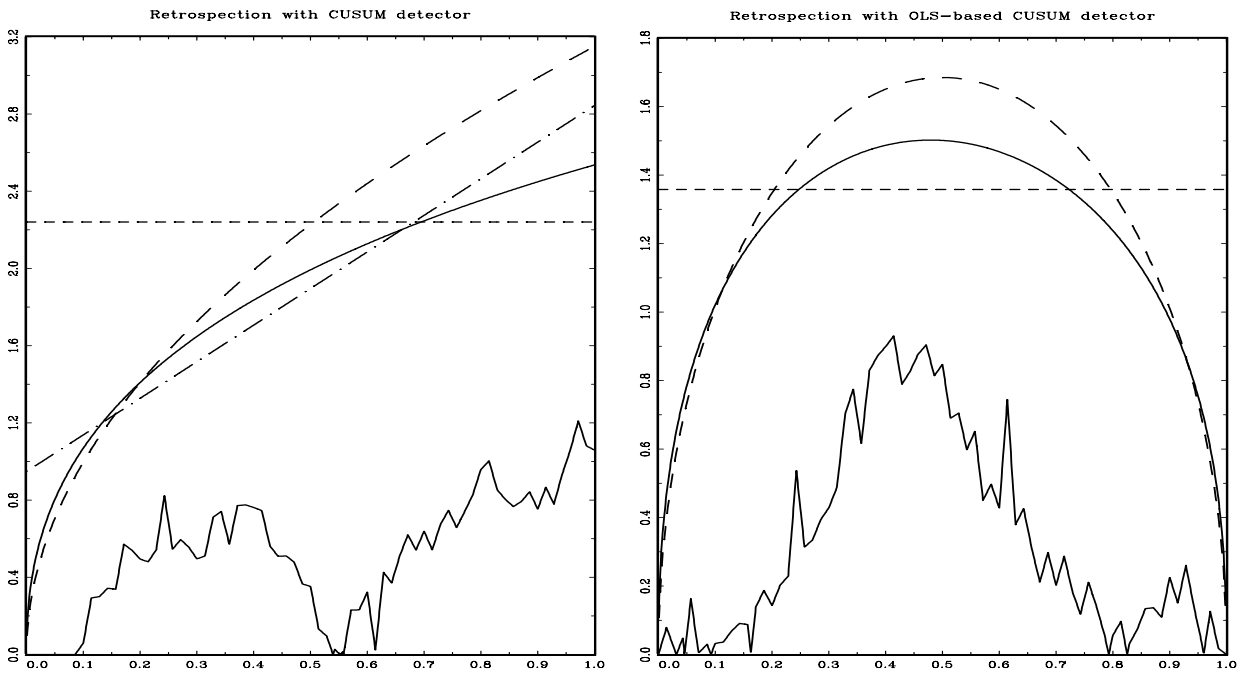

Figure 5a. CUSUM and OLS-based CUSUM retrospection with various boundaries in empirical application, shorter historical interval, size 5%

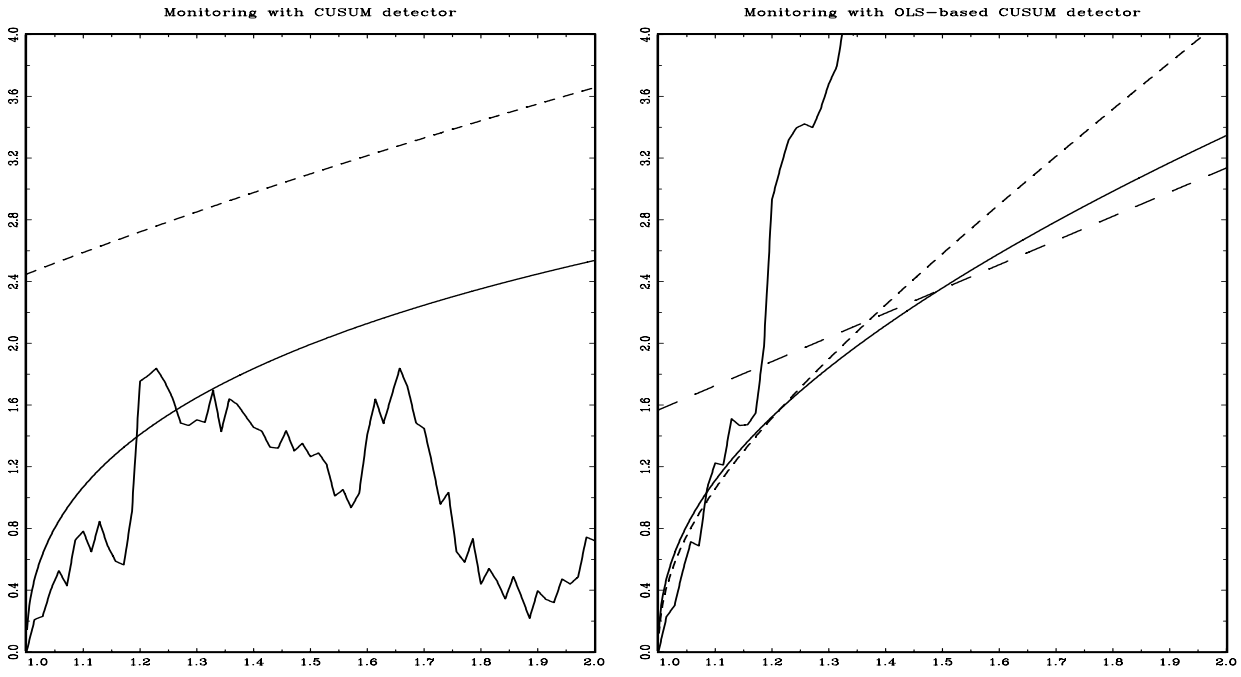

Figure 5b. CUSUM and OLS-based CUSUM monitoring with various boundaries in empirical application, horizon 2, size 5%

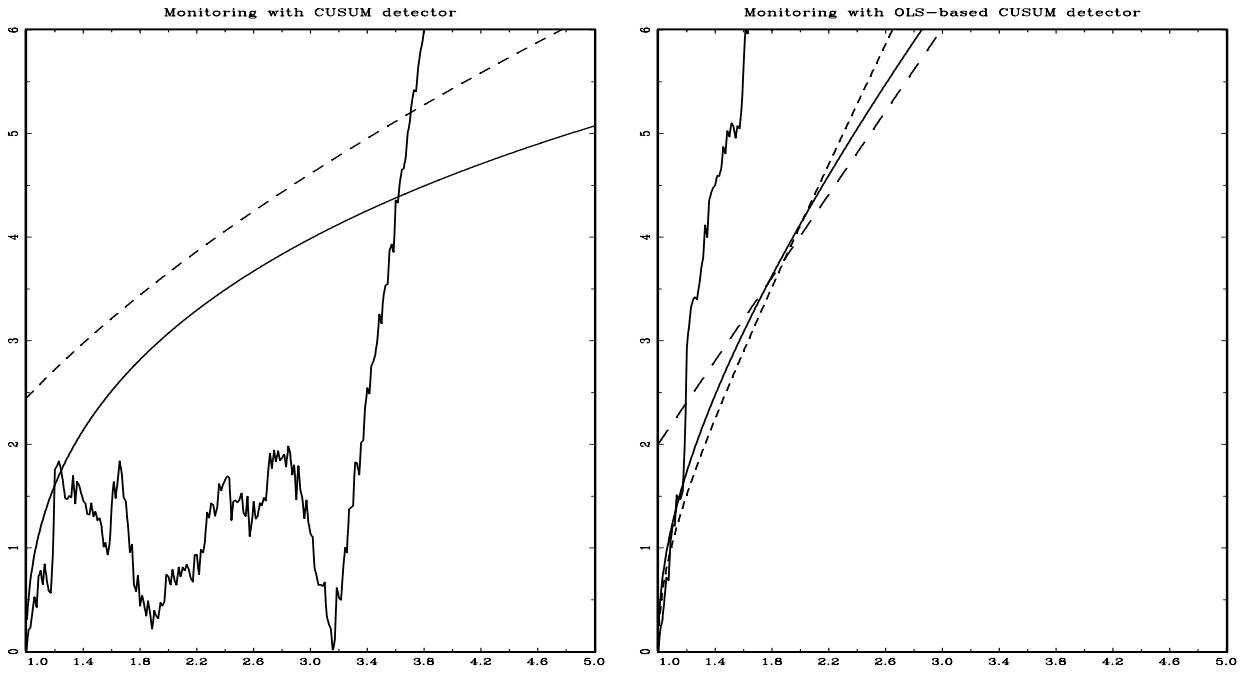

Figure 5c. CUSUM and OLS-based CUSUM monitoring with various boundaries in empirical application, horizon 5, size 5%

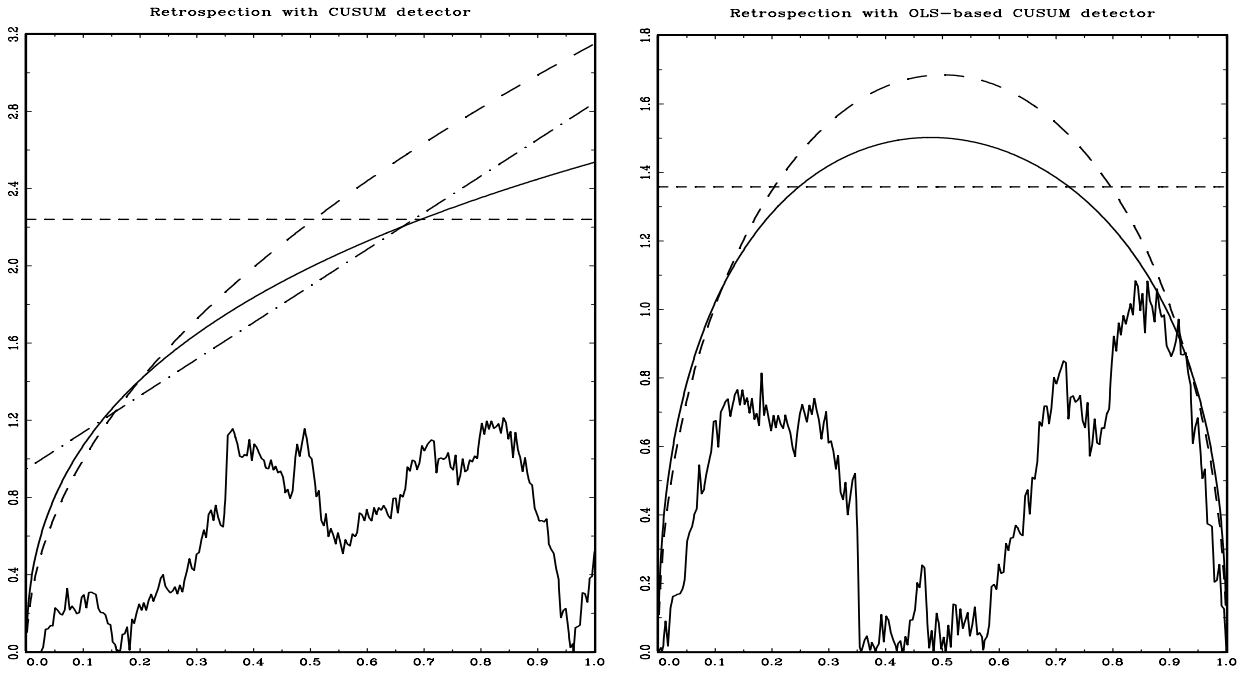

Figure 5d. CUSUM and OLS-based CUSUM retrospection with various boundaries in empirical application, longer historical interval, size 5%