Centre for Economic and Financial Research at New Economic School

April 2011

Sequential Testing with Uniformly Distributed Size

Stanislav Anatolyev
Grigory Kosenok

Working Paper No 123

CEFIR / NES Working Paper series

Sequential testing with uniformly distributed size

Stanislav Anatolyev* New Economic School

Grigory Kosenok New Economic School

April 2011

Abstract

Sequential procedures of 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 a shape of boundaries which requires that the test size be uniformly distributed over the testing period. Under this criterion, we numerically construct boundaries for 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 a context of retrospecting a historical sample and in a context of monitoring newly arriving data. We tabulate the boundaries by fitting them to certain flexible but 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, New Economic School, Nakhimovsky Pr., 47, office 1721(3), Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

From mid-seventies, in applied econometric and statistical work one could encounter applications of sequential testing tools. Sequential testing methods are usually used in the context of testing for 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 an introductory level.

Consider a linear regression framework of testing for structural stability. Let us be interested in 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 null 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 formulating a specific alternative hypothesis that assumes a particular type of non-stability of coefficients, and proceeds by constructing a test designed specifically for this alternative. Such test is expected to also have power against other alternatives than the one it is designed for. Typically, this specific alternative assumes one or more abrupt changes in coefficients at specific dates, and then a standard Wald test may be used. This leads to a usual decision rule when a scalar test statistic is compared to a critical value. A radically different approach, called *sequential* or *recursive testing*, is, while avoiding to specify 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 the sequential test is 'do not reject' if the entire trajectory stays below the boundary, and 'reject' if it crosses the boundary at least once. An example is the celebrated CUSUM test of Brown, Durbin, and Evans (1975); other examples will be given shortly. Some of existing tests are in fact a mixture of the two extreme approaches. For example, the test for a single structural break of Andrews (1994) and its extensions to multiple breaks of Bai and Perron (1998) have a

structure of a classical test, but still can be interpreted as sequential tests with a particular sequential statistic and a particular boundary (see section 2.3).

It is worthwhile to mention the relationship of sequential testing to multiple testing, i.e. testing several hypotheses (see, e.g., Romano and Wolf, 2007). Even though there are certain similarities between the two setups, they have very little in common. Under sequential testing the null is the same each period, and the number of periods is large and asymptotically infinite. Under multiple testing the nulls are typically different and even may be quite heterogenous, and the number of these nulls is limited and asymptotically fixed. Among common properties is ambiguity of a size control criterion. Another shared feature is a need for adjustment, typically upward, of critical values for each one-shot test (see a forceful argument for such adjustment for sequential tests in Inoue and Rossi, 2005). Yet another common feature is a need to take into account the dependence among one-shot test statistics. This dependence is a nuisance feature under multiple testing and may often lead to imprecise size control (e.g., the Bonferroni method), while for sequential testing the dependence is an important building block which, in particular, motivates asymptotically continuous boundaries.

More formally, the sequential statistic Q_τ, also called a *detector*, of a sequential test is computed on various subintervals indexed by τ, typically, although not necessarily, 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, the researcher gets as a by-product an idea about 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 sequential testing tools are developed for this retrospective context, in particular, Ploberger, Krämer and Kontrus (1989), Ploberger and Krämer (1992), Inclán and Tiao (1994), and others. However, starting from Chu, Stinchcombe, and White (1996), researchers

got interested in implementing sequential testing in the *monitoring* context, using data arriving in real time, i.e. for $\tau = T+1, T+2, ...$, conditional on that H_0 holds for the historical interval. The monitoring literature also includes Leisch, Hornik, and Kuan (2000), Zeileis, Leisch, Kleiber and Hornik (2005), Andreou and Ghysels (2006), Inoue and Rossi (2005) and Anatolyev (2008). In what we do in this paper we handle both retrospection and monitoring situations, placing some more weight on the latter because it is more natural for sequential testing and because it poses more interesting challenges.

The critical issue in sequential testing is a choice of which boundary to use. Consider for simplicity 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. 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, because 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 may strongly affect outcomes of the test. 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 some boundaries, such as linear monitoring ones, tend to distribute the test size more evenly over time than others do, 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 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 well also have 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 in this paper.

It is true that a chosen shape of boundaries, motivated by a particular distribution of size, will affect the power of a test. This power will of course also depend on what type of non-stability is in effect. However, recall that under sequential testing no alternative is preferred to others; if there is preference ordering over alternatives, this should have been taken into account while constructing the test statistic in the first place. A researcher's being completely agnostic to an alternative is consistent with the uniform distribution of size.

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 to be further complicated by the presence of singularities. To attain the 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 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: 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 asymptotically the detector is 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 asymptotically the detector is a Brownian bridge is more complex. Here, in contrast, boundaries are specific for the value of size in the retrospection context, and additionally for the monitoring horizon in the monitoring context.

Because the boundaries are computed numerically, we provide a user with a tabulated version of the boundaries. We handle this by fitting the computed boundaries to a certain parametric functional form, very flexible though parsimonious. The degree of fit is very high: 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. We demonstrate via simulations that the parameterized boundaries do possess the property of distributing the size uniformly.

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. Section 3 describes the method of obtaining the boundaries with a particular distribution of test size. Sections 4 and 5 report some details and give results of constructing the "uniform" boundaries for the two classes of sequential tests. In Section 6 we report asymptotic simulation results on the distribution of size using our boundaries. In Section 7 we illustrate the properties of our and alternative procedures in an empirical application. Finally, Section 8 concludes. Some proofs are contained in the Appendix.

2 Sequential testing: details

2.1 Setup and asymptotics

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

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

in case testing is one-sided or

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

in case testing is two-sided,[1] 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.

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

$$Q_\tau \xrightarrow{d} Q(r)$$

on $\mathcal{R}$ and

$$b_\tau \to b(r),$$

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

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

[1] 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 gaussian (and therefore symmetric) process.

when testing is one-sided or

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

when testing is two-sided.

While $b\left(r\right)$ is arbitrary subject to the size control requirement, the asymptotic process $Q\left(r\right)$ depends on the detector used in testing (see the next subsection for examples of detectors). Typically, $Q\left(r\right)$ is one of the following two processes:

- Wiener process W,

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

- Brownian bridge B,

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

Remark 1. 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). We conjecture that it is possible to construct "uniform" boundaries in cases when the starting point is random, but such boundaries will have a strange shape starting off from infinity at the beginning of the monitoring period. The reason is that for any boundary starting off from a finite value there is a positive probability mass of the process concentrated above this boundary, which is incompatible with uniform size distribution over a continuum. Most importantly though, such setup would be inconsistent with the monitoring paradigm where it is assumed that the historical period is stable. Our choice thus precludes some versions of monitoring detectors that have been encountered in the literature (see footnote 2).

2.2 Detectors

Let us consider examples of detectors for sequential tests that can be encountered in the literature. We list their simple versions; many exist in several variations which do not differ

in asymptotic properties. Generally, a detector is a standardized (so that it is asymptotically pivotal) cumulative sum in an expanding window, possibly contrasted with a similar measure on the whole historical interval. Without contrasting, the asymptotic process is likely a Wiener process. When there is contrasting, explicit or implicit, the asymptotic process is likely a Brownian bridge.

The classical retrospective CUSUM detector (Brown, Durbin and Evans, 1975; Ploberger, Krämer and Alt, 1988) is

$$Q_\tau = \frac{1}{\hat{\sigma}\sqrt{T-k}} \sum_{t=k+1}^{\tau} \omega_t, \quad \tau = k+1, ..., T,$$

where $\hat{\sigma}^2$ is a consistent estimate of the variance of u_τ in the regression (1), ω_t are recursive residuals

$$\omega_t = \frac{y_t - x_t'(X_{1:t-1}'X_{1:t-1})^{-1}X_{1:t-1}'Y_{1:t-1}}{\sqrt{1 + x_t'(X_{1:t-1}'X_{1:t-1})^{-1}x_t}}$$

and data matrices $X_{1:t-1}$ and $Y_{1:t-1}$ contain observations from 1 to $t-1$. Asymptotically, as $T \to \infty$,

$$Q_\tau \xrightarrow{d} W(r), \ r \in [0,1].$$

The same asymptotics is shared by the sequential analog of a t-statistic in Anatolyev (2008), for example, in a problem of testing for predictability of $g(y_t)$ by $h(x_t)$, where g and h are known functions of stationary series y_t and x_t:

$$Q_\tau = \frac{1}{\sqrt{T\hat{V}_\tau}} \sum_{t=1}^{\tau} (g(y_t) - \bar{g}(y_{1:\tau}))h(x_t),$$

where $\bar{g}(y_{1:\tau})$ is a sample average of $g(y_t)$ from $t = 1$ to $t = \tau$, and $\hat{V}_\tau$ is an estimate of asymptotic variance computed in the same window, i.e. from $t = 1$ to $t = \tau$. Chu,

Stinchcombe and White (1996) extend the CUSUM detector to the monitoring situation:[2]

$$Q_\tau = \frac{1}{\hat{\sigma}\sqrt{T-k}} \sum_{t=T+1}^{\tau} \omega_t, \quad \tau = T+1, ..., KT.$$

Asymptotically, as $T \to \infty$,

$$Q_\tau \xrightarrow{d} W(r) - W(1) \sim W(r-1), \; r \in [1, K].$$

The modern version of the retrospective CUSUM of squares detector (Brown, Durbin and Evans, 1975; Deng and Perron, 2008) is

$$Q_\tau = \sqrt{\frac{1}{\hat{\varphi}T}} \left(\sum_{t=k+1}^{\tau} \omega_t^2 - \frac{\tau}{T} \sum_{t=k+1}^{T} \omega_t^2 \right), \quad \tau = k+1, ..., T,$$

where $\hat{\varphi}$ is a consistent estimate of the long run variance of u_τ^2. A similar structure is taken by the detector in the Kokoszka and Leipus (2000) test and by that of the Inclán and Tiao (1994) test for detection of changes in variance; see also Andreou and Ghysels (2002). For all these detectors, asymptotically

$$Q_\tau \xrightarrow{d} B(r), \; r \in [0, 1]$$

as $T \to \infty$. The same asymptotics holds for the OLS-based CUSUM detector proposed by Ploberger and Krämer (1992),

$$Q_\tau = \frac{1}{\hat{\sigma}\sqrt{T}} \left(\sum_{t=k+1}^{\tau} \hat{u}_t \right),$$

where $\hat{u}_t = y_t - x_t'(X_{1:T}'X_{1:T})^{-1}X_{1:T}'Y_{1:T}$ are OLS residuals, and for the fluctuation test detector proposed by Ploberger, Krämer and Kontrus (1989)

$$Q_\tau = \frac{\tau}{\hat{\sigma}T} \left[(X_{1:T}'X_{1:T})^{1/2} \left(\hat{\beta}_\tau - \hat{\beta}_T \right) \right]_i,$$

[2] Chu, Stinchcombe and White (1996) also suggest an alternative version of the monitoring CUSUM detector

$$Q_\tau = \frac{1}{\hat{\sigma}\sqrt{T-k}} \sum_{t=k+1}^{\tau} \omega_t$$

for $\tau = T+1, ..., \infty$. This version has one objective shortcoming: it is not consistent with the monitoring paradigm where it is assumed that the historical period is stable; this version of Q_τ instead accumulates instability-driven deviations during the historical period too. As a result, its asymptotic process starts off from a random value at $r = 1$ (see Remark 1).

where $[\circ]_i$ denotes taking the i^{th} element of a vector, and $\hat{\beta}_\tau$ is an OLS estimate of β computed from the observations $k+1, k+2, ..., \tau-1, \tau$. All these detectors can be extended to the monitoring context in a natural way, with

$$Q_\tau \xrightarrow{d} B(r), \ r \in [1, K]$$

as $T \to \infty$.

2.3 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 distribute the size relatively evenly, 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 nearly linear 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 truly 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 was 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 $b^R_\alpha(r)$ such that for all $s \in [0,1]$

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

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

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

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

Remark 2. Of course, uniform distribution over a monitoring period is possible only if the monitoring horizon K is finite. 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.[3]

Figures 1a, 1b, 1c and 1d present distributions of size[4] 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

[3]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.

[4]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 *It is impossible to distribute the size uniformly 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 *It is impossible to distribute the size uniformly 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 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 $Q(r)$.

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 $r \in [0,1]$.[5] 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

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

[5] 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 trapezoid method and analytically derived asymptotic solutions near singularity points. The trapezoid method 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 et al. (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. 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.

4 Boundaries for Wiener process

4.1 Construction of boundaries

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, 20% say. As it will be shown below, 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.

Now 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 $\alpha < A$. It turns out that 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).

For simplicity, we will demonstrate this property using the retrospective one-sided case. We need to show that $b_\alpha^R(r)$ in (9) is the solution of

$$\frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{b_\alpha^R(r)^2}{2r}\right) = \alpha \int_0^r \frac{1}{\sqrt{2\pi(r-s)}} \exp\left(-\frac{\left(b_\alpha^R(r) - b_\alpha^R(s)\right)^2}{2(r-s)}\right) ds.$$

Substituting (9) here, we obtain

$$\frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{\zeta_\alpha \Psi\left(\zeta_\alpha^{-1} r\right)^2}{2r}\right) =$$

$$= \alpha \int_0^r \frac{1}{\sqrt{2\pi(r-s)}} \exp\left(-\frac{\left(\sqrt{\zeta_\alpha}\Psi\left(\zeta_\alpha^{-1} r\right) - \sqrt{\zeta_\alpha}\Psi\left(\zeta_\alpha^{-1} s\right)\right)^2}{2(r-s)}\right) ds.$$

After changing the variable of integration $s' = \zeta_\alpha^{-1} s$, or $s = \zeta_\alpha s'$, we obtain

$$\frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{\zeta_\alpha \Psi\left(\zeta_\alpha^{-1} r\right)^2}{2r}\right) =$$

$$= \alpha\zeta_\alpha \int_0^{\zeta_\alpha^{-1} r} \frac{1}{\sqrt{2\pi\left(r - \zeta_\alpha s'\right)}} \exp\left(-\frac{(\sqrt{\zeta_\alpha}\Psi\left(\zeta_\alpha^{-1} r\right) - \sqrt{\zeta_\alpha}\Psi\left(s'\right))^2}{2\left(r - \zeta_\alpha s'\right)}\right) ds'.$$

Finally by multiplying both sides of equality by $\sqrt{\zeta_\alpha}$, rearranging terms and noticing that $\alpha\zeta_\alpha = A$ we get the equation

$$\frac{1}{\sqrt{2\pi(\zeta_\alpha^{-1} r)}} \exp\left(-\frac{\Psi\left(\zeta_\alpha^{-1} r\right)^2}{2(\zeta_\alpha^{-1} r)}\right) = \int_0^{\zeta_\alpha^{-1} r} \frac{1}{\sqrt{2\pi\left(\zeta_\alpha^{-1} r - s'\right)}} \exp\left(-\frac{\left(\Psi\left(\zeta_\alpha^{-1} r\right) - \Psi\left(s'\right)\right)^2}{2\left(\zeta_\alpha^{-1} r - s'\right)}\right) A ds'$$

that defines the baseline boundary $\Psi\left(r\right)$.

Intuitively, as the size is distributed uniformly over $[0, 1]$, the baseline boundary $\Psi(r)$ accumulates exactly α over the segment $\left[0, \zeta_\alpha^{-1}\right]$. This portion of $\Psi(r)$ should be extended over the retrospective or monitoring interval and then scaled to restore the target size.

4.2 Tabulation of boundaries

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. Concretely, the baseline boundary $\Psi(r)$ is parameterized by the following functional form:

$$\Psi(r) = \exp\left(\sum_{j=0}^{J} \psi_j r^j\right) \times r^{\sum_{j=0}^{J-1} \varphi_j (\ln r)^j}. \tag{12}$$

This form turns out to be sufficiently flexible even for low values of J, and quite convenient. The convenience comes from the fact that after taking logs, $\ln \Psi(r)$ is a linear form in powers, up to J^{th}, of r and $\ln r$, which allows us to easily estimate the coefficients in (12) by least squares. Note that we sacrifice conformity to Theorems 1 and 2 for the sake of better approximation on the likely working range of r in finite samples.

Up to cubic terms corresponding to the choice $J = 3$, the parameterization of baseline boundary $\Psi(r)$ is

$$\Psi(r) = \exp\left(\psi_0 + \psi_1 r + \psi_2 r^2 + \psi_3 r^3\right) \times r^{\varphi_0 + \varphi_1 \ln r + \varphi_2 (\ln r)^2}.$$

For $A = 20\%$, the "regression" was run using the computed boundaries on a uniform grid of 5,000 values (dictated by the precision of the piecewise approximation for the boundary) of r on $[0, 1]$. The coefficients are tabulated in the following table up to four significant digits:

	One-sided	Two-sided
ψ_0	0.6607	0.6628
ψ_1	−0.3370	−0.3430
ψ_2	0.03328	0.03936
ψ_3	−0.04116	−0.04986
φ_0	0.3271	0.3282
φ_1	−0.01176	−0.01159
φ_2	−0.0003522	−0.0003435

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

$$b_\alpha^R(r) = \exp\left(\sum_{j=0}^{J} c_j r^j\right) \times r^{\sum_{j=0}^{J-1} d_j (\ln r)^j}$$

in the retrospective case and

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

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

$$\begin{aligned} c_0 &= \frac{1}{2}\ln\zeta_\alpha + \psi_0, \quad c_1 = \zeta_\alpha^{-1}\psi_1, \quad c_2 = \zeta_\alpha^{-2}\psi_2, \quad c_3 = \zeta_\alpha^{-3}\psi_3, \\ d_0 &= \varphi_0 + \ln\left(\zeta_\alpha^{-1}\right)\varphi_1 + \left(\ln\left(\zeta_\alpha^{-1}\right)\right)^2\varphi_2, \quad d_1 = \varphi_1 + 2\ln\left(\zeta_\alpha^{-1}\right)\varphi_2, \quad d_2 = \varphi_2, \end{aligned}$$

in the retrospective case, and

$$\begin{aligned} c_0 &= \frac{1}{2}\ln\xi_{\alpha,K} + \psi_0, \quad c_1 = \xi_{\alpha,K}^{-1}\psi_1, \quad c_2 = \xi_{\alpha,K}^{-2}\psi_2, \quad c_3 = \xi_{\alpha,K}^{-3}\psi_3, \\ d_0 &= \varphi_0 + \ln\left(\xi_{\alpha,K}^{-1}\right)\varphi_1 + \left(\ln\left(\xi_{\alpha,K}^{-1}\right)\right)^2\varphi_2, \quad d_1 = \varphi_1 + 2\ln\left(\xi_{\alpha,K}^{-1}\right)\varphi_2, \quad d_2 = \varphi_2, \end{aligned}$$

in the monitoring case.

4.3 Shape of boundaries

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

5 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 $r = 1$.

5.1 Construction of retrospective boundaries

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 $r = 1$.

5.2 Tabulation of retrospective boundaries

Thus, 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 at $r = 1$:

$$b_\alpha^R(r) = \exp\left(\sum_{j=0}^{J} \psi_j(\alpha)\, r^j\right) \times r^{\sum_{j=0}^{J-1} \varphi_j(\alpha)(\ln r)^j} \times (1-r)^{\sum_{j=0}^{J-1} \phi_j(\alpha)(\ln(1-r))^j}. \qquad (13)$$

The coefficients $\psi_j(\alpha)$, $\varphi_j(\alpha)$ and $\phi_j(\alpha)$ in (13) are parameterized as functions of α in the following way:

$$\psi_j(\alpha) = \psi_j^{(0)} + \psi_j^{(1)}\alpha + \psi_j^{(2)}\alpha^2 + \psi_j^{(3)} \ln\alpha + \psi_j^{(4)} (\ln\alpha)^2,$$

and similarly for $\varphi_j(\alpha)$ and $\phi_j(\alpha)$. In total then, there are $5(3J+1)$ parameters, which equals 50 in case $J = 3$.

The coefficients for the $J = 3$ are given below. The size α is unitless, i.e., for example, $\alpha = 0.05$. The "regression" was run using the computed boundaries on a uniform grid of values of α from 0.1% to 20% with a step of 0.1%. The results are reliable only for this range of sizes. As before, the grid for r contains 5,000 values uniformly distributed over $[0, 1]$.

	One-sided				
i	0	1	2	3	4
$\psi_0^{(i)}$	0.4602	−0.5542	0.2309	−0.1748	−0.007571
$\psi_1^{(i)}$	−0.2816	−1.445	0.5633	−0.06012	−0.003685
$\psi_2^{(i)}$	0.05853	0.1270	−3.135	0.01125	0.0005935
$\psi_3^{(i)}$	−0.02170	0.1858	1.223	−0.005589	−0.0003766
$\varphi_0^{(i)}$	0.2932	−0.1606	0.0009169	−0.03151	−0.001708
$\varphi_1^{(i)}$	−0.01538	−0.01785	−0.007254	−0.002907	−0.0001697
$\varphi_2^{(i)}$	−0.0005173	−0.0007062	−0.0005508	−0.0001057	-6.375×10^{-6}
$\phi_0^{(i)}$	0.2251	−0.4767	−0.4754	−0.04716	−0.002717
$\phi_1^{(i)}$	−0.02241	−0.05068	−0.06861	−0.004532	−0.0002748
$\phi_2^{(i)}$	−0.0007729	−0.001904	−0.003227	−0.0001645	-1.019×10^{-5}

	Two-sided				
i	0	1	2	3	4
$\psi_0^{(i)}$	0.6181	−0.4409	0.4119	−0.1490	−0.006098
$\psi_1^{(i)}$	−0.2241	−0.7720	0.2311	−0.04802	−0.002986
$\psi_2^{(i)}$	0.06212	0.001011	−0.5197	0.01440	0.0009509
$\psi_3^{(i)}$	−0.02084	0.1260	0.1200	−0.005842	−0.0004257
$\varphi_0^{(i)}$	0.3261	−0.1292	0.1112	−0.02467	−0.001279
$\varphi_1^{(i)}$	−0.01207	−0.01468	0.01181	−0.002155	−0.0001203
$\varphi_2^{(i)}$	−0.0003905	−0.0005949	0.0004533	-7.543×10^{-5}	-4.338×10^{-6}
$\phi_0^{(i)}$	0.2758	−0.3071	0.09167	−0.03592	−0.001990
$\phi_1^{(i)}$	−0.01724	−0.03356	0.01098	−0.003314	−0.0001934
$\phi_2^{(i)}$	−0.0005772	−0.001302	0.0004428	−0.0001170	-6.937×10^{-6}

Having a particular values 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 $[0,1]$.

5.3 Construction of monitoring boundaries

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 $r = 1$.

5.4 Tabulation of monitoring boundaries

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

$$b_\alpha^M(r) = \exp\left(\sum_{j=0}^{J} \psi_j(\gamma)(r-1)^j\right) \times (r-1)^{\sum_{j=0}^{J-1} \varphi_j(\gamma)(\ln(r-1))^j}. \tag{14}$$

The coefficients $\psi_j(\gamma)$ and $\varphi_j(\gamma)$ in (14) are parameterized as functions of γ in the same way:

$$\psi_j(\gamma) = \psi_j^{(0)} + \psi_j^{(1)}\gamma + \psi_j^{(2)}\gamma^2 + \psi_j^{(3)}\ln\gamma + \psi_j^{(4)}(\ln\gamma)^2,$$

and similarly for $\varphi_j(\gamma)$.

The coefficients for the $J = 3$ are given below. The size γ is unitless, i.e., for example, $\gamma = 0.05/(5-1) = 0.0125$. The "regression" was run using the computed boundaries on a uniform grid of values of γ from 0.001 to 0.200 with a step of 0.001, with additional constraints that $\alpha \leq 20\%$ and $K \leq 11$. The results are reliable only for this range of sizes and monitoring horizons. As before, the grid for r contains 5,000 values uniformly distributed over $[0, 1]$.

	One-sided				
i	0	1	2	3	4
$\psi_0^{(i)}$	0.2806	−0.8330	1.086	−0.3391	−0.02238
$\psi_1^{(i)}$	0.2448	−0.1961	0.04355	0.1021	0.01075
$\psi_2^{(i)}$	0.008895	−0.4043	−3.3319	0.0008031	−0.0002008
$\psi_3^{(i)}$	−0.001507	0.02280	0.1640	−0.0003989	-2.091×10^{-5}
$\varphi_0^{(i)}$	0.06110	−0.04315	0.3132	−0.1797	−0.01488
$\varphi_1^{(i)}$	−0.09482	0.09669	−0.04245	−0.04541	−0.003909
$\varphi_2^{(i)}$	−0.007929	0.01370	−0.01170	−0.003717	−0.0003222

	Two-sided				
i	0	1	2	3	4
$\psi_0^{(i)}$	0.4769	−0.8774	1.6011	−0.3012	−0.01988
$\psi_1^{(i)}$	0.2641	0.4160	−0.9827	0.1044	0.01072
$\psi_2^{(i)}$	0.008394	−0.4297	−1.295	0.0007877	−0.0001942
$\psi_3^{(i)}$	−0.001179	0.01783	0.3586	−0.0003104	-1.477×10^{-5}
$\varphi_0^{(i)}$	0.1266	−0.1507	0.7325	−0.1628	−0.01361
$\varphi_1^{(i)}$	−0.08363	0.06429	0.05871	−0.04230	−0.003669
$\varphi_2^{(i)}$	−0.007269	0.01125	−0.004592	−0.003534	−0.0003082

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 $[1, K]$.

5.5 Shape of boundaries

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 $\alpha = 5\%$). Also, they are steeper at the beginning and end of the interval, in the sense that, for example, $b_\alpha^{Uniform}(r)/b_\alpha^{Zeileis}(r) = +\infty$ as $r \to 0$ or $r \to 1$.

The relative positioning of boundaries of different type can be observed in Figures 5a–5d, see Section 7.

6 Distribution of size

Finally, Figures 4a and 4b present the results of asymptotic simulations in two situations: one-sided monitoring boundary (with $K = 5$) for the Wiener process and two-sided retrospective boundary for the Brownian bridge. It is clear that the constructed boundaries do distribute the size uniformly across the appropriate interval.

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 non-negligible expense.

7 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, for the sake of larger sample sizes.[6] 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

[6] The (seasonally adjusted) data are taken from the FRED database of the Federal Reserve Board at `http://research.stlouisfed.org/fred2` (series AHETPI, CPIAUCNS, UNRATE).

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 (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. 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.

8 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 those resulting in a detector asymptotically behaving as a Wiener process or Brownian bridge. Nevertheless, in the literature one can encounter, although much more rarely, other asymptotic processes for sequential detectors, which are usually functions of the two processes above. For example, recursive predictability tests (Inoue and Rossi, 2005) yields as an asymptotic process a squared Bessel process $W_p(r)'W_p(r)$, where $W_p(r)$ is a p-variate Wiener process; predictive testing for parameter constancy (Ghysels, Guay and Hall, 1997) yields as an asymptotic

process $B_p(r)'B_p(r) + W_{q-p}(r)'W_{q-p}(r)$, where $B_p(r)$ is a p-variate Brownian bridge. Some of such cases (those where the transitional density has a convenient analytical form) can be handled similarly to the technique we have proposed. More problematic may be MOSUM-type tests (Chu, Hornik and Kuan, 1995), where the asymptotic process is increments of the Brownian bridge $B(r) - B(r-h)$ for fixed $h \in (0,1)$ because of the presence of an additional parameter. The problems just described may constitute an agenda for future work.

Acknowledgments

We would like to thank seminar participants at various economics departments throughout the world: University of Helsinki, University of Warwick, Humboldt Universitat zu Berlin, European University Institute, University of Copenhagen, Tinbergen Institute Amsterdam, Catholic University of Leuven, Universita degli Studi dell'Insubria, Universita di Bologna, University of Haifa, Bar-Ilan University, Bilkent University, Monash University, University of New South Wales, University of Adelaide, CIREQ, ECARES, EIEF, QUT, CERGE-EI, CEU, NES and HSE. We also thank Alexey Belkin, Geert Dhaene and Franco Peracchi for useful discussions. Victoria Stepanova has provided excellent research assistance.

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*, forthcoming.

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

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

[8] 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.

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

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

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

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

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

[14] 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.

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

[16] 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.

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

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

[19] Kokoszka, P. and R. Leipus (2000) Change-point estimation in ARCH models. *Bernoulli* 6, 1–28.

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

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

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

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

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

[25] Romano, J.R. and M. Wolf (2007) Control of generalized error rates in multiple testing. *Annals of Statistics* 35, 1378–1408.

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

[27] 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),

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

These two statements imply that

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

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. $\square$

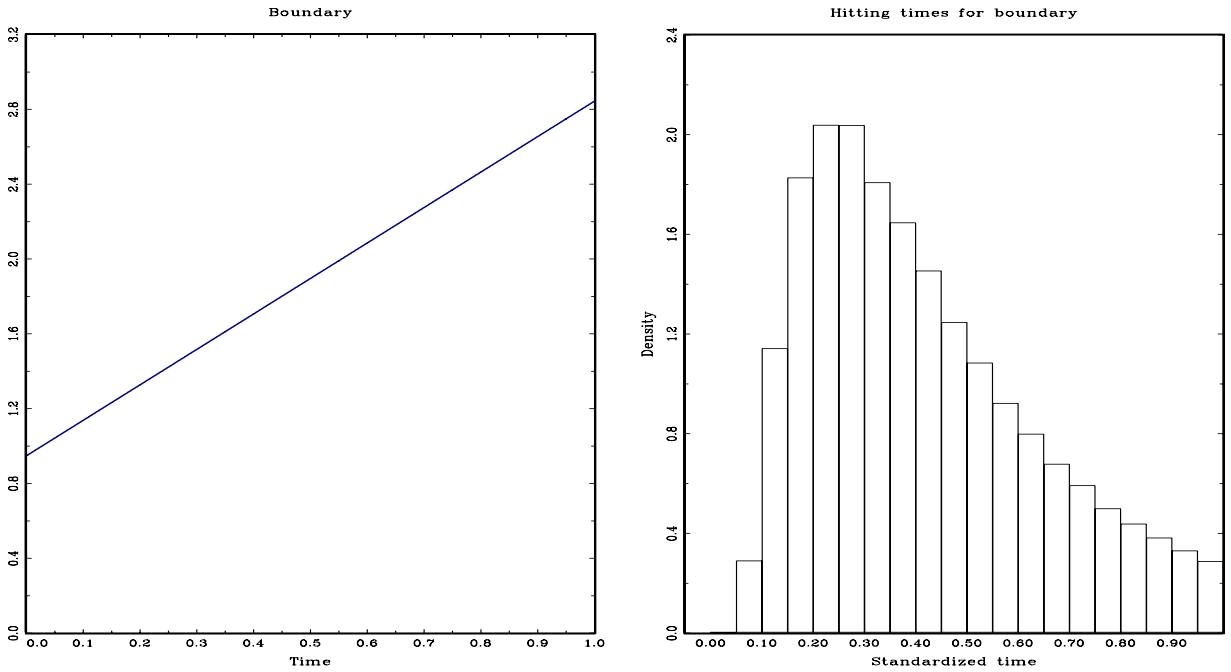

Figure 1a. Two-sided linear boundary for $W(r)$ on $[0,1]$ with distribution of size $\alpha = 5\%$

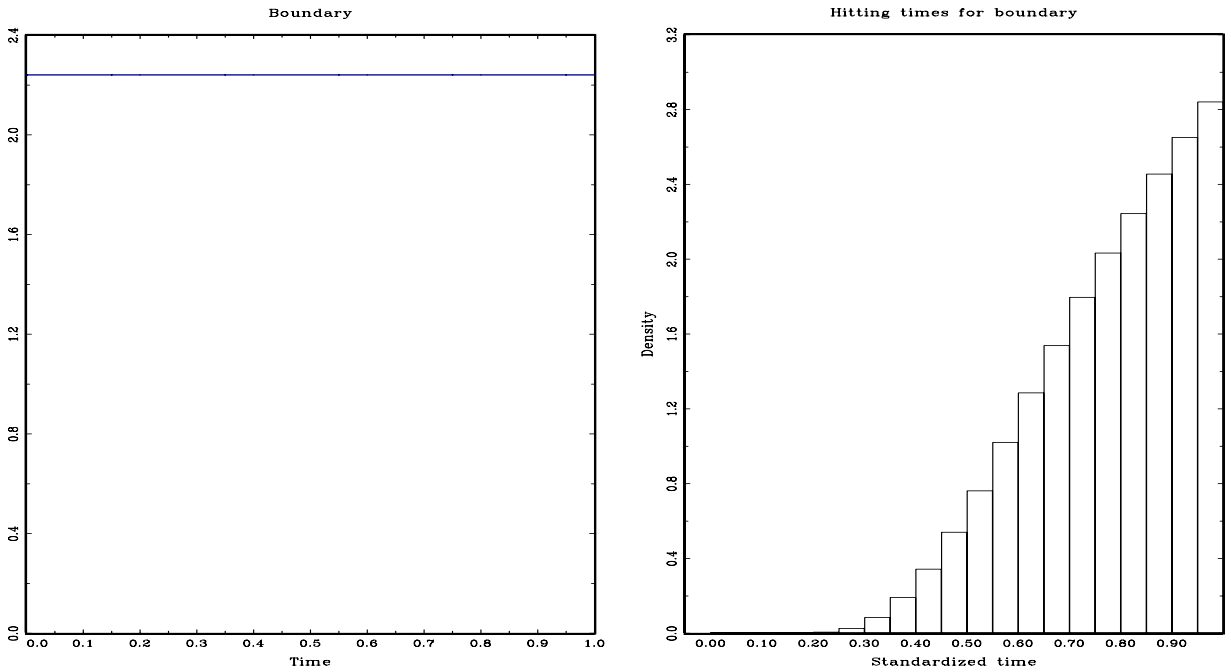

Figure 1b. Two-sided horizontal boundary for $W(r)$ on $[0,1]$ with distribution of size $\alpha = 5\%$

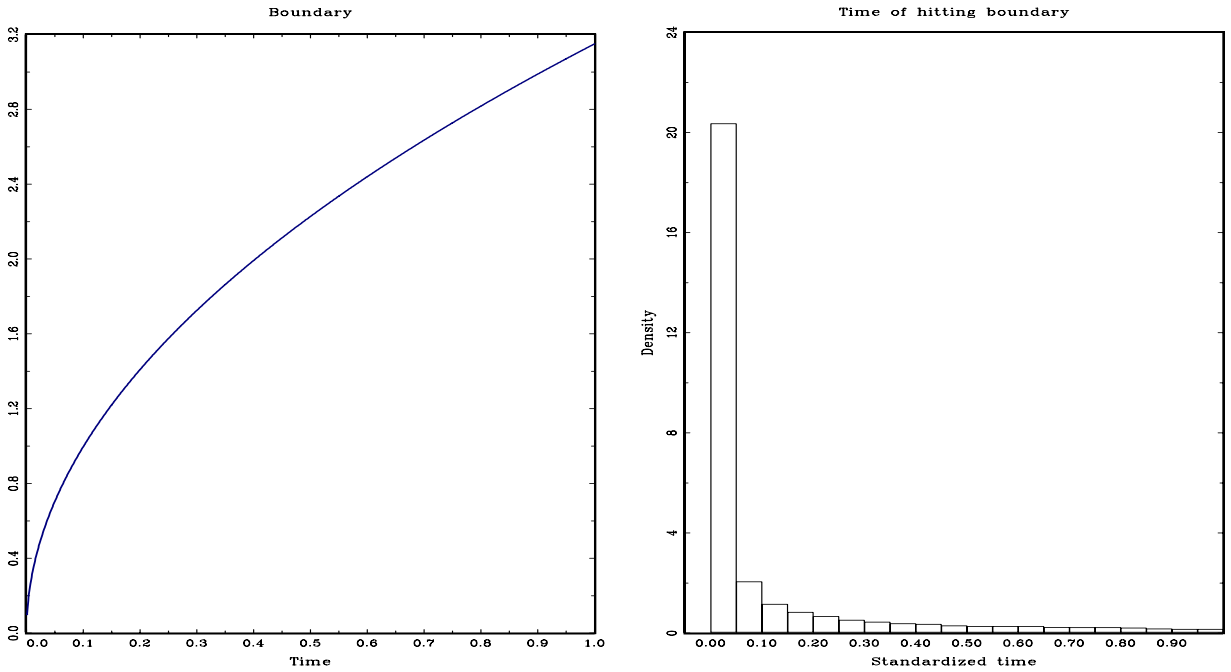

Figure 1c. Two-sided Zeileis boundary for $W(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

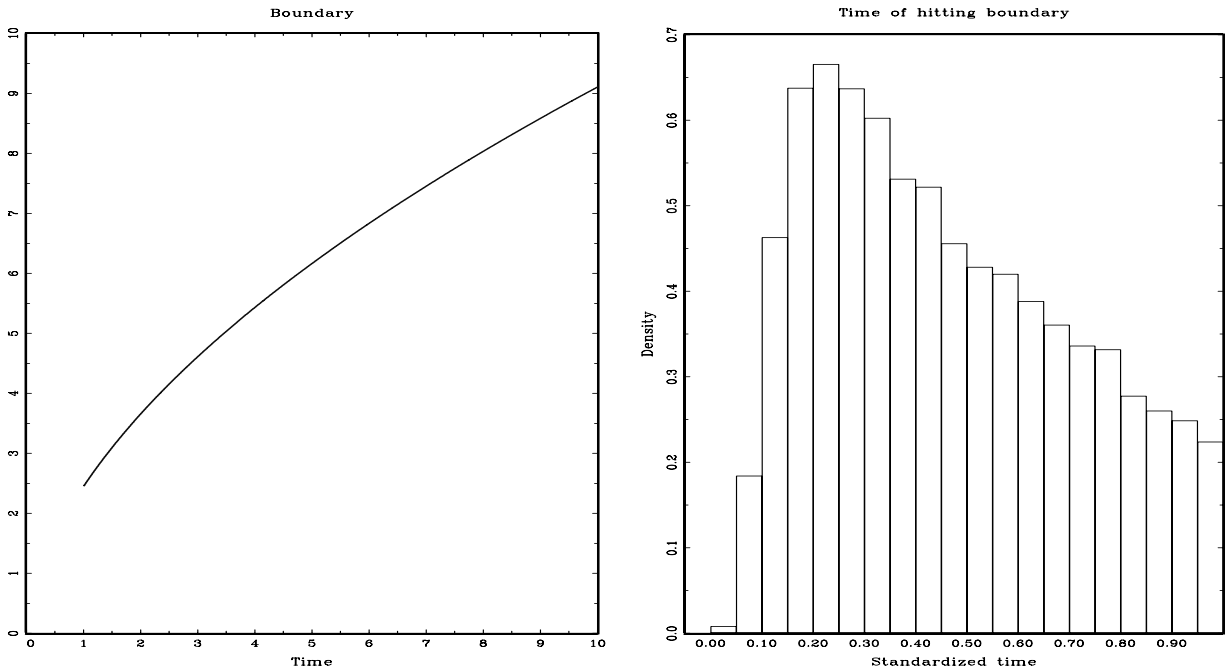

Figure 1d. Two-sided parabolic boundary for $W(r-1)$ on $[1, \infty)$ with distribution of size $\alpha = 5\%$

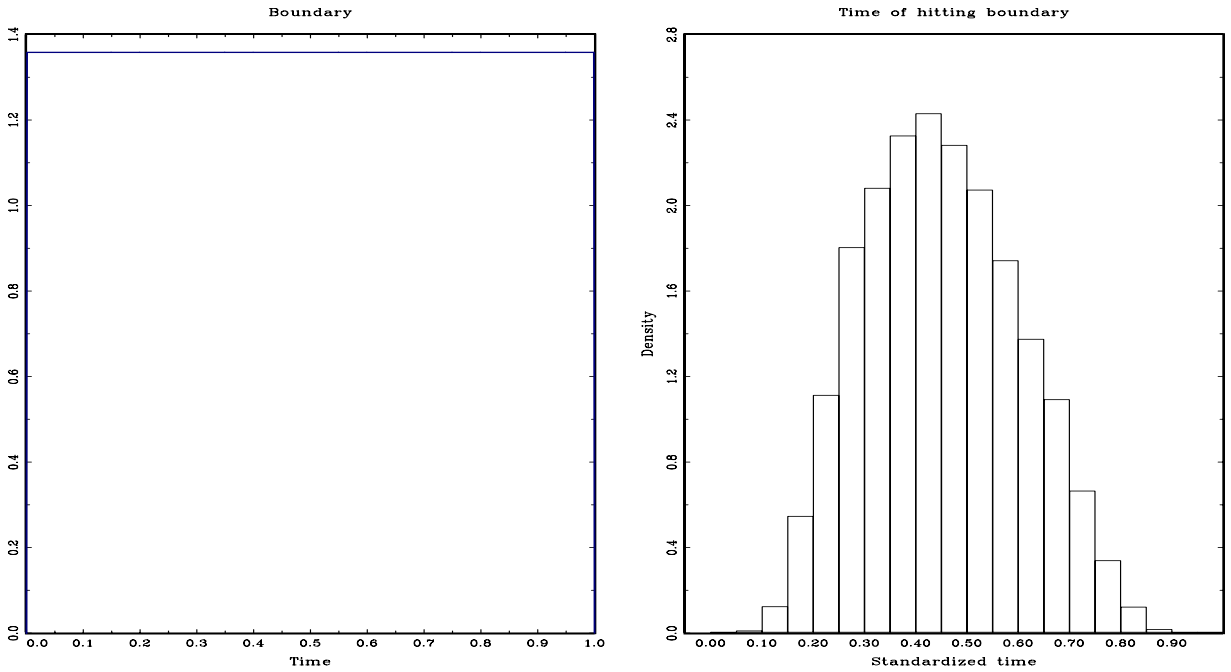

Figure 2a. Two-sided horizontal boundary for $B(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

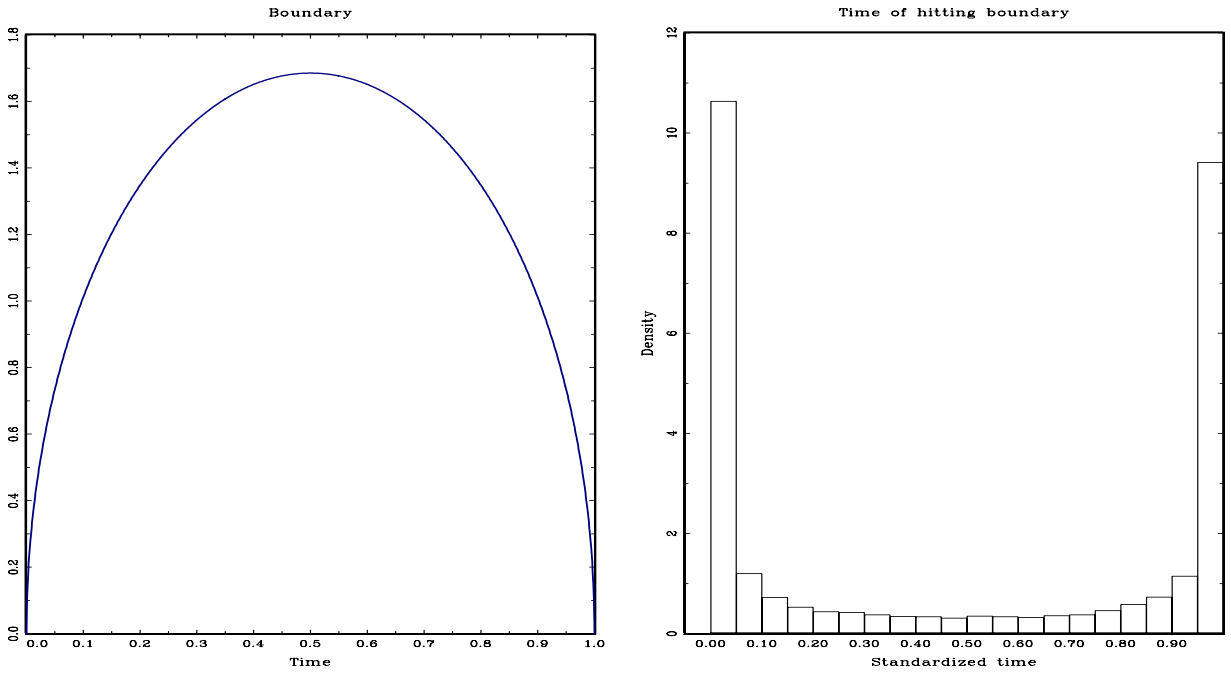

Figure 2b. Two-sided Zeileis boundary for $B(r)$ on $[0, 1]$ with distribution of size $\alpha = 5\%$

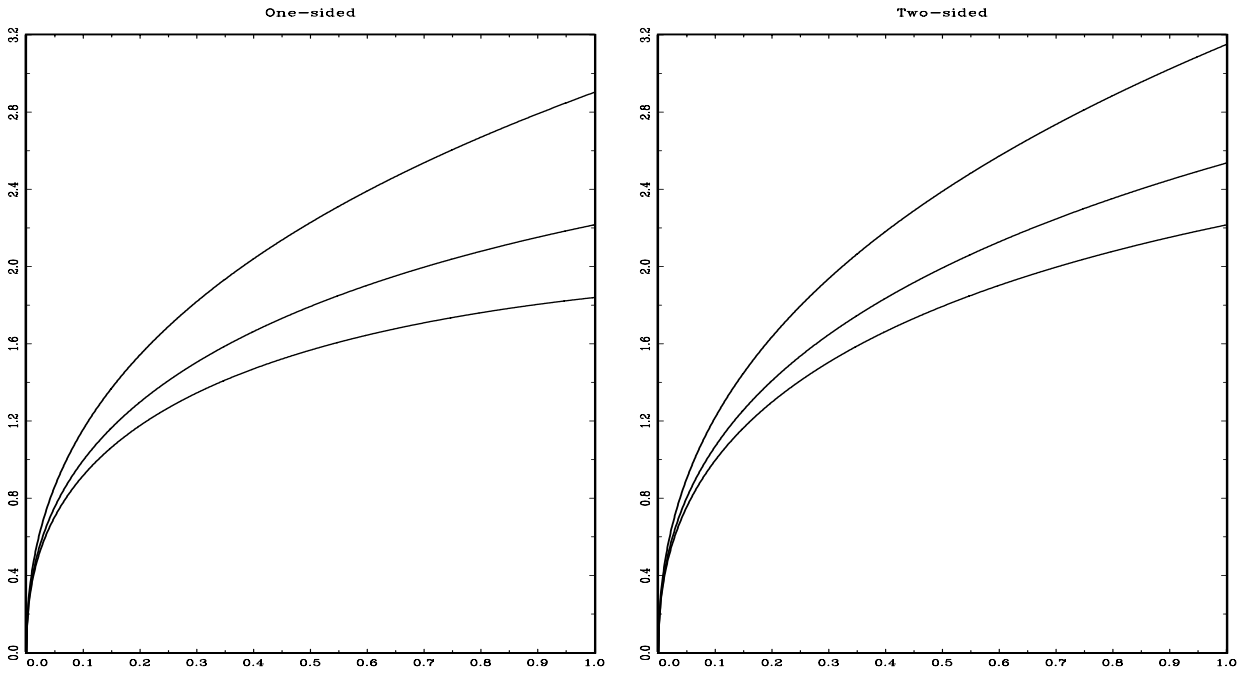

Figure 3a. Shapes of boundaries for $W(r)$ on $[0,1]$ with uniform distribution of sizes 1%, 5% and 10%

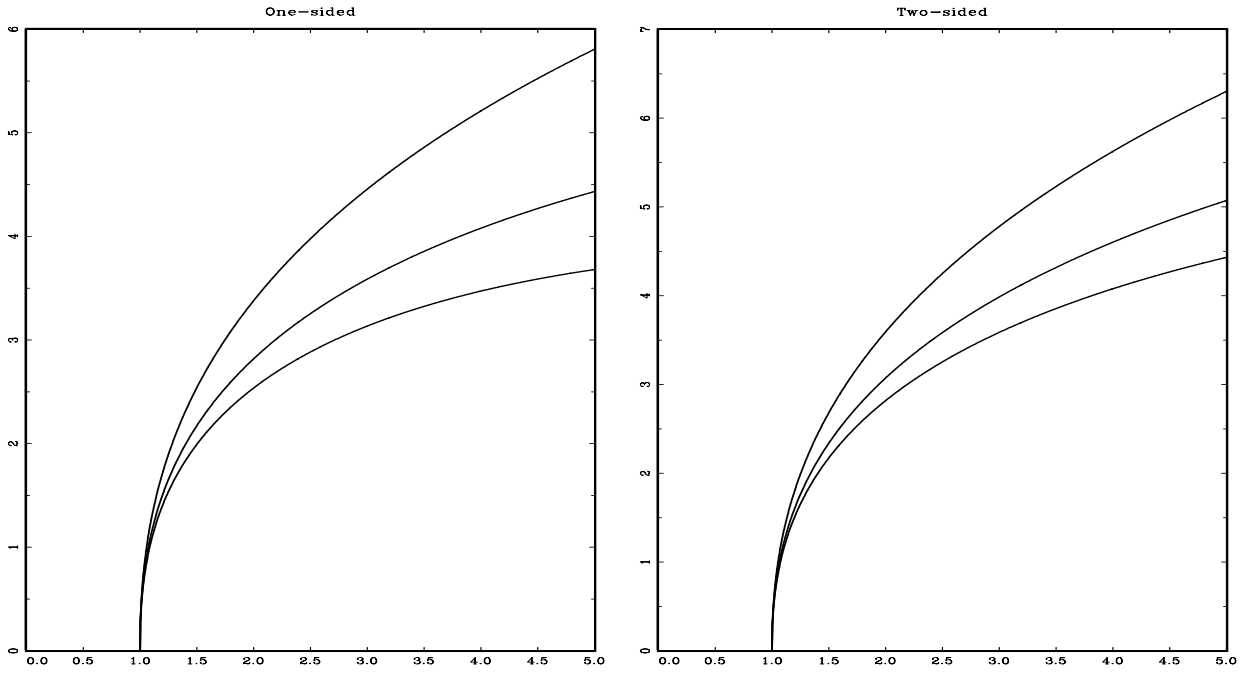

Figure 3b. Shapes of boundaries for $W(r-1)$ on $[1,5]$ with uniform distribution of sizes 1%, 5% and 10%

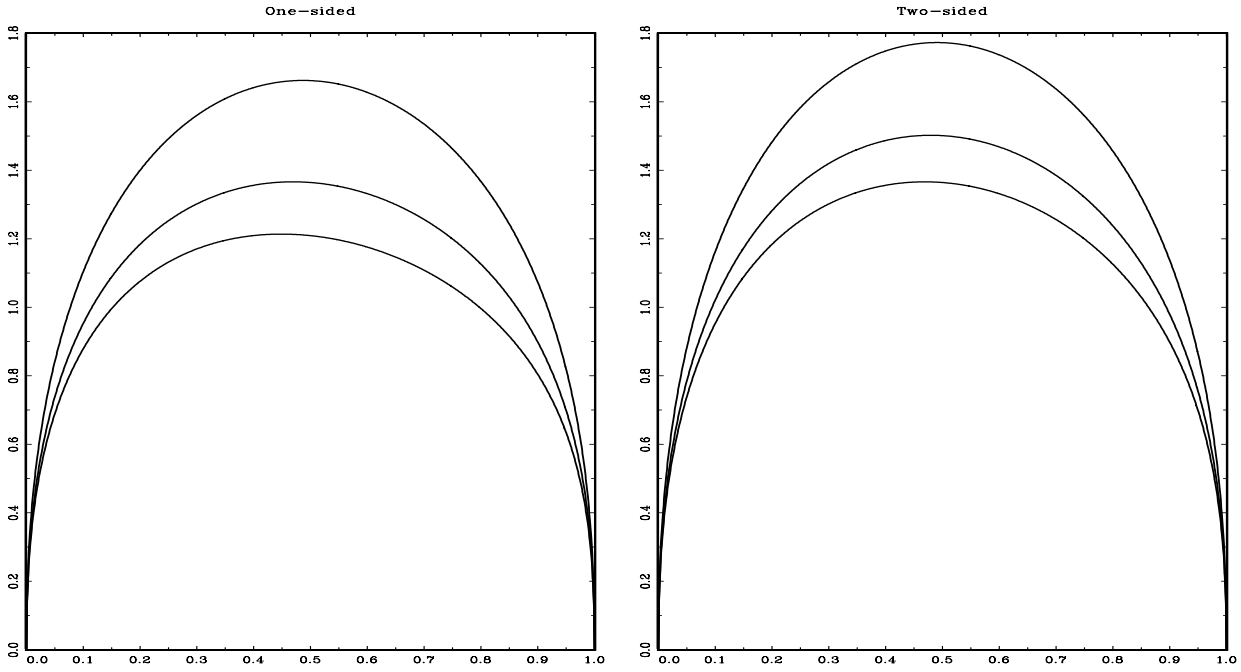

Figure 3c. Shapes of boundaries for $B(r)$ on $[0,1]$ with uniform distribution of sizes 1%, 5% and 10%

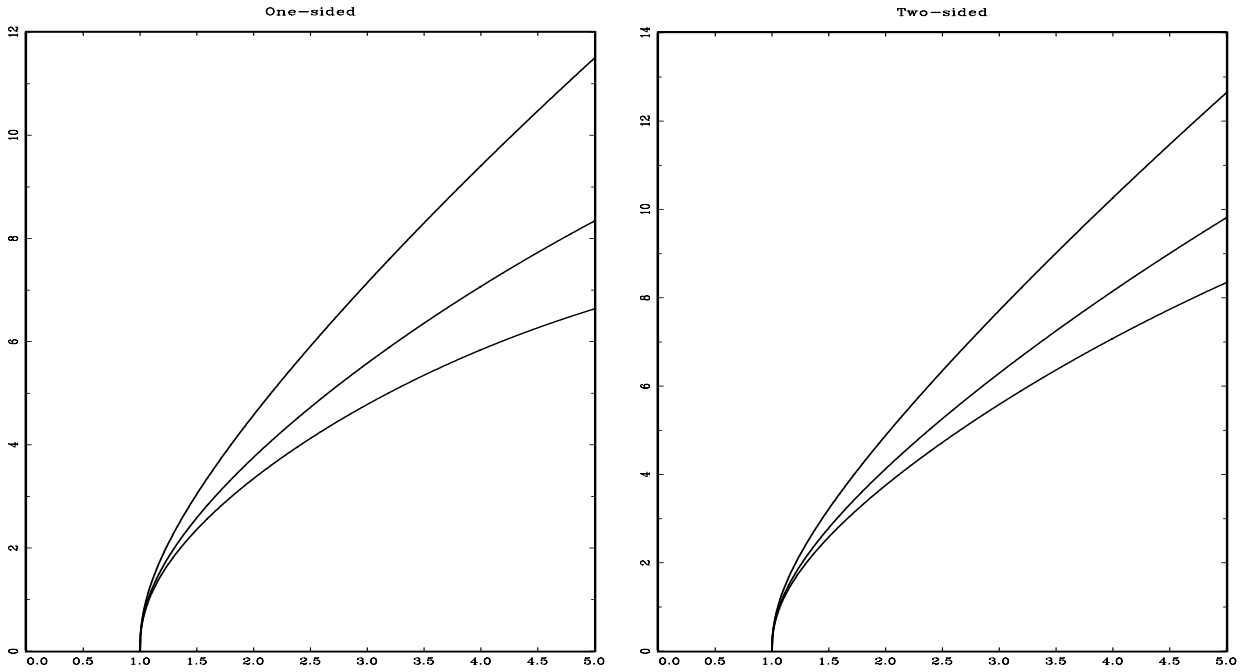

Figure 3d. Shapes of boundaries for $B(r)$ on $[1,5]$ with uniform distribution of sizes 1%, 5% and 10%

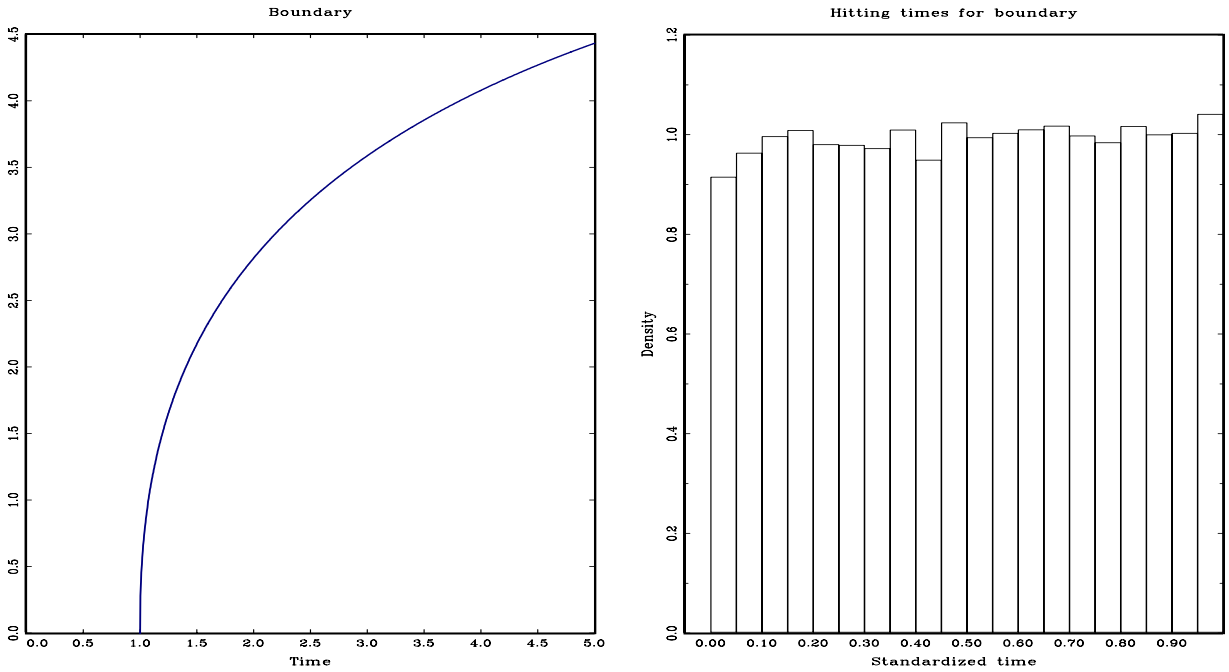

Figure 4a. One-sided boundary for $W(r-1)$ on $[1,5]$ with uniform distribution of size $\alpha = 5\%$

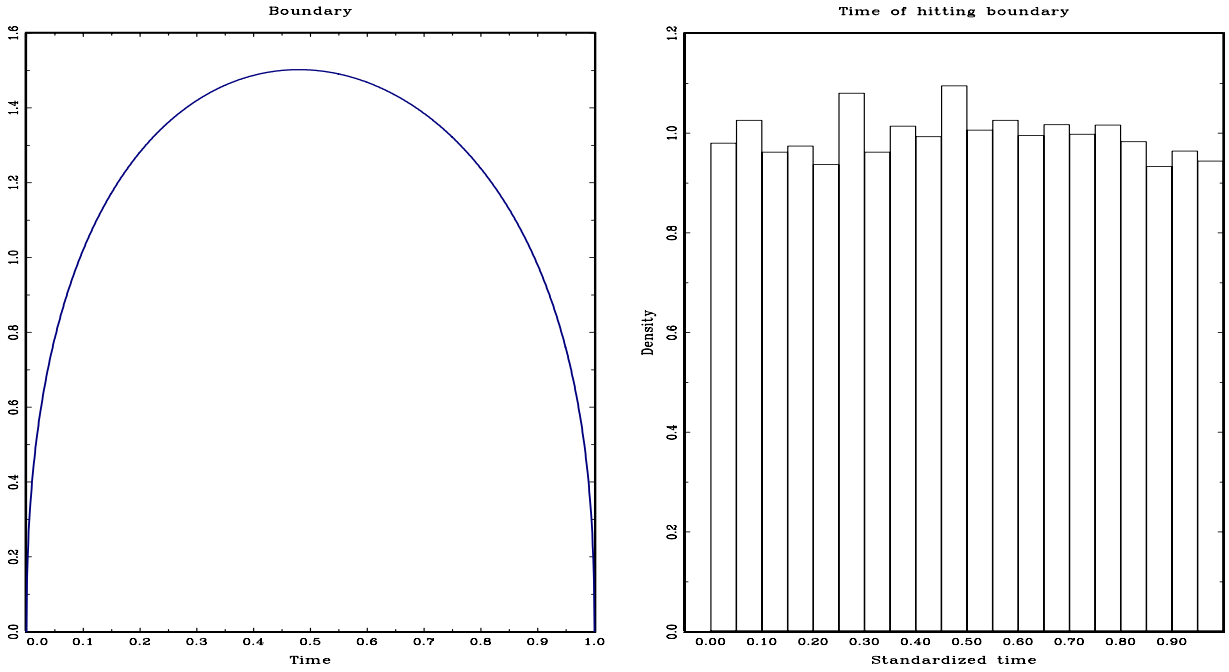

Figure 4b. Two-sided boundary for $B(r)$ on $[0,1]$ with uniform distribution of size $\alpha = 5\%$

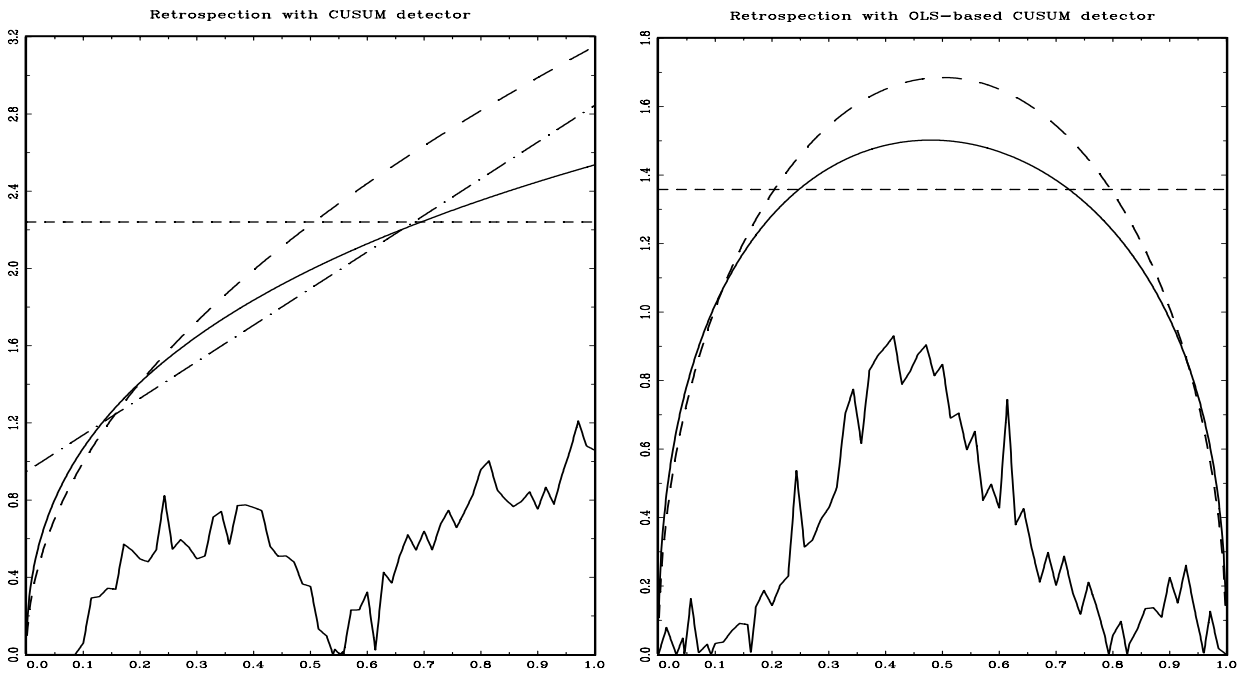

Figure 5a. CUSUM and OLS-based CUSUM retrospection with various boundaries in empirical application, shorter historical interval, size 5%

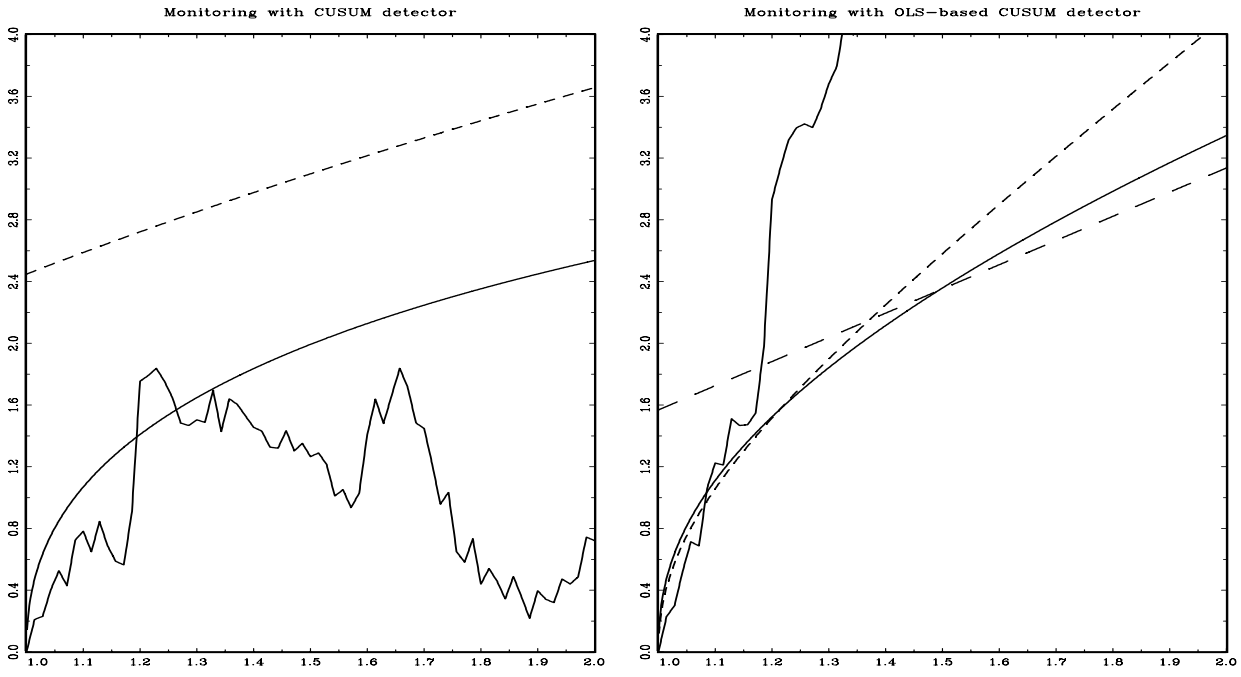

Figure 5b. CUSUM and OLS-based CUSUM monitoring with various boundaries in empirical application, horizon 2, size 5%

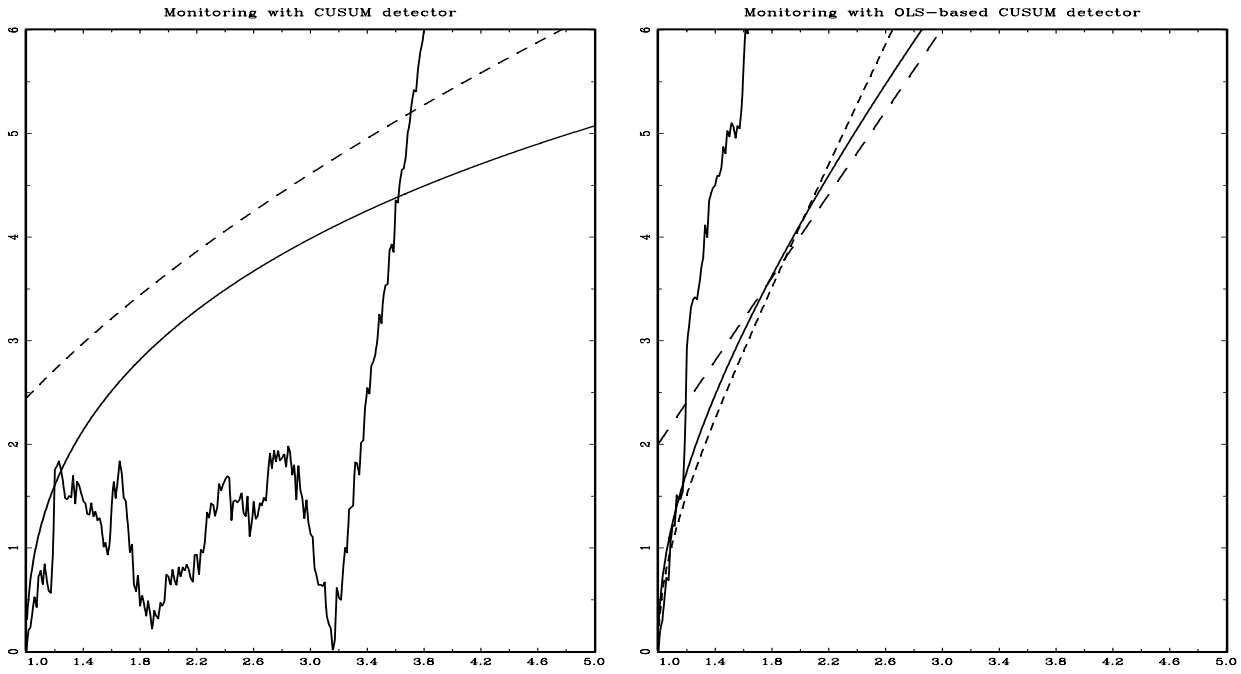

Figure 5c. CUSUM and OLS-based CUSUM monitoring with various boundaries in empirical application, horizon 5, size 5%

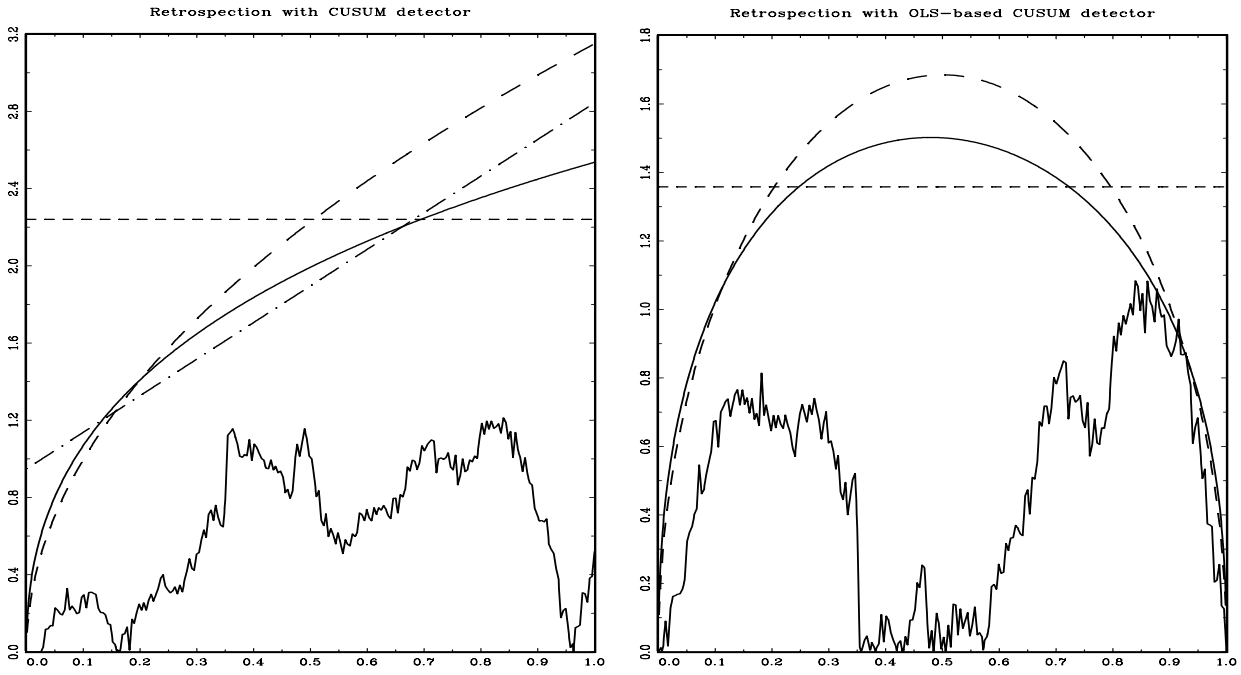

Figure 5d. CUSUM and OLS-based CUSUM retrospection with various boundaries in empirical application, longer historical interval, size 5%