Statistics and Its Interface Volume 0 (2015) 1–14

# Real-time financial surveillance via quickest change-point detection methods

Andrey Pepelyshev and Aleksey S. Polunchenko

We consider the problem of efficient financial surveillance aimed at "on-the-go" detection of structural breaks (anomalies) in "live"-monitored financial time series. With the problem approached statistically, viz. as that of *multi-cyclic* sequential (quickest) change-point detection, we propose a semi-parametric multi-cyclic change-point detection procedure to promptly spot anomalies as they occur in the time series under surveillance. The proposed procedure is a derivative of the likelihood ratio-based Shiryaev–Roberts (SR) procedure; the latter is a quasi-Bayesian surveillance method known to deliver the fastest (in the multi-cyclic sense) speed of detection, whatever be the false alarm frequency. We offer a case study where we first carry out, step by step, a preliminary statistical analysis of a set of *real-world* financial data, and then set up and devise*(a)* the proposed SR-based anomaly-detection procedure and *(b)* the celebrated Cumulative Sum (CUSUM) chart to detect structural breaks in the data. While both procedures performed well, the proposed SR-derivative, conforming to the intuition, seemed slightly better.

AMS 2000 subject classifications: Primary 62L10, 62L15; secondary 62P05.

Keywords and phrases: CUSUM chart, Financial surveillance, Sequential analysis, Shiryaev–Roberts procedure, Quickest change-point detection.

## 1. INTRODUCTION

The world's history of economic crises, including the latest and still-ongoing global financial meltdown and recession that started in 2008–2009, provides graphic evidence of the importance of efficient methods for continuous financial surveillance [7, 8]. By allowing to detect anomalous patters *early* and *reliably*, such methods form a foundation for *active* risk management [20]. This paper examines the possibility of approaching the problem of financial monitoring statistically. Specifically, the principal idea is to exploit the machinery of sequential (quickest) change-point detection. The subject is concerned with the development and evaluation of "watch dog"-type of procedures for early yet reliable detection of unanticipated changes (structural breaks) that may occur in the statistical profile of a "live"-monitored time series. For an introduction into the subject, see, e.g., [44, 59, 1, 36, 55], or [47, Part II], and the references therein.

One of the first comprehensive expositions of nonparametric change-point detection-based methods for *online* financial surveillance was offered by Brodsky and Darkhovsky [2, 3]. More recently, the machinery of Singular Spectrum Analysis (SSA) has also been utilized in [10, 23, 58, 11]. In particular, it was demonstrated via numerous case studies involving intricate real-world data that the SSA-based version of Page's [25] celebrated Cumulative Sum (CUSUM) "inspection scheme" is able to efficiently detect changes of rather complicated structure (e.g., in the frequency of a periodic component of the time series of interest).

However, nearly all of the research on the subject done to date revolves around only three change-point detection methods: the Shewhart $\bar{X}$-chart [40, 41], the CUSUM "inspection scheme" [25], and the Exponentially Weighted Moving Average (EWMA) chart [38]. Over the years, the three have *de facto* become *the* detection tools in applied sequential analysis, especially in quality control. Part of the reason is the methods' simplicity, and another part is their theoretically established strong optimality properties [24, 37, 29]. By contrast, the focus of this paper is on the Shiryaev–Roberts (SR) procedure [42, 43, 39, 44]. Although the SR procedure is only slightly "younger" than the CUSUM and EWMA charts, it has heretofore been largely neglected by practitioners as well as by statisticians. Consequently, examples of applications of the SR procedure to real-world data are extremely rare. However, the SR procedure has been recently discovered [30, 31, 45] to possess strong optimality properties in Shiryaev's [42, 43, 44] multi-cyclic setting, which is a setting adequate in many real-world applications. Motivated by this, the authors of [35, 52] have successfully applied the SR procedure in the area of cybersecurity, namely for online detection of anomalies (caused, e.g., by intrusions) in computer networks. The present paper is intended to provide yet another example of an SR-type anomaly-detection algorithm capable of operating on real-world *financial* data. Due to the exact multi-cyclic optimality of the SR procedure, the proposed algorithm is expected to compare favorably to other detection schemes, in particular the multi-cyclic CUSUM procedure.

We would like to remark that, to the best of our knowledge, the only other attempt to apply the SR procedure to

arXiv: 1509.01570

*real-world financial data* would be that made previously by Ergashev [6]. Specifically, Ergashev [6] was concerned with the problem of early detection of the "turning points" in the US business cycles. These cycles, also known as the US economic cycles, are alternating periods of recession and recovery, manifested in fluctuations of the US economic activity around its long-term potential level. Hence, the "turning points" effectively signify the onset of either recession (contraction) or recovery (expansion) of the US economy. To detect these "turning points", Ergashev [6] applied the SR procedure and the CUSUM and EWMA charts to the series of Composite Leading Indicators (CLIs); the CLIs are updated monthly by the Organisation for Economic Co-operation and Development (OECD; see on the Web at `http://www.oecd.org`) to provide early signals of "turning points" in the US business cycles. Through experiments involving the actual CLIs series, Ergashev [6] demonstrated the SR procedure to be better (i.e., quicker) at detecting the US business cycles' "turning points" than the CUSUM and EWMA charts with the same level of the "false positive" risk. In this work we too provide experimental evidence that the SR procedure might be superior to the CUSUM chart when it comes to detecting structural breaks in time series of *real-world* stock prices.

The rest of the paper is organized as follows. We start in Section 2 with a brief introduction to the area of quickest change-point detection and provide a short overview of the state-of-the-art in the field. Next, in Section 3 we offer an SR-based anomaly-detection algorithm suitable to operate on real-world data. Section 4 is devoted to a case study where we devise the proposed algorithm to perform anomaly-detection in a *real-world* financial time series. The conclusions follow in Section 5 which sums up the entire paper.

## 2. PRELIMINARY BACKGROUND ON QUICKEST CHANGE-POINT DETECTION

The aim of this section is two-fold: *(a)* to provide a short but formal introduction to the problem of quickest change-point detection and *(b)* to give a brief account of the state-of-the-art in the field. This is necessary as background for the later sections. For lack of space, we shall only consider the basic *iid* version of the quickest change-point detection problem. For a thorough treatment the general non-iid case, see, e.g., [44, 49, 34] or [47, Part II].

Suppose one is able to *sequentially* observe a time series, $\{X_n\}_{n\geq 1}$, where $X_i$'s are independent. Suppose further that the statistical structure of the series is such that $X_1,\ldots,X_\nu$ are each distributed according to a known probability density function (pdf) $f(x)$, while $X_{\nu+1},X_{\nu+2},\ldots$ each have a pdf $g(x)\not\equiv f(x)$, also known. The basic *iid* quickest change-point detection problem is to detect, as one gathers more and more data, that the baseline pdf of the data is no longer $f(x)$, and do so in an optimal manner. The challenge is that the time index $\nu$, which is referred to as the change-point, is not known in advance and may take place at any time $0\leq\nu\leq\infty$; here and onward, the notation $\nu=0$ ($\nu=\infty$) is to be understood as the case when the change is in effect from the get-go (or never, respectively). The minimax version of the problem assumes that $\nu$ is unknown (but not random). This is different from the Bayesian version of the problem which regards $\nu$ as random [42, 43, 44]. In this work we shall focus only on the minimax case.

Statistically, the problem is to sequentially test the hypotheses $\mathcal{H}_k\colon\nu=k$, $0\leq k<\infty$ (i.e., that the pdf of the observations changes at epoch $k$) against the alternative hypothesis $\mathcal{H}_\infty\colon\nu=\infty$ (i.e., that the pdf never changes); note that $\mathcal{H}_i\cap\mathcal{H}_j=\varnothing$, $i\neq j$, and that $\cup_{j\geq 0}\mathcal{H}_j=\Omega$.

The first step to test $\mathcal{H}_k$ against $\mathcal{H}_\infty$ is to construct the corresponding likelihood ratio (LR). To that end, assuming $X_1,X_2,\ldots,X_n$ have been sampled, the LR is of the form

$$\Lambda_{k:n}\triangleq\prod_{j=k+1}^{n}\Lambda_j,\;\text{ where }\;\Lambda_j\triangleq\frac{g(X_j)}{f(X_j)}$$

for $k<n$ and $\Lambda_{k:n}\equiv 1$ for $k\geq n$; the latter condition merely means that the change has not yet happened. The sequence $\{\Lambda_{k:n}\}_{1\leq k\leq n}$ has to be updated "on-the-go" incorporating new data points as they become available.

Once constructed, the LR is turned into a *detection statistic* to be subsequently used for actual decision-making. Basing the detection statistic on the LR ensures that the former is sensitive to whether the sample drawn so far is statistically homogeneous or not. There are generally two fundamentally different ways to utilize the LR to design a "good" detection statistic: either exploit the maximum likelihood principle or take the (generalized) Bayesian approach. This is shown schematically in Figure 1.

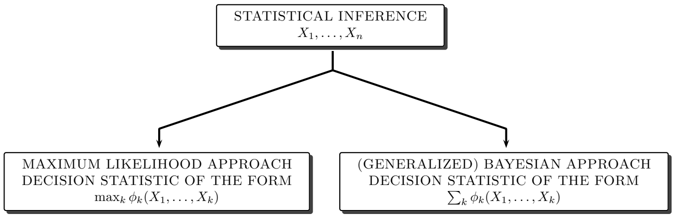

STATISTICAL INFERENCE
$X_1,\ldots,X_n$
MAXIMUM LIKELIHOOD APPROACH
DECISION STATISTIC OF THE FORM
$\max_k\phi_k(X_1,\ldots,X_k)$
(GENERALIZED) BAYESIAN APPROACH
DECISION STATISTIC OF THE FORM
$\sum_k\phi_k(X_1,\ldots,X_k)$

Figure 1: Two different approaches to statistical inference: maximum likelihood and (generalized) Bayesian.

The idea of the maximum likelihood approach is to sequentially maximize $\{\Lambda_{k:n}\}_{1\leq k\leq n}$ with respect to the change-point $\nu=k$, where $k=1,2,\ldots,n$. Specifically, the corresponding detection statistic is

$$V_n\triangleq\max_{1\leq k\leq n}\Lambda_{k:n},\;n\geq 1,\tag{1}$$

which is the famous CUSUM statistic [25]. We note that the maximization with respect to $k$ in the right-hand side of (1) is possible because the change-point $\nu = k$ is assumed *unknown* (nonrandom).

By contrast, the Bayesian approach treats the change-point as a random number, possessing a certain prior distribution [9, 42, 43, 49, 34]. However, since we agreed to assume that $\nu$ is unknown (nonrandom), the corresponding quasi-Bayesian (or generalized Bayesian) detection statistic can be defined as

$$R_n \triangleq \sum_{k=1}^{n} \Lambda_{k:n},\; n \geq 1, \tag{2}$$

i.e., $R_n$ is effectively the average of $\{\Lambda_{k:n}\}_{1\leq k\leq n}$ taken with respect to the change-point $\nu = k$, $1 \leq k \leq n$ assuming that it follows an (improper) uniform prior distribution; see, e.g., [9, 42, 43, 49, 34].

Statistics (1) and (2) are the two main choices in all of quickest change-point detection. Both lead to efficient sequential detection procedures. Specifically, a sequential detection procedure is identified with a stopping time, $T$, which is a functional of the observed data, $\{X_n\}_{n\geq 1}$. The meaning of $T$ is that after observing $X_1, \ldots, X_T$ it is declared that apparently the change is in effect. This need not be the case, and if it is not the case, then $T \leq \nu$ and the detection procedure $T$ is said to have sounded a false alarm. A "good" (i.e., optimal or nearly optimal) detection procedure is one that minimizes (or nearly minimizes) the desired detection delay penalty-function, subject to a constraint on the false alarm risk. For an overview of the major optimality criteria, see, e.g., [49, 34, 32, 55], or [47, Part II].

Let $\mathbb{P}_k(\cdot)$, $0 \leq k \leq \infty$, denote the probability measure assuming that $\nu = k$, $0 \leq k \leq \infty$ (so that $\mathbb{P}_\infty(\cdot)$ corresponds to the case when $\nu = \infty$). Let $\mathbb{E}_k[\cdot]$, $0 \leq k \leq \infty$, be the corresponding expectation.

Page [25] and then also Lorden [21] proposed to measure the "false alarm" risk through the Average Run Length (ARL) to false alarm $\mathrm{ARL}(T) \triangleq \mathbb{E}_\infty[T]$. This metric captures the average number of observations that the procedure samples before it triggers a false alarm. The higher (lower) the level of the ARL to false alarm, the lower (higher) the actual level of the "false alarm" risk.

A practical approach to quantify the detection speed is to use the "worst-case" (Supremum) Average Delay to Detection (ADD), conditional on a false alarm not having been previously occurred, i.e.,

$$\mathrm{SADD}(T) \triangleq \max_{0\leq k<\infty} \mathrm{ADD}_k(T),$$

where $\mathrm{ADD}_k(T) \triangleq \mathbb{E}_k[T-k|T > k]$, $0 \leq k < \infty$. This metric was introduced by Pollak [26].

Let

$$\Delta(\gamma) \triangleq \Big\{T\colon\; \mathrm{ARL}(T) \geq \gamma\Big\},\; \gamma > 1,$$

i.e., be the class of procedures with the ARL to false alarm of at least $\gamma > 1$, an *a priori* chosen level. Then Pollak's [26] minimax quickest change-point detection problem is to find $T_{\mathrm{opt}} \in \Delta(\gamma)$ such that $\mathrm{SADD}(T_{\mathrm{opt}}) = \inf_{T\in\Delta(\gamma)} \mathrm{SADD}(T)$ for all $\gamma > 1$. This problem is still an open one, and although there has been a continuous effort to solve it, the exact solution has been obtained in only two special cases (see [33, 51]) and, in general, only asymptotic (as $\gamma \to \infty$) solutions have been obtained so far [26, 50].

As was mentioned earlier, Page's [25] CUSUM chart has been one of the main tools for change-point detection. Part of the reason is the fact that the CUSUM chart is strictly minimax with respect to Lorden's [21] criterion for every $\gamma > 1$; see [24, 37]. The CUSUM chart is based on the maximum likelihood principle: it iteratively maximizes $\mathcal{L}_n \triangleq \log \Lambda_n$, i.e., the log-likelihood ratio (LLR), with respect to the change-point $\nu$, and stops as soon as the running maximum exceeds a certain threshold. More specifically, the CUSUM chart is based on the statistic $W_n \triangleq \max\{0, \log V_n\}$, where $V_n$ is as in (1). Note that $W_n$ satisfies the recurrence

$$W_n \triangleq \max\{0, W_{n-1} + \mathcal{L}_n\},\; n \geq 1,\; W_0 = 0. \tag{3}$$

The corresponding stopping rule is

$$\mathcal{C}_h \triangleq \min\{n \geq 1\colon W_n \geq h\}, \tag{4}$$

where $h > 0$ is a detection threshold preset so as to achieve the desired level $\gamma > 1$ of the ARL to false alarm, and thus guarantee $\mathcal{C}_h \in \Delta(\gamma)$. Since $\mathrm{ARL}(\mathcal{C}_h) \geq e^h$ for any $h > 0$ (see [21] for a proof), setting $h = h_\gamma \geq \log\gamma$ is sufficient to ensure $\mathcal{C}_h \in \Delta(\gamma)$. A more accurate approximation (mentioned, e.g., in [35]) for $\mathrm{ARL}(\mathcal{C}_h)$ is as follows:

$$\mathrm{ARL}(\mathcal{C}_h) \approx \frac{e^h}{I_g\zeta^2} - \frac{h}{I_f} - \frac{1}{I_g\zeta}, \tag{5}$$

where $I_f \triangleq -\mathbb{E}_\infty[\mathcal{L}_1]$ and $I_g \triangleq \mathbb{E}_0[\mathcal{L}_1]$ denote the Kullback–Leibler information numbers (here and throughout the rest of this section it is to be assumed that $0 < I_f < \infty$ and $0 < I_g < \infty$). The indices $I_f$ and $I_g$ that appear in the right-hand side of (5) are quantitative measures of the "contrastness" of the change, and play an important role in change-point detection.

To define $\zeta$, let $\{Z_n\}_{n\geq 0}$ be the random walk $Z_n \triangleq \sum_{j=1}^{n} \mathcal{L}_j$, $n \geq 1$, with $Z_0 = 0$. For $a \geq 0$, introduce the one-sided stopping time $\tau_a \triangleq \inf\{n \geq 1\colon Z_n \geq a\}$ and let $\kappa_a \triangleq Z_{\tau_a} - a$ denote the overshoot (i.e., the excess of $Z_n$ over the level $a$ at stopping). Then $\zeta \triangleq \lim_{a\to\infty} \mathbb{E}_0[e^{-\kappa_a}]$, which is the limiting exponential overshoot. This model-dependent constant falls within the scope of nonlinear renewal theory, and it can be shown that

$$\zeta = \frac{1}{I_g} \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k}\Big[\mathbb{P}_\infty(Z_k > 0) + \mathbb{P}_0(Z_k \leq 0)\Big]\right\}; \tag{6}$$

cf., e.g., [57, Chapters 2 & 3] and [46, Chapter VIII].

Define also $\varkappa \triangleq \lim_{a\to\infty} \mathbb{E}_0[\kappa_a]$, which is the limiting overshoot. By methods of nonlinear renewal theory it can also be shown that

$$\varkappa = \frac{\mathbb{E}_0[Z_1^2]}{2\,\mathbb{E}_0[Z_1]} + \sum_{k=1}^{\infty} \frac{1}{k}\,\mathbb{E}_0[\min\{0, Z_k\}]; \tag{7}$$

cf., e.g., [57, Chapters 2 & 3] and [46, Chapter VIII]. In practice, $\zeta$ and $\varkappa$ are usually computed numerically using (6) and (7), respectively.

It can be shown (see, e.g., [46]) that for the basic iid change-point problem $\operatorname{SADD}(\mathcal{C}_h) \equiv \mathbb{E}_0[\mathcal{C}_h]$. Let $h = h_\gamma$, where $h_\gamma$ is the solution of the equation $\operatorname{ARL}(\mathcal{C}_{h_\gamma}) = \gamma$. Then

$$\operatorname{SADD}(\mathcal{C}_{h_\gamma}) = \frac{1}{I_g}(h_\gamma + \varkappa + \beta_0) + o(1) \text{ as } \gamma \to \infty, \tag{8}$$

where $\beta_0 \triangleq \mathbb{E}_0[\min_{n\ge 0} Z_n]$. This property of the CUSUM chart is known as second order asymptotic $\operatorname{SADD}(T)$-optimality. Expansion (8) was first obtained in [4] for the single-parameter exponential family. However, it holds in a more general case as well, as long as certain mild conditions imposed on $\mathcal{L}_1$ are satisfied. See [48], where it is also shown that

$$\lim_{k\to\infty} \operatorname{ADD}_k(\mathcal{C}_{h_\gamma}) = \frac{1}{I_g}(h_\gamma + \varkappa - \beta_\infty) + o(1) \text{ as } \gamma \to \infty, \tag{9}$$

where $\beta_\infty \triangleq \lim_{n\to\infty} \mathbb{E}_\infty[Z_n - \min_{0\le k\le n} Z_k]$. In practice, constants $\beta_0$ and $\beta_\infty$ are also usually computed numerically (e.g., by Monte Carlo simulations). We also note that the two asymptotics (8) and (9) are inversely proportional to the Kullback-Leibler information number $I_g$. This number is sensitive to how faint or contrast the change is. Specifically, $I_g$ is small for faint changes, and is large otherwise. Therefore, according to (8) and (9), the average delay to detection turns out to be large for faint changes and small otherwise, which makes perfect sense.

Consider now a context in which it is of utmost importance to detect the change as quickly as possible, even at the expense of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. Put otherwise, in exchange for the assurance that the change will be detected with maximal speed, we agree to go through a "storm" of false alarms along the way (the false alarms are ensued from repeatedly applying the same detection rule, starting from scratch after each false alarm). This scenario is shown in Figure 2.

Formally, let $T_1, T_2, \ldots$ be sequential independent repetitions of the stopping time $T$, and let $\mathcal{T}_j \triangleq T_1 + T_2 + \cdots + T_j$, $j \ge 1$, be the time of the $j$-th alarm. Define $I_\nu \triangleq \min\{j \ge 1\colon \mathcal{T}_j > \nu\}$. In other words, $\mathcal{T}_{I_\nu}$ is the time of detection of a true change that occurs at $\nu$ after $I_\nu - 1$ false alarms have been raised. Write

$$\operatorname{STADD}(T) \triangleq \lim_{\nu\to\infty} \mathbb{E}_\nu[\mathcal{T}_{I_\nu} - \nu]$$

for the limiting value of the average delay to detection referred to as the *Stationary Average Delay to Detection* (STADD). The multi-cyclic change-point detection problem is to find $T_{\text{opt}} \in \Delta(\gamma)$ such that $\operatorname{STADD}(T_{\text{opt}}) = \inf_{T\in\Delta(\gamma)} \operatorname{STADD}(T)$ for every $\gamma > 1$. Since in this setup $\operatorname{ARL}(T)$ is effectively the average distance between successive false alarms, the reciprocal $1/\operatorname{ARL}(T)$ can be interpreted as the frequency of false alarms. The "intrinsic assumption" of the multi-cyclic change-point detection problem is that the process under surveillance is not expected to be affected by change "for a while", i.e., the change-point, $\nu$, is large. This is a reasonable assumption, e.g., in the area of computer network anomaly detection (see, e.g., [35, 52]) and in financial surveillance.

As has been shown in [30, 31, 45], the Shiryaev–Roberts (SR) procedure [42, 43, 39] is *exactly* optimal for every $\gamma > 1$ with respect to the stationary average detection delay $\operatorname{STADD}(T)$. Thus, in the multi-cyclic setting the SR procedure is a better alternative to the popular CUSUM chart.

The SR rule calls for stopping at epoch

$$\mathcal{S}_A \triangleq \min\{n \ge 1\colon R_n \ge A\}, \tag{10}$$

where the SR statistic $\{R_n\}_{n\ge 0}$ is given by the recursion

$$R_n = (1 + R_{n-1})\,\Lambda_n,\; n \ge 1,\; R_0 = 0; \tag{11}$$

cf. [42, 43] and [39]; here $A > 0$ is a detection threshold set *a priori* so as to ensure $\mathcal{S}_A \in \Delta(\gamma)$ for a desired $\gamma > 1$. It can be easily shown [27] that $\operatorname{ARL}(\mathcal{S}_A) \ge A$ for all $A > 0$. Hence, setting $A_\gamma = \gamma$ is sufficient to guarantee $\mathcal{S}_A \in \Delta(\gamma)$. A more accurate asymptotic approximation is $\operatorname{ARL}(\mathcal{S}_A) \approx A/\zeta$, as $A \to \infty$; see [27].

Let $R_\infty$ be a random variable that has the $\mathbb{P}_\infty$-limiting (stationary) distribution of $R_n$ as $n \to \infty$, i.e., $Q_{\text{ST}}(x) \triangleq \lim_{n\to\infty} \mathbb{P}_\infty(R_n \le x) = \mathbb{P}_\infty(R_\infty \le x)$. Let $U \triangleq \sum_{k=1}^{\infty} e^{-Z_k}$ and $\tilde{Q}(x) \triangleq \mathbb{P}_0(U \le x)$.

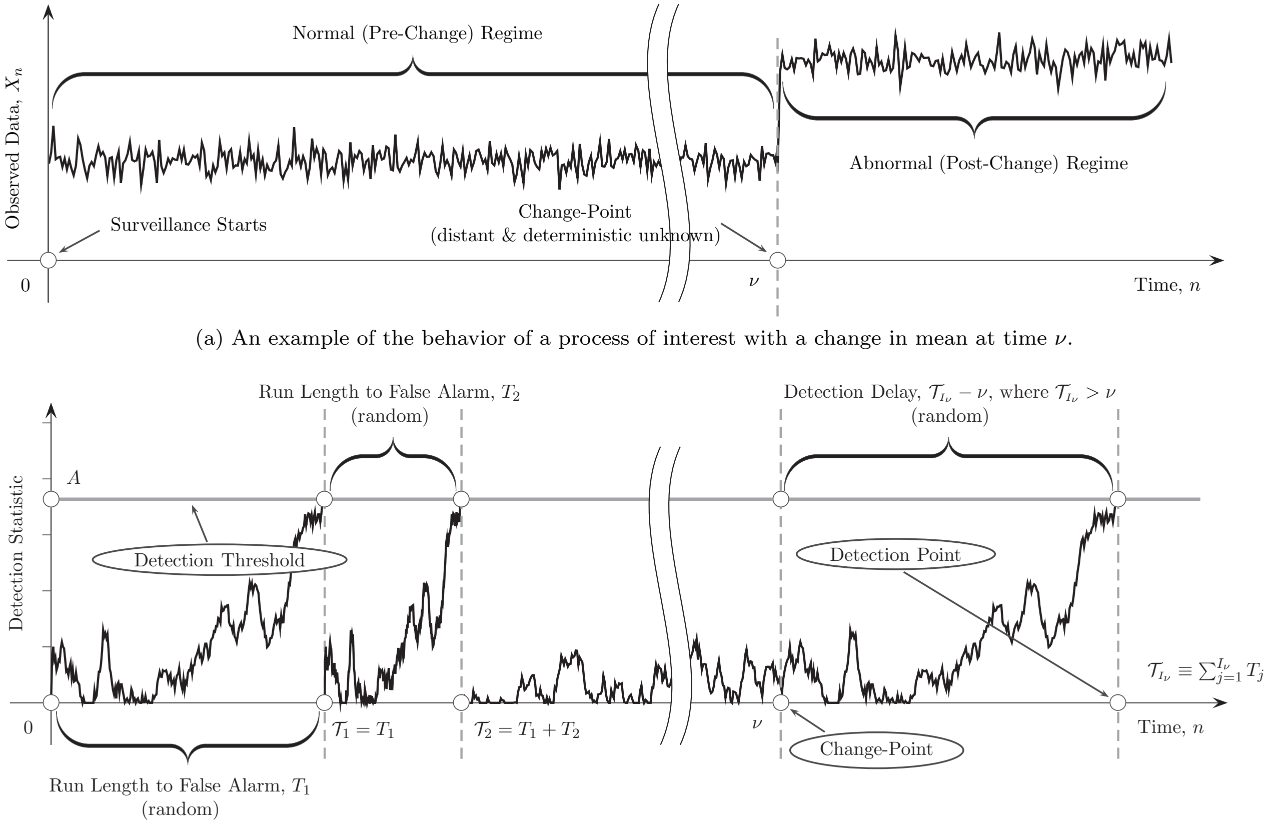
A straightforward argument shows that, for the SR procedure considered under the basic iid change-point setup, if $A = A_\gamma$ is the solution of the equation $\operatorname{ARL}(\mathcal{S}_{A_\gamma}) = \gamma$, then $\operatorname{SADD}(\mathcal{S}_{A_\gamma}) \equiv \mathbb{E}_0[\mathcal{S}_{A_\gamma}]$, and

$$\operatorname{SADD}(\mathcal{S}_{A_\gamma}) = \frac{1}{I_g}(\log A_\gamma + \varkappa - C_0) + o(1) \text{ as } \gamma \to \infty, \tag{12}$$

where

$$C_0 \triangleq \mathbb{E}[\log(1 + U)] = \int_0^\infty \log(1 + x)\, d\tilde{Q}(x);$$

cf. [50]. The asymptotic expansion (12) shows that the SR procedure is also asymptotically second-order $\operatorname{SADD}(T)$-minimax. In general, constant $C_0$ and distribution $\tilde{Q}(x)$ are amenable to numerical treatment. For cases where both can be computed analytically and in a closed form see [50] and [34].


Normal (Pre-Change) Regime
Observed Data, $X_n$
Abnormal (Post-Change) Regime
Surveillance Starts
Change-Point
(distant & deterministic unknown)
0
$\nu$
Time, $n$
Run Length to False Alarm, $T_2$
(random)
Detection Delay, $\mathcal{T}_{I_\nu} - \nu$, where $\mathcal{T}_{I_\nu} > \nu$
(random)
$A$
Detection Statistic
Detection Threshold
Detection Point
$\mathcal{T}_{I_\nu} \equiv \sum_{j=1}^{I_\nu} T_j$
0
$\mathcal{T}_1 = T_1$
$\mathcal{T}_2 = T_1 + T_2$
$\nu$
Change-Point
Time, $n$
Run Length to False Alarm, $T_1$
(random)

(a) An example of the behavior of a process of interest with a change in mean at time $\nu$.

(b) Typical behavior of the detection statistic in the multi-cyclic mode.

Figure 2: Multi-cyclic change-point detection in a stationary regime.

For the multi-cyclic setting we have

$$\mathrm{STADD}(\mathcal{S}_{A_\gamma}) = \frac{1}{I_g}(\log A_\gamma + \varkappa - C_\infty) + o(1) \text{ as } \gamma \to \infty,$$

where

$$\begin{aligned} C_\infty &\triangleq \mathbb{E}[\log(1 + R_\infty + U)] \\ &= \int_0^\infty \int_0^\infty \log(1 + x + y)\, dQ_{\mathrm{ST}}(x)\, d\tilde{Q}(y); \end{aligned}$$

cf. [50].

We conclude this section with a remark that the exact multi-cyclic optimality property of the SR procedure (11)–(10) depends heavily on the assumption that the pre- and post-change densities $f(x)$ and $g(x)$ are fully known. The consequences of setting up the SR procedure to detect the "wrong" change have been recently made clear in [5] where, apparently for the first time in the literature, it was demonstrated experimentally that, if ignored altogether, parametric uncertainty in $g(x)$ may severely affect the STADD delivered by the SR procedure: the relative loss in performance can be on the order of hundreds of percent.

## 3. APPLICATION TO FINANCIAL SURVEILLANCE

Since anomalous events in financial series happen at *unknown* points in time, and entail changes in the series' statistical properties, it is intuitively appealing to devise a quickest change-point detection method to detect the onset of such changes as rapidly as possible, while maintaining the false alarm risk at a tolerable level. This section is intended to show how quickest change-point detection can be applied to detect anomalies in "live" streams of financial data.

The main difficulty in applying either the CUSUM chart (3)–(4) or the SR procedure (11)–(10) to *real-world* data is that the pre- and post-anomaly distributions of the data are poorly understood, if known at all. As a result, any LR-based approach is effectively rendered useless. Hence, a nonparametric approach might be in order. To that end, let us first analyze how the LR exploited by both the CUSUM chart (3)–(4) and the SR procedure (11)–(10) allows the two procedures to sense the presence of a change. To that end, consider the behavior of $\mathcal{L}_n \triangleq \log \Lambda_n$ prior to the change and under the change. Before the change, the LLR has a negative expectation, i.e., $\mathbb{E}_\infty[\mathcal{L}_n] < 0$. This causes the CUSUM statistic to gravitate toward zero in the pre-change regime, and causes the SR statistic to grow slower than it would if the process had already undergone a change. However, as soon as $X_{\nu+1}$—the first "anomalous" data point—is recorded, the expectation of the LLR switches its sign to positive, i.e., $\mathbb{E}_\nu[\mathcal{L}_n] > 0$, $0 \le \nu < n$. As a result, the CUSUM statistic starts to drift away from zero up toward the detection threshold, and the SR statistic's claim rate increases compared to what it would be had there been no change. This difference in the behavior of each one of the two statistics under the pre-change regime and under the post-change regime is the main reason why the CUSUM chart and the SR procedure are able to sense the presence of a change in the observations to begin with.

The above suggests that when it is impossible to construct a LR, the latter can be replaced with a computable score function $S_n \triangleq S_n(X_1, \dots, X_n)$ such that $\mathbb{E}_\infty[S_n] < 0$ for all $n \ge 1$ and $\mathbb{E}_\nu[S_n] > 0$ for all $0 \le \nu < n$ with $n \ge 1$. This is the key element of the nonparametric approach, and in the context of quickest change-point detection this idea has been previously suggested and explored, e.g., by McDonald [22], Lai [18], Gordon and Pollak [12, 13], and recently also by Pollak [28]. A thorough exposition of the nonparametric approach to change-point detection has been offered by Brodsky and Darkhovsky [2, 3].

To be more specific, McDonald [22] suggested to base surveillance on the series of sequential ranks $U_n \triangleq \sum_{k=1}^n \mathbb{1}_{\{X_k < X_n\}}$ where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. The corresponding score function can be taken to be of the form $S_n = U_n - C$, where $C > 0$ is a design parameter selected according to the expected type of change and the desired level of the ARL to false alarm. That is, McDonald's [22] version of the CUSUM chart (3)–(4) signals an alarm according to the stopping time $\mathcal{C}_h^* = \min\{n \ge 1 \colon W_n^* \ge h\}$, where $W_n^* = \max\{0, W_{n-1}^* + S_n\}$, $n \ge 1$, and $h \ge 0$ is the detection threshold. If the observations $\{X_n\}_{n\ge1}$ are all iid, then the sequential ranks $U_n$ are approximately uniform, whatever be the observations' common baseline distribution. However, if effective the $\nu$-th data point, $X_\nu$, the baseline distribution switches to a stochastically larger distribution, the sequential ranks become larger causing the rank-based CUSUM chart to trigger an alarm. This idea of McDonald [22] was then extended to the SR procedure by Gordon and Pollak [12, 13] and by Pollak [28].

More generally, for any appropriately designed score function $S_n$, the original SR statistic $\{R_n\}_{n\ge0}$ given by (11) can be replaced with

$$\tilde{R}_n = (1 + \tilde{R}_{n-1})e^{S_n},\; n \ge 1,\; \tilde{R}_0 = 0, \tag{13}$$

so that the corresponding SR stopping time is the form

$$\tilde{\mathcal{S}}_A \triangleq \min\{n \ge 1 \colon \tilde{R}_n \ge A\}, \tag{14}$$

where $A > 0$ is the detection threshold. Likewise, for the CUSUM chart, the original CUSUM statistic $\{W_n\}_{n\ge0}$ given by (3) can be replaced with

$$\tilde{W}_n = \max\{0, \tilde{W}_{n-1} + S_n\},\; n \ge 1,\; \tilde{W}_n = 0, \tag{15}$$

so that the corresponding CUSUM stopping time becomes

$$\tilde{\mathcal{C}}_h \triangleq \min\{n \ge 1 \colon \tilde{W}_n \ge h\}, \tag{16}$$

where $h > 0$ is again the detection threshold.

In order for the score-function-based SR procedure (13)–(14) and CUSUM chart (15)–(16) to work well, the score function $S_n \triangleq S_n(X_1, \dots, X_n)$ has to be carefully designed, incorporating the type of change expected. To illustrate this point, suppose we are interested in detecting a change in both the mean and variance of the observations. Let $\mu_\infty \triangleq \mathbb{E}_\infty[X_n]$ and $\sigma_\infty^2 \triangleq \mathrm{Var}_\infty[X_n]$, and $\mu \triangleq \mathbb{E}_0[X_n]$ and $\sigma^2 \triangleq \mathrm{Var}_0[X_n]$ denote the pre- and post-anomaly mean values and variances, respectively. Introduce $\tilde{X}_n \triangleq (X_n - \mu_\infty)/\sigma_\infty$, i.e., the centered and standardized $n$-th data point. In real-world applications the pre-change parameters $\mu_\infty$ and $\sigma_\infty^2$ can usually be estimated in advance (e.g., using training data) and then periodically re-estimated to account for the nonstationary nature of the data. To deal with the uncertainty in $\mu$ and $\sigma^2$, consider the following linear-quadratic score function

$$S_n(\tilde{X}_n) = C_1\tilde{X}_n + C_2\tilde{X}_n^2 - C_3, \tag{17}$$

where $C_1$, $C_2$ and $C_3$ are design parameters; cf. [52]. Selecting $C_1, C_2$ and $C_3$ to be positive would make this score function sensitive to increases in the mean and variance. In the case when the variance either does not change at all or changes relatively insignificantly compared to the magnitude of the change in the mean, the coefficient $C_2$ may be set equal to zero. This appears to be typical for many cyber-security applications [53, 54, 52]. In the opposite case when the mean changes only slightly compared to the variance, one may take $C_1 = 0$.

Note that the score function $S_n$ given by (17) with

$$C_1 = \delta q^2, \quad C_2 = \frac{1 - q^2}{2}, \quad C_3 = \frac{\delta^2 q^2}{2} - \log q, \tag{18}$$

where $q = \sigma_\infty/\sigma$, $\delta = (\mu - \mu_\infty)/\sigma_\infty$, is optimal if the pre- and post-change distributions are Gaussian with known $\mu$ and $\sigma^2$. This is true because the score function $S_n$ given by (17) is then simply nothing but the LLR. If one has reason to believe that the time series of interest can be accurately described by the Gaussian model, then selecting $q = q_0$ and $\delta = \delta_0$ with some design values $q_0$ and $\delta_0$ would lead to decent performance of the procedure for $q < q_0$ and $\delta > \delta_0$ and optimal (i.e., best) performance for $q = q_0$ and $\delta = \delta_0$. However, it is important to emphasize that the proposed score-based "tweak" of SR procedure does not require the observations to be Gaussian, whether pre- or post-change.

For examples of score-functions that exploit SSA, see, e.g., [10, 23, 58, 11].

Another way to deal with parametric uncertainty in the observations' post-change distribution is to employ the generalized likelihood ratio (GLR) approach. However, the obvious problem with this approach is that the recursive evaluation of the running LR—either as in (1) or as in (11)—might get computationally too difficult to carry out, because now the LR has to be also maximized with respect to the unknown parameter. As a way around this, Willsky and Jones [56] and then also Lai [18, 19] suggested to restrict attention to a certain *limited* number of the most recent observations, and, based on that idea, introduced the appropriate "window-limited" modification of the CUSUM chart. The main question here, however, is how to choose the size of the window, i.e., the optimal number of the most recent observations to take into account. On the one hand, if that number is too large, the corresponding "window-limited" CUSUM statistic might still be too computationally demanding. On the other hand, basing the decision on too small a number of the latest observations is likely to lead to an increase in the detection delay. To optimize the trade-off between the computational tractability and the speed of detection, Lai [18, 19] showed that the "best" strategy is to factor in the latest $M_\gamma$ observations with $M_\gamma$ being of the order $O(\log\gamma/I_g)$ where $\gamma > 1$ is the desired level of the ARL to false alarm and $I_g \triangleq \mathbb{E}_0[\mathcal{L}_1]$ is the Kullback–Leibler information number.

# 4. A CASE STUDY

We now consider a case study where we employ the proposed change-point detection methodology to "sniff out" structural breaks in a *real-world* financial time series. Specifically, our intent is two-fold: to first provide the steps necessary to configure our change-point detection procedures and then, once the latter are properly set up, to also demonstrate and discuss their performance.

## 4.1 Data description

The time series we would like to study is the daily stock prices (at closing) of Host Hotel & Resorts, Inc. (see on the Web at www.hosthotels.com) for the period from January 3, 2000 through March 30, 2007. Host Hotel & Resorts, Inc. is the largest American lodging and real estate investment trust (or REIT) headquartered in Bethesda, Maryland, USA. An S&P 500 and Fortune 500 company, Host Hotel & Resorts, Inc. is also one of the biggest owners of luxury and upper-upscale hotels. Its hotels are operated under such reputable brand names as Marriott, Ritz–Carlton, Four Seasons, Hyatt and Hilton. Its stock is traded on the New York Stock Exchange (NYSE) under the ticker HST. Our interest in the company is due to its leading position in the industry and the significant size of its assets: as of December 31, 2014, its reported total assets were over $12 billion (with liabilities and debt totaling to about $4.6 billion) [17, p. 88].

Historical data for the HST stock for any period since the stock began trading on the NYSE are freely available on the Internet (e.g., via Yahoo.Finance; see www.finance.yahoo.com). We used the Machine Learning Data Set Repository (see on the Web at www.mldata.org). The total length of the series is $N = 1812$ data points. The choice to focus on the period between January 3, 2000 and March 30, 2007 was because the company's history was very eventful during that time period: the tragic events that took place in New York City on September 11, 2001, and the decade-long global economic unrest that followed caused considerable turbulence in the company's financial well-being. As a result, one would expect the HST stock statistical dynamics within the chosen time frame to experience multiple changes. This makes change-point analysis of the data both interesting and challenging.

## 4.2 Preliminary statistical analysis

To perform basic statistical analysis of the data, the natural point of departure would be to graph the data against time. This is done in Figure 3. A mere eye examination of the plot suggests several observations.

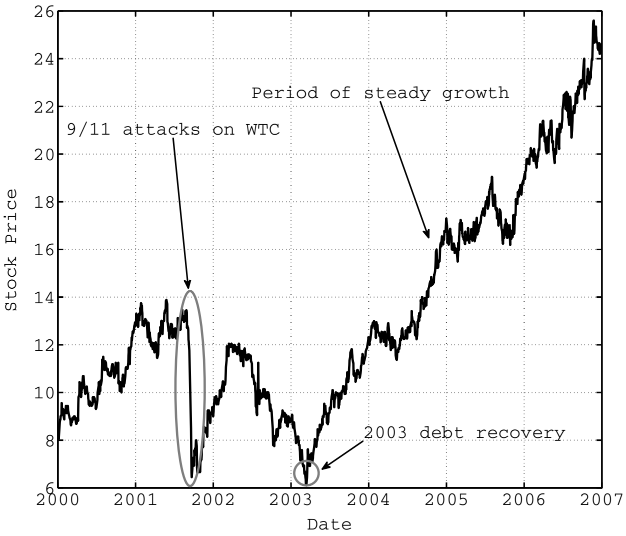


Figure 3: Daily stock prices (at closing) for the Host & Hotel Resorts, Inc. (NYSE: HST) for the period from January 3, 2003 through March 30, 2007.

First note that, as expected, the series appears to be rife with structural breaks of various scale and type. The following three are particularly notable: one occurring toward the end of the third quarter of the year 2001, followed by one more occurring around the end of the first quarter of the year 2002, followed by yet another one occurring in 2003, around the end of the first quarter.

The first of these break-points, viz. the one occurring in 2001, appears to be a crash-type event, as at that point the stock price essentially plummets, from being about $13/share right before the break to being roughly $7/share shortly after the break. The reason for such a huge loss in value is not hard to figure out: it was the result of the 9/11 terrorist attacks on the World Trade Center (WTC) Towers in New York City. Specifically, in addition to destroying the Towers, the attacks also destroyed the New York World Trade Center Marriott hotel owned and operated by Host Hotel & Resorts, Inc. To boot, the company also sustained considerable damage to its second property located nearby, the New York Marriott Financial Center hotel. However, by the end of 2001, the company received the property and business interruption insurance for the two hotels [14], and the stock began to claim up.

The second of the above three major break points, namely, the 2002 one, also appears to be a negative event in the Company's history. According to the company's 2002 annual report [15], the company's revenue for the year was negatively affected by the overall weakness of the US and global economies, which in particular resulted in business and leisure travel dropping below historic level in 2002.

The third break-point (the one occurring around the first quarter of the year 2003) appears to be a "turning point" for the company, because following this break-point, the stock begins to exhibit a consistent upward trend that lasts for years. The specific date of this "turning point" is March 14, 2003. According the company's 2003 annual report [16], 2003 was indeed a year of recovery for the company: they collected additional insurance on the hotels that were destroyed during the 9/11 attacks in 2001, sold hotels that had been found to be inefficient, and used the proceeds to substantially lessen the corporate debt.

Another observation that can be made from Figure 3 is that the HST stock appears to have a seasonal component. This should not come as a surprise, since for the hotel industry seasonal effects are common and, in fact, natural. However, dealing with such effects statistically is somewhat orthogonal to the objective of our study. Nevertheless, we would like to mention that the numerous and extensive case studies offered, e.g., in [10, 23, 11], suggest that the SSA methodology can be rather efficient in the analysis of seasonal and cyclic patterns.

To reinforce the observations made so far, Figure 4 shows the behavior of the daily returns $d_i \triangleq X_{i+1} - X_i$, $i = 1, \ldots, N-1$, on the HST stock. The returns provide a different prospective onto the behavior of the stock itself. As a matter of fact, it is the returns that are usually used as the input data to perform statistical inference on the underlying stock itself. Therefore, we also shall proceed with the returns being the series of interest.

One can clearly see a large down-pointing spike around the third quarter of the year 2001. This spike corresponds to the HST stock loosing approximately half of its value as the result of the 9/11 terrorist attacks in NYC. While this spike is extremely contrast, there is no apparent change in the daily return distribution corresponding to the 2003 structural break. Nevertheless, as we shall see shortly, the 2003 break-point *is* detectable. More importantly, in spite of the steady growth of the stock after the 2003 break-point shown in Figure 3, the behavior of the return series does not confirm any shift in the mean.

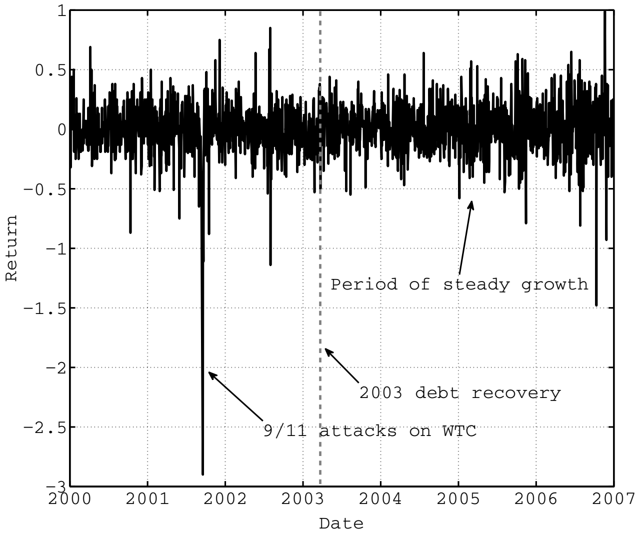


Figure 4: Daily returns on the stock (evaluated at closing) of the Host & Hotel Resorts, Inc. (NYSE: HST) for the period from January 3, 2003 through March 30, 2007.

## 4.3 Offline structural break detection

We now perform a more thorough statistical analysis of the data. Specifically, we would like to devise a statistical procedure to detect the aforementioned structural breaks. Toward this goal, the first step is to analyze the series *retrospectively* so as to not only detect the changes, but also to estimate their locations. One such "offline" change-point detection-estimation statistic is the Brodsky–Darkhovsky statistic proposed and studied in [2, 3]. The Brodsky–Darkhovsky statistic is defined as

$$(19)\quad Y_N(n) \triangleq \sqrt{\frac{n(N-n)}{N^2}}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{N-n}\sum_{i=n+1}^{N} X_i\right],$$

where $1 \leq n \leq N-1$. As can be seen from the structure of the statistic, it is effectively the difference between two sample means: one computed off the first $n \geq 1$ data points (i.e., $X_1, \ldots, X_n$), and one computed off the remaining $N-n$ data points (i.e., $X_{n+1}, \ldots, X_N$) in a chunk of $N \geq n+1$ observations $X_1, \ldots, X_N$. Therefore the statistic (19) is tailored specifically to detect deviations in the observations' mean. The actual detection procedure consists in comparing $|Y_N(n)|$ indexed by $n = 1, \ldots, N-1$ against a threshold selected according to the desired significance level. More importantly, the statistic can also be used to estimate the actual change-point, $\nu$, i.e., the time moment at which the series' baseline mean (apparently) changes. Specifically, this is accomplished by first identifying the set of values of $n$ for which $|Y_N(n)|$ is maximized, and then using any such $n$ as an estimator of the actual change-point, that is,

$$\hat{\nu}_N \triangleq \operatorname*{arg\,max}_{1\leq n\leq N-1} |Y_N(n)|\,. \tag{20}$$

It has been shown in [2] that such an estimator enjoys strong consistency (as $N \to \infty$) with exponential rate of convergence.

We have applied the Brodsky–Darkhovsky approach to the returns $\{d_i\}_{1\leq i\leq N}$, and the obtained behavior of $Y_N(n)$ for $1 \leq n \leq N-1$ is shown in Figure 5. It can be seen from the figure that the statistic exhibits a whole series of local yet fairly contrast maxima. The unique and rather strong absolute maximum occurring around the first quarter of 2003 reinforces the observation made earlier that the HST stock undergoes a structural break at that time. The specific location of the absolute maximum corresponds to March 14, 2003, which is the Brodsky–Darkhovsky estimate of the actual change-point. We note that this date is precisely the 2003 break-point.

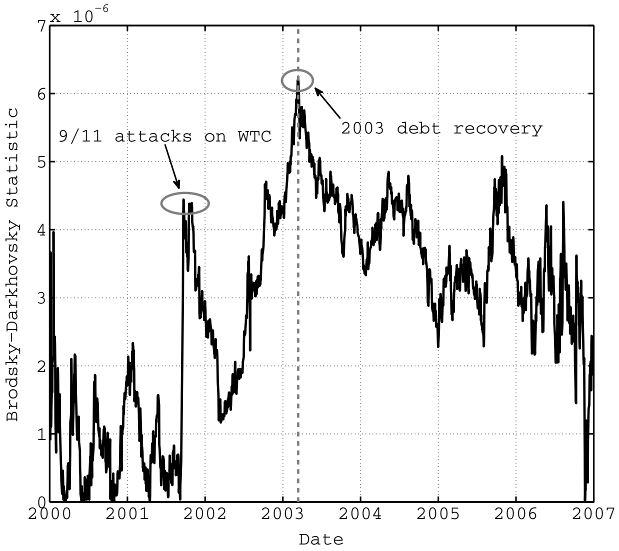


Figure 5: Behavior of the Brodsky–Darkhovsky statistic for the HST stock series.

To continue our analysis of the Brodsky–Darkhovsky statistic shown in Figure 5, the spike occurring in the second half of 2001 can be attributed to the 2001 HST stock crash caused by the 9/11 attacks in NYC. The specific value of the Brodsky–Darkhovsky estimate of this change-point is September 18, 2001, which is within the same week of the 9/11 attacks. This estimate can be refined using the following strategy. Once the absolute maximum of the Brodsky–Darkhovsky statistic is identified, the series is partitioned into two nonoverlapping segments: one composed of the observations up to the change-point and one consisting of the observations following the change-point. Then the Brodsky–Darkhovsky detection-estimation method is applied again individually to each of the two data chunks. It is argued in [2, 3] that this "divide and conquer" type of an approach also yields a strongly consistent (as the sample size gets infinitely large) estimator of the change-point.

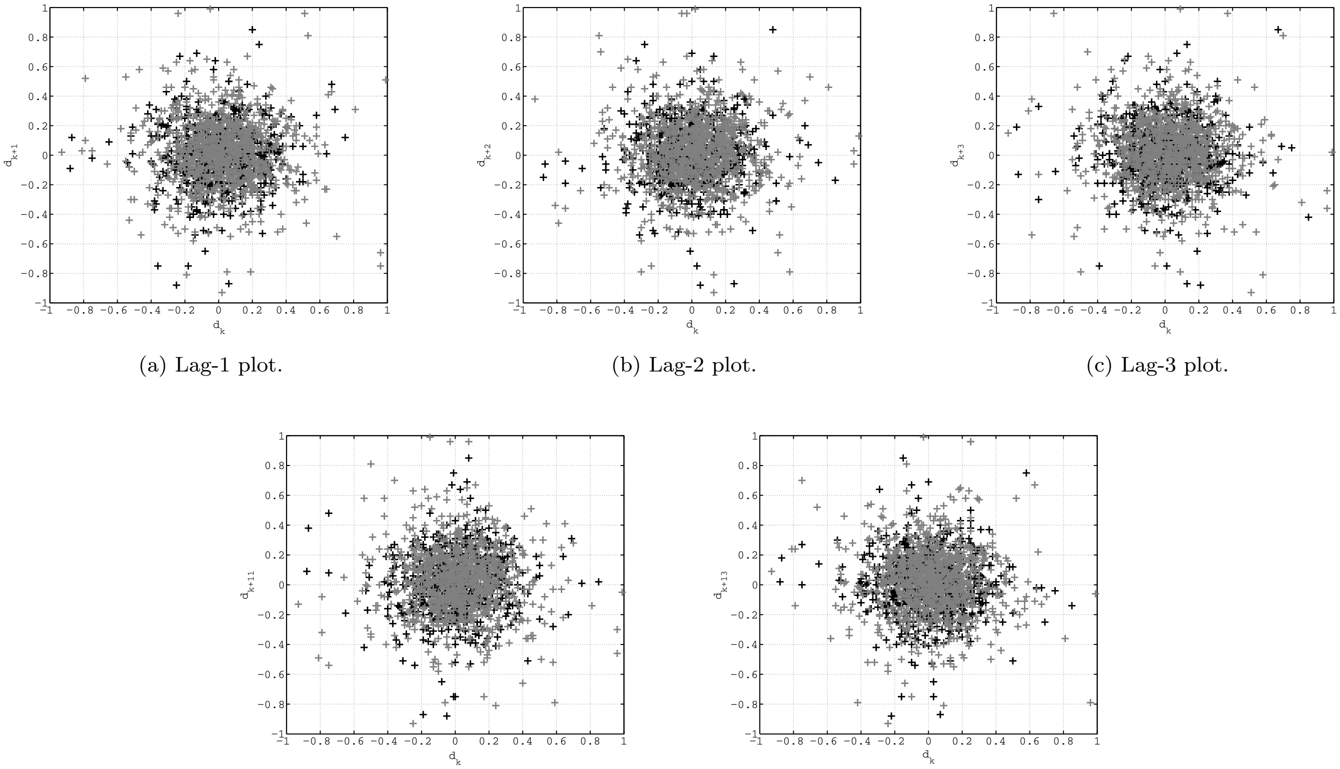
We now follow this strategy and analyze each piece of data separately. To that end, for the data segment to the left of the 2003 break-point the sample mean and standard deviation are $\hat{\mu}_\infty \approx -0.0029$ and $\hat{\sigma}_\infty \approx 0.2266$, respectively. The same sample characteristics for data segment to the right of the 2003 break-point turned out to be $\hat{\mu}_0 \approx 0.0199$ and $\hat{\sigma}_0 \approx 0.2306$. Therefore, the 2003 break-point changes not only the mean but also the variance. However, the change in the mean is far more contrast than the change in the variance. This could be part of the reason for the excellent performance of the Brodsky–Darkhovsky statistic (19).

Figures 6 show the empirical probability densities (histograms) for the returns before (see Figure 6a) and after (see Figure 6b) the 2003 event. Each of the two figures is also accompanied with a Gaussian fit with the mean and variance set to the respective estimated values. Since the two histograms are close to the Gaussian fits, there is only one conclusion to draw: the returns do behave as if they were generated by a Gaussian process.

The same conclusion can be drawn from an eye inspection of the corresponding Q-Q plots (quantile-quantile) shown in Figure 7. Specifically, the Q-Q plot for the distribution of the daily returns before the onset of the drift is shown in Figure 7a and the Q-Q plot for the distribution with the drift in effect is shown in Figure 7b. Since both plots use centered and scaled data, the fitted Gaussian distribution is the standard normal distribution. The fact that both plots are effectively a straight line is evidence of the "Gaussianness" of the return distribution before and after the drift.

Another important question to be examined about the time series at hand concerns the series' correlation structure. To that end, Figure 8 shows the correlation plot for the HST stock daily return series. Specifically, the plot distinguishes whether the data are before the 2003 event or after the 2003, and shows the autocorrelation function for the former piece in black and for the latter one in gray. It is clear from the plot that the data are essentially random throughout the entire set, as they exhibit no strong structure or correlation.

To reinforce the "no-correlation" conclusion arrived at from Figure 8, Figure 9 provides a selection of lag plots for the data, for lags equal to 1, 2, 3, 11, and 13. According to Figure 8, these are the lags at which the data correlation function may be considered statistically significant (with the level of significance being 95 %). To clear this out, the scatter plots shown in Figures 9 are to offer additional insight into the correlation structure of the time series under consideration. As in Figure 8 above, in Figure 9 the data are



(a) Lag-1 plot.

(b) Lag-2 plot.

(c) Lag-3 plot.

(d) Lag-11 plot.

(e) Lag-13 plot.

Figure 9: Lag plots (scatter plots) for the HST stock returns.

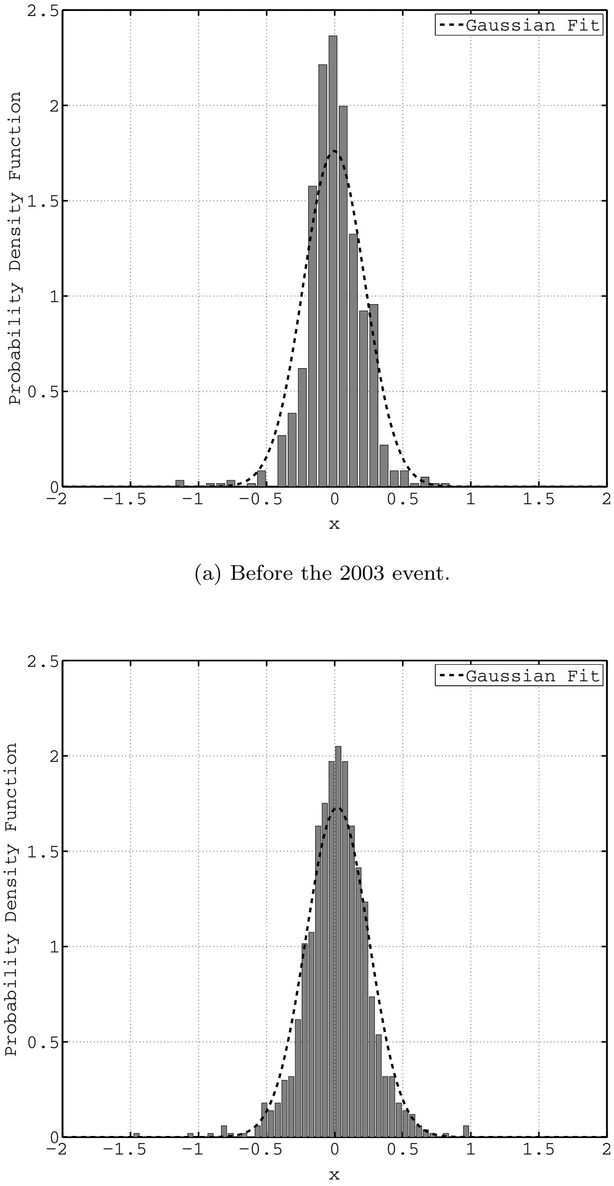
also split into two categories—before the 2003 event and after—and the two categories are distinguished using black color for the first category (before the 2003 event) and gray for the second one (after the 2003 event). The lack of any apparent patters in any of the five lag plots is an indication that the HST stock daily return series exhibits no temporal correlation.

## 4.4 Online structural break detection

We are now in a position to devise the change-point detection methodology of Section 3 to detect changes in the statistical pattern of the HST returns. To assess the performance of our detection methods, we will measure the detection delay relative to the change-point estimated by the Brodsky–Darkhovsky estimator (20) above. Recall that we are interested in comparing two score-based change-point detection procedures: the CUSUM chart given by (15)–(16) and the Shiryaev–Roberts (SR) procedure given by (13)–(14). Selecting the score function as in (17)–(18) for either procedure, we have implemented both detection methods in MATLAB, the well-known scientific computing platform developed by MathWorks, Inc. (see on the Web at http://www.mathworks.com). Since the above analysis of the data resulted in the conclusion that the data do follow a Gaussian model (before as well as after the change), to set up the detection threshold of the CUSUM chart and the SR procedure we assumed the Gaussian model with the parameters chosen as estimated in the above analysis. Via a simple Monte Carlo experiment we estimated that setting $A \approx 60$ and $h \approx 0.3$ ensures that the ARL to false alarm of either procedure is approximately 7 samples, which is roughly a week, since the timescale is working days.

The detection process is illustrated in Figure 10. Specifically, Figure 10a shows the behavior of the SR statistic in a short time window covering March 13, 2003, i.e., the date at which the HST stock underwent the change we would like to detect. Such a "zoomed-in" scale is to better illustrate the dynamics of the detection statistic around the change-point. Figure 10b shows the same but for the CUSUM statistic. We see that both procedures successfully detect the onset of the drift (occurring on March 13, 2003), and the detection delays are about one day each.

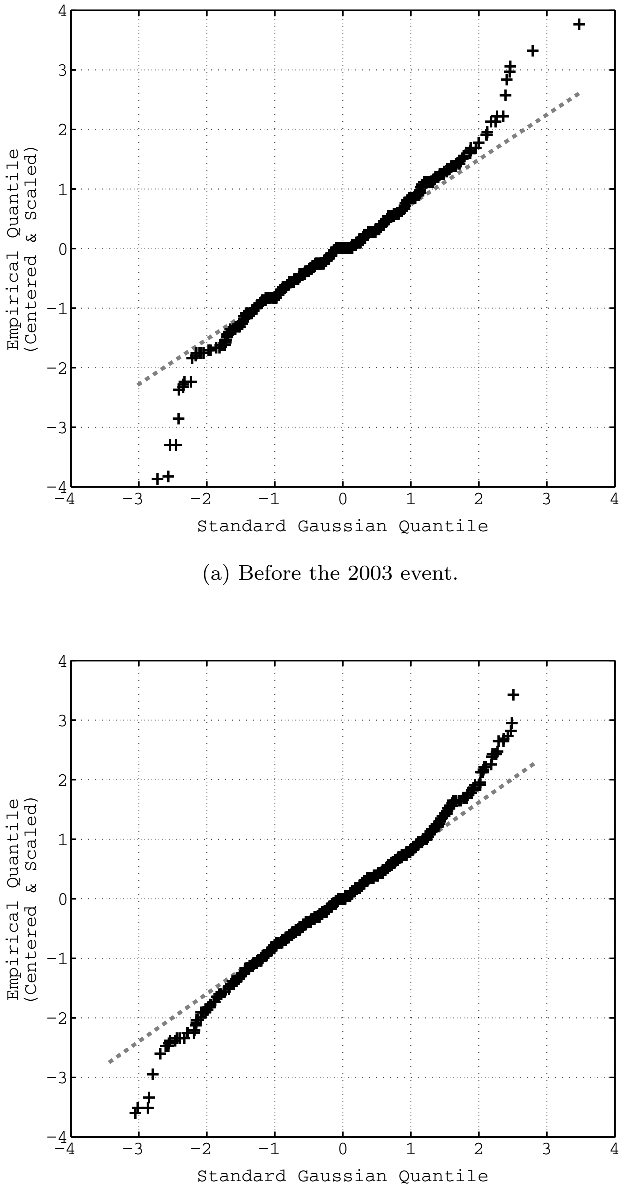
To draw a line under this section, we would like to remark that the dynamics of the CUSUM statistic is generally more informative than the dynamics of the SR statistic; compare, e.g., Figure 10b showing the CUSUM statistic and Figure 10a showing the corresponding SR statistic. Specifically, a mere eye examination of the behavior of the



(a) Before the 2003 event.

(b) After the 2003 event.

Figure 6: Empirical probability densities for the HST stock returns with Gaussian fits.



(a) Before the 2003 event.

(b) After the 2003 event.

Figure 7: Q-Q plots for the HST stock return distribution vs. the standard Gaussian distribution.

CUSUM statistic as a function of time allows not only to see whether the change has occurred or not, but to also estimate the time of its occurrence, i.e., the change-point: it is likely to be somewhere between the time instance the CUSUM statistic last hit zero and the point at which the statistic hit (or went above) the detection threshold (i.e., the point of alarm). Indeed, on the one hand, the change-point is unlikely to be past the point of alarm. On the other hand, as we discussed in the previous section, the CUSUM statistic is effectively a random walk with the "instantaneous" LLRs being the increments. Since the LLRs are, on average, negative if no change is in effect, and positive otherwise, the drift of the random walk the CUSUM chart uses for its decision-making is negative before the change and positive after. As a result, the CUSUM statistic effectively estimates zero in the pre-change regime, because zero is its reflection barrier: every time the CUSUM statistic hits zero it resets itself completely "forgetting" everything it had previously "learned" about the data. This equips the CUSUM chart with a built-in resetting mechanism: if after a suffi-

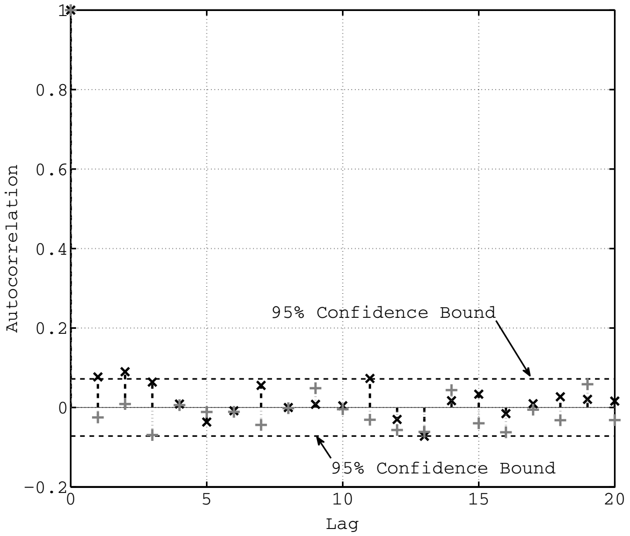
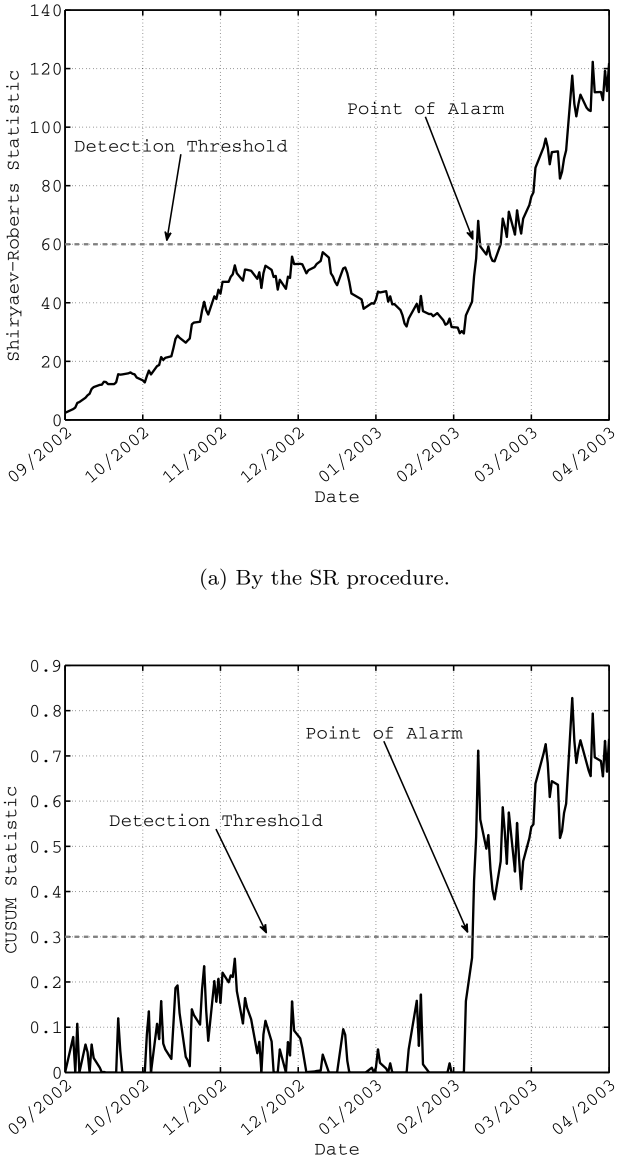


Figure 8: Autocorrelation function for the HST stock returns.

ciently long period of surveillance the data have given no indication of a change, the CUSUM statistic will likely reset itself (by hitting zero), i.e., it will discard the entire history of observations made up to that point and start completely anew. Hence, the change-point is unlikely to be to the left of the latest point at which the CUSUM statistic visited zero. This intrinsic self-restarting feature is the main reason for the exact minimax optimality (in the sense of Lorden [21]) of the CUSUM chart established in [24, 37]. By contrast, the SR statistic when plotted against time does not offer this kind of convenience of interpretation, for the SR procedure's decision-making mechanism uses entirely different principles. Nevertheless, the SR procedure is exactly multi-cyclic optimal, and the CUSUM chart is not. Therefore, when it comes to monitoring processes that are unlikely to undergo a structural break for a long period of time, so that change-point detection has to be performed in cycles, basing surveillance on the SR procedure might be a better option than going with the CUSUM chart.



(a) By the SR procedure.

(b) By the CUSUM chart.

Figure 10: Detection of the 2003 anomaly in the HST stock by the SR and CUSUM procedures.

## 5. CONCLUSION

We considered the problem of rapid but reliable anomaly detection in "live" financial data. We treated the problem statistically, viz. as that of quickest change-point detection, and proposed an anomaly-detection method that derives from the multi-cyclic (repeated) Shiryaev–Roberts (SR) detection procedure. We decided to go with this largely neglected near-coeval of the celebrated CUSUM and EWMA charts because of the strong multi-cyclic optimality properties that the SR procedure was recently discovered to have under the basic iid change-point detection setup; no such properties are exhibited by either the "good old" CUSUM "inspection scheme" or the EWMA chart. To handle real-world financial data, the proposed SR-derivative utilizes the information contained in the data in the SR-like Bayesian manner with the likelihood ratio replaced with a change-sensitive score function. This simple idea allowed the proposed procedure to preserve the low computational complexity of its prototype—the original SR procedure. More importantly, we carried out a case study where the proposed procedure was devised to detect an anomaly in a real-world financial time series, and the obtained experimental results indicated that our procedure may have also preserved the great "false alarm risk"-vs.-"detection speed" capabilities of the original SR procedure.

## ACKNOWLEDGEMENTS

The authors would like to thank the two Guest Editors, Prof. Shouyang Wang (Institute of Systems Science, Beijing, P.R. China) and Prof. Anatoly Zhigljavsky (Cardiff University, Cardiff, Wales, UK), and the Editor-in-Chief, Prof. Heping Zhang (Yale University, New Haven, Connecticut, USA), for the time and effort they invested to produce this special issue of the Journal. The authors are also personally thankful to Prof. Zhigljavsky for the invitation to contribute this work to the special issue. The constructive feedback provided by the two anonymous referees is greatly appreciated as well.

The effort of A.S. Polunchenko was supported, in part, by the Simons Foundation (www.simonsfoundation.org) via a Collaboration Grant in Mathematics (Award # 304574).

A.S. Polunchenko is also equally indebted to the Office of the Dean of the Harpur College of Arts and Sciences at the State University of New York at Binghamton for the support provided through the Dean's Research Semester Award for Junior Faculty granted for the Fall semester of 2014.

The work of A. Pepelyshev was partly supported by the St. Petersburg State University, Russia, under project # 6.38.435.2015.

*Received January 2015*

## REFERENCES

[1] Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs, NJ.

[2] Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Aplications* **243**. Kluwer Academic Publishers, Norwell, MA.

[3] Brodsky, B. E. and Darkhovsky, B. S. (2000). *Non-Parametric Statistical Diagnosis: Problems and Methods. Mathematics and Its Aplications* **509**. Kluwer Academic Publishers, Norwell, MA.

[4] Dragalin, V. P. (1994). Optimality of a generalized CUSUM procedure in quickest detection problem. In *Proceedings of the Steklov Institute of Mathematics: Statistics and Control of Random Processes*, **202** 107–120. American Mathematical Society, Providence, RI.

[5] Du, W., Polunchenko, A. S. and Sokolov, G. (2015). On Robustness of the Shiryaev–Roberts Change-Point Detection Procedure under Parameter Misspecification in the Post-Change Distribution. *Communications in Statistics—Simulation and Computation.* (in press). Available online at: http://www.tandfonline.com/doi/full/10.1080/03610918.2015.1039131.

[6] Ergashev, B. A. (2004). Sequential Detection of US Business Cycle Turning Points: Performances of Shiryayev–Roberts, CUSUM and EWMA Procedures. Available online at the Economics Working Paper Archive (EconWPA): https://ideas.repec.org/p/wpa/wuwpem/0402001.html.

[7] Frisén, M. (2008). *Financial Surveillance*. John Wiley & Sons, Inc., Hoboken, NJ.

[8] Frisén, M. (2009). Optimal Sequential Surveillance for Finance, Public Health, and Other Areas. *Sequential Analysis* **28** 310–337.

[9] Girschick, M. A. and Rubin, H. (1952). A Bayes approach to a quality control model. *Annals of Mathematical Statistics* **23** 114–125.

[10] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. A. (2001). *Analysis of Time Series Structure: SSA and related techniques. Monographs on Statistics and Applied Probability* **90**. Chapman & Hall/CRC, London, UK.

[11] Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series. Springer Briefs in Statistics.* Springer.

[12] Gordon, L. and Pollak, M. (1994). An Efficient Sequential Nonparametric Scheme for Detecting a Change of Distribution. *Annals of Statistics* **22** 763–804.

[13] Gordon, L. and Pollak, M. (1995). A Robust Surveillance Scheme for Stochastically Ordered Alternatives. *Annals of Statistics* **23** 1350–1375.

[14] Host Hotels & Resorts, Inc. (2001). US Securities and Exchange Commission, Form 10-K, Tax Year 2001.

[15] Host Hotels & Resorts, Inc. (2002). Annual Report 2002.

[16] Host Hotels & Resorts, Inc. (2003). Annual Report 2003.

[17] Host Hotels & Resorts, Inc. (2014). US Securities and Exchange Commission, Form 10-K, Tax Year 2014.

[18] Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems (with discussion). *Journal of the Royal Statistical Society. Series B. Methodological* **57** 613–658.

[19] Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory* **44** 2917–2929.

[20] Lai, T. L. and Xing, H. (2015). *Active Risk Management: Financial Models and Statistical Methods. Chapman and Hall/CRC Financial Mathematics Series.* Chapman & Hall/CRC Press, Boca Raton, FL.

[21] Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics* **42** 1897–1908.

[22] McDonald, D. (1990). A CUSUM Procedure Based on Sequential Ranks. *Journal of Naval Research* **37** 627–646.

[23] Moskvina, V. and Zhigljavsky, A. (2003). An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection. *Communications in Statistics—Simulation and Computation* **32** 319–352.

[24] Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Annals of Statistics* **14** 1379–1387.

[25] Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika* **41** 100–115.

[26] Pollak, M. (1985). Optimal detection of a change in distribution. *Annals of Statistics* **13** 206–227.

[27] Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Annals of Statistics* **15** 749–779.

[28] Pollak, M. (2010). A Robust Changepoint Detection Method. *Sequential Analysis* **29** 146–161.

[29] Pollak, M. and Krieger, A. M. (2013). Shewhart Revisited. *Sequential Analysis* **32** 230–242.

[30] Pollak, M. and Tartakovsky, A. G. (2008). Exact Optimality of the Shiryaev–Roberts Procedure for Detecting Changes in Distributions. In *Proceedings of the 2008 International Symposium on Information Theory and Its Applications* 1–6.

[31] Pollak, M. and Tartakovsky, A. G. (2009). Optimality Properties of the Shiryaev–Roberts procedure. *Statistica Sinica* **19** 1729–1739.

[32] Polunchenko, A. S., Sokolov, G. and Du, W. (2013). Quickest Change-Point Detection: A Bird's Eye View. In *Proceedings of the 2013 Joint Statistical Meetings.*

[33] Polunchenko, A. S. and Tartakovsky, A. G. (2010). On optimality of the Shiryaev–Roberts procedure for detecting a change in distribution. *Annals of Statistics* **38** 3445–3457.

[34] Polunchenko, A. S. and Tartakovsky, A. G. (2012). State-of-the-Art in Sequential Change-Point Detection. *Methodology and*

*Computing in Applied Probability* **14** 649–684.

[35] Polunchenko, A. S., Tartakovsky, A. G. and Mukhopadhyay, N. (2012). Nearly Optimal Change-Point Detection with an Application to Cybersecurity. *Sequential Analysis* **31** 409–435.

[36] Poor, H. V. and Hadjiliadis, O. (2009). *Quickest Detection*. Cambridge University Press, New York, NY.

[37] Ritov, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics* **18** 1464–1469.

[38] Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1** 239–250.

[39] Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430.

[40] Shewhart, W. A. (1925). The application of statistics as an aid in maintaining quality of a manufactured product. *Journal of the American Statistical Association* **20** 546–548.

[41] Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product. Bell Telephone Laboratories series*. D. Van Nostrand Company, Inc., Princeton, NJ.

[42] Shiryaev, A. N. (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Mathematics–Doklady* **2** 795–799. Translation from Dokl. Akad. Nauk SSSR 138:1039–1042, 1961.

[43] Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and Its Applications* **8** 22–46.

[44] Shiryaev, A. N. (1978). *Optimal Stopping Rules*. Springer-Verlag, New York, NY.

[45] Shiryaev, A. N. and Zryumov, P. Y. (2010). On the Linear and Nonlinear Generalized Bayesian Disorder Problem (Discrete Time Case). In *Optimality and Risk—Modern Trends in Mathematical Finance* (F. Delbaen, M. Rásonyi and C. Stricker, eds.) 227–236. Springer Berlin Heidelberg.

[46] Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals. Springer Series in Statistics*. Springer-Verlag, New York, NY.

[47] Tartakovsky, A., Nikiforov, I. and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability* **166**. CRC Press, Boca Raton, FL.

[48] Tartakovsky, A. G. (2005). Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. In *IEEE Conference on Decision and Control* **44** 320-325.

[49] Tartakovsky, A. G. and Moustakides, G. V. (2010). State-of-the-Art in Bayesian Changepoint Detection. *Sequential Analysis* **29** 125–145.

[50] Tartakovsky, A. G., Pollak, M. and Polunchenko, A. S. (2012). Third-order asymptotic optimality of the Generalized Shiryaev–Roberts changepoint detection procedures. *Theory of Probability and Its Applications* **56** 457–484.

[51] Tartakovsky, A. G. and Polunchenko, A. S. (2010). Minimax Optimality of the Shiryaev–Roberts Procedure. In *Proceedings of the 5th International Workshop on Applied Probability*.

[52] Tartakovsky, A. G., Polunchenko, A. S. and Sokolov, G. (2013). Efficient Computer Network Anomaly Detection by Changepoint Detection Methods. *IEEE Journal of Selected Topics in Signal Processing* **7** 4–11.

[53] Tartakovsky, A. G., Rozovskii, B. L., Blažek, R. B. and Kim, H. (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing* **54** 3372–3382.

[54] Tartakovsky, A. G., Rozovskii, B. L., Blažek, R. B. and Kim, H. (2006). Detection of intrusions in information systems by sequential changepoint methods (with discussion). *Statistical Methodology* **3** 252–340.

[55] Veeravalli, V. V. and Banerjee, T. (2013). Quickest Change Detection. In *Academic Press Library in Signal Processing: Array and Statistical Signal Processing*, (R. Chellappa and S. Theodoridis, eds.) **3** 209–256. Academic Press, Oxford, UK.

[56] Willsky, A. S. and Jones, H. L. (1976). A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control* **21** 108–112.

[57] Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[58] Zhigljavsky, A. (2009). Application of the Singular Spectrum Analysis for Change-point Detection in Time Series. In *Proceedings of the 2nd International Workshop in Sequential Methodologies*.

[59] Zhigljavsky, A. A. and Kraskovsky, A. E. (1988). *Detection of Abrupt Changes of Random Processes in Radiotechnics Problems*. St. Petersburg University Press, St. Petersburg, Russia. (in Russian).

Andrey Pepelyshev
Faculty of Mathematics
St. Petersburg State University
Peterhof, St. Petersburg 198504
Russia

School of Mathematics
Cardiff University
Cardiff, CF24 4AG
UK
E-mail address: andrey@ap7236.spb.edu

Aleksey S. Polunchenko
Department of Mathematical Sciences
State University of New York at Binghamton
Binghamton, New York 13902–6000
USA
E-mail address: aleksey@binghamton.edu
url: http://www.math.binghamton.edu/aleksey