THESIS TITLE - CHANGE POINT DETECTION

Tyler Manning-Dahan

A thesis
in
The Department
of
Engineering and Computer Science

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Applied Science
Concordia University
Montréal, Québec, Canada

August 2019

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: **Tyler Manning-Dahan**

Entitled: **Thesis Title - Change Point Detection**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

______________________________ Chair

______________________________ Examiner

______________________________ Examiner

______________________________ Examiner

______________________________ Supervisor

Approved ______________________________

Chair of Department or Graduate Program Director

__________ 20 _____ ________________________

Abstract

Thesis Title - Change Point Detection

Tyler Manning-Dahan

Text of abstract.

Acknowledgments

I would like to thank my supervisor Dr. Jia Yuan Yu for accepting me into his lab and pushing the limit of my intellectual understanding to places I have never thought possible.

I would also like to thank all my colleagues at DRW who have helped me learn about the finance industry over the past two years. I am especially grateful to Yves, Neil and Laura, who have been very patient with me and have me taught me a lot about understanding financial markets, writing clean code and building robust statistical models.

Lastly, I would like to thank my fiancée, Tanya, who initially pushed me to go back to school while I was still young and had the opportunity.

Contents

List of Figures **vi**

List of Tables **vii**

1 Introduction **1**

1.1 Motivation 1

1.1.1 Health Care 1

1.1.2 Financial Applications 2

1.2 Characteristics of the change point problem 2

1.3 Problem Formulation 4

1.3.1 Performance Measures 4

1.3.2 Other performance measures 6

1.4 Classic Algorithms 6

1.4.1 Shewart Control Chart 6

1.4.2 CUSUM 7

1.4.3 Extensions to CUSUM 7

1.4.4 Maximum Mean Discrepancy 8

1.5 Related Work 9

1.6 Our Contributions 11

2 Chapter 2 **12**

2.1 Chapter 2 Section 12

2.1.1 Chapter 2 Subsection 12

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The use of statistical control charts for detecting real-time changes in variation was pioneered by Walter Shewart in the first half of the twentieth century. Shewart was interested in reducing the unexpected causes of variation in the manufacturing processes that produced faulty manufacturing equipment [19]. Shewart's method involved charting the process measurements over time and detecting when a statistical process was no longer exhibiting an expected level of variation. Once this detection occurred, the process was stopped and was not restarted until the cause of the variation was fixed. Shewart's control charts were one of the first formal frameworks to solve the problem of detecting changes in a distribution of a sequence of random variables. This problem is now known more generally as the *change point detection problem*. Many industries make use of change-point techniques including healthcare monitoring systems, monitoring computer network traffic, and detecting regime shifts financial markets. The following are a few motivating examples.

1.1.1 Health Care

Health care is an important area for quickly detecting signal changes in heart rate monitoring [24] [20], epilepsy signal segmentation [15], and multi-modal MRI lesion detection [3] to name a few. Quickly detecting changes to a patient's health is absolutely necessary for any system to be of practical use. However, this quick detection must be balanced with high accuracy as false positives or missed detections could have life-threatening consequences.

1.1.2 Financial Applications

The application of accurate and timely change point detection is also very popular in the finance sector where shifts in asset prices can suddenly happen. Change point detection is particularly hard in financial applications because of the non-stationary data typically observed in asset price time series. Note, in the financial literature, change-points are also referred to as structural breaks, but for this thesis we will use the broader term change-points.

An on-line, quick detection technique is proposed in [18], where a modified Shiryaev - Roberts procedure is used in a case study to detect a change-point on a single stock's daily returns. They compare their non-parametric method with other classic control chart methods using speed of detection and false alarm rate as measures of performance.

Detecting changes in variance is specifically explored in [11]. The authors propose an off-line change-point algorithm that minimizes a global cost function by using an adaptive regularization function. The algorithm is applied to the absolute returns of the FTSE 100 stock index and the US dollar-Japanese Yen foreign intra-day exchange rate to detect changes in asset price volatility. The change-points identified in the FTSE 100 coincided with key market events such as the stock market crash that occurred on October 14^{th}, 1987 and breaking the 5000 price barrier in August 1997.

See section 1.3.6 of [21] for more applications to options markets and arbitrage opportunities.

1.2 Characteristics of the change point problem

A number of surveys of the literature already exist [1] [16], therefore we will not cover all existing methods but rather touch upon several, important factors to consider when tackling the change point detection problem. Across the body of literature, these factors determine what methods are available to practitioners.

The first factor is selecting between *parametric* and *non-parametric* techniques. Deciding between these two broad techniques is dependent on the prior knowledge one wants to encode into the problem. For example, if it is known that data is generated by a distribution from the exponential family of distributions, then we can subset the problem from the space of all possible distributions to a smaller space of distributions. Shewart control charts and CUSUM change-point techniques are both parametric techniques based on the Gaussian-family of distributions [17] [6]. In certain settings, it is not possible to leverage information about the data and non-parametric techniques must be used instead [4].

The second factor is deciding whether change-points should be detected *offline* or *online*. Some algorithms are off-line—also referred to as batch algorithms or retrospective or a posteriori change-point detection—and they are applied in an ex-post fashion after the dataset has been completely acquired [22]. If change-points must be detected as soon as possible, then waiting for the entire dataset to be acquired is not feasible and methods that operate on data streams as they arrive must be used. Such methods fall into the category of on-line change-point detection. The aforementioned Shewart control chart and CUSUM algorithm are both designed for data that is streamed in a real-time fashion. In the literature, on-line methods of change point detection are also referred to sequential change point detection [21]. For this thesis, we will use the terms interchangeably.

The third factor is determining if there are multiple change points or to assume there is only a single change point to detect. This is an important factor for off-line change point detection where the decision to detect one or more change-points is chosen at the outset. Detecting multiple change-points could also be relevant for the on-line case if a situation arises where the window of time series under consideration may contain more than one change point. However, most on-line change-point methods are designed to detect a single change-point at a time.

Finally, the last factor to address is determining exactly what kinds of statistical changes an algorithm should detect. Many methods focus solely on detecting changes in the mean of a distribution [12]. Some methods are more general and can detect changes in the variance or higher order moments and do not focus on any particular one. Methods like kernel change point detection can typically detect any distributional changes.

This thesis will concern itself with on-line change point detection, where data is received in a streaming nature. We assume no prior distributional characteristics on the data and operate in a completely non-parametric setting.

1.3 Problem Formulation

The basic change-point problem is set up as hypothesis test between two segments of a time series. Let $X_i, ..., X_n$ be a series of independent random variables of dimensions $d \geq 1$ be sequentially observed . Then, one of the following hypotheses holds:

$$\begin{cases} H_0 : X_1, X_2, ..., X_n \sim F_0 & \text{(no change-point occured)} \\ H_1 : X_1, X_2, X_{t_0-1} \sim F_0, X_{t_0}, X_{t_0+1}, \sim F_1 & \text{(a change-point occured).} \end{cases} \tag{1}$$

Where $i = 1, 2, ..., t_0 - 1$ and $j = t_0, ..., n$ are two distinct segments separated by change-point t_0 that is within the time series window. Furthermore, F_0, F_1 are cumulative distribution functions (CDFs) with corresponding probability density functions (PDFs), f_0, f_1. Because we are operating in a non-parametric setting, the CDFs are assumed to be completely unknown.

If there is no change in the data then we say the change time is equal to infinity and denote this probability as P^{∞} and the expectation is E^{∞}.

Many change-point detection algorithms define a statistic that is computed using each set before and after the possible change-point,t. If the statistic is above a particular threshold then time t is classified as a change-point, $\hat{\tau}$.

In the on-line scenario, the time series under consideration can be thought of as a sliding window with data constantly coming in and out of the window of interest. The size of the window is an important consideration that is typically chosen based on the problem being solved. Too small a window and the sets of data may not yield a statistically significant result. Too large of a window and the problem leans more towards and off-line model, where high volumes must be stored and several change-points may appear in a given window. If the amount of data is not a limitation then throttling the data may not be necessary.

1.3.1 Performance Measures

Because of the unsupervised nature of detecting change-points, it is difficult to evaluate the performance of change-point detection models with real world data. Many papers detail asymptotic or non-asymptotic theoretical guarantees of their proposed change-point methods. These theoretical results are typically compared across different change-point methods for benchmarking a new algorithm.

Two main issues arise when detecting change-points in a stream of data. The first is detecting a change-point when there is no actual statistical change in the observed sequence. These are typically called false positives or *false alarms* the in change-point detection literature. The false alarm rate is defined by a metric known as the *time to false alarm* (TTFA) rate.

$$TTFA = E_{\infty}[T] = E_{\theta}[N] \tag{2}$$

Where it is the expected number of observations that must be recorded before a change-point is incorrectly detected. In other words, it is the average amount of time until a change is detected given a sequence of observations with no change. Therefore, a larger value of TTFA is preferable. From a hypothesis testing perspective, this is equivalent to rejecting H_0 in cite equation in problem statement when it should not be rejected, i.e. type I error.

The second issue is not detecting a change-point when one occurs. This could be caused by detecting a change-point much too late for it to be of any use or simply missing it altogether. For quantifying this error, the worst case detection delay (WCD) metric measures how slow a model will detect a change-point in a worst case scenario. Conversely to TTFA, lower values of WCD are preferable.

$$WCD = supesssupE_t[(T-t)^+|F_{t-1}] \tag{3}$$

From a hypothesis testing perspective, this is equivalent to not rejecting H_0 in cite equation in problem statement when it should be rejected, i.e. type II error.

Balancing the TTFA and WCD of an on-line detection algorithm is crucial to for an algorithm to be of any practical use. In 1971 Lorden procedures to reacting..., it was shown asymptotically that the CUSUM algorithm provides an optimal trade-off between TTFA and WCD and ,in moustakides optimal stopping times for detecting.. 1986, it was proved optimal in the non-asymptotic case as well. Note, TTFA and WCD are also commonly referred to as ARL_0 and ARL_1 respectively where ARL stands for average run length. For clarity, we use the more explicit terms TTFA and WCD.

When detecting changes of a distribution, a user may want to quantify the size of the change in the mean by $|\mathbb{E}[X_\tau] - \mathbb{E}[X_{\tau+1}]|$ or, similarly, the size of the change in the variance by $|\text{Var}[X_\tau] - \text{Var}[X_{\tau+1}]|$.

1.3.2 Other performance measures

If labelled change-points are available for a real world dataset or a synthetic dataset, then the ground truth change-point vector, τ^*, is known. For example the *Hausdorff* metric can be used. It measures the furthest temporal distance between a predicted change-point $\hat{\tau}$ and τ^*. It is defined as:

Other standard classifier metrics can also used for comparing $\hat{\tau}$ and τ^*. This includes the F1-Score that is based on a classifier's precision and recall:

$$F_1(\hat{\tau}, \tau^*) = 2 * \frac{\text{precision*recall}}{\text{precision + recall}} \tag{4}$$

F1-Score is defined as the harmonic mean of precision and recall. Precision is defined as the ratio of true positives (TP) to the number of true positives (TP) and false positives (FP) and recall is defined as the ratio the number of true positives to the number of true positives plus the number of false negatives. F1-Score is best when F1 = 1 (perfect precision and recall) and reaches its worst value at F1 = 0. Depending on the context, any other classifier evaluation tools such as the Receiver Operating Characteristics Curve and the Precision Recall Curve may be used as well.

1.4 Classic Algorithms

Presented below are the fundamental approaches to on-line change-point detection that have been very influential.

1.4.1 Shewart Control Chart

Shewart control charts were originally designed to detect changes in the mean of a process where the values being observed are assumed to be Gaussian distributed. As the data arrives, the data is batched into samples of size N. The sample mean, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, is the then calculated and compared it to a known, true mean μ^*. If the absolute difference is greater than a threshold, then a change-point is declared at the current batch. Therefore, the decision rule is defined as,

$$|\bar{X} - \mu^*| > \kappa \frac{\sigma}{\sqrt{N}} \tag{5}$$

Where κ is a constant that controls how sensitive the algorithm is. Typically, it is set to $\kappa = 3$ as this need explanation. The true mean is assumed to known and is defined as $\mu^* = \mathbb{E}[X_i]$. In applications, the true mean can also be replaced by some target specification that a process must adhere to. Similarly, it is assumed the standard deviation, σ, is known in advance but it can also be estimated.

Tuning the hyper-parameters can drastically change the performance of the algorithm. Choosing a lower value for κ makes the control chart detect change-points more often, whereas a higher value results in less detections. The chosen sample size, N, is also critical and its effect on the performance of Shewart control charts was studied in [10].

1.4.2 CUSUM

Similar to the Shewart control chart, the CUSUM algorithm tracks a statistic over time relative to a predetermined threshold. CUSUM is best applied to a process that is already under control. It can be thought of accumulating the information of current and past samples.

The algorithm can be recursively defined by updating a statistic, S_i, after each X_i, such that:

$$\begin{cases} S_0 = 0 & \text{(Initialization)} \\ S_i = \max(0, S_{i-1} + Z_i) & \text{for i=1,2,...} \end{cases} \tag{6}$$

Where $Z_i = ln(\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)})$ and the statistic S_i is compared to a threshold h that is predetermined by the user. If $Z_i \geq h$ then a change-point is declared at time i and the algorithm is either completed or restarted. Given that the statistic only flags change-points when greater than a threshold, this algorithm only detects positive changes in the distribution. In [17], it is suggested to use two CUSUM algorithms to detect positive and negative changes in a distributional parameter.

Furthermore, it is assumed the distributions, f_0 and f_1, are known at the outset. In most applications, this is quite constraining and unrealistic. Therefore, in cases where parameters θ_0 and θ_1, maximum likelihood estimates of the parameters are usually computed.

1.4.3 Extensions to CUSUM

The filtered-derivative extension of the CUSUM introduced in X uses the change of the discrete derivative of a signal over time to detect a change-point.

In X, a fast initial response (FIR) CUSUM algorithm is proposed where the starting value of initial cumulative sums adapts over time. Instead of resetting S_0 to zero as shown above, it is add a bit more detail . This gives the algorithm a head-start in quickly detecting when a process is out of control and is especially useful for processes that don't start in control.

Finally, since CUSUM is typically better at detecting small shifts in signals and the Shewart control chart is faster at detecting larger changes, the two can be combined. The combined Shewart-CUSUM algorithm leverages the strengths of both techniques for better overall performance. See [14], [25], and [23] for more details.

1.4.4 Maximum Mean Discrepancy

Based on the kernel mean embedding , the maximum mean discrepancy (MMD) is a type of *integral probability metric* that can be used to compare two distributions [cite what IPM is from]. It is akin to the Kolmogorov distance between two distributions (which is the max norm of the difference between 2 cumulative distributions) or the Wassertstain distance.

The two-sample kernel test statistic is defined in [8] and uses the MMD as a distance measure for comparing two probability distributions. For example, suppose n samples $X = \{x_1, x_2, ..., x_n\}$ and and m samples from a different set $Y = \{y_1, y_2, ..., y_m\}$ are observed from a sample space $\mathcal{X}$. They are distributed as $X \sim \mathbb{P}$ and $Y \sim \mathbb{Q}$ respectively. The goal is to determine if in fact the two distributions are the same. The MMD is defined by a feature map $\phi : \mathcal{X} \to \mathcal{H}$

$$\mathrm{MMD}(P, Q) = ||\mathbb{E}_{X\sim P}[\phi(X)] - \mathbb{E}_{Y\sim Q}[\phi(Y)]||_{\mathcal{H}}$$

Where MMD can be understood as the distance in $\mathcal{H}$ between the kernel embeddings. Where the null hypothesis is that both samples stem from the same distribution, (i.e. $\mathbb{P} = \mathbb{Q}$) and the alternative hypothesis is that they are not drawn from the same distribution such that $\mathbb{P} \neq \mathbb{Q}$.

In [get citation], the unbiased, estimate of the squared MMD is shown to be:

$$\widehat{\mathbf{MMD}}_u^2(\mathcal{F}, X, Y) = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j=1}^{m} k\left(x_i, x_j\right) - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} k\left(x_i, y_j\right) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} k\left(y_i, y_j\right)$$

Which by law of large numbers converges to the theoretical values at a rate of XX. On its own, the MMD is not a metric but it can be one if the chosen kernel,k, is a *characteristic kernel.* That is, the kernel mean embedding is injective in the mapping into the reproducing kernel Hilbert space. This property is critical as it ensures MMD=0 if and only if $\mathbb{P} = \mathbb{Q}$. As [cite KME review] points out, there is no loss if information in this case.

For instance given the radial basis function kernel, $k(x, y) = e^{-\frac{1}{2\sigma^2}||x-y||^2}$.

REWORD: We call the function that achieves the supremum, the witness function because it is the function that witnesses the difference in the two distributions. This means that we can interpret the witness function as showing where the estimated densities of p and q are most different.

The witness function

$$f(x) = \mathbb{E}_{x' \sim p}\left[k\left(x, x'\right)\right] - \mathbb{E}_{x' \sim q}\left[k\left(x, x'\right)\right] \tag{7}$$

which can also be estimated from finite samples of data by:

$$\hat{f}(x) = \frac{1}{m}\sum_{i=1}^{m} k\left(x, x_i\right) - \frac{1}{n}\sum_{i=1}^{n} k\left(x, y_i\right) \tag{8}$$

Thus, as [need citation] points at, the witness function tracks where the densities of X and Y are most different.

1.5 Related Work

In 2005, Desobry et al. [7] developed an on-line kernel change point detection model based on single class support vector machines (ν-SVMs). The authors train a single class support vector on a past set, $\mathbf{x}_{t,1} = x_{t-m_1}, ..., x_{t-1}$ of size m_1 and train another single class support vector on a future set $\mathbf{x}_{t,2} = x_t, ..., x_{t+m_2-1}$ of size m_2. A ratio is then computed between the two sets that acts as the dissimilarity measure in Hilbert space. If the points are sufficiently dissimilar over some predetermined threshold,η, then a change point is assigned to the time spitting the two sets of data. Desobry argues that a dissimilarity measure between kernel projection of points in a Hilbert space should estimate the *density supports* rather than estimate the probability distributions of each set of points.

In 2007, Harchoui and Cappe [9] approached the off-line change point problem with a fixed number of change points by using kernel change point detection. This was further extended to an unknown number of change points in 2012 by Arlot et al. [2]. Finally, Garreau and Arlot extended this in line of research kernel change points in the off-line setting of detecting change points. Fundamentally, their method is the kernel version of the following least squares optimization problem:

$$J(\tau, \mathbf{y}) = \frac{1}{n} \sum_{k=1}^{K} (\tau) \sum (Y_i - \hat{(Y_k)})^2 + \beta \text{pen}(\tau) \tag{9}$$

The benefits of this off-line kernel change point detection is that it operates on any kind of data for which a kernel that properly reproduces a Hilbert space can be applied. For example, it can be applied to image data, histogram data, as well as $d-$dimensional vectors in $\mathbb{R}^d$. Garreau shows their KCP procedure outputs an off-line segmentation near optimal with high probability. Lastly, the authors recommend choosing the kernel based on best possible signal to noise ratio that the distribution gives based on Δ^2/M^2 . Therefore, some prior knowledge or training set is necessary for calibrating the kernel.

In the on-line setting, several methods use kernel embeddings with a two-sample hypothesis test. This is done in a similar vein to the classic CUSUM and Shewart control charts. They all make use of the maximum mean discrepancy (MMD) test statistic for a two-sample kernel hypothesis test.

In [13], the authors use the MMD hypothesis test by using a windowed approach where a fixed size of past data is compared with a fixed size of new data. They define a B-test statistic for two-sample testing for rejecting null hypothesis of first equation. The B-test statistic is a recently developed alternative to the MMD that is more efficient; it involves taking an average of the MMD over a partitioning of the data into N blocks.

More recently in [5], a kernel change-point detection method is proposed that uses deep generative models to augment the test power of the kernel two sample test statistic. The method is compared to other prominent change-point methods for on-line and off-line detection of change detection. All comparisons done on synthetic data are with piece-wise i.i.d. data. All methods are benchmarked using the AUC metric for classification performance and it is shown the KL-CPD method is competitive or better than state of the art methods such as find out which ones. However, for on-line change-point tests, the authors did not use time to detection as a metric for comparison. It would be interesting to see how this trade-off compares to competing on-line methods.

Finally, in a recent, preprint paper, a kernel CUSUM (KCUSUM) algorithm is proposed, where the classic CUSUM algorithm is adapted using the MMD statistic for on-line detection. The authors use a modified, unbiased MMD statistic that can be computed in linear time. This formulation of the MMD statistic was originally defined in section 6 of [8] as:

$$\mathrm{MMD}_l^2[\mathcal{F}, X, Y] := \frac{1}{m_2} \sum_{i=1}^{m_2} h\left(\left(x_{2i-1}, y_{2i-1}\right), \left(x_{2i}, y_{2i}\right)\right)$$

While this non-parametric approach can detect any change in the distribution of a sequence, it does struggle with more complicated distributional changes such as higher moments changes in a single dimension.

1.6 Our Contributions

Lay out the contributions of this work.

Chapter 2

Chapter 2

2.1 Chapter 2 Section

2.1.1 Chapter 2 Subsection

Subsection Test

Bibliography

[1] Samaneh Aminikhanghahi and Diane J Cook. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017.

[2] Sylvain Arlot, Alain Celisse, and Zaıd Harchaoui. Kernel change-point detection. *arXiv preprint arXiv:1202.3878*, 6, 2012.

[3] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial mri: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, 2003.

[4] E Brodsky and Boris S Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.

[5] Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. Kernel change-point detection with auxiliary deep generative models. *arXiv preprint arXiv:1901.06077*, 2019.

[6] Jie Chen and Arjun K Gupta. *Parametric statistical change point analysis: with applications to genetics, medicine, and finance.* Springer Science & Business Media, 2011.

[7] Frédéric Desobry, Manuel Davy, and Christian Doncarli. An online kernel change detection algorithm. *IEEE Trans. Signal Processing*, 53(8-2):2961–2974, 2005.

[8] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.

[9] Zaid Harchaoui and Olivier Cappé. Retrospective mutiple change-point estimation with kernels. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 768–772. IEEE, 2007.

[10] Salah Haridy, Ahmed Maged, Saleh Kaytbay, and Sherif Araby. Effect of sample size on the performance of shewhart control charts. *The International Journal of Advanced Manufacturing Technology*, 90(1-4):1177–1185, 2017.

[11] Marc Lavielle and Gilles Teyssiere. Adaptive detection of multiple change-points in asset price volatility. In *Long memory in economics*, pages 129–156. Springer, 2007.

[12] Tze-San Lee. Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice*, 4(4):643–662, 2010.

[13] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. M-statistic for kernel change-point detection. In *Advances in Neural Information Processing Systems*, pages 3366–3374, 2015.

[14] James M Lucas. Combined shewhart-cusum quality control schemes. *Journal of Quality Technology*, 14(2):51–59, 1982.

[15] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. Online bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1833–1837. IEEE, 2013.

[16] Yue S Niu, Ning Hao, Heping Zhang, et al. Multiple change-point detection: A selective overview. *Statistical Science*, 31(4):611–623, 2016.

[17] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[18] Andrey Pepelyshev and Aleksey S Polunchenko. Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570*, 2015.

[19] Walter Andrew Shewhart. *Economic control of quality of manufactured product.* ASQ Quality Press, 1931.

[20] M Staudacher, S Telser, A Amann, H Hinterhuber, and M Ritsch-Marte. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):582–596, 2005.

[21] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC, 2014.

[22] Charles Truong, Laurent Oudre, and Nicolas Vayatis. A review of change point detection methods. *arXiv preprint arXiv:1801.00718*, 2018.

[23] JO Westgard, T Groth, T Aronsson, and CH De Verdier. Combined shewhart-cusum control chart for improved quality control in clinical chemistry. *Clinical Chemistry*, 23(10):1881–1887, 1977.

[24] Ping Yang, Guy Dumont, and John Mark Ansermino. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on biomedical engineering*, 53(11):2211–2219, 2006.

[25] Emmanuel Yashchin. On the analysis and design of cusum-shewhart control schemes. *IBM Journal of Research and Development*, 29(4):377–391, 1985.