sciendo
Financial Internet Quarterly 2021, vol. 17 / no. 4, p. 27-33
10.2478/fiqf-2021-0025

# BAYESIAN ONLINE CHANGE POINT DETECTION IN FINANCE

Reza Habibi[1]

**Abstract**

It is quite common that the structure of a time series changes abruptly. Identifying these change points and describing the model structure in the segments between these change points is an important task in financial time series analysis. Change point detection is the identification of abrupt changes in the generative parameters of sequential data. In application areas such as finance, online rather than offline detection of change points in time series is mostly required, due to their use in predictive tasks, possibly embedded in automatic trading systems. However, the complex structure of the data generation processes makes this a challenging endeavor. This paper is concerned with online change point detection in financial time series using the Bayesian setting. To this end, the Bayesian posterior probability of change at a specific time is proposed and some procedures are presented for selecting the priors and estimation of parameters. Applications in simulated financial time series are given. Finally, conclusions are proposed.

**JEL classification**: G17
**Keywords**: Bayesian setting, change point, financial time series, online detection, probability of change

Received: 11.10.2021 Accepted: 01.12.2021

**Cite this:**
Habibi R. (2021) Bayesian online change point detection in finance. Financial Internet Quarterly 17 (4), pp. 27-33.



[1] Iran Banking Institute, Central Bank of Iran, Tehran, Iran, e-mail: habibi1356@gmail.com, ORCID: 0000-0001-8268-0326.

## Introduction

The problem of change-point estimation has attracted significant attention. A branch of the literature deals with the estimation of a single change-point (for a change in mean see e.g. Gombay (2008) and Gombay and Serban (2009) and references therein) while another extends it to multiple change-points with many changing parameters such as Ombao et al., (2001) who divide a time series into dyadic segments and choose the one with the minimum cost. The latter branch can be further categorized. On the one hand, the multiple change-point estimation can be formulated through an optimization task i.e. minimizing a multivariate cost function (or criterion). When the number of change-points N is unknown then a penalty is typically added e.g. the Schwarz criterion. In addition, the user can adopt certain cost functions to deal with the estimation of specific models: the least-squares for change in the mean of a series (Lavielle & Moulines, 2000), the minimum description length criterion for non-stationary time series (for example, Davis et al., 2006), the Gaussian log-likelihood function for changes in the volatility or the covariance structure of a multivariate time series, see Lavielle and Teyssiere (2006). For a comprehensive review in change point analysis in financial time series and online methods, see Xiao et al., (2018). Change point analysis in financial time series is also critical because of their importance in predictive applications. There are two different frameworks for change detection in financial time series, i.e., offline and online frameworks.

The Bayesian procedure is an influential tool for online making of statistical inferences, see Adams and MacKay (2007). The process of Bayesian online change point detection proposed by Adam and MacKay is in essence a filtering process on an infinite state hidden Markov model, in which the observed time series can be split into a set of connected segments, each segment is generated by a hidden model, called the observation model (there are infinitely many possible ways of segmentation thus infinitely many possible observation models). A "change point" is defined as the beginning time index of a new segment. Duration is defined as the length of a segment; duration is generated from a model called the duration model.

In the current paper, using the Bayesian setting and following Koop and Potter (2004), change point detection is considered in finance. To this end, let $X_i$, $i \geq 1$ denote the return of a financial asset at time $i \geq 1$. Throughout the paper, it is assumed that these variables are mutually independent and having distribution with density function $f_{\Theta i}$. Here, $\Theta_1$ is a time varying parameter e.g., the mean (level) or volatility of price. Indeed, at time $i$, parameter $\Theta_1$ is changed to new level $\Theta_1^*$ with probability of $p_i$ and remains unchanged with probability of $1 - p_1$. Let $J_i$ be a dummy variable which is zero if there is no change in $\Theta_i - 1$ with probability of $1 - p_i$ and $J_i = 1$ if a changed has occurred at $\Theta_i - 1$ with probability of $p_1$. Hence,

$$\Theta_i = (1 - J_i)\, \Theta_{i-1} + J_i \Theta_1^*,$$

where $\delta_i = \Theta_1^* - \Theta_{i-1}$ is the magnitude of change and notice that $\Theta_i = \Theta_{i-1} + J_i \delta_i$ which constitutes a random walk structure.

The above problem may also be considered as jump detection in price of a financial asset. However, a special case of the jump detection problem is the change point analysis, see Saatci et al., (2010). This type of change representation has been used by Habibi et al., (2017). Types of change point models such as at most one change point model (AMOC) at $i = k_0$ and multiple change points at $i = k_l$, $i \geq 1$ are special cases of this model by letting $J_i = 1$ for $i = k_0$ and zero otherwise for the AMOC case and $J_i = 1$ for and $i = k_l$, $i \geq 1$ zero otherwise for the multiple change point problem.

The rest of the paper is organized as follows. In the next section, first, the probability of change is proposed. Then, estimation of unknown parameters of probabilities are discussed. Applications of the proposed method in finance via simulated examples are given in section 3.

## Online change detection

In this section, the Bayesian online change point detection is proposed. Indeed, observations $X_j$, $1 \leq j \leq i$ are observed and it is interesting to know if $J_i = 1$ or $J_i = 0$? To this end, first, the Bayesian probability of change is proposed. Then, the estimation of unknown parameters of probability of change is discussed.

## Probability of change

Here, the posterior probability of change $\pi_i$ defined by

$$\pi_i = P(J_i = 1 | X_j,\ 1 \leq j \leq i)$$

as a leading indicator, is computed. This measure may be considered an early warning tool for alarming possible future changes. Notice that

$$\pi_i = \frac{p_i f_{\theta_i^*}(x_i)}{p_i f_{\theta_i^*}(x_i) + (1-p_i) f_{\theta_{i-1}}(x_i)}, \theta_i^* \neq \theta_{i-1}.$$

One can see that

$$\frac{\pi_i}{1-\pi_i} = \frac{p_i}{1-p_i}\Lambda_i, \quad \Lambda_i = \frac{f_{\theta_i^*}(x_i)}{f_{\theta_{i-1}}(x_i)}$$

where $\Lambda_i$ is the likelihood ratio measure. Then, it is seen that the logit function of $\pi_i(\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$ is given by

$$\text{logit}(\pi_i) = \text{logit}(p_i) + \log(\Lambda_i).$$

The second part of the above decomposition is $u_i = log(\Lambda_i)$ which shows the existence of change point at i-th time point as soon as $u_i$ is larger than a critical threshold. In an offline setting, the cumulative sum process of, $u_i$ i.e., $s_j = \sum_{i=1}^{j}(u_i - \bar{u}_n)$, $1 \leq j \leq n$, where $\bar{u}_n = \frac{\sum_{j=1}^{n} u_j}{n}$ is an inverted V-shaped curve which is maximized around the actual change point. However, the first term of the above-mentioned decomposition is the expert opinion regarding the existence of a change point at *i*-th time point. Indeed, $\pi_i$ is a tradeoff between real data (likelihood ratio) and the expert opinion $p_i$. As soon as, $\pi_i$ is larger than a threshold *c*, it is doubtful whether there is a change at *i*-th time point. Two main parameters in detection of change points, accurately, are the sequence of probabilities $p_i$ and threshold *c*. As follows, two procedures are proposed to this end.

*Procedure 1*. To select $p_i$'s, it is enough to have a function which is maximized at actual change points. A fast answer to this question is to assume that $logit(p_i) = \alpha\Lambda_i$, for some positive $\alpha$'s. Indeed, coefficient α is selected by expert opinion which indicates his/her belief about information that the likelihood ratio has about the location of change point. Therefore,

$$p_i = \frac{\Lambda_i^{\alpha}}{1+\Lambda_i^{\alpha}}, i \geq 1, \text{and } \text{logit}(\pi_i) = (1+\alpha)\log(\Lambda_i).$$

By this formulation, $\pi_i > c$ implies that

$$c_i = \frac{\Lambda_i^{\alpha+1}}{1+\Lambda_i^{\alpha+1}} > c.$$

Let $c_{max} := max(c_i) = c$. Hence, a Monte Carlo simulation may be applied to find distribution of $c_{max}$ and propose reasonable c's as some indicators of distribution of $c_{max}$.

*Procedure 2*. For another formulation of $p_i$, assume that $J_i$ is one if $log(\Lambda_i) > d_i$ for some pre-determined threshold $d_i$'s and zero otherwise. Indeed, $log(\Lambda_i) > d_i$ means there is a change at *i*-th time point. Here, $p_i$ is computed by $p_i = P(log(\Lambda_i) > d_i)$. The following proposition summarizes the above discussion.

Proposition 1. (1) - (4) are correct.

(1) The posterior probabilities $\pi_i$'s of having change point at *i*-th time point are given in the logit function as follows

$$log(\pi_i) = log(p_i) + log(\Lambda_i).$$

(2) Let $\alpha$ be the prior belief degree to likelihood ratio $\Lambda_i$. Then, the empirical prior probabilities $p_i$ is given by $p_i = \frac{\Lambda_i^{\alpha}}{1+\Lambda_i^{\alpha}}$, where $\Lambda_i$ is the likelihood ratio.

(3) Hence, $\pi_i = \frac{\Lambda_i^{\alpha+1}}{1+\Lambda_i^{\alpha+1}}, i \geq 1$ and $c = c_{max} = \max(c_i)$, $c_i = c_i = \frac{\Lambda_i^{\alpha+1}}{1+\Lambda_i^{\alpha+1}}$ As soon as, $\pi_i > c$, it is concluded that there is a change at *i*-th time point.

(4) For another formulation of $p_i$, let $p_i = P(log(\Lambda_i) > d_i)$ and $J_i$ is one if $log(\Lambda_i) > d_i$ and zero otherwise.

## Parameter estimation

In the previous section, Bayesian online change point detection is studied. However, the mentioned procedures contain some unknown parameters which should be estimated, in practice. In this section, the mentioned procedures are reviewed using estimated parameters and their advantages and disadvantages, and sensitivity and robustness analyses are studied.

To this end, assume that in practice, $\Theta_i$ is unknown and that $\hat{\theta}_i$ is an estimate of $\Theta_i$, e.g., maximum likelihood, least square or Bayesian estimate, based on observations $x_j$, $1 \leq j \leq i$. Assume that there exists a stochastic representation for $\hat{\theta}_i$ as follows

$$\hat{\theta}_i = (1-\lambda_i)\hat{\theta}_{i-1} + \lambda_i g(X_i) = \hat{\theta}_{i-1} + \lambda_i \hat{\delta}_i,$$

where $\hat{\delta}_i = g(X_i) - \hat{\theta}_{i-1}$, for some functions *g*'s and forgetting factor $\lambda_i \in (0,1)$. For example, when $\Theta$ plays the role of population mean and it is estimated by sample mean $\widehat{\theta}_i = \bar{x}_i$, then $\lambda_i = 1/i$. Usually, it is assumed that $\sum \lambda_i = \infty$ and $\sum \lambda_i^2 < \infty$ for making sure of the convergence issues.

When there is no change point up to *i*-th time point, then $\Lambda_i$ is close to one and it is estimated by

$$\widehat{\Lambda}_i = \frac{f_{\widehat{\theta}_i}(x_i)}{f_{\widehat{\theta}_{i-1}}(x_i)}.$$

It is easy to see that, as $\Lambda_i \to 0$, by the first order Taylor approximation, $\log(\widehat{\Lambda}_i)$ is approximated by

$$\log(\widehat{\Lambda}_i) \approx \lambda_i \delta_i \frac{\partial}{\partial \widehat{\theta}_{i-1}} \log\left(f_{\widehat{\theta}_{i-1}}(x_i)\right).$$

However, if there is a change at *i*-th time point, indeed, when $J_i$ is one since $(\widehat{\Lambda}_i) > d_i$, then $\Lambda_i$ is estimated by

$$\widehat{\Lambda}_i = \frac{f_{g(x_i)}(x_i)}{f_{\widehat{\theta}_{i-1}}(x_i)}.$$

Let $J_i$ is one if $\log(\widehat{\Lambda}_i) > d_i$ and is zero, otherwise. Here, $\widehat{\theta}_i = \widehat{\theta}_{i-1} + J_i g(X_i)$ constitutes an estimated version of the random walk process.

To compute $d_i$, it is necessary to know the null distribution of $\log(\widehat{\Lambda}_i)$. As an example, first, suppose that $X_i$ come from normal distribution with mean $\Theta_i$ and common variance $\sigma^2$. Then, $\Lambda_i = 1/i$ and $\widehat{\theta}_i = \bar{x}_i$. Then,

$$\log(\widehat{\Lambda}_i) = \lambda_i (1 - 0.5\lambda_i) (\frac{x_i - \theta_{i-1}}{\sigma})^2.$$

Also, $P(\lambda_i(1-0.5\lambda_i)\left(\frac{x_i - \widehat{\theta}_{i-1}}{\sigma})^2 < d_i\right) = 1 - \alpha$ implies that Then, $d_i = F_i^{-1}$ where $F_i$ is normal distribution with mean $\frac{\widehat{\theta}_i - \widehat{\theta}_{i-1}}{\sigma}$ and variance $\lambda_i(1-0.5\lambda_i)$. Also, $\hat{p}_i = \Phi(d_i^2) - \Phi(d_i^1)$, $d_i^2 = \frac{d_i}{\sqrt{\lambda_i(1-0.5\lambda_i)}} - \frac{\theta_i - \theta_{i-1}}{\sigma}, d_i^1 = d_i^2 - \frac{2d_i}{\sqrt{\lambda_i(1-0.5\lambda_i)}}$. The next proposition summarizes the above discussion.

Proposition 2. (1) - (2) are correct.

(1) Indicatoris $J_i$ a binary variable indicating the existence of change point at *i*-th time point with success probability $p_i = P\left(\log(\widehat{\Lambda}_i) > d_i\right)$, for some predetermined threshold $d_i$'s. Here, $\widehat{\theta}_i = \widehat{\theta}_{i-1} + J_i g(X_i)$ constitutes an estimated version of the random walk process.

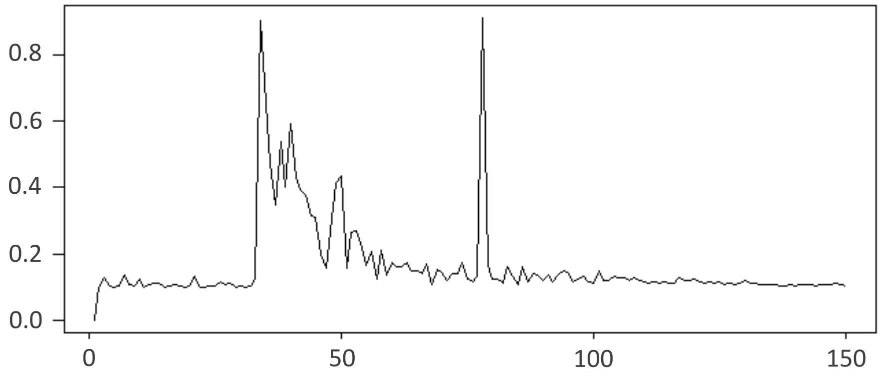
(2) Under the null hypothesis on no change point up to time point I, then the likelihood ratio $\Lambda_i$ is estimated by $\widehat{\Lambda}_i = \frac{f_{\widehat{\theta}_i}(x_i)}{f_{\widehat{\theta}_{i-1}}(x_i)}$ and approximated by $\lambda_i \delta_i \frac{\partial}{\partial \widehat{\theta}_{i-1}} \log\left(f_{\widehat{\theta}_{i-1}}(x_i)\right)$, where $\widehat{\theta}_i = (1-\lambda_i)\widehat{\theta}_{i-1} + \lambda_i g(X_i)$ where is stochastic approximation representation of $\widehat{\theta}_i$ for some functions *g*'s and forgetting factor $\Lambda_i \in (0,1)$. When, $J_i = 1$, then $\Lambda_i$ is estimated by $\widehat{\Lambda}_i = \frac{f_{g(x_i)}(x_i)}{f_{\widehat{\theta}_{i-1}}(x_i)}$.

## Financial cases

Here, some illustrative simulation cases with financial applications are given.

*Case 1*. Let $n = 150$, and $X_i$'s (return process) come from Bernoulli distribution such that at $i = 86$ the probability of success is changed from 0.1 to 0.45. It is assumed that $p_i = 0.1$ for $i \neq 86$ and it is 0.9 for 86-th observation. Let $c = 0.7$, then the first point at which $\pi_i > 0.7$ is the actual change point 86. For another example, let $n = 200$, and $X_i$'s (return process) come from Poisson distribution such that at $i = 64$ the intensity parameter is changed from 1 to 2. It is assumed that $p_i = 0.1$ for $I \neq 64$ and it is 0.9 for 64-th observation. Let $c = 0.7$, then the first point at which $\pi_i > 0.7$ is the actual change point 64. The previous cases were at most one change point (AMOC) model. Here, for a multiple change point case, let the mean of normally distributed random variables change from 1 to 2 and return to 1 at $k_0 = 34$, $k_1 = 78$ and the standard deviation is constant and equals to 0.1. Then, the time series plot of $\pi_i$ implies that there are two changes, visually. Again, the prior probability of having changes at actual change points are given as 0.9 and in other time points, this probability is given as 0.1.

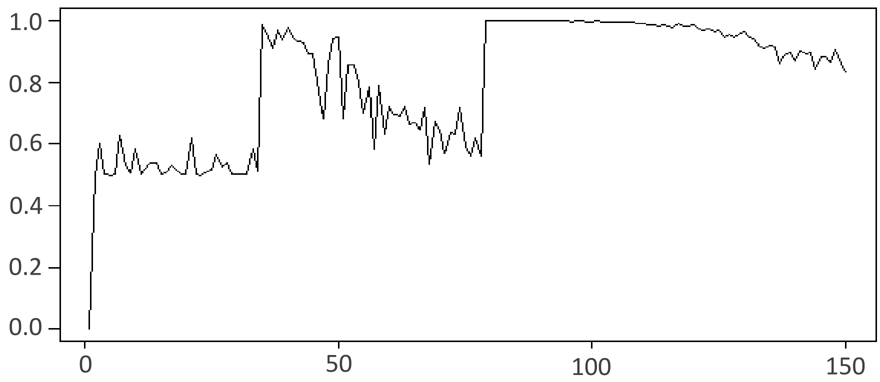
**Figure 1: Time series plot of $\pi_i$**




*Source: Own elaboration.*

Here, again, the above example 2 is considered but $p_i$'s are changed. It is assumed that $\alpha = 0.75$. Then, time series plot of $\pi_i$ is plotted as follows indicating two change points.

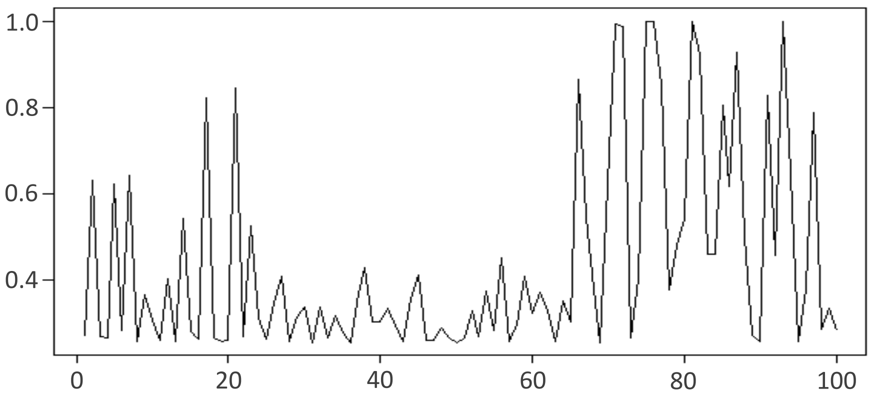
**Figure 2: Time series plot of $\pi_i$**




*Source: Own elaboration.*

*Case 2*. Volatility of a financial asset is an indicator for variability of return process of price of a given asset. Let $s_t$ be the price of a financial asset such as a stock at time $t$ and $r_t = \log(s_t) - \log(s_{t-1})$ be the logarithmic return process. Suppose that $\mu$ is the overall mean of $r_t$ and $\sigma_t^2$ is the variance of return series up to time $t$. Hence, the volatility $\sigma_t^2$ is estimated by $\hat{\sigma}_i^2 = \frac{1}{i}\sum_{j=1}^{i} x_j$, where $x_j = (r_j - \mu)^2$. Indeed, in stochastic approximation representation, then, $\lambda_i = \frac{1}{i}, g(x_j) = x_j$. Notice that $\frac{x_j}{\theta_j}$ has a central chi-squared distribution with one degree of freedom. Then, $x_j$ has scale distribution with scale parameter $\Theta_j = \sigma_j^2$. Let $n = 100$, $k_0 = 65$, $\mu = 0.001$. Assume that $\Theta_j = 0.001$ before the change and $\Theta_j = 0.007$ after the change. Plot of $\pi_i$ are given in Figure 3 indicates there is a change at 65 time point.

**Figure 3: Time series plot of $\pi_i$**



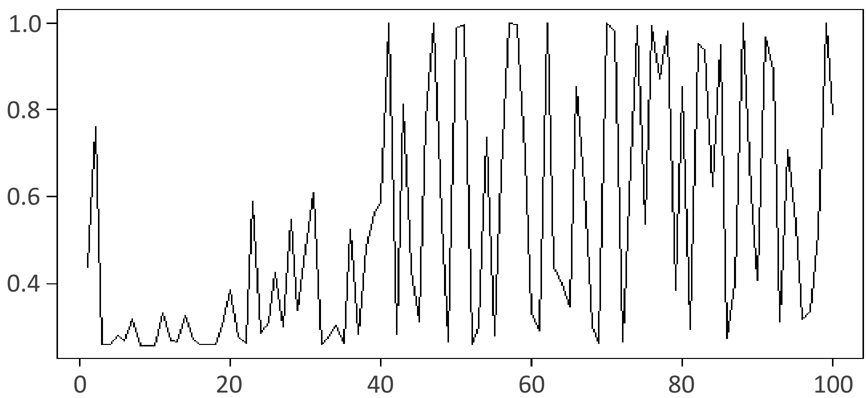

*Source: Own elaboration.*

*Case 3*. In the previous case, it was assumed that returns are independent. Here, it is assumed that $x_j = r_j - \mu$ obeys a first order autoregressive AR(1) process, given by

$$r_t - \mu = \alpha(r_{t-1} - \mu) + \varepsilon_t.$$

where $\mathcal{E}_t$ is a white noise process, normally distributed with zero mean and variance $\sigma_\epsilon^2 < \infty$. To test the existence of change in coefficient $\alpha$, the recursive least square estimate $\hat{\alpha}_t = \frac{\sum_{j=2}^{t} x_j x_{j-1}}{\sum_{j=2}^{t} x_{j-1}^2}$ is considered. Notice that $\hat{\alpha}_j = (1-\lambda_j)\hat{\alpha}_{j-1} + \lambda_j \frac{x_j}{x_{j-1}}$. Also, notice that $x_j$ given $x_{j-1}$ is normally distributed with mean $\alpha x_{j-1}$ and variance $\sigma_\epsilon^2$.. Thus, $\frac{x_j}{x_{j-1}}$ is normally distributed with mean $\alpha$ and variance $\frac{\sigma_\epsilon^2}{x_{j-1}^2}$. Let *n = 100, $k_0$ = 33,* $\sigma_\epsilon^2$ *= 0,77.* Coefficient $\alpha$ is 0.1 before the change and it is 0.75 after the change. Plot of $\pi_i$ is given in Figure 4, indicating there is a change at 33 time point.

**Figure 4: Time series plot of $\pi_i$**




*Source: Own elaboration.*

## Concluding remarks

Change point models seek to fit a piecewise regression model with unknown breakpoints to a 16 data set whose parameters are suspected of changing through of changing through time. However, the number of possible solutions to a multiple change point problem requires an efficient algorithm if long time series are to be analyzed. A sequential Bayesian change point algorithm is introduced that provides uncertainty bounds on both the number and location of change points.

The online change point detection in financial time series using the Bayesian setting is considered in the current paper. An algorithm is proposed for selecting the priors and estimation of parameters. The algorithm is able to detect the change points quickly. Simulation studies illustrate how the algorithm performs under various scenarios. Real data sets are presented to show the accuracy and performance of the proposed methods.

## References

Adams, R.P., MacKay, D.J. (2007). Bayesian Online Change Point Detection. *arXiv Preprint, arXiv:0710.3742*.

Davis, R., Lee, T., Rodriguez-Yam, G. (2006). Structural Break Estimation for Non-Stationary Time Series Models. *Journal of the American Statistical Association, 101*, 223–239.

Gombay, E. (2008). Change Detection in Autoregressive Time Series. *Journal of Multivariate Analysis, 99,* 451–464.

Gombay, E., Serban, D. (2009). Monitoring Parameter Change in Time Series Models. *Journal of Multivariate Analysis, 100,* 715–725.

Habibi, H., Howard, I., Habibi, R. (2017). Bayesian Sensor Fault Detection in Markov Jump System. *Asian Journal of Control, 19*, 1465–1481.

Koop, G.M., Potter, S.M. (2004). Forecasting and Estimating Multiple Change-point Models with an Unknown Number of Change Points. *Technical report*. USA: Federal Reserve Bank of New York.

Kurt, B.¸ Yıldız, C., Ceritli, T.Y., Sankur, B., Cemgil, A.T. (2018). A Bayesian Change Point Model for Detecting SIP-based DDoS Attacks. *Digital Signal Processing 77,* 48–62.

Lavielle, M., Moulines, E. (2000). Least-squares Estimation of an Unknown Number of Shifts in a Time Series. *Journal of Time Series Analysis, 21*, 33–59.

Lavielle, M., Teyssiere, G. (2007). Adaptive Detection of Multiple Change-points in Asset Price Volatility. *Long Memory in Economics*, 129-156.

Ombao, H., Raz, J., von Sachs, Malow, R. (2001). Automatic Statistical Analysis of Bi-variant Non-stationary Time Series. *Journal of the American Statistical Association, 96,* 543–560.

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

Saatci, Y., Turner, R.D., Rasmussen, C.E. (2010). Gaussian Process Change Point Models. In *Proceedings of the 27th International Conference on Machine Learning*. USA.

Taylor, S.J., Letham, B. (2018). Forecasting at Scale. *The American Statistician, 72*, 37–45.

Truong, C., Oudre, L., Vayatis, N. (2018). A Review of Change Point Detection Methods. *arXiv preprint, arXiv:1801.00718.*

Xiao, Z., Hu, S., Zhang, Q., Tian, X., Chen, Y., Wang, J., Chen, Z. (2018). Ensembles of Change-point Detectors: Implications for Real-time BMI Applications. *Journal of Computational Neuroscience, 10*, 1–18.