# Bayesian Approach to Phase II Statistical Process Control for Time Series

**A DISSERTATION**
**SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL**
**OF THE UNIVERSITY OF MINNESOTA**
**BY**

**Tianyang Zhou**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS**
**FOR THE DEGREE OF**
**Doctor of Philosophy**

**Douglas M. Hawkins**

**April, 2013**

# Acknowledgements

I would like to express my deepest gratitude to my academic advisor, Professor Douglas Hawkins, who made my dissertation possible with his helpful guidance and consistent patience. His erudition, rigorousness, kindness and confidence greatly helped me and encouraged me to overcome every challenge we faced during the research progress.

I would also like to thank Professor Yuhong Yang, Professor Galin Jones, Professor Peihua Qiu and Professor Robert Goldstein for sharing their ideas and research experience. They raised valuable and interesting questions to help me extend the research topic and perfect the details of my thesis.

I want to thank various people interviewing me throughout my job hunting season who talked with me about my research work. They tried to clarify and challenge me from different points of views, which made the contributions and restrictions of this work much clearer.

I would like to thank Professor John Monahan in NCSU, who showed his strong willing to help and always replied my letter with long and detailed answer and advices. He found me important technical works and helped me solve some critical problems in my thesis.

Finally I would like to thank my wife, Ange Pang, without whom I could not walk through the frustration and darkness. Thank all the people around me who offered their kind help in my journey to the PhD degree.

# Dedication

For my Dad and Mom, who are always beside me and supporting me.

For my maternal grandmother, who gave the best wishes when I was pursuing my dream.

In memory of my grandfathers and parental grandmother. You never ever walked away from the bottom of my heart.

**Abstract**

In statistical process control (SPC) problems, in-control values of parameters are required by traditional approaches. However this requirement is not realistic. New methods based on the change point model have been developed to avoid this requirement. The existing change-point methods are restricted to independent identically distributed observations, ignoring the numerous settings in which process readings are serially correlated. Furthermore, these frequentist methods are unable to make use of prior imperfect information on the parameters. In my research, I propose a Bayesian approach to the online SPC based on the change point model in an ARMA process. This approach accommodates serially correlated data, and also provides a coherent way of incorporating prior information on parameters.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Introduction of SPC Problem

Statistical process control (SPC) is an important statistical method widely used nowadays in various fields. Its traditional application is in mass-manufacturing, where abnormal performance in product process can be detected or predicted automatically, and thus waste and other undesired results can be avoided. Therefore, SPC is not only about statistics, but also about competitiveness Oakland (2008). Other successful applications are also in areas as diverse as animal sciences, biomedical areas, finance, and environmental monitoring.

SPC problems are usually statistically modeled as follows. A process being monitored is described by random variables $X_1, X_2, ...$ that are generated from a distribution $f(x|\underline{\theta} = \underline{\theta}_0)$, where $\underline{\theta}_0$ represents the values of parameters $\underline{\theta}$ when system works normally. This $\underline{\theta}_0$ is called the in-control value of $\underline{\theta}$. From time spot $\tau > 0$, there might be change in $\underline{\theta}$ from $\underline{\theta}_0$ to some other value $\underline{\theta}_1 \neq \underline{\theta}_0$, and the system is called out-of-control after time $\tau$ as long as $\underline{\theta}$ does not come back to normal value $\underline{\theta}_0$. At each time $n = 1, 2, ...$, a decision should be made on whether to give an out-of-control alarm, based on the realization of $X_1, X_2, ...X_n$. If a process is declared out of control, further investigation should be implemented to revise the system back to normal.

The classical tool of SPC is the control chart, such as Shewhart, cumulative sum

(CUSUM), and exponentially weighted moving average (EWMA). They work in the following way. At first, people carefully analyze phase I data, a historical data set with fixed size, to extract adequate information of the parameters. Any possible change of parameters should be investigated so that the reasons of change are fully understood. Then the information gathered in phase I data is used to monitor the phase II data, an endless data stream in which new observations come, and a decision of whether or not to give an out-of-control alarm should be made after each observation. Each of these control charts has two control limits, an upper control limits (UCL) and a lower control limits (LCL) that depends on the values of parameters $\underline{\theta}$. In most cases, true values of $\underline{\theta}$ are unknown, and a widely employed "plug-in" method is usually applied here. It uses estimates of parameters from phase I data as if they were the true values in the control of phase II data.

Since the initial work on these control charts of independent normal data, various extensions and improvements have been made so that more realistic and complicated models can be considered. These extensive works mainly concentrate on two aspects. Firstly, the stochastic structure should not be restricted to an independent joint distribution of the sample, because increasingly autocorrelated data appear nowadays in industry and if correlation among the data exists, some important indices will change when using those traditional methods. For example, Harris and Ross (1991) discussed the impact of autocorrelation on CUSUM and EWMA charts and demonstrated that the average run lengths (number of observations before the first alarm) of these charts are sensitive to the existence of autocorrelation.

Many methods have been developed to deal with autocorrelation. Alwan and Roberts (1988) proposed a residual chart method, in which a parametric model is established and residuals are estimated and analyzed by traditional control charts. Montgomery and Mastrangelo (1991) proposed the EWMA as a solution for monitoring an autocorrelated process. Vander Wiel (1996) studied the performance of the three traditional charts in an ARIMA model, and showed that the CUSUM has the best behavior. Zhang (1998) developed a residual chart method in which no modeling effort is required. Other literature about SPC in correlated process includes Koehler, Marks and O'Connell (2001) , Shu, Apley and Tsung (2002), Knoth and Schmid (2002), Castagliola and Tsung (2005) and Perry, Mercado and Pignatiello (2011).

The second aspect on which quality control papers focus is the pre-knowledge of in-control values of parameters. Traditional control charts usually require the true value or an accurate estimate of $\underline{\theta}_0$. However, in many scenarios this is unrealistic. Exact in-control values of the process, even good estimates, may not be available at all. Under some circumstances, for example, when phase I data are inadequate, or there is even no phase I data, the information about parameters is absent or vague. About this issue, some studies have been done with respect to the behavior of control limits using estimated parameter values, such as Jones, Champ, and Rigdon (2001) and Jones (2002). Some others proposed monitoring methods based on change point models. Hawkins, Qiu, and Kang (2003) used the two sample t-test to detect a change point in i.i.d normal process and developed an adaptive control limit sequence. Hawkins and Zamba (2005) applied the likelihood ratio test to detect change of both variance and mean in i.i.d (independent and identically distributed) normal data. Zamba and Hawkins (2006) studied quality control in multivariate normal sequence. Tsiamyrtzis and Hawkins (2008) developed a Bayesian quality control scheme for short-run $AR(1)$ time series which allows multiple change points, and Zhou, Zou, Zhang and Wang (2009) proposed a nonparametric chart by employing an EWMA control chart based on the Mann-Whitney statistic.

## 1.2 Change Point Model

The change point model is one of the most important relevant tools in quality control without detailed knowledge of the in-control status. The simplest change point model assumes that we have independent observations $X_1, X_2, ..., X_n$, where $X_i \sim f(x|\underline{\theta} = \underline{\theta}_0)$ for $i \leq \tau$, and $X_i \sim f(x|\underline{\theta} = \underline{\theta}_1)$ for $i > \tau$. That is, a change of the parameter $\underline{\theta}$ happens at time $\tau$. Some further extensions might be made such that there are multiple change points, autocorrelated in-control distribution, Markov change point structure, multiple changing parameters, etc.

Change point analysis may be traced back to Hinkley(1971), who did critical work about the frequentist approach to detection of a change point in i.i.d normal sample. A nonparametric approach is applied in Pettitt(1979). Andrews(1993) introduced Wald, Lagrange multiplier and likelihood ratio tests. Lombard and Hart(1994) and Vogelsang(1998) both considered general time series with dependent errors.

The Bayesian approach is also widely discussed by numerous papers. Compared to frequentist approaches, Bayesian method stands out in its capability of utilizing previous knowledge about all parameters in inference and prediction, especially when the data set is small. The earliest Bayesian change point model can date back to Chernoff and Zacks(1964), in which an i.i.d normal sequence with a Bernoulli change point is considered, and Gardner(1969) carefully discussed behavior of the change point in the same case. Smith(1975) concentrated on change point detection in both Binomial and Normal samples, and Booth and Smith(1982) further extended the work to some simple cases in time series modeling and discussed Bayesian hypothesis tests. West and Harrison(1986) embedded change point analysis in a dynamic Bayesian forecasting framework. Naylor and Woodward(1993) examined Bayesian significance test based on this model in actual data. McCulloch and Tsay(1993) extended Chernoff and Zacks to the autoregressive model, and their work further generalized to ARFIMA by Ray and Tsay(2001). Albert and Chib(1993) talked about a slightly different scenario in which a change of model structure is described as a Markov chain on a state space. Chib(1998) combined the state space change point model with Chernoff and Zacks' early paper to consider the situation with change point probability related to current state. Lai and Xing(2011) studied multiple change points in exponential family. Other interesting Bayesian change point models are described in Raftery and AkmanSource(1986), Wang and Zivot(2000), and Son and Kim(2005).

From the literature review, most previous work focused on the shift of mean, not only because it is of greatest interest, but also because it is the easiest case to start with. A few other papers study the changes of variance, such as Inclan and Tiao(1994), Giordani and Kohn(2006) and Davis, Lee and Rodriguez-Yam(2006). For general change, Dobigeon, Tourneret and Davy(2007) discussed the segmentation problem in $AR(p)$ model, which includes change of the whole autoregression structure.

## 1.3 Quality Control Based on Change Point Model

Bringing results from the change point model back to quality control scenario is straightforward, but the differences between change point detection in phase I and phase II data generate some further problems. In the framework of the change point model, the data set is often considered to have fixed size, but in phase II quality control, the data size increases endlessly. Therefore a couple of questions arise.

The first question is: should we model the process with single or multiple change points? Indeed, multiple change points fit reality better than single change point, but create an almost unsolvable puzzle since the first change point work in Chernoff and Zacks (1964). They first gave a general solution for multiple change points, but then turned to a simple case of AMOC (at most one change point) because of the resulting computational nightmare: the likelihood function will contain $2^n$ many terms if at each time there is positive probability for the change point to appear. Therefore usually methods dealing with exact estimation under multiple change point assumption only work for short-run processes, like the work in Tsiamyrtzis (2005). About this issue, Barry and Hartigan(1992) somewhat solved this problem. Their partition model has some excellent properties so that exact estimate of nearest change points and current parameter values can be completed with time consumption in order $O(n^3)$, while approximated estimate has $O(n^2)$ time consumption.

In the quality control case, however, AMOC is preferred for another reason. The reason is that for phase II data with ongoing observations, a change point is very likely to be detected soon after it occurs unless the change is small and so arguably unimportant. Therefore in this thesis, we will focus on the AMOC case.

The second question arising from the difference between the fixed-sample change point model and phase II quality control is the decision rule: when is the best time to give the alarm? For a fixed sample size, one can easily find reasonable model selection approaches, such as the odds ratio and Bayes factor. Suppose $X_1, X_2, ..., X_n$ have been collected and the posterior distribution of the change point location is derived, then one can compute the odds ratio

$$\lambda_{1,0}(\underline{X}) = \frac{P(M_1|\underline{X})}{P(M_0|\underline{X})}$$

or the Bayes factor

$$B_{1,0}(\underline{X}) = \frac{P(M_1|\underline{X})}{P(M_0|\underline{X})} \cdot \frac{P(M_0)}{P(M_1)}$$

where $M_1$ represents the model with change point, while $M_0$ is the model with no change point. Then the alarm can be given if $B_{1,0}(\underline{X}) \geq C$, (See Moreno, Casella and Garcia-Ferrer (2004)) or $\lambda_{1,0}(\underline{X}) > C$. These are reasonable schemes, but in phase II data when new observations are coming in an endless stream, we need to adapt this rule in order to fit the context of quality control.

Regarding this issue, Hawkins, Qiu, and Kang (2003) proposed a principle: If the process is under control, the probability of giving an alarm conditioned on there being no alarm before should be a constant. This is a nice property because in-control average run length can be calculated and therefore the behavior of such a control chart is traceable. This criterion is also applied in Hawkins and Zamba (2005), where SPC with shift in variance is studied, and Zamba and Hawkins (2005), in which a multivariate change point model is considered. In all these three papers, parametric models are assumed and used to define two-sample tests. Then control charts with tables of control limits based on these test statistics are computed through simulation.

Another candidate decision rule is based on loss functions. Taking the mean change of i.i.d normal process for example, a sensible loss function can be $f(|\delta|)I(t > \tau)$ per unit time, where $t$ is the current time, $\tau$ is the time when change point happens, $\delta$ is the shift of mean, and $f(\cdot)$ is a convex function defined on $[0, \infty)$. The purpose of quality control is to minimize the expected loss in the future. It can be proved that if $f(\cdot)$ is a constant function, then the corresponding best decision is consistent with the odds ratio approach.

If the loss function and decision rule are complicated, however, this dynamic programming problem will take exponential time to solve. In this thesis, we will propose a straightforward and simpler decision rule. This rule can be easily implemented and has a close connection with HQK and related papers.

## 1.4 Contribution and Organization of This Thesis

In summary, this thesis contributes to the SPC problem in the following aspects:

1. Proposes a general time series model (ARMA) which allows autocorrelation and correlation among noise terms.

2. Investigates a change of every parameter, providing a more flexible framework for the change point model and the SPC based on it.

3. Proposes a Bayesian approach which allows the utilization of informative yet incomplete knowledge about in control and out of control parameters.

4. Investigates the influence of hyperparameters on the behavior of this Bayesian SPC method.

The following chapters are organized as follows. Chapter 2 gives the set up of models and notation, as well as theoretical results. Chapter 3 discusses the decision rule. Chapter 4 gives a comparison and connection between our Bayesian approach and some other important methods. Chapter 5 shows simulation results of run length behavior. Chapter 6 talks about some computational issues. Chapter 7 gives a real example and analyzes it using the Bayesian SPC approach. Chapter 8 talks about the future work.

# Chapter 2

# Bayesian Change Point Model in Phase I Data

## 2.1 Change Point Model

Consider a stationary and invertible $ARMA(p,q)$ model:

$$(1-\sum_{i=1}^{p}\phi_i B^i)X_t = (1+\sum_{j=1}^{q}\theta_j B^j)\varepsilon_t \tag{2.1}$$

where

$$B^i X_t = X_{t-i}$$

is the backward operator, and $\varepsilon_t$'s are i.i.d $\mathcal{N}(\mu, \sigma^2)$. Possible change points may exist on any parameter, including the stationary mean $\mu$, variance $\sigma^2$ and regression or moving average coefficient $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_p)'$ or $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_q)'$. Because we only consider change point models under the scenario of stochastic process control, we assume at most one change point.

Under this circumstance, the $ARMA(p,q)$ model with at most one change point can be formulated as:

$$(1-\sum_{i=1}^{p}\phi_{1i}B^i)X_t=(1+\sum_{j=1}^{q}\theta_{1j}B^j)\varepsilon_t \quad \text{for } t\le T$$

$$(1-\sum_{i=1}^{p}\phi_{2i}B^i)X_t=(1+\sum_{j=1}^{q}\theta_{2j}B^j)\varepsilon_t \quad \text{for } t> T \tag{2.2}$$

where $\varepsilon_t$ i.i.d $\sim \mathcal{N}(\mu_1,\sigma_1^2)$ for $t\le T$ and $\varepsilon_t$ i.i.d $\sim \mathcal{N}(\mu_2,\sigma_2^2)$ for $t>T$. In this formulation, there are two sets of parameters on the left and right of the change point $T$: $(\mu_k,\sigma_k,\boldsymbol{\phi}_k,\boldsymbol{\theta}_k)$, $k=1,2$. If we do not allow some of them to change, for example, if the variance remains the same after $T$, we can set $\sigma_1=\sigma_2$ accordingly.

A matrix expression for the change point model is more concise and clear. For a fixed number of observations $X_1,X_2,...,X_n$ with $n>\max(p,q)$, let $\boldsymbol{X}_i^j=(X_i,...,X_j)'$, $\boldsymbol{\varepsilon}_i^j=(\varepsilon_i,...,\varepsilon_j)'$,

$$\Phi_k=\begin{pmatrix} -\phi_{kp} & \cdots & -\phi_{k2} & -\phi_{k1} & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0\\ 0 & -\phi_{kp} & \cdots & -\phi_{k2} & -\phi_{k1} & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0\\ \vdots & & \ddots & & & & \ddots & & \ddots & & & & \ddots & \vdots\\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & -\phi_{kp} & \cdots & -\phi_{k2} & -\phi_{k1} & 1 \end{pmatrix}$$

and

$$\Theta_k=\begin{pmatrix} \theta_{kq} & \cdots & \theta_{k2} & \theta_{k1} & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0\\ 0 & \theta_{kq} & \cdots & \theta_{k2} & \theta_{k1} & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0\\ \vdots & & \ddots & & & & \ddots & & \ddots & & & & \ddots & \vdots\\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \theta_{kq} & \cdots & \theta_{k2} & \theta_{k1} & 1 \end{pmatrix}$$

with $k=1,2$. As for the sizes, $\Phi_1$ is $T\times(T+p)$, $\Phi_2$ is $(n-T)\times(n-T+p)$, $\Theta_1$ is $T\times(T+q)$, and $\Theta_2$ is $(n-T)\times(n-T+q)$.

Now (2.2) can be expressed as

$$\Phi_1\boldsymbol{X}_{-p}^T=\Theta_1\boldsymbol{\varepsilon}_{-q}^T$$

$$\Phi_2\boldsymbol{X}_{T+1-p}^n=\Theta_2\boldsymbol{\varepsilon}_{T+1-q}^n \tag{2.3}$$

In fact, in this formulation we need to assume knowledge about $\boldsymbol{X}_{-p}^0=(X_{-p},...,X_0)'$ and $\boldsymbol{\varepsilon}_{-q}^0=(\varepsilon_{-q},...,\varepsilon_0)'$.

## 2.2 Likelihood Function

Now we derive the posterior distribution of the parameter set $(\mu_k, \sigma_k, \boldsymbol{\phi}_k, \boldsymbol{\theta}_k)$. Firstly we need to write down the likelihood function. For an $ARMA(p,q)$ with change point model formulated as (2.3), the likelihood function is:

$$L(\boldsymbol{X}_1^n|\mu_1,\mu_2,\sigma_1,\sigma_2,\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,T) \propto$$

$$(\frac{1}{\sigma_1^2})^{\frac{T}{2}} \exp[-\frac{1}{2\sigma_1^2}(\Phi_1\boldsymbol{X}_{-p}^T - \Theta_1\mu_1\mathbf{1}_T)'(\Theta_1\Theta_1')^{-1}(\Phi_1\boldsymbol{X}_{-p}^T - \Theta_1\mu_1\mathbf{1}_T)]$$

$$\cdot(\frac{1}{\sigma_2^2})^{\frac{n-T}{2}} \exp[-\frac{1}{2\sigma_2^2}(\Phi_2\boldsymbol{X}_{T+1-p}^n - \Theta_2\mu_2\mathbf{1}_{n-T})'(\Theta_2\Theta_2')^{-1}(\Phi_2\boldsymbol{X}_{T+1-p}^n - \Theta_2\mu_2\mathbf{1}_{n-T})] \tag{2.4}$$

If our assumption of prehistory knowledge is $\boldsymbol{\varepsilon}_{-q}^0 = \mathbf{0}$, the likelihood function (2.4) can be written as

$$L(\boldsymbol{X}_1^n|\mu_1,\mu_2,\sigma_1,\sigma_2,\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,T) \propto (\frac{1}{\sigma_1^2})^{\frac{T}{2}} \exp[-\frac{1}{2\sigma_1^2}||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T - \mu_1\mathbf{1}_T||_2^2]$$

$$\cdot(\frac{1}{\sigma_2^2})^{\frac{n-T}{2}} \exp[-\frac{1}{2\sigma_2^2}||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n - \mu_2\mathbf{1}_{n-T}||_2^2] \tag{2.5}$$

where $\Theta_1^*$ and $\Theta_2^*$ are the right-most square submatrices of $\Theta_1$ and $\Theta_2$:

$$\Theta_k^* = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \theta_{k1} & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \theta_{kq} & \cdots & \theta_{k2} & \theta_{k1} & 1 \end{pmatrix}$$

and the sizes are $T \times T$ and $(n-T)\times(n-T)$, respectively.

## 2.3 Prior Distribution

A natural choice of prior distribution is the conjugate prior distribution. However, this scheme is not realistic in that (1) The inverse of $\Theta^*$ has a very complicated form even if $q = 1$, and (2) we require stationarity and invertibility of this time series. Therefore, we adopt a quasi-conjugate prior distribution, which leaves the $ARMA$ coefficients $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ apart while concentrating on $\sigma$ and $\mu$. For example, if the change may occur in any parameter and the model is characterized as in (2.2) and (2.3), the choice of joint prior distribution for $(\mu, \sigma)$ will be

$$\pi(\mu_1, \sigma_1, \mu_2, \sigma_2) \propto (\frac{1}{\sigma_1^2})^{a_1+2} \exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1 - \mu_{10})^2 - \frac{b_1}{\sigma_1^2})(\frac{1}{\sigma_2^2})^{a_2+2} \exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2 - \mu_{20})^2 - \frac{b_2}{\sigma_2^2}) \tag{2.6}$$

Detailed choices of prior distribution in different cases will be discussed in the following sections.

The choice of prior distribution for $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ is more complicated. We are restricting our discussion to stationary and invertible process. Therefore, all roots for equations:

$$1 - \sum_{i=1}^{p} \phi_i z^i = 0 \quad \text{and} \quad 1 + \sum_{i=1}^{q} \theta_i z^i = 0$$

should be outside the unit ball. In most literature, this restriction is denoted by $(\boldsymbol{\phi}, \boldsymbol{\theta}) \in C_p \times C_q$, where $C_p$ and $C_q$ are subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$ which enable the unit ball condition to hold. Generally when $n \geq 3$ the structure of such subsets is very complicated. However, from Barndorff-Nielsen and Schou(1973), there is a one-to-one differentiable correspondence between $\boldsymbol{\phi}$ and the partial autocorrelation function $\mathbf{r} = (r_1, ..., r_p)$ of the $ARMA(p, q)$ process. Therefore, the condition of stationarity and invertibility is equivalent to $|r_i| < 1$ for $i = 1, 2, ..., p$. Notes from Piccolo(1982) and Monahan(1984) also gave a closed form recursive formula for such a transformation:

$$\phi_{i,n} = \phi_i \quad i = 1, ..., n$$

$$\phi_{i,k-1} = \frac{\phi_{i,k} - \phi_{k,k}\phi_{k-i,k}}{1 - \phi_{k,k}^2} \quad i = 1, ..., k-1, \ k = 1, ..., n$$

$$r_i = \phi_{i,i} \quad i = 1, ..., n$$

and for the inverse transformation:

$$\phi_{i,i} = r_i \quad i = 1, ..., n$$

$$\phi_{i,k} = \phi_{i,k-1} + \phi_k \phi_{k-i,k-1} \quad i = 1, ..., k-1, \ k = 1, ..., n$$

$$\phi_i = \phi_{i,n} \quad i = 1, ..., n$$

and an explicit expression for the Jacobian of $d\boldsymbol{\phi}/dr$:

$$J = \prod_{i=1}^{p} (1 - r_i^2)^{[\frac{1}{2}(i-1)]} \prod_{j=1}^{[\frac{1}{2}p]} (1 + r_{2j}) \tag{2.7}$$

Setting up a prior distribution on $C_p$ is the same as assigning a prior distribution on $\mathbf{r}$. It is reasonable to choose an i.i.d transformed beta distribution as a start, i.e.

$$\pi(\mathbf{r}) \propto \prod_{i=1}^{p} (1 + r_i)^{\alpha_i} (1 - r_i)^{\beta_i} \tag{2.8}$$

Similarly, for $\boldsymbol{\theta}$, the unit root rule can be written as:

$$1 - \sum_{i=1}^{q} (-\theta_i) z^i = 0$$

which has similar form as for $\boldsymbol{\phi}$, so we can use an analogous scheme to sample $\boldsymbol{\theta}$.

Some special cases own their simple prior distributions for $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. For example, for $ARMA(1,1)$ sequence, the prior distribution can be set up with

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}) \propto (1 + \phi)^{\alpha} (1 - \phi)^{\beta} (1 + \theta)^{\gamma} (1 - \theta)^{\delta} \tag{2.9}$$

which is essentially a scaled beta distribution: $(\phi + 1)/2 \sim Beta(\alpha, \beta)$ and $(\theta + 1)/2 \sim Beta(\gamma, \delta)$. In the future, we will directly denote $\phi \sim Beta(\alpha, \beta)$ and $\theta \sim Beta(\gamma, \delta)$, without special clarification. In $ARMA(2,2)$ sequence, the prior distribution can be set up with

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}) \propto (1 + \phi_2)^{\alpha_1} (1 - \phi_2)^{\beta_1} (1 - \frac{\phi_1}{1 - \phi_2})^{\alpha_2} (1 + \frac{\phi_1}{1 - \phi_2})^{\beta_2}$$
$$\cdot (1 + \theta_2)^{\gamma_1} (1 - \theta_2)^{\delta_1} (1 - \frac{\theta_1}{1 + \theta_2})^{\gamma_2} (1 + \frac{\theta_1}{1 + \theta_2})^{\delta_2}$$

For higher order $ARMA$ model, we can first sample $r_i$'s from (2.8), transform to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, and then apply a Monte Carlo method to evaluate the posterior distribution.

The prior distribution of $T$ can have several different forms. A natural choice is the discrete uniform distribution, i.e. $P(T=j)=1/n,\ j=1,2,...,n$. Another choice is the geometric distribution, i.e.

$$P(T=j)=\begin{cases} p_0(1-p_0)^{j-1} & 1\leq j\leq n-1 \\ (1-p_0)^{n-1} & j=n \end{cases}$$

This prior distribution guarantees that as $n$ increases, each time spot less than $n$ has fixed mass. In chapter 3, we will discuss the choice of $\pi(T)$ in the context of Phase II process control.

## 2.4 Posterior Distribution

We will discuss the posterior distribution under several change point cases: the change of mean $\mu$, change of variance $\sigma^2$ and change of time series coefficients $(\boldsymbol{\phi},\boldsymbol{\theta})$. There are two states (change or no change) for each category, so we have $2^3-1=7$ different cases.

**Change of All Parameters**

The choice of prior distribution is:

$$\pi(\mu_1,\sigma_1,\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\mu_2,\sigma_2,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T)\propto(\frac{1}{\sigma_1^2})^{a_1+2}\exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1-\mu_{10})^2-\frac{b_1}{\sigma_1^2})$$

$$\cdot(\frac{1}{\sigma_2^2})^{a_2+2}\exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2-\mu_{20})^2-\frac{b_2}{\sigma_2^2})\cdot\pi(\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2)\pi(T)$$

Combining this prior with the likelihood function, we have the joint posterior distribution

$$f(\mu_1,\sigma_1,\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\mu_2,\sigma_2,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T|\boldsymbol{X}_1^n)\propto L(\boldsymbol{X}_1^n|\mu_1,\mu_2,\sigma_1,\sigma_2,\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,T)$$

$$\cdot\pi(\mu_1,\sigma_1,\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\mu_2,\sigma_2,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T)$$

We will be interested in the marginal distribution of $T$. So we have

$$f(\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T|\boldsymbol{X}_1^n)=\int_{\mu_1,\sigma_1,\mu_2,\sigma_2}f(\mu_1,\sigma_1,\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\mu_2,\sigma_2,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T|\boldsymbol{X}_1^n)d\mu_1d\sigma_1d\mu_2d\sigma_2$$

$$\propto\int_{\mu_1,\sigma_1,\mu_2,\sigma_2}(\frac{1}{\sigma_1^2})^{\frac{T}{2}}\exp[-\frac{1}{2\sigma_1^2}||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T-\mu_1\mathbf{1}_T||_2^2](\frac{1}{\sigma_1^2})^{a_1+2}\exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1-\mu_{10})^2-\frac{b_1}{\sigma_1^2})$$

$$\cdot(\frac{1}{\sigma_2^2})^{\frac{n-T}{2}}\exp[-\frac{1}{2\sigma_2^2}||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n-\mu_2\mathbf{1}_{n-T}||_2^2](\frac{1}{\sigma_2^2})^{a_2+2}\exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2-\mu_{20})^2-\frac{b_2}{\sigma_2^2})$$

$$\cdot\pi(\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2)\pi(T)d\mu_1 d\sigma_1 d\mu_2 d\sigma_2$$

$$\propto \pi(\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2)\pi(T)(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{T}{2}+a_1)\Gamma(\frac{n-T}{2}+a_2)$$

$$\cdot[\frac{1}{2}(S_{2,T}^-+\tau_1^2\mu_{10}^2+2b_1-\frac{(S_{1,T}^-+\tau_1^2\mu_{10})^2}{T+\tau_1^2})]^{-\frac{T}{2}-a_1}[\frac{1}{2}(S_{2,T}^++\tau_2^2\mu_{20}^2+2b_2-\frac{(S_{1,T}^++\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})]^{-\frac{n-T}{2}-a_2} \tag{2.10}$$

where

$$S_{1,T}^- = 1_T'\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T, \quad S_{2,T}^- = ||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T||^2$$

$$S_{1,T}^+ = 1_{n-T}'\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1}^n, \quad S_{2,T}^+ = ||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1}^n||^2$$

and $\Gamma(x)=\int_0^\infty t^{x-1}\exp(-t)dt$ is the gamma function.

The marginal posterior distribution of $T$ can be derived from:

$$f(T|\boldsymbol{X}_1^n)=\int_{\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2} f(\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2,T|\boldsymbol{X}_1^n)d\boldsymbol{\phi}_1 d\boldsymbol{\theta}_1 d\boldsymbol{\phi}_2 d\boldsymbol{\theta}_2$$

and this can be evaluated by numerical methods which will be discussed at the end of this chapter.

### 2.4.1 Same $\sigma$, Change of $\mu$, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$

In this case, $\sigma^2$ does not change, i.e. $\sigma_1=\sigma_2=\sigma$, and there is change of $\mu$ and $(\boldsymbol{\phi},\boldsymbol{\theta})$. The likelihood function becomes:

$$L(\boldsymbol{X}_1^n|\mu_1,\mu_2,\sigma,\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,T)$$

$$\propto(\frac{1}{\sigma^2})^{\frac{n}{2}}\exp[-\frac{1}{2\sigma^2}(||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T-\mu_1\mathbf{1}_T||_2^2+||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n-\mu_2\mathbf{1}_{n-T}||_2^2)]$$

The choice of prior distribution is:

$$\pi(\mu_1,\mu_2,\sigma^2,T)\propto(\frac{1}{\sigma^2})^{a+2}\exp(-\frac{\tau_1^2}{2\sigma^2}(\mu_1-\mu_{10})^2-\frac{\tau_2^2}{2\sigma^2}(\mu_2-\mu_{20})^2-\frac{b}{\sigma^2})\pi(T)$$

i.e.

$$\sigma^2\sim IG(a,b), \quad \mu_1|\sigma^2\sim N(\mu_{10},\frac{\sigma^2}{\tau_1^2}), \quad \mu_2|\sigma^2\sim N(\mu_{20},\frac{\sigma^2}{\tau_2^2}), \quad \mu_1\perp\mu_2|\sigma^2$$

The posterior distribution is given by:

$$f(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2, T|\boldsymbol{X}_1^n) \propto \pi(T)\pi(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2)(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}$$

$$\cdot[\frac{1}{2}(S_{2,T}^- + \tau_1^2\mu_{10}^2 - \frac{(S_{1,T}^- + \tau_1^2\mu_{10})^2}{T+\tau_1^2} + S_{2,T}^+ + \tau_2^2\mu_{20}^2 - \frac{(S_{1,T}^+ + \tau_2^2\mu_{20})^2}{n-T+\tau_2^2}) + b]^{-\frac{n}{2}-a} \quad (2.11)$$

where

$$S_{1,T}^- = 1_T'\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T, \quad S_{2,T}^- = ||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T||^2$$

$$S_{1,T}^+ = 1_{n-T}'\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1}^n, \quad S_{2,T}^+ = ||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1}^n||^2$$

The marginal posterior distribution for $T$ can be evaluated by numerical integration.

### 2.4.2 Same $\mu$, Change of $\sigma$, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$

In this case, $\mu$ does not change, i.e. $\mu_1 = \mu_2 = \mu$, and there is change of $\sigma$ and $(\boldsymbol{\phi}, \boldsymbol{\theta})$. The likelihood function becomes:

$$L(\boldsymbol{X}_1^n|\mu, \sigma_1, \sigma_2, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, T) \propto (\frac{1}{\sigma_1^2})^{\frac{T}{2}} \exp[-\frac{1}{2\sigma_1^2}||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T - \mu\mathbf{1}_T||_2^2]$$

$$\cdot(\frac{1}{\sigma_2^2})^{\frac{n-T}{2}} \exp[-\frac{1}{2\sigma_2^2}||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n - \mu\mathbf{1}_{n-T}||_2^2]$$

The choice of prior distribution is:

$$\pi(\mu, \sigma_1^2, \sigma_2^2, T) \propto (\frac{1}{\sigma_1^2})^{a_1+1} exp(-\frac{b_1}{\sigma_1^2})(\frac{1}{\sigma_2^2})^{a_2+1} exp(-\frac{b_2}{\sigma_2^2}) \exp[\frac{(\mu-\mu_0)^2}{2\sigma_0^2}]\pi(T)$$

i.e.

$$\sigma_1^2 \sim IG(a_1, b_1), \quad \sigma_2^2 \sim IG(a_2, b_2), \quad \mu \sim N(\mu_0, \sigma_0^2), \quad \mu \perp \sigma_1^2, \sigma_2^2$$

The posterior distribution:

$$f(\mu, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, T|X) \propto \Gamma(a_1+\frac{T}{2})\Gamma(a_2+\frac{n-T}{2})(\frac{1}{2}\sum_{i=1}^{T}||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T - \mu\mathbf{1}_T||_2^2 + b_1)^{-(a_1+\frac{T}{2})}$$

$$\cdot(\frac{1}{2}||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n - \mu\mathbf{1}_{n-T}||_2^2 + b_2)^{-(a_2+\frac{n-T}{2})} \exp[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2]$$

$$\cdot\pi(T)\pi(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2) \quad (2.12)$$

The marginal posterior distribution for $T$ can be evaluated by numerical integration.

### 2.4.3 Same $\mu$ and $\sigma$, Change of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$

In this case, $\mu$ and $\sigma$ do not change, and there is change of $(\boldsymbol{\phi}, \boldsymbol{\theta})$. The likelihood function becomes:

$$L(\boldsymbol{X}_1^n|\mu, \sigma, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, T)$$
$$\propto (\frac{1}{\sigma^2})^{\frac{n}{2}} \exp[-\frac{1}{2\sigma^2}(||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T - \mu\mathbf{1}_T||_2^2 + ||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n - \mu\mathbf{1}_{n-T}||_2^2)]$$

The choice of prior distribution is:

$$\pi(\mu, \sigma^2, T) \propto (\frac{1}{\sigma^2})^{a+1} \exp(-\frac{\tau^2}{2\sigma^2}(\mu - \mu_0)^2 - \frac{b}{\sigma^2})\pi(T)$$

i.e.

$$\sigma^2 \sim IG(a, b), \quad \mu|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\tau^2})$$

The posterior distribution is given by:

$$f(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2, T|\boldsymbol{X}_1^n) \propto [S_2^- + S_2^+ + \tau^2\mu_0^2 - \frac{(S_1^- + S_1^+ + \tau^2\mu_0^2)^2}{\tau^2 + 1}]\pi(T)\pi(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2) \tag{2.13}$$

Where $S_1^-$, $S_1^+$, $S_2^-$, $S_2^+$ are defined the same as above. $f(T|\boldsymbol{X}_1^n)$ can be evaluated through Monte Carlo method.

### 2.4.4 Same $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$

This case is similar to the sections 2.4.1, 2.4.2 and 2.4.3, except that $\boldsymbol{\phi}_1 = \boldsymbol{\phi}_2 = \boldsymbol{\phi}$, $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}$, and $S_{1,T}^-$, $S_{2,T}^-$, $S_{1,T}^+$ and $S_{2,T}^+$ are defined as

$$S_{1,T}^- = 1_T'\Theta^{*-1}\Phi\boldsymbol{X}_{-p}^T, \quad S_{2,T}^- = ||\Theta^{*-1}\Phi\boldsymbol{X}_{-p}^T||^2$$

$$S_{1,T}^+ = 1_{n-T}'\Theta^{*-1}\Phi\boldsymbol{X}_{T+1}^n, \quad S_{2,T}^+ = ||\Theta^{*-1}\Phi\boldsymbol{X}_{T+1}^n||^2$$

**Change of $\mu$ and $\sigma$**

The prior distribution is:

$$\pi(\mu_1, \sigma_1, \mu_2, \sigma_2, T) \propto (\frac{1}{\sigma_1^2})^{a_1+2} \exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1 - \mu_{10})^2 - \frac{b_1}{\sigma_1^2})$$

$$\cdot(\frac{1}{\sigma_2^2})^{a_2+2} \exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2 - \mu_{20})^2 - \frac{b_2}{\sigma_2^2})\pi(T)$$

The posterior distribution is given by:

$$f(\boldsymbol{\phi},\boldsymbol{\theta},T|\boldsymbol{X}_1^n) \propto \pi(\boldsymbol{\phi},\boldsymbol{\theta})\pi(T)(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{T}{2}+a_1)\Gamma(\frac{n-T}{2}+a_2)$$

$$\cdot[\frac{1}{2}(S_{2,T}^{-}+\tau_1^2\mu_{10}^2+2b_1-\frac{(S_{1,T}^{-}+\tau_1^2\mu_{10})^2}{T+\tau_1^2})]^{-\frac{T}{2}-a_1}[\frac{1}{2}(S_{2,T}^{+}+\tau_2^2\mu_{20}^2+2b_2-\frac{(S_{1,T}^{+}+\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})]^{-\frac{n-T}{2}-a_2} \tag{2.14}$$

as in 2.4.1.

**Same $\sigma$ and Change of $\mu$**

In this case, $\sigma_1=\sigma_2=\sigma$, and the choice of prior distribution is:

$$\pi(\mu_1,\mu_2,\sigma^2,T) \propto (\frac{1}{\sigma^2})^{a+2}\exp(-\frac{\tau_1^2}{2\sigma^2}(\mu_1-\mu_{10})^2-\frac{\tau_2^2}{2\sigma^2}(\mu_2-\mu_{20})^2-\frac{b}{\sigma^2})\pi(T)$$

The posterior distribution is given by:

$$f(\boldsymbol{\phi},\boldsymbol{\theta},T|\boldsymbol{X}_1^n) \propto \pi(\boldsymbol{\phi},\boldsymbol{\theta})\pi(T)(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}$$

$$\cdot[\frac{1}{2}(S_{2,T}^{-}+\tau_1^2\mu_{10}^2-\frac{(S_{1,T}^{-}+\tau_1^2\mu_{10})^2}{T+\tau_1^2}+S_{2,T}^{+}+\tau_2^2\mu_{20}^2-\frac{(S_{1,T}^{+}+\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})+b]^{-\frac{n}{2}-a} \tag{2.15}$$

as in 2.4.2.

**Same $\mu$ and Change of $\sigma$**

In this case, $\mu_1=\mu_2=\mu$, and the choice of prior distribution is:

$$\pi(\mu,\sigma_1^2,\sigma_2^2,T) \propto (\frac{1}{\sigma_1^2})^{a_1+1}exp(-\frac{b_1}{\sigma_1^2})(\frac{1}{\sigma_2^2})^{a_2+1}exp(-\frac{b_2}{\sigma_2^2})\exp[\frac{(\mu-\mu_0)^2}{2\sigma_0^2}]\pi(T)$$

The posterior distribution:

$$f(\mu,\boldsymbol{\phi},\boldsymbol{\theta},T|X) \propto \pi(\boldsymbol{\phi},\boldsymbol{\theta})\pi(T)\Gamma(a_1+\frac{T}{2})\Gamma(a_2+\frac{n-T}{2})(\frac{1}{2}\sum_{i=1}^{T}||\Theta_1^{*-1}\Phi_1\boldsymbol{X}_{-p}^T-\mu\mathbf{1}_T||_2^2+b_1)^{-(a_1+\frac{T}{2})}$$

$$\cdot(\frac{1}{2}||\Theta_2^{*-1}\Phi_2\boldsymbol{X}_{T+1-p}^n-\mu\mathbf{1}_{n-T}||_2^2+b_2)^{-(a_2+\frac{n-T}{2})}\exp[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2] \tag{2.16}$$

as in 2.4.3.

## 2.5 Numerical Integration

In Chapter 2.4, we face the problem of integrating

$$f(T|\boldsymbol{X}_1^n) = \int_{\boldsymbol{\phi}_1,\boldsymbol{\theta}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_2} f(\boldsymbol{\phi}_1, \boldsymbol{\theta}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_2, T|\boldsymbol{X}_1^n) d\boldsymbol{\phi}_1 d\boldsymbol{\theta}_1 d\boldsymbol{\phi}_2 d\boldsymbol{\theta}_2$$

or similar posterior distribution functions. For cases when analytical integration is impossible, we can apply numerical integration methods. To evaluate a one-dimensional integration with a finite range of integration variable, we can sum up the value of $f$ in finite integration points applying quadrature rules, such as rectangle, trapezoidal, or Newton-Cotes rule. For example, an $AR(1)$ model without change of $\phi$ (but $\mu$ and $\sigma$ have possible changes) should apply this finite summation method because this is univariate case with $-1 < \phi < 1$. To evaluate an integration with higher dimension, Monte Carlo integration (Hammersley and Handscomb (1964)) is a better tool, because if we still use the one-dimensional method and Fubini Theorem to evaluate multi-dimensional integration, the number of integration points needed is exponentially high, and we encounter the so called "curse of dimensionality".

The Monte Carlo integration works as follows. Suppose we would like to integrate $f(\boldsymbol{x})p(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, ..., x_m) \in \mathbb{R}^m$, and suppose random vector $\boldsymbol{X} \sim p(\boldsymbol{x})$, where $p(\boldsymbol{x})$ is a known distribution. Then

$$\int_{\mathbb{R}^m} f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = E_p(f(\boldsymbol{x}))$$

If we can sample from distribution $p(\boldsymbol{x})$ and obtain a sample $\{\boldsymbol{X_j}\}_{j=1}^N$, the definite integral can be approximated by

$$\int_{\mathbb{R}^m} \widehat{f(\boldsymbol{x})p(\boldsymbol{x})} d\boldsymbol{x} \approx \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{X_j})$$

The variance of the Monte Carlo integration estimate is

$$\text{Var}[\frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{X_j})] = \frac{1}{N}\text{Var}[f(\boldsymbol{X_1})]$$

Therefore, if

$$\sigma_f^2 = \text{Var}[f(\boldsymbol{X_1})] := \int_{\mathbb{R}^m} f^2(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} - [\int_{\mathbb{R}^m} f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}]^2 < \infty$$

Then the $\sqrt{N}$-convergence is guaranteed:

$$\frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{X_j}) - \int_{\mathbb{R}^m} f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = O_p(N^{-1/2})$$

Compared to method with uniform integration points, Monte Carlo integration is desired in that for high dimensional integration, the convergence rate is not related to the dimension $d$, and thus we can avoid the curse of dimensionality.

An improvement of Monte Carlo integration is Quasi-Monte Carlo integration, which differs from Monte Carlo integration only in the way the random sample is generated. Monte Carlo integration uses i.i.d samples, while Quasi-Monte Carlo integration uses low-discrepancy random numbers. An example is the Halton sequence: For instance if we would like to sample $N$ points uniformly from $[0,1]^d$, we can express an integer $n$ $(1 \le n \le N)$ by $b_s$-ary $(1 \le s \le d)$ representation:

$$n = \sum_{i=0}^{k_n} a_i(n) b_s^i$$

Then we choose the $n$th random number to be

$$g(n) = (g_1(n), ..., g_d(n))$$

where

$$g_s(n) = \sum_{i=0}^{k_n} a_i(n) b_s^{-i-1} \quad \text{for } s = 1, ..., d$$

Any other random vector sequence can be therefore generated by a transformation from this Halton sequence.

Another improvement is importance sampling. The idea is that if we would like to evaluate the integral of $f(x)$ by a Monte Carlo method, we want to sample more points from the region where $f(x)$ is more volatile, and to sample fewer points from the region where $f(x)$ is more smooth. To show this, suppose we approximate

$$\int_{\mathbb{R}^d} h(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{R}^d} \frac{h(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})d\boldsymbol{x} = E_g(\frac{h(\boldsymbol{x})}{g(\boldsymbol{x})})$$

by

$$\hat{E}_g(\frac{h(\boldsymbol{x})}{g(\boldsymbol{x})}) = \frac{1}{N}\sum_{i=1}^{N}\frac{h(\boldsymbol{X_i})}{g(\boldsymbol{X_i})}$$

where $g(\boldsymbol{x})$ is a pdf. Then

$$Var(\hat{E}_g) = \frac{1}{N}Var(\frac{h(\boldsymbol{x})}{g(\boldsymbol{x})})$$

If $g(\boldsymbol{x}) = h(\boldsymbol{x})/\int h(\boldsymbol{x})d\boldsymbol{x}$, $Var(\hat{E}_g) = 0$. However, this is unrealistic, since we no longer need the Monte Carlo method if we can set up $g$ in this way. Nevertheless, we can still use a $g$ which can make $h/g$ smoother than $h$ itself to reduce the variance.

# Chapter 3

# Phase II Data and Decision Rule

## 3.1 Phase II Data

In SPC problems, phase II data form an endless sequence where analysis is conducted as the data set size increases. In this scenario, a decision (alarm or no alarm) is made each time a new observation is available, such that a reasonable loss function is minimized. This loss function must take into account the type I error, that an alarm is given when process is in control, and the type II error, that an alarm fails to be given when the process is out of control.

From the argument in chapter 2, we see that in phase II data, the assumption of at most one change point is reasonable. Therefore, we directly model this scenario so that only one change will occur to the sequence. This change can be change of mean ($\mu$), variance ($\sigma^2$), or coefficients of $ARMA$ model ($\boldsymbol{\phi}$ and $\boldsymbol{\theta}$). However, only one time $T$ exists in the whole sequence such that some parameters change at time $T$.

The Bayesian phase II SPC problem is modeled as follows:

1. The parameters $\mu_j$, $\sigma_j$, $\boldsymbol{\phi}_j$, and $\boldsymbol{\theta}_j$ ($j = 1, 2$), are generated from their corresponding prior distribution.

2. The change point $T$ is generated from a prior distribution $\pi(T)$ on positive integer set $\mathbb{N}$.

3. For $n = 1, 2, ...,$ observations $X_1,\ X_2, ...$ are obtained from an $ARMA(p,q)$ model characterized in chapter 2.1. After $X_n$ is observed, a decision $D_n \in \{0, 1\}$ is made based on observations $(X_1, X_2, ..., X_n)$. Here 1 represents alarm decision, and 0 represents no alarm decision.

4. If no alarm is given, the process will keep going. A new observation is read and a decision problem for time spot $n+1$ comes. If an alarm is given, the process will stop and deep investigation is conducted.

Numerous choices for $\pi(T)$ are legitimate, as long as they satisfy $\sum_{T=1}^{\infty} \pi(T) < \infty$. Among them the geometric($p$) distribution is a very natural one, since it is equivalent to a time-homogeneous Markov Chain model, in which there are two states: 0 (in control) and 1 (out of control), and the transition matrix is

$$P = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}$$

This formulation assumes a memoryless change point pattern, i.e., the risk rate is constant as $T$ increases. A more complicated choice is the discrete Weibull distribution, where

$$\pi(T) = (1-p)^{(T-1)^\beta} - (1-p)^{T^\beta}, \qquad T = 1, 2, ..., \quad \beta > 1$$

This distribution describes the wear-out effect in that the risk rate

$$\frac{\pi(T)}{\sum_{i=T}^{\infty} \pi(i)} = \frac{(1-p)^{(T-1)^\beta} - (1-p)^{T^\beta}}{(1-p)^{(T-1)^\beta}} = 1 - (1-p)^{T^\beta - (T-1)^\beta}$$

is increasing as $T$ grows.

One modification that might be considered is to revert the prior above, i.e, instead of use $\pi(T) = p_T = p(1-p)^{T-1}$, we use $p_T = p(1-p)^{n-T-1}$, for $1 \le T \le n-1$, and $p_n = (1-p)^{n-1}$ stays the same. This is to take into account the fact that at time $n$, we are more inclined to believe that there is a change point undetected more recently than long ago. If we are not willing to incline the weight of prior, we can instead use $p_T = [1-(1-p)^{n-1}]/(n-1)$, which is corresponding to an discrete uniform distribution over the first $n-1$ time spots.

## 3.2 Loss Function and Bayesian Decision Rule

The first approach of a Bayesian decision rule is based on a loss function. In fact, we can obtain the odds ratio and Bayes factor of comparing the models:

$$M_0: \ T \geq n$$

$$M_1: \ T < n$$

for each $n$. Here $M_0$ represents the case that there is no change point, or the process is in control, and $M_1$ represents the case that there is one change point, or the process is out of control. A traditional way of decision making, under each fixed $n$, is to give an alarm if $\lambda_{1,0}(\boldsymbol{X}_1^n) > C$, where $\lambda_{1,0}$ is the odds defined in chapter 1, and $C$ is a constant determined by the user. This method works best if the loss function is defined as

$$L(D, M) = \begin{cases} L_1 & D = 1, M = M_0 \\ L_2 & D = 0, M = M_1 \\ 0 & \text{otherwise} \end{cases}$$

where $M$ denotes the model and $D$ denotes the decision. $L_1$ and $L_2$ are two positive loss constants. In this case, the best decision rule $D^*$ satisfies:

$$D^* = \arg\min_D E(L(D, M)|\boldsymbol{X}_1^n) = \begin{cases} L_1 P(M_0|\boldsymbol{X}_1^n) & D = 1 \\ L_2 P(M_1|\boldsymbol{X}_1^n) & D = 0 \end{cases}$$

and therefore $D^* = D$ if and only if $L_1 P(M_0|\boldsymbol{X}_1^n) < L_2 P(M_1|\boldsymbol{X}_1^n)$, i.e, $\lambda_{1,0}(\boldsymbol{X}_1^n) > L_1/L_2$.

In phase II statistical process control, however, this may not be the best decision rule, because the loss function is not defined in the same way as when the data size is fixed. Different from a fixed sample, a phase II sequence is an on-going sequence, and therefore makes the loss function dynamic: the loss function in step $n$ depends on the loss functions thereafter. For example, suppose at step $n$ we make a decision of not giving the alarm and keeping monitoring, then there are two possible status in step $n+1$: the decision made at $n$ is right or wrong, two status for the process, in- or out-of-control, and thus two loss functions. This approach was considered by Naylor and Woodward (1992). They pointed out that although this method is correct theoretically, it does not work practically since the number of cases needing to be taken care of is exponentially large.

A compromise version of this approach is to simplify the future loss expression. Let us consider a scenario in real application, that losses include the defective products and false alarm. If an alarm is made, the loss would be the cost of check $C_0$ plus the potential cost for repair $C_1$ should there be a change point occurring before. If the alarm is not made, the loss would be the cost generated from defective products $C_d$, which may also include prospective loss in the future if the control system fails to give alarm under out-of-control status. The loss function at time $n$ can be expressed as:

$$L_n(D(\boldsymbol{X}_1^n)) = \begin{cases} C_0 + C_1 I(M_1) & \text{if } D(\boldsymbol{X}_1^n) = 1 \\ C_d I(M_1) & \text{if } D(\boldsymbol{X}_1^n) = 0 \end{cases}$$

The best decision $D^*(\boldsymbol{X}_1^n)$ will minimize the expected loss in the future:

$$D^*(\boldsymbol{X}_1^n) = \min_{D(\boldsymbol{X}_1^n) \in \{0,1\}} E[L_n(D(\boldsymbol{X}_1^n))]$$

It is easy to see that

$$E[L_n(D(\boldsymbol{X}_1^n))|\boldsymbol{X}_1^n] = \begin{cases} C_0 + C_1 \Pr(M_1|\boldsymbol{X}_1^n) & \text{if } D(\boldsymbol{X}_1^n) = 1 \\ C_d \Pr(M_1|\boldsymbol{X}_1^n) & \text{if } D(\boldsymbol{X}_1^n) = 0 \end{cases}$$

and $D(\boldsymbol{X}_1^n) = 1$ is the best decision if and only if $C_0 + C_1\Pr(M_1|\boldsymbol{X}_1^n) < C_d\Pr(M_1|\boldsymbol{X}_1^n)$, i.e.

$$\lambda_{1,0}(\boldsymbol{X}_1^n) = \frac{\Pr(M_1|\boldsymbol{X}_1^n)}{\Pr(M_0|\boldsymbol{X}_1^n)} > \frac{C_0}{C_d - C_1 - C_0}$$

In real application, we have

$$\lambda_{1,0}(\boldsymbol{X}_1^n) = \frac{\sum_{T<n} f(T|\boldsymbol{X}_1^n)}{\sum_{T \geq n} f(T|\boldsymbol{X}_1^n)} > \frac{C_0}{C_d - C_1 - C_0}$$

This decision rule actually to monitor the process until $\lambda_{1,0}$ exceeds a constant threshold.

## 3.3 Discriminant Analysis

Another approach emerges from the perspective of discriminant analysis. In a multi-class classification problem, if a classification rule is generated from training data set, the traditional discriminant analysis will assign a new observation $\boldsymbol{Y}$ to class $c \in \mathcal{C} = \{1, 2, ..., C\}$ when $c$ provides the maximum likelihood $L(c|\boldsymbol{Y})$. In the Bayesian framework, if there exists a well defined prior distribution $\pi$ on the class set $\mathcal{C}$, the discriminant analysis will assign observation $\boldsymbol{Y}$ to class $c$ which maximizes the posterior $\pi(c)L(c|\boldsymbol{Y})$.

In the framework of Bayesian phase II SPC problem, we view the change point $T$ as a fixed time spot preselected from a prior distribution $\pi(T)$. Under this circumstance, there are $n$ classes in $\mathcal{C} = \{1, 2, ..., n\}$. The $\boldsymbol{Y}$ here is gradually revealed by new observations of $\boldsymbol{X}$. At time $n$, the classification rule is based on a maximum posterior distribution of $T$ over $\mathcal{C}$.

Finally, the decision rule of signaling is as follows. At time $n$, if the assigned class is after $n$, no alarm is given and a new observation is read. If the assigned class is before $n$, an alarm is given, i.e.

$$D_n(\boldsymbol{X}_1^n) = \begin{cases} 0 & \text{If } T^* \geq n \\ 1 & \text{If } T^* < n \end{cases}$$

where

$$T^* = \arg \max_{1 \leq T \leq N} f(T|\boldsymbol{X}_1^n)$$

In the rest chapters of this thesis, we will focus on the decision rule based on discriminant analysis. This is not only because this method does not require another round of calibration or simulation study of the preselected threshold, but also because the behavior of the SPC method will be essentially an extension of multiple classical SPC methods developed before. We will show this in the following chapter.

# Chapter 4

# Comparison and Connection with Other Methodologies

Several classical control charts in SPC are mentioned in Chapter 1. Among them, CUSUM, HQK and Hawkins and Zamba are most representative and most relevant to the method used in our Bayesian SPC in time series model. In this chapter, we give a brief introduction to these three SPC methods and discuss their connection with our new method.

## 4.1 CUSUM

CUSUM is a famous control chart proposed by E. S. Page in the 1950's. Originally this method dealt with an i.i.d normal sequence in that it takes the record of a value $S$ defined as:

$$S_0 = 0, \quad \text{and} \quad S_{n+1} = \max(0, S_n + X_n - k_n)$$

where $k_n$ is a reference value defined by the user. The alarm is given at time $n$ if $S_n > C_\alpha$, where $C_\alpha$ is a threshold used to adjust average run length (ARL). Essentially, the idea of CUSUM is to give alarm at time $n$ as soon as $n$ satisfies:

$$\max_{1 \le t \le n} \prod_{i=t}^{n} \frac{f_1(X_i)}{f_0(X_i)} \ge c_\alpha$$

where $c_\alpha$ is a constant tuning parameter, $f_0$ is the in control distribution, and $f_1$ is the out of control distribution. This equation is essentially the same as:

$$\frac{\prod_{i=1}^{t} f_0(x_i) \prod_{i=t+1}^{T} f_1(X_i)}{\prod_{i=1}^{T} f_0(X_i)} < c_\alpha \qquad \text{for every } 0 < t \leq T, \quad T < n$$

$$\text{and } \frac{\prod_{i=1}^{t} f_0(x_i) \prod_{i=t+1}^{n} f_1(X_i)}{\prod_{i=1}^{n} f_0(X_i)} \geq c_\alpha \qquad \text{for some } 0 < t \leq n \tag{4.1}$$

In our Bayesian SPC methodology, if the stochastic process is i.i.d normal, and we know the real means and variances before and after the change point, i.e, the prior distribution for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ concentrates at a point, and for each time $n$ when we have $n$ observations, $p_t = 1/(n-1+c_\alpha)$ for $0 < t < n$ and $p_n = c_\alpha/(n-1+c_\alpha)$, the posterior distributions for change point $T$ are:

$$f(T|\boldsymbol{X}_1^n) \propto p_T \prod_{i=1}^{T} f_0(x_i) \prod_{i=T+1}^{n} f_1(X_i)$$

where $f_0$ and $f_1$ are density functions for $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. The SPC will give alarm at time $n$ if

$$f(T|\boldsymbol{X}_1^n) < f(n|\boldsymbol{X}_1^n)$$

i.e.

$$p_T \prod_{i=1}^{T} f_0(x_i) \prod_{i=T+1}^{n} f_1(X_i) < p_n \prod_{i=1}^{n} f_0(x_i) \tag{4.2}$$

which is actually the same as the CUSUM formula (4.1).

Therefore, under the normality assumption, our Bayesian SPC method is an extension of the CUSUM under the ARMA model and Bayesian framework, and the CUSUM is a special case when we know the in control and out of control parameters, while the prior of $p_T$ is specially selected so that it is flat over every possible change point, and the ratio of $p_T/p_n$ is constant.

## 4.2 HQK

Compared to the CUSUM scheme, HQK handled the case completely in the other end: no information at all is available for in control and out of control parameters. This paper proposed a method that utilizes the two sample t-test statistic:

$$t_{T,n} = \sqrt{\frac{T(n-T)}{n}} \frac{\bar{X}^- - \bar{X}^+}{\sqrt{V_{T,n}/(n-2)}}$$

where $\bar{X}^- = \sum_{i=1}^{T} X_i/T$, $\bar{X}^+ = \sum_{i=T+1}^{n} X_i/(n-T)$, and

$$V_{T,n} = \sum_{i=1}^{T}(X_i - \bar{X}^-)^2 + \sum_{i=T+1}^{n}(X_i - \bar{X}^+)^2$$

The stopping rule in HQK consists of a control limit sequence $h_{n,\alpha}$ computed through simulation to maintain a constant rejection rate under in control status, i.e, alarm is given as soon as $t_{T,n} > h_{n,\alpha}$ for some $T < n$. In fact, the square of $t$ statistic is an F-statistic:

$$t_{T,n}^2 = F_{T,n} = \frac{V_n - V_{T,n}}{V_{T,n}/(n-2)}$$

where

$$V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = V_{T,n} + \frac{T(n-T)}{n}(\bar{X}^- - \bar{X}^+)^2$$

for every $T$ and $n$, so the stopping rule can be rewritten as

$$\frac{V_{T,n}}{V} \leq (1 + \frac{h_{n,\alpha}}{n-2})^{-1} \tag{4.3}$$

In our Bayesian framework, if we only consider a mean change in i.i.d scenario, the posterior distribution of the change point $T$ is

$$f(T|\boldsymbol{X}_1^n) \propto \frac{p_T V_{T,n,\tau_1,\tau_2,b}^{-\frac{n}{2}-a}}{\sqrt{(T+\tau_1^2)(n-T+\tau_2^2)}} \qquad \text{for } T < n \text{ and } f(n|\boldsymbol{X}_1^n) \propto \frac{p_n V_{n,\tau_1,b}^{-\frac{n}{2}-a}}{\sqrt{(n+\tau_1^2)(\tau_2^2)}}$$

where

$$V_{T,n,\tau_1,\tau_2,b} = (S_{2,T}^- + \tau_1^2\mu_{10}^2 - \frac{(S_{1,T}^- + \tau_1^2\mu_{10})^2}{T+\tau_1^2} + S_{2,T}^+ + \tau_2^2\mu_{20}^2 - \frac{(S_{1,T}^+ + \tau_2^2\mu_{20})^2}{n-T+\tau_2^2}) + 2b$$

$$V_{n,\tau_1,b} = (S_{2,n}^- + \tau_1^2\mu_{10}^2 - \frac{(S_{1,n}^- + \tau_1^2\mu_{10})^2}{n+\tau_1^2}) + 2b$$

According to our stopping rule, an alarm is given as soon as:

$$f(T|\boldsymbol{X}_1^n) > f(n|\boldsymbol{X}_1^n) \qquad \text{for some } T$$

i.e.

$$\frac{V_{T,n,\tau_1,\tau_2,b}}{V_{n,\tau_1,b}} \leq \left(\frac{p_T^2(n+\tau_1^2)\tau_2^2}{p_n^2(T+\tau_1^2)(n-T+\tau_2^2)}\right)^{1/(n+2a)} \tag{4.4}$$

It is not hard to see that

$$\lim_{\tau_1\to 0,\tau_2\to 0,b\to 0} V_{T,n,\tau_1,\tau_2,b} = V_{T,n} \qquad a.s$$

$$\lim_{\tau_1\to 0,b\to 0} V_{n,\tau_1,b} = V_n \qquad a.s$$

Therefore in a special case where $a = b = \tau_1 = 0$, we can pick up a $\tau_2$ that is very close to 0 so the left hand side is arbitrarily close to that of HQK, and right hand side can be adjusted through $p_T$ and $p_n$ so that it is exactly the same as that of HQK.

In summary, we can find a setting in our Bayesian SPC scheme so that it is close to HQK.

## 4.3 Hawkins and Zamba: Shift of Variance

Hawkins and Zamba (2005) proposed a method to detect the shift of variance, assuming an i.i.d normal sequence with unknown $\mu$ and $\sigma$, and a change point model:

$$X_t \sim \begin{cases} N(\mu_1, \sigma_1^2) & 1 \leq t \leq T \\ N(\mu_2, \sigma_2^2) & T < t \leq n \end{cases}$$

They assume potential changes of mean and variance, but only care about the change of $\sigma$, i.e, the hypothesis test of the likelihood ratio test is:

$$H_0: \ \sigma_1 = \sigma_2 \qquad\qquad H_1: \ \sigma_1 \neq \sigma_2$$

Hawkins and Zamba (2005) used the likelihood ratio test statistic:

$$G_{T,n} = [(T-1)\log(\frac{V_n/(n-2)}{V_{0,T}/(T-1)}) + (n-T-1)\log(\frac{V_n/(n-2)}{V_{T,n}/(n-T-1)})]/C$$

where $V_n = \sum_{i=1}^n (X_i - \bar{X})^2$,

$$C = 1 + [(k-1)^{-1} + (n-k-1)^{-1} - (n-2)^{-1}]/3$$

$$V_{k,m} = \sum_{i=k+1}^{m} (X_i - \frac{1}{m-k}\sum_{i=k+1}^{m} X_i)^2 \qquad \text{for every } k \text{ and } m.$$

A tuning parameter $h_{n,\alpha}$ is set up such that the alarm is given as soon as $G_{T,n} > h_{n,\alpha}$ for some $2 \leq T \leq n-2$.

Hawkins and Zamba's approach essentially considers change for both the mean and variance, since they use the separate sample means for both left and right segments. However, only variance change will be detected. Later in chapter 5, we will see that only a mean change in i.i.d standard normal sequence will not trigger reaction of this method at all.

In our Bayesian model considering change of both mean and variance, if the prior set up is $a_1 = a_2 = 1$ and $b_1 = b_2 = \tau_1 = \tau_2 = 0$, the model for change of both mean and variance has posterior distribution:

$$f(T|X) \propto p_T T^{-\frac{1}{2}}(n-T)^{-\frac{1}{2}}\Gamma(\frac{T-1}{2})\Gamma(\frac{n-T-1}{2})V_{0,T}^{-\frac{T-1}{2}}V_{T,n}^{-\frac{n-T-1}{2}}$$

and

$$f(n|X) \propto p_n n^{-\frac{1}{2}}\Gamma(\frac{n-1}{2})V_n^{-\frac{n-1}{2}}$$

The alarm is given as soon as $f(T|X) > f(n|X)$ for some $T < n$, i.e.

$$\frac{V_n^{n-1}}{V_{0,T}^{T-1}V_{T,n}^{n-T-1}} > \frac{p_n^2 T(n-T)[\Gamma(\frac{n-1}{2})]^2}{p_T^2 n[\Gamma(\frac{T-1}{2})\Gamma(\frac{n-T-1}{2})]^2}$$

which is the same as

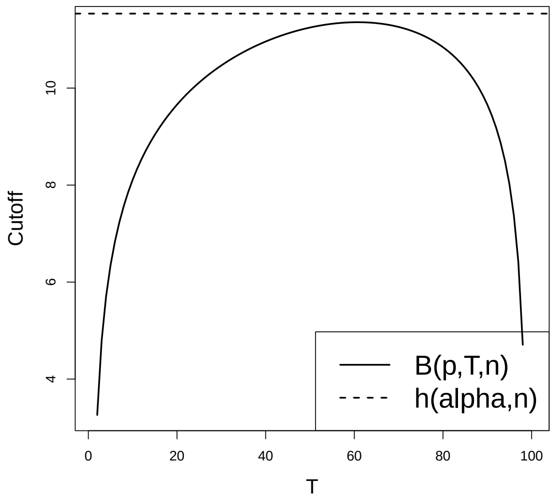
$$[(T-1)\log(\frac{V_n}{V_{0,T}}) + (n-T-1)\log(\frac{V_n}{V_{T,n}})] > \log(\frac{p_n^2 T(n-T)[\Gamma(\frac{n-1}{2})]^2}{p_T^2 n[\Gamma(\frac{T-1}{2})\Gamma(\frac{n-T-1}{2})]^2 V_n})$$

i.e

$$[(T-1)\log(\frac{V_n/(n-2)}{V_{0,T}/(T-1)}) + (n-T-1)\log(\frac{V_n/(n-2)}{V_{T,n}/(n-T-1)})]/C(T,n)$$

$$> \log(\frac{p_n^2 T(n-T)[\Gamma(\frac{n-1}{2})]^2(T-1)^{T-1}(n-T-1)^{n-T-1}}{p_T^2 n[\Gamma(\frac{T-1}{2})\Gamma(\frac{n-T-1}{2})(n-2)^{n-2}]^2 V_n})/C(T,n) \qquad (4.5)$$

We define the right hand side of (4.5) to be $B_{p,T,n}$. At $n = 100$, we have the comparison of $B_{p,T,n}$ with $h_{\alpha,n}$:




assuming $p_T = p(1-p)^{n-T-1}$ for $1 < T < n$, $p_n = (1-p)^{n-1}$, $p = \alpha = 0.01$ and $V_n \approx n$. Moreover, if we can freely select $p_T$ and $p_n$, for every $n$, we can find proper $p_T$ and $p_n$ such that the Bayesian SPC methodology is close to the Hawkins and Zamba method, assuming $V_n = O_p(n)$.

In summary, in some special settings of the prior information, especially those assuming plat prior for some parameters of interest, and some special settings of the $\pi(T)$, the Bayesian SPC method can have a decision rule close to the CUSUM, HQK and Hawkins and Zamba method.

# Chapter 5

# Simulation Study

## 5.1 Conditional Alarm Rate

In this section, we will consider five models of most interest:

1. The model assuming i.i.d normal sequence with mean change only.
2. The model assuming i.i.d normal sequence with variance change only.
3. The model assuming i.i.d normal sequence with mean and/or variance change only.
4. The model assuming $AR(1)$ sequence with change of mean, variance or autoregression coefficient.
5. The model assuming $ARMA(1,1)$ sequence with change of mean, variance, autoregression coefficient or moving average coefficient.

Our simulation will focus on two aspects. The first one is the conditional alarm rate. Hawkin, Qiu and Kang (2003) suggests a reasonable way of selecting the control limit sequence $h_{n,\alpha}$. This sequence is selected so that the conditional probability of false alarm (in the healthy process without any change point, the probability of an alarm, given that there is no alarm given before, also referred as hazard or alarm rate in HQK) is a specified constant $\alpha$. In this section, we will investigate the behavior of conditional rejection probability under the five basic models.

Another measure that people often focus on in SPC problems is average run length (ARL). ARL can be used to measure the behavior of an SPC method. In most cases, it is desirable to have a large in control ARL and a short out of control ARL. Moreover, it is necessary to know the mechanism with which ARL can be adjusted by the hyperparameters. In the following simulations, we will also show how those hyperparameters can affect ARL, and we will compare the ARL curves of different models versus magnitude of change.

### 5.1.1 I.I.D Normal Sequence

The i.i.d normal sequence is the simplest scenario in this thesis. In the notation in chapter 2, an i.i.d normal sequence has $p = q = 0$, and $\boldsymbol{\phi} = \boldsymbol{\theta} = 0$.

**Change of Mean Only**

According to chapter 2.4.5, the choice of prior distribution is:

$$\pi(\mu_1, \mu_2, \sigma^2, T) \propto (\frac{1}{\sigma^2})^{a+2} \exp(-\frac{\tau_1^2}{2\sigma^2}(\mu_1 - \mu_{10})^2 - \frac{\tau_2^2}{2\sigma^2}(\mu_2 - \mu_{20})^2 - \frac{b}{\sigma^2})\pi(T)$$

i.e.

$$\sigma^2 \sim IG(a, b), \quad \mu_1|\sigma^2 \sim N(\mu_{10}, \frac{\sigma^2}{\tau_1^2}), \quad \mu_2|\sigma^2 \sim N(\mu_{20}, \frac{\sigma^2}{\tau_2^2})$$

Therefore the $\tau_1$ and $\tau_2$ are the accuracy of our prior estimate of $\mu_1$ and $\mu_2$. We have an explicit analytical expression for the posterior distribution of $T$:

$$f(T|X_1^n) \propto p_T(T + \tau_1^2)^{-\frac{1}{2}}(n - T + \tau_2^2)^{-\frac{1}{2}}$$
$$\cdot[\frac{1}{2}(S_{2,T}^- + \tau_1^2\mu_{10}^2 - \frac{(S_{1,T}^- + \tau_1^2\mu_{10})^2}{T + \tau_1^2} + S_{2,T}^+ + \tau_2^2\mu_{20}^2 - \frac{(S_{1,T}^+ + \tau_2^2\mu_{20})^2}{n - T + \tau_2^2}) + b]^{-\frac{n}{2}-a}$$

where

$$S_{1,T}^- = \sum_{i=1}^{T} X_i, \quad S_{2,T}^- = \sum_{i=1}^{T} X_i^2, \quad S_{1,T}^+ = \sum_{i=T+1}^{n} X_i, \quad S_{2,T}^+ = \sum_{i=T+1}^{n} X_i^2$$

We carry out the following simulation study:

1. Generate an i.i.d standard normal sequence $\{X_i\}_{i=1}^{300}$.

2. Set up prior information: $a = 3$ and $b = 2$, i.e. $\sigma^2 \sim IG(3, 2)$. This is a prior with $E(\sigma^2) = 1$ and $Var(\sigma^2) = 2$. $\mu_{10} = \mu_{20} = 0$. $\tau_1 = \tau_2 = 1$.

3. Read in the first observation $X_1$. $n = 1$.

4. If $n \leq 300$, read in the next observation $X_{n+1}$. $n$ increases by 1.

5. Set up new prior distribution for $T$: $p_T = p(1-p)^{n-T}$, where $p = 0.01$. $p_n = \sum_{k=n}^{\infty} p_k = (1-p)^{n-1}$. Then calculate $f(T|X_1, ..., X_n)$.

6. If $f(n|X_1, ..., X_n) \geq f(T|X_1, ..., X_n)$ for any $T = 1, 2, ..., n-1$, go to step 4. Else give alarm, take record of the current $n$, and go to step 1.

The above procedure is repeated for 10,000 times. The Figure 5.1 shows the conditional probability which shows the estimated alarm rate given that there is no alarm before.

It is worth noting that the estimated conditional probability has increasing variance as $n$ increases. Therefore we use kernel smooth to reduce its variance, especially for large $n$. From this graph, we can observe a roughly constant rejection rate for $n < 300$, and a slight fluctuation for $0 < n < 100$. Specifically, it takes the process several steps (for this particular example, 10 steps) to reach its stable rejection rate.

**Change of Variance**

Account to Chapter 2.4.5, The choice of prior distribution is

$$\pi(\mu, \sigma_1^2, \sigma_2^2, T) \propto (\frac{1}{\sigma_1^2})^{a_1+1} exp(-\frac{b_1}{\sigma_1^2})(\frac{1}{\sigma_2^2})^{a_2+1} exp(-\frac{b_2}{\sigma_2^2}) \exp[\frac{(\mu-\mu_0)^2}{2\sigma_0^2}] \cdot p_T$$

i.e.

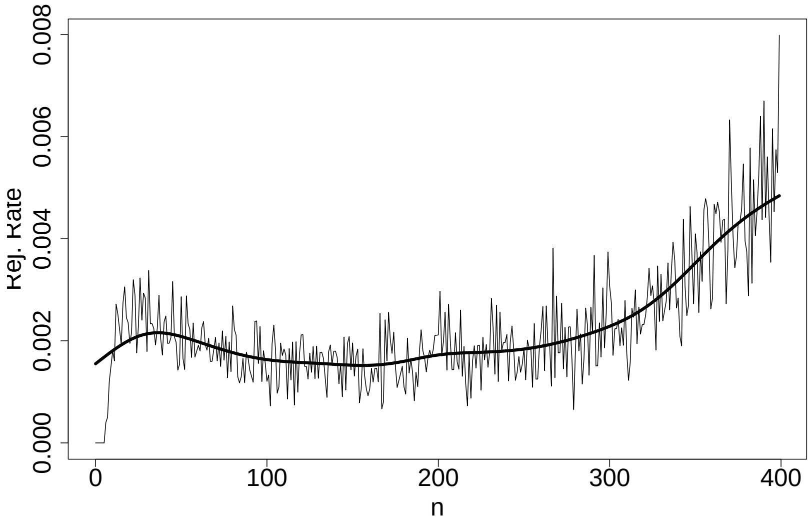
$$\sigma_1^2 \sim IG(a_1, b_1), \quad \sigma_2^2 \sim IG(a_2, b_2), \quad \mu \sim N(\mu_0, \sigma_0^2), \quad \mu \perp \sigma_1^2, \sigma_2^2$$

The posterior distribution is:

$$f(\mu, T|X) \propto p_T \Gamma(a_1 + \frac{T}{2}) \Gamma(a_2 + \frac{n-T}{2})(\frac{1}{2}\sum_{i=1}^{T}(X_i - \mu)^2 + b_1)^{-(a_1+\frac{T}{2})}$$

$$\cdot(\frac{1}{2}\sum_{i=T+1}^{n}(X_i - \mu)^2 + b_2)^{-(a_2+\frac{n-T}{2})} \exp[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2]$$

Figure 5.1: Conditional rejection rate for mean change only model




and $f(T|X)$ can be evaluated through Monte Carlo integration.
Settings of prior and model information are as follows:

1. Generate an i.i.d standard normal sequence $\{X_i\}_{i=1}^{300}$.

2. Set up prior information: $a_1 = a_2 = 3$ and $b_1 = b_2 = 2$, i.e. $\sigma_2^2, \sigma_1^2 \sim IG(3, 2)$. $\mu_0 = 0$ and $\sigma_0 = 5$.

3. Read in the first observation $X_1$. $n = 1$.

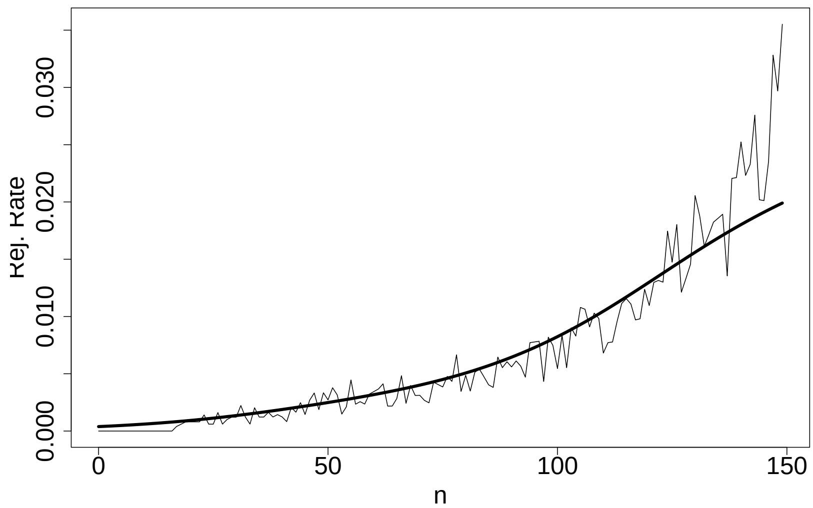
4. If $n \leq 300$, read in the next observation $X_{n+1}$. $n$ increases by 1.

5. Set up new prior distribution for $T$: $p_T = p(1-p)^{n-T}$, where $p = 0.015$. $p_n = \sum_{k=n}^{\infty} p_k = (1-p)^{n-1}$. Then calculate $f(T|X_1, ..., X_n)$.

6. If $f(n|X_1, ..., X_n) \geq f(T|X_1, ..., X_n)$ for any $T = 1, 2, ..., n-1$, go to step 4. Else give alarm, take record of the current $n$, and go to step 1.

This procedure is repeated for 5,000 times. The Figure 5.2 shows the conditional probability which shows the estimated conditional probability of alarm given that there is no alarm before. From this graph we observe an increase rejection rate, which is different from the previous model, and the number of warm up steps is also more than the previous model.

Figure 5.2: Conditional rejection rate for variance change only model




**Change of Mean and Variance**

According to chapter 2, the choice of prior is:

$$\pi(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, T) \propto (\frac{1}{\sigma_1^2})^{a_1+2} \exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1 - \mu_{10})^2 - \frac{b_1}{\sigma_1^2})$$

$$\cdot(\frac{1}{\sigma_2^2})^{a_2+2} \exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2 - \mu_{20})^2 - \frac{b_2}{\sigma_2^2})\pi(T)$$

i.e.

$$\sigma_1^2 \sim IG(a_1, b_1), \quad \sigma_2^2 \sim IG(a_2, b_2), \quad \mu_1|\sigma_1^2 \sim N(\mu_{10}, \frac{\sigma_1^2}{\tau_1^2}), \quad \mu_2|\sigma_2^2 \sim N(\mu_{20}, \frac{\sigma_2^2}{\tau_2^2})$$

and the posterior distribution of $T$ is explicitly available:

$$f(T|X) \propto p_T(T + \tau_1^2)^{-\frac{1}{2}}(n - T + \tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{T}{2} + a_1)\Gamma(\frac{n-T}{2} + a_2)$$

$$\cdot[\frac{1}{2}(S^-_{2,T}+\tau_1^2\mu_{10}^2+2b_1-\frac{(S^-_{1,T}+\tau_1^2\mu_{10})^2}{T+\tau_1^2})]^{-\frac{T}{2}-a_1}[\frac{1}{2}(S^+_{2,T}+\tau_2^2\mu_{20}^2+2b_2-\frac{(S^+_{1,T}+\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})]^{-\frac{n-T}{2}-a_2}$$

specifically:

$$f(n|X) \propto p_n(n+\tau_1^2)^{-\frac{1}{2}}(\tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{n}{2}+a_1)\Gamma(a_2)[\frac{1}{2}(S^-_{2,n}+\tau_1^2\mu_{10}^2+2b_1-\frac{(S^-_{1,n}+\tau_1^2\mu_{10})^2}{n+\tau_1^2})]^{-\frac{n}{2}-a_1}b_2^{-a_2}$$

The settings of prior and model information are as follows:

1. $p_T = p(1-p)^{n-T}$, where $p = 0.01$. When first $n$ observations are available, $p_n = \sum_{k=n}^{\infty} p_k = (1-p)^{n-1}$.

2. $a_1 = a_2 = 3$, $b_1 = b_2 = 2$, $\mu_{10} = \mu_{20} = 0$, and $\tau_1 = \tau_2 = 1$.

The procedure is repeated 10,000 times. The Figure 5.3 shows the conditional probability which shows the estimated conditional probability of alarm given that there is no alarm before.

This graphs demonstrates similar pattern as the mean change only model: the rejection rate is stable between $0 < n < 300$, and there is a small fluctuation for $0 < n < 100$. It also takes a few steps for the process to reach the stable rejection rate of 0.002.

### 5.1.2 $AR(1)$ Model

An $AR(1)$ model with change point at $T$ can be written as:

$$Y_n = \begin{cases} \phi_1 Y_{n-1} + X_n & n \leq T \\ \phi_2 Y_{n-1} + X_n & n > T \end{cases}$$

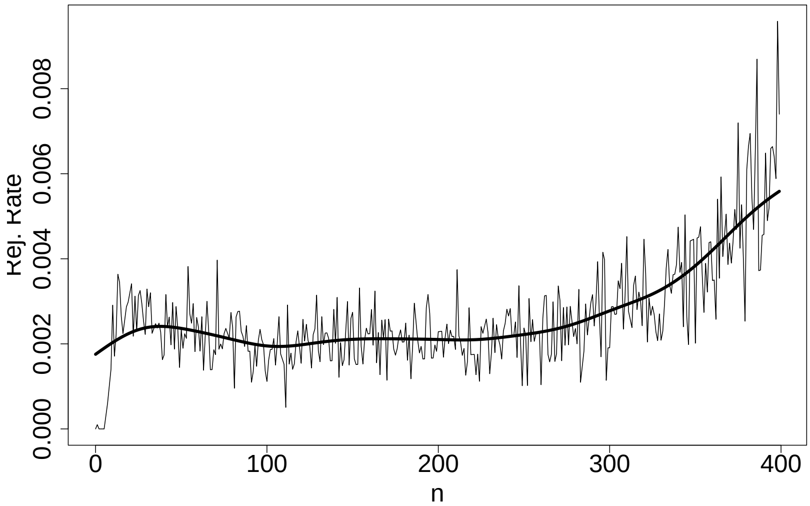
and

$$X_n \sim \begin{cases} N(\mu_1, \sigma_1^2) & n \leq T \\ N(\mu_2, \sigma_2^2) & n > T \end{cases}$$

i.e, we allow change of any parameter. We use $Y_n$ to represent the $AR(1)$ sequence so that $X_n$ can be consistently used to denote i.i.d random variables. According to Chapter 2.4, the choice of prior is

$$\pi(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, T) \propto (\frac{1}{\sigma_1^2})^{a_1+2}\exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1-\mu_{10})^2 - \frac{b_1}{\sigma_1^2})$$

Figure 5.3: Conditional rejection rate for mean and variance change only model




$$\cdot(\frac{1}{\sigma_2^2})^{a_2+2}\exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2-\mu_{20})^2-\frac{b_2}{\sigma_2^2})\pi(T)$$

while $\phi_1 \sim Beta(\alpha_1, \beta_1)$ and $\phi_2 \sim Beta(\alpha_2, \beta_2)$ are independent. The posterior distribution of $T$ is:

$$f(\phi_1,\phi_2,T|Y) \propto p_T(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{T}{2}+a_1)\Gamma(\frac{n-T}{2}+a_2)$$

$$\cdot[\frac{1}{2}(S_{2,T}^-+\tau_1^2\mu_{10}^2+2b_1-\frac{(S_{1,T}^-+\tau_1^2\mu_{10})^2}{T+\tau_1^2})]^{-\frac{T}{2}-a_1}[\frac{1}{2}(S_{2,T}^++\tau_2^2\mu_{20}^2+2b_2-\frac{(S_{1,T}^++\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})]^{-\frac{n-T}{2}-a_2}$$

where

$$S_{1,T}^- = \sum_{i=2}^{T}(Y_i-\phi_1 Y_{i-1}), \quad S_{2,T}^- = \sum_{i=2}^{T}(Y_i-\phi_1 Y_{i-1})^2,$$

$$S_{1,T}^+ = \sum_{i=T+1}^{n}(Y_i-\phi_2 Y_{i-1}), \quad S_{2,T}^+ = \sum_{i=T+1}^{n}(Y_i-\phi_2 Y_{i-1})^2$$

and

$$f(T|Y) \propto \int_{\phi_1,\phi_2} f(\phi_1,\phi_2,T|Y)(1+\phi_1)^{\alpha_1}(1-\phi_1)^{\beta_1}(1+\phi_2)^{\alpha_2}(1-\phi_2)^{\beta_2}d\phi_1 d\phi_2$$

For this 2-dimensional integration, we can evaluate the integration by Monte Carlo simulation:

$$f(T|Y) \approx \sum_{j=1}^{N} f(\phi_{1j}, \phi_{2j}, T|Y)$$

where $(\phi_{1j}, \phi_{2j})$ are samples from $Beta(\alpha_1, \beta_1) \times Beta(\alpha_2, \beta_2)$. The sample size $N$ is set to be 100 in our simulation.
We carry out the following simulation study:

1. Generate an $AR(1)$ sequence with length=300, $\mu = 0$, $\sigma = 1$ and $\phi = 0.5$.

2. $p_T = p(1-p)^{n-T}$, where $p = 0.01$. When first $n$ observations are available, $p_n = \sum_{T=n}^{\infty} p_T = (1-p)^{n-1}$.

The test is carried with hyperparameters:

1. $a_1 = a_2 = 3$ and $b_1 = b_2 = 2$, i.e. $\sigma_1^2, \sigma_2^2$ i.i.d $\sim IG(3,2)$. This is a prior with $E(\sigma_j^2) = 1$ and $Var(\sigma_j^2) = 1$, $j = 1, 2$.

2. $\mu_{10} = \mu_{20} = 0$. $\tau_1 = \tau_2 = 1$.

3. $\alpha_1 = \alpha_2 = 1$ and $\beta_1 = \beta_2 = 1$, i.e. $\phi_1, \phi_2$ i.i.d $\sim Unif(-1, 1)$

4. Monte Carlo sample size $N = 100$.

The procedure is repeated 5,000 times. The Figure 5.4 shows the conditional probability which shows the estimated conditional probability of alarm given that there is no alarm before.

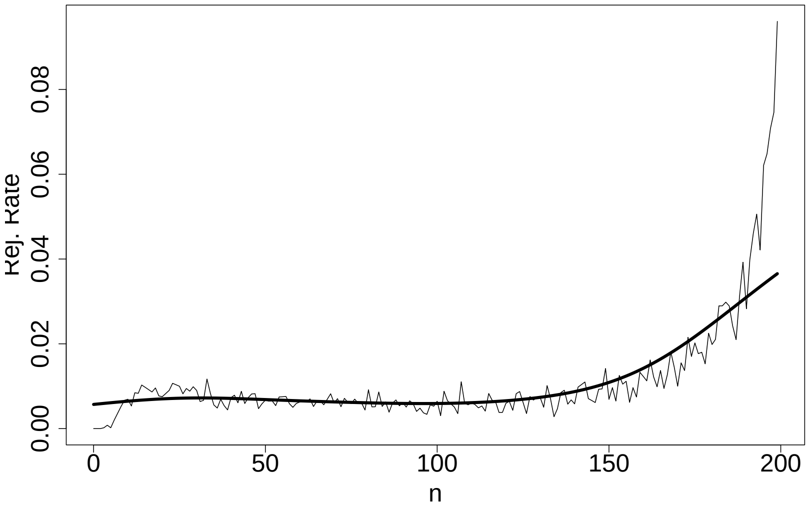
This graph demonstrates similar pattern as mean and variance change only model: A stable rejection rate around 0.008 for $0 < n < 150$, a rocketing rejection rate after $n > 200$, and a few steps at beginning for the rejection rate to reach this level.

### 5.1.3 ARMA(1,1) Model

Under an ARMA(1,1) model,

$$Y_n = \begin{cases} \phi_1 Y_{n-1} + X_n + \theta_1 X_{n-1} & n \le T \\ \phi_2 Y_{n-1} + X_n + \theta_2 X_{n-1} & n > T \end{cases}$$

Figure 5.4: Conditional rejection rate for AR(1) model




where $Y_n$ are observations, $\phi \in (-1, 1)$ and $\theta \in (-1, 1)$ guarantee the stationarity and invertibility, and

$$X_n \sim \begin{cases} N(\mu_1, \sigma_1^2) & n \leq T \\ N(\mu_2, \sigma_2^2) & n > T \end{cases}$$

In the ARMA(1,1) model, we need to assume initial values, for example, here we assume $Y_1 = X_1$ and the inductive equation starts to work from $n = 2$. Then it is easy to solve the $X_n$ by:

$$X_n = \begin{cases} Y_1 & n = 1 \\ Y_n - \phi_1 Y_{n-1} - \theta_1 X_{n-1} & 1 < n \leq T \\ Y_n - \phi_2 Y_{n-1} - \theta_2 X_{n-1} & n > T \end{cases} \tag{5.1}$$

According to the previous result, the choice of prior distribution is:

$$\pi(\mu_1, \sigma_1^2, \phi_1, \theta_1, \mu_2^2, \sigma_2, \phi_2, \theta_2) \propto (1 - \phi_1)^{\alpha_1} (1 + \phi_1)^{\beta_1} (1 - \phi_2)^{\alpha_2} (1 + \phi_2)^{\beta_2}$$

$$\cdot (1 - \theta_1)^{\gamma_1} (1 + \theta_1)^{\delta_1} (1 - \theta_2)^{\gamma_2} (1 + \theta_2)^{\delta_2}$$

$$\cdot (\frac{1}{\sigma_1^2})^{a_1+2} \exp(-\frac{\tau_1^2}{2\sigma_1^2}(\mu_1 - \mu_{10})^2 - \frac{b_1}{\sigma_1^2})(\frac{1}{\sigma_2^2})^{a_2+2} \exp(-\frac{\tau_2^2}{2\sigma_2^2}(\mu_2 - \mu_{20})^2 - \frac{b_2}{\sigma_2^2})$$

The posterior distribution of $T$ is:

$$f(\phi_1,\theta_1,\phi_2,\theta_2,T|Y) \propto p_T(T+\tau_1^2)^{-\frac{1}{2}}(n-T+\tau_2^2)^{-\frac{1}{2}}\Gamma(\frac{T}{2}+a_1)\Gamma(\frac{n-T}{2}+a_2)$$

$$[\frac{1}{2}(S_{2,T}^{-}+\tau_1^2\mu_{10}^2+2b_1-\frac{(S_{1,T}^{-}+\tau_1^2\mu_{10})^2}{T+\tau_1^2})]^{-\frac{T}{2}-a_1}\cdot[\frac{1}{2}(S_{2,T}^{+}+\tau_2^2\mu_{20}^2+2b_2-\frac{(S_{1,T}^{+}+\tau_2^2\mu_{20})^2}{n-T+\tau_2^2})]^{-\frac{n-T}{2}-a_2}$$

where

$$S_{1,T}^{-}=\sum_{i=2}^{T}X_i,\quad S_{2,T}^{-}=\sum_{i=2}^{T}X_i^2,\quad S_{1,T}^{+}=\sum_{i=T+1}^{n}X_i,\quad S_{2,T}^{+}=\sum_{i=T+1}^{n}X_i^2$$

and the $X_i$'s are solved inductively in (5.1). If $\phi$ and $\theta$ are unknown, we can handle the integration by Monte Carlo simulation.
We carry out the following simulation study:

1. Generate an $ARMA(1,1)$ sequence with length=300, $\mu=0$, $\sigma=1$, $\phi=0$ and $\theta=0$.

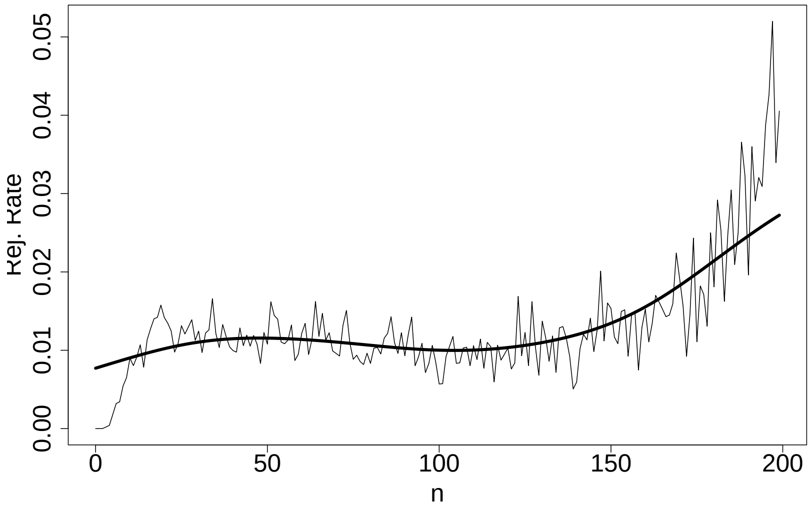
2. $p_T = p(1-p)^{n-T}$, where $p=0.02$. When first $n$ observations are available, $p_n=\sum_{T=n}^{\infty}p_T=(1-p)^{n-1}$.

The hyperparameters are all the same as AR case except $\gamma_1=\delta_1=\gamma_2=\delta_2=1$. The Figure 5.5 shows the conditional rejection rate when applying this ARMA(1,1) model to test an i.i.d standard normal sequence. This graph demonstrates similar pattern as AR(1) model: A stable rejection rate around 0.01 for $0<n<120$, and a few (about 20) steps at beginning for the rejection rate to reach this level.

## 5.2 Average Run Length

Average Run Length (ARL) is an important index of the behavior of a quality control method. Two types of ARL are usually studied: in control ARL (IC ARL) and out of control ARL (OOC ARL). In control ARL is defined as the expected number of observations until an alarm is given, provided that the sequence is in control all the time. Out of control ARL is defined as the expected number of observations from a change point to the alarm. The in control ARL measures the type I error, while out of control ARL measures the type II error. In practice, people anticipate long in control ARL and

Figure 5.5: Conditional rejection rate for ARMA(1,1) model




short out of control ARL.

As were mentioned in chapter 2, we can have 7 change point models for general $ARMA(p, q)$ model. For some particular cases, such as the i.i.d or simple $AR(1)$ case, various change point models are available. However, most of them are very alike each other, different in only a few details. Therefore in this section, we will mainly consider five methods: $\mu$ change only (M model) with i.i.d model assumption, $\sigma$ change only (S model) with i.i.d model assumption, $\mu$ and $\sigma$ change only (MS model) with i.i.d model assumption, $\mu$, $\sigma$ and $\phi$ change only (MSP model) with $AR(1)$ model assumption, and change of all parameters (MSPT model) with $ARMA(1, 1)$ model assumption.

### 5.2.1 In Control ARL

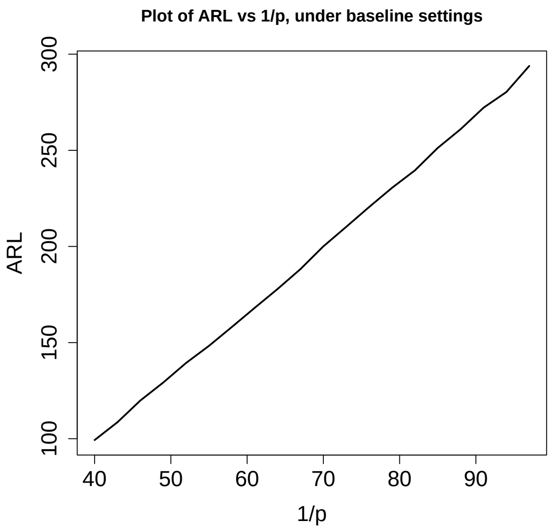
Because of the invariance of our method, without loss of generality, we only consider the special case in which in control status is the standard normal distribution. Since there are many hyperparameters, we focus on the study of the marginal relationship between ARL and $p$ under different accuracy levels of the setting of hyperparameters.

**M model:**

In the mean change only model, we first plot the in control average run length versus different settings of $p$, under the baseline setting: $\mu_{10} = \mu_{20} = 0$, $\tau_1 = \tau_2 = 1$, $a = 3$, $b = 2$, as shown in Figure 5.6.

Notice that the x-axis is $1/p$ instead of $p$. From the graph we can see an almost perfect

Figure 5.6: Plot of in control ARL vs $1/p$ for M model



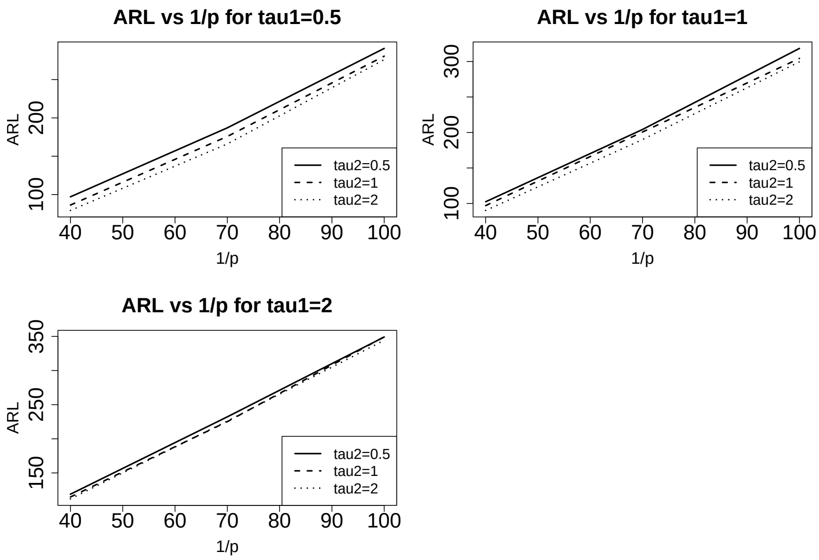

linear relationship with little curvature exists between the average run length and the reciprocal of $p$.

Recall that in the mean change only model, $\tau_1$ and $\tau_2$ represent the accuracy of the prior information for $\mu_1$ and $\mu_2$. Small $\tau_1$ and $\tau_2$ indicate the lack of precise information of $\mu_1$ and $\mu_2$ and large $\tau_1$ and $\tau_2$ show our strong belief of their ranges of value. In the following study, we examine how $\tau_1$, $\tau_2$ and $p$ jointly affect in control average run length. We carry out simulation study with the mean-change-only model on i.i.d sequence, with other settings remaining the same as baseline except that $\tau_1$ goes through $(0.5, 1, 2)$, $\tau_2$ goes through $(0.5, 1, 2)$ and $p$ goes through $(40, 70, 100)$. The numbers outside parenthesis are average run lengths, while the numbers in the parenthesis are the corresponding standard errors.

Table 5.1: ARL for different $p$, $\tau_1$ and $\tau_2$ in M model

| $\tau_1$ | $1/p$ | $\tau_2 = 0.5$ | $\tau_2 = 1$ | $\tau_2 = 2$ |
|---|---|---|---|---|
| | 40 | 97 (1.9) | 86.3 (1.5) | 79.3 (1.2) |
| 0.5 | 70 | 187 (3.8) | 175.8 (3.1) | 166 (2.5) |
| | 100 | 290.6 (5.8) | 280.5 (4.9) | 276 (3.6) |
| | 40 | 102.4 (2) | 97 (1.6) | 90.1 (1.2) |
| 1 | 70 | 204.2 (3.9) | 200.6 (3.2) | 190 (2.4) |
| | 100 | 318.4 (5.9) | 304.3 (5) | 299.5 (3.9) |
| | 40 | 118.8 (2.1) | 114.7 (1.7) | 112 (1.2) |
| 2 | 70 | 232.2 (4.2) | 225.1 (3.6) | 225.9 (2.6) |
| | 100 | 349.1 (6.5) | 349.4 (5.4) | 344.4 (4.1) |

Figure 5.7: Plot of in control ARL vs $1/p$ for M model



We plot this table in three graphs corresponding to different $\tau_1$ values for a better view in Figure 5.7.

From the table and the graphs, we see that

1. The linear relationship between in control ARL and $1/p$ is strong for all $\tau_1$ and $\tau_2$ combinations. It is hard to observe any curvature from the graphs.

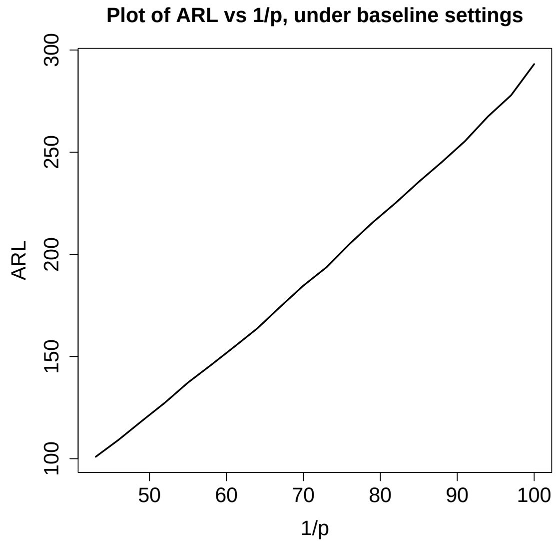
2. In control average run length is positively related to $\tau_1$, the precision of prior distribution of in control mean $\mu_1$ and negatively related to $\tau_2$, the precision of the prior distribution of out of control mean $\mu_2$. It is worth noting that we set up $\mu_{20} = 0$ before. If $\tau_2$ is large, the method tends to believe that there is a change point and after that, the mean of sequence is somewhere close to 0. If instead we set up $\mu_{20} \neq 0$, the relationship between average run length and $\tau_2$ will become positive.

3. We can see that the standard error of ARL estimates increases with $1/p$ and $\tau_1$ and decreases with $\tau_2$.

**MS model:**

In the mean and variance change only model, the plot of in control average run length versus different settings of $p$, under the baseline setting: $\mu_{10} = \mu_{20} = 0$, $\tau_1 = \tau_2 = 1$, $a_1 = a_2 = 3$, $b_1 = b_2 = 2$ is shown in Figure 5.8.

Figure 5.8: Plot of in control ARL vs $1/p$ for MS model




We observe a linear relationship with little curvature. The table and the plot of in control ARL with different $p$, $\tau_1$ and $\tau_2$ combinations is in Table 5.2 and Figure 5.9.

We have similar observations:

1. The linear trend between average run length and $1/p$ is strong with little curvature.

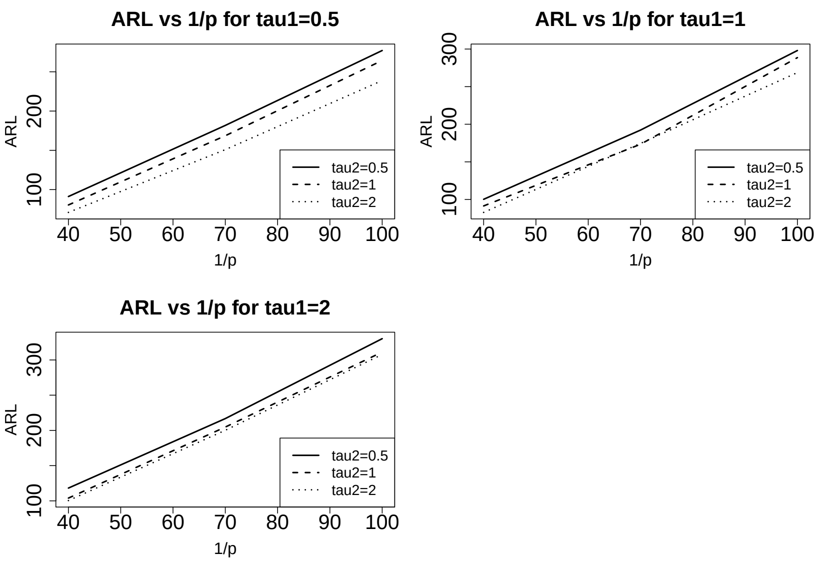
2. The average run length is positively related to $1/p$ and $\tau_1$, and negatively related to $\tau_2$.

3. The standard error of our ARL estimates increases with $1/p$ and $\tau_1$, and negatively related to $\tau_2$.

Table 5.2: ARL for different $p$, $\tau_1$ and $\tau_2$ in MS model

| $\tau_1$ | $1/p$ | $\tau_2 = 0.5$ | $\tau_2 = 1$ | $\tau_2 = 2$ |
|---|---|---|---|---|
| 0.5 | 40 | 91.2 (1.8) | 80.4 (1.5) | 71 (1.2) |
| | 70 | 181.7 (3.5) | 168.6 (3) | 150.9 (2.4) |
| | 100 | 277.1 (5.4) | 264.4 (4.6) | 238.9 (3.8) |
| 1 | 40 | 100.2 (1.9) | 91.4 (1.6) | 82.7 (1.2) |
| | 70 | 192.2 (3.8) | 173.4 (3.2) | 174.6 (2.5) |
| | 100 | 298.1 (5.8) | 288.6 (4.9) | 268.7 (4) |
| 2 | 40 | 118.1 (2.1) | 103.9 (1.8) | 100.4 (1.4) |
| | 70 | 216.8 (4.2) | 204.7 (3.5) | 200.2 (2.8) |
| | 100 | 330.2 (6.3) | 311.2 (5.6) | 307.8 (4.4) |

Figure 5.9: Plot of in control ARL vs $1/p$ for MS model

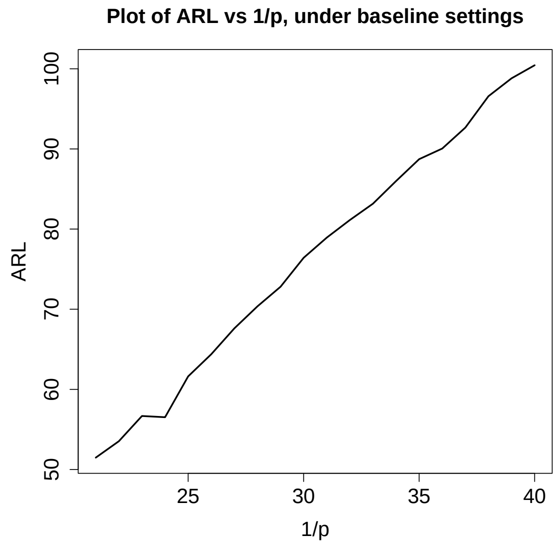


**S model:**

In the variance change only model, the plot of in control average run length versus different settings of $p$, under the baseline setting: $\mu_0 = 0$, $\sigma_0 = 5$, $a_1 = a_2 = 3$, $b_1 = b_2 = 2$ is in Figure 5.10.

From this graph, we still observe a strong linear relationship between ARL and $1/p$.

Figure 5.10: Plot of in control ARL vs $1/p$ for S model

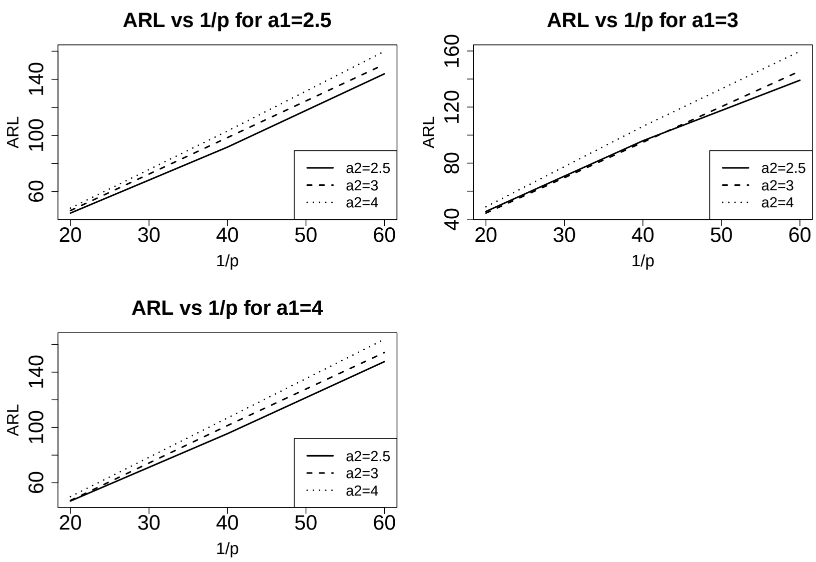



To examine how precision of prior distribution will affect average run length, we select pairs of $a_1$ and $a_2$ from 2.5, 3 and 4. Here we guarantee that $b_1 = a_1 - 1$ and $b_2 = a_2 - 1$ so that the mean of prior distribution for $\sigma_1^2$ and $\sigma_2^2$ will always be 1, and the standard deviation of their prior distribution change in 2, 1 and 0.5. The table of in control ARL and its plot are shown in Table 5.3 and Figure 5.11.

Table 5.3: ARL for different $p$, $a_1$ and $a_2$ in S model

| $a_1$ | $1/p$ | $a_2 = 2.5$ | $a_2 = 3$ | $a_2 = 4$ |
|---|---|---|---|---|
| 2.5 | 20 | 44.7 (0.4) | 46.5 (0.4) | 48.2 (0.4) |
| | 40 | 91.7 (0.9) | 98.4 (0.9) | 103.1 (0.9) |
| | 60 | 143.9 (1.6) | 150.7 (1.6) | 159.9 (1.5) |
| 3 | 20 | 45.5 (0.4) | 44.4 (0.3) | 48.8 (0.3) |
| | 40 | 96 (0.9) | 94.9 (0.9) | 106.2 (0.9) |
| | 60 | 139.1 (1.5) | 145.7 (1.6) | 159.8 (1.6) |
| 4 | 20 | 46.8 (0.3) | 47.2 (0.3) | 50 (0.3) |
| | 40 | 95.5 (0.9) | 101.3 (0.9) | 106.9 (0.9) |
| | 60 | 147.6 (1.5) | 154.2 (1.5) | 163.8 (1.5) |

Figure 5.11: Plot of in control ARL vs $1/p$ for S model



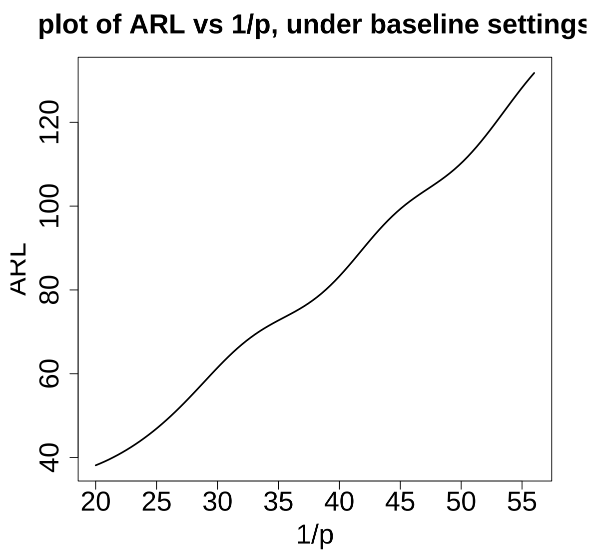
We have observations:

1. The linear relationship between ARL and $1/p$ is strong with little curvature.
2. The ARL is positively related to $1/p$, $a_1$, and $a_2$.
3. The standard error of our ARL estimate is positively related to $1/p$, but has little relationship with $a_1$ and $a_2$.

**MSP model:**

In the MSP model, the plot of in control average run length versus different settings of $p$, under the baseline setting: $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 1$, and $\beta_2 = 1$ is shown in graph 5.12

Figure 5.12: Plot of in control ARL vs $1/p$ for MSP model




Because of the simulation error, we apply kernel smooth to obtain a better view. From this graph, we can still observe a linear relationship between ARL and $1/p$. To examine how precision of prior distribution for $\phi$ will affect average run length, we select pairs of $\alpha_1$ and $\alpha_2$ from 1, 6 and 24. Here we guarantee that $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2$ so that the mean of prior distribution for $\phi_1$ and $\phi_2$ will always be 0, and the standard deviation of their prior distribution change in 0.28, 0.14 and 0.07. The table of ARL and its plot are shown in Table 5.4 and Figure 5.13.

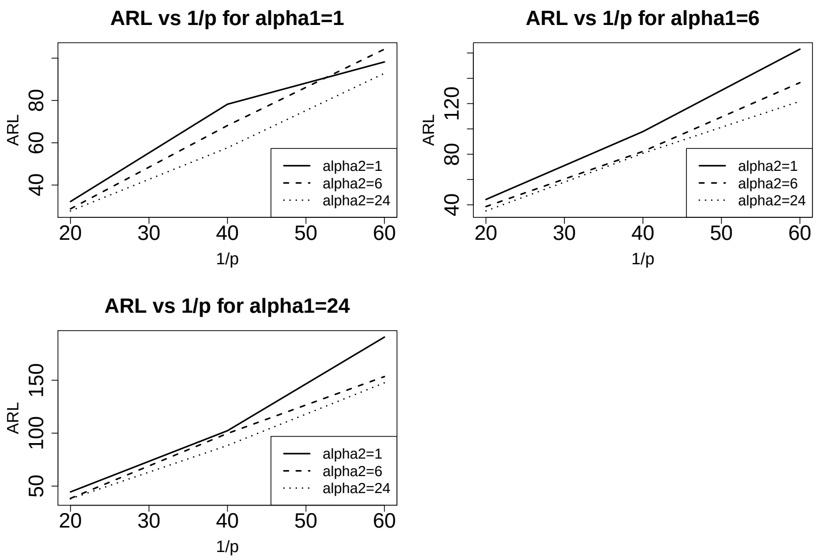
We have the following observations:

1. The ARL has a roughly linear relationship with $1/p$, with a little bit curvature.
2. The ARL is positively related to $1/p$ and $\alpha_1$, and negatively related to $\alpha_2$. Here $\alpha_1$ and $\alpha_2$ can be regarded as precision, because the variance of our prior of $\phi$ is proportional to $1/(2\alpha + 1)$.
3. The standard error of our ARL estimate is positively related to $1/p$ and $\alpha_1$, and negatively related to $\alpha_2$.

Table 5.4: ARL for different $p$, $\alpha_1$ and $\alpha_2$ in MSP model

| $\alpha_1$ | $1/p$ | $\alpha_2 = 1$ | $\alpha_2 = 6$ | $\alpha_2 = 24$ |
|---|---|---|---|---|
| 1 | 20 | 32.2 (0.6) | 28.8 (0.5) | 27.8 (0.4) |
| 1 | 40 | 78.2 (1.6) | 68.1 (1.3) | 57.6 (1) |
| 1 | 60 | 98.2 (2) | 104.3 (2) | 92.8 (1.7) |
| 6 | 20 | 44.2 (0.8) | 38.6 (0.7) | 35.2 (0.6) |
| 6 | 40 | 97.9 (2.1) | 82.2 (1.5) | 80.7 (1.4) |
| 6 | 60 | 163 (3.3) | 136.6 (2.6) | 121.8 (2.2) |
| 24 | 20 | 44.5 (0.9) | 38.4 (0.7) | 38 (0.6) |
| 24 | 40 | 102.3 (2.1) | 99.7 (1.8) | 88.4 (1.6) |
| 24 | 60 | 190.8 (4.2) | 153.5 (3) | 147.6 (2.6) |

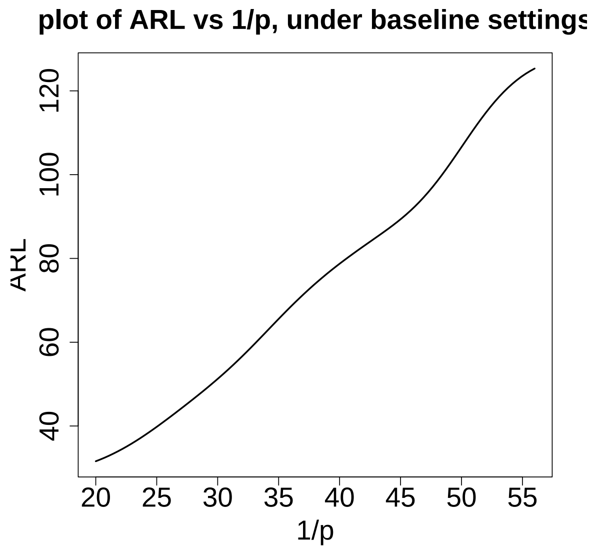
Figure 5.13: Plot of in control ARL vs $1/p$ for M model



**MSPT model:**

In the MSPT model, the plot of in control average run length versus different settings of $p$, under the baseline setting: $p = 0.02$, $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\delta_1 = 1$, $\delta_2 = 1$, is shown in Figure 5.14.

Figure 5.14: Plot of in control ARL vs $1/p$ for MSPT model

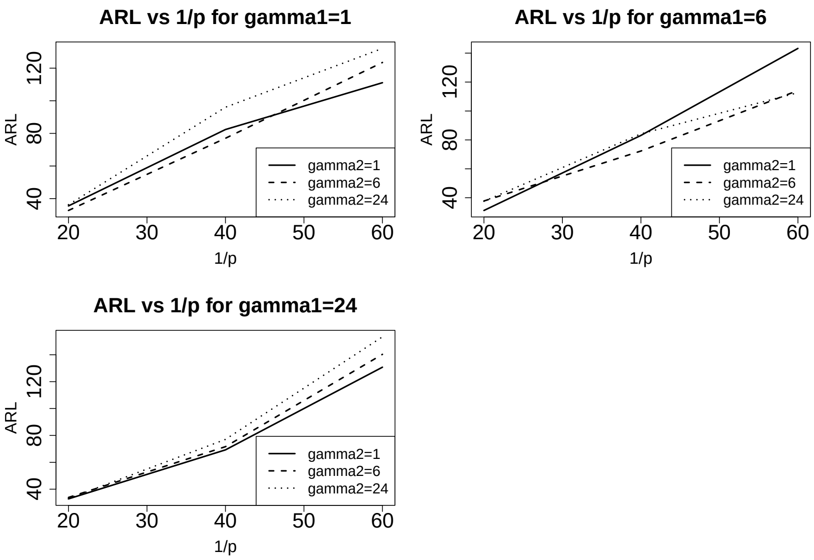



Because of the simulation error, we apply kernel smooth to obtain a better view. From this graph, we can still observe a linear relationship between ARL and $1/p$. To examine how precision of prior distribution for $\theta$ will affect average run length, we select pairs of $\gamma_1$ and $\gamma_2$ from 1, 6 and 24. Here we also guarantee that $\delta_1 = \gamma_1$ and $\delta_2 = \gamma_2$. The ARL table and its plot are shown in Table 5.5 and Figure 5.15.

Table 5.5: ARL for different $p$, $\gamma_1$ and $\gamma_2$ in MSPT model

| $\gamma_1$ | $1/p$ | $\gamma_2 = 1$ | $\gamma_2 = 6$ | $\gamma_2 = 24$ |
|---|---|---|---|---|
| | 20 | 35.5 (0.7) | 32.7 (0.7) | 36.3 (0.7) |
| 1 | 40 | 82.4 (1.6) | 77 (1.7) | 96 (2) |
| | 60 | 111.1 (2.7) | 123.5 (2.6) | 132.1 (3) |
| | 20 | 31.2 (0.6) | 37.8 (0.7) | 37.6 (0.7) |
| 6 | 40 | 83.1 (1.9) | 72.3 (1.6) | 84.2 (1.7) |
| | 60 | 143.3 (3.1) | 114.2 (2.5) | 112.7 (2.4) |
| | 20 | 32.8 (0.6) | 33.8 (0.6) | 32.9 (0.6) |
| 24 | 40 | 69.2 (1.7) | 71.6 (1.3) | 77 (1.6) |
| | 60 | 130.7 (3.1) | 140.3 (2.9) | 153.3 (3.1) |

Figure 5.15: Plot of in control ARL vs $1/p$ for MSPT model



We have the following observations:

1. The ARL has a roughly linear relationship with $1/p$, with a little bit curvature.
2. The ARL is positively related to $1/p$, but it is hard to see its relationship with $\gamma_1$ and $\gamma_2$.
3. The standard error of our ARL estimate is positively related to $1/p$.

### 5.2.2 Out of Control ARL

Another interesting feature of the different SPC methods is the behavior of out-of-control average run length (OOC-ARL). In this section, we compare the OOC-ARL of all five methods. Each method is tested under four circumstances: i.i.d standard normal sequence with a jump of mean, a jump of variance, a jump of $\phi$ and a jump of $\theta$ at $T = 15$.

The baseline settings of the five methods are:

- M model: $p = 0.024$, $a = 3$, $b = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$.
- S model: $p = 0.024$, $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$.
- MS model: $p = 0.025$, $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_0 = 0$, $\sigma_0 = 5$.
- MSP model: $p = 0.02$, $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$.
- MSPT model: $p = 0.02$, $a_1 = 3$, $b_1 = 2$, $a_2 = 3$, $b_2 = 2$, $\mu_{10} = 0$, $\mu_{20} = 0$, $\tau_1 = 1$, $\tau_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\delta_1 = 1$, $\delta_2 = 1$.

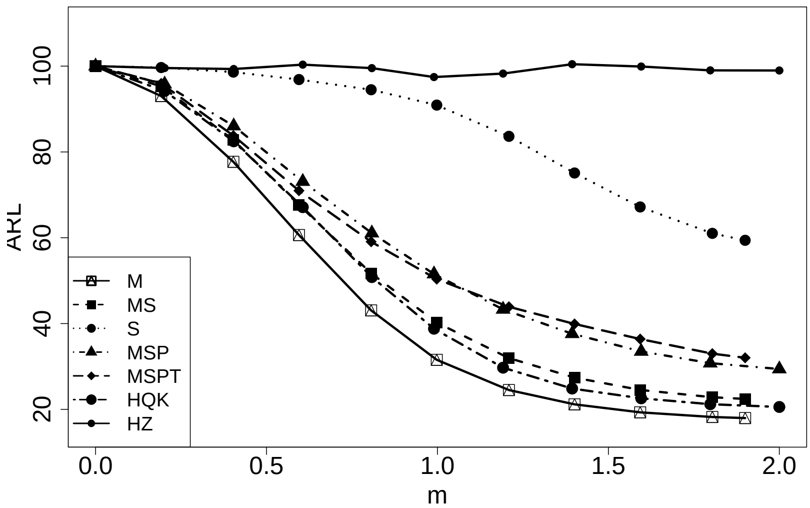
An important purpose of selecting these baseline settings is to obtain a similar in-control average run length, so that comparison of OOC-ARL is meaningful.

In addition to the four models we developed in this thesis, we also compare the OOC-ARL curves with those in HQK (Hawkins, Qiu and Kang (2003)) and HZ (Hawkins and Zamba (2005)). Again, the HQK is intended to detect a mean change and HZ is designed to detect change of variance.

**Jump of Mean**

The study is designed as follows: an i.i.d standard normal sequence is generated for $T < 15$. Starting from $T = 15$, there is a jump of $\mu$ from 0 to $\tilde{\mu}$, where $\tilde{\mu} \in [0, 2)$, with an increment of 0.1.

Figure 5.16: Out of control ARL vs different after-jump value of $\mu$




The Figure 5.16 shows the best OOC-ARL behavior coming from M model, MS model and HQK model. MSP model does a little worse, S model gives a much slower reaction of change, and HZ method does not have reaction on mean change.

These finds are according to our intuition about the properties of all these models: M model, MS model, and MSP model are expected to detect a jump of $\mu$. However, since the more complicated model allows for a wider variety of shifts, such as $\sigma^2$ and $\phi$, it is to be expected that it will be less effective than a model specifically designed for the shift of $\mu$.

Although the S model is not designed to react to a jump of mean, it has lower ARL as the mean changes. This is because we set up the prior distribution so that this method knows the true mean is somewhere around 0 ($\mu_0 = 0$, $\sigma_0 = 5$). As mean change increases, the method will find more outliers, which result in a smaller ARL.

HQK method also has good performance here, while HZ method acts as expected: it should not detect any mean change, since it is based on a two sample F statistic.

Finally, it is worth noting that M model and MS model have almost the same ARL behavior. While the M model performs a little better, there is only a slight performance penalty for expanding to to allow for variance changes as well.

**Jump of Variance**

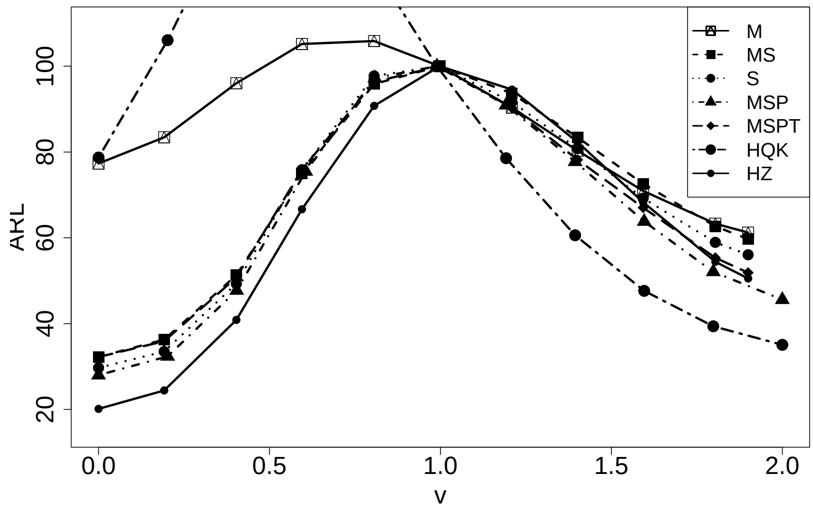
The study is designed as follows: an i.i.d standard normal sequence is generated for $T < 15$. Starting from $T = 15$, there is a jump of $\sigma$ from 1 to $\tilde{\sigma}$, where $\tilde{\sigma} \in [0, 2)$, with an increment of 0.1.

From Figure 5.17 we observe a bias of most OOC-ARL curves: when variance is a little bit less than 1, the OOC-ARL is even longer than the in-control ARL. However, the M model and HQK model are the most severe ones. When variance decreases, there will be less outliers, and thus less alarms given by the two methods. This is another proof of the similarity between these two models. For the other models, we cannot tell too much difference among them, but when variance decreases, HZ model has a slightly better performance than other models.

**Jump of $\phi$**

The study is designed as follows: an i.i.d standard normal sequence is generated for $T < 15$. Starting from $T = 15$, there is a jump of $\phi$ from 0 to $\tilde{\phi}$, where $\tilde{\phi} \in (-1, 1)$, with an increment of 0.1. Here we rescale the innovation to make sure the marginal mean

Figure 5.17: Out of control ARL vs different after-jump value of $\sigma$




and variance are the same before and after the jump, i.e,

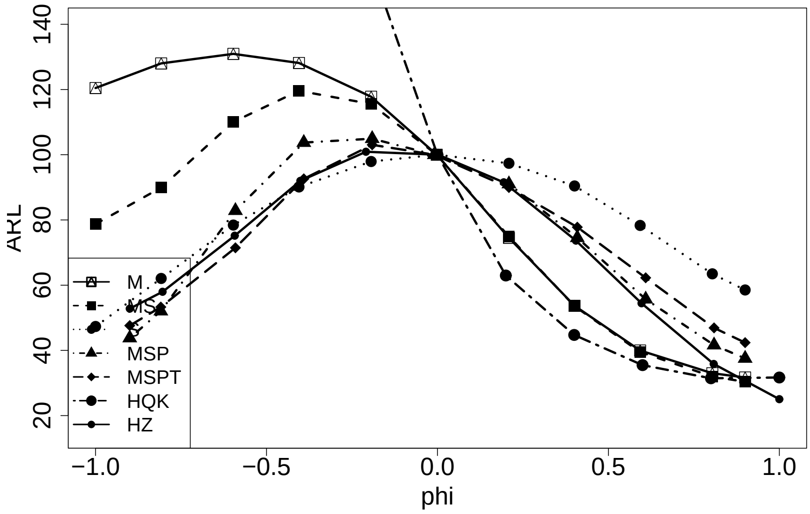
$$Y_k = X_k \qquad \text{for } k \leq n$$

and

$$Y_k = \tilde{\phi} Y_{k-1} + X_k \sqrt{1 - \tilde{\phi}^2}$$

From the Figure 5.18 we find some interesting features. First of all, the HQK model performs terribly when $\phi$ reduce to a negative value, and both M model and MS model show strong bias from the in-control scenario ($\phi = 0$). However, when $\phi$ jumps to positive value, these three methods have better performance than the other methods. In fact, when $\phi$ jumps to a positive value, the sequence becomes more positively autocorrelated, making the sample mean in a short period of time more deviated from 0. The ARL hence tends to be smaller when $\phi$ increases. However, if $\phi$ jumps to a negative value, the sequence becomes more negatively autocorrelated, making the sample mean in every short period of time more stable and closer to 0. In the extreme case, when $\phi$

Figure 5.18: Out of control ARL vs different after-jump value of $\phi$




gets close to -1, OOC-ARL for HQK will be close to infinity.

The MS model is able to detect the change if $\phi$ jumps to a large negative value. The decrease of ARL for very negative $\tilde{\phi}$ results from the increase of sample variance in a short period of time after the jump, but if the $\tilde{\phi}$ is not small enough, the performance is poor.

From the graph, the S model reacts to the change of $\phi$ symmetrically, because an increase or decrease of $\phi$ will cause a decrease or increase of sample variance in a short period of time in our scenario (recall that we rescaled the AR(1) sequence). The MSP model and MSPT model are better than S model when $\phi$ jumps to positive value and worse than S model when $\phi$ jumps to small negative value, because these two models still consider a change of $\mu$, making their performances somewhere between S model and M/MS model.

**Jump of $\boldsymbol{\theta}$**

The study is designed as follows: an i.i.d standard normal sequence is generated for

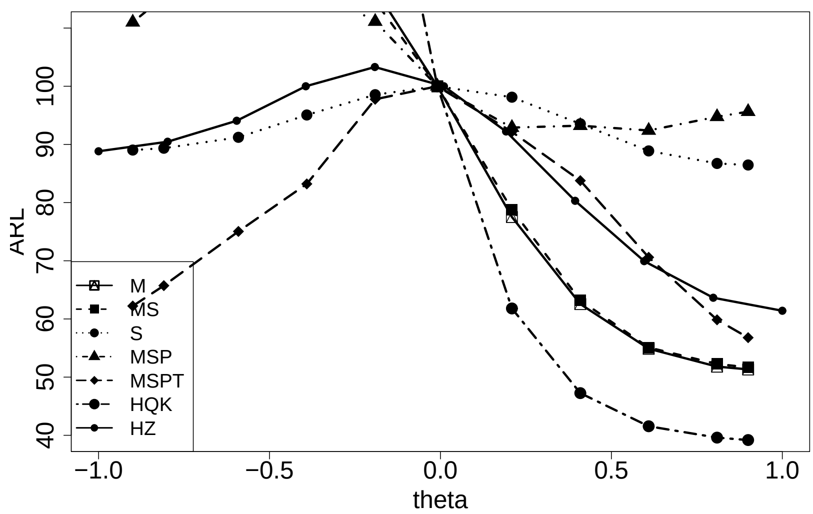
$T < 15$. Starting from $T = 15$, there is a jump of $\theta$ from 0 to $\tilde{\theta}$, where $\tilde{\phi} \in (-1, 1)$, with an increment of 0.1. As in the previous case, we also rescale the innovation to make sure the marginal mean and variance are the same before and after the jump, i.e,

$$Y_k = X_k \qquad \text{for } k \leq n$$

and

$$Y_k = \frac{X_k + \tilde{\theta} X_{k-1}}{\sqrt{1 + \tilde{\theta}^2}}$$

Figure 5.19: Out of control ARL vs different after-jump value of $\theta$




From Figure 5.19 we have several interesting discoveries from this graph. The models that focus on mean change (M model, MS model and HQK method) have good performance when $\theta$ jumps to positive values. The reason is similar to the case with jump of $\phi$: if $\theta$ jumps to positive value, the autocorrelation becomes $\theta > 0$, so that sample mean in a short period of time will dramatically deviate from in control level. However, these models perform terribly when $\theta$ jumps to negative value, because in those cases the sequence has negative autocorrelation, and thus the sample mean after change point tends to be more stable.

Our S model has a slight reaction to the change of $\theta$, due to the change of sample variance in a short period of time after the jump. This effect is similar to that in the case with jump of $\sigma$. The HZ model performs better than S model when $\theta$ jumps to positive value, but worse than S model when $\theta$ jumps to negative value.

Our MSP model has little reaction to change point if $\theta$ jumps to positive value, and performs even worse when $\theta$ jumps to negative value. As for the MSPT model, it has the best performance when $\theta$ jumps to negative value, and it has a symmetric ARL curve, except a little bias for $-0.2 < \theta < 0$. Therefore if we lack the information of the jump direction of $\theta$, the MSPT will be the most ideal choice.

# Chapter 6

# Algorithm and Practical Problems

## 6.1 Time and Space Complexity

Suppose we have a phase II sequence $X_1, X_2, ...,$ and the model we apply does not require Monte Carlo integration, such as those in 2.4.5 with known $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. Then every time when there is a new observation $X_n$, the naive algorithm has time complexity $O(n^2)$ and space complexity $O(1)$, since for each $T = 1, 2, ..., n$, we have to compute $S_{1,T}^-$, $S_{2,T}^-$, $S_{1,T}^+$, and $S_{2,T}^+$, which takes $2n$ steps to sum up. This can be improved to an algorithm with $O(n)$ time complexity and $O(n)$ space complexity by keeping records of the sums. Let $S_{j,T,n}^-$ and $S_{j,T,n}^+$ denote the sums at $n$ observations, where $j = 1, 2$. Then

$$S_{j,T,n+1}^- = S_{j,T,n}^-$$

$$S_{j,T,n+1}^+ = S_{j,T,n}^+ + X_{n+1}^j$$

where $S_{j,T,n}^-$ or $S_{j,T,n}^+$ is 0 if not defined before. We update and take record of the sums after each new observation.

## 6.2 Problem of Underflow

If the numerical integration is not needed, we always suggest using the log-posterior distribution, i.e. $\log(f(T|\boldsymbol{X}))$ instead of $f(T|\boldsymbol{X})$ itself. This is because $f(T|\boldsymbol{X})$ can have small positive numbers very easily, and therefore cause an underflow. For example, $\exp(-4000)$ is common in our method but this number is simply 0 in double type. If we just compare log-values, this problem can be avoided. However, if numerical integration is needed, we have to sum up a series of $f(T|\boldsymbol{X})$ evaluated at different samples. The series of $f$ might all be very small. In this case, we can still compare the log of the sum, using the following method.

Suppose we have sequence $A_1, A_2, ..., A_N$ with very small positive values like $\exp(-4000)$, and $a_i = \log(A_i)$, $i = 1, ..., N$ are easier to store in computer. Then we have

$$\log(\sum_{i=1}^{N} A_i) = \log(\sum_{i=1}^{N} \exp(a_i)) = \log[\exp(a_0) \sum_{i=1}^{N} \exp(a_i - a_0)] = a_0 + \log[\sum_{i=1}^{N} \exp(a_i - a_0)]$$

where

$$a_0 = \max_{1 \le i \le N} a_i$$

and the last summation will not cause the underflow problem, since all the components are positive and the maximum of them is 1.

## 6.3 Calibration of Hyperparameters

The Bayesian SPC method works well if a relevant and informative prior can be used. Usually people form prior distributions from two sources: subjective belief and objective experience. If subjective belief is applied, qualitative features of in control and out-of-control status, mean and variance of each parameter, as well as the in control and out of control ARL can be used as reasonable guidelines. For example, we can set $p = 0.05$ if we think the process generally runs out of control in about 20 time units, and we can set $\alpha_1 = \beta_1 = 3$ and $\alpha_2 = \beta_2 = 0.3$ if we think the in control correlation is weak and out of control correlation is strong.

If phase I data is available, we can find the settings for hyperparameters by fitting them from the phase I data. For example, if we assume a model with all parameters changed before and after the change point, we can fit maximum likelihood estimators $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\phi}}_2, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ from the likelihood function as in Chapter 2.4.1:

$$L(\boldsymbol{X}_1^n | \mu_1, \mu_2, \sigma_1, \sigma_2, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, T)$$

and the hyperparameters, such as $\mu_{10}$, $\mu_{20}$, $\tau_1$, $\tau_2$, $a_1$, $a_2$, $b_1$, $b_2$ can be selected so that the moments (usually first and second moments are enough) of the priors coincide with those estimated parameters. Finally, $p$ can be used to adjust the in-control ARL.

Another method of finding hyperparameters setting from phase I data is to integrate out parameter set

$$(\mu_1, \mu_2, \sigma_1, \sigma_2, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, T)$$

with respect to their prior distribution in the likelihood function, and get a "hyper" likelihood function that only consists of hyperparameters. Then find the fitted hyperparameters through an maximization procedure. This approach is natural, but it is not easy to implement. Neither the integration nor the maximization is trivial. If close form "hyper" likelihood function is not available, this approach is not desired method.

In the next chapter, we will demonstrate the procedure to select prior information and calibrate hyperparameters in a real example.

# Chapter 7

# A Real Example of Application of Bayesian SPC Method

There is currently a wide interest in the topic of climate change and global warming, thought to have started some time after World War II. The increasing emit of green house gas is believed to dramatically raise the world average temperature, and therefore change the climate. For example, the warmer atmosphere can evaporate more water and bring the world more storms and precipitation.

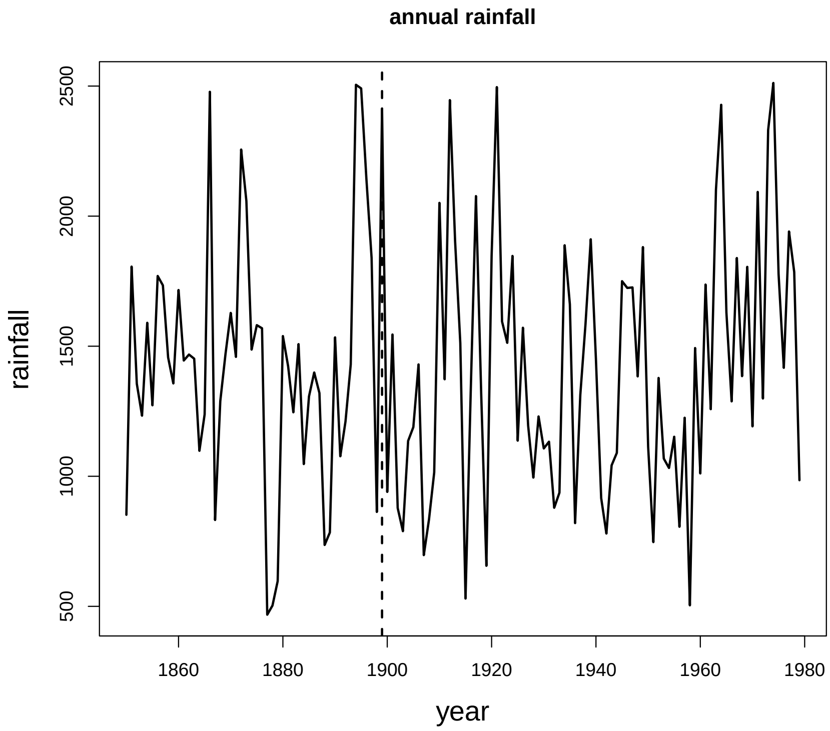
We demonstrate our procedure using 130 years' annual rainfall data from 1849 to 1979 in Fortaleza, Brazil (Hipel and McLeod, 1994). The first 50 years precede the more recent increase in atmospheric carbon dioxide, and so should provide good prior information about the rainfall patterns in the absence of any changes. The plot of this time series is in Figure 7.1.

We assume there is no change of underlying model for the first 50 years. The plots of autocorrelation function and partial autocorrelation function suggest an $MA(1)$ model:

$$X_n = \varepsilon_n + \theta\varepsilon_{n-1} + \nu$$

where $\varepsilon_n$ are i.i.d with $N(0, \sigma^2)$. Fitting this model by minimizing the conditional sum of squared residuals gives estimates $\hat{\theta} = 0.29$ with standard error 0.14. For the selection of prior distribution of in control $\theta$, we assume $\pi(\theta) \propto (1+\theta)^\gamma(1-\theta)^\delta$, where $\gamma, \delta \geq 0$

Figure 7.1: Rain Fall Data for Fortaleza (mm), 1849-1979




and this is essentially $Beta(\gamma, \delta)$. The corresponding Beta distribution can be set up so that its mean equals $(1+\hat{\theta})/2 = 0.645$ and standard deviation equals $se(\hat{\theta})/2 = 0.07$. We can solve $\gamma$ and $\delta$ by

$$\frac{\gamma}{\gamma+\delta} = 0.645$$

$$\frac{\gamma\delta}{(\gamma+\delta)^2(\gamma+\delta+1)} = 0.07^2$$

and we get $\gamma = 30.24$ and $\delta = 16.65$.

We also have $\hat{\nu} = 1468.53$ with standard error 83.80, and $\hat{\sigma^2} = 465.7^2 = 216874$. In our own model we use

$$X_n = (\varepsilon_n + \mu) + \theta(\varepsilon_{n-1} + \mu)$$

So $\mu = \frac{1}{1+\theta}\nu$. We can substitute $\theta$ for $\hat{\theta}$ and get $\mu \sim N(1138.63, 64.97^2)$. Using our notation in chapter 2, the prior for $\mu$ can be set up to $\mu_0 = 1138.63$ and

$\tau = 465.70/64.97 = 7.17$, recall that the prior for $\mu$ given $\sigma$ is a normal distribution, with precision $\tau$.

Under the frequentist model and the normal assumption, the $\hat{\sigma^2}$ follows a Gamma distribution, but we need to fit a inverse Gamma distribution, and the Gamma distribution has $\sigma^2$ as parameter. However, we can use a bootstrap sample of $\hat{\sigma^2}$ to estimate its first and second moments and match them with inverse Gamma parameters. In this example, we get $a = 21.50$ and $b = 4342304$.

All the hyperparameters we estimated above are for the in-control parameters. For the out-of-control parameters, or the right side of the change point, we have no idea where will they change. A reasonable idea is to replicate them, i.e, $\gamma_2 = \gamma_1 = 30.24$, $a_2 = a_1 = 21.50$, etc. However, if we are more confident that there will be a change of any parameters, but are still not sure about the direction of the change, we can simply raise the variance of that prior distribution. For example, we can set up $\tau_2 = \tau_1/10$ if we believe that $\mu$ is going to change. We can also set up $a_2 = a_1/5$ and $b_2 = b_1/5$ if we believe that $\sigma$ is going to change.

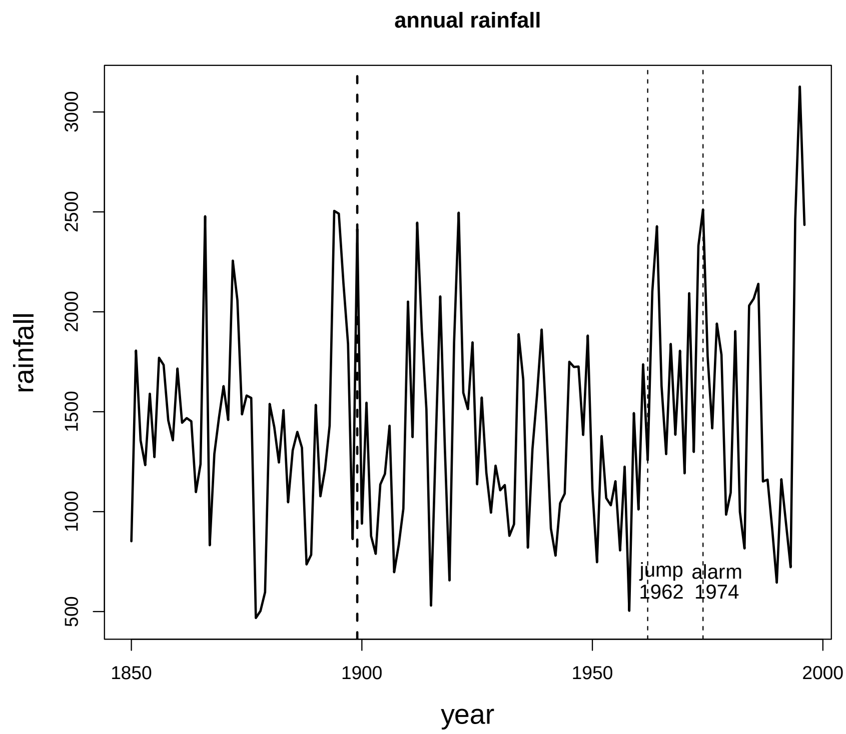
Finally, the user should decide $p$ based on the knowledge about precipitation. For example, we might assume $p = 0.02$ for an average of 50 years of the first occurrence of the next change point of annual rainfall in Fortaleza.

We carry out the following SPC on the phase II data (rain fall data for Fortaleza after 1900), applying the following setting of hyperparameters: $\mu_{10} = \mu_{20} = 1138.627$, $\tau_1 = 7.17 = 10\tau_2$, $a_1 = 21.50 = 10a_2$, $b_1 = 4342304 = 10b_2$, $\gamma_1 = \gamma_2 = 30.24$, and $\delta_1 = \delta_2 = 16.65$. We also set up $p = 0.02$.

This SPC method does not detect any change of any parameter we are considering. We also apply the $ARMA(1,1)$ (MSPT) and $AR(1)$ (MSP) models, using the same approach to determine those hyperparameters. Neither of them gives any alarm till the end of the rainfall sequence.

Finally, we apply the M model and MS model for detection of mean or variance change. The result of these tests are shown in Figure 7.2. For the MS model, the process gives an alarm at year 1974, indicating a change point occurring at year 1972, while the M model gives an alarm at 1974, but indicating a change point at 1962. In fact, if we run more steps after the alarm is given, these two methods agree on a change point at 1962 from then on.

Figure 7.2: Result of Bayesian SPC on Fortaleza Rainfall Data




From the graph, we can observe a shift of mean away from the original mean of around 1500. However, there is a disagreement between the models assuming i.i.d sequence and the models considering autocorrelation. An explanation for this discrepancy is that the marginal mean did shift, but the models considering autocorrelation treat it as normal phenomenon under the influence of autocorrelation.

In this graph, we also include the annual rainfall data for several years after 1979, from Manly (2009) and some online data from Fortaleza Aeroporto weather station. There is a reversion of the average rainfall during the late 1970s and early 1980s, and the positive shift of average precipitation during 1960s looks like a temporary change, implying that there is no long lasting fundamental change of the average rainfall in Fortaleza till 1980s.

# Chapter 8

# Future Work

One important extension of our current work is to the multivariate case. The most study on multivariate SPC is to monitor the change of mean vector. Suppose there is a multivariate normal sequence $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, ...)$, where $\boldsymbol{X}_i$ is a $p \times 1$ random vector following a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. After time $T$, the mean $\boldsymbol{\mu}$ jumps from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1$.

Multivariate SPC monitoring mean chagne has long been studied since the first related paper, Hotelling (1947). The popular methods include Hotelling's $T^2$ chart, the MEWMA (Multivariate exponentially-weighted moving average) chart, and the MCUSUM (Multivariate Cumulative Sum) chart. In some recent important works regarding unknown or vague information about mean vector and jump direction, Zamba and Hawkins (2006) proposed a $h_{\alpha,n}$ chart similar to that in Hawkins, Qiu and Kang (2003), based on the Hotelling's $T^2$ statistic. This method is useful when prior information of in-control and out-of-control mean is completely unknown. Liu and Qiu (2011) incorporated prior information into the posterior mean of the direction of changing vector.

Embedding the Bayesian framework of our thesis into multivariate case is not very straightforward, but instead bearing some critical issues. First issue is how many parameters, and thus how many degrees of freedom should we set up. Apparently, if we assume an arbitrary structure of $\boldsymbol{\Sigma}$, $p(p+1)/2$ parameters for $\boldsymbol{\Sigma}$ have to be estimated.

For simplicity, one can first assume diagonal structure of $\boldsymbol{\Sigma}$ =diag$(\sigma_1^2, \sigma_2^2, ..., \sigma_p^2)$, or even $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$. In fact, following this set up, the marginal posterior distribution of the change point $T$ will be similar to what we derived in this thesis. One can even derive the variance and mean-variance change only model based on this assumption.

If the independent assumption for each individual vector $\boldsymbol{X}_i$ is released, theoretical work and computation might be much harder. One should carefully consider how to set up the prior information of $\boldsymbol{\Sigma}$ in a reasonable way so that $\boldsymbol{\Sigma}$ is always ensured to be positive definite, and the rest derivation of posterior distribution is workable theoretically and computationally. Probably some medium-strong assumption should still be made on the correlation structure of $\boldsymbol{X}_i$.

An even challenging part of our extension is the Bayesian multivariate SPC on autocorrelated sequence, say multivariate $ARMA(1,1)$ model:

$$\boldsymbol{X}_i = \boldsymbol{\Phi}\boldsymbol{X}_{i-1} + \boldsymbol{\Theta}\boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\Phi}$ is a $p \times p$ matrix making the sequence $\boldsymbol{X}$ stationary, and $\boldsymbol{\Theta}$ is a $p \times p$ matrix making the sequence invertible. The biggest issue, like discussed above, would be dimensionality of $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$. If no structure assumption is made, there will be $p^2$ parameters to fit. The possible solution might be as that for the $\boldsymbol{\Sigma}$: to assume a structure, but leave the scale as unknown parameter.

# References


[1] J. H. Albert and S. Chib (1993): Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics,* Vol. 11, No. 1, pp. 1-15

[2] L. C. Alwan and H. V. Roberts (1988), Time-Series Modeling for Statistical Process Control. *Journal of Business & Economic Statistics*, Vol. 6, pp. 87-95.

[3] D. Andrews (1993): Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, Vol. 61, No. 4, pp. 821-856

[4] M. M. Barbieri and C. Conigliani (1998): Bayesian Analysis of Autoregressive Time Series with Change Points. *Statistical Methods & Applications*. Volume 7, Number 3, 243-255

[5] O. Barndorff-Nielsen and G. Schou (1973): On the parametrization of Autoregressive Models by Partial Autocorrelations. *Journal of Multivariate Analysis* 3, 408-419.

[6] D. Barry and J. A. Hartigan (1992): Product Partition Models for Change Point Problems. *The Annals of Statistics*, Vol. 20, No. 1, 260-279

[7] D. Barry and J. A. Hartigan (1993): A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association,* Vol. 88, No. 421, pp. 309-319

[8] N. B. Booth and A. F. M. Smith (1982): A Bayesian Approach to Retrospective Identification of Change Points *Journal of Econometrics* 19 7-22.

[9] P. Castagliola and F. Tsung (2005): Autocorrelated SPC for Non-Normal Situations *Quality and Reliability Engineering International*, Vol. 21, pp. 131-161.

[10] B. P. Carlin, A. E. Gelfand and A. F. M. Smith(1991): Hierarchical Bayesian Analysis of Changepoint Problems. *Applied Statistics*, Vol. 41, No. 2, pp. 389-405

[11] H. Chernoff and S. Zacks (1964): Estimating the Current Mean of a Normal Distribution Which Is Subjected to Changes in Time. *The Annals of Mathematical Statistics*, Vol. 35, No. 3, pp. 999-1018

[12] S. Chib (1998): Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86 pp. 221-241

[13] R. A. David, T. C. Lee and G. A. Rodriguez-Yam (2006): Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association*, Vol. 101, No. 473

[14] R. A. Davis, T. C. Lee and G. A. Rodriguez-Yam (2008): Break Detection for a Class of Nonlinear Time Series Models *Journal of Time Series Analysis,* Vol. 29, No. 5

[15] N. Dobigeon, J. Tourneret and M. Davy (2007): Joint Segmentation of Piecewise Constant Autoregressive Processes by Using a Hierarchical Model and a Bayesian Sampling Approach *IEEE Transactions on Signal Processing*, Vol. 55, No. 4

[16] P. Fearnhead (2006): Exact and Efficient Bayesian Inference for Multiple Changepoint Problems *Stat Comput*, Vol. 16, pp. 203-213.

[17] P. Fearnhead and Z. Liu (2007): On-line Inference for Multiple Change Points Problems *Journal of the Royal Statistical Society: Series B*, Vol. 69, No. 4, pp. 589-605.

[18] P. Fearnhead and Z. Liu (2011): Efficient Bayesian Analysis of Multiple Changepoint Models with Dependence Across Segments *Stat Comput*, Vol. 21, pp. 217-229.

[19] P. Giordani and R. Kohn (2006): Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models *Sveriges Riksbank Working Paper Series* No. 196

[20] L. A. Gardner (1969): On Detecting Changes in the Mean of Normal Variates. *The Annals of Mathematical Statistics*, Vol. 40, No. 1, pp. 116-126

[21] A. E. Gelfand and A. F. M. Smith (1990): Sampling-Based Approaches to Calculating Marginal Densities *Journal of the American Statistical Association*, Vol. 85, No. 410, pp. 398-409.

[22] S. Geman and D. Geman (1984): Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Journal of Applied Statistics*, Vol. 20, No. 6, pp. 721-741.

[23] J. M. Hammersley, D.C. Handscomb (1964): Monte Carlo Methods. *Methuen*, 1964

[24] T. J. Harris and W. H. Ross(1991): Statistical Process Control Procedures for Correlated Observations. *Canadian Journal of Chemical Engineering*, Vol. 69, pp. 48-57.

[25] D. M. Hawkins, P. Qiu, C. W. Kang (2003): The Changepoint Model for Statistical Process Control *Journal of Quality Technology*, Vol. 35, No. 4, pp. 355-366.

[26] Hotelling, H., (1947), Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights, C.Eisenhart et. al. pp.111-184.

[27] D. M. Hawkins and K. D. Zamba (2005): A change-point model for a shift in variance *Journal of Quality Technology*, Vol. 37, No. 1, pp. 21-31.

[28] D. M. Hawkins and K. D. Zamba (2005): Statistical Process Control for Shifts in Mean or Variance Using a Changepoint Formulation *Technometrics*, Vol. 47, No. 2, pp. 164-173.

[29] D. V. Hinkley (1970): Inference About the Change-Point in a Sequence of Random Variables *Biometrika*, Vol. 57, No. 1, pp. 1-17

[30] K. W. Hipel and A. I. McLeod (1994): Time Series Modelling of Water Resources and Environmental Systems.

[31] B. Hultblad and S. Karlsson (2007): Bayesian simultaneous determination of structural breaks and lag lengths *Studies in Nonlinear Dynamics & Econometrics*, Vol. 12, No. 3, Article 4.

[32] C. Inclan and G. C. Tiao (1994): Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance *Journal of the American Statistical Association,* Vol. 89, No. 427, pp. 913-923

[33] M. C. Jones (1986): Randomly Choosing Parameters from the Stationarity and Invertibility Region ofAutoregressive-Moving Average Models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 36, No. 2, pp. 134-138

[34] L. A. Jones (2002): The Statistical Design of EWMA Control Charts With Estimated Parameters *American Society for Quality*, Vol. 34, No. 3, pp. 277-288.

[35] L. A. Jones, C. W. Champ, S. E. Rigdon (2001): The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters *American Society for Quality*, Vol. 43, No. 2, pp. 156-167.

[36] P. Jong and J. Penzer (2000): Diagnosing Shocks in Time Series *Journal of the American Statistical Association*. 93, 796-806

[37] S. Knoth and W. Schmid (2002): Monitoring the Mean and the Variance of a Stationary Process. *Statistica Neerlandica*, Vol. 56, No. 1, pp. 77-100.

[38] A. B. Koehler, N. B. Marks, R. T. O'Connell (2001): EWMA Control Charts for Autoregressive Processes *The Journal of the Operational Research Society*, Vol. 52, No. 6, pp. 699-707.

[39] T. L. Lai (1995): Sequential Changepoint Detection in Quality Control and Dynamical Systems *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, No. 4, pp. 613-658.

[40] T. L. Lai (2001): Sequential Analysis: Some Classical Problems and New Challenges. *Statistica Sinica*, Vol. 11, pp. 303-408.

[41] T. L. Lai and H. Xing (2011): A Simple Bayesian Approach to Multiple Changepoints. *Statistica Sinica* 21, 539-569

[42] W. Liu and P. Qiu (2011): A Multivariate CUSUM Chart Accommodating Prior Information About Potential Shifts.

[43] F. Lombard and J. D. Hart (1994): The Analysis of Change-point Data with Dependent errors. *Lecture Notes-Monograph Series*, Vol. 23, Change-Point Problems, pp. 194-209

[44] I. B. MacNeill (1974): Tests for Change of Parameter at Unknown Times and Distributions of Some RelatedFunctionals on Brownian Motion *The Annals of Statistics*, Vol. 2, No. 5, pp. 950-962

[45] B. Manly (2009): Statistics for Environmental Science and Management, Second Edition

[46] R. E. McCulloch and R. S. Tsay (1993): Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series. *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 968-978

[47] A. Mohammad-Djafari and O. Feron (2006): A Bayesian approach to change points detection in time series. *International Journal of Imaging Systems and Technology, Special Issue: Computer Vision*. Vol 16, Issue 5, pages 215-221

[48] J. F. Monahan (1984): A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, Vol. 71, No. 2, pp. 403-404

[49] D. C. Montgomery and C. M. Mastrangelo (1991): Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology*, Vol. 23, No. 3, pp. 179-204.

[50] E. Moreno, G. Casella and A. Garcia-Ferrer (2004): An Objective Bayesian Analysis of the Change Point Problem. *Stochastic Environmental Research and Risk Assessment*, Vol. 19, No. 3, pp. 191-204

[51] J. C. Naylor and P. W. Woodward (1993): An Application to Bayesian Methods in SPC. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 42, No. 4, *Special Issue: Conference on Practical Bayesian Statistics*, pp. 461-469

[52] Oakland (2008): Statistical Process Control, Six Edition.

[53] E. S. Page (1955): A Test for a Change in a Parameter Occurring at an Unknown Point *Biometrika*, Vol. 42, No. 3/4, pp. 523-527

[54] E. S. Page (1961): Cumulative Sum Chart *Technometrics*, Vol. 3, No. 1, pp. 1-9

[55] M. Perry, G. Mercado and J. Pignatiello (2011): Phase II Monitoring of Covariance Stationary Autocorrelated Processes. *Quality and Reliability Engineering International*, Vol. 27, No. 1, pp. 35-45

[56] A. N. Pettitt (1979): A Non-Parametric Approach to the Change-Point Problem. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* Vol. 28, No. 2, pp. 126-135

[57] D. Piccolo (1982): The size of the stationarity and invertibility region of an autoregressive-moving average process. *J. Time Ser. Anal., 3, 245-247.*

[58] A. E. Raftery and V. E. AkmanSource (1986): Bayesian Analysis of a Poisson Process with a Change-Point *Biometrika,* Vol. 73, No. 1, pp. 85-89

[59] B. Ray and R. Tsay (2001): Bayesian Methods for Change-point Detection in Long-range Dependent Processes. *Journal of Time Series Analysis*, Vol. 23, pp. 687-706

[60] L. Shu, D. W. Apley and F. Tsung (2002): Autocorrelated Process Monitoring Using Triggered Cuscore Charts. *Quality and Reliability Engineering International*, Vol. 18, pp. 411-421.

[61] A. F. M. Smith (1975): A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables *Biometrika*, Vol. 62, No. 2, pp. 407-416

[62] Y. S. Son and S. W. Kim (2005): Bayesian single change point detection in a sequence of multivariate normal observations. *Statistics*, Vol. 39, No. 5, 373-387

[63] D. A. Stephens (1994): Bayesian Retrospective Multiple-Changepoint Identification. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 43, No. 1, pp. 159-178

[64] L. Tierney (1994): Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, Vol. 22, No. 4, pp. 1701-1728.

[65] D. H. Timmer and J. J. Pignatiello (2003): Change Point Estimates for the Parameters of an AR(1) Process *Quality and Reliability Engineering International*, Vol. 19, pp. 355-369.

[66] J. Tourneret, M. Doisy and M. Lavielle: Bayesian off-line detection of multiple change-points corrupted by multiplicative noise: application to SAR image edge detection *Signal Processing*. Volume 83, Issue 9, pp. 1871-1887

[67] P. Tsiamyrtzis, D. M. Hawkins (2005): A Bayesian Scheme to Detect Changes in the Mean of a Short-Run Process. *Technometrics*. 47(4): 446-456

[68] P. Tsiamyrtzis and D. M. Hawkins (2008): A Bayesian EWMA Method to Detect Jumps at the Start-up Phase of a Process *Quality and Reliability Engineering International*, Vol. 24, pp. 721-735.

[69] J. Wang and E. Zivot (2000): A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance. *Journal of Business & Economic Statistics*, Vol. 18, No. 3, pp. 374-386

[70] M. West and P. J. Harrison (1986): Monitoring and Adaptation in Bayesian Forecasting Models. *Journal of the American Statistical Association*, Vol. 81, No. 395, pp. 741-750

[71] S. A. Vander Wiel (1996): Monitoring Processes That Wander Using Integrated Moving Average Models *Technometrics*, Vol. 38, No. 2, pp. 139-151.

[72] T. J. Vogelsang (1998): Testing for a Shift in Mean Without Having to Estimate Serial-Correlation Parameters. *Journal of Business & Economic Statistics*, Vol. 16, No. 1, pp. 73-80

[73] K. D. Zamba and D. M. Hawkins (2006): A multivariate change-point model for statistical process control *Technometrics*, Vol. 48, No. 4, pp. 539-549.

[74] N. F. Zhang (1998): A Statistical Control Chart for Stationary Process Data *Technometrics*, Vol. 40, No. 1, pp. 24-38.

[75] C. Zhou, C. Zou, Y. Zhang and Z. Wang (2009): Nonparametric control chart based on change-point model *Stat Papers*, Vol 50, Springer, pp. 13-28.