Spring 2018

New Bayesian methods for quality control applications

Baosheng He
University of Iowa

NEW BAYESIAN METHODS FOR QUALITY CONTROL APPLICATIONS

by

Baosheng He

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Industrial Engineering
in the Graduate College of
The University of Iowa

May 2018

Thesis Supervisor: Professor Yong Chen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Baosheng He

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Industrial Engineering at the May 2018 graduation.

Thesis Committee: _____
Yong Chen, Thesis Supervisor

_____
Mary Kathryn Cowles

_____
Joyee Ghosh

_____
Andrew Kusiak

_____
Amaury Lendasse

To my family

ACKNOWLEDGEMENTS

First I would like to express my deepest appreciation to my advisor, Professor Yong Chen, for his patient guidance and support in my PhD career. Without my advisor, it is impossible I can reach this point. I am sincerely thankful for his inspiration and detailed teaching in the nearly 4 years. I have been being impressed and motivated by his elegant and thoughtful work in research.

I would also say my gratitude to my committee members. Thank you, Professor Andrew Kusiak, Professor Amaury Lendasse, Professor Kate Cowles, and Professor Joyee Ghosh, for your friendliness and valuable comments for my thesis.

I thank all of my friends I met in the department, Yuxing Hou, Bo Sun, Guanglin Xu and Xudong Zhang, and Yiheng Liu and Hongda Zhang from Statistics and Actuarial Science department for their help and caring in my life and research.

Many thanks to the department administrative staff, Karla Stout, Tara Hoadley, and Deborah Hampton, who were always there to help.

In my life in Iowa city, my most enjoyable entertainment is to play soccer games with my teammates in Team Hawkloooong. I thank all my teammates.

Last but never the least, I would like to thank my parents, my sister, and my lovely wife Yuanyuan. Their love makes everything here worthwhile.

ABSTRACT

In quality control applications, the most basic tasks are monitoring and fault diagnosis. Monitoring results determines if diagnosis is required, and conversely, diagnostic results aids better monitoring design. Quality monitoring and fault diagnosis are closely related but also have significant difference. Essentially. monitoring focus on *online* changepoint detection, whilst the primary objective of diagnosis is to identify fault root causes as an *offline* method. Several critical problems arise in the research of quality control: firstly, whether process monitoring is able to distinguish systematic or assignable faults and occasional deviation; secondly, how to diagnose faults with coupled root causes in complex manufacturing systems; thirdly, if the changepoint and root causes of faults can be diagnosed simultaneously.

In Chapter 2, we propose a novel Bayesian statistical process control method for count data in the presence of outliers. That is, we discuss how to discern out of control status and temporary abnormal process behaviors in practice, which is incapable for current SPC methodologies. In this work, process states are modeled as latent variables and inferred by the sequential Monte Carlo method. The idea of Rao-Blackwellization is employed in the approach to control detection error and computational cost. Another contribution of this work is that our method possesses self-starting characteristics, which makes the method a more robust SPC tool for discrete data. Sensitivity analysis on monitoring parameter settings is also implemented to provide practical guidelines.

In Chapter 3, we study the diagnosis of dimensional faults in manufacturing. A novel Bayesian variable selection oriented diagnostic framework is proposed. Dimensional fault sources are not explicitly measurable; instead, they are connected with dimensional measurements by a generalized linear mixed effect model, based on which we further construct a hierarchical quality-fault model to conduct Bayesian inference. A reversible jump Markov Chain Monte Carlo algorithm is developed to estimate the approximate posterior probability of fault patterns. Such diagnostic procedure is superior over previous studies since no numeric regularization is required for decision making. The proposed Bayesian diagnosis can further lean towards sparse fault patterns by choosing suitable priors, in order to handle the challenge from the diagnosability of faults. Our work considers the diagnosability in building dimensional diagnostic methodologies. We explain that the diagnostic result is trustworthy for most manufacturing systems in practice. The convergence analysis is also implemented, considering the trans-dimensional nature of the diagnostic method.

In Chapter 4 of the thesis, we consider the diagnosis of multivariate linear profile models. We assume liner profiles as piece-wise constant. We propose an integrated Bayesian diagnostic method to answer two problems: firstly, whether and when the process is shifted, and secondly, in which pattern the shift occurs. The method can be applied for both Phase I and Phase II needs. For Phase I diagnosis, the method is implemented with no knowledge of in control profiles, whereas in Phase II diagnosis, the method only requires partial observations. To identify exactly which profile components deviate from nominal value, the variability of the value of pro-

file components is marginalized out through a fully Bayesian approach. To address computational difficulty, we implement Monte Carlo Method to alternatively inspect between spaces of changepoint positions and fault patterns. The diagnostic method is capable to be applied under multiple scenarios.

PUBLIC ABSTRACT

Quality control (SPC) is a broad concept in industries. The most critical problems are about statistical process control (i.e., monitoring) and fault diagnosis. Process is defined as a combination of materials, methods, equipment, etc. Therefore all processes have inherent statistical variability. For example, variations are generated in multiple manufacturing stages and finally transmitted forward to products. A process is in control if its variability stays within specified limits; otherwise it is out of control. Based on measurements of products, monitoring is applied to detect if the process is out of control, whereas diagnosis is required to probe into fault causes if nonconforming products are detected.

In this thesis, we discuss three practical yet fully studied topics. Firstly, we study for count data processes, how to develop a SPC tool to conduct robust monitoring in the presence of outliers. Secondly, we study how to make diagnosis on variation faults in manufacturing processes, considering such faults could be theoretically not diagnosable. Thirdly, we design a compound diagnostic framework for both out of control time and fault patterns.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Applications of quality control widely exist in industries (Zhou et al., 2004; Montgomery, 2009), signal processing (Wang and Kuo, 2007), medical surveillance (Tsiamyrtzis and Hawkins, 2005; Cooper et al., 2006; Woodall et al., 2006), etc. Addressing uncertainties of quality properties by data-driven methods receives intensive interests. Bayesian quality control methodologies have been used for decades (Ingleby and Lorenc, 1993; Calabrese, 1995; Colosimo and Del Castillo, 2006). However, such work are mainly restricted in statistical process control (SPC) studies (Montgomery, 2009), and on Gaussian processes. In this thesis, we propose novel Bayesian methods to extend the application of Bayes theorems in quality control area.

In Chapter 2, we develop a Bayesian self-staring SPC method to monitor count processes possibly undergoing outliers. The main objective is to distinguish outliers and true out of control status. In Chapter 3 we propose a Bayesian diagnostic practice for dimensional integrity in manufacturing, aided by the knowledge of the diagnosability of dimensional faults. In chapter 4, we combine the identification of process shift and process faults in a Bayesian framework, by which both Phase I and Phase II diagnosis of multivariate processes can be completed.

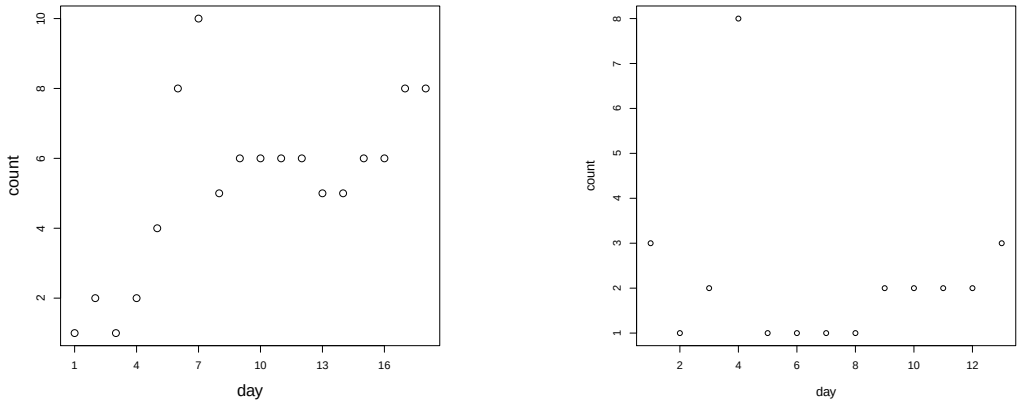
1.1 SPC on Count Process and Outliers

Traditional SPC methods, such as the cumulative sum (CUSUM) and exponentially weighted moving average(EWMA) control charts, rely on a large phase I

data set to estimate in-control process parameters (Montgomery, 2009). However, for many applications sufficient phase I data are not available either because collecting massive historical data is costly (Castagliola et al., 2013) or because processes under monitoring are inconsistent (Tsiamyrtzis and Hawkins, 2005). Self-starting process monitoring frameworks have been developed when phase I data are unavailable or very limited: Hawkins (1987) suggested a CUSUM scheme to detect shifts on mean μ and variance σ^2 for data from normal or at least asymptotically normal distributions. Quesenberry (1991b) proposed a Shewhart type methodology for $i.i.d$ normally distributed data sequence. Li et al. (2010) used sequential likelihood ratio tests to detect both mean and variance shifts for normally distributed data. Bayesian schemes are also used for self-starting process monitoring, with the uncertainty of process parameters represented by suitable priors. Tsiamyrtzis and Hawkins (2005, 2008) provided a closed form solution of the posterior distribution of process means affected by random walks and abrupt jumps. Apley (2012) dealt with a similar problem with a graphical implementation. Compared with these works for continuous data, the self-starting monitoring approaches for count processes are quite limited. Quesenberry (1991a, 1995b) built shewhart-type charts based on normal approximations for Poisson and Binomial data, and the monitoring performance can be improved if CUSUM and EWMA setups are employed. Shen et al. (2015) proposed to inspect possible mean shifts of Poisson data by heuristic tests. But the method requires historical data to initialize the monitoring.

Outliers with values notably deviated from the in-control state are commonly

observed in many real world processes. As explained by Hawkins et al. (2003), outliers are generated by isolated special causes. Here we use an example from asthma patient health care monitoring to illustrate the motivation for this research. Asthma patients use a cellphone App to record the time when they need the assistance from rescue inhalers, and the records are synchronized to a terminal for medical staff. The daily inhaler usage counts are used as indicators of their health condition. Once abnormal increase of the usage count is observed, doctors would consider medical intervention. Patients' records undergoes surges caused by factors other than symptom aggravation, e.g., errors in data transmission and air condition deterioration. Fig. 1.1 shows the record of two patients, referred to as PA and PB. The count sequence of PA can be clearly split into two segments divided by a change at day 6; however, for PB, we notice the count suddenly raises to a peak at day 4 but immediately goes back to normal afterwards, which is found to be caused by the patient's misoperation after we checked the original database, i.e., an outlier. For more examples in SPC, see Zhang and Albin (2009), and Jensen et al. (2006) for examples in SPC, and Zhang (1998) and Steiner et al. (2009) for more general applications. Such outliers are more common in early process readings (Jones-Farmer et al., 2014). Therefore if self-starting SPC methods are applied, for example, in Phase I analysis to reduce the need for potential costly samples, outliers should be treated with concern. However, outliers are not well regarded in existing self-starting monitoring frameworks, and outliers are often dealt with in a retrospective manner (Jones-Farmer et al., 2014). This paper is inspired to fill the lack of self-starting SPC methods for count data, particularly those being



(a) PA (b) PB

Figure 1.1: Asthma Patient Inhaler Usage Count

robust to outliers without sacrificing the fast detection ability to sustained shifts. To reach this goal, we first propose a hidden Markov model (HMM) for the transition of the process states. Then we propose to use a particle filter (Arulampalam et al., 2002; Del Moral et al., 2006), which is a Bayesian sequential Monte Carlo (SMC) method, to estimate the posterior probability of the process jumping to the out of control state. And a simple but useful decision rule is used to make decisions on the state of the process. An efficient Rao-Blackwellization strategy is developed to improve the efficiency of the particle filter method. A Bayesian scheme is used here since first it fits the self-starting problem naturally (Tsiamyrtzis and Hawkins, 2005), and secondly, the HMM formulation also facilitates Bayesian inference.

1.2 Dimensional Integrity and Fault Diagnosis

In manufacturing area, dimensional quality measures the integrity between the actual product dimensions and the design nominal. Dimensional faults refer to that the variability the dimensional measurements of products exceeds the tolerance limit. The excessive variability could be introduced from multiple variation sources in assembly and machining stages, e.g., positioning of fixture locators, alignment of machine tools, and deformation or deterioration of compliant parts (Zhou and Chen, 2005). To obtain close inspection on product dimensional quality, the measurement data is collected by automatic in-process sensing devices, e.g., the optical coordinate measuring machine (OCMM) widely built in autobody assembly lines (Zhou et al., 2003). The collected data are then used to help ensure the dimensional integrity.

SPC (Montgomery, 2009) techniques are used to detect the deviation of the dimensional quality (see Wu* and Tian (2005), Castagliola (2005), Castagliola and Maravelakis (2011), etc). However, SPC itself has no diagnostic capability to locate the root causes of the deviation. In principle the diagnosis of the nonconformity of the dimensional integrity is closely related with variance components analysis (VCA). Research in recent years attempted to satisfy the dire need for the dimensional fault diagnosis. Shi and Ceglarek (1996) proposed to identify a single fault through principle component analysis (PCA). Apley and Shi (1998) discussed the modeling for fixture variation related faults in panel assembly and suggested a least sum of square variance estimation diagnostic algorithm, and further Zhou and Chen (2005) compared the characteristics of different variance estimators for diagnostic purpose. Zhou et al.

(2003) illustrated the diagnosability of dimensional faults in multistage assembly processes. Zhou et al. (2004) suggested a hypothesis testing approach based on minimal norm quadratic unbiased estimators (MINQUE) of variance components. Besides, to incorporate engineering expertise into the diagnosis, Bayesian approaches were also developed as alternatives; for example, Bastani et al. (2013) used an enhanced relevance vector machine (RVM), i.e., an empirical Bayesian algorithm, to obtain sparse estimate of the scale of the variation shift.

A common assumption among the aforementioned diagnostic approaches is that the quality measurements and process variation are related by mixed linear models, in which potential dimensional fault sources are accounted as random effects. Although these studies provides insights to diagnose dimensional faults from product measurements, the application is restricted because: firstly, the reported methods require all fault sources to be identifiable (Zhou et al., 2003; Zhou and Chen, 2005), which could be violated if variation sources outnumber the dimension of quality measurements; secondly, for hypothesis tests or variance estimation oriented diagnosis, it is usually difficult to determine the cut-off threshold value; thirdly, given multiple potential fault sources, the credibility of different diagnostic results is desired, which is inaccessible from these diagnostic frameworks. Therefore, considering the inadequacy of current study in dimensional faults diagnosis, in this thesis we propose a sparse Bayesian diagnosis, upon which variation sources of the shift are evaluated through Markov Chain Monte Carlo (MCMC) variable selection. More specifically, in the proposed diagnostic algorithm, fault patterns, i.e., set of variation sources of

the shift, of different dimensionalities are evaluated, so we resort to a reversible jump MCMC (RJ-MCMC) framework; afterwards, the fault patterns perceived as likely root causes are further checked according to their diagnosability. Compared with other diagnostic methods, the proposed approach has three advantages: firstly, the method estimates the marginal posterior probability of each fault pattern, rather than process variance component, so numeric regularization (Bastani et al., 2013) is not concerned; secondly, the diagnosability property can be integrated into the diagnosis; thirdly, the method prefers sparse results, which fit practical realities more (Bastani et al., 2013; Li and Chen, 2016).

1.3 Diagnosis of Multivariate Linear Profiles

In many practical SPC applications, the process quality is better characterized and summarized by the relationship between the response variable and one or more explanatory variables, rather than the specific univariate or multivariate distribution of the concerned quality variables. Therefore our concern transfers from the specific quality distribution to the relationship, i.e., the so called profiles. Usually, a simple but effective way is to use linear profiles to model the relationship. Therefore the study focusing on linear profiles receives the main interest in past decades (see, e.g, Woodall et al. (2004), Woodall (2007), Kim et al. (2003), Kang and Albin (2000), and Mahmoud et al. (2007)). Most of the previous work devotes to constructing efficient monitoring frameworks for linear profiles. Kang and Albin (2000) proposed two control charts for Phase II monitoring for simple linear profiles, namely a multivari-

ate T^2 chart and and a combined chart of an exponentially weighted moving average (EWMA)chart and a range chart. Zhu and Lin (2009) considered monitoring the slope of simple linear profiles in Phase I and Phase II. Mahmoud et al. (2007) considered changepoint methods for detecting changes in simple linear regression models. For multivariate linear profiles, or multivariate linear regression models, Kazemzadeh et al. (2008) and Mahmoud (2008) proposed Phase I methods for profiles represented by polynomial and multiple regression models, respectively, and Zou et al. (2007a) proposed a multivariate EWMA chart to monitor general linear profiles in Phase II.

The aforementioned work are all motivated by the requirement of monitoring, whilst the diagnosis of linear profiles, i.e., identifying the root causes, are not fully discussed. Statistical methods are designed for multivariate quality processes more commonly. Barton and Gonzalez-Barreto (1996) proposed the representation of a quality measurement vector as a linear combination of pre-identified fault patterns plus a residual. The fault diagnosis was achieved through online monitoring of the coefficients of the linear model; Runger et al. (2007) further studied how to estimate those coefficients considering whether the process-oriented effects occur only as special causes or also as common causes of variation. Some other traditional statistical methods attempt to locate the shifted profile components by interpreting and decomposing Hotelling's T^2-type statistic; for example, see Li et al. (2008). On the contrary, for the diagnosis of linear profiles, the available work is limited. Zou et al. (2011) proposed a LASSO-based diagnostic framework, which could be use to estimate the possible shifted profile components. Zou et al. (2007a) also suggested a diagnostic

aids for linear profile monitoring, based on parametric tests. However, the current diagnostic methods for linear profiles have several shortcomings: firstly, they depends on exact knowledge or accurate estimate of the shift position, or changepoint, of the process, and this assumption could be problematic in practice, and secondly, approximation and asymptotic properties are used in Zou et al. (2011), so the diagnostic effectiveness needs further examination when the dimension of the profile parameters is high, and the same concern happens for using F test on single profile parameters in Zou et al. (2007a) to screen shifted parameters. To satisfy the practical desire for multivariate linear profile diagnosis, we propose a novel Bayesian framework integrating the diagnosis of the changepoint and shifted profile parameters. We assume that the quality measurements follow a multivariate linear regression model, i.e., the regression coefficients are the profile parameters; further, a changepoint model is used to explain how the process is shifted on the profile parameters. Compared with the previous work, our approach is able to address the inference on the joint uncertainties of the change model and the changepoint altogether. In addition, our diagnosis employs a Bayesian variable selection idea, so it can provide *a posteriori* estimates of the change models (or fault patterns), rather than the statistical significance of single profile parameters.

The remaining of the thesis is organized as follows. In chapter 2, we provide the technique details of implementing the RBPF on monitoring Poisson and Binomial sequences, with robustness to the presence of outliers. In chapter 3, we consider dimensional faults modeled as random effects in manufacturing, and discuss how the

Bayesian diagnosis obtains sparse estimates of fault patterns. In chapter 4, we develop a Bayesian diagnosis on piece-wise multivariate linear profile models, which aims to identify the shift position and pattern of profile coefficients.

CHAPTER 2
A BAYESIAN STATISTICAL PROCESS CONTROL METHOD FOR COUNT DATA IN THE PRESENCE OF OUTLIERS

Most self-starting SPC methods consider only sustained shifts. However, in practice outliers could appear as temporary deviations from the in-control state, which brings challenges to most self-starting control charts. In this paper, we propose a new online Bayesian framework to detect the out of control state for count data in the presence of outliers. Based on a hidden Markov model (HMM) for process states, we are able to use a PF approach for robust process monitoring. We demonstrate how a RBPF can be used to control computational costs and improve inference efficiency. Based on the results of simulation and case study, the method is more preferable to traditional self-starting SPC methods.

The remainder of this chapter is organized as follows. In Section 2.1 we formally define the proposed statistical models. In Section 2.2, following a brief introduction of particle filter (PF) scheme and Rao-Blackwellization, we develop the Rao-Blackwellized particle filter (RBPF) for Poisson and Binomial data respectively. In Section 2.3, we illustrate the advantages of our method over traditional self-starting SPC methods via simulations. In Section 2.4 we further demonstrate the value of the RBPF by two applications on real data sets. Conclusions are given in Section 2.6.

2.1 Statistical Model

We respectively consider two popular count data models, namely Poisson and Binomial sequence:

$$y_t \sim Poiss(\theta_t) \qquad t = 1, 2, \cdots, T, \tag{2.1a}$$

$$y_t \sim Binom(n, \theta_t) \qquad t = 1, 2, \cdots, T, \tag{2.1b}$$

where n is the constant sample size parameter known to us, and θ_t is unknown in both models. Suppose the value of θ_t is determined by the latent process state $S_t \in \{0, 1, 2\}$, which represents "in control", "outlier" and "out of control" states respectively, and conditioning on the process states S_t, θ_t follows:

$$\theta_t = \begin{cases} \theta_{IC} & \text{if} \quad S_t = 0 \\ X_t & \text{if} \quad S_t = 1 \\ \theta_{OC} & \text{if} \quad S_t = 2 \end{cases}, \tag{2.2}$$

in which θ_{IC}, θ_{OC} are unknown constants representing in control and out of control values of θ_t, respectively, and $X_t \sim D(\theta; \boldsymbol{\xi}_D)$ ($\boldsymbol{\xi}_D$ are hyperparameters). We assume that the process state S_t propagate as a Markov chain with $p_0 = P(S_t = 1 | S_{t-1} = 0)$, $p_1 = P(S_t = 2 | S_{t-1} = 0)$, and $r = P(S_t = 2 | S_{t-1} = 1)$. Therefore the transition matrix of the Markov chain $\{S_t, t = 1, 2, \cdots\}$ can be represented as:

$$\mathbf{P} = (P_{ij}) = \begin{pmatrix} 1 - p_0 - p_1 & p_0 & p_1 \\ r & 1 - r & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

where

$$P(S_t = 2|S_{t-1} = 2) = 1, \tag{2.4a}$$

$$P(S_t = 2|S_{t-1} = 1) = 0, \tag{2.4b}$$

i.e., $S_t = 2$ is an absorbing state and the transition $S_{t-1} = 1 \rightarrow S_t = 2$ is not allowed. This is because: firstly, unless assignable causes are removed by external intervention, the out of control state cannot be reversed, that is, naturally $S_t = 2$ is a closed or absorbing state, and secondly, outliers immediately followed by out of control observations could be treated as out of control as well. Equations (2.1) to (2.2) connect measurements y_t, an intermediate variable θ_t, and an latent state S_t hierarchically as an HMM. In this context, the HMM is considered since its flexibility in modeling nonstationary processes (see examples in speech recognition (Rabiner and Juang, 1986), bioinformatics (Durbin et al., 1998), and novelty detection (Chatzis et al., 2009)). And this advantage facilitates the Bayesian filtering procedure discussed in Section 2.2.

2.2 Bayesian Monitoring Framework

Our primary goal of monitoring is to detect the occurrence of the out of control state in self-starting mode, based on sequentially observed $y_1, y_2, \cdots$. Among existing self-starting SPC methods, Bayesian frameworks naturally fit the above hierarchical modeling assumption. By contrast, frequentist methods based on sequential hypothesis testing could make false decisions on outliers (Rocke, 1989; Bakir, 2006).

Given $P(S_{t-1}|y_{1:t-1})$, $P(S_t|y_{1:t})$ can be calculated sequentially by:

$$P(S_t|y_{1:t}) \propto f(y_t|y_{1:t-1}, S_t) \times P(S_t|y_{1:t-1}), \tag{2.5}$$

where $f(y_t)$ is the likelihood function. For the right side of Equation (2.5), we have:

$$P(S_t|y_{1:t-1}) = \sum_{S_{t-1}} P(S_t|S_{t-1})P(S_{t-1}|y_{1:t-1}), \tag{2.6}$$

and

$$f(y_t|y_{1:t-1}, S_t) = \int f(y_t|\theta_t)\pi_\theta(\theta_t|y_{1:t-1}, S_t)d\theta_t, \tag{2.7}$$

where $P(S_t|S_{t-1})$ is given in Equation (2.3) and $\pi_\theta(\theta_t|y_{1:t-1}, S_t)$ is the posterior of θ_t by choosing a prior $\pi_\theta(\theta; \boldsymbol{\xi}_0)$ with hyperparameters $\boldsymbol{\xi}_0$. In fact, calculating $\pi_\theta(\theta_t|y_{1:t-1}, S_t)$ requires enumerating all possible transition routes of $S_{1:t-1}$, and the computational complexity is $O(3^t)$ up to t; meanwhile the closed form result of the integral in Equation (2.7) may not be available. Therefore exact Bayesian inference based on Equations (2.5) to (2.7) is generally intractable. As mentioned in the introduction section, we propose an approximate inference method based on particle filtering, which is an SMC method. Particle filter (PF) methods implement sequential Monte Carlo approximations to keep the computational cost grow as $O(t)$ (Doucet et al., 2001). In the following, we briefly introduce the standard PF method and the idea of Rao-Blackwellization to improve efficiency of particle filtering. Then we present the detailed implementation of Rao-Blackwellized particle filtering for Poisson and Binomial data.

2.2.1 Standard Particle Filtering

For inference of S_t, the process state at time t, a standard PF generates N particles represented by $\{\mathscr{P}_t^i = (S_t^i, \boldsymbol{\theta}_t^i)\}_{i=1}^N$, where S_t^is approximate posterior distribution of S_t, and the three components of $\boldsymbol{\theta}_t^i = (\theta_{0,t}^i, \theta_{1,t}^i, \theta_{2,t}^i)$ approximate posterior distributions of θ_{IC}, X_t, and θ_{OC}, respectively. The standard PF algorithm at time $t+1$ is given as follows:

1. generate N *i.i.d* samples of $\theta_{0,0}^i, \theta_{2,0}^i \sim \pi_\theta(\theta; \boldsymbol{\xi}_0)$ and $\theta_{1,0}^i \sim D(\theta; \boldsymbol{\xi}_D)$ for the initial particles $\{\mathscr{P}_0^i = (S_0^i = 0, \boldsymbol{\theta}_0^i)\}_{i=1}^N$, and assign each $\mathscr{P}_0^i$ with weight $\frac{1}{N}$;

2. assume at time t we have N particles $\{\mathscr{P}_t^i = (S_t^i, \boldsymbol{\theta}_t^i)\}_{i=1}^N$ with weights $\{w_t^i\}_{i=1}^N$ sum up to 1;

3. at time $t+1$, generating the descendant particles $\{\widetilde{\mathscr{P}}_{t+1}^i = (\tilde{S}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i)\}_{i=1}^N$ by transiting the component S_t^i in each $\mathscr{P}_t^i$ to $\tilde{S}_{t+1}^i$ with probability $P(\tilde{S}_{t+1}^i | S_t^i)$; $\tilde{\boldsymbol{\theta}}_{t+1}^i$ is the same as $\boldsymbol{\theta}_t^i$ except that an independent sample of the component $\tilde{\theta}_{1,t+1}^i$ is generated from $D(\theta; \boldsymbol{\xi}_D)$ if $\tilde{S}_{t+1}^i = 1$;

4. update the weights of the descendant particles by:

$$\begin{aligned} \tilde{w}_{t+1}^i &\propto w_t^i \times f(y_{t+1} | \widetilde{\mathscr{P}}_{t+1}^i) \\ &= w_t^i \times f(y_{t+1} | \tilde{S}_{t+1}^i, \tilde{\boldsymbol{\theta}}_{t+1}^i), \end{aligned} \tag{2.8}$$

   where $f(y_{t+1} | \widetilde{\mathscr{P}}_{t+1}^i)$ is the likelihood function of y_{t+1};

5. resample N particles $\mathscr{P}_{t+1}^i$ from $\widetilde{\mathscr{P}}_{t+1}^i$ with replacement according to the weight $\tilde{w}_{t+1}^i$, based on available resampling procedures including multinomial sampling,

stratified sampling, etc. (Tillé, 2011); then reassign $\tilde{w}_{t+1}^i$ of each resampled particle to be $\frac{1}{N}$;

6. approximately marginalize the posterior of the process state by:

$$P(S_{t+1} = l|y_{1:t+1}) \approx \sum_{i=1}^{N} \delta(S_{t+1}^i = l)/N, \tag{2.9}$$

where δ is the indicator function and $l = 0, 1, 2$.

The process is believed to be out of control if $P(S_{t+1} = 2|y_{1:t+1}) \geq M$. Based on our experiences, $M = 0.9$ is a reasonable setting. For all implementations of our PF algorithms in this paper, we use $M = 0.9$. All the initial particles have $S_0^i = 0$ and equal weights.

2.2.2 Rao-Blackwellized Particle Filter

The standard PF method is highly inefficient for our problem because of two reasons: (1) variance inflation: for each particle at t, only one descendant is generated at $t + 1$, which causes only a few particles at $t + 1$ having nontrivial weights and the variance of the particles' weights increasing with time (Arulampalam et al., 2002; Doucet et al., 2001; Chopin and Pelgrin, 2004). (2) degeneracy issue: in our problem two components of $\boldsymbol{\theta}_t^i$ are never renewed since the initialization, which results in lack of mixing properties of $\boldsymbol{\theta}_t^i$ (Chopin, 2007). Rao-Blackwellization is a strategy to improve efficiency of standard PF method based on marginalization. First, we exactly marginalize out the three possible transitions for S_{t+1} as given by $P(S_{t+1} = j|S_t), j = 0, 1, 2$, by generating all possible descendant particles for S_{t+1}, which leads to variance reduction of particles' weights (Chopin and Pelgrin, 2004; Chopin, 2007). Secondly,

by choosing conjugate priors for θ_{IC} and θ_{OC} (which are available for the Poisson and Binomial models), θ_t in Equation (2.7) can be integrated out analytically. Such operation can be done recursively so that the degeneracy issue is removed inherently because the samples/particles for $\boldsymbol{\theta}_t^i$s are not needed anymore. Assume conjugate priors $\pi_\theta(\theta;\boldsymbol{\xi}_{IC,0})$, $\pi_\theta(\theta;\boldsymbol{\xi}_{OC,0})$, and $D(\theta;\boldsymbol{\xi}_D)$ for θ_{IC}, θ_{OC}, and X_t respectively, and denote $\boldsymbol{\xi}_{IC,t}$ and $\boldsymbol{\xi}_{OC,t}$ as the *a posteriori* parameters for θ_{IC} and θ_{OC} at t. The Rao-Blackwellized particle filtering algorithm is presented as follows:

1. suppose one initial particle $\mathscr{P}_t^i = (S_0^1, \boldsymbol{\zeta}_0^1)$ with $S_0^1 = 0$, $\boldsymbol{\zeta}_0^1 = \{\boldsymbol{\xi}_{IC,0}^i, \boldsymbol{\xi}_{OC,0}^i\}$;

2. at t, assume we have N particles $\{\mathscr{P}_t^i = S_t^i, \boldsymbol{\zeta}_t^i\}_{i=1}^N$ with weights $\{w_t^i\}_{i=1}^N$, and $\boldsymbol{\zeta}_t^i = \{\boldsymbol{\xi}_{IC,t}^i, \boldsymbol{\xi}_{OC,t}^i\}$;

3. at $t+1$, depending on possible value of S_{t+1}, each $\mathscr{P}_t^i$ generates at most three new particles denoted by:

$$\widetilde{\mathscr{P}}_{t+1}^{j_0} = (\tilde{S}_{t+1}^{j_0} = 0, \tilde{\boldsymbol{\zeta}}_{t+1}^{j_0})$$
$$\widetilde{\mathscr{P}}_{t+1}^{j_1} = (\tilde{S}_{t+1}^{j_1} = 1, \tilde{\boldsymbol{\zeta}}_{t+1}^{j_1})$$
$$\widetilde{\mathscr{P}}_{t+1}^{j_2} = (\tilde{S}_{t+1}^{j_2} = 2, \tilde{\boldsymbol{\zeta}}_{t+1}^{j_2}),$$

and therefore $R(N) = \sum_i 3I(S_t^i = 0) + 2I(S_t^i = 1) + I(S_t^i = 2)$ new descendants $\{\widetilde{\mathscr{P}}_{t+1}^j\}_{j=1}^{R(N)}$ are generated in total. If $\tilde{S}_{t+1}^j = 0$ or 2, one corresponding component of $\tilde{\boldsymbol{\zeta}}_{t+1}^j$ needs to be derived analytically by conjugacy, and the other one remains unchanged. If $\tilde{S}_{t+1}^j = 1$, $\tilde{\boldsymbol{\zeta}}_{t+1}^j$ remains unchanged;

4. update the weights by:

$$\begin{aligned}\tilde{w}^j_{t+1} &\propto w^i_t \times P(\tilde{S}^j_{t+1}|S^i_t) \times f(y_{t+1}|\widetilde{\mathscr{P}}^j_{t+1}) \\ &= w^i_t \times P(\tilde{S}^j_{t+1}|S^i_t) \times f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t);\end{aligned} \tag{2.10}$$

5. use the optimal resampling method proposed by Fearnhead and Liu (2007) to resample N particles $\{\mathscr{P}^i_{t+1}\}^N_{i=1}$ with weights $\{w^i_{t+1}\}^N_{i=1}$ from $\{\widetilde{\mathscr{P}}^j_{t+1}\}^{R(N)}_{j=1}$. Based on Fearnhead and Liu (2007), this resampling method guarantees no duplicated particles are resampled and maximizes the diversity of particles. The detailed procedure of the optimal resampling method is given in Appendix A;

6. approximately marginalize the posterior of the process state by:

$$P(S_{t+1} = l|y_{1:t+1}) \approx \sum_1^N w^i_{t+1}\delta(S^i_{t+1} = l), \tag{2.11}$$

The out of control state detection decision is still made by checking if $P(S_{t+1} = 2|y_{1:t+1}) \geq M$ based on Equation (2.11). For all initial particles $\boldsymbol{\zeta}^i_0 = (\boldsymbol{\xi}_{IC,0}, \boldsymbol{\xi}_{OC,0})$. In some applications, although the exact value of θ_{IC} and θ_{OC} are unknown, it is reasonable to assume a relation between the out of control and in control parameters, e.g., $\theta_{OC} = k\theta_{IC}$ $(k > 1)$, where k is a known constant. If this ratio relation is assumed, the Rao-Blackwellized PF (RBPF) algorithm presented in this section can still be applied if we set $\boldsymbol{\zeta}^i_t = \boldsymbol{\xi}^i_t$ and $\boldsymbol{\xi}^i_0 = \boldsymbol{\xi}_{IC,0}$.

For implementing the RBPF in monitoring Poisson and Binomial data, we choose the priors of θ_{IC} and θ_{OC} to be the Gamma and Beta distributions respectively. Steps 2 and 3 in the RBPF algorithm are conducted based on the following lemmas.

Lemma 2.1. *For $y_t \sim Poiss(\theta_t)$, and a priori $\theta_{IC} \sim Gamma(\alpha_{IC,0}, \beta_{IC,0})$, $\theta_{OC} \sim Gamma(\alpha_{OC,0}, \beta_{OC,0})$, and $X_t \sim Gamma(\alpha_D, \beta_D)$, suppose in step 2 of the RBPF algorithm $\widetilde{\mathscr{P}}_{t+1}^j$ is a descendant of $\mathscr{P}_t^i$, and $\boldsymbol{\xi}_{IC,t}^i = (\alpha_{IC,t}, \beta_{IC,t})$, $\boldsymbol{\xi}_{OC,t}^i = (\alpha_{OC,t}, \beta_{OC,t})$, then:*

1. *if $\tilde{S}_{t+1}^j = 0$,*

   $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (\alpha_{OC,t+1}, \beta_{OC,t+1}) = \boldsymbol{\xi}_{OC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (\alpha_{IC,t+1}, \beta_{IC,t+1})$, *where*

$$\begin{aligned} \alpha_{IC,t+1} &= \alpha_{IC,t} + y_{t+1}, \\ \beta_{IC,t+1} &= 1/(\frac{1}{\beta_{IC,t}} + 1), \end{aligned} \tag{2.12}$$

   and

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{IC,t}^i = (\alpha_{IC,t}, \beta_{IC,t})) \\ &\propto \frac{\Gamma(\alpha_{IC,t+1})\beta_{IC,t+1}^{\alpha_{IC,t+1}}}{\Gamma(\alpha_{IC,t})(\beta_t)^{\alpha_{IC,t}}}; \end{aligned} \tag{2.13}$$

2. *if $\tilde{S}_{t+1}^j = 2$*

   - *if $S_t^i = 0$,*

     $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (\alpha_{IC,t+1}, \beta_{IC,t+1}) = \boldsymbol{\xi}_{IC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (\alpha_{OC,t+1}, \beta_{OC,t+1})$, *where*

$$\begin{aligned} \alpha_{OC,t+1} &= \alpha_{OC,0} + y_{t+1}, \\ \beta_{OC,t+1} &= 1/(\frac{1}{\beta_{OC,0}} + 1), \end{aligned} \tag{2.14}$$

     and

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{OC,0} = (\alpha_{OC,0}, \beta_{OC,0})) \\ &\propto \frac{\Gamma(\alpha_{OC,t+1})\beta_{OC,t+1}^{\alpha_{OC,t+1}}}{\Gamma(\alpha_{OC,0})(\beta_t)^{\alpha_{OC,0}}}; \end{aligned} \tag{2.15}$$

- *if* $S_t^i = 2$,

  $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (\alpha_{IC,t+1}, \beta_{IC,t+1}) = \boldsymbol{\xi}_{IC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (\alpha_{OC,t+1}, \beta_{OC,t+1})$, *where*

$$\begin{aligned} \alpha_{OC,t+1} &= \alpha_{OC,t} + y_{t+1}, \\ \beta_{OC,t+1} &= 1/(\frac{1}{\beta_{OC,t}} + 1), \end{aligned} \tag{2.16}$$

  and

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{IC,t}^i = (\alpha_{OC,t}, \beta_{OC,t})) \\ &\propto \frac{\Gamma(\alpha_{OC,t+1})\beta_{OC,t+1}^{\alpha_{OC,t+1}}}{\Gamma(\alpha_{OC,t})(\beta_t)^{\alpha_{OC,t}}}; \end{aligned} \tag{2.17}$$

3. *if* $\tilde{S}_{t+1}^j = 1$,

   $\tilde{\boldsymbol{\zeta}}_{t+1}^j = \boldsymbol{\zeta}_t^i$, *and*

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_D = (\alpha_D, \beta_D)) \\ &\propto \frac{\Gamma(\alpha_D + y_{t+1})(\frac{1}{\beta_D} + 1)^{-(\alpha_D + y_{t+1})}}{\Gamma(\alpha_D)(\beta_D)^{\alpha_D}}. \end{aligned} \tag{2.18}$$

Lemma 2.1 is based on the results for the Poisson-Gamma Bayesian model (Gelman et al., 2014). Such results when $\theta_{OC} = k\theta_{IC}$ is given in Lemma 2.2.

Lemma 2.2. *For* $y_t \sim Poiss(\theta_t)$, $\theta_{OC} = k\theta_{IC}$, *and under the same prior settings for* θ_{IC} *in Lemma 2.1, suppose in step 2 of the RBKF algorithm* $\widetilde{\mathscr{P}}_{t+1}^j$ *is a descendant of* $\mathscr{P}_t^i$, *and* $\boldsymbol{\xi}_t^i = (\alpha_t, \beta_t)$, *then:*

1. *if* $\tilde{S}_{t+1}^j = 0$ *or* $\tilde{S}_{t+1}^j = S_t^i = 2$,

   $\tilde{\boldsymbol{\xi}}_{t+1}^j = (\alpha_{t+1}, \beta_{t+1})$, *where*

$$\begin{aligned} \alpha_{t+1} &= \alpha_t + y_{t+1}, \\ \beta_{t+1} &= 1/(\frac{1}{\beta_t} + 1), \end{aligned} \tag{2.19}$$

and

$$f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t) = f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\xi}^i_t = (\alpha_t, \beta_t)) \propto \frac{\Gamma(\alpha_{t+1})\beta^{\alpha_{t+1}}_{t+1}}{\Gamma(\alpha_t)(\beta_t)^{\alpha_t}}; \tag{2.20}$$

2. *if* $\tilde{S}^j_{t+1} = 1$,

   $\tilde{\boldsymbol{\xi}}^j_{t+1} = \boldsymbol{\xi}^i_t$, *and*

$$f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t) = f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\xi}_D = (\alpha_D, \beta_D)) \propto \frac{\Gamma(\alpha_D + y_{t+1})(\frac{1}{\beta_D} + 1)^{-(\alpha_D + y_{t+1})}}{\Gamma(\alpha_D)(\beta_D)^{\alpha_D}}; \tag{2.21}$$

3. *if* $\tilde{S}^j_{t+1} = 2$ *and* $S^i_t = 0$,

   $\tilde{\boldsymbol{\xi}}^j_{t+1} = (\alpha_{t+1}, \beta_{t+1})$, *where*

$$\begin{aligned} \alpha_{t+1} &= \alpha_t + y_{t+1}, \\ \beta_{t+1} &= 1/(\frac{1}{k\beta_t} + 1), \end{aligned} \tag{2.22}$$

   and:

$$f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t) = f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\xi}^i_t = (\alpha_t, \beta_t)) \propto \frac{\Gamma(\alpha_{t+1})\beta^{\alpha_{t+1}}_{t+1}}{\Gamma(\alpha_t)(k\beta_t)^{\alpha_t}}. \tag{2.23}$$

Equations (2.22) to (2.23) can be proved based on the fact that if $\theta_t \sim Gamma(\alpha_t, \beta_t)$, then $k\theta_t \sim Gamma(\alpha_t, k\beta_t)$. Note that from Lemma 2.2, when $\tilde{S}^j_{t+1} = 2$, $\tilde{\boldsymbol{\xi}}^j_{t+1}$ represents the $\boldsymbol{\xi}_{OC,t}$ in 2.1. For $y_t \sim Binom(n, \theta_t)$, the corresponding results used in steps 2 and 3 of the RBKF algorithm are given in Lemma 2.3 and Lemma 2.4.

Lemma 2.3. *For $y_t \sim Binom(n, \theta_t)$, and the priors $\theta_{IC} \sim Beta(a_{IC,0}, b_{IC,0})$, $\theta_{OC} \sim Beta(a_{OC,0}, b_{OC,0})$ and $X_t \sim Gamma(a_D, b_D)$, suppose in step 2 of the RBKF algorithm $\widetilde{\mathscr{P}}_{t+1}^j$ is a descendant of $\mathscr{P}_t^i$, and $\boldsymbol{\xi}_{IC,t}^i = (a_{IC,t}, b_{IC,t})$ $\boldsymbol{\xi}_{OC,t}^i = (a_{OC,t}, b_{OC,t})$, then:*

1. *if $\tilde{S}_{t+1}^j = 0$,*

   $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (a_{OC,t+1}, b_{OC,t+1}) = \boldsymbol{\xi}_{OC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (a_{IC,t+1}, b_{IC,t+1})$, *where*

   $$\begin{aligned} a_{IC,t+1} &= a_{IC,t} + y_{t+1}, \\ b_{IC,t+1} &= b_{IC,t} + n - y_{t+1}, \end{aligned} \tag{2.24}$$

   and

   $$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{IC,t}^i = (a_{IC,t}, b_{IC,t})) \\ &\propto \frac{B(a_{IC,t+1}, b_{IC,t+1})}{B(a_{IC,t}, b_{IC,t})}, \end{aligned} \tag{2.25}$$

   where $B(\cdot)$ is the beta function;

2. *if $\tilde{S}_{t+1}^j = 2$,*

   - *if $S_t^j = 0$,*

     $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (a_{IC,t+1}, b_{IC,t+1}) = \boldsymbol{\xi}_{IC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (a_{OC,t+1}, b_{OC,t+1})$, *where*

     $$\begin{aligned} a_{OC,t+1} &= a_{OC,0} + y_{t+1}, \\ b_{OC,t+1} &= b_{OC,0} + n - y_{t+1}, \end{aligned} \tag{2.26}$$

     and

     $$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{OC,0} = (\alpha_{OC,0}, \beta_{OC,0})) \\ &\propto \frac{B(a_{OC,t+1}, b_{OC,t+1})}{B(a_{OC,0}, b_{OC,0})} \end{aligned} \tag{2.27}$$

- *if* $S_t^j = 2$,

  $\tilde{\boldsymbol{\xi}}_{IC,t+1}^j = (a_{IC,t+1}, b_{IC,t+1}) = \boldsymbol{\xi}_{IC,t}^i$ *and* $\tilde{\boldsymbol{\xi}}_{OC,t+1}^j = (a_{OC,t+1}, b_{OC,t+1})$, *where*

$$\begin{aligned} a_{OC,t+1} &= a_{OC,t} + y_{t+1}, \\ b_{OC,t+1} &= b_{OC,t} + n - y_{t+1}, \end{aligned} \tag{2.28}$$

  and

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_{OC,t}^i = (a_{OC,t}, b_{OC,t})) \\ &\propto \frac{B(a_{OC,t+1}, b_{OC,t+1})}{B(a_{OC,t}, b_{OC,t})}, \end{aligned} \tag{2.29}$$

3. *if* $\tilde{S}_{t+1}^j = 1$,

   $\tilde{\boldsymbol{\zeta}}_{t+1}^j = \boldsymbol{\zeta}_t^i$, *and*

$$\begin{aligned} f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\zeta}_t^i) &= f(y_{t+1}|\tilde{S}_{t+1}^j, \boldsymbol{\xi}_D = (a_D, b_D)) \\ &\propto \frac{B(a_D + y_{t+1}, b_D + n - y_{t+1})}{B(a_D, b_D)}. \end{aligned} \tag{2.30}$$

Similar to 2.1, for Lemma 2.3, readers can also refer to the Beta-Binomial model Gelman et al. (2014). The results corresponding to $\theta_{OC} = k\theta_{IC}$ for Binomial distribution are given in Lemma 2.4.

Lemma 2.4. *For* $y_t \sim Binom(n, \theta_t)$, $\theta_{OC} = k\theta_{IC}$, *and under the same settings for the priors in Lemma 2.3, suppose in step 2 of the RBKF algorithm* $\widetilde{\mathscr{P}}_{t+1}^j$ *is a descendant of* $\mathscr{P}_t^i$ *and* $\boldsymbol{\xi}_t^i = (a_t, b_t)$, *then:*

1. *if* $\tilde{S}_{t+1}^j = 0$ *or* $S_t^i = \tilde{S}_{t+1}^j = 2$,

   $\tilde{\boldsymbol{\xi}}_{t+1}^j = (a_{t+1}, b_{t+1})$, *where*

$$\begin{aligned} a_{t+1} &= a_t + y_{t+1}, \\ b_{t+1} &= b_t + n - y_{t+1}, \end{aligned} \tag{2.31}$$

and

$$f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t) = f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\xi}^i_t = (a_{IC,t}, b_{IC,t})) \propto \frac{B(a_{t+1}, b_{t+1})}{B(a_t, b_t)}, \tag{2.32}$$

2. *if* $\tilde{S}^j_{t+1} = 1$,

$\tilde{\boldsymbol{\zeta}}^j_{t+1} = \boldsymbol{\zeta}^i_t$, *and*

$$f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\zeta}^i_t) = f(y_{t+1}|\tilde{S}^j_{t+1}, \boldsymbol{\xi}_D = (a_D, b_D)) \propto \frac{B(a_D + y_{t+1}, b_D + n - y_{t+1})}{B(a_D, b_D)}. \tag{2.33}$$

If $\tilde{S}^j_{t+1} = 2$ and $S^i_t = 0$, since the consistence of the distributions of θ_{IC} and $k\theta_{IC}$ is unavailable for Binomial distribution, we propose a convenient approximation method for the case when $\theta_{OC} = k\theta_{IC}$. We propose using $Beta(\hat{a}_t, \hat{b}_t)$ as a second order moment-based approximation for the distribution of $k\theta_{OC}$, where $(\hat{a}_t, \hat{b}_t)$ can be calculated by solving the following equations:

$$\begin{aligned} \frac{\hat{a}_t}{\hat{a}_t + \hat{b}_t} &= \frac{ka_t}{a_t + b_t}, \\ \frac{\hat{a}_t\hat{b}_t}{(\hat{a}_t + \hat{b}_t)^2(\hat{a}_t + \hat{b}_t + 1)} &= \frac{k^2 a_t b_t}{(a_t + b_t)^2(a_t + b_t + 1)}; \end{aligned} \tag{2.34}$$

Based on this approximation, steps 2 and 3 of the RBKF algorithm for the case when $\theta_{OC} = k\theta_{IC}$ can be implemented by substituting $\hat{\boldsymbol{\xi}}_t = (\hat{a}_t, \hat{b}_t)$ for $\boldsymbol{\xi}_{OC,0}$ in Equations (2.26) and (2.27).

Readers may be curious about choosing the Bayes factor (Kass and Raftery, 1995) to define the decision rule. Actually, we tested the feasibility of a decision rule based on using the following Bayes factor:

$$BF = \frac{P(S_t = 2|y_{1:t})/P(S_t \neq 2|y_{1:t})}{P(S_t = 2)/P(S_t \neq 2)}.$$

However, because $P(S_t = 2|y_{1:t})$ has small value before the shift occurs, a slight change on the ratio could cause substantial difference in BF, making it difficult to propose a proper robust cutoff value for the decision rule.

2.3 Simulation Study

In this section, we investigate the performance of the RBPF via simulation studies. To demonstrate the advantage of the proposed Bayesian approach, two existing self-starting schemes, namely the Q-CUSUM and Q-EWMA charts introduced in Quesenberry (1995b,a) are also tested for comparison purpose. The monitoring performance is assessed by two performance measures:

- the false alarm rate ($\text{FAR}_{L_{IC}}$)—the probability of generating a false alarm within the first L_{IC} observations before the shift, and
- the detection delay ($\text{DD}_{L_{IC}}$)—the average out of control run length, i.e., the average number of out of control observations required to detect the shift, when the length of the pre-shift period equals L_{IC}.

The Poisson and Binomial model parameters used in this simulation study are given in Table 2.1. To implement the RBPF monitoring, the same weakly informa-

Table 2.1: Settings of Simulation Process

	Poisson model	Binomial model
in control	$\theta_{IC} = 7$	$n = 50, \theta_{IC} = 0.07$
out of control	$\theta_{OC,1} = 1.6 * \theta_{IC}$ or $\theta_{OC,2} = 2 * \theta_{IC}$	$n = 50, \theta_{OC,1} = 1.6 * \theta_{IC}$ or $\theta_{OC,2} = 2 * \theta_{IC}$

tive prior is used for $\pi_\theta(\theta; \boldsymbol{\xi}_{IC,0})$, $\pi_\theta(\theta; \boldsymbol{\xi}_{OC,0})$ and $D(\theta; \boldsymbol{\xi}_D)$, which is $Gamma(3, 3)$ for Poisson and $Beta(1, 1)$ for Binomial data. Meanwhile, the number of particles is $N = 300$ and p_0 and r in $\mathbf{P}$ are set as 0.05 and 0.95, respectively. For Q-charts, we adopt the same settings as in Quesenberry (1995a,b); that is, the reference value $k_s = 0.75$ (Q-CUSUM) and the smoothing parameter $\lambda = 0.25$ (Q-EWMA). Such settings are demonstrated to have satisfactory performance in detecting shifts of various scales in the reference. The RBPF and Q-charts are compared under various parameter settings. For each parameter settings, 20000 simulation runs are used. To appropriately address the trade-off between false alarm rates and detection delays, we set $\text{FAR}_{L_{IC}} = 0.05$ for all methods when no outlier exists in the pre-shift period, by tuning p_1 in $\mathbf{P}$ for the RBPF and the control limit H_s in Q-charts. With respect to the computational efficiency, the RBPF requires 30 milliseconds to process a dataset of 100 observations on a Intel Core i7 3.40GHz workstation, whereas Q-charts require 10 milliseconds.

2.3.1 Performance Comparison: Scenario I

In scenario I, we consider the simplest situation, where an outlier is manually is added at the midpoint (i.e., $L_{IC}/2$) of the pre-shift segment. Two different lengths of pre-shift period ($L_{IC} = 50$ and $L_{IC} = 100$) are considered. Outliers are randomly generated based on the corresponding count model with mean $\delta\theta_{IC}$; $\delta = 1, 1.5, 2, 2.5, 3, 4$ and $\delta = 1$ represents the case when no outlier exists.

Poisson data

Table 2.2 compares the $\text{FAR}_{L_{IC}}$ and $\text{DD}_{L_{IC}}$ performance of RBPF, Q-CUSUM, and Q-EWMA for Poisson data. $\text{FAR}_{L_{IC}}$ is not affected by the observations after the pre-shift segment, so it is not specified with respect to the value of θ_{OC}. As discussed earlier, we set $\text{FAR}_{L_{IC}}$ to be close to 0.050 for all three methods when there is no outlier in the pre-shift period ($\delta = 1$). When an outlier exits in the pre-shift period ($\delta > 1$), however, for all three methods the $\text{FAR}_{L_{IC}}$ are different from 0.050 due to the impact of the outlier. For a robust method, we hope the outlier will not cause substantially high FAR. From Table 3, the FAR performance of the RBPF method is the most robust to the existence of outliers, with $\text{FAR}_{L_{IC}}$ no greater than 0.061 for all cases. In contrast, the FAR of both Q-charts are quite sensitive to outliers. Particularly, the $\text{FAR}_{L_{IC}}$ of the Q-charts are increased to as high as above 0.8 and 0.7 for $L_{IC} = 50$ and $L_{IC} = 100$, respectively. Such a high false alarm rate makes it practically impossible for the self-starting Q-charts to bypass the outliers in order to detect the sustained shift in the out-of-control state. It is also interesting to note that, for the RBPF, when the outlier magnitude is large (e.g., $\delta = 4$) the $\text{FAR}_{L_{IC}}$ is in fact smaller than 0.050. The reason for this is that when delta is large, the RBPF can detect the outlier easily. And the obvious outlier divides the pre-shift period into two shorter periods of length $L_{IC}/2$, where the false alarm rate over each shorter period is much less than that of the longer period, leading to smaller $\text{FAR}_{L_{IC}}$ for the RBPF.

For detection delay ($\text{DD}_{L_{IC}}$) under both values of θ_{OC}, when there is no outlier ($\delta = 1$), the RBPF and Q-charts have very similar performance. However, for the

Table 2.2: Comparison scenario I: Poisson

Method	δ	$\text{FAR}_{L_{IC}}$		$\text{DD}_{L_{IC}}$			
		$L_{IC}=50$	$L_{IC}=100$	$L_{IC}=50$		$L_{IC}=100$	
				$\theta_{OC,1}$	$\theta_{OC,2}$	$\theta_{OC,1}$	$\theta_{OC,2}$
RBPF	1	0.050	0.050	6.18	2.76	6.93	3.14
	1.5	0.061	0.052	6.45	2.83	7.09	3.17
	2	0.055	0.057	6.64	2.89	7.16	3.20
	2.5	0.051	0.051	6.60	2.89	7.16	3.22
	3	0.041	0.045	6.66	2.84	7.13	3.19
	4	0.032	0.035	6.55	2.83	7.05	3.19
Q-CUSUM	1	0.049	0.049	6.27	3.17	6.68	3.48
	1.5	0.059	0.061	6.55	3.23	6.85	3.52
	2	0.112	0.072	6.73	3.28	6.91	3.54
	2.5	0.218	0.153	7.04	3.42	6.88	3.60
	3	0.412	0.289	7.30	3.45	7.05	3.66
	4	0.823	0.709	8.20	3.50	7.38	3.68
Q-EWMA	1	0.050	0.049	6.72	3.12	7.14	3.33
	1.5	0.061	0.058	6.92	3.25	7.22	3.42
	2	0.13	0.102	7.37	3.32	7.43	3.42
	2.5	0.266	0.196	7.87	3.44	7.69	3.57
	3	0.442	0.373	8.49	3.45	7.75	3.53
	4	0.803	0.745	9.96	3.64	8.12	3.62

cases when the outlier exists, the value of $\text{DD}_{L_{IC}}$ of the RBPF is consistently smaller than that of the Q-charts. As δ varies from 1.5 to 4, the detection delay of the Q-charts is increased steadily and significantly, whereas that of the RBPF is just increased slightly. This is because when the Q-charts are applied, the undetected outliers biases the estimate of in control parameters, which in turn hurt the effectiveness in detecting the out of control state.

This demonstrates that the proposed method can not only avoid false alarms caused by outliers, but also provide much more robust performance in detection of the out-of-control state.

Table 2.3: Comparison scenario I: Binomial

Method	δ	$\text{FAR}_{L_{IC}}$		$\text{DD}_{L_{IC}}$			
		$L_{IC}=50$	$L_{IC}=100$	$L_{IC}=50$		$L_{IC}=100$	
				$\theta_{OC,1}$	$\theta_{OC,2}$	$\theta_{OC,1}$	$\theta_{OC,2}$
RBPF	1	0.050	0.050	9.88	4.81	13.02	5.79
	1.5	0.052	0.051	10.17	4.92	13.44	6.09
	2	0.054	0.046	10.43	5.03	13.69	6.14
	2.5	0.050	0.042	10.74	5.10	13.91	6.23
	3	0.044	0.034	11.27	5.18	14.22	6.23
	4	0.028	0.022	12.26	5.27	14.61	6.38
Q-CUSUM	1	0.050	0.050	15.21	6.42	14.05	6.65
	1.5	0.054	0.051	16.64	6.88	14.30	7.03
	2	0.057	0.058	16.78	6.91	14.55	7.18
	2.5	0.077	0.070	18.8	7.05	14.74	7.23
	3	0.108	0.076	19.79	7.32	15.40	7.26
	4	0.223	0.15	23.37	7.56	15.92	7.37
Q-EWMA	1	0.050	0.050	15.62	6.52	18.50	6.92
	1.5	0.056	0.052	23.31	6.82	19.14	6.95
	2	0.077	0.059	25.15	6.96	19.55	7.05
	2.5	0.136	0.096	27.17	7.06	20.05	7.16
	3	0.193	0.164	29.27	7.46	20.54	7.26
	4	0.452	0.361	31.57	7.79	20.95	7.46

Binomial data

Similar patterns appear in the results of monitoring Binomial data, as shown in Table 2.3. When $\delta = 4$, the false alarm rate increases to 0.223 and 0.452 for Q-CUSUM and Q-EWMA ($L_{IC} = 50$), respectively, whereas that of the RBPF is less than 0.054 for all δs. The RBPF also outperforms the Q-charts in terms of detection delay, regardless of the existence of the outlier. In addition, compared with the results in Table 2.2, the outlier causes more substantial detrimental effects on the $\text{DD}_{L_{IC}}$ of the Q-charts, especially when especially when $\theta_{OC} = \theta_{OC,1}$; for example, when $\delta = 4$ and $L_{IC} = 50$, the detection delay is increased to 23.37 for Q-CUSUM and 31.57 for Q-EWMA, respectively, whereas the RBPF has a detection delay of 12.26.

2.3.2 Performance Comparison: Scenario II

In this section we considering outliers occurring randomly within the pre-shift segment. Each observation in the pre-shift segment is generated as an outlier with probability $p = 0.025$ and 0.05. Moreover, considering the independence among the scales of outliers, we generate the mean of each outlier from a hyperprior. Note here the hyperprior is only used in generating outliers and is not involved in monitoring stage. To investigate the robustness of the three methods to possibly multiple outliers, we set $L_{IC} = 100$, $\theta_{OC} = \theta_{OC,1}$.

Poisson data

For the Poisson model, the mean of outliers is generated from $Gamma(\cdot; \boldsymbol{\xi}')$. $\boldsymbol{\xi}'$ is set as $(96, 0.25)$ to make the value of outliers apparently different from that of in control observations.

Monitoring under scenario II is more challenging for the three SPC methods. From Table 2.4, the false alarm rates for the PBRF are higher than that under scenario I because under scenario II the outliers are expected to occur much more frequently. For example, when $p = 0.05$, in average five outliers are expected in the pre-shift period. Nonetheless, the RBPF still has much lower false alarm rates compared with Q-CUSUM and Q-EWMA charts, which have $\text{FAR}_{L_{IC}} = 0.842$ and 0.853 respectively for $p = 0.05$. The RBPF also demonstrates superior detection speed. From Table 2.4, the $\text{DD}_{L_{IC}}$ of the RBPF is no greater than 7.43 for all cases whereas that value of the Q-charts can increase to 12.94. So far in the simulation studies we have

Table 2.4: Comparison Scenario II: Poisson

Method	p	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$
RBPF	0.025	0.085	7.29
	0.05	0.134	7.43
Q-CUSUM	0.025	0.640	7.54
	0.05	0.842	9.58
Q-EWMA	0.05	0.691	8.07
	0.05	0.853	12.94

Table 2.5: Scenario II: RBPF with Knowledge of k Available (Poisson)

Method	p	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$
RBPF	0.025	0.065	5.96
	0.05	0.094	6.16

applied the RBPF without assuming any knowledge of the shift magnitude k. In Section 2.2.2, we also present the RBPF algorithm if the knowledge of k is available. Next we test the performance of the RBPF algorithm incorporating the knowledge of $k = 1.6$ for data generated under setting of scenario II. From Table 2.5, we see that utilizing the knowledge of k in the RBPF method benefits both $\text{FAR}_{L_{IC}}$ and $\text{DD}_{L_{IC}}$. Being able to effectively utilize the shift magnitude information is another advantage of the RBPF over traditional self-starting SPC methods such as the Q-charts. We also notice that even when knowledge of k is not available, the proposed RBPF method can still provide satisfactory performance for most cases we tested.

Binomial data

For the Binomial model, we choose the hyperprior as $Beta(6, 30)$, and the results are presented in Table 2.6. Similar to the results in Table 2.4, the Q-charts generate false alarms with significantly higher probability than the RBPF methods. Also, for the Q-charts, the adverse impact of outliers on $\mathrm{DD}_{L_{IC}}$ is more significant. For example, the detection delays of the RBPF are no greater than 15.42, whereas that increases to 32.21 for Q-EWMA. The results when k is known as shown in Table 2.7 leads to the same conclusion as for Poisson data.

Table 2.6: Comparison Scenario II: Binomial

Method	p	$\mathrm{FAR}_{L_{IC}}$	$\mathrm{DD}_{L_{IC}}$
RBPF	0.025	0.055	14.22
	0.05	0.079	15.42
Q-CUSUM	0.025	0.088	15.61
	0.05	0.132	17.60
Q-EWMA	0.025	0.143	27.39
	0.05	0.212	32.21

Table 2.7: Scenario II: RBPF with Knowledge of k Available (Binomial)

Method	p	$\mathrm{FAR}_{L_{IC}}$	$\mathrm{DD}_{L_{IC}}$
RBPF	0.025	0.047	12.10
	0.05	0.052	13.27

2.3.3 Sensitivity Analysis of Parameters Setting of RBPF

An important practical question arises when using the RBPF, that is how to properly select the parameters. The main concern focus on (p_0, p_1, r) in $\mathbf{P}$. The choice of (p_0, p_1, r) affects $\text{FAR}_{L_{IC}}$ and $\text{DD}_{L_{IC}}$. For r, we typically assume that outliers occur more likely in an isolated way so we set $r \gg 1-r$. Based on our experience, given p_0, p_1, the monitoring performance is similar for any $r \in [0.95, 1)$. In this paper, we fix $r = 0.95$. In practice, a desired $\text{FAR}_{L_{IC}}$ should be specified by practitioners. Then we fix the value of p_0 or p_1 (p_0 in this paper) and the other parameter is determined by the specified using tuning simulations. To study the effects of p_0, we conduct simulation tests for various values of p_0 under scenario 1 and tune the value of p_1 correspondingly so that $\text{FAR}_{L_{IC}} = 0.05$ for $L_{IC} = 50$, $\delta = 1, 2, 2.5, 3$ and $\theta_{OC} = \theta_{OC,1}$.

Table 2.8 presents the monitoring performance with respect to different settings of p_0. It can be seen that, when there is an outlier, a larger value of p_0 leads to lower false alarm rate but higher detection delay. However, for all the tested p_0s, the RBPF outperforms the Q-charts (see Table 2.2). Similar trends are observed fro Binomial data, as shown in Table 2.9. For Binomial data, under all tested values of p_0, the $\text{FAR}_{L_{IC}}$ and $\text{DD}_{L_{IC}}$ are much better than those of the Q-charts.

2.4 Case Study

In this section, we implement the SPC methods to two real world examples. In the first example, we revisit the two data series shown in Fig. 1.1, which were collected for monitoring of asthma patient health status. Due to confidentiality considerations,

any information that might reveal the identity of patients have been removed and count data has been rescaled. Underdispersion can be observed within the count data of both the patients, so we choose to model the series by the Binomial model. We use $n = 24$, the number of hours during a day, as the sample size of the Binomial distribution since patients usually use inhalers at most once per hour. All the three SPC methods are designed with $FAR_{L_{IC}} = 0.05$ for $L_{IC} = 50$. The RBPF detects the out of control state at the 7th day for PA, whereas it does not signal for PB. The posterior probability of the out of control state at each time is given in Fig. 2.1 (up to the time when monitoring stops), and the detected shift is marked by the red vertical line. Compared with the RBPF, the Q-charts also detect the shift for PA at the same sample; however, they generate a false alarm at day 4 for PB. Therefore, the RBPF provides more prominent ability in the health management of asthma patients.

The second example is the bi-monthly aggregated counts of adverse events related to loss of therapy for one specific neurostimulator, as shown in Fig. 2.2. The data can be retrieved from the FDA MAUDE database Xu et al. (2015) assumes the data follows Poisson distribution and indicates a shift happened in June 2011

Table 2.8: Performance with Multiple p_0: Poisson

	$\delta = 1$		$\delta = 2$		$\delta = 2.5$		$\delta = 3$	
p_0	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$
0.10	0.05	6.65	0.045	7.09	0.039	7.12	0.033	7.08
0.08	0.05	6.50	0.049	6.97	0.044	6.95	0.038	6.82
0.06	0.05	6.26	0.055	6.74	0.048	6.79	0.039	6.72
0.05	0.05	6.18	0.061	6.64	0.055	6.60	0.041	6.66
0.03	0.05	5.89	0.066	6.29	0.075	6.33	0.060	6.49
0.01	0.05	5.73	0.074	6.30	0.090	6.40	0.082	6.25

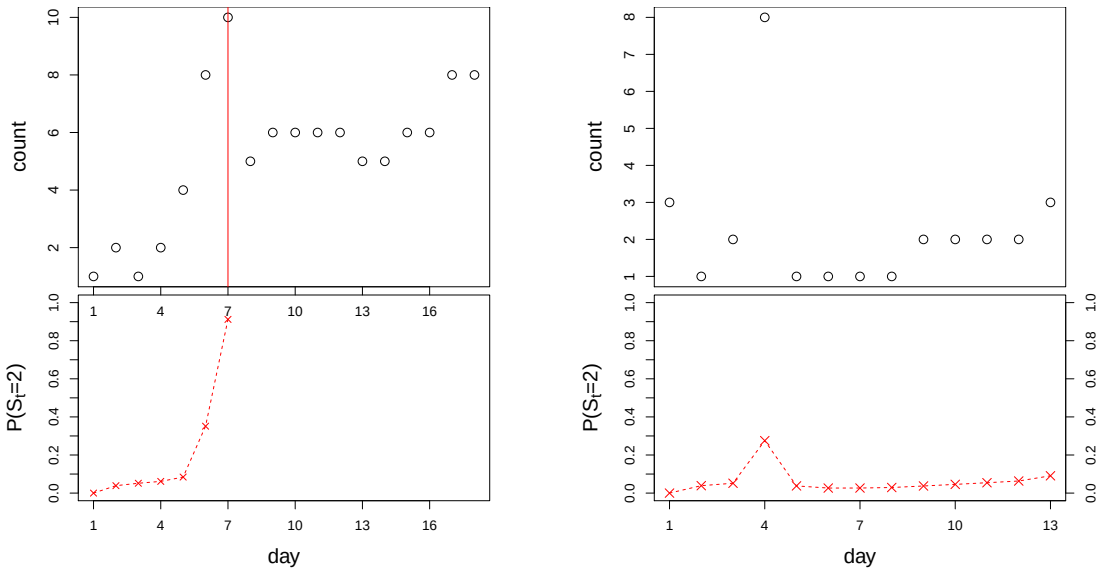
Table 2.9: Performance with Multiple p_0: Binomial

	$\delta = 1$		$\delta = 2$		$\delta = 2.5$		$\delta = 3$	
p_0	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$	$\text{FAR}_{L_{IC}}$	$\text{DD}_{L_{IC}}$
0.10	0.05	10.21	0.045	10.86	0.037	11.45	0.027	12.26
0.08	0.05	10.18	0.044	10.81	0.041	11.39	0.032	11.93
0.06	0.05	10.01	0.050	10.55	0.045	10.98	0.040	11.55
0.05	0.05	9.88	0.052	10.43	0.049	10.74	0.044	11.27
0.03	0.05	9.67	0.055	10.21	0.056	10.60	0.055	10.86
0.01	0.05	9.59	0.061	10.27	0.072	10.54	0.085	10.75

(marked by the red vertical line in Fig. 2.2). It can be seen that, before the shift at June 2011, there are two outliers at December 2009 and June 2010, respectively, which are marked by a circle in Fig. 2.2. To apply the RBPF, we choose $p_0 = 0.1$ and $Gamma(6, 10)$ as the hyperprior prior, and p_1 is tuned empirically as in Section 2.3, such that the RBPF has $FAR_{L_{IC}} = 0.05$ for $L_{IC} = 20$ and $\theta_{IC} = 60$. The posterior probability of being out of control is also provided in Fig. 2.2, against the original data. The RBPF detects the shift at the 3rd observation after the shift. The Q-charts are also adjusted accordingly. However, the Q-charts generate false alarms at the outliers.

2.5 Conclusions

We have proposed an online Bayesian framework to detect the abrupt shift in a sequence of count data. We consider outliers as challenges to SPC monitoring, and then use a hidden Markov chain to model the real process. Since implementing exact inference is not allowed here because the explosive computing cost, a RBPF with optimal resampling procedure becomes our choice, and it is performed on count data



(a) PA (b) PB

Figure 2.1: RBPF Monitoring of Asthma Patient Inhaler Usage Count

with procedures developed in Section 2.2.2. In addition, choosing weakly-informative prior for the process mean permits the self-starting property. The performance of the proposed RBPF has been evaluated through simulation studies by comparing with two traditional self-starting control charts. The results indicate that the proposed method has significant advantage over the traditional ones. Specifically, the proposed method controls the false alarm rate at an acceptable level, and it can be further improved if the knowledge of shift scale is available; the Q-charts, however, is incompetent to identify outliers from the shift. In addition, we have conducted the sensitivity analysis to provide guidelines to set parameters of the RBPF properly. Apart from the numeric investigation, we also have verified the practical value of the RBPF on

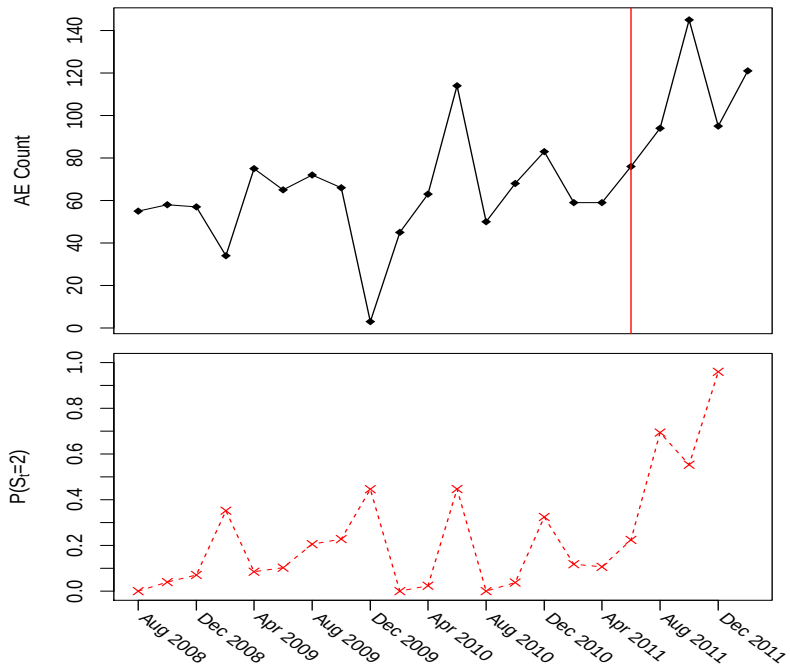


Figure 2.2: Adverse Event Count of Loss of Therapy for Neurostimulator

several real world applications.

2.6 Conclusions

We have proposed an online Bayesian framework to detect the abrupt shift in a sequence of count data. We consider outliers as challenges to SPC monitoring, and then use a hidden Markov chain to model the real process. Since implementing exact inference is not allowed here because the explosive computing cost, a RBPF with optimal resampling procedure becomes our choice, and it is performed on count data with procedures developed in Section 2.2.2. In addition, choosing weakly-informative prior for the process mean permits the self-starting property. The performance of the proposed RBPF has been evaluated through simulation studies by comparing with two traditional self-starting control charts. The results indicate that the proposed method has significant advantage over the traditional ones. Specifically, the proposed method controls the false alarm rate at an acceptable level, and it can be further improved if the knowledge of shift scale is available; the Q-charts, however, is incompetent to identify outliers from the shift. In addition, we have conducted the sensitivity analysis to provide guidelines to set parameters of the RBPF properly. Apart from the numeric investigation, we also have verified the practical value of the RBPF on several real world applications.

CHAPTER 3
PROCESS VARIATIONAL FAULTS DIAGNOSIS USING BAYESIAN METHOD

The diagnosis of dimensional faults in manufacturing is widely concerned but yet fully addressed. The diagnosis is challenging because: firstly, process variation sources are not explicitly accessible to dimensional measurements of products, and secondly, the diagnosability property of the fault. Considering the inadequacy of the diagnosis based on variance estimation, we propose a sparse Bayesian diagnostic framework. Essentially, a Bayesian variable selection is implemented by employing a reversible jump Markov Chain Monte Carlo algorithm to estimate the posterior distribution of fault patterns. We discuss the diagnosability and sparsity of the fault,and demonstrate the diagnostic efficacy of the proposed method on simulation data and practical applications.

The rest part of this Chapter is organized as follows. In Section 3.1, we construct the fault-quality model for diagnostic purpose. In Section 3.2 and 3.3, we demonstrate the background of the RJ-MCMC algorithm and the detailed implementation of the proposed diagnostic method. The diagnosability issue of variance deviation faults and the concept of sparse fault patterns are explained in Section 3.4. In Section 3.5 and 3.6, we illustrat the performance of the proposed method via simulation tests and an industrial case study. Conclusions are presented in Section 3.7.

3.1 Fault-quality Model

The first step is to develop a fault-quality diagnostic model that links process variations and product dimensional quality data. Linear fault-quality models are adopted to describe the variation propagation in multistage manufacturing (Zhou et al., 2003; Zhou and Chen, 2005). In this thesis, our priority is the diagnosis of dimensional faults. Therefore, the development of the linear fault-quality model from chaining state-space models of manufacturing stages is not explained here, and we just give the final mixed linear model. Suppose the dimensional measurement vector of product i is $\boldsymbol{Y}_i^{m\times 1}$, and process variation can be represented by the random effects vector $\boldsymbol{\beta}_i^{q\times 1}$, and the linear relationship is then:

$$\boldsymbol{Y}_i = \boldsymbol{B}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \cdots, N, \tag{3.1}$$

where $\boldsymbol{B}^{m\times p}$ is the system matrix known from manufacturing process kinetics, and $\boldsymbol{\epsilon}_i^{m\times 1} \sim Norm(\mathbf{0}, \sigma^2\boldsymbol{I})$ represents the measurement noise and unmodeled process errors. Discussions on $\boldsymbol{B}$ and $\boldsymbol{\beta}_i$ in manufacturing applications can be found in Apley and Shi (1998) and Zhou et al. (2003). Assume the random effects $\boldsymbol{\beta}_i \sim Norm(\mathbf{0}, \boldsymbol{\Sigma})$, and we also assume that the components in $\boldsymbol{\beta}_i$ be independent as in Zhou et al. (2003) for simplicity but without losing reasonableness, i.e., $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)$ where σ_j^2s are the variance components being diagnosed. $\boldsymbol{\Sigma}$ can be decomposed by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} \sigma_{0_1}^2 + \lambda_1^2 & & \\ & \ddots & \\ & & \sigma_{0_p}^2 + \lambda_p^2 \end{bmatrix}, \tag{3.2}$$

where $\sigma_{0_j}^2$s $(j = 1, 2, \cdots, p)$ are the design tolerance of σ_j^2s correspondingly, and λ_js represent the increased dimensional variability caused by faults ($\lambda_j > 0$ if σ_j^2 undergoes shift; otherwise $\lambda_j = 0$); usually $\sigma_{0_j} \ll \lambda_j$ if $\lambda_j > 0$. Denote $\boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{0_1}^2, \sigma_{0_2}^2, \cdots, \sigma_{0_p}^2)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$, then the covariance matrix of $\boldsymbol{Y}_i$ is expressed by:

$$\begin{aligned} \text{cov}(\boldsymbol{Y}_i) &= \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T + \sigma^2\boldsymbol{I} \\ &= \boldsymbol{B}\boldsymbol{\Sigma}_0\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \sigma^2\boldsymbol{I}. \end{aligned} \tag{3.3}$$

To interpret how $\boldsymbol{\Lambda}$ affects $\text{cov}(\boldsymbol{Y}_i)$, the dimension quality of product i, Equation (3.3) can be further rewritten as:

$$\begin{aligned} \text{cov}(\boldsymbol{Y}_i) &= \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T + \sigma^2\boldsymbol{I} \\ &= \boldsymbol{B}\boldsymbol{\Sigma}_0\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \sigma^2\boldsymbol{I} \\ &= \boldsymbol{B}\boldsymbol{\Sigma}_0\boldsymbol{B}^T + \sigma^2\boldsymbol{I} + \sum_{j=1}^{p} \boldsymbol{F}_j\lambda_j^2, \end{aligned} \tag{3.4}$$

and

$$\boldsymbol{F}_j = \boldsymbol{B}_{:j}\boldsymbol{B}_{:j}^T \quad j = 1, 2, \cdots, p, \tag{3.5}$$

where $\boldsymbol{B}_{:j}$ is the jth column of $\boldsymbol{B}$. Our target is to diagnose on which variance components faults occur given faults are detected by SPC tools. In other words, we should confirm the fault pattern $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_p]^T$, which is defined as an indicator vector such that $\gamma_j = 1$ if $\lambda_j > 0$ and 0 otherwise.

3.2 Hierarchical Bayesian Model

Before we introduce the Bayesian diagnosis, we first define the prior distributions for the unknown parameters in Equation (3.4) in a hierarchical order. We

assume that σ_j^2s undergo shift with probability p_{0j} independently like Steward et al. (2016) and Li and Chen (2016), so:

$$\pi_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) = \prod_{j=1}^{p} p_{0j}^{\gamma_j}(1-p_{0j})^{1-\gamma_j}, \tag{3.6}$$

and in this work we assume the deviation happens with equal chance on σ_j^2s, i.e., $p_{0j} = p_0$, and other options could be used if strongly suggested by historical experience. Next, assume λ_js independently and follow:

$$\pi_{\lambda}(\lambda_j|\gamma_j = 1) = Norm^{+}(\mu_{\lambda}, \sigma_{\lambda}^2), \tag{3.7}$$

where $Norm^{+}(\cdot; \mu_{\lambda}, \sigma_{\lambda})$ is a left truncated normal distribution with mean μ_{λ} and variance σ_{λ} so that λ_js are bounded as nonnegative. To complete the setting for priors, we set

$$\pi_{\sigma^2}(\sigma^2) = Inv\text{-}Gamma(c_0, d_0), \tag{3.8}$$

where (c_0, d_0) are hyperparameters. The settings of the priors will be discussed more specifically later. Let $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ be the vector containing nonzero λ_js, the posterior probability of a specific $\boldsymbol{\gamma}$ given $\boldsymbol{D} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T, \cdots, \boldsymbol{Y}_n^T]^T$ can then be marginalized from

$$P(\boldsymbol{\gamma}|\boldsymbol{D}) \propto \int_{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}} \int_{\sigma^2} L(\boldsymbol{D}|\sigma^2, \boldsymbol{\Sigma}_0, \boldsymbol{\Lambda}, \boldsymbol{\gamma}) \pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_{\boldsymbol{\gamma}}) \pi_{\sigma^2}(\sigma^2) \pi_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\sigma^2 d\boldsymbol{\lambda}_{\boldsymbol{\gamma}}, \tag{3.9}$$

where $L(\cdot)$ is the likelihood and $\pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_{\boldsymbol{\gamma}})$ is the joint prior of $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$.

3.3 Bayesian Variable Selection Based Diagnosis

Directly implementing diagnosis from Equation (3.9) is not feasible because: firstly, integrating over $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ and σ^2 is intractable; secondly, even the integration can

be handled numerically, the exhaustive calculation of $P(\boldsymbol{\gamma}|\boldsymbol{D})$ needs to consider all 2^p possible fault patterns, so the computational load is unaffordable when p is large. Therefore we have to adopt the MCMC idea. However, unlike usual MCMC applications, the inference of $\boldsymbol{\gamma}$ requires traversing across parameter spaces with varying dimensionality, i.e., the length of λ_γ changes as the Markov chain moves. So here we need to consider the reversible jump Markov Chain Monte Carlo (RJ-MCMC).

3.3.1 Reversible Jump Markov Chain Monte Carlo Method

The RJ-MCMC method is an extension of the Metropolis–Hastings Markov Chain Monte Carlo(MH-MCMC) algorithm. Suppose we consider a series of possible models M_k, $k \in \mathcal{K}$, of which the parameter vector $\boldsymbol{\theta}_k \in \boldsymbol{\theta}_k$ respectively. We abbreviate the pair $(k, \boldsymbol{\theta}_k)$ as $\boldsymbol{x}_k$. For example, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, in regular MH-MCMC, direct move between them is not allowed if they are of different dimensions, say, n_i and n_j, since the so called "dimension matching" requirement is violated (Green, 1995); however, in the RJ-MCMC framework this problem is solved by constructing a bijection with $\boldsymbol{\mu}_1 \in \mathbb{R}^{r_i \times 1}$ and $\boldsymbol{\mu}_2 \in \mathbb{R}^{r_j \times 1}$ given from densities g, g' respectively, such that $(\boldsymbol{\theta}_2, \boldsymbol{\mu}_2) = h(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)$ and $n_i + r_i = n_j + r_j$ (i.e., so called "dimensionality-matching"); $h(\cdot)$ is the bijection function. Denote the interested posteriors as $\pi(\boldsymbol{x}_1)$ and $\pi(\boldsymbol{x}_2)$, and the jumping proposal between two models is $q(\cdot|\cdot)$, Green (1995) and Green and Hastie (2009) showed that to achieve the required detailed balance:

$$\begin{aligned} &\int \pi(\boldsymbol{x}_1) g(\boldsymbol{\mu}_1) q(\boldsymbol{x}_1|\boldsymbol{x}_2) \alpha(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{\theta}_1 d\boldsymbol{\mu}_1 = \\ &\int \pi(\boldsymbol{x}_2) g'(\boldsymbol{\mu}_2) q(\boldsymbol{x}_2|\boldsymbol{x}_1) \alpha(\boldsymbol{x}_2, \boldsymbol{x}_1) d\boldsymbol{\theta}_2 d\boldsymbol{\mu}_2, \end{aligned} \tag{3.10}$$

$\alpha(\boldsymbol{x}_1, \boldsymbol{x}_2)$, the acceptance probability of the move $\boldsymbol{x}_1 \to \boldsymbol{x}_2$, should be:

$$\alpha(\boldsymbol{x}_1, \boldsymbol{x}_2) = \min\{1, \frac{\pi(\boldsymbol{x}_2)g'(\boldsymbol{\mu}_2)q(\boldsymbol{x}_2|\boldsymbol{x}_1)}{\pi(\boldsymbol{x}_1)g(\boldsymbol{\mu}_1)q(\boldsymbol{x}_1|\boldsymbol{x}_2))} \times \left|\frac{\partial(\boldsymbol{\theta}_2, \boldsymbol{\mu}_2)}{\partial(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)}\right|\}, \tag{3.11}$$

where the Jacobian factor is from the transformation from $(\boldsymbol{\theta}_2, \boldsymbol{\mu}_2)$ to $(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)$ and obviously depends on $h(\cdot)$; and the acceptance probability of the reverse move is:

$$\alpha(\boldsymbol{x}_2, \boldsymbol{x}_1) = \min\{1, \frac{\pi(\boldsymbol{x}_1)g(\boldsymbol{\mu}_1)q(\boldsymbol{x}_1|\boldsymbol{x}_2))}{\pi(\boldsymbol{x}_2)g'(\boldsymbol{\mu}_2)q(\boldsymbol{x}_2|\boldsymbol{x}_1)} \times \left|\frac{\partial(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)}{\partial(\boldsymbol{\theta}_2, \boldsymbol{\mu}_2)}\right|\}. \tag{3.12}$$

For simplicity, in such moves commonly r_i or r_j are set as 0; for example, if $n_i < n_j$, then only $\boldsymbol{\mu}_1$ is needed, and then Equation (3.11) are changed to:

$$\alpha(\boldsymbol{x}_1, \boldsymbol{x}_2) = \min\{1, \frac{\pi(\boldsymbol{x}_2)q(\boldsymbol{x}_2|\boldsymbol{x}_1)}{\pi(\boldsymbol{x}_1)g(\boldsymbol{\mu}_1)q(\boldsymbol{x}_1|\boldsymbol{x}_2))} \times \left|\frac{\partial(\boldsymbol{\theta}_2)}{\partial(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)}\right|\}, \tag{3.13}$$

and the acceptance probability of the reverse move is:

$$\alpha(\boldsymbol{x}_2, \boldsymbol{x}_1) = \min\{1, \frac{\pi(\boldsymbol{x}_1)g(\boldsymbol{\mu}_1)q(\boldsymbol{x}_1|\boldsymbol{x}_2))}{\pi(\boldsymbol{x}_2)q(\boldsymbol{x}_2|\boldsymbol{x}_1)} \times \left|\frac{\partial(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)}{\partial(\boldsymbol{\theta}_2)}\right|\}. \tag{3.14}$$

For detailed theoretical discussion of the RJ-MCMC, readers can refer to Brooks et al. (2003b) and Green and Hastie (2009).

3.3.2 Dignosis of Variance Components via RJ-MCMC

To diagnose $\boldsymbol{\gamma}$, we apply the RJ-MCMC method to generate R samples of the unknown variance components in Equation (3.4), denoted by $(\boldsymbol{\lambda}_{\gamma^{(1)}}, \sigma^2_{(1)})$, $(\boldsymbol{\lambda}_{\gamma^{(2)}}, \sigma^2_{(2)})$, ..., $(\boldsymbol{\lambda}_{\gamma^{(R)}}, \sigma^2_{(R)})$ and calculate:

$$P(\boldsymbol{\gamma}|\boldsymbol{D}) \approx \frac{\sum_{k=1}^{R} I(\boldsymbol{\gamma}^{(k)} = \boldsymbol{\gamma})}{R}, \tag{3.15}$$

where I is the indicator function.

To smoothly move the Markov chain on fault patterns with different dimensionality, here we adopt the the "birth-death" transition mode in Brooks et al. (2003b). That is, for the jump from $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ to $\boldsymbol{\lambda}_{\boldsymbol{\gamma}'}$, it is restricted to three categories:

- $\boldsymbol{\gamma} = \boldsymbol{\gamma}'$;
- $\boldsymbol{\gamma}'$ is different from $\boldsymbol{\gamma}$ by one component, say the jth one, such that $\boldsymbol{\gamma}_{-j} = \boldsymbol{\gamma}'_{-j}$ whereas $\gamma_j = 0$ and $\gamma'_j = 1$, i.e., $\#\boldsymbol{\gamma}' = \#\boldsymbol{\gamma} + 1$ (denote $\#\boldsymbol{\lambda}$ as the number of 1s in $\boldsymbol{\lambda}$) or $\{\boldsymbol{\lambda}_{\boldsymbol{\gamma}'}\} = \{\lambda'_j\} \cup \{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}\}$;
- $\boldsymbol{\gamma}'$ is different from $\boldsymbol{\gamma}$ at the jth component, such that $\boldsymbol{\gamma}_{-j} = \boldsymbol{\gamma}'_{-j}$ whereas $\gamma_j = 1$ and $\gamma'_j = 0$ i.e., $\#\boldsymbol{\gamma}' = \#\boldsymbol{\gamma} - 1$ or $\{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}\} = \{\lambda_j\} \cup \{\boldsymbol{\lambda}_{\boldsymbol{\gamma}'}\}$.

Therefore, $\boldsymbol{\gamma}'$ is either same as $\boldsymbol{\gamma}$, or a neighboring pattern of $\boldsymbol{\gamma}$. Denote the three above types of transition by 0, 1, -1, each of which is assumed to happen with equal probability. To keep the dimension matching, given transition 1 or -1, a component in γ, say γ_j, is accordingly chosen with its value switched, and we further assume γ_j is chosen uniformly at random among the potential candidates, e.g., $p = 3$ and $\boldsymbol{\gamma} = [1, 0, 0]^T$, $P(\gamma_j = \gamma_2) = P(\gamma_j = \gamma_3) = 0.5$ conditional on transition 1. The jumping probability between variance shift models then follows:

$$q(\boldsymbol{\gamma}'|\boldsymbol{\gamma}) = \begin{cases} \frac{1}{3} & \text{for transition 0} \\ \frac{1}{3\times(p-\#\boldsymbol{\gamma})} & \text{for transition 1} \\ \frac{1}{3\times\#\boldsymbol{\gamma}} & \text{for transition -1} \end{cases} . \tag{3.16}$$

Note that the transition is limited to two options on boundary patterns (i.e., $\boldsymbol{\gamma} = \mathbf{0}^{1\times p}$ or $\mathbf{1}^{1\times p}$) with equal chance. The complete move of the Markov chain alternatively

updates $\boldsymbol{\lambda}_\gamma$ and σ^2_γ to $\boldsymbol{\lambda}_{\gamma'}$ and $\sigma^2_{\gamma'}$ respectively by a Metropolis-Hastings sampler:

1. conditional the noise variance σ^2_γ and the transition type from $\{0, 1, -1\}$:

   - for transition 0, $\boldsymbol{\lambda}_{\gamma'}$ is generated from a proposal distribution $q_{\boldsymbol{\lambda}}(\cdot|\cdot)$ so that the acceptance ratio is:

$$r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) = \frac{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_{\gamma'}) \times \pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_{\gamma'}) \times q_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_\gamma|\boldsymbol{\lambda}_{\gamma'})}{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_\gamma) \times \pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_\gamma) \times q_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_{\gamma'}|\boldsymbol{\lambda}_\gamma)}, \tag{3.17}$$

     and such move is accepted with probability $\alpha(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) = \min(r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}), 1)$. For $q_{\boldsymbol{\lambda}}(\cdot|\cdot)$, we assume $\boldsymbol{\lambda}_{\gamma'}$ is generated from $Norm^+(\boldsymbol{\lambda}_\gamma, \tilde{\sigma}^2_\lambda \boldsymbol{I})$;

   - for transition 1, a "new" variance component mapping $\boldsymbol{u}_1$ in Equation (3.13) is generated, denoted as u. Assuming $g(u) = Norm^+(\tilde{\mu}_\lambda, \tilde{\sigma}^2_\lambda)$, the acceptance ratio follows:

$$\begin{aligned} r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) =& \frac{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_{\gamma'}) \times \pi_\lambda(u) \times q(\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma}|\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')}{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_\gamma) \times g(u) \times q(\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')|\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma})} \\ & \times \left| \frac{\partial(\boldsymbol{\lambda}_{\gamma'})}{\partial(\boldsymbol{\lambda}_\gamma, \lambda_j)} \right| \end{aligned} \tag{3.18}$$

     where

$$\frac{q(\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma}|\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')}{q(\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')|\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma})} = \frac{1/\#\boldsymbol{\gamma}'}{1/(p - \#\boldsymbol{\gamma})}$$

     and

$$\left| \frac{\partial(\boldsymbol{\lambda}_{\gamma'})}{\partial(\boldsymbol{\lambda}_\gamma, u)} \right| = 1,$$

     since the bijection between $\boldsymbol{\lambda}_{\gamma'}$ and $(\boldsymbol{\lambda}_\gamma, \lambda_j)$ is an identity map. The acceptance probability of $\boldsymbol{\lambda}_\gamma \to \boldsymbol{\lambda}_{\gamma'}$ is $\alpha(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) = \min(1, r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}))$;

   - for transition -1, a component in $\boldsymbol{\lambda}_\gamma$, say u again, is randomly chosen and removed, which is equivalent to adding u to $\boldsymbol{\lambda}_{\gamma'}$. Therefore, similar to

Equation (3.14), we have

$$r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) = \frac{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_{\gamma'}) \times g(u) \times q(\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma}|\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')}{L(\boldsymbol{D}|\sigma^2_{\boldsymbol{\lambda}_\gamma}, \boldsymbol{\lambda}_\gamma) \times \pi_\lambda(u) \times q(\boldsymbol{\lambda}_{\gamma'}, \boldsymbol{\gamma}')|\boldsymbol{\lambda}_\gamma, \boldsymbol{\gamma})} \times \left|\frac{\partial(\boldsymbol{\lambda}_{\gamma'}, u)}{\partial(\boldsymbol{\lambda}_\gamma)}\right|, \tag{3.19}$$

and we accept the move by $\alpha(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}) = \min(1, r(\boldsymbol{\lambda}_\gamma, \boldsymbol{\lambda}_{\gamma'}))$.

2. The update $\sigma^2_\gamma \to \sigma^2_{\gamma'}$ is straightforward after we sample $\boldsymbol{\lambda}_{\gamma'}$. We choose the proposal density $q_\sigma(\sigma^2_{\gamma'}|\sigma^2_\gamma) = \pi_\sigma(\sigma^2_{\gamma'})$, i.e., the prior distribution of σ^2; that is,

$$r(\sigma^2_\gamma, \sigma^2_{\gamma'}) = \frac{L(\boldsymbol{D}|\sigma^2_{\gamma'}, \boldsymbol{\lambda}_{\gamma'}) \times \pi_\sigma(\sigma^{-2}_{\gamma'}) \times \pi_\sigma(\sigma^2_\gamma)}{L(\boldsymbol{D}|\sigma^{-2}_\gamma, \boldsymbol{\lambda}_{\gamma'}) \times \pi_\sigma(\sigma^2_\gamma) \times \pi_\sigma(\sigma^2_{\gamma'})} = \frac{L(\boldsymbol{D}|\sigma^2_{\gamma'}, \boldsymbol{\lambda}_{\gamma'})}{L(\boldsymbol{D}|\sigma^{-2}_{\gamma'}, \boldsymbol{\lambda}_\gamma)} \tag{3.20}$$

and the acceptance probability is $\alpha(\sigma^2_\gamma, \sigma^2_{\gamma'}) = \min(1, r(\sigma^2_\gamma, \sigma^2_{\gamma'}))$.

To improve the acceptance rate, elementwise move with random order can be used on $\boldsymbol{\lambda}_\gamma$ for within-model transition (i.e., transition 0). Algorithm 3.1 gives the pseudocode for the implementation.

3.4 Diagnosability and Sparse Fault Pattern

3.4.1 Diagnosability of Variance Components

Naturally,the *maximum a posteriori* estimate of $\boldsymbol{\gamma}$ is considered as the diagnostic result, so we expect $P(\boldsymbol{\gamma}|\boldsymbol{D})$ to be distinctive enough. However, this cannot be guaranteed in many situations. To reach the correct diagnosis, we need the assistance of the identifiability, or diagnosability of variance components. Reader can refer to Zhou et al. (2003) for thorough discussion, and here we just adopt the related concepts under our specific problem framework. The Bayesian diagnosis is guaranteed to

Algorithm 3.1 RJMCMC based algorithm for variance components selection

Data: $\boldsymbol{Y}_i$, $i = 1, \cdots, N$

Input: $\boldsymbol{B}$, $(\mu_\lambda, \sigma_\lambda^2)$, (c_0, d_0), $(\tilde{\mu}_\lambda, \tilde{\sigma}_\lambda^2)$, and MCMC sample size R

Output: MCMC samples

Initialization: randomly assign $\boldsymbol{\lambda}_{\gamma^{(0)}}, \sigma_{\gamma^{(0)}}$

for $t = 1$ **to** R **do**

fix $\sigma^2_{\gamma^{(t-1)}}$ and randomly choose from transition 0, 1, -1

if transition 0 **then**

sample $\boldsymbol{\lambda}_{\gamma'}$ of same dimension as $\boldsymbol{\lambda}_{\gamma^{(t-1)}}$

if $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'}) \geq 1$ **then**

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$

else

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$ with probability $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'})$; otherwise $\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma^{(t-1)}}$

end if

else if transition 1 **then**

construct $\boldsymbol{\gamma}'$ by randomly switching one zero component $\gamma_j^{(t-1)}$ in $\boldsymbol{\gamma}^{(t-1)}$ to 1, and accordingly extending $\boldsymbol{\lambda}_{\gamma^{(t-1)}}$ to $\boldsymbol{\lambda}_{\gamma'}$ by λ_j in Equation (3.18)

if $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'}) \geq 1$ **then**

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$

else

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$ with probability $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'})$; otherwise $\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma^{(t-1)}}$

end if

else

construct $\boldsymbol{\gamma}'$ by randomly switching one nonzero component $\gamma_j^{(t-1)}$ in $\boldsymbol{\gamma}^{(t-1)}$ to 0

if $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'}) \geq 1$ **then**

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$

else

$\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma'}$ with probability $\alpha(\boldsymbol{\lambda}_{\gamma^{(t-1)}}, \boldsymbol{\lambda}_{\gamma'})$; otherwise $\boldsymbol{\lambda}_{\gamma^{(t)}} \leftarrow \boldsymbol{\lambda}_{\gamma^{(t-1)}}$

end if

end if

given $\boldsymbol{\lambda}_{\gamma^{(t)}}$ and sample $\sigma^2_{\gamma'}$

if $\alpha(\sigma^{-2}_{\gamma^{(t-1)}}, \sigma^2_{\gamma'}) \geq 1$ **then**

$\sigma^2_{\gamma^t} \leftarrow \sigma^2_{\gamma'}$

else

$\sigma^2_{\gamma^t} \leftarrow \sigma^2_{\gamma'}$ with probability $\alpha(\sigma^{-2}_{\gamma^{(t-1)}}, \sigma^2_{\gamma'})$; otherwise $\sigma^2_{\gamma^t} \leftarrow \sigma^{-2}_{\gamma^{(t-1)}}$

end if

end for

Take the initial R_b samples as burn-in period, and $P(\boldsymbol{\gamma}|\boldsymbol{D}) \approx \frac{\sum_{t=R_b+1}^{R} I(\gamma^{(t)}=\gamma)}{R-R_b}$

detected variance shift faults which are uniquely diagnosable, as defined in Definition 3.1.

Definition 3.1. Denote $\boldsymbol{\lambda} = \mathbf{1}^{p\times 1}\boldsymbol{\Lambda}$ (i.e., the diagonal of $\boldsymbol{\Lambda}$).λ_j is uniquely diagnosable, if $\forall\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2$ s.t.:

$$\lambda_{1_j} \neq \lambda_{2_j} \Rightarrow \mathrm{cov}(\boldsymbol{Y}_i)|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_1} \neq \mathrm{cov}(\boldsymbol{Y}_i)|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_2}. \tag{3.21}$$

That is, shift on uniquely diagnosable variance components changes the covariance of measurement $\boldsymbol{Y}_i$; if all variance components are uniquely diagnosable, the fault-quality model discussed above is called fully diagnosable; otherwise, it is partially diagnosable. Obviously, fully diagnosable models are most favored in practice in terms of diagnostic complexity. Denote

$$\boldsymbol{H} = (\boldsymbol{B}^T\boldsymbol{B}) \circ (\boldsymbol{B}^T\boldsymbol{B}) \tag{3.22}$$

where $\circ$ is the Hadamard product operator. The fully diagnosable property is determined by Lemma 3.1.

Lemma 3.1. *The fault-quality model defined by Equations* (3.1) *to* (3.3) *is fully diagnosable if and only if* $\boldsymbol{H}$ *is of full rank.*

Lemma 3.1 is proved in Appendix B. The full diagnosability can be more explicitly interpreted by associating it with the rank of $\boldsymbol{B}$:

Lemma 3.2. *If* $\boldsymbol{B}$ *is of full rank, then the fault-quality model defined by Equations* (3.1) *to* (3.3) *is fully diagnosable variance faults.*

The proof of Lemma 3.2 is given in Appendix C. The contrapositive of Lemma 3.2 explains how Definition 3.1 could be violated, though not intuitively: in practice, restricted by economic and technical reasons, usually the number of quality measurements on one product is likely less than that of dimensional variability sources, i.e., $m < p$ and inherently $\boldsymbol{B}$ is not full rank, which is the necessary condition for being partially diagnosable on variance faults. For partially diagnosable situations, the change on $\text{cov}(\boldsymbol{Y}_i)$ can only reveal that on a combination of variance components as a whole, rather separable individual ones. To incorporate such circumstances, we extend the concept of diagnosable faults in a more generalized way as below:

Definition 3.2. Given fault pattern $\boldsymbol{\gamma}$, a linear combination $\boldsymbol{f}^T\boldsymbol{\lambda}_\gamma$, $\boldsymbol{f} \in \mathbb{R}^{\#\gamma\times 1}$ is diagnosable, if $\forall\ \boldsymbol{\lambda}_\gamma^1,\ \boldsymbol{\lambda}_\gamma^2$, such that

$$\boldsymbol{f}^T(\boldsymbol{\lambda}_\gamma^1) \neq \boldsymbol{f}^T(\boldsymbol{\lambda}_\gamma^2) \Rightarrow \text{cov}(\boldsymbol{Y}_i)|_{\boldsymbol{\lambda}_\gamma=\boldsymbol{\lambda}_\gamma^1} \neq \text{cov}(\boldsymbol{Y}_i)|_{\boldsymbol{\lambda}_\gamma=\boldsymbol{\lambda}_\gamma^1}, \tag{3.23}$$

and $\{\boldsymbol{\lambda}_\gamma\}$ is a diagnosable class of variance faults.

It can be seen that Definition 3.1 is a special case of Definition 3.2 when $\boldsymbol{f}$ contains only one nonzero element. Faults in the same diagnosable class are said to be "coupled". The coupling of variance faults possibly can be further decomposed, and we further define the minimal diagnosable class:

Definition 3.3. Diagnosable class $\{\boldsymbol{\lambda}_\gamma\}$ is a minimal diagnosable class (MDC) is no strict subsets of it are also diagnosable classes.

Therefore MDCs are the minimal fault units that can be diagnosed, and the cardinality of an MDC is referred to as the *degree*. The relation among fault

sources can be more intuitively interpreted from Equations (3.4) and (3.5): denote $\sum_{j=1}^{p} \boldsymbol{F}_j \lambda_j^2 = V(\boldsymbol{\lambda})$, a fully diagnosable model requires that $\forall \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2$ such that $\boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_2 \Rightarrow V(\boldsymbol{\lambda}_1) \neq V(\boldsymbol{\lambda}_2)$. Rao and Kleffe (1988) indicate that the model is fully diagnosable if and only if $\boldsymbol{F}_j$s are linearly independent; otherwise, for example, if $\boldsymbol{F}_1 = \boldsymbol{F}_2 + \boldsymbol{F}_3$, then $\boldsymbol{\lambda}_\gamma = \lambda_1 = 1$ or $\boldsymbol{\lambda}_\gamma = [\lambda_2, \lambda_3]^T = [1/\sqrt{2}, 1/\sqrt{2}]^T$ give equal $\text{cov}(\boldsymbol{Y}_i)$, so the two fault patterns are not discernible. Theorem 2 and Corollary 1 in (Zhou et al., 2003) illustrate howo to search MDCs by $\boldsymbol{H}^r$, the the reduced row echelon form (RREF) of $\boldsymbol{H}$: each nonzero row of $\boldsymbol{H}^r$ defines an MDC containing the components of $\boldsymbol{\lambda}$ corresponding to the nonzero element of the row; additionally, all MDCs can be enumerated by thoroughly permuting columns of $\boldsymbol{H}$.

3.4.2 Sparse Fault Pattern

Diagnosing partially diagnosable models is challenging because the diagnosis cannot distinguish true fault patterns from the equivalent ones, which are coupled with the true pattern, just conditional on dimensional measurement data from products. However, for most reliable manufacturing lines, the number of faults should be small. Based on this characteristics, we propose the concept of sparse faults. That is, $\boldsymbol{\gamma}$ is sparse if the rest k components in $\boldsymbol{\lambda\gamma}$ excluding all uniquely diagnosable components follow:

- rule 1: they are all contained by an MDC of degree larger than $2k$, or
- rule 2: they can be split to multiple subsets, each of which is separately contained by MDCs as described in rule 1, and the MDCs are mutually *not con-*

nected; two MDCs are *not connected* if they have no common components (Zhou et al., 2003).

So basically, $\boldsymbol{\gamma}$s contain more 0s are more likely to be sparse. Essentially, our Bayesian diagnosis is a Stochastic Search Variable Selection (SSVS) (Gilks et al., 1995), and the sparse property of faults can be utilized by setting proper p_0, i.e., $p_0 < 0.5$, in implementation. In Section 3.5, we can see how our diagnosis benefits from such setting if faults are truly sparse, whereas it makes mistakes if otherwise.

3.5 Simulation Study

In this section, we designed four situations among which fault-diagnosis models vary from fully to partially diagnosable of different fault coupling structures, and under each situation we tested how the effectiveness of the Bayesian diagnosis on multiple fault patterns. For all simulations, we chose fault-quality models to be underdetermined, i.e., $m < p$ for $\boldsymbol{B}$, to make the study closer to real circumstances mentioned before. $\boldsymbol{B}$ is constructed by firstly randomly generating elements $Unif(-1, 1)$ and secondly conducting column operations to get required rank (except for the example in 3.5.4). We further set $\boldsymbol{\Sigma}_0 = \mathbf{0}^{p\times p}$ for simplicity and $\sigma = 0.5$. Conditional on $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}_\gamma$, $N = 50$ dimensional measurements was generated as product data from Equation (3.1) for diagnostic usage in each example. Because diagnosis is done *offline*, so it is reasonable that historical experience or expertise can assist setting priors, e.g., Li and Chen (2016) choose the prior of variance shift through an empirical Bayesian approach, and (Steward et al., 2016) choose the maximum likelihood estimate (MLE) as

prior hyperparameters. We chose $\mu_\lambda = 5$ and $\sigma_\lambda = 3$, which makes $\pi_{\boldsymbol{\lambda}}$ fairly informative. Suppose σ^2 can be estimated from historical data, we set $c_0 = \frac{N}{2}$, $d_0 = \frac{c_0-1}{c_0}\sigma^2$, which makes the *a priori* mean of σ^2 equals its true value. Finally, we set $p_0 = 0.1$ to lean toward sparse faults.

To implement Algorithm 3.1, we chose $R_b = 5000$ MCMC samples for the burn-in purpose and $R = 15000$ subsequent samples to approximately marginalize $P(\boldsymbol{\gamma}|D)$; additionally, we set the ancillary jumping proposal parameters $(\tilde{\mu}_\lambda, \tilde{\sigma}^2_\lambda) = (0, 3)$ for high acceptance rate, which was verified in tuning experiments. For each diagnostic task, we ran the simulation 1000 times for 1000 sets of randomly generated $\boldsymbol{D}$. The *maximum a posteriori* (MAP) fault pattern is determined as the diagnostic results. The effectiveness of Algorithm 3.1 is assessed by the average $P(\boldsymbol{\gamma}|D)$ calculated from the 1000 repetitions and the proportion that the MAP fault pattern is true, which are denoted by $\hat{P}(\boldsymbol{\gamma}|\boldsymbol{D})$ and r_T.

3.5.1 Fully Diagnosable Case

We chose $m = 5$, $P = 7$, and $\boldsymbol{B}$ was constructed without column operations so that $\boldsymbol{H}$ is of full rank, and therefore all components in $\boldsymbol{\lambda}$ are uniquely diagnosable. The diagnostic effectiveness was investigated under 3 sets of faults:

- $\boldsymbol{\lambda}_{\gamma_{\text{FD1}}}$: $\lambda_2 = 2$;
- $\boldsymbol{\lambda}_{\gamma_{\text{FD2}}}$: $[\lambda_1 = 2.5, \lambda_7 = 1.8]^T$;
- $\boldsymbol{\lambda}_{\gamma_{\text{FD3}}}$: $[\lambda_1 = 2.5, \lambda_4 = 1.8, \lambda_7 = 2]^T$;
- $\boldsymbol{\lambda}_{\gamma_{\text{FD4}}}$: $[\lambda_1 = 2, \lambda_2 = 2.5, \lambda_5 = 1.8, \lambda_7 = 3.2]^T$.

Table 3.1: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{FD1}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(0,1,0,0,0,0,0)$	0.989	1
$(0,1,1,0,0,0,0)$	0.002	
$(1,1,0,0,0,0,0)$	0.002	

Table 3.2: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{FD2}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,0,0,0,0,1)$	0.978	1
$(1,0,0,0,0,1,1)$	0.006	
$(1,0,1,0,0,1,1)$	0.006	

Table 3.3: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{FD3}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,0,1,0,0,1)$	0.984	1
$(1,0,0,1,1,0,1)$	0.0046	
$(1,0,1,1,0,0,1)$	0.0045	

Table 3.4: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{FD4}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,1,0,0,0,0,1)$	0.982	1
$(1,1,1,0,0,0,1)$	0.0056	
$(1,1,0,1,0,0,1)$	0.0051	

The diagnostic results are given in Tables 3.1-3.4, where $*$ marks the true fault pattern. For conciseness, we omit the possible fault patterns diagnosed with trivial average posterior probabilities (same for other tests). The results show that the true faults are correctly diagnosed in all 1000 runs, and meanwhile the average posterior probability $\hat{P}(\boldsymbol{\gamma}|\boldsymbol{D})$ has dominating value (≥ 0.978) at the true fault pattern.

3.5.2 MDC of Degree 4

For this example, we chose $m = 4$ and $q = 7$, and then we constructed $\boldsymbol{B}$ with rank equals 2. Such $\boldsymbol{B}$ makes $\text{Rank}(\boldsymbol{H}) \leq 4$ so it is a partially diagnosable scenario.

We provide

$$\boldsymbol{H}^r = \left[\begin{array}{ccc|cccc} & & & -0.75 & -2.75 & -6.03 & 0 \\ & \boldsymbol{I}^{3\times 3} & & 7 & 15 & 37.08 & 0 \\ & & & 1.31 & 10.31 & 17.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \multicolumn{7}{c}{\mathbf{0}^{3\times 7}} \end{array}\right],$$

where $\boldsymbol{I}$ is the identity matrix. Without showing the exhaustive search process of all MDCs, it can be proved that: λ_7^2 is uniquely diagnosable, and every other 4 components in $\boldsymbol{\Lambda}$ create an MDC of degree 4. The tested shifts are:

- $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc41}}}$: $\lambda_2 = 2$;
- $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc42}}}$: $[\lambda_1 = 1.5, \lambda_7 = 2.5]^T$;
- $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc43}}}$: $[\lambda_1 = 1.5, \lambda_4 = 2, \lambda_7 = 2.5]^T$;
- $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc44}}}$: $[\lambda_1 = 1.5, \lambda_2 = 2.5, \lambda_7 = 2]^T$.

$\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc41}}}$ just contains a single non-uniquely diagnosable fault; $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc42}}}$ is a pattern including single faults either uniquely diagnosable or not; $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc43}}}$ and $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc44}}}$ both contain a single uniquely diagnosable fault and coupled faults. The diagnosis simulation results of each pattern can be summarized respectively as:

1. $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc41}}}$: Table 3.5 shows $\boldsymbol{\gamma}_{\text{mdc41}}$ is correctly diagnosed as the MAP estimate. Although λ_2 is not uniquely diagnosable, $\boldsymbol{\gamma}_{\text{mdc41}}$ is sparse, since all MDCs which contains λ_1 have degree 4. Therefore $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc41}}}$ is distinguished by the Bayesian diagnosis from other equivalent fault patterns as we expect;

2. $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$: $\boldsymbol{\gamma}_{\text{mdc42}}$ is sparse from rule 1. From Table 3.6 we can see the diagnosis of λ_7 is not affected by λ_1, and it is diagnosed in all nontrivial candidate patterns. This property is quite practical because it illustrates that uniquely diagnosable faults are guaranteed to be diagnosed regardless of the existence of other non-uniquely diagnosable faults. As for λ_1, the sparse characteristic helps the method capture it, which has been discussed already.

3. $\boldsymbol{\lambda}_{\gamma_{\text{mdc43}}}$: this fault pattern is challenging for the diagnosis because $\boldsymbol{\gamma}_{\text{mdc432}}$ is not strictly sparse. As shown in Table 3.7, the portion of $\boldsymbol{\gamma}_{\text{mdc432}}$ being correctly diagnosed is below 0.5 (0.44). However, the most likely nonuniquely diagnosable faults given by the diagnosis are all contained in the same MDC $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. Therefore the diagnostic result still provide useful information; that is, if λ_2 and λ_3 are excluded by inspection, then we know the true faults would likely be those coupled with the diagnostic results within the same MDC. Readers may doubt that λ_5 and λ_6 should also be considered if λ_2 and λ_3 are diagnosed, because they are also coupled together within the same MDC; theoretically this is justifiable, but the inspection should be arranged with lower priority since the two components are diagnosed with trivial probability. The diagnosis on λ_7 has been explained;

4. $\boldsymbol{\lambda}_{\gamma_{\text{mdc44}}}$: λ_1 and λ_2 are correctly diagnosed with probability higher than 0.5. However, here the probability of correct diagnosis is not significantly higher than that otherwise, since $\boldsymbol{\gamma}_{\text{mdc44}}$ is not a sparse pattern.

Table 3.5: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc41}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(0,1,0,0,0,0,0)$	0.996	1
$(1,1,0,0,0,0,0)$	0.001	
$(0,1,1,0,0,0,0)$	0.001	

Table 3.6: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,0,0,0,0,1)$	0.973	0.997
$(1,1,0,0,0,0,1)$	0.02	
$(1,0,1,0,0,0,1)$	0.01	

Table 3.7: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc43}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$(0,1,1,0,0,0,1)$	0.46	
$*(1,0,0,1,0,0,1)$	0.41	0.44
$(1,1,1,0,0,0,1)$	0.0032	
$(1,1,0,0,0,1,1)$	0.0029	

Table 3.8: Diagnosis under FP_{mdc44}

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,1,0,0,0,0,1)$	0.51	0.58
$(1,0,0,0,0,0,1)$	0.32	
$(1,0,0,1,0,0,1)$	0.088	
$(1,0,1,0,0,0,1)$	0.017	

3.5.3 MDC of Degree 6

Fault patterns satisfy sparsity with larger opportunity if they have smaller number of components, and that create larger chance for us to diagnose them. Reversely, the more components in fault patterns, the more difficult for the diagnosis. To manifest this, in this example we construct a fault-diagnosis model with $m = 5$ and $p = 10$, under which the largest MDC contains 6 components, such that:

$$\boldsymbol{H}^r = \left[\begin{array}{c|ccccc|c} & -0.75 & -0.25 & 0.07 & -1.74 & -5.31 & \\ \boldsymbol{I}^{3\times 3} & 7 & 5 & -0.12 & 18.53 & 46.25 & \boldsymbol{0}^{2\times 2} \\ & 1.31 & 0.31 & 0.053 & 2.79 & 9.83 & \\ \hline & & & \boldsymbol{0}^{7\times 8} & & & \boldsymbol{I}^{2\times 2} \\ & & & & & & \boldsymbol{0}^{5\times 2} \end{array}\right].$$

Given the The true faults:

- $\boldsymbol{\lambda}_{\gamma_{\text{mdc61}}}$: $[\lambda_1 = 1.5, \lambda_3 = 2.5]^T$;

Table 3.9: Diagnosis under $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc61}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,1,0,0,0,0,0,0,0)$	0.694	0.83
$(0,0,1,0,0,1,0,0,0,0)$	0.11	
$(0,1,1,0,0,1,0,0,0,0)$	0.062	
$(0,0,0,1,0,1,0,0,0,0)$	0.039	

Table 3.10: Diagnosis under $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc62}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$(0,0,0,0,0,1,0,1,0,0)$	0.27	
$(0,0,1,0,0,0,1,0,0,0)$	0.22	
$(0,0,1,1,0,0,0,0,0,0)$	0.21	
$(0,0,0,1,0,1,0,0,0,0)$	0.16	

- $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_{\text{mdc62}}}$: $[\lambda_3 = 2.5, \lambda_4 = 3.2, \lambda_5 = 1.5]^T$.

The simulation results are accordingly can be interpreted as the aforementioned diagnostic characteristics: $\boldsymbol{\gamma}_{\text{mdc62}}$ is correctly diagnosed with probability 0.83 as it is a sparse pattern, whereas $\boldsymbol{\gamma}_{\text{mdc62}}$ is totally missed in the diagnosis. Table 3.10 indicates that faults in the MDCs containing λ_j, $j = 3, \cdots, 8$, should be thoroughly inspected, so as for other similar situations (i.e., no distinct clue implying true fault patterns). Nonetheless, non-sparse patterns are often contained by MDCs of relatively small degree, so exhaustive search would be affordable.

3.5.4 MDCs of Degree 3

If the degree of an MDC is low, then it is difficult to facilitate the diagnosis by the sparsity; for example, if every two variance components forms an MDC, then even a single fault cannot be discerned from others. Fortunately, sparse fault patterns still have chance to exist for low degree MDCs. In this example, the method is applied to the situation in which the whole fault set can be split to 2 groups *not connected*, for

each of which the largest MDC have a degree of 3. Given $\boldsymbol{B}$ as a 5×10 matrix:

$$\boldsymbol{B} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & & & & & \\ \sqrt{2} & \sqrt{3} & 2 & 2 & 0 & & \mathbf{0}^{3\times 5} & & & \\ & & & & & & & & & \\ & \mathbf{0}^{3\times 5} & & & & 1 & 1 & 1 & 1 & 0 \\ & & & & & 0 & 1 & \sqrt{2} & \sqrt{3} & 2 \end{bmatrix},$$

we have:

$$\boldsymbol{H}^r = \left[\begin{array}{c|c|cccc} & 5.54 & & & & \\ \boldsymbol{I}^{4\times 4} & -4.64 & & & \mathbf{0}^{4\times 5} & \\ & 0.10 & & & & \\ & 0.61 & & & & \\ \hline & & 1 & 0 & 0 & 0.12 & 1.63 \\ \mathbf{0}^{6\times 5} & & 0 & 1 & 0 & 1.63 & -8.90 \\ & & 0 & 0 & 1 & -0.75 & 7.27 \\ \hline & & \multicolumn{5}{c}{\mathbf{0}^{2\times 5}} \end{array}\right].$$

It can be proved by permuting columns of $\boldsymbol{B}$ that components λ_1 and λ_6 are separately contained in MDCs that are not connected (e.g., $\{\lambda_1, \lambda_2, \lambda_5\}$ and $\{\lambda_6, \lambda_9, \lambda_{10}\}$ respectively), and the same for λ_2 and λ_9. From rule 2 in Section 3.4.2, we know $\{\sigma_1^2, \lambda_6^2\}$ are $\{\lambda_2^2, \lambda_9^2\}$ are both sparse patterns even their size is less than the half of the degree of the MDCs contained by. We assign the true shift sets $\boldsymbol{\lambda}_{\gamma_{\text{mdc31}}}$ and $\boldsymbol{\lambda}_{\gamma_{\text{mdc31}}}$ as:

- $\boldsymbol{\lambda}_{\gamma_{\text{mdc31}}}$: $[\lambda_1 = 2, \lambda_6 = 2]^T$;
- $\boldsymbol{\lambda}_{\gamma_{\text{mdc31}}}$: $[\lambda_2 = 2, \lambda_9 = 2]^T$.

The results are given in Tables 3.11 and 3.12. It can be seen that the true fault patterns stand out with dominating average posterior probabilities in the diagnosis. This example extends the scenario in which sparse faults still apply although MDCs are of small degree.

3.5.5 MCMC Convergence Analysis

The efficiency of MCMC methods should also be concerned in practice. In this section we provide empirical evaluation for the convergence perfomance of Algorithm 3.1. However, the convergence assessment is a contentious issue even for fixed-dimensional cases, and the transdimensional nature of Algorithm 3.1 adds additional difficulties, i.e., how we judge convergence both within each model and across models with respect to posterior model probabilities. The Markov chain constructed by Algorithm 3.1 is ergodic, that is, $P(\boldsymbol{\gamma}, \boldsymbol{\lambda}|\boldsymbol{D})$-irreducible and aperiodic. This can be provided by Lemma 10.11 in Robert and Casella (2005), and the detailed proof is omitted here since it is not our primal concern. The convergence is guaranteed; however, how fast and well the chain converges requires further assessment. Before we give more quantitative opinions of the convergence, we visualize the transition process of the Markov chains in diagnosing $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$, to intuitively prove the method

Table 3.11: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc31}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,0,0,0,1,0,0,0,0)$	0.971	0.99
$(1,1,0,0,0,1,0,0,0,0)$	0.009	
$(1,0,1,0,0,1,0,0,0,0)$	0.006	
$(1,0,0,0,0,1,0,0,0,1)$	0.004	

Table 3.12: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc32}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(0,1,0,0,0,0,0,0,1,0)$	0.80	0.89
$(1,0,1,0,0,0,0,0,1,0)$	0.17	
$(0,0,1,0,0,0,0,0,1,0)$	0.017	
$(0,1,1,0,0,0,0,0,1,0)$	0.008	

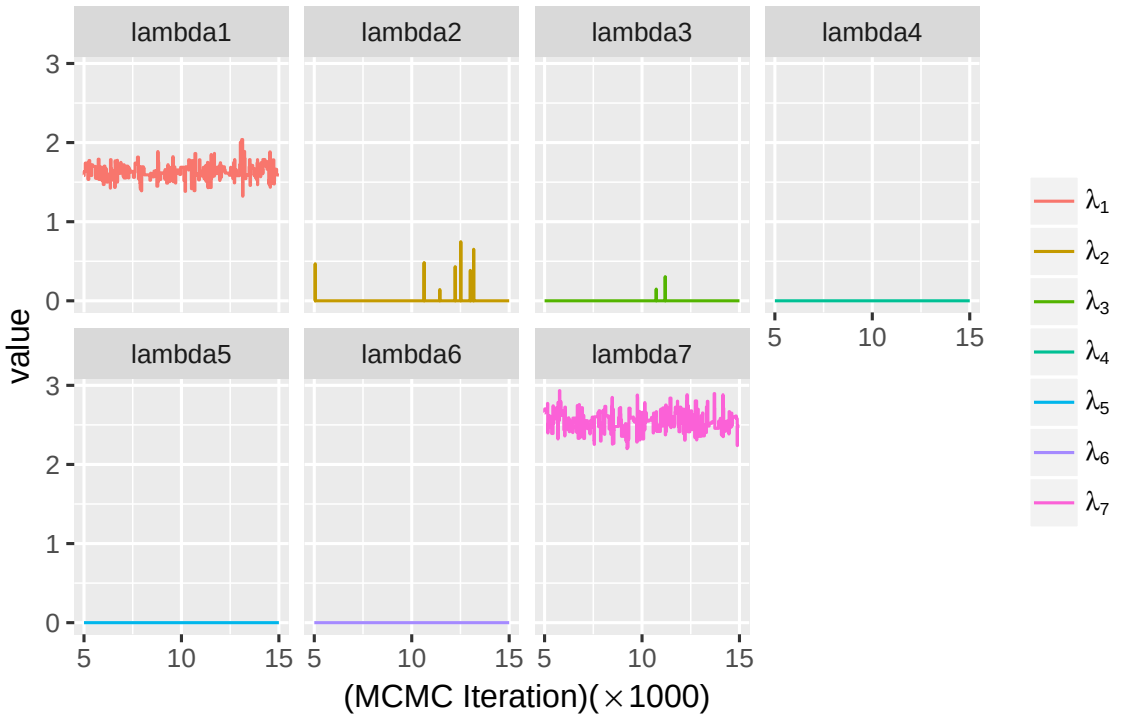


Figure 3.1: Convergence of MCMC Sampling of λ_js

efficiently converges. After the burn-in period of 5000 samples, from Figure 3.1 we see that the Markov chains of the samples of λ_1 and λ_2 have fast mixing, and the only samples near the true shift value are accepted. On the contrary, for other fault components, the samples series indicate they are discarded by the sampler approximately with probability 1. The diagnosis from the MCMC completes within 15000 iterations, and the total sampling time on a Intel Core i7 3.40GHz platform is less than 15 second. Therefore, Algorithm 3.1 provides acceptable diagnosis cost, and it could also give more information other than fault patterns, such as shift magnitudes, if the true fault pattern is sparse. For more rigorous convergence assessments for Algorithm 3.1, since we concern model components $\boldsymbol{\gamma}$ over estimating $\boldsymbol{\lambda}_j$s, and

therefore we here choose the assessment method suggested in Brooks et al. (2003a). The method proposed various ideas based on the sample path of the model indicator under the assumption that replicated chains that have converged would generate similar posterior model probabilities; two hypothesis tests, namely Chi-squared and Kolmogorov-Smirnov tests are used to testify the assumption. Here we generated 6 replicated chains (i.e., run Algorithm 3.1 by 6 times), and still we checked the convergence performance after 5000 burn-in iterations. Since the Kolmogorov-Smirnov test requires pair-wise comparisons, and therefore for the labor-saving consideration, we chose the Anderson–Darling test to compare the empirical posterior model distributions from each chain simultaneously. The p-values of the two tests with iteration length is given in Figure 3.2, from which we can see that the p-values is quite close to 1 and that indicates no evidence to reject that Algorithm 3.1 has converged after the burn-in stage.

3.5.6 Prior Sensitivity Analysis

To evaluate the robustness of the method, here we also implement the sensitivity analysis on the priors. π_γ and π_λ are affect the diagnosis most but cannot be easily determined. Therefore in this section we chose $\sigma_\lambda = 10$ and $p_0 = 0.4$ to make the priors more noninformative, which adds more challenges to the diagnosis. We selected $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$ and $\boldsymbol{\lambda}_{\gamma_{\text{mdc43}}}$ as examples, so the diagnosability would not affect the result as nuisance factor. The diagnosis results are given in Table 3.13 and 3.14. We can tell that with the more weakly informative priors, for $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$, the posterior

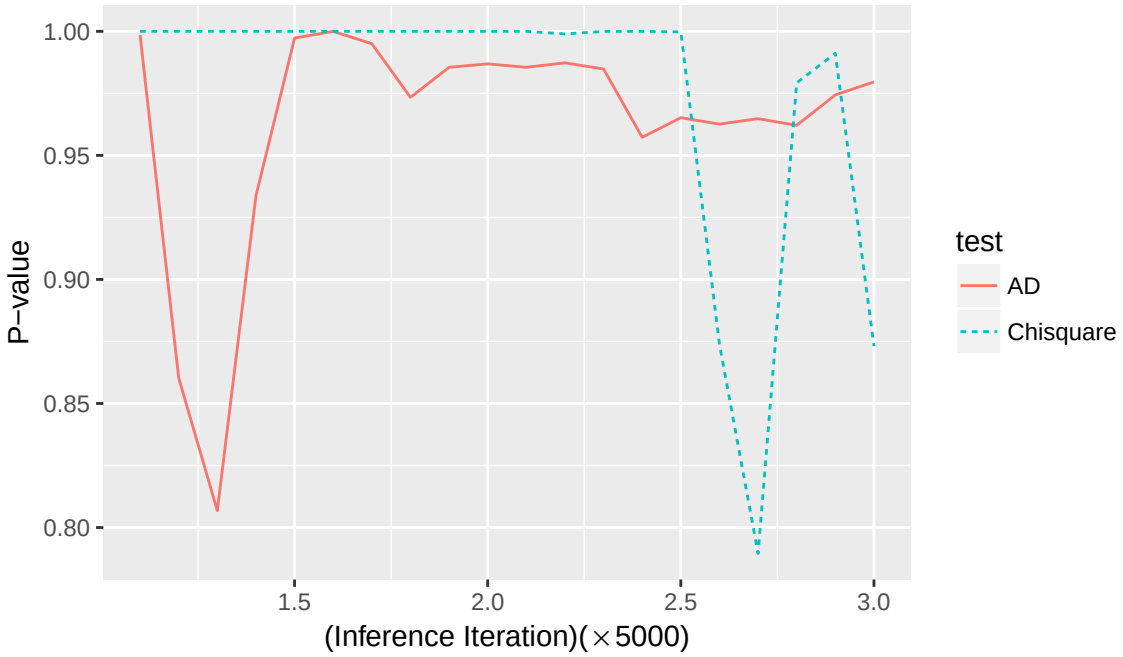


Figure 3.2: Convergence Assessment by Tests

Table 3.13: Prior Sensitivity Analysis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc41}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(1,0,0,0,0,0,1)$	0.844	0.988
$(1,1,0,0,0,0,1)$	0.084	
$(1,0,1,0,0,0,1)$	0.0314	

Table 3.14: Prior Sensitivity Analysis under $\boldsymbol{\lambda}_{\gamma_{\text{mdc42}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$(0,1,1,0,0,0,1)$	0.40	
$*(1,0,0,1,0,0,1)$	0.23	0.28
$(1,1,1,0,0,0,1)$	0.16	

probability of the true fault pattern is less spiky, but nonetheless the diagnostic result is not affected, whereas for $\boldsymbol{\lambda}_{\gamma_{\text{mdc43}}}$, such priors yield no significantly different result than before, and the routine of combining knowledge on the identifiablity and the Bayesian diagnosis still could lead to correct answer here. Therefore, the method does not significantly rely on the priors, and such feature is favorable in practice.

3.6 Case Study

We consider the assembly process example in Zhou et al. (2003) as a practical problem to evaluate the performance of the Bayesian diagnosis. The 2-dimensional panel assembly example is simplified from an autobody assembly process. As shown in Figure 3.3, the process consists of 3 stages, and in each stage, active locators are potentially subject to wearing, abrading, potential improper installation, etc, and faults on 2-way and 4-way locators would cause fixtures deviate on X dimension and X-Z dimensions respectively. To assess if excessive dimensional variation exist, 5 coordinate sensors are placed at positions M_i, $i = 1, \cdots, 5$, and each can provide measurements regarding the accuracy of fixture locators on both X and Z dimensions. Within each stage, the number of the potential fixture deviation components is 6, 9, 3 respectively, and then 18 in total. Therefore, the assembly process can be described by the model in Equation (3.1) with $p = 18$ and $m = 10$.

To assess how well the methods perform in practical circumstances, we adopted the idea used in the simulation study. That is, we generated observations randomly as true manufacturing measurements, and then implemented the method for diagnosis aim. Same priors used in simulations were applied here. We modified some design parameters from the original work and reconstruct $\boldsymbol{B}^{10\times 18}$ as shown in Appendix D. The engineering meaning of $\boldsymbol{B}$ is explained in the original work, if readers have

interest. To investigate the diagnosability of faults, $\boldsymbol{H}^r$ is calculated:

$$\boldsymbol{H}^r = \left[\begin{array}{c|cccccc} \boldsymbol{I}^{12\times12} & \multicolumn{6}{c}{\boldsymbol{0}^{12\times6}} \\ \hline & 1 & 0 & 0 & 1.21 & 0.21 & 0.70 \\ & 0 & 1 & 0 & 5.76 & -4.39 & -13.15 \\ \boldsymbol{0}^{6\times12} & 0 & 0 & 1 & -0.074 & 0.91 & 1.23 \\ & \multicolumn{6}{c}{\boldsymbol{0}^{3\times6}} \end{array}\right],$$

from which we know $\lambda_{1\cdots,12}$ are uniquely diagnosable, and the degree of the largest MDCs is 4, e.g., $\{\lambda_{13}, \lambda_{16}, \lambda_{17}, \lambda_{18}\}$. From the previous results we know that the diagnosis of uniquely diagnosable faults is reliable, so here we just concern the diagnosis of components coupled with others. The true shift sets tested are:

- $\boldsymbol{\lambda}_{\gamma_{\text{case1}}}$ $\boldsymbol{\lambda}_{\gamma_{\text{case1}}}$: $\lambda_{13} = 1.5$;
- $\boldsymbol{\lambda}_{\gamma_{\text{case2}}}$: $[\lambda_{13} = 1.5, \lambda_{14} = 2.5]^T$.

The performance of the Bayesian diagnosis method is assessed in the same way as before and given in Table 3.15 and 3.16. From Table 3.15 we can see the Bayesian estimation of sparse fault pattern is quite accurate as we expect, and the knowledge of MDCs help us narrow searching scope to $\lambda_{14,\cdots,18}$ with respect to the clue given in in Table 3.16, as discussed in Section 3.5. Notably, the diagnostic tasks in this case study are beyond the capability of other diagnostic methods mentioned in Section 1.2. In terms of this, our method is the only choice for such multi-stage manufacturing process.

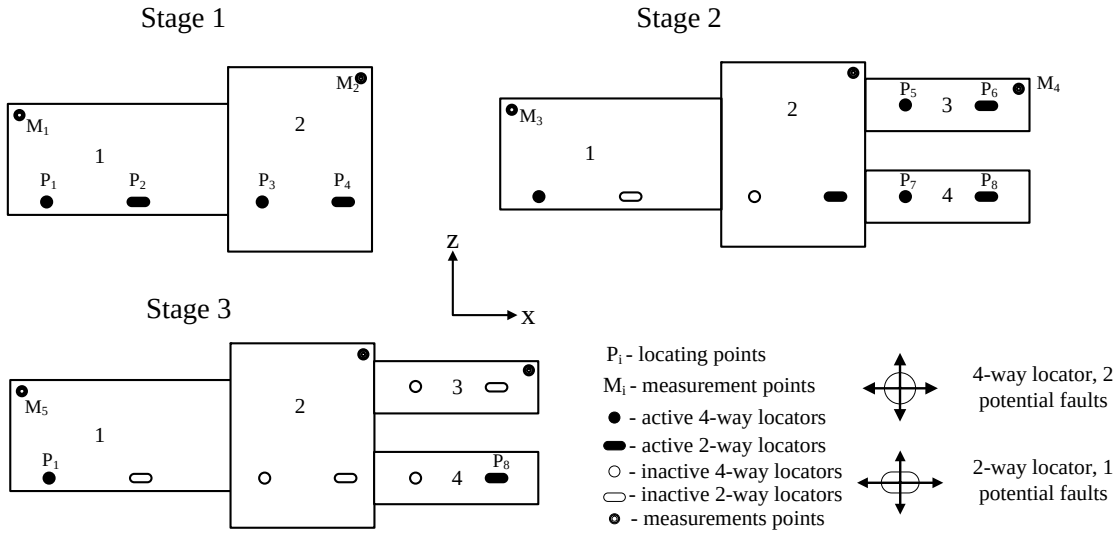


Figure 3.3: 2-Dimensional Panel Assembly

Table 3.15: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{case1}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$*(\mathbf{0}^{1\times12}, 1, 0, 0, 0, 0, 0)$	0.89	0.93
$(\mathbf{0}^{1\times12}, 0, 0, 0, 1, 0, 0)$	0.070	
$(\mathbf{0}^{1\times12}, 1, 1, 0, 1, 0, 0)$	0.0015	
$(\mathbf{0}^{1\times12}, 1, 0, 0, 1, 0, 0)$	0.006	

Table 3.16: Diagnosis under $\boldsymbol{\lambda}_{\gamma_{\text{case2}}}$

$\boldsymbol{\gamma}$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_T
$(\mathbf{0}^{1\times12}, 0, 0, 0, 1, 0, 0)$	0.48	
$*(\mathbf{0}^{1\times12}, 1, 1, 0, 0, 0, 0)$	0.24	0.24
$(\mathbf{0}^{1\times12}, 0, 0, 0, 1, 0, 1)$	0.019	
$(\mathbf{0}^{1\times12}, 0, 0, 0, 1, 1, 0)$	0.016	

3.7 Conclusion

We have developed a new Bayesian variable selection based diagnostic method for the fault-quality model defined in Section 3.1. The implementation of the method is affected by the diagnosability of fault sources; however, the concept of MDCs helps improve the opportunity to identify true fault patterns, if the assumption on the sparsity holds.

The RJ-MCMC methodology is adopted in our study to estimate the posterior

probability of fault patterns containing varying numbers of components. Guidelines are provided for the selection of hyperparameters in the Bayesian hierarchical model. Numerical examples are used to demonstrate the effectiveness of the diagnosis approach.

CHAPTER 4
A BAYESIAN METHOD FOR CHANGEPOINT DETECTION AND DIAGNOSIS OF MULTIVARIATE LINEAR PROFILES

The SPC of a process characterized by a linear profile model is studied intensively in recent decades. However, such research is not well extended to multivariate linear profile diagnosis. Being motivated by the insufficiency, we propose a novel Bayesian diagnostic framework to serve both Phase I and Phase II diagnostic usage. The Bayesian framework integrates the diagnosis of linear profiles with respect to the changepoint and change model; that is, it solves two concerns: firstly, whether and when the process is shifted, and secondly, in which pattern the shift occurs. Essentially the framework is a variable selection approach, and the diagnostic decision is made based on the posterior probabilities. A Markov Chain Monte Carlo algorithm is suggested to complete the Bayesian analysis. We represent simulation tests and a case study to demonstrate the effectiveness of the Bayesian diagnosis.

The proposed approach can be applied to i) inspect inhomogeneous items in raw data and extract in control information from it (Phase I) or to ii) retrospectively diagnostic analysis following online monitoring signals (Phase II). Multiple numerical examples demonstrate the diagnostic effectiveness under various situations. The rest part of this chapter is organized as follows. Section 4.1 introduces the changepoint model for multivariate quality data. In Section 4.2 we develop in detail the Bayesian diagnostic procedures. Numerical experiments and a calibration case study are provided in Section 4.3 and Section 4.4 respectively. Finally Section 4.5 summarizes the

chapter and our contributions.

4.1 Changepoint Model for Multivariate Linear Profile

Consider N multivariate observations $\boldsymbol{Y}_i = [Y_{i1}, \cdots, Y_{ij}, \cdots, Y_{im}]^T$, $i = 1, \cdots, N$, and a fixed explanatory variables $\boldsymbol{B}_i$ are associated by a linear model

$$\boldsymbol{Y}_i = \boldsymbol{B}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \tag{4.1}$$

where $\boldsymbol{\beta}_i^{p\times 1} = [\beta_i^{(1)}, \beta_i^{(2)}, ..., \beta_i^{(p)}]^T$ is the p dimension profile parameter of concern, and $\boldsymbol{\epsilon}_i^{m\times 1}$s are measurement error or process variation independently generated from $Norm(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. We assume $\boldsymbol{B}$ is of full rank, and without loss of generality, $\boldsymbol{B}$ is supposed to be centralized over columns and has no intercept column. Denote $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{IC}$ if the profile is in statistical control or $\boldsymbol{\beta}_{OC}$ otherwise. A changepoint model is considered to illustrate the appearance of the shift at τ; that is:

$$\boldsymbol{Y}_i|\boldsymbol{\beta}_i, \sigma^2 \sim Norm(\boldsymbol{B}\boldsymbol{\beta}_i, \sigma^2\boldsymbol{I}), \tag{4.2}$$

where $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{IC}$ for $i \leq \tau$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{OC}$ otherwise. In many applications, however, a more reasonable assumption is that $\boldsymbol{\beta}_{IC}$ is partially shifted. That is, suppose there exists a p dimension indicator vector $\boldsymbol{\gamma} = [\gamma^{(1)}, ..., \gamma^{(p)}]^T$ consisting of 1s and 0s, such that:

$$\begin{aligned} \beta_{OC}^{(j)} &= \beta_{IC}^{(j)} \quad \text{if } \gamma^{(j)} = 0, \\ \beta_{OC}^{(j)} &\neq \beta_{IC}^{(j)} \quad \text{if } \gamma^{(j)} = 1, \end{aligned} \tag{4.3}$$

for $j = 1, \cdots, p$. So $\boldsymbol{\gamma}$ represents the fault pattern.

4.2 Bayesian Diagnostic Approach for τ and $\boldsymbol{\gamma}$

In this section, we provide the technical details to implement the Bayesian diagnosis of τ and $\boldsymbol{\gamma}$ in both Phase I and Phase II diagnostic applications.

4.2.1 Prior Setting

A hierarchical model is constructed to incorporate the prior information of the variables, e.g., σ^2, $\boldsymbol{\beta}_{IC,OC}$, τ and $\boldsymbol{\gamma}$. Firstly, $\boldsymbol{Y}_i$ follows

$$\boldsymbol{Y}_i|\boldsymbol{\beta}_{IC},\boldsymbol{\beta}_{OC},\sigma^2 \sim \begin{cases} Norm(\boldsymbol{B}\boldsymbol{\beta}_{IC},\sigma^2\boldsymbol{I}) & \text{if } i \leq \tau, \\ Norm(\boldsymbol{B}\boldsymbol{\beta}_{OC},\sigma^2\boldsymbol{I}) & \text{if } i > \tau. \end{cases} \tag{4.4}$$

The prior for σ^2 is chosen to be a inverse gamma distribution

$$\pi_{\sigma^2}(\sigma^2) = IG(\frac{\nu_1}{2},\frac{\nu_2}{2}), \tag{4.5}$$

where the hyperparameters ν_1 and ν_2 could be chosen to make $\pi_{\sigma^2}(\sigma^2)$ reasonably informative, since the knowledge of measurement noise is often available from historical experience of the monitoring system, and $\pi_{\sigma^2}(\sigma^2)$ is chosen as a conjugate prior of the likelihoods in Equation (4.4) for computational convenience. The access to $\boldsymbol{\beta}_{IC}$ depends on the applications; for example, when implementing a Phase I analysis, the exact value of $\boldsymbol{\beta}_{IC}$ should not be available, and the same for a Phase II study following self-starting multivariate profile monitoring (see the ongoing work introduced in Zou et al. (2007b)). Under such situations, we assign Zellner's g prior for $\boldsymbol{\beta}_{IC}$:

$$\boldsymbol{\beta}_{IC}|g \sim Norm(\boldsymbol{0}, g\sigma^2(\boldsymbol{B}^T\boldsymbol{B})^{-1}). \tag{4.6}$$

For $\boldsymbol{\beta}_{OC}$, we previously have already explained that the shift only changes the profile parameter partially. So we can decompose $\boldsymbol{\beta}_{IC}$ and $\boldsymbol{\beta}_{OC}$ into two partial vec-

tors respectively as $\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}$, $\boldsymbol{\beta}_{OC}^{\bar{\boldsymbol{\gamma}}}$ and $\boldsymbol{\beta}_{IC}^{\boldsymbol{\gamma}}$, $\boldsymbol{\beta}_{IC}^{\bar{\boldsymbol{\gamma}}}$, each of which contains the profile components accordingly; for example, if $p = 4$ and $\boldsymbol{\gamma} = [1, 1, 0, 0]^T$ is given, then $\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}} = [\beta_{OC}^{(1)}, \beta_{OC}^{(2)}]^T$, $\boldsymbol{\beta}_{OC}^{\bar{\boldsymbol{\gamma}}} = [\beta_{OC}^{(3)}, \beta_{OC}^{(4)}]^T = [\beta_{IC}^{(3)}, \beta_{IC}^{(4)}]^T$, and the decomposition rule labeled by $\boldsymbol{\gamma}$ also applies to other vectors and matrices (see $\boldsymbol{B}_{\boldsymbol{\gamma}}$ later as an example) in this chapter. The prior of $\boldsymbol{\beta}_{OC}$ given $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{IC}$ is then a combination:

$$\begin{aligned} \boldsymbol{\beta}_{OC}^{\bar{\boldsymbol{\gamma}}} &= \boldsymbol{\beta}_{IC}^{\bar{\boldsymbol{\gamma}}}, \\ \boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}} | g &\sim Norm(\mathbf{0}, g\sigma^2(\boldsymbol{B}_{\boldsymbol{\gamma}}^T \boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}), \end{aligned} \tag{4.7}$$

where $\boldsymbol{B}_{\boldsymbol{\gamma}}$ is a partial matrix with columns picked from $\boldsymbol{B}$ to match $\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}$. Zellner's g priors are adopted because their simplicity and computational tractability in calculating the posterior distribution of $(\tau, \boldsymbol{\gamma})$ (Zellner, 1986). The hyperparameter g in the priors of $\boldsymbol{\beta}_{IC}$ and $\boldsymbol{\beta}_{OC}$ should be appropriately determined. Multiple criteria are offered to select proper g, e.g., $g = N$ for the unit information prior, $g = p^2$ from a minimax perspective, $g = \min(N, p^2)$ for the benchmark prior, or calculating g from the local empirical Bayes approach. Nonetheless, the famous *Bartlett's Paradox* and *Information Paradox* causes less optimum inference by using a g fixed in value (Liang et al., 2008). To avoid the shortcomings of the general g-priors, we propose to use the so called "hyper g-prior", i.e., a mixture of Zellner's g-Priors by assuming g has the prior

$$\pi_g(g) = \frac{a-2}{2}(1+g)^{-a/2}, \quad g > 0, \tag{4.8}$$

which can be also expressed as $\frac{g}{1+g} \sim Beta(1, \frac{a}{2} - 1)$. $a > 2$ guarantees that $\pi_g(g)$ is proper, and such setting is required in later work. Equations (4.4) to (4.11) will be used for the fully Bayesian inference framework in the following sections. Changepoint

detection or diagnosis algorithms are often prone to overfitting (Haynes et al., 2017); that is, adding changepoints which don't truly exist. In this work, such diagnostic mistake is referred to the false shift discovery, or type I error. Therefore, the prior $\pi_\tau(\tau)$ has perceived importance to control the type I error rate. We construct two different $\pi_\tau(\tau)$s for Phase I and Phase II approaches respectively. That is,

- for Phase I

$$\pi_\tau(\tau) = \begin{cases} p_0(1-p_0)^{\tau-1} & 1 \leq \tau \leq N-1 \\ (1-p_0)^{N-1} & \tau = N \end{cases} \tag{4.9}$$

- and for Phase II

$$\pi_\tau(\tau) \propto e^{\lambda I(\tau=N)} \quad 0 \leq \tau \leq N, \tag{4.10}$$

where p_0 and λ are hyperparameters. $\pi_\tau(\tau)$ is constructed with different forms and supports for Phase I and Phase II diagnosis on account of both interpretation purpose and distinctions between the two applications. In Phase I analysis, the process starts from in control state, and we assume the in control run length follows a Geometric distribution, i.e., suppose the shift happens with probability p_0 at each time given the process is in control by the time, and $\tau = N$ indicates the process has no shift; by doing the , the $\pi_\tau(\tau)$ is eligible to incorporate the knowledge of average in control run length (AIRL). For Phase II usage, $\tau = 0$ means the shift happens before the very beginning observation, whereas $\tau = N$ means no shift happens and the diagnosis could be introduced after false alarms generated in monitoring. So λ should be determined by the type I error probability or false alarm rate of the precedent monitoring. For

simplicity we assume that each component in $\boldsymbol{\beta}_{IC}$ undergoes shifts independently and identically with probability ω_0. Therefore the prior of $\boldsymbol{\gamma}$ can be expressed as

$$\pi_{\gamma}(\boldsymbol{\gamma}) = \prod_{j=1}^{p} (1-\omega_0)^{1-\gamma^{(j)}} \omega_0^{\gamma^{(j)}} \tag{4.11}$$

and $\pi_{\gamma}(\boldsymbol{\gamma})$ becomes a uniform prior when $\omega_0 = 0.5$.

4.2.2 Posterior Distribution for τ and $\boldsymbol{\gamma}$

Since $\tau = N$ or $\boldsymbol{\gamma} = \mathbf{0}$ should indicate there is no shift happens to the profile, so to keep the independent assumption between τ and $\boldsymbol{\gamma}$ consistently valid, we define that $\beta_{OC} = \beta_{IC}$ given $\tau = N$ and $\boldsymbol{\gamma} \neq \mathbf{0}$, or $\tau \neq N$ and $\boldsymbol{\gamma} = \mathbf{0}$. We derive the posterior distributions of $(\tau, \boldsymbol{\gamma})$ depending on whether $\boldsymbol{\beta}_{IC}$ is known.

$\boldsymbol{\beta}_{IC}$ is unknown (Phase I)

Denote $\boldsymbol{D} = [\boldsymbol{Y}_1^T, ...\boldsymbol{Y}_N^T]^T$, we have

$$P(\boldsymbol{\gamma}, \tau | g, \boldsymbol{D}) \propto L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g)\pi_{\tau}(\tau)\pi_{\gamma}(\boldsymbol{\gamma}), \tag{4.12}$$

and the uncertainty of $\boldsymbol{\beta}_{IC}$, $\boldsymbol{\beta}_{OC}$ has to be considered in calculating $L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g)$ by

$$L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g) = \iiint L(\boldsymbol{D}|\boldsymbol{\beta}_{IC}, \boldsymbol{\beta}_{OC}, \sigma^2, \tau, \boldsymbol{\gamma}, g)\pi_{\sigma^2}(\sigma^2)\pi_{\beta_{OC}^{\gamma}}(\boldsymbol{\beta}_{OC}^{\gamma}|g)\, d\sigma^2\, d\boldsymbol{\beta}_{IC}\, d\boldsymbol{\beta}_{OC}^{\gamma}. \tag{4.13}$$

Fortunately the integral in Equation (4.13) has closed form derivations based on the value of τ; that is:

- if no shift happens, i.e., $\tau = N$, or $\boldsymbol{\gamma} = \mathbf{0}$, the likelihood is calculated as:

$$\begin{aligned} L(\boldsymbol{D}|g) &= \iint L(\boldsymbol{D}|\sigma^2, \boldsymbol{\beta}_{IC})\pi_{\sigma^2}(\sigma^2)\, d\sigma^2\, d\boldsymbol{\beta}_{IC} \\ &\propto (1+gN)^{-\frac{p}{2}} (\frac{\nu_1'}{2})^{-\frac{\nu_2'}{2}}, \end{aligned} \tag{4.14}$$

and

$$\begin{aligned}\frac{\nu_1'}{2} &= \frac{\nu_1}{2} + \frac{mN}{2} \\ \frac{\nu_2'}{2} &= \frac{\nu_2}{2} + (\sum_{i=1}^{N} \boldsymbol{Y}_i^T \boldsymbol{Y}_i - \frac{g}{1+gN} \sum_{i=1}^{N} \boldsymbol{Y}_i^T \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{Y}_i)/2.\end{aligned} \tag{4.15}$$

and

- Otherwise, $\boldsymbol{Y}_{1:\tau}$ and $\boldsymbol{Y}_{\tau+1:N}$ are not mutually independent, and $L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g)$ should be decomposed by:

$$L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g, \sigma^2) = L(\boldsymbol{Y}_{1:\tau}|g, \sigma^2) L(\boldsymbol{Y}_{\tau+1:N}|\boldsymbol{\gamma}, \tau, \boldsymbol{Y}_{1:\tau}, g, \sigma^2), \tag{4.16}$$

and then we have

$$L(\boldsymbol{D}|g, \boldsymbol{\gamma}, \tau) \propto \frac{\left|\boldsymbol{\Sigma}_{\tau^-}^{\bar{\gamma}}\right|^{-\frac{1}{2}} \left|(\boldsymbol{B}_{\boldsymbol{\gamma}}^T \boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}\right|^{-\frac{1}{2}} g^{\frac{1}{2}\sum_{j=1}^{p} I(\gamma^{(j)}=0)} (1+g\tau)^{-\frac{p}{2}}}{\left|\boldsymbol{\Sigma}_{\tau^+}\right|^{-\frac{1}{2}}} (\frac{\nu_1'}{2})^{-\frac{\nu_2'}{2}} \tag{4.17}$$

where $|\cdot|$ is the determinant of a matrix; for $\sigma^2\boldsymbol{\Sigma}_{\tau^-}^{\gamma}$ and $\sigma^2\boldsymbol{\Sigma}_{\tau^+}$, it can be proved that

$$\boldsymbol{\beta}_{IC}^{\bar{\gamma}}|g, \boldsymbol{\gamma}, \boldsymbol{Y}_{1:\tau} \sim Norm(\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}}, \sigma^2\boldsymbol{\Sigma}_{\tau^-}^{\bar{\gamma}}),$$

$$\tilde{\boldsymbol{\beta}}_{OC}|g, \boldsymbol{\gamma}, \boldsymbol{Y}_{\tau+1:N} \sim Norm(\boldsymbol{\mu}_{\tau^+}, \sigma^2\boldsymbol{\Sigma}_{\tau^+}),$$

where $\tilde{\boldsymbol{\beta}}_{OC} = [\boldsymbol{\beta}_{OC}^{\bar{\gamma}}, \boldsymbol{\beta}_{OC}^{\gamma}]^T$, and

$$\begin{aligned}\boldsymbol{\Sigma}_{\tau^-} &= \frac{g}{1+g\tau}(\boldsymbol{B}^T\boldsymbol{B})^{-1} \\ \boldsymbol{\mu}_{\tau^-} &= \frac{g}{1+g\tau}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T \sum_{i=1}^{\tau} \boldsymbol{Y}_i \\ \boldsymbol{\Sigma}_{\tau^+} &= \begin{pmatrix} (\boldsymbol{\Sigma}_{\tau^-}^{\bar{\gamma}})^{-1} + (N-\tau)(\boldsymbol{B}_{\bar{\gamma}}^T\boldsymbol{B}_{\bar{\gamma}}) & (N-\tau)\boldsymbol{B}_{\bar{\gamma}}^T\boldsymbol{B}_{\gamma} \\ (N-\tau)\boldsymbol{B}_{\gamma}^T\boldsymbol{B}_{\bar{\gamma}} & \frac{g(N-\tau)+1}{g}(\boldsymbol{B}_{\gamma}^T\boldsymbol{B}_{\gamma}) \end{pmatrix}^{-1} \\ \boldsymbol{\mu}_{\tau^+} &= \boldsymbol{\Sigma}_{\tau^+}(\tilde{\boldsymbol{B}} \sum_{i=\tau+1}^{N} \boldsymbol{Y}_i + (\boldsymbol{\Sigma}_{\tau^-}^{\bar{\gamma}})^{-1}\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}})\end{aligned} \tag{4.18}$$

where the subscript τ^- or τ^+ represents the posterior distribution is conditional on the data up to τ or N respectively, and $\tilde{\boldsymbol{B}}$ is formed by permuting the columns of $\boldsymbol{B}$ corresponding to $\tilde{\boldsymbol{\beta}}_{OC}$. The $\frac{\nu_1'}{2}$ and $\frac{\nu_2'}{2}$ in Equation (4.17) are then defined as:

$$\begin{aligned}\frac{\nu_1'}{2} &= \frac{\nu_1}{2} + \frac{mN}{2} \\ \frac{\nu_2'}{2} &= \frac{\nu_2}{2} + (\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}})^T(\boldsymbol{\Sigma}_{\tau^-}^{\bar{\gamma}})^{-1}\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}} - \sum_{i=1}^{N}\boldsymbol{Y}_i^T\boldsymbol{Y}_i \\ &\quad + \frac{g}{1+g\tau}\sum_{i=1}^{\tau}\boldsymbol{Y}_i^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{Y}_i - \boldsymbol{\mu}_{\tau^+}^T\boldsymbol{\Sigma}_{\tau^+}^{-1}\boldsymbol{\mu}_{\tau^+}.\end{aligned} \tag{4.19}$$

Related proof is provided in Appendix E.

$\boldsymbol{\beta}_{IC}$ is known (Phase II)

The knowledge of $\boldsymbol{\beta}_{IC}$ could facilitate the diagnosis, because it helps calculate the likelihoods more precisely. The joint posterior probability of the concerned variables can be expressed as:

$$P(\boldsymbol{\gamma}, \tau | g, \boldsymbol{\beta}_{IC}, \boldsymbol{D}) \propto L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, \boldsymbol{\beta}_{IC}, g)\pi_\tau(\tau)\pi_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}). \tag{4.20}$$

The likelihood $L(\cdot)$ is calculated by the integral:

$$L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau, g, \boldsymbol{\beta}_{IC}) = \iint L(\boldsymbol{D}|\boldsymbol{\beta}_{IC}, \boldsymbol{\beta}_{OC}, \sigma^2, \tau, \boldsymbol{\gamma}, g)\pi_{\sigma^2}(\sigma^2)\pi_{\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}}(\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}|g)\, d\sigma^2\, d\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}. \tag{4.21}$$

As in the discussion of $\pi_\tau(\tau)$, we should be aware of that the process could have at most three states. That is:

- If shift happens before the very beginning observation, i.e., $\tau = 0$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{OC}$

for $i = 1, \cdots, N$. Therefore

$$\begin{aligned} L(\boldsymbol{D}|\boldsymbol{\gamma}, \tau = 0, g, \boldsymbol{\beta}_{IC}) &= \iint L(\boldsymbol{D}|, \boldsymbol{\beta}_{OC}, \sigma^2, \tau = 0, \boldsymbol{\gamma}, g) \pi_{\sigma^2}(\sigma^2) \pi_{\beta^{\boldsymbol{\gamma}}_{OC}}(\boldsymbol{\beta}^{\boldsymbol{\gamma}}_{OC}|g)\, d\sigma^2\, d\boldsymbol{\beta}^{\boldsymbol{\gamma}}_{OC} \\ &\propto (1+gN)^{-\frac{\sum_{j=1}^{p} I(\gamma^{(j)}=1)}{2}} (\frac{\nu_1'}{2})^{-\frac{\nu_2'}{2}}, \end{aligned} \tag{4.22}$$

where

$$\begin{aligned} \frac{\nu_1'}{2} &= \frac{\nu_1}{2} + \frac{mN}{2} \\ \frac{\nu_2'}{2} &= \frac{\nu_2}{2} + \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}} \boldsymbol{\beta}^{\tilde{\boldsymbol{\gamma}}}_{IC})^T (\boldsymbol{I} + (\frac{g}{1+gN})(\boldsymbol{B}^T_{\boldsymbol{\gamma}} \boldsymbol{B}_{\boldsymbol{\gamma}})^{-1})(\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}} \boldsymbol{\beta}^{\tilde{\boldsymbol{\gamma}}}_{IC}). \end{aligned} \tag{4.23}$$

The above results come from the conjugate property of $\pi_{\sigma^2}(\sigma^2)$, and

- if no shift happens, the likelihood is only conditional on the value of $\boldsymbol{\beta}_{IC}$, i.e.,

$$\begin{aligned} L(\boldsymbol{D}|\boldsymbol{\beta}_{IC}) &= \int L(\boldsymbol{D}|\sigma^2, \boldsymbol{\beta}_{IC}) \pi_{\sigma^2}(\sigma^2)\, d\sigma^2 \\ &\propto (\frac{\nu_1'}{2})^{-\frac{\nu_2'}{2}}, \end{aligned} \tag{4.24}$$

where

$$\begin{aligned} \frac{\nu_1'}{2} &= \frac{\nu_1}{2} + \frac{mN}{2} \\ \frac{\nu_2'}{2} &= \frac{\nu_2}{2} + \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{\beta}_{IC})^T (\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{\beta}_{IC}), \end{aligned} \tag{4.25}$$

and

- if $0 < \tau < N$, the likelihood is recalculated as

$$\begin{aligned} L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}, g) &= \int L(\boldsymbol{D}|\sigma^2, \tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}, \boldsymbol{\beta}_{OC}) \pi_{\beta^{\boldsymbol{\gamma}}_{OC}}(\boldsymbol{\beta}^{\boldsymbol{\gamma}}_{OC}|g) \pi_{\sigma^2}(\sigma^2)\, d\boldsymbol{\beta}^{\boldsymbol{\gamma}}_{OC}\, d\sigma^2 \\ &\propto (1+g(N-\tau))^{-\frac{\sum_{j=1}^{p} I(\gamma^{(j)}=1)}{2}} (\frac{\nu_1'}{2})^{-\frac{\nu_2'}{2}}, \end{aligned} \tag{4.26}$$

and

$$\begin{aligned}
\frac{\nu_1'}{2} &= \frac{\nu_1}{2} + \frac{mN}{2} \\
\frac{\nu_2'}{2} &= \frac{\nu_2}{2} + (\sum_{i=1}^{\tau}(\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{\beta}_{IC})^T(\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{\beta}_{IC}) \\
&\quad + \sum_{i=\tau+1}^{N}(\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{\beta}_{IC}^{\tilde{\boldsymbol{\gamma}}})^T(\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{\beta}_{IC}^{\tilde{\boldsymbol{\gamma}}}) \\
&\quad - \frac{g}{1+g(N-\tau)}\sum_{i=\tau+1}^{N}(\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{\beta}_{IC}^{\tilde{\boldsymbol{\gamma}}})^T\boldsymbol{B}_{\boldsymbol{\gamma}}(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}\boldsymbol{B}_{\boldsymbol{\gamma}}^T(\boldsymbol{Y}_i - \boldsymbol{B}_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{\beta}_{IC}^{\tilde{\boldsymbol{\gamma}}}))/2,
\end{aligned} \tag{4.27}$$

where $(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}$ exists because $\boldsymbol{B}$ is a full rank matrix, and so as $\boldsymbol{B}_{\boldsymbol{\gamma}}$. See the proof of Equation (4.27) in Appendix E.

4.2.3 Diagnosis by MCMC

Notice that all the above conditional posteriors of τ and $\boldsymbol{\gamma}$ are conditional on g. Integrating g out from the posteriors, however, either requires calculating Gaussian hypergeometric functions, which could incur numerical overflow (Liang et al., 2008) problems, or almost impossible. Besides, the exhaustive calculation of $P(\boldsymbol{\gamma}, \tau|\boldsymbol{D}, g)$ is quite time consuming or infeasible for problems of large size data or high dimension profile parameters, since we have $(n-1) \times 2^p$ possible combinations of $(\boldsymbol{\gamma}, \tau)$. Therefore, we decide to implement MCMC methodologies to calculate the empirical posterior distribution of $(\tau, \boldsymbol{\gamma})$. Particularly, we implement a Gibbs sampler to

alternatively sample τ, $\boldsymbol{\gamma}$ and τ from their conditional posterior distributions:

$$\begin{aligned} P(\tau|\boldsymbol{D}, g, \boldsymbol{\gamma}) &\propto \pi_\tau(\tau) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, g) \\ P(\boldsymbol{\gamma}|\boldsymbol{D}, g, \tau) &\propto \pi_\gamma(\boldsymbol{\gamma}) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, g) \\ \pi_g(g|\boldsymbol{D}, \tau, \boldsymbol{\gamma}) &\propto \pi_g(g) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, g), \end{aligned} \tag{4.28}$$

for Phase I diagnosis, or

$$\begin{aligned} P(\tau|\boldsymbol{D}, g, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}) &\propto \pi_\tau(\tau) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}, g) \\ P(\boldsymbol{\gamma}|\boldsymbol{D}, g, \tau, \boldsymbol{\beta}_{IC}) &\propto \pi_\gamma(\boldsymbol{\gamma}) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}, g) \\ \pi_g(g|\boldsymbol{D}, \tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}) &\propto \pi_g(g) L(\boldsymbol{D}|\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}_{IC}, g) \end{aligned} \tag{4.29}$$

for Phase II diagnosis. Although the conditional distributions in Equations (4.28) and (4.29) are known up to a constant, they are neither of explicit forms nor commonly used distributions. Therefore, we implement Metropolis-in-Gibbs sampling procedures; that is, τ, $\boldsymbol{\gamma}$ and g are sampled from the conditional posterior distribution sequentially, whilst such sampling is completed via a Metropolis-Hasting (MH) proposal kernel. Denote the initial samples as $(\tau^{(0)}, \boldsymbol{\gamma}^{(0)}, g^0)$, each of which could be sampled from the priors defined before. Suppose the samples at the tth iteration are $(\tau^{(t)}, \boldsymbol{\gamma}^{(t)}, g^{(t)})$, then they move to new values at $t+1$ by:

- generate $\tau^{(*)} \sim q_\tau(\tau^{(*)}|\tau^{(t)})$. $q_\tau(\cdot|\tau^{(t)})$ could either be generated randomly from a discrete uniform distribution on the support of $\pi_\tau(\tau)$ (i.e., free move) or from a neighborhood of $\tau^{(t)}$ of length $d_{\tau^{(t)}}$ except $\tau^{(t)}$ (i.e., local move). In our work, we adopt the local move with $d_{\tau^{(t)}} = 5$. The acceptance ratio is

$$r(\tau^{(t)}, \tau^{(*)}) = \frac{\pi_{\tau^{(*)}} L(\boldsymbol{D}|\tau^{(*)}, g^{(t)}, \boldsymbol{\gamma}^{(t)}) / q_\tau(\tau^{(*)}|\tau^{(t)})}{\pi_{\tau^{(t)}} L(\boldsymbol{D}|\tau^{(t)}, g^{(t)}, \boldsymbol{\gamma}^{(t)}) / q_\tau(\tau^{(t)}|\tau^{(*)})}, \tag{4.30}$$

and we accept the move, i.e., $\tau^{(t+1)} = \tau^{(*)}$, by probability $\min(1, r(\tau^{(t-1)}, \tau^{(*)}))$;

- generate $\boldsymbol{\gamma}^{(*)} \sim q_{\gamma}(\boldsymbol{\gamma}^{(*)}|\boldsymbol{\gamma}^{(t)})$. $\boldsymbol{\gamma}^{(*)}$ is sampled by using the *"rev.jump"* sampling algorithm in Zeugner (2011), and see the reference for details. The acceptance ratio for such proposal is:

$$r(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\gamma}^{(*)}) = \frac{\pi_{\gamma}(\boldsymbol{\gamma}^{(*)}) L(\boldsymbol{D}|\boldsymbol{\gamma}^{(*)}, g^{(t)}, \tau^{(t+1)}) / q_{\gamma}(\boldsymbol{\gamma}^{(*)}|\boldsymbol{\gamma}^{(t)})}{\pi_{\gamma}(\boldsymbol{\gamma}^{(t)}) L(\boldsymbol{D}|\boldsymbol{\gamma}^{(t)}, g^{(t)}, \tau^{(t+1)}) / q_{\gamma}(\boldsymbol{\gamma}^{(t)}|\boldsymbol{\gamma}^{(*)})}, \tag{4.31}$$

and then $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(*)}$ with probability $\min(1, r(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\gamma}^{(*)}))$;

- generate $g^{(*)}$ from $\pi_g(\cdot)$ (equivalent to generating $\frac{g^{(*)}}{1+g^{(*)}}$ from $Beta(1, a/2-1)$ and then resolving $g^{(*)}$). The acceptance ratio is simply

$$r(g^{(t)}, g^{(*)}) = \frac{L(\boldsymbol{D}|g^{(*)}, \boldsymbol{\gamma}^{(t+1)}, \tau^{(t+1)})}{L(\boldsymbol{D}|g^{(t)}, \boldsymbol{\gamma}^{(t+1)}, \tau^{(t+1)})}, \tag{4.32}$$

and the move is accepted with probability $\min(1, r(g^{(t)}, g^{(*)}))$.

The above procedures are for Phase I diagnosis and assume that $\boldsymbol{\beta}_{IC}$ is unknown, and the implementation should be straightforward for Phase II diagnosis. During the sample, τ and $\boldsymbol{\gamma}$ are sampled independently, and we allow the existence of $\tau^{(t)} = N$ and $\boldsymbol{\gamma}^{(t)} \neq \mathbf{0}$ or $\tau^{(t)} \neq N$ and $\boldsymbol{\gamma}^{(t)} = \mathbf{0}$. By doing so, the restriction bound between τ and $\boldsymbol{\gamma}$ is released, and this facilitates the mixing efficiency of the Markov chain. The diagnosis algorithm in this chapter is summarized as Algorithm 4.1.

Algorithm 4.1 MCMC algorithm for multivariate linear profile diagnosis

Data: $\boldsymbol{Y}_i$, $i = 1, \cdots, N$

Input: $\boldsymbol{B}$, hyperparameters $(\nu_1, \nu_2, a, \omega_0)$, and the iteration number R

Output: MCMC samples

Initialization: randomly generate $(\tau^{(0)}, \boldsymbol{\gamma}^{(0)}, g^{(0)})$

for $t + 1 = 1$ **to** R **do**

generate $\tau^{(*)} \sim q_\tau(\tau^{(*)}|\tau^{(t)})$, and take

$$\tau^{(t+1)} = \begin{cases} \tau^{(t)} & \text{with probability} \quad 1 - r(\tau^{(t)}, \tau^{(*)}) \\ \tau^{(*)} & \text{with probability} \quad r(\tau^{(t)}, \tau^{(*)}) \end{cases}$$

by calculating $r(\tau^{(t)}, \tau^{(*)})$ from Equation (4.30);

generate $\boldsymbol{\gamma}^{(*)} \sim q_\gamma(\boldsymbol{\gamma}^{(*)}|\boldsymbol{\gamma}^{(t)})$, and take

$$\boldsymbol{\gamma}^{(t+1)} = \begin{cases} \boldsymbol{\gamma}^{(t)} & \text{with probability} \quad 1 - r(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\gamma}^{(*)}) \\ \boldsymbol{\gamma}^{(*)} & \text{with probability} \quad r(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\gamma}^{(*)}) \end{cases}$$

by calculating $r(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\gamma}^{(*)})$ from Equation (4.31);

generate $g^{(*)} \sim \pi_g(\cdot)$, and take

$$g^{(t+1)} = \begin{cases} g^{(t)} & \text{with probability} \quad 1 - r(g^{(t)}, g^{(*)}) \\ g^{(*)} & \text{with probability} \quad r(g^{(t)}, g^{(*)}) \end{cases}$$

by calculating $r(g^{(t)}, g^{(*)})$ from Equation (4.32);

end for

Diagnostic result: the *maximum a posteriori* (MAP) estimate of $(\tau, \boldsymbol{\gamma})$

As we explained before, since since $\tau = N$ and $\boldsymbol{\gamma} = 0$ equivalently indicate the profile is in control, so for all samples with either $\tau^{(t)} = N$ or $\boldsymbol{\gamma}^{(t)} = \mathbf{0}$, they can be aggregated as $\tau = N$ and $\boldsymbol{\gamma} = 0$ for interpretation convenience, and then we have $P(\tau = N|\boldsymbol{D}) = P(\boldsymbol{gamma} = \mathbf{0}|\boldsymbol{D})$.

4.3 Performance study

In this section, the performance of our Bayesian diagnosis is evaluated by simulations under different scenarios. The diagnosis is tested for both Phase I and Phase II usage under multiple scenarios. We choose $m = 12$, $p = 8$, and the elements of $\boldsymbol{B}$ are randomly sampled from Unif$(-1, 1)$ for all simulation tests. $\boldsymbol{Y}_i$s are then generated from Equation (4.4) with $\sigma = 1$. For the priors settings in Section 4.2.1, we choose $\nu_1 = N$, and $\nu_2 = \nu_1\sigma^2$, so that the prior mean is close to the true measurement variance. To assure the proper $\pi_g(g|\cdot)$ in Equations (4.28) and (4.29), we choose $a = 3$ as explained for $\pi_g(g)$. Further, we choose $\omega_0 = 0.5$ so π_γ is noninformative. The setting of $\pi_\tau(\tau)$ will be explained later according to specific diagnostic requirements.

The MCMC diagnostic algorithm utilizes the first $R_b = 5000$ MCMC iterations for burning and the following $R_i = 2 \times 2^p \times N$ iterations for inference. To assess the reliability of the Bayesian diagnosis, each specific diagnostic task is replicated by 1000 times with 1000 sets of randomly generated $\boldsymbol{D}$. Two statistics are used to evaluate the diagnostic performance: firstly, the average of the posterior probabilities of τ and $\boldsymbol{\gamma}$ approximately calculated from the MCMC stage, which are denoted as $\hat{P}(\tau|\boldsymbol{D})$ and $\hat{P}(\boldsymbol{\gamma}|\boldsymbol{D})$, and secondly, r_τ and r_γ, the respective ratio of the MAP estimates of τ and $\boldsymbol{\gamma}$ being true.

4.3.1 Diagnostic Performance on In Control Process

Before we assess the Bayesian diagnostic effectiveness, it is necessary to investigate how well Algorithm 4.1 controls generating type I error. The idea is similar

to studies of in control monitoring characteristics of SPC tools, e.g., AIRL, based on which parameters of SPC tools are designed.

Phase I In Control Diagnosis

Phase I analysis is used to provide knowledge of $\boldsymbol{\beta}_{IC}$ for Phase II applications. In quality control area, we are interested in (i) if $\boldsymbol{\beta}_t$ is inconsistent within $\boldsymbol{D}$, and if not then (ii) on which components of $\boldsymbol{\beta}_{IC}$ the shift happens. The Bayesian diagnosis is performed on $\boldsymbol{D}$ of length $N = \{40, 80, 160\}$, representing short, medium, and long process runs. As mentioned before, p_0 should be determined by the stability of the process. We choose $p_0 = 1/N$ so the AIRL is assumed to be N. However, it should clarified that p_0 is intrinsically independent of the sample size N, and in this work we set p_0 as so mainly because we make no assumption on the AIRL, which should be available in practice. The Bayesian diagnosis is conducted on $\boldsymbol{D}$ of length $N = \{40, 80, 160\}$, representing short, medium, and long process runs.

The results are given in Tables 4.1 and 4.2 (the symbol $*$ indicates the true value). It can be seen that r_τ and r_γ is controlled no less than 0.99 and 0.995 respectively. Therefore, the type I error of diagnosing τ is controlled effectively. Meanwhile, the empirical estimated value of the posterior probability of $(\tau, \boldsymbol{\gamma}) = (0, \mathbf{0})$ is more than 100 times larger than that of other candidate diagnostic results. In addition, since $\tau = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$ are equivalent, so either of them being the MAP estimate leads to the judgment of $\boldsymbol{D}$ is in control.

Table 4.1: Phase I In Control Diagnosis of τ

(a) $N = 40$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*40$	0.361	0.99
2	0.023	
3	0.023	
1	0.022	
4	0.022	

(b) $N = 80$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*80$	0.358	0.98
10	0.011	
4	0.011	
2	0.011	
7	0.011	

(c) $N = 160$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*160$	0.407	0.995
31	0.005	
4	0.005	
6	0.005	
6	0.005	

Table 4.2: Phase I In Control Diagnosis of $\boldsymbol{\gamma}$

(a) $N = 40$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.361	0.999
$1,1,0,1,1,1,0,1$	0.003	
$1,0,0,1,1,1,1,1$	0.003	
$1,1,1,0,1,1,1,0$	0.003	
$0,1,1,1,0,1,0,1$	0.003	

(b) $N = 80$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.358	0.995
$1,1,0,1,1,1,1,0$	0.003	
$1,1,1,1,0,1,0,1$	0.003	
$1,1,0,1,0,1,1,1$	0.003	
$1,1,1,0,1,0,1,1$	0.003	

(c) $N = 160$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.407	0.995
$0,1,1,0,1,0,1,1$	0.003	
$1,1,1,0,1,0,1,1$	0.003	
$0,1,0,1,1,0,1,1$	0.003	
$0,0,0,1,1,0,0,0$	0.003	

Phase II In Control Diagnosis

The Phase II diagnosis should be conducted after the SPC monitoring signals. That is, the diagnosis is employed as a retrospective procedure. Usually $\boldsymbol{\beta}_{IC}$ is accessible in Phase II stages, so out of control data are mainly concerned. Therefore, for diagnostic purpose, only a small fragment of $\boldsymbol{D}$ should be collected before the monitoring signal to save diagnostic cost. The number of observations in the small fragment can be chosen as the average out of control run length of the SPC monitor-

ing; meanwhile, the process is commonly allowed to continue for a short period after the signal to collect enough out of control observations. In Phase II diagnostic tests, we set $N = 10, 20, 40$. For the $\pi_\tau(\tau)$ in Equation (4.10), λ is set by assuming the false alarm rate equals 0.1 for the monitoring, i.e.,

$$\frac{e^\lambda}{e^\lambda + N} = 0.1,$$

and it can be adjusted to counter the robustness of the monitoring in practice. It needs to be mentioned that if the τ is diagnosed as 0, more observations ahead of the signal position are required to track the exact location of τ. The Phase II diagnostic results on in control observations are given in Tables 4.3 and 4.4, from which we $\tau = 0$ is correctly diagnosed in 80%, 89%, 94.2% runs respectively, whereas all $r_{\boldsymbol{\gamma}}$s in Table 4.4 are greater than 0.98. Since $\tau = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$ are exchangable, so either of them being the MAP estimate leads to the correct diagnosis. Moreover, it should be noted that such diagnostic results are conditional on the SPC signaling incorrectly, and the monitoring false alarm rate should be considered to estimate the true type I error. For example, the true type I error when $N = 40$ should be approximately $0.1 \times (1 - r_{\boldsymbol{\gamma}})$.

4.3.2 Diagnostic Performance on Inhomogeneous Process

After the in control tests, we now assess how Algorithm 4.1 performs on diagnosing the shift position and pattern. The observations generated here contains two parts, i.e., in control and out of control segments. N, σ^2 and priors are set as in Section 4.3.1. Two out of control profiles are considered, namely $\boldsymbol{\beta}_{OC}^{\gamma_1} = [0, -1.5, 1.8, 0, 0, 2.2]^T$

Table 4.3: Phase II In Control Diagnosis of τ

(a) $N = 10$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*10$	0.307	0.8
0	0.135	
9	0.104	
8	0.080	
7	0.069	

(b) $N = 20$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*20$	0.363	0.89
19	0.080	
18	0.064	
17	0.055	
16	0.044	

(c) $N = 40$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
$*40$	0.447	0.942
0	0.049	
2	0.036	
4	0.029	
3	0.025	

Table 4.4: Phase II In Control Diagnosis of $\boldsymbol{\gamma}$

(a) $N = 10$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.307	0.982
$0,0,0,1,0,0,1,0$	0.004	
$0,0,0,0,0,0,1,1$	0.004	
$0,0,0,0,1,0,1,0$	0.004	
$0,0,0,1,1,0,0,0$	0.004	

(b) $N = 20$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.362	0.984
$0,0,0,0,1,0,0,1$	0.004	
$0,0,0,0,1,1,0,0$	0.004	
$0,0,1,0,1,0,0,0$	0.004	
$0,0,0,1,0,0,0,1$	0.004	

(c) $N = 40$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,0,0,0,0,0,0$	0.447	0.986
$0,0,0,1,0,1,0,0$	0.004	
$0,0,0,0,0,1,1,0$	0.004	
$0,0,0,1,0,0,1,0$	0.004	
$0,0,0,0,0,0,1,1$	0.004	

or $\boldsymbol{\beta}_{OC}^{\gamma_2} = [0,0,2,0,0,0,0,0]^T$.

Phase I Diagnosis of τ and $\boldsymbol{\gamma}$

Assuming in practice the shift could happen at arbitrary time, $\tau = \tau_0$ is randomly generated from $1 : N-1$ for each of the 1000 runs, and $\boldsymbol{Y}_{1,\cdots,N}$ are then generated accordingly. The simulation results are given in Tables 4.5 - 4.10. Since τ is *not* consistent within the 1000 runs, so we provided the average approximate

Table 4.5: Phase I Diagnosis of τ with $N = 40$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.985	0.998

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.898	0.994

posterior probability $\hat{P}(\tau = \tau_0|\boldsymbol{D})$. The correctness of diagnosing τ is close to 1, which indicate the diagnosis is sensitive to the changepoint. The diagnosis of $\boldsymbol{\gamma}$ is affected by the amount of available observations and the scale of $\boldsymbol{\beta}_{OC}$. The value of $r_{\boldsymbol{\gamma}}$ is less plausible compared with that for in control diagnosis, especially when $\boldsymbol{\beta}_{OC} = \boldsymbol{\beta}_{OC}^{\gamma_2}$ and $N = 40$. However, it should be noted that: firstly, the main target of Phase I analysis is to provide knowledge of $\boldsymbol{\beta}_{IC}$, so identifying τ is more crucial for the Phase I diagnosis; secondly, for $\boldsymbol{\beta}_{OC} = \boldsymbol{\beta}_{OC}^{\gamma_2}$, the diagnosis still obtains the correct fault pattern in at least 67.5% of the 1000 simulations, so for shifts happen on more than one profile components, our algorithm can diagnose the fault with reasonable probability; thirdly, we notice that the only deviated profile component in $\boldsymbol{\beta}_{OC}^{\gamma_2}$ is included in all estimates of $\boldsymbol{\gamma}$ with nontrivial posterior probability ($\boldsymbol{\beta}_{OC}^{\gamma_1}$ as well), as shown in Tables 4.6, 4.8 and 4.10, so the inspection after the diagnosis has large chance to identify the true fault. Therefore, the Phase I diagnosis can provide required insight to screen in control observations and profiles.

Phase II Diagnosis of τ and $\boldsymbol{\gamma}$

Considering for the Phase II diagnosis, the majority of inspected observations should be out of control (if the signal is true), in the simulation here $\tau = \tau_0$ is

Table 4.6: Phase I Diagnosis of $\boldsymbol{\gamma}$ with $N = 40$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,1,0,0,1,0,0$	0.223	0.675
$0,1,1,1,0,1,0,0$	0.074	
$0,1,1,0,0,1,0,1$	0.070	
$0,1,1,0,1,1,0,0$	0.068	
$0,1,1,0,0,1,1,0$	0.064	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.047	0.206
$0,0,1,0,0,0,1,0$	0.034	
$0,0,1,0,0,1,0,0$	0.033	
$0,0,1,1,0,0,0,0$	0.33	
$1,0,1,0,0,0,0,0$	0.033	

randomly generated from $0 : N/4$. Tables 4.11 - 4.16 show the Phase II diagnostic effectiveness is remarkable. τ is correctly diagnosed in (almost) all of the 1000 simulations, with dominating posterior probabilities ($\hat{P}(\tau = \tau_0|\boldsymbol{D}) \geq 0.789$). Meanwhile, as we expected, the true value of $\boldsymbol{\gamma}$ is diagnosed with much higher correctness than in Section 4.3.2; for example, $r_\gamma = 0.574$ on just one profile component fault given $N = 10$ observations, and the value increases to 0.816 when 40 observations are available.

The results suggest Algorithm 4.1, when implemented as a Phase II diagnostic tool, works appropriately even with small amount of data, and this could be an appealing property for short run applications. One needs to be mentioned is if the *MAP* τ is 0, more observations ahead of $\boldsymbol{D}$ are required to track the exact location of τ.

Table 4.7: Phase I Diagnosis of τ with $N = 80$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.986	0.998

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.776	0.954

4.3.3 Sparse Diagnosis on $\boldsymbol{\gamma}$

From the previous tests we see that the diagnosis on τ is reliable in most cases, and concerns mainly lay on diagnosing $\boldsymbol{\gamma}$ if only a single fault happens on profile parameters and small amount of data is available. That is, the diagnosis is capable in capturing the inhomogeneity among observations precisely, but it could require extra assistance to identify the fault pattern exactly. In fact, for most reliable system, the number of faults on profile components is small when the process is out of control, so the diagnosis could consider scrutinizing sparse faults with higher priority, particularly if the diagnosis is unable to provide exclusive conclusion of $\boldsymbol{\gamma}$.

To make the diagnosis more parsimonious on $\boldsymbol{\gamma}$, we can choose a more informative π_γ, rather than the uniform one used before. Here repeated the diagnostic tests on $\boldsymbol{\beta}_{OC} = \boldsymbol{\beta}_{OC}^{\gamma_2}$ and $N = 40$ (Phase I) and $N = 10$ (Phase II), considering such situations are the most challenging. The diagnosis of τ is not affected so the results are not given. The diagnostic results of $\boldsymbol{\beta}_{OC}^{\gamma_2}$ with the sparse prior of $\boldsymbol{\gamma}$ are given in Tables 4.17 and 4.18. From which we can see that the diagnostic correctness of $\boldsymbol{\gamma}$ in the 1000 simulations is increased to 0.888 and 0.968 for the Phase I and Phase II tests

Table 4.8: Phase I Diagnosis of $\boldsymbol{\gamma}$ with $N = 80$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,1,0,0,1,0,0$	0.296	0.744
$0,1,1,1,0,1,0,0$	0.080	
$0,1,1,0,0,1,1,0$	0.078	
$0,1,1,0,0,1,0,1$	0.075	
$0,1,1,0,0,1,1,0$	0.074	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.061	0.27
$0,0,1,1,0,0,0,0$	0.041	
$0,0,1,0,0,1,0,0$	0.039	
$1,0,1,0,0,0,0,0$	0.038	
$0,0,1,0,0,0,0,1$	0.038	

Table 4.9: Phase I Diagnosis of τ with $N = 160$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.983	0.999

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.846	0.998

Table 4.10: Phase I Diagnosis of $\boldsymbol{\gamma}$ with $N = 160$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,1,0,0,1,0,0$	0.377	0.83
$1,1,1,0,0,1,0,0$	0.082	
$0,1,1,0,0,1,1,0$	0.075	
$0,1,1,0,0,1,0,1$	0.075	
$0,1,1,0,1,1,0,0$	0.075	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.143	0.37
$0,0,1,0,0,1,0,0$	0.056	
$0,0,1,0,0,0,0,1$	0.055	
$0,1,1,0,0,0,0,0$	0.053	
$0,0,1,0,1,0,0,0$	0.052	

respectively, and the corresponding posterior distribution of $\boldsymbol{\gamma}|\boldsymbol{D}$ is also more "spiky" at the true pattern than it is in previous tests. The results prove that utilizing the sparse prior of $\boldsymbol{\gamma}$ helps improve the diagnostic effectiveness on sparse faults. In addition, such improvement would not undermine the diagnostic capability on non-sparse faults, which in fact have larger chance to be identified. We also should emphasize that the invariably high correctness of diagnosing τ makes improving the diagnostic effectiveness with respect to $\boldsymbol{\gamma}$ smooth.

4.4 Case Study

Parker et al. (2001) introduced a calibration case at NASA Langley Research Center, in which a multivariate linear regression model is used to relate 6 response variables (electrical measurement of forces and moments) and 6 explanatory variables (forces and moments). By removing the intercepts, the calibration model can be

Table 4.11: Phase II Diagnosis of τ with $N = 10$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.985	0.998

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.789	1

Table 4.12: Phase II Diagnosis of $\boldsymbol{\gamma}$ with $N = 10$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,1,0,0,1,0,0$	0.385	0.844
$1,1,1,0,0,1,0,0$	0.079	
$0,1,1,0,0,1,1,0$	0.078	
$0,1,1,1,0,1,0,0$	0.077	
$0,1,1,0,0,1,0,1$	0.077	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.131	0.574
$1,0,1,0,0,0,0,0$	0.052	
$0,1,1,0,0,0,0,0$	0.051	
$0,0,1,1,0,0,0,0$	0.051	
$0,0,1,0,1,0,0,0$	0.050	

supposed to:

$$\begin{bmatrix} Y_{i,1} \\ Y_{i,2} \\ Y_{i,3} \\ Y_{i,4} \\ Y_{i,5} \\ Y_{i,6} \end{bmatrix} = \begin{bmatrix} 10 & -0.01 & -0.03 & 0.26 & 0 & 0.03 \\ 0.24 & 21.01 & -0.09 & 0.03 & -0.12 & 0.01 \\ 0.09 & 0.01 & 6.81 & 0.04 & 0.02 & -0.03 \\ 0 & 0 & 0 & 10.53 & -0.47 & 0.21 \\ -0.021 & 0 & 0.01 & 0.02 & 7 & -0.34 \\ 0.04 & 0 & -0.01 & 0.18 & -0.34 & 11.46 \end{bmatrix} \begin{bmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \beta_{i,3} \\ \beta_{i,4} \\ \beta_{i,5} \\ \beta_{i,6} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}, \quad (4.33)$$

where $\epsilon_{i.} \overset{iid}{\sim} Norm(0, \sigma^2)$. For the calibration aim, the value of response is measured

Table 4.13: Phase II Diagnosis of τ with $N = 20$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.991	1

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.985	1

by a network of strain gauges. The value of $\boldsymbol{\beta}_{IC}$ is obtained from calibration experiment. Therefore here we only consider applying the Bayesian diagnosis on Phase II stages. Suppose $\boldsymbol{\beta}_{IC} = [0, 1.5, 0, 1, 0, -1]^T$, $\sigma^2 = 1$, and $N = 20$ measurement data is available. We consider two out of control profiles, namely $\boldsymbol{\beta}_{OC} = [0, 2, 0, 1, 1, -2]^T$ and $\boldsymbol{\beta}_{OC} = [1, 1.5, 0, 1, 0, -1]^T$. We investigate the diagnosis on this application by following the same simulation settings in Section 4.3.2.

The results is given in Tables 4.19 and 4.20, from which we see that both τ and $\boldsymbol{\gamma}$ are diagnosed with high correctness.

4.5 Conclusion

We have proposed a Bayesian diagnostic approach for multivariate linear profile models. We have illustrated the diagnosis as an integration of offline changepoint detection and stochastic search variable selection for identifying fault patterns. The novel diagnosis has several advantages over existing diagnostic frameworks: firstly, it is eligible to provide estimates on both the changepoint and shifted profiles simultaneously, whereas most other diagnostic tools for profiles requires the exact position

Table 4.14: Phase II Diagnosis of $\boldsymbol{\gamma}$ with $N = 20$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	$r_{\boldsymbol{\gamma}}$
$*0, 1, 1, 0, 0, 1, 0, 0$	0.504	0.894
$0, 1, 1, 0, 1, 1, 0, 0$	0.074	
$1, 1, 1, 0, 0, 1, 0, 0$	0.073	
$0, 1, 1, 0, 0, 1, 0, 1$	0.071	
$0, 1, 1, 0, 0, 1, 1, 0$	0.070	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	$r_{\boldsymbol{\gamma}}$
$*0, 0, 1, 0, 0, 0, 0, 0$	0.337	0.818
$0, 0, 1, 0, 0, 0, 1, 0$	0.064	
$0, 0, 1, 0, 0, 0, 0, 1$	0.064	
$0, 0, 1, 0, 1, 0, 0, 0$	0.062	
$0, 0, 1, 1, 0, 0, 0, 0$	0.062	

Table 4.15: Phase II Diagnosis of τ with $N = 40$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.998	1

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.877	0.998

Table 4.16: Phase II Diagnosis of $\boldsymbol{\gamma}$ with $N = 40$

(a) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_1}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,1,0,0,1,0,0$	0.682	0.954
$0,1,1,0,0,1,0,1$	0.055	
$0,1,1,0,1,1,0,0$	0.054	
$0,1,1,0,0,1,1,0$	0.052	
$0,1,1,1,0,1,0,0$	0.049	

(b) Diagnosis under $\boldsymbol{\beta}_{OC}^{\gamma_2}$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.344	0.816
$0,0,1,0,0,1,0,0$	0.068	
$1,0,1,0,0,0,0,0$	0.064	
$0,0,1,0,0,0,1,0$	0.062	
$0,0,1,0,0,0,1,0$	0.061	

of changepoint is given (Mahmoud et al., 2007; Zou et al., 2007a), so the proposed diagnostic method here can be conducted more economically in practice; secondly, the Bayesian diagnosis is able to handle multivariate linear profiles by estimating the fault pattern $\boldsymbol{\gamma}$ via a MCMC procedure; thirdly, the Bayesian diagnosis can be flexibly employed for both Phase I and Phase II applications, and to the best of our knowledge, there is no other alternatives own such capability. The simulation study illustrate that by setting priors reasonably the proposed Bayesian method provide reliable diagnostic insights of out of control information whilst controlling the type I error rate. Finally, we have demonstrated the diagnostic effectiveness on a calibration application example.

Table 4.17: Phase I Diagnosis of $\boldsymbol{\gamma}$ with Sparse Prior

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.529	0.888
$0,1,1,0,0,0,0,0$	0.055	
$1,0,1,0,0,0,0,0$	0.055	
$0,0,1,1,0,0,0,0$	0.053	
$0,0,1,0,1,0,0,0$	0.051	

Table 4.18: Phase II Diagnosis of $\boldsymbol{\gamma}$ with Sparse Prior

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,0,1,0,0,0,0,0$	0.755	0.968
$1,0,1,0,0,0,0,0$	0.037	
$0,0,1,0,0,1,0,0$	0.034	
$0,0,1,0,0,0,1,0$	0.032	
$0,0,1,0,0,0,0,1$	0.031	

Table 4.19: Diagnosis of τ for Calibration Application

(a) $\boldsymbol{\beta}_{OC} = [1, 1.5, 0, 1, 0, -1]^T$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	0.998	1

(b) $\boldsymbol{\beta}_{OC} = [0, 2, 0, 1, 1, -2]^T$

τ	$\hat{P}(\tau\|\boldsymbol{D})$	r_τ
τ_0	1	1

Table 4.20: Diagnosis of $\boldsymbol{\gamma}$ for Calibration Application

(a) $\boldsymbol{\beta}_{OC} = [1, 1.5, 0, 1, 0, -1]^T$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*1,0,0,0,0,0$	0.675	0.852
$1,0,1,0,0,0$	0.035	
$1,0,0,0,1,0$	0.035	
$1,0,0,1,0,0$	0.034	
$1,0,0,0,0,1$	0.034	

(b) $\boldsymbol{\beta}_{OC} = [0, 2, 0, 1, 1, -2]^T$

$\boldsymbol{\gamma}^T$	$\hat{P}(\boldsymbol{\gamma}\|\boldsymbol{D})$	r_γ
$*0,1,0,0,1,1$	0.850	0.933
$0,1,0,1,1,1$	0.051	
$1,1,0,0,1,1$	0.049	
$0,1,1,0,1,1$	0.022	
$0,1,1,1,1,1$	0.016	

APPENDIX A
OPTIMAL RESAMPLING IN SECTION 2.2.2

Suppose N particles need to be resampled.

1. given particles $\{\widetilde{\mathscr{P}}_{t+1}^{j}\}_{j=1}^{R(N)}$ with weights $\{\tilde{w}_{t+1}^{j}\}_{j=1}^{R(N)}$ sum to 1. Calculate c, such that $\sum_{j=1}^{R(N)} min(c\tilde{w}_{t+1}^{j}, 1) = N$, and c is unique;

2. for $j = 1, \ldots, R(N)$, assume L particles satisfy $\{\tilde{w}_{t+1}^{j}\} \geq 1/c$, then we keep these particles $\widetilde{\mathscr{P}}_{t+1}^{j}$ with their original weights.

3. use the stratified sampling method of Carpenter et al. (1999) to resample $N-L$ from the remaining $R(N)-L$ particles, and assign weights $1/c$ to them.

A.1 Stratified Sampling

The stratified sampling algorithm is implemented as follows: let weights of $R(N)-L$ particles be $\{\tilde{w}_{t+1}^{j}\}_{j=1}^{R(N)-L}$, and to sample $N-L$ particles from them:

1. let $U = \sum_{j=1}^{R(N)-L} \tilde{w}_{t+1}^{j}$ and u be randomly sampled form $Unif[0, U]$;

2. let $u = u - \tilde{w}_{t+1}^{j}$. Only if $u \leq 0$ then keep the particle, and set $u = u - \tilde{w}_{t+1}^{j}$

3. repeat step 2 until $N-L$ particles in total are sampled

APPENDIX B
PROOF OF LEMMA 3.1

By Theorem 1 in Zhou et al. (2003): if $\boldsymbol{\theta}_j = \sigma_j^2$ is uniquely diagnosable, then $\boldsymbol{p}_j \in R(\boldsymbol{H})$, and $\boldsymbol{p}_j$ is a identity vector with the jth element equals 1. Therefore, if the fault-quality model is fully diagnosable, then $\boldsymbol{H}$ is of full rank.

APPENDIX C
PROOF OF LEMMA 3.2

A process is fully diagnosable on mean shift if and only if $\boldsymbol{B}$ is of full rank (Li and Chen, 2016), that is, $\text{rank}(\boldsymbol{B}) = p$, and then $\text{rank}(\boldsymbol{B}^T\boldsymbol{B}) = \text{rank}(\boldsymbol{B}) = p$; since $\boldsymbol{B}^T\boldsymbol{B}$ is positive definite, so by the Schur product theorem (Li et al., 2007) we know $\text{rank}(\boldsymbol{H}) = p$ and all variance faults are identifiable. If $\boldsymbol{B}$ is not full rank, and $m < p$, then $\text{rank}(\boldsymbol{B}) \leq m$, so $\boldsymbol{H}$ is positive semidefinite and $m \leq \text{rank}(\boldsymbol{H}) \leq m^2$ (Bapat and Sunder, 1985; Rao and Kleffe, 1988), and thus unless the rank of $\boldsymbol{H}$ is achieved, the fully diagnosable property is uncertain.

APPENDIX D
B IN CASE STUDY

$$
B = \left[\begin{array}{ccccccccc|ccc}
1 & 0.785 & -0.785 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.143 & -0.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0. & 0 & 1 & 1.1 & -1.1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2.2 & -1.26 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.401 & -0.786 & 0 & 0 & 0.38 & 1 & 0.385 & -0.385 & 1 & 0.385 & -0.385 \\
0 & 0.073 & -0.143 & 0 & 0 & 0.07 & 0 & 1.02 & -0.07 & 0 & 1.02 & -0.07 \\
\hline
\multicolumn{9}{c|}{\multirow{4}{*}{$\mathbf{0}^{4\times 6}$}} & 0 & 0 & 0 \\
 & & & & & & & & & 0 & 0 & 0 \\
 & & & & & & & & & -1 & -0.096 & 0 \\
 & & & & & & & & & 0 & 0.057 & 0
\end{array}\right]
$$

$$
\left[\begin{array}{ccc|cccccc}
\multicolumn{3}{c|}{\mathbf{0}^{6\times 3}} & \multicolumn{6}{c}{\mathbf{0}^{6\times 6}} \\
\hline
1 & 0.60 & -1.60 & \multicolumn{6}{c}{\multirow{2}{*}{$\mathbf{0}^{2\times 6}$}} \\
0 & 2.48 & -1.48 & & & & & & \\
\hline
\multicolumn{3}{c|}{\multirow{2}{*}{$\mathbf{0}^{2\times 3}$}} & 1 & 0.24 & -0.24 & 1.24 & -0.096 & -0.096 \\
 & & & 0 & -0.142 & 1.142 & -0.142 & 1.050 & -1.157
\end{array}\right]
$$

APPENDIX E
DERIVATIONS OF $L(\boldsymbol{D}|G, \boldsymbol{\gamma}, \tau)$

E.1 Derivation of Equations (4.14) and (4.17)

- for Equation (4.14) Because $\tau = 0$, so $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{IC}$ for $i = 1, \cdots, N$. Therefore

$$\boldsymbol{\beta}_{IC}|\boldsymbol{D}, g, \sigma^2, \tau = 0 \sim Norm((\frac{1}{g} + N)^{-1}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\sum_{i=1}^{N}\boldsymbol{Y}_i, (\frac{1}{g} + N)^{-1}\sigma^2(\boldsymbol{B}^T\boldsymbol{B})^{-1}).$$

So

$$L(\boldsymbol{D}|g, \tau = 0, \sigma^2) \propto (\sigma^2)^{-\frac{mN}{2}}(\frac{1}{1+gN})^{\frac{p}{2}}\exp(-\frac{||\boldsymbol{D}||^2}{2\sigma^2} + \frac{g}{1+gN}\frac{(\boldsymbol{B}^T\sum_{i=1}^{N}\boldsymbol{Y}_i)^T(\boldsymbol{B}^T\boldsymbol{B})^{-1}(\boldsymbol{B}^T\sum_{i=1}^{N}\boldsymbol{Y}_i)}{2\sigma^2}),$$

where $||\boldsymbol{D}||^2 = \sum_{i=1}^{N}||\boldsymbol{Y}_i||^2$. Since we use the conjugate prior $\pi_\sigma^2(\sigma^2)$, so it has

$$\sigma^2|\boldsymbol{D}, g, \tau = 0 \sim IG(\frac{\nu_1'}{2}, \frac{\nu_2'}{2}),$$

where $\frac{\nu_1'}{2}$, $\frac{\nu_1'}{2}$ are defined in Equation (4.15) and then we have the results in Equation (4.13).

- for Equation (4.17)

Since

$$\boldsymbol{\beta}_{IC}|\boldsymbol{Y}_{1,\cdots}, g, \sigma^2, \tau \sim Norm(\boldsymbol{\mu}_{\tau^-}, \boldsymbol{\Sigma}_{\tau^-}),$$

it has

$$L(\boldsymbol{Y}_{1,\cdots,\tau}|g, \sigma^2) \propto (\sigma^2)^{-\frac{m\tau}{2}}(\frac{1}{1+g\tau})^{\frac{p}{2}}\exp(-\frac{\sum_{i=1}^{\tau}||\boldsymbol{Y}_i||^2}{2\sigma^2} + \frac{g}{1+g\tau}\frac{(\boldsymbol{B}^T\sum_{i=1}^{\tau}\boldsymbol{Y}_i)^T(\boldsymbol{B}^T\boldsymbol{B})^{-1}(\boldsymbol{B}^T\sum_{i=1}^{\tau}\boldsymbol{Y}_i)}{2\sigma^2}).$$

Then

$$\beta_{OC}|\boldsymbol{Y}_{1,\cdots,\tau},g,\boldsymbol{\gamma}\sim Norm(\begin{pmatrix}\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}}\\ \mathbf{0}\end{pmatrix},\sigma^2\begin{bmatrix}(\frac{1}{g}+N)^{-1}(\boldsymbol{B}^T\boldsymbol{B})_{\bar{\gamma}}^{-1} & \mathbf{0}\\ \mathbf{0} & g(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}\end{bmatrix}),$$

so

$$\tilde{\boldsymbol{\beta}}_{OC}|\boldsymbol{D},g,\tau,\boldsymbol{\gamma}\sim Norm(\boldsymbol{\mu}_{\tau^+},\sigma^2\boldsymbol{\Sigma}_{\tau^+}),$$

and then

$$\begin{aligned}L(\boldsymbol{Y}_{\tau+1,\cdots,N}|g,\tau,\boldsymbol{\gamma},\sigma^2,\boldsymbol{Y}_{1,\cdots,\tau})\propto(\sigma^2)^{-\frac{m(N-\tau)}{2}}&\frac{\left|\begin{bmatrix}(\frac{1}{g}+N)^{-1}(\boldsymbol{B}^T\boldsymbol{B})_{\bar{\gamma}}^{-1} & \mathbf{0}\\ \mathbf{0} & g\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}\end{bmatrix}\right|^{-\frac{1}{2}}}{|\boldsymbol{\Sigma}_{\tau^+}|^{-\frac{1}{2}}}\\ &\times\exp(-\frac{\sum_{i=\tau+1}^N||\boldsymbol{Y}_i||^2+(\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}})^T(\frac{1}{g}+N)(\boldsymbol{B}^T\boldsymbol{B})_{\bar{\gamma}}\boldsymbol{\mu}_{\tau^-}^{\bar{\gamma}}}{\sigma^2}\\ &-\frac{\boldsymbol{\mu}_{\tau^+}^T(\boldsymbol{\Sigma}_{\tau^+})^{-1}\boldsymbol{\mu}_{\tau^+}}{\sigma^2}).\end{aligned}$$

Therefore, from

$$L(\boldsymbol{D}|g,\tau,\boldsymbol{\gamma},\sigma^2)=L(\boldsymbol{Y}_{1,\cdots,\tau}|g,\sigma^2)L(\boldsymbol{Y}_{\tau+1,\cdots,N}|g,\tau,\boldsymbol{\gamma},\sigma^2,\boldsymbol{Y}_{1,\cdots,\tau}),$$

Equation (4.13) can be derived by using the conjugacy of $\pi_\sigma^2(\sigma^2)$ again.

E.2 Derivation of Equation (4.26)

Given $\boldsymbol{\beta_{IC}}$, then we have

$$\begin{aligned}\boldsymbol{\beta}_{OC}^{\boldsymbol{\gamma}}|g,\boldsymbol{Y}_{\tau+1,\cdots,N},\sigma^2\sim Norm((N-\tau+\frac{1}{g})^{-1}(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1})\sum_{\tau+1}^N(\boldsymbol{Y}_i-\boldsymbol{B}_{\bar{\gamma}}\boldsymbol{\beta}_{IC}^{\bar{\gamma}}),\\ \sigma^2(N-\tau+\frac{1}{g})^{-1}(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1})\end{aligned}$$

So then we have

$$L(\boldsymbol{D}|g,\tau,\boldsymbol{\gamma},\sigma^2,\boldsymbol{\beta}_{IC}) \propto (g(N-\tau)+1)^{-\frac{\sum_{j=1}^{p}(I(\gamma(j)=1))}{2}}(\sigma^2)^{-\frac{mN}{2}}$$
$$\times \exp(-\frac{\sum_{i=1}^{\tau}||\boldsymbol{Y}_i-\boldsymbol{B}\boldsymbol{\beta}_{IC}||^2}{2\sigma^2}-\frac{\sum_{i=\tau+1}^{N}||\boldsymbol{Y}_i-\boldsymbol{B}_{\tilde{\gamma}}\boldsymbol{\beta}_{IC}^{\tilde{\gamma}}||^2}{2\sigma^2}$$
$$+\frac{g}{1+g(N-\tau)}\frac{\boldsymbol{\mu}_{\beta_{IC}^{\tilde{\gamma}}}^T\boldsymbol{\Sigma}_{\beta_{IC}^{\tilde{\gamma}}}^{-1}\boldsymbol{\mu}_{\beta_{IC}^{\tilde{\gamma}}}}{2\sigma^2})$$

where $\boldsymbol{\mu}_{\beta_{IC}^{\tilde{\gamma}}}$ and $\boldsymbol{\Sigma}_{\beta_{IC}^{\tilde{\gamma}}}$ are the a posteriori mean and variance of $\boldsymbol{\beta}_{IC}^{\tilde{\gamma}}$, and

$$\boldsymbol{\mu}_{\beta_{IC}^{\tilde{\gamma}}} = \frac{g}{1+g(N-\tau)}(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}\sum_{i=\tau+1}^{N}(\boldsymbol{Y}_i-\boldsymbol{B}_{\tilde{\gamma}})$$
$$\boldsymbol{\Sigma}_{\beta_{IC}^{\tilde{\gamma}}} = \frac{g}{1+g(N-\tau)}(\boldsymbol{B}_{\boldsymbol{\gamma}}^T\boldsymbol{B}_{\boldsymbol{\gamma}})^{-1}$$

REFERENCES

Apley, D. and Shi, J. (1998). Diagnosis of multiple fixture faults in panel assembly. *Journal of Manufacturing Science and Engineering*, 120(4):793–801.

Apley, D. W. (2012). Posterior distribution charts: a bayesian approach for graphically exploring a process mean. *Technometrics*, 54(3):279–293.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188.

Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics—Theory and Methods*, 35(4):743–757.

Bapat, R. and Sunder, V. (1985). On majorization and schur products. *Linear algebra and its applications*, 72:107–117.

Barton, R. R. and Gonzalez-Barreto, D. R. (1996). Process-oriented basis representations for mltltivariate process diagnostics. *Quality Engineering*, 9(1):107–118.

Bastani, K., Kong, Z., Huang, W., Huo, X., and Zhou, Y. (2013). Fault diagnosis using an enhanced relevance vector machine (rvm) for partially diagnosable multistation assembly processes. *IEEE Transactions on Automation Science and Engineering*, 10(1):124–136.

Brooks, S., Giudici, P., and Philippe, A. (2003a). Nonparametric convergence assessment for mcmc model selection. *Journal of Computational and Graphical Statistics*, 12(1):1–22.

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003b). Efficient construction of reversible jump markov chain monte carlo proposal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):3–39.

Calabrese, J. M. (1995). Bayesian process control for attributes. *Management science*, 41(4):637–645.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):2–7.

Castagliola, P. (2005). A new s2-ewma control chart for monitoring the process variance. *Quality and Reliability Engineering International*, 21(8):781–794.

Castagliola, P., Celano, G., Fichera, S., and Nenes, G. (2013). The variable sample size t control chart for monitoring short production runs. *The International Journal of Advanced Manufacturing Technology*, 66(9-12):1353–1366.

Castagliola, P. and Maravelakis, P. E. (2011). A cusum control chart for monitoring the variance when parameters are estimated. *Journal of Statistical Planning and Inference*, 141(4):1463–1478.

Chatzis, S. P., Kosmopoulos, D. I., and Varvarigou, T. A. (2009). Robust sequential data modeling using an outlier tolerant hidden markov model. *IEEE transactions on pattern analysis and machine intelligence*, 31(9):1657–1669.

Chopin, N. (2007). Dynamic detection of change points in long time series. *Annals of the Institute of Statistical Mathematics*, 59(2):349–366.

Chopin, N. and Pelgrin, F. (2004). Bayesian inference and state number determination for hidden markov models: an application to the information content of the yield curve about inflation. *Journal of Econometrics*, 123(2):327–344.

Colosimo, B. M. and Del Castillo, E. (2006). *Bayesian process monitoring, control and optimization*. CRC Press.

Cooper, D., Verlander, N., Smith, G., Charlett, A., Gerard, E., Willocks, L., and O'brien, S. (2006). Can syndromic surveillance data detect local outbreaks of communicable disease? a model using a historical cryptosporidiosis outbreak. *Epidemiology & Infection*, 134(1):13–20.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press.

Fearnhead, P. and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):589–605.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Green, P. J. and Hastie, D. I. (2009). Reversible jump mcmc. *Genetics*, 155(3):1391–1403.

Hawkins, D. M. (1987). Self-starting cusum charts for location and scale. *The Statistician*, pages 299–316.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of quality technology*, 35(4):355.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017). Computationally efficient change-point detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26(1):134–143.

Ingleby, N. B. and Lorenc, A. C. (1993). Bayesian quality control using multivariate normal distributions. *Quarterly Journal of the Royal Meteorological Society*, 119(513):1195–1225.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, 38(4):349.

Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H., and Champ, C. W. (2014). An overview of phase i analysis for process improvement and monitoring. *Journal of Quality Technology*, 46(3):265–280.

Kang, L. and Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of quality Technology*, 32(4):418.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2008). Phase i monitoring of polynomial profiles. *Communications in Statistics—Theory and Methods*, 37(10):1671–1686.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology*, 35(3):317.

Li, H.-B., Huang, T.-Z., Shen, S.-Q., and Li, H. (2007). Lower bounds for the minimum eigenvalue of hadamard product of an m-matrix and its inverse. *Linear Algebra and its applications*, 420(1):235–247.

Li, J., Jin, J., and Shi, J. (2008). Causation-based tˆ sup 2ˆ decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology*, 40(1):46.

Li, S. and Chen, Y. (2016). A bayesian variable selection method for joint diagnosis of manufacturing process and sensor faults. *IIE Transactions*, 48(4):313–323.

Li, Z., Zhang, J., and Wang, Z. (2010). Self-starting control chart for simultaneously monitoring process mean and variance. *International Journal of Production Research*, 48(15):4537–4553.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Mahmoud, M. A. (2008). Phase i analysis of multiple linear regression profiles. *Communications in Statistics—Simulation and Computation®*, 37(10):2106–2130.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007). A change point method for linear profile data. *Quality and Reliability Engineering International*, 23(2):247–268.

Montgomery, D. C. (2009). *Statistical quality control*, volume 7. Wiley New York.

Parker, P. A., Morton, M., Draper, N., and Line, W. (2001). A single-vector force calibration method featuring the modern design of experiments. In *Proceedings of the American institute of aeronautics and astronautics 39th aerospace sciences meeting & exhibit*, Reno, Nevada.

Quesenberry, C. (1991a). Spc q charts for a binomial parameter p: short or long runs. *Journal of Quality Technology*, 23(3):239–246.

Quesenberry, C. P. (1991b). Spc q charts for start-up processes and short or long runs. *Journal of quality technology*, 23(3):213–224.

Quesenberry, C. P. (1995a). On properties of binomial q charts for attributes. *Journal of Quality Technology*, 27(3):204–213.

Quesenberry, C. P. (1995b). On properties of poisson q charts for attributes. *Journal of quality technology*, 27(4):293–303.

Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16.

Rao, C. R. and Kleffe, J. (1988). *Estimation of variance components and applications*, volume 3. North Holland.

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York.

Rocke, D. M. (1989). Robust control charts. *Technometrics*, 31(2):173–184.

Runger, G. C., Barton, R. R., Del Castillo, E., and Woodall, W. H. (2007). Optimal monitoring of multivariate data for fault patterns. *Journal of Quality Technology*, 39(2):159.

Shen, X., Tsui, K.-L., Woodall, W. H., and Zou, C. (2015). Self-starting monitoring scheme for poisson count data with varying population sizes. *Technometrics*, 58(just-accepted):1–33.

Shi, J. and Ceglarek, D. (1996). Fixture failure diagnosis for autobody assembly using pattern recognition. *Journal of Engineering for Industry*, 118:55.

Steiner, S. H., Variyath, A. M., et al. (2009). A multivariate robust control chart for individual observations. *Journal of Quality Technology*, 41(3):259.

Steward, R. M., Rigdon, S. E., and Pan, R. (2016). A bayesian approach to diagnostics for multivariate control charts. *Journal of Quality Technology*, 48(4):303.

Tillé, Y. (2011). *Sampling algorithms*. Springer.

Tsiamyrtzis, P. and Hawkins, D. M. (2005). A bayesian scheme to detect changes in the mean of a short-run process. *Technometrics*, 47(4):446–456.

Tsiamyrtzis, P. and Hawkins, D. M. (2008). A bayesian ewma method to detect jumps at the start-up phase of a process. *Quality and Reliability Engineering International*, 24(6):721–735.

Wang, C.-H. and Kuo, W. (2007). Identification of control chart patterns using wavelet filtering and robust fuzzy clustering. *Journal of Intelligent Manufacturing*, 18(3):343–350.

Woodall, W. H. (2007). Current research on profile monitoring. *Production*, 17(3):420–425.

Woodall, W. H., Mohammed, M. A., Lucas, J. M., Watkins, R., et al. (2006). The use of control charts in health-care and public-health surveillance/discussion/discussion/discussion/discussion/discussion/discussion/rejoinder. *Journal of Quality Technology*, 38(2):89.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, 36(3):309.

Wu*, Z. and Tian, Y. (2005). Weighted-loss-function cusum chart for monitoring mean and variance of a production process. *International Journal of Production Research*, 43(14):3027–3044.

Xu, Z., Kass-Hout, T., Anderson-Smits, C., and Gray, G. (2015). Signal detection using change point analysis in postmarket surveillance. *pharmacoepidemiology and drug safety*, 24(6):663–668.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243.

Zeugner, S. (2011). Bayesian model averaging with bms. *Tutorial to the R-package BMS 1e30*.

Zhang, H. and Albin, S. (2009). Detecting outliers in complex profiles using a $\chi 2$ control chart method. *IIE Transactions*, 41(4):335–345.

Zhang, N. F. (1998). A statistical control chart for stationary process data. *Technometrics*, 40(1):24–38.

Zhou, S. and Chen, Y. (2005). A comparison of process variation estimators for in-process dimensional measurements and control. *Journal of dynamic systems, measurement, and control*, 127:69.

Zhou, S., Chen, Y., and Shi, J. (2004). Statistical estimation and testing for variation root-cause identification of multistage manufacturing processes. *IEEE Transactions on Automation Science and Engineering*, 1(1):73–83.

Zhou, S., Ding, Y., Chen, Y., and Shi, J. (2003). Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics*, 45(4):312–325.

Zhu, J. and Lin, D. K. (2009). Monitoring the slopes of linear profiles. *Quality Engineering*, 22(1):1–12.

Zou, C., Jiang, W., and Tsung, F. (2011). A lasso-based diagnostic framework for multivariate statistical process control. *Technometrics*, 53(3):297–309.

Zou, C., Tsung, F., and Wang, Z. (2007a). Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics*, 49(4):395–408.

Zou, C., Zhou, C., Wang, Z., and Tsung, F. (2007b). A self-starting control chart for linear profiles. *Journal of Quality Technology*, 39(4):364.