SELF-STARTING METHODS IN BAYESIAN STATISTICAL PROCESS CONTROL & MONITORING

Konstantinos Bourazas

A thesis submitted in fulfilment
of the requirements for the degree
of Doctor of Philosophy

Athens University of Economics and Business
Department of Statistics

August, 2021

Copyright

Supervisor

Dr Panagiotis Tsiamyrtzis Signature:_____

Thesis committee

Dr Nikolaos Demiris Signature:_____

Dr Ioannis Ntzoufras Signature:_____

External examination committee

Dr Giovanna Capizzi Signature:_____

Dr Subha Chakraborti Signature:_____

Dr Bianca Maria Colosimo Signature:_____

Dr Stylianos Psarakis Signature:_____

Acknowledgements

I would like to thank from the heart my supervisor Panagiotis Tsiamyrtzis for invaluable help and inspiration, Frederic Sobas for essential knowledge transfer and contribution and finally my parents and my intimates for precious support and advice. I would like also to thank Dr. Demiris and Dr. Ntzoufras for serving as members in my PhD thesis committee over the years. Warm thanks to all members of the external examination committee (Dr. Capizzi, Dr Chakraborti, Dr Colosimo and Dr Psarakis) for reviewing this work and for their feedback. Furthermore, I would like to thank the international company Instrumentation Laboratory (IL), Bedford, Massachusetts, USA, for the partial funding of this research.

Abstract

In this dissertation, the center of attention is in the research area of Bayesian Statistical Process Control and Monitoring (SPC/M) with emphasis in developing self-starting methods for short horizon data. The aim is in detecting a process disorder as soon as it occurs, controlling the false alarm rate, and providing reliable posterior inference for the unknown parameters. Initially, we will present two general classes of methods for detecting parameter shifts for data that belong to the regular exponential family. The first, named Predictive Control Chart (PCC), focuses in transient shifts (outliers) and the second, named Predictive Ratio CUSUM (PRC), in permanent shifts. In addition, we present an online change point scheme available for both univariate or multivariate data, named Self-starting Shiryaev (3S). It is a generalization of the well-known Shiryaev's procedure, which will utilize the cumulative posterior probability that a change point has been occurred. An extensive simulation study along with a sensitivity analysis evaluate the performance of the proposed methods and compare them against standard alternatives. Technical details, algorithms and general guidelines for all methods are provided to assist in their implementation, while applications to real data illustrate them in practice.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A brief description

Statistical Process Control/Monitoring (SPC/M) is an effective area of Statistics that is applied in a plethora of disciplines, like: industrial processes, medical laboratories, economics, image analysis, geophysics etc.. It includes all those methods that deal with the quick and valid detection of any disorder in an ongoing process. More specifically, its main aim is to detect when a process deteriorates from its In Control (IC) state, where only natural causes of variation are observed, to the Out Of Control (OOC) state, where exogenous to the process variation is present (Deming, 1986). In essence, the OOC state represents the situation where a change is experienced in at least one IC process setting. The implementation of SPC/M methods is primarily (but not exclusively) performed via control charts, which are statistical tools used monitoring a process and examining whether it runs under statistical stability (IC state) or an assignable cause of variation is present (OOC state). Typically, a control chart is a time series representation of the data (or of a functional form of them) and performs sequential decision making, where we judge a point as "alarm" (indicating process transition to the OOC state), when it plots beyong the control limits.

The concept of control charts initiated by Shewhart in 1926 at Bell Labs (Shewhart, 1926) and over the years numerous types of control charts have been developed for different types of data and different OOC scenarios. Most of them are within the frequentist based approach, with the most representative being $\bar{x}, R, s, c, p$ etc. (also known as Shewhart-type charts), along with CUSUM (Page, 1954) and EWMA (Roberts, 1959). Montgomery (2020) and Oakland (2019) presented the most important frequentist processes, while Hawkins and Olwell (1998) provided an analytical review of CUSUM charts. In the nonparametric field, Qiu and Li (2011) proposed several types of distribution free CUSUM charts, while Chakraborti and Graham (2019) presented an overview of the literature on nonparametric control charts for one dimensional data. From the Bayesian point of view, the online change point model proposed by Shiryaev (1963) and its modification (Roberts, 1966) are the most dominant univariate procedures in the area. A significant amount of Bayesian procedures have been presented in Colosimo and Del Castillo (2006), providing a summary of the Bayesian subregion. Finally, Qiu (2014), Kenett and Zacks, S. (2021) and Tartakovsky et al. (2014) presented methods of more than one approach, i.e. frequentist, nonparametric or Bayesian.

In spite of the fact that univariate procedures are well established, in many cases in SPC/M, we are interested in testing and monitoring simultaneously more than one variables or quality characteristics. In these cases, the application of univariate control charts to each recorded variable is known to be suboptimal, even when these variables are independent. Firstly, when the number of recorded variables is large (something that happens more and more often in the big data era) handling a vast amount of univariate charts is cumbersome and very tricky, especially in handling issues like the the overall false alarm rate (i.e. Type I error) in the decision making. Secondly and most importantly, testing these variables independently will be misleading, as associations between the variables are not taken into account. For example an alarm in a marginal distribution monitoring, does not necessarily imply an

alarm in the joint distribution and the inverse, i.e. an alarm in the joint distribution monitoring does not imply an alarm to some of the marginal distribution control charts. Interestingly enough the above is valid even when the recorded variables are independent. Furthermore, it should be noted that, apart from testing, attributing the alarm to a specific process characteristic is of major importance in multivariate settings.

Hotelling's (1947) T^2 based control chart is the most prominent representative of multivariate control charts. Multivarite generalizations of CUSUM and EWMA were also proposed by Crosier (1988) Lowry et al. (1992) respectively. Mason and Young (2002) and Bersimis et al. (2007) provide a review of multivariate control chart methods. From a nonparametric point of view, Conover et al. (2019) presented the multivariate sequential Normal scores, while several multivariate methods exist in the books of Lauro et. al (2012) and Qiu (2014). The Bayesian approach in Multivariate Statistical Process Control and Monitoring (MSPC/M) is rather restricted. Triantafyllopoulos (2006) developed a control chart based on the sequential Bayes Factors for monitoring multivariate autocorrelated processes. Makis (2008, 2009) formulated a multivariate Bayesian scheme in the optimal stopping framework, for monitoring the mean vector in infinite and finite processes. Furthermore, Zou et al. (2011) proposed a practical LASSO-based diagnostic procedure of the responsible factors for deterioration in high-dimensional procedures and Feltz and Shiau (2001) implemented empirical Bayes process monitoring techniques.

1.2 Standard setup and issues

The majority of the proposed SPC/M methods require two phases (I/II). Phase I is the training phase, where independent IC data are gathered and the goal is to perform calibration of the monitoring scheme, i.e. to derive reliable estimates of the distribution and the unknown parameter(s). Phase II follows and it is the

testing phase, where new observations are collected and compared against the IC standards that established in phase I. Thus, phase I plays a crucial role, as the phase II effective performance will heavily depend on the successful analysis of phase I. Phase I is typically retrospective (offline), while phase II is a prospective (online) or monitoring analysis. Chakraborti et al. (2008) provided a detailed review on the retrospective control charts for univariate variables in phase I along with the appropriate false alarms metrics, while Woodall and Montgomery (1999) presented a plethora of methods in SPC/M area, discussing about open problems and issues in this area.

It is well documented that the phase I/II separation has certain restrictions, which are mainly related with the assumptions and operations while in phase I. Jones-Farmer et al. (2014), presented a detailed overview of phase I methods, exploring the major issues and developments in this domain. For more recent development on the implementation of the phase I analysis and its effect to phase II performance, one can refer Atalay et al. (2020) and Dasdemir et al. (2016) and the references therein. The major issue in phase I analysis is that a large amount of independent IC samples is needed in order to provide reliable estimates of the unknown parameter(s). This is a serious constrain for short runsand for processes that require online decision making from the start of the process (like in the monitoring of medical type data). However, even if the assumption of a large enough initial sample is met, the estimation error for the parameter(s) of interest is typically not taken into account. In most cases, only the point estimates are used and this negatively affects the phase II performance, by increasing the actual false alarm rate. Regarding the phase I sample size requirements in eliminating the effect of estimation error one can refer to Jensen et al. (2006), Zhang et al. (2013, 2014) and Lee et al. (2013) along with the references therein.

Apart from the estimation error, another consequence of the one-off plugged in estimates is that IC information which is available from phase II data is wasted. This issue is of primary importance with two extensions: a practical, as we can improve

the performance of a chart using more information, and a philosophical, as we do not utilize the available information from all the data. Another important issue in phase I is the assumptions' violation. Phase I data are considered to be "clean" data from the stable state of the process. However, a shift in a parameter may occur during phase I, regardless of how well designed or carefully employed a process is. Undetected violations jeopardize the performance in both phases I and II. It is worth noting that this risk increases as the size of the sample under study increases. Finally, retrospective analysis in phase I implies that a disorder might be detected long after its occurrence which is troublesome, especially in processes that online decision making is needed.

1.3 Self-starting control charts

Self-starting methods in SPC/M have been proposed to mitigate the problems arising from phase I/II separation. The developed methodology not only relaxes the necessity of a sufficiently large "clean" dataset in order to provide reliable estimates for the unknown parameter(s), but also provides testing from the early start of a process without any preliminary calibration. Essentially, in a parametric setting with a known distribution, the estimation of the unknown parameter(s) is performed while we test of whether the process is under a stable (IC) state. Hawkins (1987) introduced the term "self-starting" to describe CUSUM schemes for detecting persistent shifts in location or scale parameters of Normal data, without the need of a phase I exercise. Since then, numerous of self-starting control charts have been developed and they are widely used, as their setup offers a framework that is very attractive in real world problems.

Nevertheless, there does not seem to be a concrete definition in the literature of what can be called "self-starting" and what not. As a result many methods self-identify as self-starting, but they are not in practice, as they require the existence of a

preliminary phase or a reference sample for the parameters' estimates. Consequently, we will provide Definition 1.3.1, aiming to establish the basic principles that a control chart needs to have in order to be characterized as "self-starting" and we will follow these principles over the development of this dissertation.

Definition 1.3.1. A control chart will be called as self-starting if:

- it can provide testing, without the need of a preliminary training phase,
- it is online, i.e. to be able to raise an alarm the moment it happens, not retrospectively and
- the IC and the OOC states contain at least one unknown parameter.

From now on in this dissertation, we will characterize a method as self-starting based on Definition 1.3.1.

1.4 Categorization of self-starting methods

The collection and recording of existing methods in a research area is of great significance. On one hand, we discover the already developed methods, avoiding to "reinventing the wheel" and on the other hand we learn about the subareas or unsolved problems, which need further investigation. Castillo et al. (1996) reviewed and commented on the control methods for short production runs, while, more recently, Marques et al. (2015) classified the short run methods and provided a decision model for the choice of the most appropriate method regarding the ongoing process. In a similar spirit, we will start this thesis by providing an extensive literature review for self-starting methods. Trying to list the self-starting methods, a rough categorization could be based on three pillars:

- Dimension: univariate or multivariate,

- Type of shift: transient or permanent,
- Type of approach: frequentist, Bayesian or nonparametric.

A chart is referred as univariate if only a single variable (feature) is involved, while a multivariate chart corresponds to a procedure that controls more than one variable (feature) simultaneously. Concerning the types of a shift, they represent standard causal variations in SPC/M, i.e. the OOC states most often considered in practice. Consequently, the majority of SPC/M methods are designed to efficiently detect them. A transient shift is typically of large size and corresponds to an outlying observation, that is to say an isolated unusual value. An outlier may indicate a sample peculiarity, a data entry error or another impermanent problem. On the contrary, permanent shifts are usually of small or medium size and they are systematic changes to at least one parameter of a procedure. They can by of various types, such as step changes, scale shifts, linear trends, rotations etc.. In Figure 1.4.1, we provide a paradigm of an outlier and location shift in univariate and bivariate sequences.

Concerning the type of approach, the methods, referred as frequentist or Bayesian, are parametric methods via the corresponding school of thought, while for the nonparametric methods, there is no distributional assumption. In the next subsections, we will provide an analytical review of the main methods that are or can be considered as self-starting, classifying them to the above categories. It is important to highlight that some methods may belong to more than one category concerning the type of shift, as they are capable to detect both transient and permanent shifts. Figure 1.4.2 depicts the categorization graphically, as a treemap.

1.4.1 Univariate self-starting methods

We start the literature review with the frequentist univariate schemes and specifically with those that are focused on detecting transient shifts (outliers). We could not start with anything other than Q statistics, as they dominate the region. Quesen-

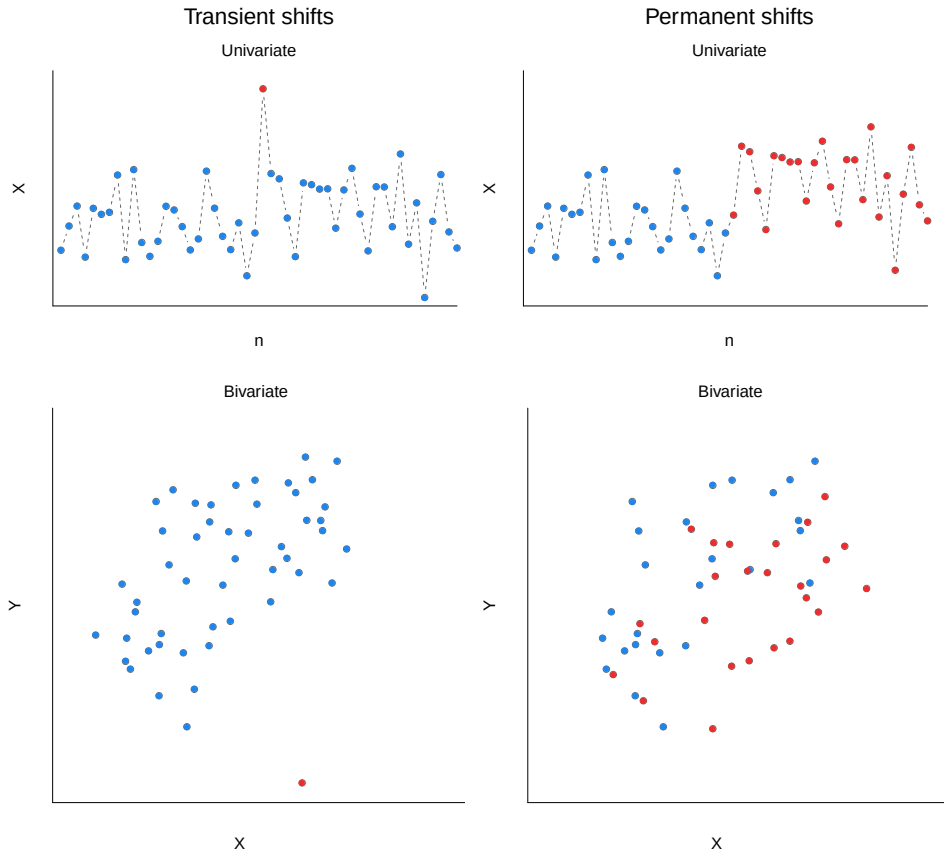


Figure 1.4.1: An illustration of a transient and a permanent shift in univariate and bivariate data. The IC and the OOC data are in blue and red respectively.

berry (1991a, 1991b, 1991c, 1995d) introduced the Q charts for detecting transient shifts in shorts runs for Normal, Poisson, Binomial and Geometric data respectively. The Q chart procedure is based on the sequentially updated Q statistic, that does not require a preliminary phase. Assuming normality, Castillo and Montgomery (1994) presented modifications that enhance the detection properties and He et al. (2008) proposed two schemes to alleviate bias issues in Q charts. Completing the review of Q-based charts, Ravichandran (2019) identified his proposal as self-starting, because updated Q statistic is used, although an IC sample for the initial estimates is required. Methods with many similarities with the aforementioned, are the t chart (Gu et al., 2014) and the methodology of Korzenowski et al. (2015), both implemented to monitor the process mean in multi-variety and small batch production runs. In the

same context, a variable sample size t control chart for monitoring short production runs was proposed by Castagliola et al. (2013). For non-normal data, Zhang et al. (2017) constructed a Shewhart-type control chart, named *Beta* chart for monitoring the Weibull shape parameter. Recently, Dogu and Noor-ul-Amin (2021) proposed a self-starting chart for detecting anomalies in exponentially distributed time between events (TBE) data. It should be noted that in the latter method, the proposed statistic can be used in an EWMA or a CUSUM in detecting persistent shifts, a type of process disturbance that we will review next.

Retaining the focus on the frequentist approach, we continue with the permanent shifts. Worsley (1979, 1986) pioneering work, introduced the ratio test statistic in modeling a mean change point. Similarly, the likelihood ratio and the change point formulation play the role of the "frame of reference" for the models of Gombay (2003), Mei (2006) and Dessein and Cont (2013). Hawkins's work has a prominent place in the sequential change point detection. He introduced numerous methods, related to the detection of shifts in the mean or/and the variance of a process, like Hawkins (1977), Hawkins et al. (2003), Hawkins and Zamba (2005a, 2005b). Special note should be made for the self-starting CUSUM (SSC) for location and scale (Hawkins, 1987) and the SSCs using the Q statistics (Hawkins and Olwell, 1998) that are widely used. In relation to the CUSUM-type charts, Atwi et al. (2011) proposed a CUSUM-based method for hidden Markov models, while Tercero-Gómez et al. (2014) developed a SSC for detecting changes in the process mean, combined with a maximum likelihood estimation for the change point. Furthermore, Siegmund and Venkatraman (1995) suggested a CUSUM-type change point model, which is based on a generalized likelihood ratio statistic for detecting a change in a normal mean with known variance.

Next, we will review various hybrid control chart suggestions that typically combining more than one procedures, enhancing the power in detecting of permanent shifts. The common element of these "alloys" of methods in this paragraph is an EWMA,

which is adopted by all of them. Specifically, Li et al. (2010) proposed a self-starting control chart, based on the likelihood ratio test (LRT) and the EWMA for simultaneous monitoring of the mean and the variance. In addition, Li and Wand (2010) and Li et al. (2010) proposed a hybrid chart focused on mean shifts, which combines a SSC using the Q statistic with an adaptive EWMA scheme for the magnitude of the shift. Again based on Q statistic, Roes et al. (1999) developed the alternative $Q(R)$ chart, estimating the standard deviation via the mean moving range, while modelling the cost of low-volume processes. An EWMA chart for t statistic is demonstrated in different versions for short runs by Chang and Sun (2016), Wang et al. (2020) and Song et al. (2020). Of particular interest is the bootstrap based monitoring scheme for Poisson count data with varying population sizes (Shen et al., 2016). Furthermore, Castillo and Montgomery (1995) suggested a Kalman Filter process control scheme for short runs, which essentially acts like an EWMA. In closing, it is noteworthy that of the outlier detection methods, the Geometric Q chart (Quesenberry, 1995d) and the monitoring scheme of Korzenowski et al. (2015) can be applied for permanent shifts, along with the TBE model (Dogu and Noor-ul-Amin 2021) after certain modifications, as we have already mentioned.

From the Bayesian perspective, several methods have been proposed for identifying isolated shifts. The common denominator in all of them is the use of the predictive distribution, which represents the conditional likelihood of the future observable(s) given the available data (i.e. all the unknown parameters have been integrated out). The most general methodology in this category of models is the Predictive Control Chart (PCC) by Bourazas et al. (2021), which is partly presented in this dissertation. PCC offers a unifying, closed form mechanism, that is capable to monitor for transient shifts, data from any (ontinuous or discrete) distribution as long as it is a member of the regular exponential family. In the same framework, the methods of Ali and Riaz (2020) and Bayarri and García-Donato (2005) were specialized in monitoring the dispersion of Normal and Poisson count data respectively. Kumar

and Chakraborti (2017) along with Ali (2020) suggested methods for monitoring the exponentially distributed TBE data. The work of Tsiamyrtzis et al. (2015) for a sequence of Normal data was the "harbinger" of PCC, while in similar context, Olwell (1996, 1997) proposed methods to handle Normal and Poisson data respectively.. Additionally, the Cumulative Bayes Factor (CBF) by West (1986) and West and Harrison (1986) are memory based method, but the Bayes Factor (BF) can be an indicator for the presence of an isolated shift as well.

Plenty of methods for investigating permanent shifts are in the Bayesian "reservoir", including the aforementioned CBF and the Normal linear dynamic models also by West and Harrison (1996). More research in the area of sequential model comparison, monitoring and forecasting was performed by Harrison and West (1987, 1991), Harrison and Veerapen (1994) and Harrison (1999). By the same token, Geweke and Amisano (2010) applied the cumulative predictive Bayes Factor for model comparisons in economics. Still in the Bayesian perspective the work of Tsiamyrtzis and Hawkins (2005, 2008, 2010, 2019) with change point models for the mean of Normal or Poisson data, further boosted the Bayesian modelling in short runs. By the same logic, a three-state sequential algorithm was developed, modelling the three states of an epidemic (Zamba et al., 2013).

Change point models have occupied the Bayesian statisticians for many decades. Smith (1975) adopted the change point formulation to provide inference for sequences of Normal or Binomial data. In 1990s, the work of Wasserman and Sudjianto (1993) with a Bayesian second ordered dynamic linear model and of Wasserman (1994, 1995), who developed Bayesian EWMA schemes, were novel for detecting mean shifts in short runs. Later, Crowder and Eshleman (2001) investigated an adaptive filtering approach to monitor low volume autocorrelated data. Howley et al. (2009) proposed a CUSUM-type monitoring of clinical indicators, based on the posterior predictive of Bernoulli trials and in the same spirit Toubia-Stucky et al. (2012) developed a memory based procedure for the proportion of non-conforming items

in a process. We sould highlight in particular the highly cited online change point by Adams and MacKay (2007), who proposed a machine learning algorithm for the estimation of a change point based on posterior distribution of the length of the last IC sequence, after the last disorder. Following a different approach, Lai and Xing (2010) extended the Shiryaev's Bayesian online change point model, while assuming pre/post change unknown parameters and providing corresponding stopping rules. Recently, Noor-ul-Amin and Noor (2021) and Noor et al. (2020) proposed Bayesian self-starting EWMAs for specific distributions in the exponential family. Regarding the previously described methods, a Bayesian change detection model is provided along with a frequentist alternative in Atwi et al. (2011), while an EWMA and a CUSUM modification is proposed for the TBE chart in Ali (2020).

Through the prism of nonparametric statistics, only a handful of the developed self-starting methods are capable to detect isolated jumps, especially when we have short production runs. Their most characteristic representative is the control chart proposed by Alloway and Raghavachari (1991), which is based on the Hodges-Lehmann estimator and the Wilcoxon signed rank statistic. Moreover, Conover et al. (2018) used Sequential Normal Scores (SNS) for developing a Shewhart type chart, along with CUSUM and EWMA type modifications for persistent shifts. We could add to this category the recursive segmentation and permutation (RS/P) method of Capizzi and Masarotto (2013). Although it has been developed for the effective detection of mean and/or scale shifts, it can identify isolated jumps for subgrouped data.

Along with the latter method, which is suitable for persistent shifts, Capizzi and Masarotto (2012) also proposed a hybrid scheme named CUSCORE-type chart, using an EWMA adaptation. Furthermore, a sequential nonparametric test, based on windowed Kolmogorov - Smirnov statistics (Madrid Padilla et al., 2019), have been suggested to handle univariate data, while Gombay (2004) proposed sequential testing strategies based on U statistics and Wiener process approximations. Several methods have adapted a CUSUM-type detection scheme using nonparametric statis-

tics. For instance, the two stage procedure for prospective change-point detection and the retrospective estimation of the detected location of the change, via a modified Kolmogorov-Smirnov test (Brodsky, 2010). In the same regard the independent works of Li (2021) and Hou and Yu (2020), proposing nonparametric CUSUMs for the detection of any type of distributional change.

The sequential ranks have a special place in the nonparametric approach, as they are widely used in change point detection. Apart from the aforementioned work, Conover et al. (2017) analytically discussed about sequential Normal scores. In the same spirit with the latter, the control chart of Villanueva-Guerra et al. (2017) is focused on scale shifts, based on squared ranks. Some of the initial proposals applied the idea of sequential ranks like in Reynolds (1975), where a truncated version of sequential tests with linear barriers was used. Furthermore, Bakir and Reynolds (1979) and Amin and Searcy (1991) used the Wilcoxon signed-rank statistics, while Bhattacharya and Frierson (1981) adopted a partial weighted sum of sequential ranks. McDonald (1990) proposed a CUSUM-type chart for detecting an abrupt change form the sampling distribution to a stochastically dominating one. In more recent research, Liu et al. (2013, 2014, 2015) developed CUSUM-type and EWMA-type self-starting schemes for process monitoring, based on the sequential ranks. Regarding methods that are capable in detecting both location and scale changes simultaneously, Lombard and Van Zyl (2018) suggested the signed rank CUSUMs and Ross et al. (2011) developed a nonparametric model for a stream of random variables.

1.4.2 Multivariate self-starting methods

In this subsection we deal with multivariate processes, which in their vast majority are frequentist based. Starting from the transient shifts, Quesenberry (2001) proposed a snapshot chart, where we plot all the univariate Q statistics for all the measured variables on a particular production unit on one chart. Despite the fact

that it is a valuable attempt to provide a summary chart for high dimensional data, it is still suboptimal, as it ignores possible associations between the variables. Khoo and Quah (2002) and Khoo et al. (2005) took it one step further, as they proposed the extension of Q chart for individual or subgrouped multivariate Normal data. Also, they demonstrated their use for memory based charts. It is important to mention that to the best of our knowledge in the literature, there are neither Bayesian nor nonparamteric self-starting procedure for detecting multivariate outlying observations. Combining it with the fact that there are only a few methods from the frequentist approach, we can easily conclude that this is a virgin area for the self-starting procedures in short runs and not only.

Dealing with the permanent shifts, the most cited paper is the work of Zamba and Hawkins (2006), where a frequentist change point model was proposed. The introduced model is based on an unknown-parameter likelihood ratio test for detecting a step change in the mean vector of multivariate Normal data. Also, the work of Sullivan and Jones (2002) is widely mentioned, where a self-starting multivariate EWMA for the mean vector was demonstrated, using the Q update formulas for the sample mean vector and covariance matrix. Remaining in the EWMA framework, Hawkins and Maboudou-Tchao (2007) suggested the use of recursive residuals in a self-starting chart. The recursive residuals were initially introduced in a CUSUM procedure by Brown et al. (1975), while they were also used by Taleb and Arfa (2012) in a modification of CUSUM and by Capizzi and Masarotto (2010). In the latter, two robust control charts were proposed, a CUSUM-type and an EWMA-type, which were based on a CUSCORE procedure for monitoring the unknown mean of a multivariate Normal distribution. Furthermore, Li et al. (2017) integrated a multivariate spatial rank test with the EWMA charting for monitoring sparse multivariate mean shifts.

The EWMA-type charts are numerous in the literature, but the first reported self-starting multivariate EWMA is by Quesenberry (1997), involving the computation

of a Hotelling T^2 statistic. Also, a special mention should be made of one of the first models for detecting a change in the multivariate normal mean by Srivastava and Worsley (1986). Recently, the work by Yang and Qiu (2021) is also worth mentioning; they propose a variable-sampling control chart, where the intervals are determined by the covariate information. Even though the chart needs initial estimates for the IC parameters, they are recursively updated in the monitoring process. In the same pattern, Zhang et al. (2012) developed a hybrid cumulative count of conforming chart for monitoring high-quality processes. Returning to the self-starting methods that satisfy the criteria of Definition 1.3.1, Maboudou-Tchao and Hawkins (2011) formulated a high-dimensional control chart to monitor changes in both the location and the scale of Normal data. For the mean vector shifts in Normal data, Li at al. (2014) proposed a self-starting chart, providing change point estimate, Zantek et al. (2006) suggested a CUSUM-type for monitoring multistage manufacturing systems. Apropos of the linear profiles, Zou et al. (2007), Amiri et al. (2016) and Xia and Tsung (2019) developed self-starting schemes for detecting changes in the regression parameters, based on the recursive residuals or a Wald-type sequential statistic. Working in the same spirit, Aminnayeri and Sogandi (2016) and Khosravi and Amiri (2019) proposed logistic regression profile control charts to monitor the relationship between a Bernoulli or a Binomial response variable and explanatory variables. Amirkhani et al. (2018) used a CUSUM-type control chart based on the residual values of the accelerated failure time regression model for monitoring the survival times of patients. Further, Pazhayamadom et al. (2013, 2016) demonstrated a harvest control rule for monitoring limited data fisheries, based on a self-starting CUSUM.

Only a limited number of methods can be found in the Bayesian arena; Zeng and Zou (2011) proposed a change point model for the patient outcomes in healthcare. Specifically, they formulated the change detection as model-selection problem and the decision making relies on Bayes Factors. Furthermore, Hou et al. (2020) developed

a self-starting monitoring scheme by fitting a piecewise linear model in monitoring the process mean. Continuing with the nonparametric approach, Paynabar et al. (2016) developed a potentially distribution free change point model for phase I monitoring analysis of multichannel profiles, while Xue and Qiu (2020) suggested a nonparametric CUSUM chart for monitoring multivariate serially correlated processes. Although, the latter method is not self-starting based on Definition 1.3.1, as it needs a batch (even small) of IC data to initiate the chart, before the initial estimates can be sequentially updated during the online process.

1.4.3 Supplementary research

The research on an area is not only concerned with proposing new innovative methods, but also investigating the application and properties of existing methods. Hence, supplementary research papers are added to an area, improving our knowledge about the already proposed methodologies. Quenseberry (1995a, 1995b, 1995c, 2000) explored to greater depth the properties of Q statistics for Normal, Poisson and Binomial data respectively and illustrated geometric Q chart for nosocomial infection surveillance. Furthermore, Keefe et al. (2015) discussed about the IC performance of self-starting charts, emphasizing on self-starting CUSUM and Q chart for Normal data, conditioned on the already observed data. On the same wavelength, Zantek (2005, 2006, 2008) and Zantek and Nestler (2009) investigated the run-length performance of the same type of charts, facilitating the understanding of Q statistics behaviour and the derivation of an appropriate design. Remaining in Q statistics, there are several publications on their application in practice, such as the evaluation of the performance of Q-based charts by Theroux et al. (2014) for short production runs in aerospace manufacturing. In the same framework, Snoussi et al. (2005) and Kawamura et al. (2013) applied Q statistics to the residuals of a time series model for detecting anomalies, while Lampreia et al. (2012) applied a modified version of Q statistics for the vibration monitoring of repairable systems.

For the rest methods of the area, the self-starting CUSUM proposed by Hawkins was tested in real-world applications, as it was used for the assessment of antihypertensive responses by Cornélissen et al. (1997) and for the development of a self-balancing approach for debris cleanup operations using data from an Atlantic hurricane by Fetter and Rakes (2011). Furthermore, Sobas et al. (2014, 2020) discussed about the application of Bayesian techniques to medical lab processes. On a different topic, Celano et al. (2012a, 2012b, 2013) discussed in detail about the performance and the economic design of t control charts in short runs and Lang (2019) provided analytical tables of the control limits for a nonparametric adaptive CUSUM based on sequential ranks. An analytical comparative study of already proposed methods always helps in their assessment under OOC scenarios. In this context, Yu et al. (2020) compared different EWMA approaches for a self-starting forecasting process, while Dogu and Kim (2020) staged a comparison between self-starting methods for individual multivariate observations. Finally, we cannot forget the important work of Pollak and Siegmund (1991), who introduced three stopping rules for a change in the unknown mean of a univariate Normal process and of Baron (2001, 2004), who discussed about the stopping rules for Bayesian sequential change point models.

1.5 Thesis structure

In this thesis, the focus is on the effective online detection of process disorders and the reliable inference for the unknown process parameter(s) of short horizon data, without the requirement of any calibration phase. We will attempt to achieve this goal adopting the Bayesian perspective in developing appropriate self-starting control charts. Namely, we propose a general Bayesian method named the Predictive Control Chart (PCC) focused on the detection of large transient shifts (outliers). In the same philosophy, we propose Predictive Ratio CUSUM (PRC) focusing on detecting medium/small persistent parameter shifts. In addition, we introduce the

Self-Starting Shiryaev (3S), a general online change point model available for univariate (U3S) or multivariate (M3S) processes.

Analytically, in Chapter 2, we provide the PCC initial assumptions and derive the modelling along with the necessary formulas for several discrete and continuous univariate distributions that belong to the regular exponential family. Along with them, we provide guidelines regarding choices of prior distributions and the options that allow the use of possibly available historical data, via a power prior mechanism. Further, we present the possibility of employing a Fast Initial Response (FIR) PCC, which enhances its performance during the early stages of the process. An extended simulation study follows including a comparison against a frequentist based competing method, a sensitivity analysis and a robustness study examining both prior and model type misspecifications. The PCC application to real data wraps this Chapter, where a continuous (Normal) and a discrete (Poisson) real-data case from a medical lab and an industrial setting respectively, are being explored.

The PRC derivation is presented in Chapter 3. Precisely, we provide the PRC design, including the recursive formulas for several univariate discrete and continuous distributions that belong to the regular exponential family, along with the related decision making and a FIR option. A simulation study for detecting persistent parameter shifts in Normal, Poisson and Binomial data follows, where we evaluate the PRC performance against frequentist and Bayesian competitors, while examining a prior sensitivity. At the end of this chapter, we provide a PRC illustration to real data, where a continuous (Normal) and a discrete (Poisson) cases are examined.

Chapter 4 refers to the 3S detection scheme. Specifically, we present the assumptions and the general model structure, which allow its application to any type of data, regardless of the dimension. In addition, we discuss about the posterior inference for the process parameters and about the 3S stopping time, introducing an adaptive decision limit. Apart from the description of the general form, we provide the details and the properties of the submodels of the univariate (U3S) and multivariate (M3S)

of Normal data. The performance of the proposed models is examined via extensive simulations in a comparative study against frequentist and nonparametric methods for both univariate and multivariate data, along with a sensitivity analysis. Three applications to well documented data demonstrate U3S and M3S, where a mean or a variance shift for univariate data and a translocation shift for multivariate data are experienced. Finally, the conclusions of this thesis along with the suggested future research work are presented in Chapter 5, while all the necessary technical details are provided in Appendices.

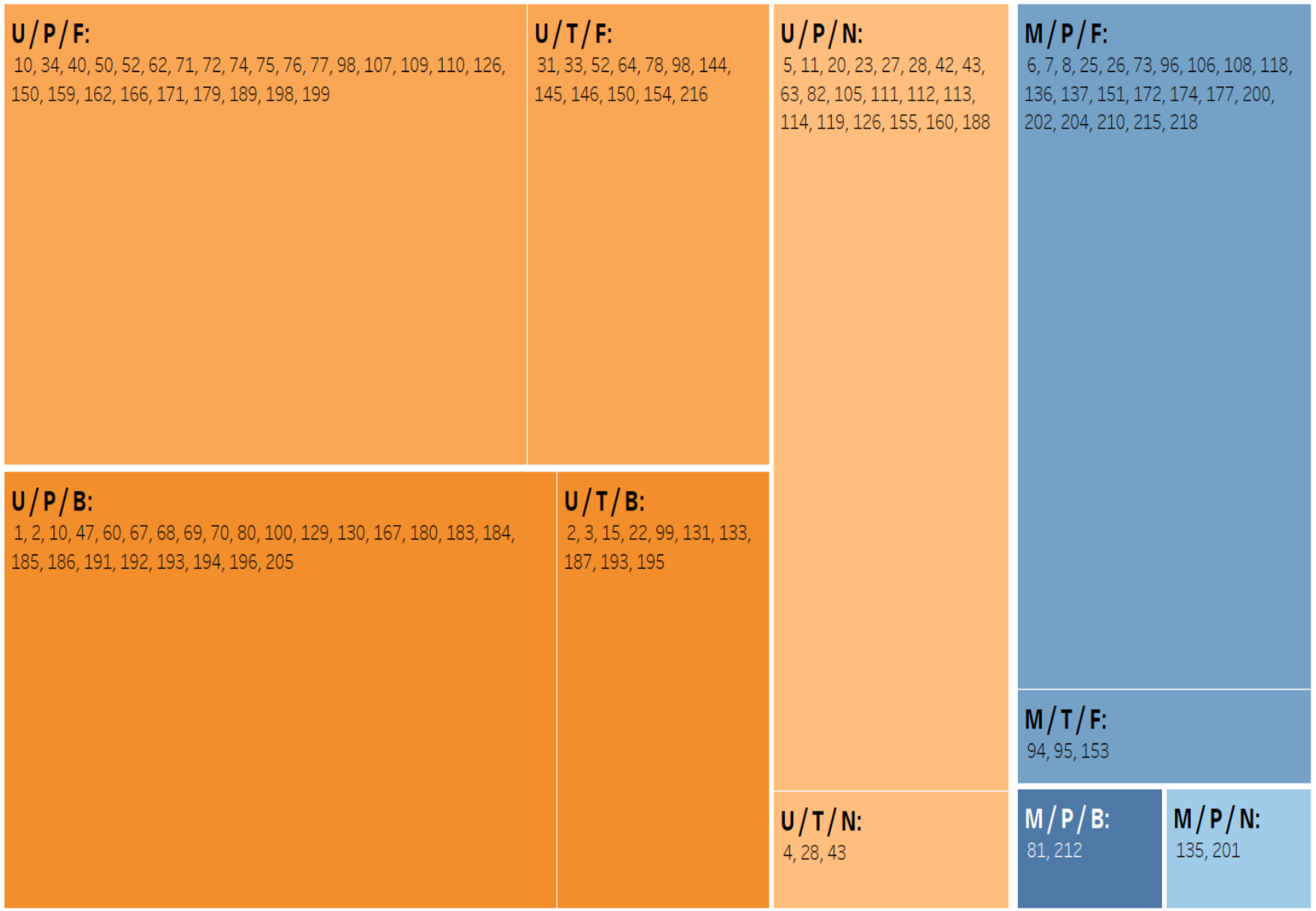


Figure 1.4.2: A treemap which summarizes all the self starting methods based on the dimension of the process data (Univariate in orange, Multivariate in blue), the statistical approach (frequentist, nonparametric or Bayesian), and type of shift which is of interest (Transient or Permanent). The numbers refer to the ordered number of a publication in the references.

Chapter 2

Predictive Control Chart (PCC)

2.1 PCC Theoretical background

Being in the Bayesian framework, our goal is to utilize the available prior information and provide a control chart with enhanced performance compared to existing self-starting frequentist-based methods. The Predictive Control Chart (PCC) were initially proposed in the M.Sc. thesis of Bourazas (2014) and Kiagias (2014) for continuous and discrete cases respectively and published in Bourazas et al. (2021). PCC is formed by the predictive distribution and it will provide a sequentially updated region against which every new observable will be plotted. Observations falling outside the predictive region will ring an alarm triggering further investigation and potentially some form of corrective action. In this chapter, we extend PCC in four ways. By using power priors, by defining rigorously the IC region, by developing a fast initial response (FIR) scheme and by providing two applications to real data.

2.1.1 PCC for k-parameter regular exponential family (k-PREF)

Initially, we need to derive the predictive distribution (Geisser, 1993), which depends on the likelihood of the observed univariate data. From a process under study, we sequentially obtain the data $\boldsymbol{X} = (x_1, \ldots, x_n)$, which we consider to be a random sample from the distribution $X_j|\boldsymbol{\theta}$, where $X_j, j = 1, \ldots, n$, is univariate, while the unknown parameter $\boldsymbol{\theta}$ can be either univariate or multivariate, e.g. $X_j|\theta \sim Bin(N_j, \theta)$, $X_j|\theta \sim P(\theta)$, $X_j|\boldsymbol{\theta} \sim N(\theta_1, \theta_2^2)$ etc. Our main interest is in detecting in an online fashion and without employing a phase I exercise, the presence of large transient shifts on the unknown parameter(s) $\boldsymbol{\theta}$. We assume that the likelihood, is a member of the univariate k-parameter regular exponential family (denoted from this point on as k-PREF), and by following Bernardo and Smith (2000), it can be written as:

$$f(\boldsymbol{X}|\boldsymbol{\theta}) = \left[\prod_{j=1}^{n} g(x_j)\right] [c(\boldsymbol{\theta})]^n \exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \sum_{j=1}^{n} h_i(x_j)\right\}, \tag{2.1.1}$$

where $g(x_j) \geq 0$, $h_1(x_j), \ldots, h_k(x_j)$ are real-valued functions of the univariate observation x_j that do not depend on $\boldsymbol{\theta}$, while $c(\boldsymbol{\theta}) \geq 0$ and $\eta_1(\boldsymbol{\theta}), \ldots, \eta_k(\boldsymbol{\theta})$ are real-valued functions of the unknown parameter(s) $\boldsymbol{\theta}$ that cannot depend on $\boldsymbol{X}$. PCC will be developed for any likelihood that belongs to the k-PREF, providing a general platform where binary (Binomial), count (Poisson, Negative Binomial) or various continuous (Normal, Gamma, Lognormal etc.) univariate data can be analyzed using the same methodology.

The prior distribution is of key importance in the Bayesian approach. Since in practice, historical data (of the same or a similar process, not to be confused with phase I data) are typically available, we recommend the use of power priors (Ibrahim and Chen, 2000), which offer a framework to incorporate past data (when available)

in the mechanism of forming the prior distribution. The power prior is derived by:

$$\pi\left(\boldsymbol{\theta}|\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)\propto f\left(\boldsymbol{Y}|\boldsymbol{\theta}\right)^{\alpha_0}\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}\right), \tag{2.1.2}$$

where $\boldsymbol{Y}=(y_1,\ldots,y_{n_0})$ refers to a vector of historical univariate data (under the same distribution law $f(\cdot|\boldsymbol{\theta})$ that the current data obey), $0\leq\alpha_0\leq 1$ is a scalar parameter, $\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}\right)$ is the initial prior for the unknown parameter(s) and $\boldsymbol{\tau}$ is the vector of the initial prior hyperparameters. The (fixed) parameter, α_0, controls the power prior's tail heaviness and consequently the influence of the historical data on the posterior distribution. Essentially, α_0 represents the probability of the historical data being compatible with the current observations and at the extremes $\alpha_0=0$ or 1, the historical data will be ignored or taken fully into account (just as the current data) respectively. A typical value for α_0 is $1/n_0$, which conveys the weight of a single observation to the prior information. In general, α_0 should be determined by the relevance of past with current data and how likely is the past data to provide reliable estimates for the unknown parameters (depending on the size n_0). For relevant historical data but with small (large) n_0 it is recommended to use $\alpha_0<1/n_0$ ($\alpha_0>1/n_0$). It should be noted that the power priors are robust in conflicts of historical and current data, as they use only the sufficient statistic of the past data.

Generalizing the power prior concept, we could either assume α_0 is unknown (modeled by a prior distribution) or we could allow the use of multiple historical data: if $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are historical data from different sources weighted by α_0 and β_0 respectively, then the power prior is proportional to:

$$\pi\left(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{Z},\alpha_0,\beta_0,\boldsymbol{\tau}\right)\propto f\left(\boldsymbol{Y}|\boldsymbol{\theta}\right)^{\alpha_0}f\left(\boldsymbol{Z}|\boldsymbol{\theta}\right)^{\beta_0}\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}\right). \tag{2.1.3}$$

It is worth mentioning that, Ibrahim et al. (2003), proved that the power prior is 100% efficient in the sense that the ratio of the output to input information is equal to one, with respect to Zellner's information rule (see Zellner, 1988).

In a subjective Bayesian manner, $\pi_0(\cdot)$ should reflect all available information regarding the unknown parameter(s) before the data become available and its form can be derived from prior knowledge, expert's opinion etc. From an objective Bayesian point of view and under the scenarios of lacking any prior knowledge, one can adopt a weakly informative or even non-informative initial prior, such as flat (uniform) prior, Jeffreys (Jeffreys, 1961) or reference (Bernardo, 1979, Berger et al., 2009) prior (see also the discussion regarding prior elicitation in Subsection 2.1.2).

To preserve closed form solutions for all scenarios, when implementing PCC, we will adopt a conjugate prior for $\pi_0(\boldsymbol{\theta}|\boldsymbol{\tau})$, which always exists for any likelihood that is a member of the k-PREF (Bernardo and Smith, 2000) and its form is given by:

$$\pi_0(\boldsymbol{\theta}|\boldsymbol{\tau}) = [K(\boldsymbol{\tau})]^{-1} [c(\boldsymbol{\theta})]^{\tau_0} \exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta})\tau_i\right\}, \tag{2.1.4}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (parameter space) and $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_k)$ is the $(k+1)$-dimensional vector of the initial prior hyperparameters, such that:

$$K(\boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} [c(\boldsymbol{\theta})]^{\tau_0} \exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta})\tau_i\right\} d\boldsymbol{\theta} \quad < \infty. \tag{2.1.5}$$

The conjugate prior, $\pi_0(\boldsymbol{\theta}|\boldsymbol{\tau})$, is also a member of the exponential family. The choice of the hyperparameters $\boldsymbol{\tau}$ will reflect the prior knowledge, ranging from highly informative to vague and even non-informative choices. Non-conjugate choices of the initial prior are allowed, at the cost of not having PCC in closed form but evaluated numerically. Lemma 2.1.1 provides the form of a conjugate power prior.

Lemma 2.1.1. For any vector of historical data $\boldsymbol{Y} = (y_1, \ldots, y_{n_0})$ of the same form with the current data, a conjugate $\pi_0(\boldsymbol{\theta}|\boldsymbol{\tau})$ will lead to a conjugate power prior of the form:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}) \propto \pi_0(\boldsymbol{\theta}|\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y})), \tag{2.1.6}$$

where $\boldsymbol{t}_{n_0}(\boldsymbol{Y}) = \left(n_0, \sum_{l=1}^{n_0} h_1(y_l), \ldots, \sum_{l=1}^{n_0} h_k(y_l) \right)$ is a $(k+1)$-dimensional vector, with $\boldsymbol{Y} = (y_1, \ldots, y_{n_0})$ referring to the vector of historical univariate data.

Proof

For a likelihood $f(\cdot|\boldsymbol{\theta})$, being a member of the k-PREF, the conjugate prior is given by (2.1.4) and the normalizing constant, $K(\boldsymbol{\tau})$ is given by (2.1.5) (for discrete $\boldsymbol{\theta}$, we replace the integral sign by summation). Then for the historical data $\boldsymbol{Y} = (y_1, ..., y_{n_0})$, sampled from the same member of the k-PREF as the likelihood, $f(\cdot|\boldsymbol{\theta})$, the power prior will become:

$$\begin{aligned}
\pi\left(\boldsymbol{\theta}|\boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) &\propto f\left(\boldsymbol{Y}|\boldsymbol{\theta}\right)^{\alpha_0} \pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}\right) \\
&= \left[\prod_{l=1}^{n_0} g(y_l)\right]^{\alpha_0} [c(\boldsymbol{\theta})]^{\alpha_0 n_0} \exp\left\{\alpha_0 \sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \sum_{l=1}^{n_0} h_i(y_l)\right\} \times \\
&\quad \times [K(\boldsymbol{\tau})]^{-1} [c(\boldsymbol{\theta})]^{\tau_0} exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \tau_i\right\} \\
&= [K(\boldsymbol{\tau})]^{-1} \left[\prod_{l=1}^{n_0} g(y_l)\right]^{\alpha_0} [c(\boldsymbol{\theta})]^{\tau_0 + \alpha_0 n_0} \exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \left(\tau_i + \alpha_0 \sum_{l=1}^{n_0} h_i(y_l)\right)\right\} \\
&\propto [c(\boldsymbol{\theta})]^{\tau_0 + \alpha_0 n_0} \exp\left\{\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \left(\tau_i + \alpha_0 \sum_{l=1}^{n_0} h_i(y_l)\right)\right\} \\
&\propto \pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y})\right),
\end{aligned}$$

where $\boldsymbol{t}_{n_0}(\boldsymbol{Y}) = \left(n_0, \sum_{l=1}^{n_0} h_1(y_l), \ldots, \sum_{l=1}^{n_0} h_k(y_l) \right)$ is a $(k+1)$-dimensional vector, with $\boldsymbol{Y} = (y_1, ..., y_{n_0})$ referring to the vector of historical data.

Q.E.D.

Theorem 2.1.1 provides, in closed form, the posterior and predictive distributions of any likelihood that belongs to the k-PREF:

Theorem 2.1.1. For any likelihood belonging to the k-PREF (2.1.1) and an initial conjugate prior (2.1.4) via a power prior (2.1.1) mechanism we have:

(i) The posterior distribution of the unknown parameter(s) $\boldsymbol{\theta}$:

$$p\left(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)=\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}+\alpha_0\boldsymbol{t}_{n_0}(\boldsymbol{Y})+\boldsymbol{t}_n(\boldsymbol{X})\right), \tag{2.1.7}$$

where $\boldsymbol{t}_n(\boldsymbol{X})=\left(n,\sum_{j=1}^{n}h_1(x_j),\ldots,\sum_{j=1}^{n}h_k(x_j)\right)$ is a $(k+1)$-dimensional vector, with $\boldsymbol{X}=(x_1,\ldots,x_n)$ being the observed univariate data.

(ii) The predictive distribution of the single future univariate observable X_{n+1}:

$$f\left(X_{n+1}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)=\frac{K\left(\boldsymbol{\tau}+\alpha_0\boldsymbol{t}_{n_0}(\boldsymbol{Y})+\boldsymbol{t}_n(\boldsymbol{X})+\boldsymbol{t}_1(X_{n+1})\right)}{K\left(\boldsymbol{\tau}+\alpha_0\boldsymbol{t}_{n_0}(\boldsymbol{Y})+\boldsymbol{t}_n(\boldsymbol{X})\right)}g(X_{n+1}), \tag{2.1.8}$$

where $\boldsymbol{t}_1(X_{n+1})=(1,h_1(X_{n+1}),\ldots,h_k(X_{n+1}))$ is a $(k+1)$-dimensional vector, function of the future observable X_{n+1}.

Proof

(i) Once the current data $\boldsymbol{X}=(x_1,...,x_n)$ become available, we will be able to derive the posterior distribution of the unknown parameter(s) $\boldsymbol{\theta}$, using Bayes theorem:

$$\begin{aligned}
p\left(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right) &\propto f\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}|\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)\\
&\propto f\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}+\alpha_0\boldsymbol{t}_{n_0}(\boldsymbol{Y})\right)\\
&= \left[\prod_{j=1}^{n}g(x_j)\right][c(\boldsymbol{\theta})]^n\exp\left\{\sum_{i=1}^{k}\eta_i(\boldsymbol{\theta})\sum_{j=1}^{n}h_i(x_j)\right\}\times\\
&\quad\times\left[K(\boldsymbol{\tau})\right]^{-1}\left[\prod_{l=1}^{n_0}g(y_l)\right]^{\alpha_0}[c(\boldsymbol{\theta})]^{\tau_0+\alpha_0 n_0}\\
&\quad\times\exp\left\{\sum_{i=1}^{k}\eta_i(\boldsymbol{\theta})\left(\tau_i+a_0\sum_{l=1}^{n_0}h_i(y_l)\right)\right\}\\
&\propto [c(\boldsymbol{\theta})]^{\tau_0+\alpha_0 n_0+n}\exp\left\{\sum_{i=1}^{k}\eta_i(\boldsymbol{\theta})\left(\tau_i+a_0\sum_{l=1}^{n_0}h_i(y_l)+\sum_{j=1}^{n}h_i(x_j)\right)\right\}\\
&\propto \pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}+\alpha_0\boldsymbol{t}_{n_0}(\boldsymbol{Y})+\boldsymbol{t}_n(\boldsymbol{X})\right),
\end{aligned}$$

where $\boldsymbol{t}_n(\boldsymbol{X})=\left(n,\sum_{j=1}^{n}h_1(x_j),\ldots,\sum_{j=1}^{n}h_k(x_j)\right)$ is a $(k+1)$-dimensional vector, with $\boldsymbol{X}=(x_1,...,x_n)$ being the observed data. This is a member of exponential family,

and specifically of the same distribution form as the initial prior (as expected since we use a conjugate prior).

(ii) We have that the predictive distribution of a future observable will be given by:

$$
\begin{aligned}
&f\left(X_{n+1}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right) = \\
&= \int_{\boldsymbol{\Theta}} f\left(X_{n+1}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right) d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\Theta}} \left[g(X_{n+1}) c(\boldsymbol{\theta}) \exp\left\{ \sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) h_i(X_{n+1}) \right\} \right] \times \left[\left[K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y}) + \boldsymbol{t}_n(\boldsymbol{X})\right) \right]^{-1} \right. \\
&\qquad \left. [c(\boldsymbol{\theta})]^{\tau_0 + \alpha_0 n_0 + n} \exp\left\{ \sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \left(\tau_i + a_0 \sum_{l=1}^{n_0} h_i(y_l) + \sum_{j=1}^{n} h_i(x_j) \right) \right\} \right] d\boldsymbol{\theta} \\
&= \left[K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y}) + \boldsymbol{t}_n(\boldsymbol{X})\right) \right]^{-1} g(X_{n+1}) \times \\
&\quad \times \int_{\boldsymbol{\Theta}} [c(\boldsymbol{\theta})]^{\tau_0 + \alpha_0 n_0 + n + 1} \exp\left\{ \sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) \left(\tau_i + a_0 \sum_{l=1}^{n_0} h_i(y_l) + \sum_{j=1}^{n} h_i(x_j) + h_i(X_{n+1}) \right) \right\} d\boldsymbol{\theta} \Rightarrow
\end{aligned}
$$

$$
f\left(X_{n+1}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right) = \frac{K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y}) + \boldsymbol{t}_n(\boldsymbol{X}) + \boldsymbol{t}_1(X_{n+1})\right)}{K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_{n_0}(\boldsymbol{Y}) + \boldsymbol{t}_n(\boldsymbol{X})\right)} g(X_{n+1}),
$$

where $\boldsymbol{t}_1(X_{n+1}) = (1, h_1(X_{n+1}), \ldots, h_k(X_{n+1}))$ a $(k+1)$-dimensional vector, function of the future observable X_{n+1}. Note that the vectors $\boldsymbol{t}_{n_0}(\boldsymbol{Y}), \boldsymbol{t}_n(\boldsymbol{X})$ and $\boldsymbol{t}_1(X_{n+1})$ refer to the respective sufficient statistics for the power prior and the likelihood.

Q.E.D.

PCC construction will be based on the predictive distribution and it can start as soon as $n = 2$ (except when we have Normal likelihood with both parameters unknown, $\alpha_0 = 0$ and we use the reference prior, where PCC starts at $n = 3$). The exact form of the predictive distribution (under conjugate prior), for various likelihood choices (either discrete or continuous data), used commonly in SPC/M, can be found in Table 2.1.1. To unify notation in the table, we denote by $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{X}) = (y_1, \ldots, y_{n_0}, x_1, \ldots, x_n)$ the vector of historical and current univariate data, $\boldsymbol{w} = (\alpha_0, \ldots, \alpha_0, 1, \ldots, 1)$ the vector of weights corresponding to each element d_j in $\boldsymbol{D}$ and finally we call $N_D = n_0 + n$ the length of the data vector $\boldsymbol{D}$.

Distribution	**Likelihood:** $f(\boldsymbol{x}\|\boldsymbol{\theta})$	**Initial Prior:** $\pi_0(\boldsymbol{\theta}\|\boldsymbol{\tau})$	**Predictive:** $f(x_{n+1}\|\boldsymbol{D},\boldsymbol{w},\boldsymbol{\tau})$
Poisson (s_i=rate)	$X_i\|\theta \sim P(\theta \cdot s_i)$	$\theta \sim G(c,d)$	$f(x_{n+1}\|\boldsymbol{D},\boldsymbol{w},\boldsymbol{\tau},s_1,\ldots,s_n,s_{n+1}) = NBin\left(c+\sum_{j=1}^{N_D} w_j d_j \ , \ \frac{s_{n+1}}{d+\sum_{j=1}^{N_D} w_j s_j + s_{n+1}}\right)$
Binomial (N_i=trials)	$X_i\|\theta \sim Bin(N_i,\theta)$	$\theta \sim Beta(a,b)$	$f(x_{n+1}\|\boldsymbol{D},\boldsymbol{w},\boldsymbol{\tau},N_1,\ldots,N_n,N_{n+1}) = BetaBin\left(a+\sum_{j=1}^{N_D} w_j d_j \ , \ b+\sum_{j=1}^{N_D} w_j N_j - \sum_{j=1}^{N_D} w_j d_j \ , \ N_{n+1}\right)$
Negative Binomial	$X_i\|\theta \sim NBin(r,\theta)$	$\theta \sim Beta(a,b)$	$NBinBeta\left(a+r\sum_{j=1}^{N_D} w_j \ , \ b+\sum_{j=1}^{N_D} w_j d_j \ , \ r\right)$
Normal (known variance)	$X_i\|\theta \sim N(\theta,\sigma^2)$	$\theta \sim N(\mu_0,{\sigma_0}^2)$	$N\left(\frac{\sigma^2\mu_0+{\sigma_0}^2\sum_{j=1}^{N_D} w_j d_j}{\sigma^2+{\sigma_0}^2\sum_{j=1}^{N_D} w_j} \ , \ \frac{{\sigma_0}^2\sigma^2}{\sigma^2+{\sigma_0}^2\sum_{j=1}^{N_D} w_j}+\sigma^2\right)$
Normal (known mean)	$X_i\|\theta^2 \sim N(\mu,\theta^2)$	$\theta^2 \sim IG(a,b)$	$t_{\left(2a+\sum_{j=1}^{N_D} w_j\right)}\left(\mu \ , \ \frac{2b+\sum_{j=1}^{N_D} w_j (d_j-\mu)^2}{2a+\sum_{j=1}^{N_D} w_j}\right)$
Normal (both unknown)	$X_i\left\|\left(\theta_1,{\theta_2}^2\right)\right. \sim N(\theta_1,{\theta_2}^2)$	$\left(\theta_1,{\theta_2}^2\right) \sim NIG\ (\mu_0,\lambda,a,b)$	$t_{\left(2a+\sum_{j=1}^{N_D} w_j\right)}\left(\frac{\lambda\mu_0+\sum_{j=1}^{N_D} w_j d_j}{\lambda+\sum_{j=1}^{N_D} w_j} \ , \ \frac{\left(2b+\sum_{j=1}^{N_D} w_j\left(d_j-\frac{\sum_{j=1}^{N_D} w_j d_j}{\sum_{j=1}^{N_D} w_j}\right)^2+\lambda\sum_{j=1}^{N_D} w_j\left(\frac{\sum_{j=1}^{N_D} w_j d_j}{\sum_{j=1}^{N_D} w_j}-\mu_0\right)^2\frac{1}{\lambda+\sum_{j=1}^{N_D} w_j}\right)\left(\lambda+\sum_{j=1}^{N_D} w_j+1\right)}{\left(2a+\sum_{j=1}^{N_D} w_j\right)\left(\lambda+\sum_{j=1}^{N_D} w_j\right)}\right)$
Gamma (known shape)	$X_i\|\theta \sim G(\alpha,\theta)$	$\theta \sim G(c,d)$	$CompG\left(\alpha \ , \ c+\alpha\sum_{j=1}^{N_D} w_j \ , \ d+\sum_{j=1}^{N_D} w_j d_j\right)$

Weibull (known shape)	$X_i\|\theta^\kappa \sim W(\theta,\kappa)$	$\theta^\kappa \sim IG(a,b)$	$Burr\left(\kappa, a+\sum_{j=1}^{N_D} w_j, \left(b+\sum_{j=1}^{N_D} w_j d_j^\kappa\right)^{1/\kappa}\right)$
Inverse Gamma (known shape)	$X_i\|\theta \sim IG(a,\theta)$	$\theta \sim G(c,d)$	$GB2\left(-1, \left(d+\sum_{j=1}^{N_D} w_j/d_j\right)^{-1}, a, c+a\sum_{j=1}^{N_D} w_j\right)$
Pareto (known minimum)	$X_i\|\theta \sim Pa(m,\theta)$	$\theta \sim G(c,d)$	$exGPD\left(\frac{d+\sum_{j=1}^{N_D} w_j log(d_j/m)}{m\cdot\left(c+\sum_{j=1}^{N_D} w_j\right)}, \left(c+\sum_{j=1}^{N_D} w_j\right)^{-1}\right)$
Lognormal (known σ^2)	$X_i\|\theta \sim LogN(\theta,\sigma^2)$	$\theta \sim N(\mu_0,{\sigma_0}^2)$	$LogN\left(\frac{\sigma^2\mu_0+{\sigma_0}^2\sum_{j=1}^{N_D} w_j log(d_j)}{\sigma^2+{\sigma_0}^2\sum_{j=1}^{N_D} w_j}, \frac{{\sigma_0}^2\sigma^2}{\sigma^2+{\sigma_0}^2\sum_{j=1}^{N_D} w_j}+\sigma^2\right)$
Lognormal (known μ)	$X_i\|\theta^2 \sim LogN(\mu,\theta^2)$	$\theta^2 \sim IG(a,b)$	$Logt_{\left(2a+\sum_{j=1}^{N_D} w_j\right)}\left(\mu, \frac{2b+\sum_{j=1}^{N_D} w_j(log(d_j)-\mu)^2}{2a+\sum_{j=1}^{N_D} w_j}\right)$
Lognormal (both unknown)	$X_i\|\left(\theta_1,{\theta_2}^2\right) \sim LogN(\theta_1,{\theta_2}^2)$	$\left(\theta_1,{\theta_2}^2\right) \sim NIG\left(\mu_0,\lambda,a,b\right)$	$Logt_{\left(2a+\sum_{j=1}^{N_D} w_j\right)}\left(\frac{\lambda\mu_0+\sum_{j=1}^{N_D} w_j log(d_j)}{\lambda+\sum_{j=1}^{N_D} w_j}, \frac{\left(2b+\sum_{j=1}^{N_D} w_j\left(log(d_j)-\frac{\sum_{j=1}^{N_D} w_j log(d_j)}{\sum_{j=1}^{N_D} w_j}\right)^2+\lambda\sum_{j=1}^{N_D} w_j\left(\frac{\sum_{j=1}^{N_D} w_j log(d_j)}{\sum_{j=1}^{N_D} w_j}-\mu_0\right)^2\frac{1}{\lambda+\sum_{j=1}^{N_D} w_j}\right)\left(\lambda+\sum_{j=1}^{N_D} w_j+1\right)}{\left(2a+\sum_{j=1}^{N_D} w_j\right)\left(\lambda+\sum_{j=1}^{N_D} w_j\right)}\right)$

Table 2.1.1: The predictive distribution using an initial conjugate prior in a power prior mechanism for some of the distributions typically used in SPC/M, which also belong to the k-PREF. $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{X}) = (y_1, \ldots, y_{n_0}, x_1, \ldots, x_n)$ is the vector of historical and current univariate data, $\boldsymbol{w} = (\alpha_0, \ldots, \alpha_0, 1, \ldots, 1)$ are the weights corresponding to each element d_j of $\boldsymbol{D}$ and $N_D = n_0 + n$.

2.1.2 Prior elicitation

The big advantage of PCC is the use of (typically available) prior information, which allows to decrease the uncertainty of the unknown parameter(s) $\boldsymbol{\theta}$, improving the performance (with respect to false alarms and detection power), especially at the early stages. The speed at which this uncertainty decreases is inversely related to the information that the prior distribution carries. When strong opinion about the unknown parameter(s) is available and located accurately (i.e. we have highly informative initial prior placed at the parameter space where the unknown parameter is), then the PCC performance will be optimal, i.e. the false alarm tolerance at the nominal level and quite high detection power. Nevertheless, a highly informative prior miss-placed on the parameter space (with respect to where the true unknown $\boldsymbol{\theta}$ is), will have as result to get an extremely high False Alarm Rate (FAR), until sufficient information from the data moves the posterior to the area where the true $\boldsymbol{\theta}$ lies. Thus, a general recommendation is to avoid having a highly informative initial prior distribution (to eliminate the risk of inflated false alarms if miss-placed). Wang et al. (2018) developed effective numerical methods for exploring reasonable choices of an informative prior distribution.

From the above it becomes evident that the elicitation of the hyper-parameters $\boldsymbol{\tau}$ play an important role to PCC. There are two different ways that one can proceed: being subjective or objective. In the latter we use non-informative priors and in a sense we leave the data to carry the information. In the former we use a low/medium (but not high) informative prior distribution. Such a prior will carry more information compared to the objective priors (reducing the posterior variability of $\boldsymbol{\theta}$) enhancing the PCC performance, especially at the start of the process. Furthermore, as the size of the data increases, the influence of the low/medium information prior is washing-out.

In the case where no prior information for $\boldsymbol{\theta}$ exists, or a user prefers to follow an

objective prior approach, then the hyper-parameters determination should be chosen with caution, especially when we do not have historical data to use in a power prior (i.e. $\alpha_0 = 0$). Various classes of non-informative priors exist like:

- **Flat prior**: a uniform prior equally weighting all possible values of the unknown parameter.
- **Jeffreys prior**: a prior that is closed under parameter transformations.
- **Reference prior**: a function that maximizes some measure of distance (e.g. Hellinger) or divergence (e.g. Kullback-Leibler) between the posterior and prior, as data become available.

A list of Jeffreys and reference initial priors that can be used for likelihoods that are members of the k-PREF are given in Table 2.1.2.

Likelihood $f(\cdot\|\boldsymbol{\theta})$	**Initial Reference/Jeffreys Prior** $\pi_0(\boldsymbol{\theta}\|\boldsymbol{\tau})$
$P(\theta \cdot s_i)$	$\pi_0(\theta) \propto \frac{1}{\sqrt{\theta}} \equiv G(1/2, 0)$
$Bin(N_i, \theta)$	$\pi_0(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}} \equiv Beta(1/2, 1/2)$
$NBin(r, \theta)$	$\pi_0(\theta) \propto \frac{1}{\theta\sqrt{(1-\theta)}} \equiv Beta(0, 1/2)$
$W(\theta, \kappa)$	$\pi_0(\theta^\kappa) \propto \frac{1}{\theta^\kappa} \equiv IG(0, 0)$
$G(a, \theta)$, $IG(a, \theta)$, $Pa(m, \theta)$	$\pi_0(\theta) \propto \frac{1}{\theta} \equiv G(0, 0)$
$N(\theta, \sigma^2)$, $LogN(\theta, \sigma^2)$	$\pi_0(\theta) \propto c \equiv N(0, +\infty)$
$N(\mu, \theta^2)$, $LogN(\mu, \theta^2)$	$\pi_0(\theta^2) \propto \frac{1}{\theta^2} \equiv IG(0, 0)$
$N(\theta_1, \theta_2^2)$, $LogN(\theta_1, \theta_2^2)$	$\pi_0^R(\theta_1, \theta_2^2) \propto \frac{1}{\theta_2^2} \equiv NIG(0, 0, -1/2, 0)$, $\pi_0^J(\theta_1, \theta_2^2) \propto \frac{1}{\theta_2^3} \equiv NIG(0, 0, 0, 0)$

Table 2.1.2: Initial Reference (R) and Jeffreys (J) prior distributions. For univariate $\boldsymbol{\theta}$ the two classes of non-informative priors coincide.

When we need to choose an "objective" prior we should aim to satisfy the following properties: have the minimal possible influence in the process, do not decrease the reflexes of PCC and attempt to have stable false alarm performance. Based on this proposal we will next provide more specific details along with some guidelines for the likelihoods of Normal, Poisson and Binomial that are most common in SPC/M studies.

For the $N(\theta_1, \theta_2^2) - NIG(\mu_0, \lambda, a, b)$ model, we have to carefully determine the parameters of the Inverse Gamma (i.e. a and b). For example, the prior $NIG(0, \epsilon, \epsilon, \epsilon)$ (which converges to Jeffreys prior as $\epsilon \to 0$) gives higher density at values of θ_2^2 which are close to 0. Thus, it becomes very informative, increasing drastically the false alarms especially for large values of θ_2^2. Similar results hold for $NIG(0, \epsilon, 1/2, \epsilon)$ and $NIG(0, \epsilon, 1, \epsilon)$, where the mean of the marginal posterior of θ_2^2 is the MLE and the unbiased estimator respectively. On the other hand, a flatter prior like $NIG(0, \epsilon, \epsilon, 1)$ may overestimate θ_2^2 reducing the reflexes of PCC. Generally, we recommend to choose a value for the hyper-parameter $a > 2$, so that the mean and the variance of the prior Inverse Gamma is defined. In different cases, the prior parameters have to be determined carefully.

For the $P(\theta_3) - Gamma(c, d)$ model, the initial prior $Gamma(\epsilon, \epsilon)$ seems not to be a good choice. Despite that the posterior mean is the MLE, this prior may increase the number of false alarms, especially when θ_3 is close to 0. In that case, if $x_n = 0$, then no-alarm zone of PCC, which will be defined as Highest Predictive Mass ($HPrM$) region Rn+1 in Subsection 2.1.3, will shrink to a short region. In general we found that small values for both of the hyper-parameters c and d (e.g. less than 1/3) tend to affect R_{n+1} in the same manner, even when the prior mean is correctly located.

For $Bin(N, \theta_4) - Beta(a, b)$ model we propose to avoid $Beta(\epsilon, \epsilon)$, which converges to Haldane's prior (Haldane, 1932) as $\epsilon \to 0$, where the posterior mean is equal to the MLE, as we will have inflated false alarms. Also, we suggest to avoid small values for both of the hyper-parameters a and b (e.g. less than 1/3), especially if θ_4 is close

to 0 as we will have inflated false alarms (just as we had in the Poisson-Gamma respective case). In contrary, the flat $Beta(1,1)$, equally weighting all values of θ_4, will have the posterior mode to be the MLE and provide weak information, inflating the predictive. Thus, the detection performance of PCC will be affected.

Generally, reference priors (Bernardo, 1979, Berger et al., 2009) and neutral priors (Kerman, 2011) provide a stable start to PCC under total prior ignorance. Our proposal though, when some information about the unknown parameters exists, is to adopt a medium/low volume information prior $\pi_0(\boldsymbol{\theta}|\boldsymbol{\tau})$ which will enhance the PCC performance (compared to non-informative choices) and its effect will be removed once a short sequence of data becomes available.

2.1.3 HPrD/M region and Type I error

The PCC is based on the sequentially updated form of the predictive distribution, which is used to determine a region denoted by R_{n+1}, where the future observable (X_{n+1}) will most likely be, as long as the process is stable (i.e. no changes occurred). The region R_{n+1} will be the $100(1-\alpha)\%$ Highest Predictive Density/Mass (HPrD/M) region, which is the unique shortest region, that minimizes the absolute difference with the predetermined coverage. For notational convenience we will adopt the name HPrD, even for cases in which the predictive distribution is discrete, where we derive the Highest Predictive Mass (HPrM) region.

The definiton of HPrD/M is as follows:

Definition 2.1.1. Assume the set R^c which contains the values of the predictive density (or mass) function, which are greater than a threshold c, i.e.:

$$R^c = \{x_{n+1} : f(x_{n+1}|\boldsymbol{D}, \boldsymbol{w}, \boldsymbol{\tau}) \geq c\}. \tag{2.1.9}$$

The HPrD/M region will be given by minimizing the absolute difference of a highest

predictive probability from a significance level $1-a$, for all the possible values of c. Specifically:

$$R_{n+1} = \min_{R^c} \left| \int_{R^c} f(x_{n+1}|\boldsymbol{D}, \boldsymbol{w}, \boldsymbol{\tau}) - (1-a) \right|, \tag{2.1.10}$$

for the discrete case, we replace the integral sign by summation.

R_{n+1} will be the shortest region with the smallest absolute difference from the probability $1-a$. In other words, it minimizes the Lebesque measure $m(R^c)$ for continuous cases or the corresponding measure $l(R^c) = \sum_i \delta_{x_i} \left(f(x_i|\boldsymbol{D}, \boldsymbol{w}, \boldsymbol{\tau}) \geq c \right)$ for discrete cases, where $\delta_{x_i}(\cdot)$ represents the Dirac delta function.

For continuous distributions the HPrD region is calculated just like the Highest Posterior Density (HPD) region in Bayesian analysis (see for example Carlin and Louis, 2009), where instead of the posterior, we use the predictive distribution and the minimum value of the absolute difference will be 0. For discrete predictive distributions, typically we will not be able to obtain a region that has the exact coverage probability $1-\alpha$. In this case the HPrD/M can be obtained by starting from the mode of the predictive distribution and continue adding sequentially the next most probable values of the predictive distribution, until we get sufficiently close (minimizing the absolute difference) to the predetermined coverage level $1-\alpha$. Algorithm 1 provides the details in how to derive the HPrM region for a discrete predictive distribution and Figure 2.1.1 provides an illustration.

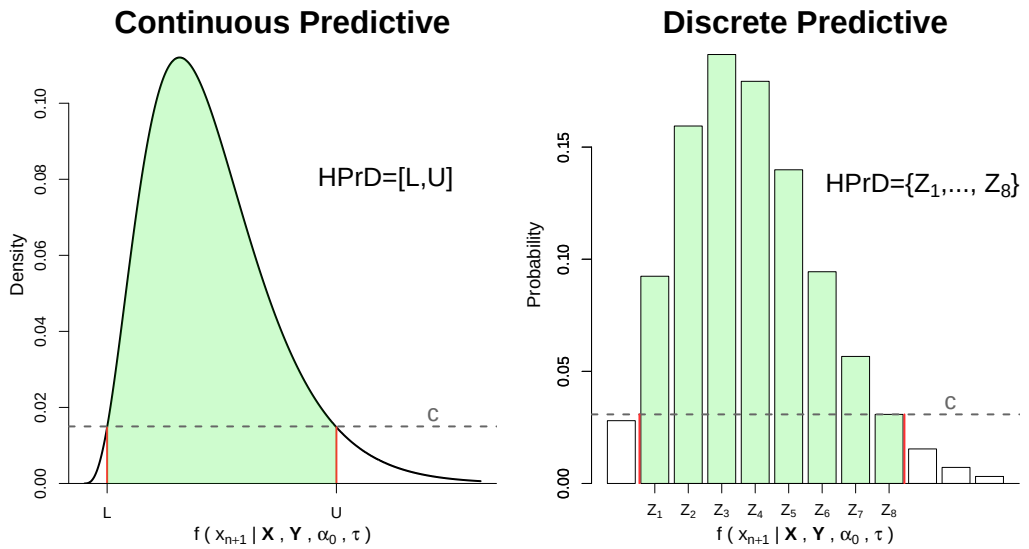


Figure 2.1.1: The HPrD/M region (R_{n+1}) for continuous (left panel) and discrete (right panel) data.

Algorithm 1 HPrM algorithm for a discrete distribution

Set p_i the i^{th} decreasing ordered probability of $f\left(X_{n+1}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)$, e.g. p_1 is the max
Set $z_i = \arg\{p_i\}$, i.e. the argument(s) where p_i get their values
$n \leftarrow 1$ { *initial values* }
$sum\_probs \leftarrow 0$
$diff \leftarrow 1$
$HPrM \leftarrow \emptyset$
$stop \leftarrow 0$
while $stop = 0$
    $sum\_probs \leftarrow sum\_probs + p_n$
    if $|sum\_probs - (1-a)| < diff$
        $HPrM \leftarrow \{HPrM, z_n\}$
        $diff \leftarrow |sum\_probs - (1-a)|$
        $n \leftarrow n + 1$
    else
        $stop \leftarrow 1$
$HPrM \leftarrow sort\{HPrM\}$

We should also note here that in symmetric discrete predictive distributions (like a Beta Binomial with $\alpha = \beta$), the HPrM region might not be unique, as there might exist two regions that achieve the minimum of absolute difference (we can choose at

random).

PCC will plot the sequentially updated HPrD/M region versus time, providing the "in control" region of the next data point and thus give an alarm if a new observable does not belong to the respective HPrD/M region. For unimodal predictive distributions, the region R_{n+1} will be an interval for continuous distributions, or a set with consecutive numbers for the discrete case, while for a multimodal predictive, R_{n+1} might be formed as a union of non-overlapping regions.

Regarding the false alarm tolerance, the (predetermined) parameter $0 < \alpha < 1$, also known as False Alarm Rate (FAR), will reflect our tolerance to false alarms and consequently the detection power. The proposed PCC can be viewed as a sequential (multiple) hypothesis testing procedure, where at each time point n we draw the HPrD/M region (R_{n+1}) for the future observable, so that if no changes occurred in the process (IC state), the probability to raise an alarm is: $P\left(X_{n+1} \notin R_{n+1} | IC\right) = \alpha$. We suggest two metrics in selecting α, depending on whether we know or not in advance the number of data points, N, that PCC will be used for (in short runs or Phase I studies) and/or whether N is large.

If we have a (known) fixed horizon of N data points, for which PCC will be employed and N is not too large (typically up to a few dozens), then we suggest to control the Family Wise Error Rate ($FWER$), which expresses the probability of raising at least one false alarm out of a pre-determined number of N hypothesis tests. This is identical to the concept of False Alarm Probability (FAP) introduced by Chakraborti et al. (2008) for phase I analysis. Among various proposals in controlling $FWER$, we adopt the Šidák's correction (Šidák, 1967), which is slightly more powerful than the popular Bonferroni's correction (Dunn, 1961). Šidák's correction assumes independence across tests and is more conservative in the presence of positive dependence, compared with independent tests. If we define V to be the number of false alarms observed in a PCC, applied on N observations in total, i.e. $n = 1, \ldots, N$, from the IC state of the distribution ($0 \leq V \leq N - 1$, when PCC starts at n=2), then the

Šidák's correction (assuming independence) will provide:

$$FWER = P\left(V \geq 1\right) = 1 - P\left(V = 0\right) = 1 - P\left(\bigcap_{i=2}^{N} \{X_i \in R_i | IC\}\right)$$

$$= 1 - \prod_{i=2}^{N} P\left(X_i \in R_i | IC\right) = 1 - \left(1-\alpha\right)^{N-1} \Rightarrow \alpha = 1 - \left(1 - FWER\right)^{\frac{1}{N-1}}. \quad (2.1.11)$$

So, once we know N and we set the desirable $FWER$, we can obtain the parameter α needed in deriving the HPrD/M regions, R_{n+1}. It is evident that as N increases, α decreases and approaches zero, it leads to an extremely conservative decision scheme, that will reduce the OOC detection power.

We recommend to use the above approach, as long as $\alpha \geq 10^{-3}$, even though this can be adjusted depending on the type of process we monitor. However, in the cases where N is either unknown in advance or it is too large, then we suggest to derive α using the metric of IC Average Run Length (ARL_0). Following Montgomery (2020), this corresponds to the desired average number of data points that we will plot in the PCC before a false alarm occurs, given that the process is under the IC state. As N increases, the updated posterior distribution gets more informative (offering consistent estimates of the unknown parameters) and thus the resulting hypothesis tests will tend to be nearly independent. Then, the value of the desired (predetermined) ARL_0 will be approximately:

$$ARL_0 \approx \frac{1}{\alpha} \quad \Rightarrow \quad \alpha \approx \frac{1}{ARL_0}. \quad (2.1.12)$$

Based on either (2.1.11) or (2.1.12), we predetermine the coverage level $100(1-\alpha)\%$ that the HPrD/M region (R_{n+1}) will have.

2.1.4 Fast Initial Response (FIR) PCC

One of the most serious issues in self-starting methods, is the weak response to early shifts (Goedhart et al., 2017, Capizzi and Masarotto, 2019). The Fast Initial Response (FIR) feature is typically used to improve the performance of the standard charts for early shifts in a process. Lucas and Crosier (1982) were the first to propose a FIR feature for CUSUM, while Steiner (1999) introduced the FIR EWMA by narrowing the control limits. In the latter, the time dependent effect of the FIR adjustment, decreases exponentially with time and becomes negligible after a few observations. Precisely, Steiner's adjustment is given by:

$$\mathrm{FIR}_{adj} = 1 - (1 - f)^{1+\mathrm{a}(t-1)}, \tag{2.1.13}$$

where $\mathrm{a} > 0$ is a smoothing parameter, t is the current number of hypotheses tests performed and $0 < f < 1$ represents the proportion of the adjusted limit over the initial test (i.e. $t = 1$).

As the PCC uses control limits, much like the EWMA, we will adopt Steiner's adjustment for a time-varying narrowing of the R_{n+1} region in the start of the process. Despite the head-start the FIR option can provide to PCC, we should make sure that we do not significantly inflate the false alarms. Thus, the FIR parameters should be selected by taking into account the false alarm behavior of PCC, which depends on the prior settings, especially when the volume of available data is small. If an extremely informative prior (near point mass) is used, then the PCC behavior acts like a typical Shewhart chart, as the resulting R_{n+1} region is not essentially updated by new observations. On the other hand, if a non-informative prior, like the initial reference prior without historical IC data, is selected, then the FAR depends only on the (iid) data. As a result for these two cases, the observed FAR will meet the predetermined standards (even from the very first hypothesis testing) and therefore we should avoid the use of a FIR adjustment (or otherwise the observed FAR will

be inflated).

However, in the case of a weakly informative prior, the R_{n+1} region is quite wide (as we combine prior and likelihood uncertainty), but at the same time the prior distribution provides beneficial information for the IC state. Combining these two facts, the first IC data points are more likely to be plotted within the R_{n+1} region. This will result in a temporarily smaller (from what is anticipated) FAR, especially for the very early tests at the start of a process. Thus, we could use a FIR adjustment without a negative effect on the predetermined expected number of false alarms. We propose to be somewhat conservative and use $f = 0.99$, i.e. the adjusted R_{n+1} region will be the 99% of the original for the first test and a $= (-3/log_{10}(1-f) - 1)/4$, i.e. the adjusted R_{n+1} region will be the 99.9% of the original at the fifth test. We should note that t is the current number of tests, not the number of observations, as for the first (or the second) observation PCC does not provide a test.

2.2 PCC decision making

The major role of PCC is to control a process and identify transient large shifts (outliers), in an online fashion and without a phase I exercise. As such, PCC performs a hypothesis test as each new data point x_{n+1} becomes available and raises an alarm when $x_{n+1} \notin R_{n+1}$, indicating that the new observable is not in agreement with what is anticipated from the predictive distribution (that was built from the previous data and the prior distribution). The endpoints of R_{n+1}, formed from the predictive distribution, play the role of the control limits of the chart. The range of these limits reflect the variability of the predictive distribution, which is known to depend on both the length of the available data and the precision of the prior distribution. For a weakly informative prior the range will be wider at the start of the process and as more data become available it will become more narrow and eventually stabilize, washing out the effect of the prior. Figure 2.2.1 provides illustrations of PCC for data

streams of length 30 that come from a continuous (Normal data with both parameters unknown) and two discrete (Poisson and Binomial) cases, when the process is either IC or has a large isolated shift at location 15 (OOC scenario).

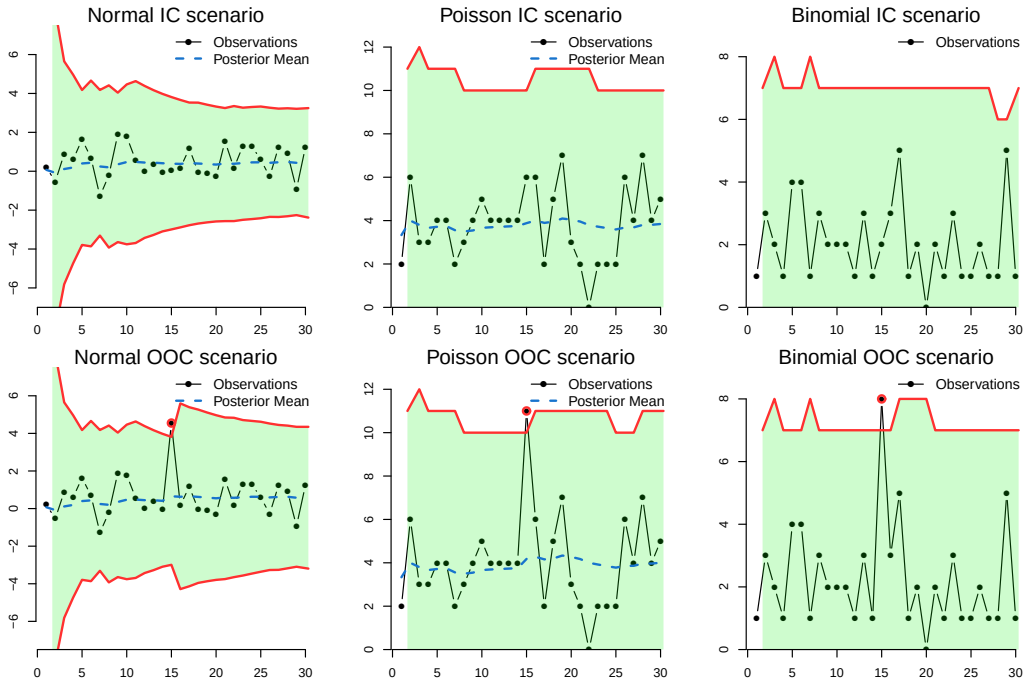


Figure 2.2.1: The IC and OOC illustration of PCC for i.i.d. Normal, Poisson and Binomial data. For the IC Normal data $X_i|\left(\theta_1,\theta_2^2\right) \sim N\left(\theta_1=0,\theta_2^2=1\right)$ and for the OOC case we sample $X_{15} \sim N(4,1)$. The initial prior was $\left(\theta_1,\theta_2^2\right) \sim NIG\left(\mu=0,\lambda=2,a=1,b=0.8\right)$. For the IC Poisson data $X_i|\theta_3 \sim P\left(\theta_3=4\right)$. For the OOC case $X_{15} \sim P(10)$, while $\theta_3 \sim G\left(c=8,d=2\right)$. For the IC Binomial data $X_i|\theta_4 \sim Bin\left(N=20,\theta_4=0.1\right)$. For the OOC case $X_{15} \sim Bin(20,0.368)$, while $\theta_4 \sim Beta\left(a=0.5,b=4.5\right)$. In all cases, α needed to derive the $100(1-\alpha)\%$ HPrD/M (R_{n+1}) was selected to satisfy $FWER=0.05$ for $N=30$ observations.

As can be seen in Figure 2.2.1, the limits tend to become more narrow and finally stabilize when the size of the data increases, forming a more informative posterior distribution of the unknown parameter(s). The outlying observations in all scenarios are plotted outside the R_{n+1} region, hence raising an alarm. The region R_{n+1} is formed online, after the data point x_n becomes available, and so when we get an alarm (i.e. $x_{n+1} \notin R_{n+1}$), the suggestion is to stop the process, perform some root cause analysis to identify external sources of variation, possibly have an intervention

and finally restart the PCC (the posterior we had right before the alarm can act as the new prior, or the previous IC data can be used in the power prior mechanism). However, if we will not react to an alarm, due to the Bayesian dynamic update mechanism, the isolated change detected will be absorbed. As a consequence, the posterior and predictive distribution will have inflated variance leading to wider R_{n+1} regions. In the OOC scenarios in Figure 2.2.1 we observe that the R_{n+1} regions are wider at time 16 due to the "no action" policy at the alarm for time 15. This effect is reduced with time but it is still present until observation 30, where the R_{n+1} is wider compared to the respective region of the IC data.

The PCC methodology with all possible options is synopsized in a flowchart in Figure 2.2.2 and in pseudocode in Algorithm **??**.

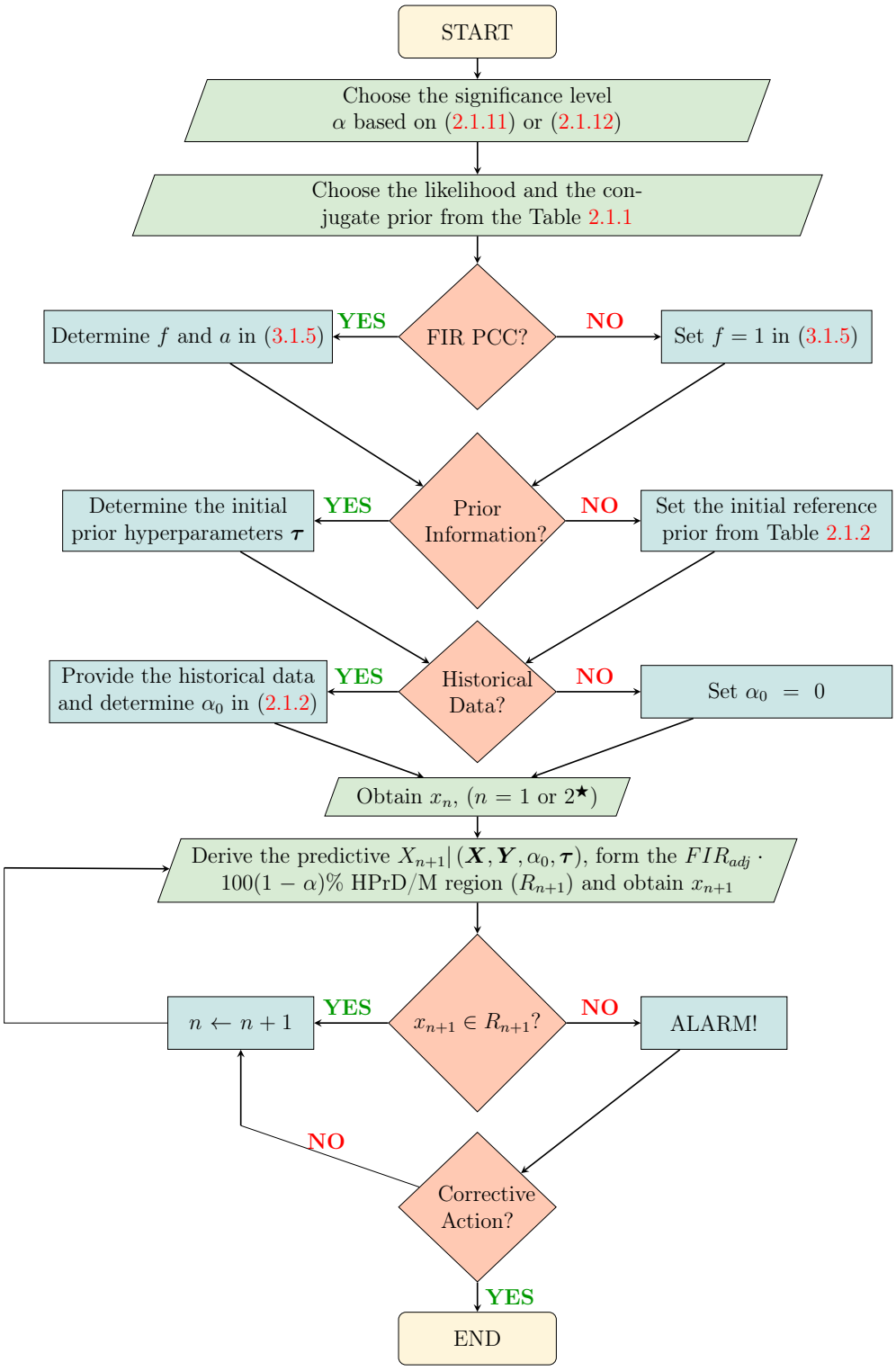


Figure 2.2.2: PCC flowchart. A parallelogram corresponds to an input/output information, a decision is represented by a rhombus and a rectangle denotes an operation after a decision making. In addition, the rounded rectangles indicate the beginning and end of the process.

Algorithm 2 PCC algorithm

1: Select the significance level α, based on $FWER$ or ARL_0 { *FAR* }
2: Choose the data distribution and the conjugate prior density for $\boldsymbol{\theta}$ { *distributions* }
3: **Is FIR-PCC of interest?** { *FIR* }
4: **YES**
5: Determine the parameters f and a
6: **NO**
7: Set $f = 1$
8: **Is prior information available?** { *initial prior* $\pi_0(\cdot)$ }
9: **YES**
10: Determine the hyperparameters of the initial prior $\boldsymbol{\tau}$
11: **NO**
12: Set the initial reference/Jeffeys prior (see Table 2.1.2)
13: **Are prior data available?** { *power prior* }
14: **YES**
15: Provide the historical data $\boldsymbol{Y}$ and determine α_0
16: **NO**
17: Set $\alpha_0 = 0$
18: Once the data point x_n $(n \geq 1^{\star})$ arrives, derive the predictive distribution of next observable $X_{n+1}|\,(\boldsymbol{X}, \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau})$
19: Derive the $FIR_{adj} \cdot 100(1-\alpha)\%$ HPrD/M region, obtain x_{n+1} and draw it { R_{n+1} }
20: **if** $x_{n+1} \in R_{n+1}$ { *test* }
21: $n \leftarrow n + 1$
22: **goto** 18
23: **else** { *alarm!* }
24: **if** you do not make a corrective action
25: **then goto** 21
26: **else**
27: **end**

*For the Normal - NIG model using the initial reference prior and $\alpha_0 = 0$ we need $n = 2$ to initiate PCC, while for all other cases PCC starts at after x_1 becomes available.

Apart from controlling a process, PCC can be used for monitoring the unknown parameter(s). As we showed in Theorem 2.1.1, before deriving the predictive distribution at each time point, we first obtain the posterior distribution for the unknown

parameter(s). Decision theory can be used to provide loss function based optimal point/interval estimates and/or hypothesis testing for each parameter. For example, using the squared error loss function, the Bayes rule (optimal point estimate) is known to be the mean of the posterior distribution (Carlin and Louis, 2009), i.e. we have a (sequentially updated) point estimate of the unknown process parameter(s). To illustrate this option, in Figure 2.2.1, we additionally plot the posterior mean estimate of θ_1 for the Normal and θ_3 for the Poisson cases.

Finally, PCC summarizes the predictive distribution through a region, but other forecasting options (like point estimates) are straightforward to derive as well using decision theory.

2.3 Comparative study and sensitivity analysis

The PCC is developed in a general framework, allowing its use for any likelihood that belongs to the k-PREF. In traditional SPC/M, significant amount of work has been dedicated for Normal, Poisson and Binomial data. When the goal is to detect transient large shifts in a short run process of individual univariate data, without employing a phase I calibration stage, the frequentist based Q charts developed by Quesenberry (1991a,b,c) are probably the most prominent representative methods for Normal, Binomial and Poisson data respectively. In absence of phase I parameter estimates, the Q charts provide a self-starting monitoring method, where calibration and testing happens simultaneously, aiming to detect process disturbances (OOC states) in an online fashion.

2.3.1 Competing methods

In this Subsection, we will present the Q chart procedure for Normal, Poisson and Binomial data, i.e. the methods that we will compare against the proposed PCC. We

denote as $\Phi^{-1}(\cdot)$ the inverse of the standard Normal CDF, $G_\nu(\cdot)$ the Student-t CDF with ν degrees of freedom, $B(\cdot)$ the Binomial CDF and $H(\cdot)$ the Hypergeometric CDF. Briefly, the Q statistics are:

- Q statistic (Quessnberry, 1991a) for the Normal data assuming both μ and σ^2 unknown, i.e. $X_i|\theta \sim N(\theta_1, \theta_2^2)$

$$Q_{n+1} = \Phi^{-1}\left\{ G_{n-1}\left(\frac{X_{n+1} - \sum_{j=1}^{n} X_j}{\sqrt{\sum_{j=1}^{n} (x_j - \bar{x}_n)^2 / (n-1)}} \right) \right\}, \quad n = 2, 3, ...$$

- Q statistic (Quessnberry, 1991c) for the Poisson data assuming rate λ unknown and inspected units s_i known, i.e. $X_i|\theta \sim P(\theta \cdot s_i)$

$$Q_{n+1} = \Phi^{-1}\left\{ B\left(X_{n+1}; \sum_{j=1}^{n+1} X_j, \frac{s_{n+1}}{\sum_{j=1}^{n+1} s_j} \right) \right\}, \quad n = 1, 2, ...$$

- Q statistic (Quessnberry, 1991b) for the Binomial data assuming probability p unknown and size N known, i.e. $X_i|\theta \sim Bin(N_i, \theta)$

$$Q_{n+1} = \Phi^{-1}\left\{ H\left(X_{n+1}; \sum_{j=1}^{n+1} X_j, N_{n+1}, \sum_{j=1}^{n} N_j \right) \right\}, \quad n = 1, 2, ...$$

For all the above cases, the Q statistics are tested against the standard Normal distribution. It is worth mentioning that only for the Normal case the Q statistics for the Normal case are independently and identically distributed $N(0, 1)$ random variables. For the Poisson and the Binomial case the Q statistics approximated by the standard Normal probabilities. Such approximations are known to be questionable for small values of the unknown parameters. The Lower Control Limit (LCL) and

the Upper Control Limit (UCL) for Q statistics are:

$$LCL = \Phi^{-1}(a_1) \quad \text{and} \quad UCL = \Phi^{-1}(1 - a_2),$$

where a_1 and a_2 are the false alarm rate from the lower and the upper side respectively, usually equal.

2.3.2 Simulation results

Quesenberry (1991a) presented three versions of Q chart for Normal data, where either a parameter is known or both unknown (we ignore the scenario that both parameters are known). For these three scenarios and their relation to the respective PCCs we have the following Lemma.

Lemma 2.3.1. All three versions of Q chart for Normal data are special cases of the respective PCCs, when the initial prior is the reference prior and we do not make use of a power prior option (i.e. $\alpha_0 = 0$).

Proof

Following Quesenberry (1991a) the Q chart in all three cases of the Normal distribution, makes use at each data point x_{n+1}, of the statistic Q_{n+1}. For PCC we set $\alpha_0 = 0$, eliminating the power prior part regarding the past data ($\boldsymbol{Y}$) and in each case we set the hyperparameters $\boldsymbol{\tau}$, so that we have the respective reference prior for the unknown parameter(s). We will show that controlling Q_{n+1} statistic is identical to controlling PCC's standardized predictive residual:

$$PR_{n+1} = \frac{X_{n+1} - \hat{\mu}_n}{\hat{\sigma}_n}$$

where, $\hat{\mu}_n$ and $\hat{\sigma}_n$ are the mean and standard deviation respectively of the predictive distribution of $X_{n+1}|\,(\boldsymbol{X}, \boldsymbol{Y}, \alpha_0 = 0, \boldsymbol{\tau}) \equiv X_{n+1}|\,(\boldsymbol{X}, \boldsymbol{\tau})$. Denoting by $\Phi^{-1}(\cdot)$ the inverse of the standard normal CDF and $G_\nu(\cdot)$ the Student-t CDF with ν degrees of

freedom we get:

Case I: μ unknown, σ^2 known.

We have: $X_i|\theta \sim N(\theta, \sigma^2)$ and the reference prior is $\pi(\theta) \propto c \equiv N(0, +\infty)$. Then the predictive distribution will be:

$$X_{n+1}|(\boldsymbol{X}, \boldsymbol{\tau}) \sim N\left(\bar{x}_n\ ,\ \frac{n+1}{n}\sigma^2\right) \Rightarrow PR_{n+1} = \frac{X_{n+1} - \bar{x}_n}{\sqrt{\frac{n+1}{n}}\sigma} = Q_{n+1} \sim N(0,1).$$

Case II: μ known, σ^2 unknown.

We have: $X_i|\theta^2 \sim N(\mu, \theta^2)$ and the reference prior is $\pi(\theta^2) \propto 1/\theta^2 \equiv IG(0,0)$. Then the predictive distribution will be:

$$X_{n+1}|(\boldsymbol{X}, \boldsymbol{\tau}) \sim t_{n-1}\left(\mu\ ,\ \frac{\sum_{j=1}^{n}(x_j - \mu)^2}{n}\right) \Rightarrow PR_{n+1} = \frac{X_{n+1} - \mu}{\sqrt{\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{n}}} \sim t_{n-1}.$$

Transformating the PR_{n+1} we get: $\Phi^{-1}\{G_n(PR_{n+1})\} = Q_{n+1} \sim N(0,1)$.

Case III: μ unknown and σ^2 unknown.

We have: $X_i|(\theta_1, \theta_2^2) \sim N(\theta_1, \theta_2^2)$ and the reference prior is $\pi(\theta_1, \theta_2^2) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$. Then the predictive distribution will be:

$$X_{n+1}|(\boldsymbol{X}, \boldsymbol{\tau}) \sim t_{n-2}\left(\bar{x}_n\ ,\ \frac{\sum_{j=1}^{n}(x_j - \bar{x}_n)^2}{n-1}\right) \Rightarrow PR_{n+1} = \frac{X_{n+1} - \bar{x}_n}{\sqrt{\frac{\sum_{j=1}^{n}(x_j-\bar{x}_n)^2}{n-1}}} \sim t_{n-2}.$$

Transformating again the PR_{n+1} we get: $\Phi^{-1}\{G_{n-1}(PR_{n+1})\} = Q_{n+1} \sim N(0,1)$.

For cases II and III, as the functions $\Phi^{-1}(\cdot)$ and $G_\nu(\cdot)$ are injective, it is identical to control PR_{n+1} or Q_{n+1}.

Q.E.D.

The proof shows that the Normal Q charts (in all three cases) are identical to the

respective PCC when neither prior information (i.e. use of reference prior) nor historical data are available. What happens though when prior information and/or historical data do exist? In such scenarios, the posterior distribution will be more informative, enhancing the predictive distribution, which will boost the PCC performance. For discrete data (Poisson and Binomial) the Q charts use the uniform minimum variance unbiased (UMVU) estimation of the cumulative distribution function of the process, thus we lose ability to compare analytically against the respective exact discrete PCC.

In what follows we will perform a simulation study to examine the performance of Q charts against PCC when we have $N = 30$ data points from $N(\theta_1, \theta_2^2)$, $P(\theta_3)$ or $Bin(20, \theta_4)$ distributions. We will design charts to have a $FWER = 0.05$ at the last observation $N = 30$ (using Šidák correction). We will compare the running $FWER(k) = 1 - P\left(\bigcap_{i=2}^{k} \{X_i \in R_i | IC\}\right)$ of Q charts and PCC at each of the $k = 2, \ldots, 30$ data points, when we simulate IC sequences from $N(\theta_1 = 0, \theta_2^2 = 1)$, $P(\theta_3 = 2)$ and $Bin(20, \theta_4 = 0.1)$ respectively (see Keefe et al., 2015 for more details regarding the conditional IC performance of self-starting control charts). To examine the OOC detection power of Q charts and PCC we will use the IC sequences generated and introduce large isolated shifts at one of the locations: 5 (early), 15 (middle) or 25 (late). The size of the shifts that we will consider are:

- Normal mean: $\delta_N = \{2.5\theta_2 \text{ or } 3\theta_2\} = \{2.5 \text{ or } 3\}$, i.e. OOC states come from $N(2.5, 1)$ or $N(3, 1)$.
- Poisson mean (or variance): $\delta_P = \{2.5\sqrt{\theta_3} \text{ or } 3\sqrt{\theta_3}\} = \{2.5\sqrt{2} \text{ or } 3\sqrt{2}\}$, i.e. OOC states come from $P(2 + 2.5\sqrt{2}) = P(5.536)$ or $P(2 + 3\sqrt{2}) = P(6.243)$.
- Binomial probability of success: $\delta_B = \left\{2.5\sqrt{\frac{\theta_4(1-\theta_4)}{N}} \text{ or } 3\sqrt{\frac{\theta_4(1-\theta_4)}{N}}\right\} = \left\{2.5\sqrt{\frac{0.1(1-0.1)}{20}} \text{ or } 3\sqrt{\frac{0.1(1-0.1)}{20}}\right\}$, i.e. OOC states come from $Bin(20, 0.268)$ or $Bin(20, 0.301)$.

For detection, we will record the cases that a chart provides an alarm at the exact

time that the shift was introduced. Specifically, these cases will be denoted as the OOC Detection (OOCD), where $OOCD(k') = P\left(\{X_{k'} \notin R_{k'}|OOC\} \bigcap_{i=2}^{k'-1} \{X_i \in R_i|IC\}\right)$ and $k' = \{5, 15, 25\}$. Both $FWER(k)\%$ for IC data (at each time 2, ..., 30) and $OOCD(k')\%$ at locations 5, 15 or 25 will be estimated over 100,000 iterations.

PCC will require to define a prior distribution and so within this simulation study we will take advantage to examine the sensitivity of the PCC performance for various prior settings. Precisely, for each setup described above, we will make use of two initial priors (reference and weakly informative) and two values for the α_0 parameter (0 or $1/n_0$) representing the absence or presence of n_0 historical data $\boldsymbol{Y}$ (we will use $n_0 = 10$ historical data from the IC likelihood). Therefore, for each scenario we will compare the Q chart against one of the four possible versions of PCC (with/without prior knowledge, with/without historical data). The initial priors $\pi_0(\cdot|\boldsymbol{\tau})$, which we will employ are (see Figure 2.3.1):

- Normal: reference prior $\pi_0(\theta_1, \theta_2^2) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ or the weakly informative $NIG(0, 2, 1, 0.8)$.
- Poisson: reference prior $\pi_0(\theta_3) \propto 1/\sqrt{\theta_3} \equiv G(1/2, 0)$ or the weakly informative $G(4, 2)$.
- Binomial: reference prior $\pi_0(\theta_4) \propto 1/\sqrt{\theta_4(1-\theta_4)} \equiv Beta(1/2, 1/2)$ or the weakly informative $Beta(0.5, 4.5)$.

The simulation findings are summarized graphically in Figure 2.3.2 and analytically in Table 2.3.1, where we observe that overall PCC outperforms Q chart. Starting from the false alarms in the case of Normal data, both methods reach the nominal 5% at time $N = 30$, but at all time points k, the $FWER(k)$ of PCC is always smaller. For both discrete cases, the Q chart's $FWER(k)$ becomes unacceptably high, something that is caused from the fact that the true parameter values are near (even though not too close) to the parameter space boundary, which in conjunction with the UMVU estimation, inflates drastically the false alarms (the closer we get to the parameter boundary the worst the Q chart's performance regarding false alarms).

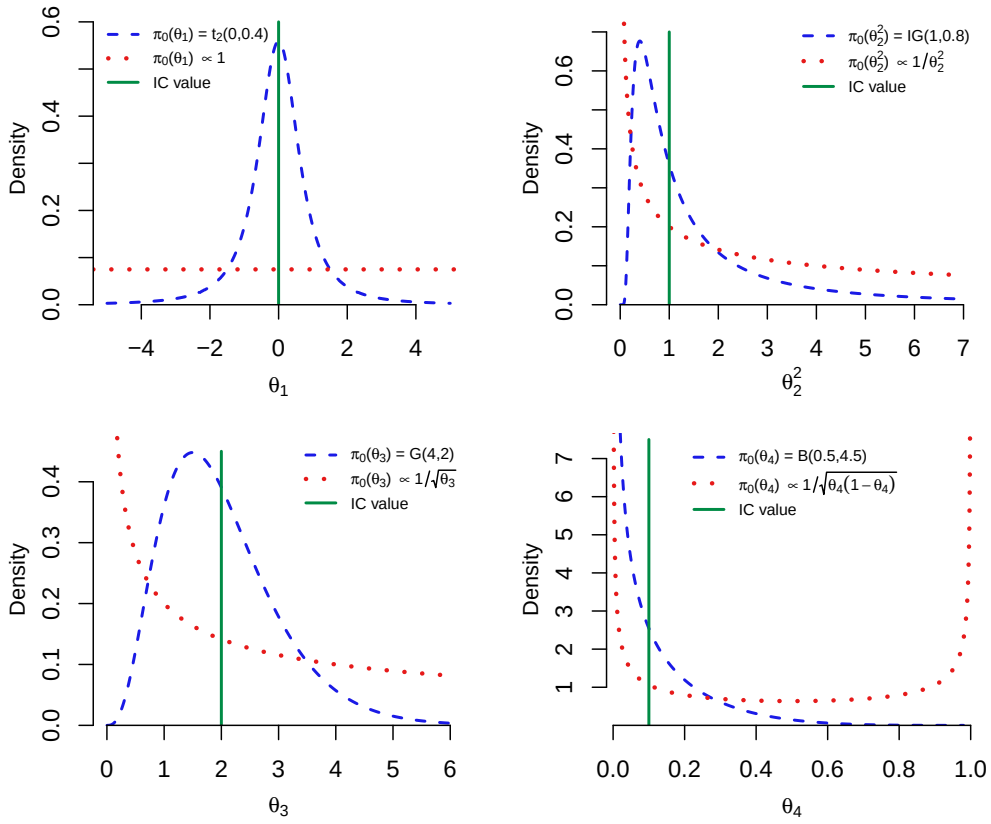


Figure 2.3.1: The initial reference (i.e. non-informative) and the weakly informative prior distributions used in the simulation study, along with the IC values (as vertical segments) for the parameters $\theta_1, \theta_2^2, \theta_3$ and θ_4 of the simulation study.

Finally, the extremely small $FWER(k)$ observed for PCC in the first 5 data points motivates the use of the FIR-PCC described in Section 2.1.4.

For the Normal data, the simulations verify Lemma 2.3.1, as the Q chart and the PCC with reference prior and no historical data have identical performance. Moving to the detection power, as it is measured by $OOCD(k')$, both methods improve as the size of the shift increases (from 2.5 to 3 sd) or the shift delays its appearance (from $k' = 5$ to 15 to 25), just as it was expected. Especially for the shifts at time 5, PCC greatly outperforms Q charts thanks to the head-start from the prior and/or the historical data. Focusing at each location of the shift, we observe that as we move from Q chart to PCC with reference prior and next to PCC with weakly

informative prior the performance improves (quite significantly for some scenarios). When relevant historical data are available, through the power prior mechanism, they further boost the performance. The somewhat competitive performance of Q chart in one of the Binomial scenarios should be considered in conjunction with its quite high FWER, when compared to the one achieved by PCC (see also Table 2.3.2, where the FWER of PCCs is increased to align with the one that Q chart can achieve in the Poisson and Binomial cases, offering a straightforward comparison of detection power). In summary, PCC appears more powerful to the respective Q charts in detecting isolated shifts in short runs of individual data.

Focusing on the performance of PCC at location $k' = 5$, we observe that in the Normal scenario we have smaller power compared to the respective setting in Poisson or Binomial (as we move k' to higher values, the differences vanish). This is caused from the fact that in the Normal scenario we have two unknown parameters as opposed to the Poisson and Binomial cases where each has only one unknown parameter (a PCC built using four data points for a setting with two unknown parameters will be a lot more challenging, as opposed to a setting with only one unknown parameter). A Normal PCC scheme with either the mean or the variance being known would radically improve the performance reaching (or even overcoming) the levels achieved in the Poisson and Binomial. The effect of the two unknown parameters (Normal) versus the single unknown parameter (Poisson and Binomial) is responsible in the performance of PCC_1 to PCC_4 in detecting outliers at $k' = 25$. With one unknown parameter, the information collected from the 24 in control data points has significantly reduced the posterior (and predictive) uncertainty, shrinking the effect of the prior and providing a near uniform performance. For the Normal case though the posterior (and predictive) uncertainty at $k' = 25$ remains non-negligible, allowing the prior setting to play some role and differentiate the performance across the four versions of PCCs (in general the more the data the higher the shrinkage of the prior's effect). Concluding, we should note that PCC was shown to be more powerful in

detecting large isolated shifts compared to Q chart. The relative performance of Q chart to PCC remains the same when we use medium or small shifts, with detection power dropping as the size of the isolated shift decreases.

	Jump	k'	Q chart $OOCD(k')\%$ $(FWER\%)$	PCC_1 $OOCD(k')\%$ $(FWER\%)$	PCC_2 $OOCD(k')\%$ $(FWER\%)$	PCC_3 $OOCD(k')\%$ $(FWER\%)$	PCC_4 $OOCD(k')\%$ $(FWER\%)$
Normal	0σ		(5.049)	(5.049)	(4.347)	(4.776)	(4.932)
	2.5σ	5	1.901	1.901	1.492	4.205	6.271
		15	12.791	12.791	14.249	17.433	18.407
		25	17.025	17.025	17.691	20.005	20.371
	3σ	5	2.873	2.873	2.816	9.024	12.556
		15	22.809	22.809	24.914	30.112	31.426
		25	30.095	30.095	31.021	34.410	34.880
Poisson	$0\sqrt{\lambda}$		(18.283)	(4.515)	(4.192)	(4.409)	(4.320)
	$2.5\sqrt{\lambda}$	5	12.437	12.696	14.793	16.265	16.928
		15	17.220	18.196	18.660	19.052	19.302
		25	17.704	19.164	19.180	19.510	19.623
	$3\sqrt{\lambda}$	5	18.185	19.185	21.984	24.240	25.204
		15	24.930	26.826	27.434	27.972	28.345
		25	25.740	28.153	28.196	28.683	28.823
Binomial	$0\sqrt{\frac{p(1-p)}{N}}$		(17.878)	(4.387)	(3.991)	(4.852)	(4.381)
	$2.5\sqrt{\frac{p(1-p)}{N}}$	5	14.079	15.848	15.540	16.111	17.008
		15	20.057	18.845	19.319	20.084	20.067
		25	20.284	19.878	20.035	19.839	20.315
	$3\sqrt{\frac{p(1-p)}{N}}$	5	21.646	24.078	24.098	24.509	26.039
		15	29.469	28.765	29.353	30.207	30.213
		25	29.952	30.165	30.389	30.117	30.703

Table 2.3.1: The FWER for $N = 30$ (in parenthesis) and the outlier detection power at $k' = \{5, 15, 25\}$, of the Q chart against PCC under a reference prior (PCC_1), a reference prior with historical data (PCC_2), a weakly informative prior (PCC_3) and a weakly informative prior with historical data (PCC_4). The results refer to Normal, Poisson and Binomial data.

	Jump	k'	Q chart $OOCD(k')\%$ $(FWER\%)$	PCC_1 $OOCD(k')\%$ $(FWER\%)$	PCC_2 $OOCD(k')\%$ $(FWER\%)$	PCC_3 $OOCD(k')\%$ $(FWER\%)$	PCC_4 $OOCD(k')\%$ $(FWER\%)$
Poisson	$0\sqrt{\lambda}$		(18.283)	(16.498)	(15.646)	(16.550)	(16.183)
	$2.5\sqrt{\lambda}$	5	18.185	34.295	35.388	38.820	39.221
		15	24.930	38.634	39.192	39.899	40.388
		25	25.740	37.823	38.215	38.456	38.679
	$3\sqrt{\lambda}$	5	12.437	25.410	26.138	28.906	29.157
		15	17.220	28.657	29.108	29.736	30.166
		25	17.704	28.181	28.440	28.692	28.869
Binomial	$0\sqrt{\frac{p(1-p)}{N}}$		(17.878)	(16.606)	(15.383)	(17.950)	(16.682)
	$2.5\sqrt{\frac{p(1-p)}{N}}$	5	21.646	38.442	38.898	38.345	40.992
		15	29.469	40.947	42.666	42.406	43.004
		25	29.952	40.052	41.283	40.589	41.210
	$3\sqrt{\frac{p(1-p)}{N}}$	5	14.079	28.073	28.037	27.982	29.906
		15	20.057	29.549	30.984	30.920	31.351
		25	20.284	29.040	30.053	29.662	30.039

Table 2.3.2: The FWER for $N = 30$ (in parenthesis) and the outlier detection power at $k' = \{5, 15, 25\}$, of the Q chart against PCC under a reference prior (PCC_1), a reference prior with historical data (PCC_2), a weakly informative prior (PCC_3) and a weakly informative prior with historical data (PCC_4). The results refer to Poisson and Binomial data, where PCC has aligned FWER with the one achieved by Q chart.

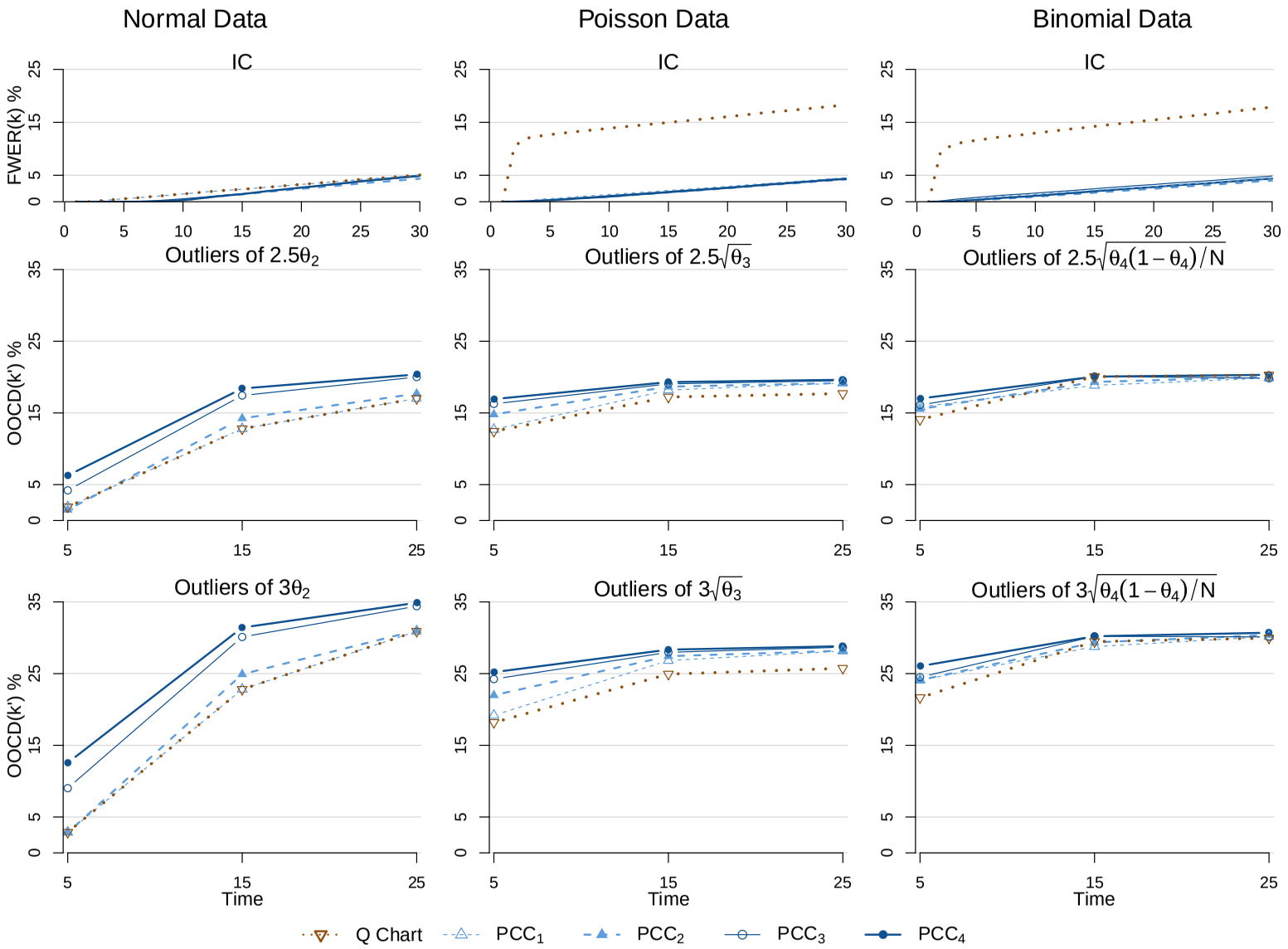


Figure 2.3.2: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 30$ (top row) and the $OOCD(k')$ at $k' = 5, 15$ or 25, of the Q chart and PCC under a reference prior (PCC_1), a reference prior with historical data (PCC_2), a weakly informative prior (PCC_3) and a weakly informative prior with historical data (PCC_4), when we have outliers of 2.5 (middle row) or 3 (bottom row) standard deviations. Columns 1 to 3 refer to the Normal, Poisson and Binomial cases respectively.

2.4 PCC Robustness

Apart from checking the prior sensitivity that was done in Subsection 2.3.2, we will also examine how robust the suggested PCC performance is to possible model type misspecifications. For the PCC construction we assume that the observed data are iid observations from a specific likelihood. In this section, we will examine how robust is the PCC performance when:

(a) we violate the assumption of independence (i.e. the data are correlated)

(b) the assumed likelihood function is invalid (i.e. data are generated from a different random variable from the one assumed in the PCC construction).

Regarding (a) we will use a Normal (with both parameters unknown) PCC implementation, but the actual data will be generated as sequentially dependent Normal data via an autoregressive (AR) model: $X_n = c + \phi X_{n-1} + \epsilon_n$ with $c = 0$ and $\epsilon_n \sim N(0,1)$. To examine various degrees of dependence we will use $\phi = -0.4, 0.4$ (moderate) or 0.8 (high). For the outlying observations we will set $c = 2.5$ or 3, in order to introduce shifts of size of 2.5σ or 3σ respectively, at one of the locations 5, 15 or 25 (just as we did in Section 2.3.2).

For (b) we will examine the following scenarios:

- Use a Normal based PCC (both parameters unknown) while the data are generated from a Student t_7 distribution, i.e. we have heavier tails (t_7 is symmetric, with the same mean but 40% inflated variance compared with the standard Normal).

- Use a Normal based PCC (both parameters unknown) while the data are generated from a $Gumbel\,(\mu = -0.5, \beta = 0.8)$ distribution, i.e. we have skewed data ($Gu\,(-0.5, 0.8)$ has approximately the same mean and variance with the standard Normal, but it has positive skewness ≈ 1.14).

- Use a Poisson based PCC while the data are generated from a $NBin\,(r=6, p=1/4)$ distribution, i.e. we have over-dispersed data ($NBin\,(6,1/4)$ has the same mean with $P(2)$, but its variance is $\approx 33\%$ inflated).

The aforementioned likelihoods are illustrated in Figure 2.4.1.

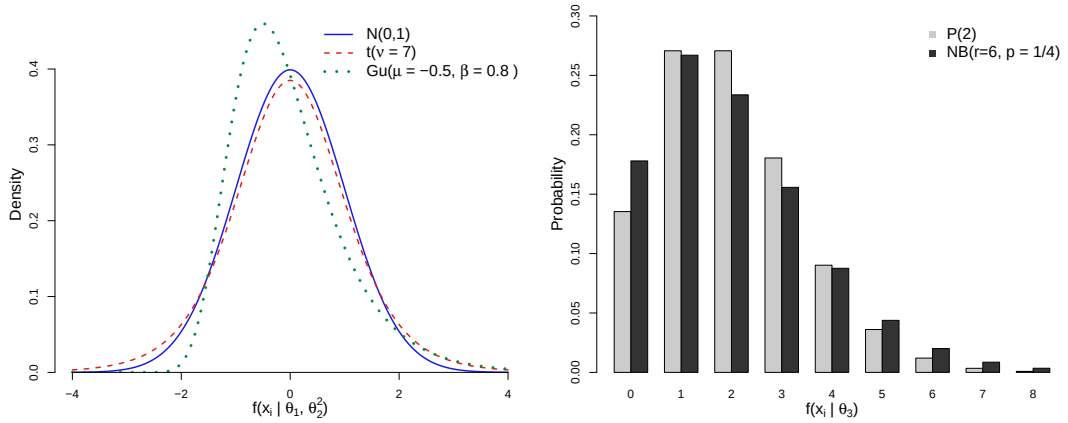


Figure 2.4.1: The various misspecification of the PCC distributional forms regarding the continuous (left panel) and discrete (right panel) data generation mechanisms.

For this misspecification scenario, we generate the OOC data from the introduced distributions in a manner that the isolated large shifts will correspond to either 2.5 or 3 standard deviations, again at locations 5, 15 or 25 (similar to what we had in Section 2.3.2). Precisely:

- Student t: OOC states come from $t_7\left(\mu = 2.5\cdot\sqrt{7/5}, \sigma = 1\right)$ or $t_7\left(\mu = 3\cdot\sqrt{7/5}, \sigma = 1\right)$.
- Gumbel: OOC states come from $Gu\,(-0.5+2.5, 0.8)$ or $Gu\,(-0.5+3, 0.8)$.
- Negative Binomial: OOC states come from $NBin\,(6\cdot 2.5, 1/4)$ or $NBin\,(6\cdot 3, 1/4)$.

The prior distributions (reference prior and weakly informative) along with the use or not of $n_0 = 10$ historical data (power prior with $\alpha_0 = 0$ or $1/n_0$) will be identical to the ones used in Section 2.3.2.

Figures 2.4.2 and 2.4.3 summarize graphically the results of Tables 2.4.1 and 2.4.2, regarding the performance ($FWER(k)$ and $OOCD(k')$ are as defined in Section 2.3.2)

for independence and distributional misspecifications respectively. In the former, we observe that PCC is almost unaffected in the presence of moderate autocorrelation. For highly dependent data ($\phi = 0.8$ or larger), PCC is somewhat less robust as it decreases its detection power and slightly increases the FWER percentages, however still achieving noticeable performance, especially at the early stages thanks to the IC prior information.

In the distributional violation scenarios (Figure 2.4.3), we observe that PCC retains its high detection percentages in all cases. However, the $FWER(k)$ is significantly inflated. This can be explained by considering the shape discrepancies among the assumed and actual likelihood functions, where IC values are somewhat outlying under the misspecified assumed model (a more strict α value in determining the HPrD/M region would reduce the $FWER(k)$ in such scenarios at the cost of somewhat reducing power).

	Jump	k'	PCC_1 $OOCD(k')\%$ $(FWER\%)$	PCC_2 $OOCD(k')\%$ $(FWER\%)$	PCC_3 $OOCD(k')\%$ $(FWER\%)$	PCC_4 $OOCD(k')\%$ $(FWER\%)$
A R (1), $\phi = -0.4$	$0sd$		(4.420)	(3.293)	(4.711)	(4.480)
	$2.5sd$	5	1.421	0.511	4.038	4.789
		15	9.822	10.369	14.050	14.441
		25	13.289	13.794	15.995	16.270
	$3sd$	5	2.059	1.066	8.092	9.880
		15	17.294	18.516	24.093	24.776
		25	23.557	24.446	27.724	28.185
A R (1), $\phi = 0.4$	$0sd$		(6.319)	(4.135)	(5.530)	(5.026)
	$2.5sd$	5	2.535	0.531	4.082	4.755
		15	12.724	12.915	16.640	16.669
		25	15.511	15.943	18.120	18.308
	$3sd$	5	3.671	1.155	8.615	10.138
		15	21.836	22.571	28.115	28.342
		25	26.773	27.656	30.740	31.135
A R (1), $\phi = 0.8$	$0sd$		(9.218)	(5.637)	(7.226)	(6.795)
	$2.5sd$	5	3.098	0.347	3.135	3.854
		15	11.237	10.191	12.407	12.121
		25	10.341	10.509	11.668	11.640
	$3sd$	5	4.591	0.857	6.508	7.904
		15	17.783	16.820	20.031	19.832
		25	16.488	16.931	18.619	18.712

Table 2.4.1: The FWER at $N = 30$ (in parenthesis) and the outlier detection power at $k' = \{5, 15, 25\}$ for the Normal distribution for PCC with both parameters being unknown, when we actually have data from an AR(1) process. PCC process is under a reference prior (PCC_1), a reference prior with historical data (PCC_2), a weakly informative prior (PCC_3) and a weakly informative prior with historical data (PCC_4).

	Jump	k'	PCC_1 $OOCD(k')\%$ $(FWER\%)$	PCC_2 $OOCD(k')\%$ $(FWER\%)$	PCC_3 $OOCD(k')\%$ $(FWER\%)$	PCC_4 $OOCD(k')\%$ $(FWER\%)$
Student$(df=7)$	$0sd$		(15.338)	(14.425)	(19.128)	(19.361)
	$2.5sd$	5	2.543	1.366	8.282	9.606
		15	14.576	15.417	19.861	20.468
		25	17.560	18.313	20.427	20.847
	$3sd$	5	3.782	2.737	15.511	18.167
		15	25.243	27.059	33.409	34.462
		25	30.435	31.765	34.518	35.183
Gumbel$(-0.5, 0.8)$	$0sd$		(21.903)	(19.583)	(23.849)	(23.227)
	$2.5sd$	5	3.488	1.245	6.320	6.953
		15	15.614	15.528	18.505	18.180
		25	16.654	17.021	18.387	18.333
	$3sd$	5	4.911	2.279	10.943	12.150
		15	27.444	25.030	29.539	29.259
		25	26.648	27.426	29.420	29.549
Neg. Bin$(6, \frac{1}{4})$	$0sd$		(17.526)	(16.761)	(17.686)	(17.543)
	$2.5sd$	5	11.626	12.478	13.976	14.055
		15	14.766	15.035	15.442	15.504
		25	14.499	14.601	14.772	14.848
	$3sd$	5	19.709	21.374	23.701	24.010
		15	24.251	24.690	25.254	25.351
		25	23.790	23.997	24.171	24.290

Table 2.4.2: The FWER at $N = 30$ (in parenthesis) and the outlier detection power at $k' = \{5, 15, 25\}$ for the Normal distribution for PCC violating the distributional assumption. Panel 1 and 2 refer to the Normal PCC with both parameters being unknown while the data come from a Student or Gumbel distribution respectively. In panel 3 we assume Poisson based PCC while the data are from a Negative Binomial. PCC process is under a reference prior (PCC_1), a reference prior with historical data (PCC_2), a weakly informative prior (PCC_3) and a weakly informative prior with historical data (PCC_4).

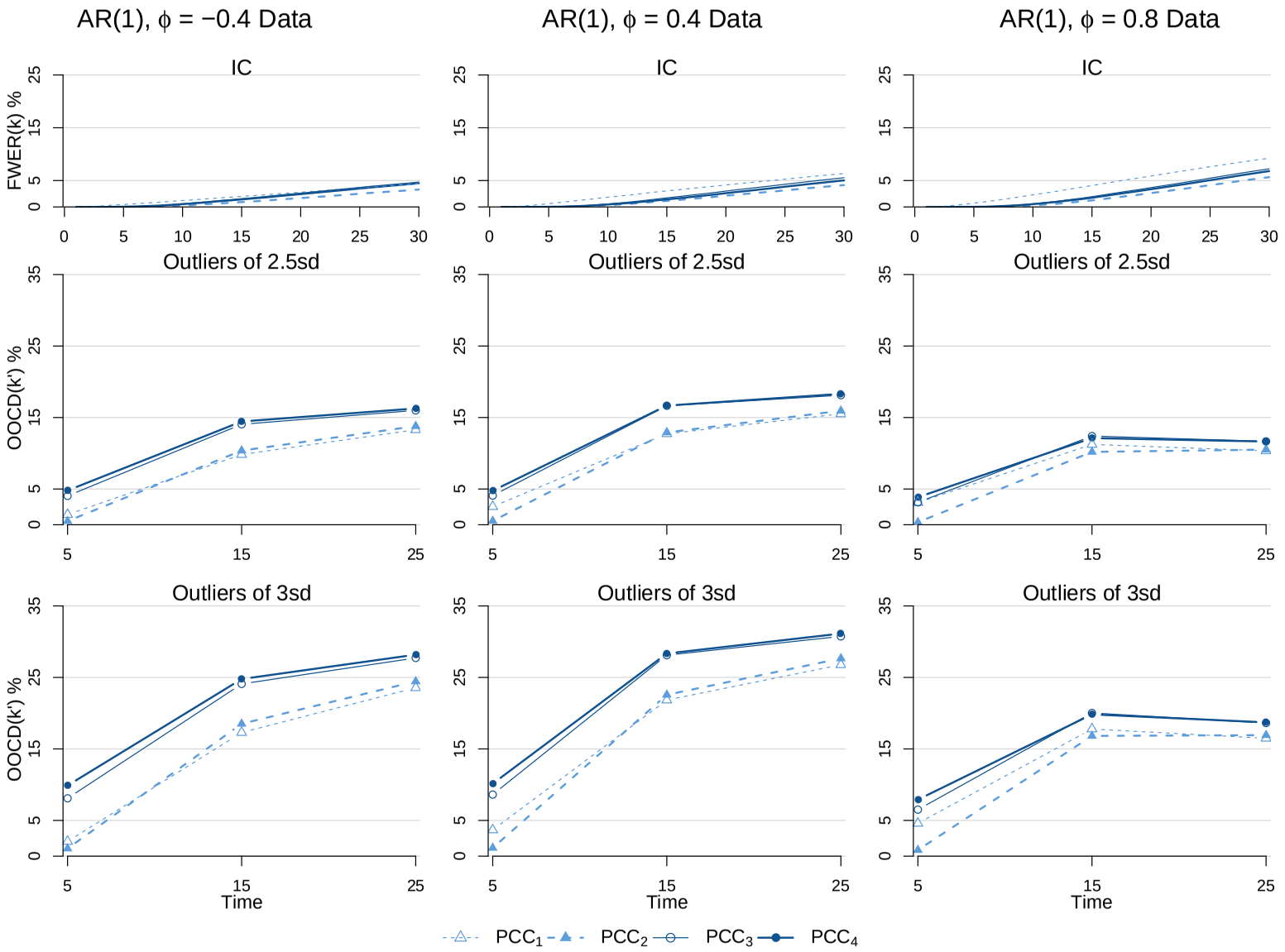


Figure 2.4.2: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 30$ (top row) and the $OOCD(k')$ at $k' = 5, 15$ or 25 and size of 2.5 (middle row) or 3 (bottom row) standard deviations for the Normal distribution PCC with both parameters being unknown, when we actually have data from an AR(1) process. A reference or weakly informative prior and the presence or absence of historical data is considered. Columns 1 to 3 refer to the various degrees of autocorrelation.

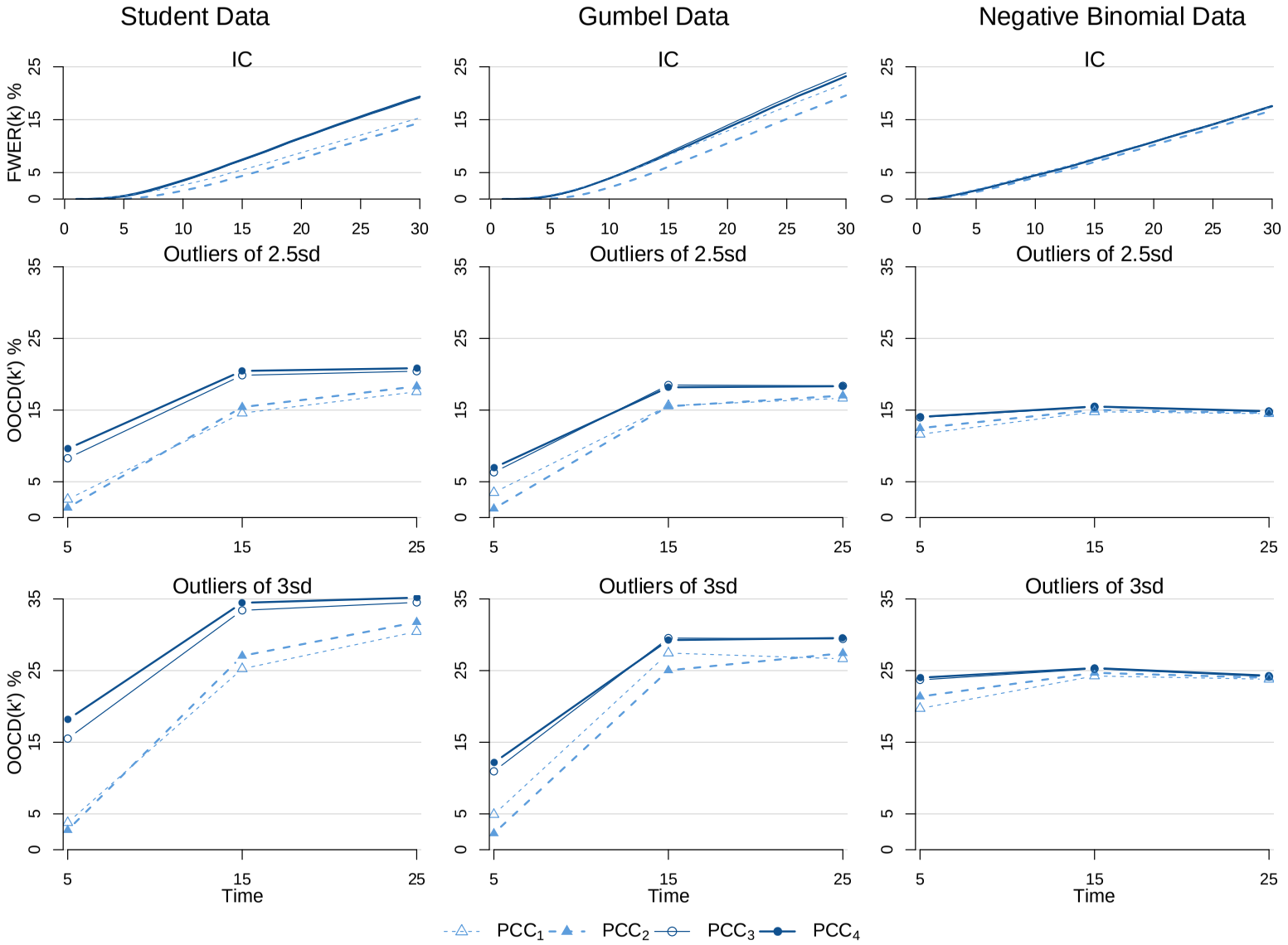


Figure 2.4.3: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 30$ (top row) and the $OOCD(k')$ at $k' = 5, 15$ or 25, of PCC under a reference or weakly informative prior and in the presence or absence of historical data, when we have outliers of 2.5 (middle row) or 3 (bottom row) standard deviations. Columns 1 and 2 refer to the Normal PCC with both parameters being unknown while the data come from a Student or Gumbel distribution respectively. In column 3 we assume Poisson based PCC while the data are from a Negative Binomial.

Finally, for both the violation schemes, it is worth mentioning that PCC detection seems to be stabilized and not necessarily improved when the outliers occur at location 25. This can be attributed to the contaminated estimates of the unknown parameters from the data that violate the PCC assumptions, as well as the fact that the influence of the prior is decreased. Overall, the PCC appears to be robust when we violate the assumptions, as its performance is somewhat reduced but noticeably far from collapsing.

2.5 PCC real data application

2.5.1 PCC application to Normal data

In this section we will illustrate the use of PCC in practice. Specifically, we will apply the proposed PCC methodology in two real data sets (one for continuous and one for discrete data). Regarding the continuous case, we will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory. We are interested in the variable "activated Partial Thromboplastin Time" (aPTT), measured in seconds. APTT is a blood test that characterizes coagulation of the blood. It is a routine clotting time test and can be used as a diagnosis of bleeding risk (e.g. aPTT value is higher in patients with hemophilia or Willebrand disease) or for unfractionated heparin treatment monitoring. We gathered 30 daily normal IQC observations (X_i) from a medical lab (see Table 2.5.1), where $X_i|\left(\theta_1, {\theta_2}^2\right) \sim N\left(\theta_1, {\theta_2}^2\right)$. Notice that these data are based on control samples and in regular practice will become available sequentially. The goal is to accurately detect any transient parameter shift of large size, as this will have an impact on the reported patient results. Thus, it is of major importance to perform on-line monitoring of the process without a phase I exercise. Via available prior information, we elicit the prior $\pi_0\left(\theta_1, {\theta_2}^2|\boldsymbol{\tau}\right) \sim NIG\left(29.6, 1/7, 2, 0.56^2\right)$. Furthermore, there were $n_0 = 30$

historical data (from a different reagent) available (see Table 2.5.1), with $\bar{\mathbf{y}} = 30.18$ and $var(\mathbf{y}) = 0.32$. We set $\alpha_0 = 1/30$ and combining these two sources of information we get the power prior $\pi\left(\theta_1, {\theta_2}^2 | \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) \sim NIG\left(30.1, 8/7, 5/2, 0.7^2\right)$. To examine prior sensitivity we will also use as initial prior the reference prior $\pi_0\left(\theta_1, {\theta_2}^2 | \boldsymbol{\tau}\right) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ (to declare a-priori ignorance) and so we will get two versions of PCC (one for each initial prior). Figure 2.5.1 provides the two versions of PCC (continuous/dotted limits for weakly informative/reference prior) along with a plot of the historical data and the marginal distributions of the mean (θ_1) and variance (θ_2^2) at the end of the data collection.

$y_1 - y_{15}$	30.4	29.9	30.1	30.2	31.2	30.7	30.6	29.6	29.3	30.2	30.4	30.3	29.5	29.9	30.2
$x_1 - x_{15}$	30.8	30.2	30.9	30.2	30.5	30.4	30.9	30.2	30.3	30.1	30.6	29.9	30.5	29.8	30.5
$y_{16} - y_{30}$	29.9	30.5	29.7	30.7	29.9	29.6	30.1	30.1	29.9	30.1	29.9	29.9	29.7	32.2	30.6
$x_{16} - x_{30}$	28.8	30.3	30.4	30.6	30.2	30.8	30.7	31.0	30.3	30.7	30.2	30.3	30.6	30.4	30.2

Table 2.5.1: The aPTT (in seconds) internal quality control observations of the historical $\boldsymbol{Y} = (y_1, y_2, \ldots, y_{30})$ and the current $\boldsymbol{X} = (x_1, x_2, \ldots, x_{30})$ data.

Specifically, for each parameter we plot the marginal weakly informative initial, $\pi_0\left(\cdot|\boldsymbol{\tau}\right)$, power, $\pi\left(\cdot|\boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right)$, priors and the posterior distribution, $p\left(\cdot|\boldsymbol{X}, \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right)$. We should emphasize that despite the fact that we provide the plots at the end of the data sequence, in practice the PCC chart and each of the two posterior distributions will start being plotted at observation 2 and 1 respectively and will be sequentially updated every time a new observable becomes available. PCC provides an alarm at location 16, indicating that there was a transient large shift during that day. This would call for checking the process at that date and if an issue was found then we would take some corrective action, initiate the PCC and reanalyze all the patient samples that were received between days 15 (no alarm) and 16 (alarm). In the present study, no action was taken and the process continued to operate. As a result, the PCC limits were inflated right after the alarm, but this effect was gradually absorbed as more IC data become available. We also note (as expected) that the use of the reference prior provides wider limits, especially at the early stage of the process, making the chart less responsive. Finally, the marginal posterior distri-

butions can be used to draw inference regarding the unknown parameters, at each time point.

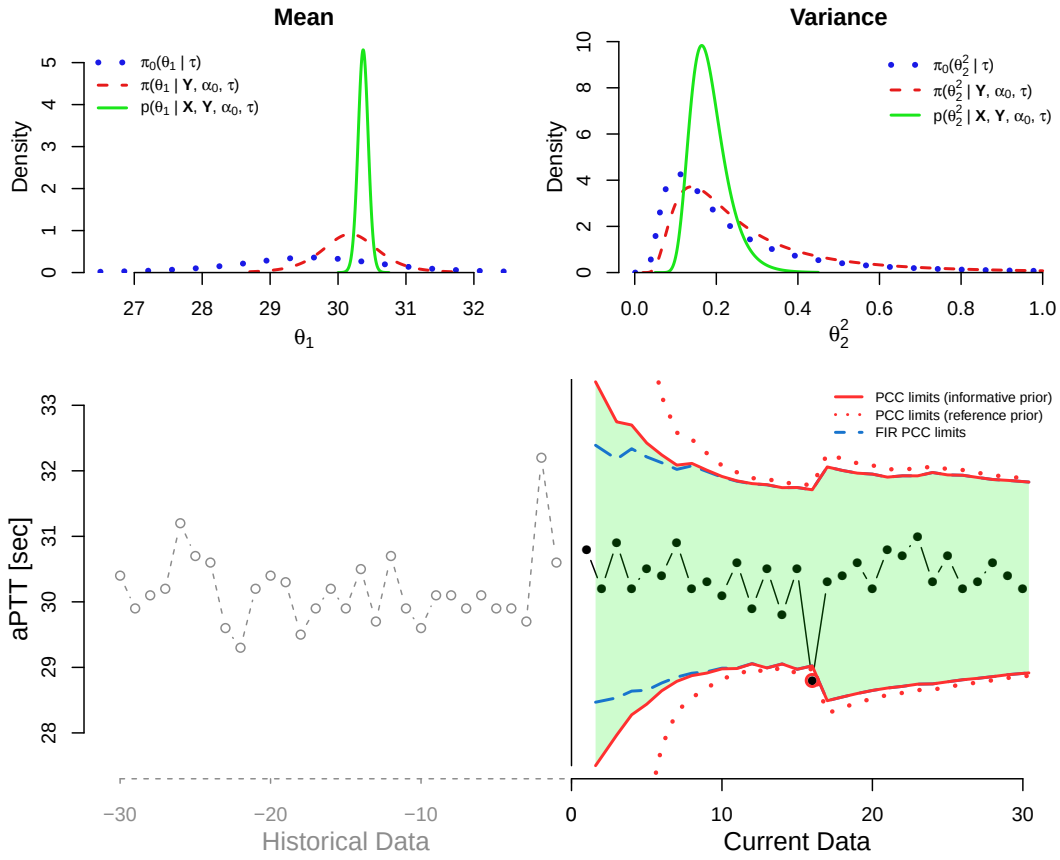


Figure 2.5.1: The PCC application on Normal data. At the upper panels (left and right), we have the marginal distributions for the mean and the variance respectively. With the dotted, dashed and solid lines we denote the initial prior, the power prior and the posterior after gathering all the current data respectively. At the lower panels, we provide the time series of the historical data (open circles on left) and of the current data (solid points on the right). The solid lines represent the limits of PCC, the dotted lines are the limits of PCC under prior ignorance, i.e. using the initial reference prior and the dash lines correspond to the FIR adjustment, setting $f = 0.99$ and $\mathrm{a} = (-3/log_{10}(1-f)-1)/4 = 0.125$.

2.5.2 PCC application to Poisson data

Next, we provide an illustration of PCC for discrete (Poisson) data. The data come from Hansen and Ghare (1987) and were also analyzed by Bayarri and García-Donato

(2005). They refer to the number of defects (x_i), per inspected number of units (s_i), encountered in a complex electrical equipment of an assembly line. We have 25 counts (see Table 2.5.2) arriving sequentially that we will model using the Poisson distribution with unknown rate parameter, i.e. $X_i|\theta \sim P(\theta \cdot s_i)$. In contrast to the previous application, neither prior information regarding the unknown parameter nor historical data exist. Therefore, we use the reference prior as initial prior for θ, i.e. $\pi_0(\theta|\boldsymbol{\tau}) \propto 1/\sqrt{\theta} \equiv G(1/2, 0)$ and we also set $\alpha_0 = 0$ for the power prior term.

Inspected units $(s_1 - s_{13})$	4	7	5	7	7	7	6	7	7	6	8	6	3
Defect counts $(x_1 - x_{13})$	17	23	24	27	32	33	18	28	29	31	39	29	30
Inspected units $(s_{14} - s_{25})$	8	9	6	7	5	7	3	6	8	8	7	8	
Defect counts $(x_{14} - x_{25})$	31	21	26	20	24	29	15	32	20	24	24	14	

Table 2.5.2: Number of defects (x_i) and inspected units (s_i) per time point ($i = 1, 2, \ldots, 25$), in an assembly line of an electrical equipment.

In Figure 2.5.2, we provide the initial prior and posterior distributions, the plot of the data, (daily rate of defects i.e. total number of defects per number of inspected units and number of inspected units) and the Poisson based PCC (the wavy form of the limits is caused by the variation in the number of inspected units we have per day).

Similarly to what we mentioned earlier, the posterior and the PCC will start from times 1 and 2 respectively and will be updated sequentially, every time a new data point becomes available, offering online inference in controlling the process. PCC raises two alarms, at locations 13 and 25. In the former, the observed rate (30/3=10) seems to be higher (process degradation) from what it was expected from the process as it was evolving till that time, while the latter indicates that the observed rate (14/8=1.75) was smaller from what PCC was anticipating (process improvement). Similar to the previous application, the fact that the alarms were kept in the process inflated the subsequent limits. At last, online inference regarding the unknown Poisson rate parameter is available via its (sequentially updated) posterior distribution.

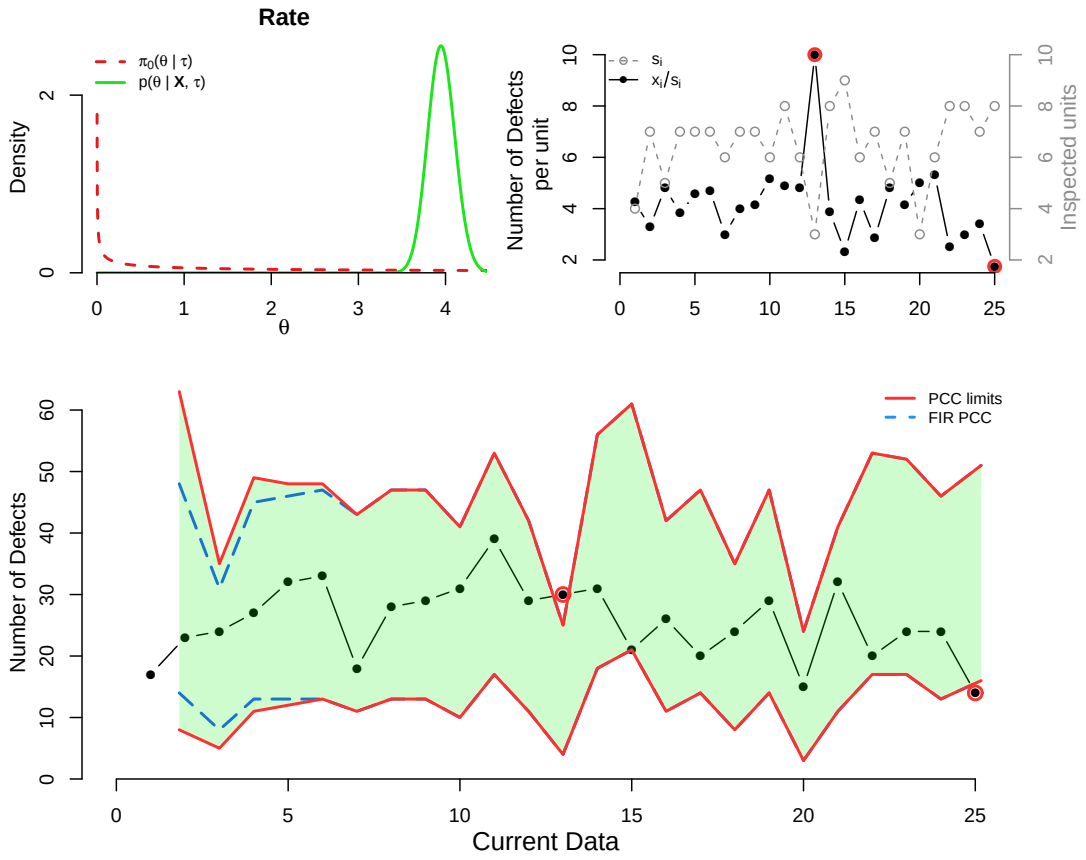


Figure 2.5.2: The PCC application on Poisson data. At the upper left panel we have the distributions for the rate parameter. With the dashed and solid lines we denote the prior and posterior distributions respectively, after gathering all the available data. At the upper right panel, we provide the number of inspected units s_i (dashed line) and the number of defects per size x_i/s_i, i.e the rate of defects (solid line), whereas at the lower panel we present the PCC implementation. Specifically, solid lines correspond to the standard PCC process, while the dashed represent the PCC based on FIR adjustment, setting $f = 0.95$ and $\mathrm{a} = \left(-3/log_{10}(1-f) - 1\right)/4 \approx 0.326$.

Chapter 3

Predictive Ratio CUSUM (PRC)

3.1 PRC Theoretical background

In this chapter, the focus is on detecting medium/small persistent parameter shifts in short horizon data. In the literature there are two standard methods that could be employed in such a setup: the frequentist self-starting CUSUM (SSC) of Hawkins and Olwell (1998) and the Bayesian Cumulative Bayes Factor (CBF) of West (1986) and West and Harrison (1986). We will propose a self-starting Bayesian scheme named Predictive Ratio Cusum (PRC), which much like the PCC methodology, will be based on the use of the predictive distribution.

PRC will provide an enhanced Bayesian analogue of SSC and at the same time provide an antagonistic to CBF method which will differentiate from CBF in three ways. Firstly, PRC will examine specific alternative hypotheses as OOC scenarios (much like it is done for traditional CUSUM in SPC/M), in contrast to the diffused West's CBF (neutral) alternatives. Secondly, PRC will be formulated for various discrete and continuous distributions that are members of the regular exponential family, providing a closed form mechanism (i.e. easy to be used in practice), capable to examine a variety of standard OOC scenarios considered in SPC/M. Last but most impor-

tantly, by proposing a procedure that derives a decision making threshold based on the false alarm tolerance (something that both CBF and SSC are typically lacking), again allowing its straightforward use to real life problems. In the same spirit to PCC, in this chapter, we will provide the technical details, develop a FIR option, evaluate the performance against competition, examine topics regarding sensitivity and robustness and conclude with illustrations to real datasets.

3.1.1 PRC for k-parameter regular exponential family (k-PREF)

As we mentioned, PRC is a Bayesian CUSUM type chart, comparing the IC state against an OOC scenario. It is of great importance to clarify that the null state (IC) is not fixed, but sequentially updated, every time a new data point arrives. Likewise, the alternative (OOC) scenario cannot be fixed, but it should be constructed sequentially and designed suitably in order to increase the detection power. West (1986), suggested to derive a neutral alternative (OOC) hypothesis scenario, by intervening to the most recent posterior parameters, $\boldsymbol{\tau}_n$, in such a way that we reserve the same location, but we inflate the variance, getting a more diffused (spread out) predictive distribution. Despite the indisputable convenience of that choice, there is significant room for improvement, at least within the SPC/M methodological framework. In particular, the adoption of an alternative informative scenario with shifted parameters, which simply yields a benchmark of the OOC state, can greatly improve detection power. Typically, the kind of persistent shifts that we aim to detect (like a mean jump, a variance/rate inflation, etc.), can be predetermined and arise from the nature of the process along with what is considered as process deterioration/improvement. This is a well known strategy in SPC/M, where charts can be built with a specific OOC state in mind (like the traditional CUSUM).

For PRC, we keep the same general distributional setup, as we had in Chapter 2

for PCC, where the likelihood belongs to the k-PREF family, and we use a conjugate power prior. The posterior parameters $\boldsymbol{\tau}_n$ in Theorem 2.1.1 summarize all the information regarding the unknown parameter(s) $\boldsymbol{\theta}$ at time n, as they consist of the initial prior hyperparameters and the sufficient statistics of the current and the (possibly available) historical data. Our recommendation in PRC is to adopt informative OOC scenarios (typically used in SPC/M), for the unknown parameter(s) $\boldsymbol{\theta}$, resulting an intervention to the most recent posterior distribution parameters $\boldsymbol{\tau}_n$. In this manner, we provide an antagonistic methodology of CBFs and simultaneously develop an effective Bayesian alternative of the SSC.

The choice of the unknown parameter shifts, will be expressed in a way that preserves conjugacy, allowing closed form solutions, while reflecting our perspective for the OOC state. For most of the cases, where the posterior distribution (or the posterior marginal, if $\boldsymbol{\theta}$ is multivariate) is a member of a location or scale family, we will consider shifts that represent location or scale transformation of the unknown parameter respectively. This will guarantee that we remain in the same distribution with updated parameters $\boldsymbol{\tau}_n'$, derived as simple location or scale transformations of the IC state posterior parameters $\boldsymbol{\tau}_n$. The only exception occurs in the case of a Beta posterior (resulting in Binomial and Negative Binomial likelihood settings), which is neither location nor scale family. For the Beta posterior, our proposal will be to introduce the OOC shift, not on θ but on the expected posterior odds, i.e. $E_{\theta|\boldsymbol{X}}\left[\theta/(1-\theta)\right]$. Table 3.1.1 reports the IC and OOC states of the unknown parameter(s) $\boldsymbol{\theta}$, along with the relevant interpretation, for various likelihood choices from the k-PREF that are commonly used in SPC/M.

As it was mentioned earlier, PRC will be based on the predictive distribution of the next unseen data $f\left(X_{n+1}|\boldsymbol{X},\boldsymbol{Y},\alpha_0,\boldsymbol{\tau}\right)$, or for simplicity $f\left(X_{n+1}|\boldsymbol{X}_n\right)$. In Theorem 2.1.1, if we will replace the current posterior distribution, $\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}_n\right)$, with the OOC posterior, $\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}_n'\right)$, corresponding to the shifted parameter scenario, we will derive

the shifted (OOC) predictive distribution:

$$f'\left(X_{n+1}|\boldsymbol{X}_n\right)=\int\limits_{\boldsymbol{\Theta}} f\left(X_{n+1}|\boldsymbol{\theta}\right)\pi_0\left(\boldsymbol{\theta}|\boldsymbol{\tau}_n'\right)d\boldsymbol{\theta}=\frac{K\left(\boldsymbol{\tau}_n'+\boldsymbol{t}_f(X_{n+1})\right)}{K\left(\boldsymbol{\tau}_n'\right)}g(X_{n+1}) \quad (3.1.1)$$

PRC is based on the sequential comparison (via their ratio) between the current predictive distribution $f\left(X_{n+1}|\boldsymbol{X}_n\right)$, which includes all the relevant information from the process up to the current time, and the corresponding shifted predictive, $f'\left(X_{n+1}|\boldsymbol{X}_n\right)$, representing the OOC shifted parameter scenario. The ratio of the shifted predictive over the current predictive for X_{n+1} will be:

$$L_{n+1}=\frac{f'\left(X_{n+1}|\boldsymbol{X}_n\right)}{f\left(X_{n+1}|\boldsymbol{X}_n\right)}=\frac{\frac{K\left(\boldsymbol{\tau}_n'+\boldsymbol{t}_f(X_{n+1})\right)}{K\left(\boldsymbol{\tau}_n'\right)}g(X_{n+1})}{\frac{K\left(\boldsymbol{\tau}_n+\boldsymbol{t}_f(X_{n+1})\right)}{K\left(\boldsymbol{\tau}_n\right)}g(X_{n+1})}=\frac{K\left(\boldsymbol{\tau}_n'+\boldsymbol{t}_f(X_{n+1})\right)\cdot K\left(\boldsymbol{\tau}_n\right)}{K\left(\boldsymbol{\tau}_n+\boldsymbol{t}_f(X_{n+1})\right)\cdot K\left(\boldsymbol{\tau}_n'\right)}, \quad (3.1.2)$$

In general, the predictive distribution becomes available after the first observation, except when we have Normal/Lognormal likelihood with both parameters unknown and total prior ignorance (i.e. no historical data, so $\alpha_0=0$ and we use the non-informative reference prior as initial prior), where the predictive requires two observations to become proper. PRC will build up evidence by monitoring the log-ratio of predictive densities, $log(L_{n+1})$, using a CUSUM. Precisely, starting with $S_1=0$ (or $S_1=S_2=0$, when we have two unknown parameters and total prior ignorance), the one sided PRC statistic at time $n+1$ will be:

$$S_{n+1}=\max\{0,S_n+log\left(L_{n+1}\right)\} \quad \text{or} \quad S_{n+1}=\min\{0,S_n-log\left(L_{n+1}\right)\} \quad (3.1.3)$$

when we are interested in detecting upward or downward shifts respectively. Controlling S_{n+1} is performed in the same spirit as in traditional CUSUM, where an alarm is raised when the cumulative statistic exceeds an appropriately selected threshold value (also known as decision interval). Thus, the suggested control chart, will plot

S_{n+1} versus the order of the data, having a horizontal line at height h to denote the predetermined decision threshold, which we will derive in Section 3.2. An alarm will be ringed, each time the statistic S_{n+1} will plot beyond h.

From a Bayesian perspective, (3.1.2) is simply the predictive Bayes Factor at time $n+1$, comparing the OOC model $(\mathcal{M}_1 : f'(X_{n+1}|\boldsymbol{X}_n))$ against the IC model $(\mathcal{M}_0 : f(X_{n+1}|\boldsymbol{X}_n))$, i.e. $L_{n+1} = B_{10}^{n+1}$. Therefore, the statistic S_{n+1} can be written as:

$$\begin{aligned} S_{n+1} &= \max\left\{0, S_n + log\left(B_{10}^{n+1}\right)\right\} = \max\left\{0, \sum_{i=\kappa}^{n} log\left(B_{10}^{i+1}\right)\right\} \quad \text{or} \\ S_{n+1} &= \min\left\{0, S_n - log\left(B_{10}^{n+1}\right)\right\} = \min\left\{0, \sum_{i=\kappa}^{n} -log\left(B_{10}^{i+1}\right)\right\} \end{aligned} \tag{3.1.4}$$

for the upward or downward shifts respectively, where κ $(1 \leq \kappa \leq n)$ is the last time for which the monitoring statistic was equal to zero (i.e. $S_\kappa = 0$ and $\forall \ell > \kappa$ we have $|S_\ell| > 0$). In other words, S_{n+1} represents the most recent cumulative logarithmic Bayes Factor evidence, a quantity that is known in the Bayesian decision theory framework to provide a summary of evidence for the alternative (OOC) $\mathcal{M}_1$ against the (IC) null $\mathcal{M}_0$ model (West, 1986).

The designed OOC parameter shifts, along with the exact formula of the $log(L_{n+1})$ statistic used in PRC, can be found in Table 3.1.1, for various likelihood choices (of discrete and continuous univariate data) that belong to the k-PREF and are commonly used in SPC/M. To unify notation, we denote by $\boldsymbol{D}_n = (\boldsymbol{Y}, \boldsymbol{X}_n) = (y_1, \ldots, y_{n0}, x_1, \ldots, x_n)$ the vector of historical and current data, $\boldsymbol{w} = (\alpha_0, \ldots, \alpha_0, 1, \ldots, 1)$ the vector of weights corresponding to each element d_j of $\boldsymbol{D}_n$ and we call $N_D = n_0 + n$ the length of the data vector $\boldsymbol{D}_n$. Technical details regarding the derivation of all these PRC models are available in Appendix A.

Likelihood Prior	**OOC State**	**Interpretation**	$log\left(L_{n+1}\right)$
$\boldsymbol{P}(\theta \cdot s_i)$ $G(c,d)$	$k \cdot \theta$	$k>1$: $(k-1)100\%$ increase in θ $k<1$: $(1-k)100\%$ decrease in θ	$(\hat{c}_n + x_{n+1})\, log\dfrac{\hat{d}_n + s_{n+1}}{\hat{d}_n/k + s_{n+1}} + x_{n+1} \cdot logk$
	$\hat{c}_n = c + \sum_{j=1}^{N_D} w_j d_j, \qquad \hat{d}_n = d + \sum_{j=1}^{N_D} w_j s_j$		
$\boldsymbol{Bin}(N_i,\theta)$ $Beta(a,b)$	$k \cdot E_{\theta\mid\boldsymbol{x}}\left(\dfrac{\theta}{1-\theta}\right)$	$k>1$: $(k-1)100\%$ increase in expected odds of θ $k<1$: $(1-k)100\%$ decrease in expected odds of θ	$log\dfrac{B\left(k\cdot\hat{a}_n+\hat{b}_n, N_{n+1}\right)\cdot B\left(\hat{a}_n, x_{n+1}\right)}{B\left(\hat{a}_n+\hat{b}_n, N_{n+1}\right)\cdot B\left(k\cdot\hat{a}_n, x_{n+1}\right)}$
	$\hat{a}_n = a + \sum_{j=1}^{N_D} w_j d_j, \qquad \hat{b}_n = b + \sum_{j=1}^{N_D} w_j N_j - \sum_{j=1}^{N_D} w_j d_j$		
$\boldsymbol{NBin}(r,\theta)$ $Beta(a,b)$	$k \cdot E_{\theta\mid\boldsymbol{x}}\left(\dfrac{\theta}{1-\theta}\right)$	$k>1$: $(k-1)100\%$ increase in expected odds of θ $k<1$: $(1-k)100\%$ decrease in expected odds of θ	$log\dfrac{B\left(k\cdot\hat{a}_n+\hat{b}_n, r+x_{n+1}\right)\cdot B\left(\hat{a}_n, r\right)}{B\left(\hat{a}_n+\hat{b}_n, r+x_{n+1}\right)\cdot B\left(k\cdot\hat{a}_n, r\right)}$
	$\hat{a}_n = a + r\sum_{j=1}^{N_D} w_j, \qquad \hat{b}_n = b + \sum_{j=1}^{N_D} w_j d_j$		

$\boldsymbol{N}(\theta, \sigma^2)$ $N(\mu_0, \sigma_0^2)$	$\theta + k \cdot \sigma$	$k > 0$: increase in θ size of $k \cdot \sigma$ $k < 0$: decrease in θ size of $k \cdot \sigma$	$\left(z_{n+1} - \frac{k}{2} \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}\right) \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}$

$$\hat{\mu}_n = \frac{\sigma^2 \mu_0 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j d_j}{\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j}, \qquad \hat{\sigma}_n^2 = \frac{{\sigma_0}^2 \sigma^2}{\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j}, \qquad z_{n+1} = \frac{x_{n+1} - \hat{\mu}_n}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}$$

$\boldsymbol{N}(\mu, \theta^2)$ $IG(a, b)$	$k \cdot \theta^2$	$k > 1$: $(k-1)100\%$ increase in θ^2 $k < 1$: $(1-k)100\%$ decrease in θ^2	$(\hat{a}_n + 1/2) \cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$

$$\hat{a}_n = a + \frac{\sum_{j=1}^{N_D} w_j}{2}, \qquad \hat{b}_n = b + \frac{\sum_{j=1}^{N_D} w_j \left(d_j - \mu\right)^2}{2}, \qquad z_{n+1} = \frac{x_{n+1} - \mu}{\sqrt{\hat{b}_n / \hat{a}_n}}$$

$\boldsymbol{N}(\theta_1, \theta_2^2)$ $NIG(\mu_0, \lambda, a, b)$	$\theta_1 + k \cdot \hat{\theta}_2$	$k > 0$: increase in θ size of $k \cdot \hat{\theta}_2$ $k < 0$: decrease in θ size of $k \cdot \hat{\theta}_2$	$(\hat{a}_n + 1/2) \cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + \left(z_{n+1} - k \cdot \hat{\lambda}_n / (\hat{\lambda}_n + 1)\right)^2}$
$\boldsymbol{N}(\theta_1, \theta_2^2)$ $NIG(\mu_0, \lambda, a, b)$	$k \cdot \theta_2^2$	$k > 1$: $(k-1)100\%$ increase in θ_2^2 $k < 1$: $(1-k)100\%$ decrease in θ_2^2	$(\hat{a}_n + 1/2) \cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$

$$\hat{\mu}_n = \frac{\lambda\mu_0 + \sum_{j=1}^{N_D} w_j d_j}{\lambda + \sum_{j=1}^{N_D} w_j}, \quad \hat{\lambda}_n = \lambda + \sum_{j=1}^{N_D} w_j, \quad \hat{a}_n = a + \frac{\sum_{j=1}^{N_D} w_j}{2}, \quad \hat{b}_n = b + \frac{\lambda\mu_0^2 + \sum_{j=1}^{N_D} w_j d_j^2}{2} - \frac{\left(\lambda\mu_0 + \sum_{j=1}^{N_D} w_j d_j\right)^2}{2\left(\lambda + \sum_{j=1}^{N_D} w_j\right)}, \quad \hat{\theta}_2 = \sqrt{\frac{\hat{b}_n}{\hat{a}_n}}, \quad z_{n+1} = \frac{x_{n+1} - \hat{\mu}_n}{\sqrt{\frac{\left(\hat{\lambda}_n + 1\right) \cdot \hat{b}_n}{\hat{\lambda}_n \cdot \hat{a}_n}}}$$

$\boldsymbol{G}(\alpha,\theta)$ $G(c,d)$	$k\cdot\theta$	$k>1$: $(k-1)100\%$ increase in θ $k<1$: $(1-k)100\%$ decrease in θ	$(\hat{c}_n+\alpha)\cdot log\frac{\hat{d}_n+x_{n+1}}{\hat{d}_n+k\cdot x_{n+1}}+\alpha\cdot logk$
$\hat{c}_n=c+\alpha\sum_{j=1}^{N_D}w_j,\qquad \hat{d}_n=d+\sum_{j=1}^{N_D}w_jd_j$			
$\boldsymbol{Weibull}(\theta,\kappa)$ $IG(a,b)$	$k\cdot\theta^{\kappa}$	$k>1$: $(k-1)100\%$ increase in θ^{κ} $k<1$: $(1-k)100\%$ decrease in θ^{κ}	$(\hat{a}_n+1)\cdot log\frac{\hat{b}_n+x_{n+1}^{\kappa}}{\hat{b}_n+x_{n+1}^{\kappa}/k}-logk$
$\hat{a}_n=a+\sum_{j=1}^{N_D}w_j,\qquad \hat{b}_n=b+\sum_{j=1}^{N_D}w_jd_j^{\kappa}$			
$\boldsymbol{IG}(\alpha,\theta)$ $G(c,d)$	$k\cdot\theta$	$k>1$: $(k-1)100\%$ increase in θ $k<1$: $(1-k)100\%$ decrease in θ	$(\hat{c}_n+\alpha)\cdot log\frac{\hat{d}_n\cdot x_{n+1}+1}{\hat{d}_n\cdot x_{n+1}+k}+\alpha\cdot logk$
$\hat{c}_n=c+\alpha\sum_{j=1}^{N_D}w_j,\qquad \hat{d}_n=d+\sum_{j=1}^{N_D}\frac{w_j}{d_j}$			
$\boldsymbol{Pareto}(m,\theta)$ $G(c,d)$	$k\cdot\theta$	$k>1$: $(k-1)100\%$ increase in θ $k<1$: $(1-k)100\%$ decrease in θ	$(\hat{c}_n+1)\cdot log\frac{\hat{d}_n+log(x_{n+1}/m)}{\hat{d}_n+k\cdot log(x_{n+1}/m)}+logk$
$\hat{c}_n=c+\sum_{j=1}^{N_D}w_j,\qquad \hat{d}_n=d+\sum_{j=1}^{N_D}\frac{w_j}{log(d_j/m)}$			

$\boldsymbol{logN}(\theta, \sigma^2)$ $N(\mu_0, \sigma_0^2)$	$\theta + k \cdot \sigma$	$k > 0$: increase in θ size of $k \cdot \sigma$ $k < 0$: decrease in θ size of $k \cdot \sigma$	$\left(z_{n+1} - \frac{k}{2} \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}\right) \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}$

$$\hat{\mu}_n = \frac{\sigma^2 \mu_0 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j log\,(d_j)}{\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j}, \qquad \hat{\sigma}_n^2 = \frac{{\sigma_0}^2 \sigma^2}{\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j}, \qquad z_{n+1} = \frac{log(x_{n+1}) - \hat{\mu}_n}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}$$

$\boldsymbol{logN}(\mu, \theta^2)$ $IG(a, b)$	$k \cdot \theta^2$	$k > 1$: $(k-1)100\%$ increase in θ^2 $k < 1$: $(1-k)100\%$ decrease in θ^2	$(\hat{a}_n + 1/2) \cdot log \frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$

$$\hat{a}_n = a + \frac{\sum_{j=1}^{N_D} w_j}{2}, \qquad \hat{b}_n = b + \frac{\sum_{j=1}^{N_D} w_j \left(log\,(d_j) - \mu\right)^2}{2}, \qquad z_{n+1} = \frac{log(x_{n+1}) - \mu}{\sqrt{\hat{b}_n / \hat{a}_n}}$$

$\boldsymbol{logN}(\theta_1, \theta_2^2)$ $NIG(\mu_0, \lambda, a, b)$	$\theta_1 + k \cdot \hat{\theta}_2$	$k > 0$: increase in θ size of $k \cdot \hat{\theta}_2$ $k < 0$: decrease in θ size of $k \cdot \hat{\theta}_2$	$(\hat{a}_n + 1/2) \cdot log \frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + \left(z_{n+1} - k \cdot \hat{\lambda}_n / (\hat{\lambda}_n + 1)\right)^2}$
$\boldsymbol{logN}(\theta_1, \theta_2^2)$ $NIG(\mu_0, \lambda, a, b)$	$k \cdot \theta^2$	$k > 1$: $(k-1)100\%$ increase in θ_2^2 $k < 1$: $(1-k)100\%$ decrease in θ_2^2	$(\hat{a}_n + 1/2) \cdot log \frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$

$$\hat{\mu}_n = \frac{\lambda\mu_0 + \sum_{j=1}^{N_D} w_j log(d_j)}{\lambda + \sum_{j=1}^{N_D} w_j}, \quad \hat{\lambda}_n = \lambda + \sum_{j=1}^{N_D} w_j, \quad \hat{a}_n = a + \frac{\sum_{j=1}^{N_D} w_j}{2}, \quad \hat{b}_n = b + \frac{\lambda\mu_0^2 + \sum_{j=1}^{N_D} w_j \left(log(d_j)\right)^2}{2} - \frac{\left(\lambda\mu_0 + \sum_{j=1}^{N_D} w_j log(d_j)\right)^2}{2\left(\lambda + \sum_{j=1}^{N_D} w_j\right)}, \quad \hat{\theta}_2 = \sqrt{\frac{\hat{b}_n}{\hat{a}_n}}, \quad z_{n+1} = \frac{log(x_{n+1}) - \hat{\mu}_n}{\sqrt{\frac{\left(\hat{\lambda}_n + 1\right) \cdot \hat{b}_n}{\hat{\lambda}_n \cdot \hat{a}_n}}}$$

Table 3.1.1: The PRC scheme using an initial conjugate prior in a power prior, for some of the univariate distributions typically used in SPC/M, belonging to the k-PREF. $\boldsymbol{D}_n = (\boldsymbol{Y}, \boldsymbol{X}_n) = (y_1, \ldots, y_{n_0}, x_1, \ldots, x_n)$ is the vector of historical and current data, $\boldsymbol{w} = (\alpha_0, \ldots, \alpha_0, 1, \ldots, 1)$ are the weights corresponding to each element d_j of $\boldsymbol{D}_n$ and $N_D = n_0 + n$.

3.1.2 Fast Initial Response (FIR) PRC

It is well known that the self-starting memory based charts have a weak response to shifts arriving early in the process. Failing to react to an early shift will lead to its absorption, contaminating the calibration and reducing the testing performance. Lucas and Crosier (1982) were the first to introduce the Fast Initial Response (FIR) feature for CUSUM, by adding a constant value to the initial cumulative statistic, enhancing its reaction to very early shifts in the process. Steiner (1999) introduced the FIR EWMA by narrowing its control limits, with the effect of this adjustment decreasing exponentially fast. For PRC, we propose an exponentially decreasing adjustment, multiplied to the statistic $log(L_{n+1})$. Specifically, the adjustment (inflation) will be:

$$\text{FIR}_{adj} = 1 + f \cdot d^{(t-1)}, \tag{3.1.5}$$

where t is the time of the examined predictive ratio, $f \geq 0$ represents the proportion of the inflation for the PRC statistic, $log(L_{n+1})$, when $t = 1$ and $0 < d < 1$ is a smoothing parameter, specifying the exponential decay of the adjustment (the smaller the d the fastest the decay). As the first predictive ratio is available for the second observation, we have $t = 1$ when $n = 2$. The only exception is when we have two unknown parameters and total prior ignorance (i.e. use of initial reference prior and $\alpha_0 = 0$ due to lack of historical data), where we get $t = 1$ when $n = 3$.

The proposed FIR adjustment is more flexible compared to the fixed constant of Lucas and Crosier (1982) FIR-CUSUM, as it allows to control the influence, by tuning the initial parameters (f, d) providing a better interpretation. The FIR option can improve the performance at the early start, but the choice of the adjustment parameters must be prudent, avoiding to inflate significantly the false alarm rate. The expected number of false alarms for PRC will depend on the prior settings, especially when the volume of available data is small. Our suggestion is to be somewhat conservative, especially when a weakly informative prior is used, so that the

FIR adjustment will not seriously affect the predetermined expected number of false alarms. In general, we recommend to use $(f, d) = (1/2, 3/4)$, where the adjusted $log(L_{n+1})$ will be inflated by 50% for the first test, while the inflation will be only 5% at the ninth test (the choice of these parameters will reflect the user's based needs).

3.2 PRC design and Inference

3.2.1 Tuning the PRC

PRC is simply a sequential hypothesis testing procedure, where two competing states of the predictive are compared via their log-predictive ratio, within a memory based (CUSUM) control scheme. In the ratio, the denominator refers to the running (considered as IC) predictive model, while the numerator is the intervened (considered as OOC) competing model. Our goal is to detect a transition from the IC to the OOC model, as soon as it occurs, while keeping the false alarms at a low predetermined level.

For the classical CUSUM process, where both IC and OOC models have all parameters known, certain optimality properties have been derived (like in Moustakides, 1986 or Ritov, 1990) along with theoretical results regarding the choice of the design parameters. Namely, numerical algorithms have been developed to compute the IC Average Run Length, ARL_0 (i.e. the expected number of observations before the occurrence of the first false alarm), as in Brook and Evans (1972). However, such algorithms are not applicable to self-starting setups, where both the IC and OOC distributions include unknown parameter(s) that we estimate online (i.e. these distributions are not fixed, but sequentially updated).

When PRC alarms, the process should be stopped and examined thoroughly (triggering a potential corrective action), preventing further contaminated data from joining

the calibration step. We will define as stopping time T of a PRC, tuned for an upward shift:

$$T = \inf\{n+1 : S_{n+1} \geq h\} \tag{3.2.1}$$

where $n+1 \geq 2$, except the special case with two unknown parameters and complete prior ignorance, where we have $n+1 \geq 3$, while $h > 0$ is a preselected constant to guarantee a predetermined false alarm standard (for downward shifts in (3.2.1) we have $S_{n+1} \leq h$, with $h < 0$). The choice of h reflects on the tolerance that we have on false alarms, measured via either the Family Wise Error Rate ($FWER$), for a fixed and not too long horizon of N data points or the In Control Average Run Length (ARL_0), when we have an unknown or a large N scenario. Due to the general form of PRC's mechanism, that allows hosting any distribution from the k-PREF, there is no single optimal strategy in selecting h. In what follows, we will provide specific guidelines for the selection of h, utilizing the distributional setup under study.

Scenario 1: *the predictive distribution is a location-scale family.*
In this case we will derive h via the standard predictive distribution (i.e. the distribution with location=0 and scale=1). Then, at each step of PRC we will perform the same location-scale transformation to both the IC and OOC predictive distributions, so that each time the IC, $f(X_{n+1}|\boldsymbol{X}_n)$, becomes the standard predictive (note that the location-scale transformation will be different at each step). The transformed predictives will be used in the ratio (3.1.2). From the distributions in Table 3.1.1 the location-scale predictive is valid for the Normal and logarithmic transformed Lognormal likelihood cases, where the logarithm of the standardized predictive ratio (3.1.2) is tabulated. Algorithms 3 and 4 can be used to derive h, when we work with either the $FWER$ for a fixed horizon of N data or when we use ARL_0 metric respectively. Furthermore, in Appendix B we provide a table with the derived h threshold values for various choices of $(N, FWER)$ or ARL_0 values, combined with specific OOC shift sizes k, when we have total prior ignorance (i.e. use of initial reference prior and no historical data).

Algorithm 3 Determine PRC's decision limit h based on $FWER$

Define the length of the data, N, for which PRC will be employed {*initial input*}
Define the $FWER$ that we aim to have at the N-th data point
Define the vector $\boldsymbol{\tau}'_n$, which represents the OOC disturbance that we wish to detect
Define the number of iterations, I, used in the empirical estimation
if {predictive distribution is a location-scale family} **then**
    $f(X)$ = the standard distribution { *loc.=0, sc.=1 and* $df_n = 2\hat{a}_n$ *if* $X \sim t$ }
else
    $f(X)$ = the marginal (prior predictive) distribution from (11)
end if
Generate a matrix D of dimension $I \times N$ with random numbers from $f(X)$
Set S to be a matrix of dimension $I \times N$ filled with zeros
Set M to be a vector of dimension I filled with NAs
for {i in $1 : I$}
    for {n in $1 : (N-1)$}
        $L_{n+1} \leftarrow \dfrac{f'(D[i,n+1] \mid D[i,1],\ldots,D[i,n])}{f(D[i,n+1] \mid D[i,1],\ldots,D[i,n])}$ {*Predictive ratio*}
        $S[i,n+1] = \max\{0, S[i,n] + log(L_{n+1})\}$
        (or S[i,n+1] = $\min\{0, S[i,n] - log(L_{n+1})\}$ for downward shifts) {*PRC statistic*}
    end for
    $M[i] \leftarrow \max\{S[i,]\}$
    (or $M[i] \leftarrow \min\{S[i,]\}$ for downward shifts)
end for
$H \leftarrow \hat{F}_I^{-1}(1 - FWER)$
(or $H \leftarrow \hat{F}_I^{-1}(FWER)$ for downward shifts) { $\hat{F}_I(x) = \dfrac{1}{I}\sum_{i=1}^{I} \mathbb{1}\{M[i] < x\}$ }
if {predictive distribution is a location-scale family} **then**
    $h \leftarrow H$ {*empirical estimate of* h}
else
    $h_m \leftarrow H$ {*marginal based (conservative) empirical estimate of* h}
end if

Algorithm 4 Determine PRC's decision limit h based on ARL_0

1: Define the ARL_0 that you aim to have
2: Define the numerical tolerance tol, which represents the maximum of error estimate
3: Define the vector $\boldsymbol{\tau}'_n$, which represents the OOC disturbance that you wish to detect
4: Define the number of iterations I, used in the empirical estimation
5: **if** {predictive distribution is a location scale family} **then**
6: $f(X)$ = the standard distribution { *loc.=0, sc.=1 and* $df_n = 2\hat{a}_n$ *if* $X \sim t$ }
7: **else**
8: $f(X)$ = the marginal (prior predictive) distribution from (11)
9: **end if**
10: **start function** $ARL(h)$
11: Set M to be a vector of dimension I filled with NAs
12: **for** $\{i$ in $1 : I\}$
13: Set $S \leftarrow 0$
14: Set $n \leftarrow 1$
15: Generate $x_n \sim f(X)$
16: **while** $\{S < h$ (or $S > h$ for downward shifts)$\}$
17: Generate $x_{n+1} \sim f(X)$
18: $L_{n+1} \leftarrow \dfrac{f'(x_{n+1}|x_1, \ldots, x_n)}{f(x_{n+1}|x_1, \ldots, x_n)}$ { *Predictive ratio* }
19: $S \leftarrow \max\{0, S + log(L_{n+1})\}$
(or $S \leftarrow \min\{0, S - log(L_{n+1})\}$ for downward shifts) { *PRC statistic* }
20: Set $n \leftarrow n + 1$
21: **end while**
22: $M[i] \leftarrow n$
23: **end for**
24: **return** $\{ARL(h) \leftarrow \bar{M}\}$ { $\bar{M} = \dfrac{1}{I}\sum_{i=1}^{I} M[i]$ }
25: **end function** $ARL(h)$
26: Set $h_1 = 2$ (or $h_1 = -2$ for downward shifts) the first initial value for h (or h_m)
27: Get $ARL(h_1)$ { *use of function* $ARL(h)$}
28: **if** $\{|ARL(h_1) - ARL_0| < tol\}$
29: $H \leftarrow h_1$
30: **goto** 48
31: **end if**
32: Set $h_2 = 4$ (or $h_1 = -4$ for downward shifts) the second initial value for h (or h_m)
33: Get $ARL(h_2)$ { *use of function* $ARL(h)$}
34: **if** $\{|ARL(h_2) - ARL_0| < tol\}$
35: $H \leftarrow h_2$
36: **goto** 48
37: **end if**

$H \leftarrow h_2 + (ARL_0 - ARL(h_2)) \cdot \dfrac{(h_2 - h_1)}{(ARL(h_2) - ARL(h_1))}$ { *regula falsi estimate*}
Get $ARL(H)$ { *use of function ARL(h)*}
while $\{|ARL(H) - ARL_0| > tol\}$
    $H \leftarrow h_2 + (ARL_0 - ARL(h_2)) \cdot \dfrac{(h_2 - h_1)}{(ARL(h_2) - ARL(h_1))}$ { *regula falsi estimate*}
    Get $ARL(H)$ { *use of function ARL(h)*}
    $h_1 \leftarrow h_2$
    $h_2 \leftarrow H$
end while
if {predictive distribution is a location scale family} **then**
    $h \leftarrow H$ { *empirical estimation*}
else
    $h_m \leftarrow H$ { *marginal based (conservative) empirical estimation*}
end if

Scenario 2: *the predictive is not location-scale family, but we have an informative prior.*

The unknown parameter(s) and the lack of standardization (since we do not have location-scale family) prevent from deriving the sampling predictive distribution as in scenario 1. Our suggestion is to use the marginal (prior predictive) distribution to generate imaginary data. Using the power prior (2.1.3), the general form of the marginal distribution will be available in closed form:

$$f(X|\boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} f(X|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{Y}, \alpha_0, \boldsymbol{\tau})d\boldsymbol{\theta} = \frac{K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_h(\boldsymbol{Y}) + \boldsymbol{t}_f(X)\right)}{K\left(\boldsymbol{\tau} + \alpha_0 \boldsymbol{t}_h(\boldsymbol{Y})\right)} g(X) \tag{3.2.2}$$

The marginal, is a compound distribution of the likelihood and the prior, with the unknown parameter(s) being integrated out. It has heavier tails (greater variance) compared to the likelihood, leading to an estimated decision limit $h_m (\neq h)$ that will result a more conservative $FWER$ or ARL_0 metric. Essentially, the likelihood based threshold h is a limiting case of the marginal-based threshold h_m, when the prior variance tends to zero. Thus, we can generate imaginary data from the marginal, in order to control either the $FWER$ or the ARL_0 and derive the h_m threshold from Algorithm 3 or 4 respectively.

An important issue in this proposal, is that the prior needs to be informative, otherwise the marginal would be too diffused compared to the likelihood, resulting an upper or lower bound h_m that will be too conservative (i.e. $|h_m|$ will seriously overestimate $|h|$, decreasing significantly the false alarms and the detection power). We propose to measure the discrepancy of the likelihood over the marginal variance by:

$$E_{\boldsymbol{\theta}}\left(\frac{Var(X|\boldsymbol{\theta})}{Var(X|\boldsymbol{Y},\alpha_0,\boldsymbol{\tau})}\right)=\rho \tag{3.2.3}$$

The ratio parameter $\rho \leq 1$ expresses the expected underdispersion of the likelihood variance versus the marginal variance. When $\rho \to 1$ (i.e. we use a highly informative prior), then the marginal is a reliable representative of the likelihood, resulting $h_m \to h$. After an extensive simulation study, we recommend to use the marginal distribution approach only when the distributional setting roughly satisfies $\rho \geq 0.9$. Table 3.2.1 provides the formulas for estimating ρ for each of the likelihoods reported in Table 3.1.1 (that do not fall in the location-scale family treated by scenario 1), where for the power prior term, we assume the historical data $\boldsymbol{Y} = (y_1, \ldots, y_{n_0})$, that are weighted by α_0 (for no historical data, set $\alpha_0 = 0$).

In Figure 3.2.1 we provide some illustration of the achieved $FWER$ and ARL_0 metrics in a Poisson and a Binomial likelihood scenario with varying ρ and parameter values, when the OOC scenario was set to shifts of size $k = 2$ (i.e. double the Poisson rate parameter or double the expected odds in the Binomial). The designed performance metrics were $FWER = 5\%$ for a sequence of $N = 50$ observation or $ARL_0 = 100$ while the achieved values in each case were obtained by averaging over 100,000 iterations of the IC process. As the prior becomes more informative, we have $h_m \to h$, where h corresponds to the (unknown) threshold with the designed $FWER$ or ARL_0 performance. The convergence of h_m to h will depend primarily on the value of ρ and to a smaller degree on the actual parameter values.

Scenario 3: *neither the predictive is location-scale family nor we have an informa-*

Likelihood $f(\cdot\|\boldsymbol{\theta})$	**Initial Prior** $\pi_0(\boldsymbol{\theta}\|\boldsymbol{\tau})$	**Expected Ratio (3.2.3)** ρ
$P(\theta \cdot s_i)$	$G(c,d)$	$1-\frac{1}{d+\alpha_0\sum_{j=1}^{n_0}s_j+1}$
$Bin(N_i,\theta)$	$Beta(a,b)$	$1-\frac{N}{a+b+\alpha_0\sum_{j=1}^{n_0}N_j+N}$
$NBin(r,\theta)$	$Beta(a,b)$	$1-\frac{r}{b+\alpha_0\sum_{j=1}^{n_0}y_j-1+r}$
$G(\alpha,\theta)$	$G(c,d)$	$1-\frac{\alpha}{c+\alpha\cdot(\alpha_0 n_0+1)-1}$
$W(\theta,\kappa)$	$IG(\alpha,\beta)$	$1-\frac{\Gamma^2\left(1+\frac{1}{\kappa}\right)\left[\Gamma\left(\alpha+\alpha_0 n_0\right)\Gamma\left(\alpha+\alpha_0 n_0-\frac{2}{\kappa}\right)+\Gamma^2\left(1-\frac{1}{\kappa}\right)\right]}{\Gamma\left(\alpha+\alpha_0 n_0\right)\Gamma\left(\alpha+\alpha_0 n_0-\frac{2}{\kappa}\right)\Gamma\left(1+\frac{2}{\kappa}\right)+\Gamma^2\left(1+\frac{1}{\kappa}\right)\Gamma^2\left(1-\frac{1}{\kappa}\right)}$
$IG(\alpha,\theta)$	$G(c,d)$	$1-\frac{\alpha-2}{c+\alpha\cdot\alpha_0 n_0-1+\alpha}$
$Pa(m,\theta)$	$G(c,d)$	$1-\frac{Var_\theta\left(\frac{\theta}{\theta-1}\right)}{E_\theta\left(\frac{\theta}{(\theta-1)^2(\theta-2)}\right)+Var_\theta\left(\frac{\theta}{\theta-1}\right)}$

Table 3.2.1: The expected ratio of the variance of the likelihood $f(X|\boldsymbol{\theta})$ over the variance of the marginal $f(X|\boldsymbol{Y},\alpha_0,\boldsymbol{\tau})$, defined in (3.2.3)

tive prior.

When our distributional setup does not conform with either scenario 3 or 4, we face the most challenging case. Since we do not have a reliable way to estimate h using imaginary data, we will make use of the predictive Bayes factor to form some evidence based limits for the charting statistics S_{n+1}. In (3.1.4) we expressed S_{n+1} as the zero truncated cumulative logarithmic Bayes Factor, which measures the evidence of the alternative model $\mathcal{M}_1$ (OOC) against the null $\mathcal{M}_0$ (IC). In addition, under the assumption that the IC and OOC models are equally probable a-priori, i.e. $P(\mathcal{M}_0) = P(\mathcal{M}_1)$, then B_{10}^{n+1} is simply the posterior odds of the two models. Kass and Raftery (1995), following Jeffreys (1961), provided an analytical interpre-

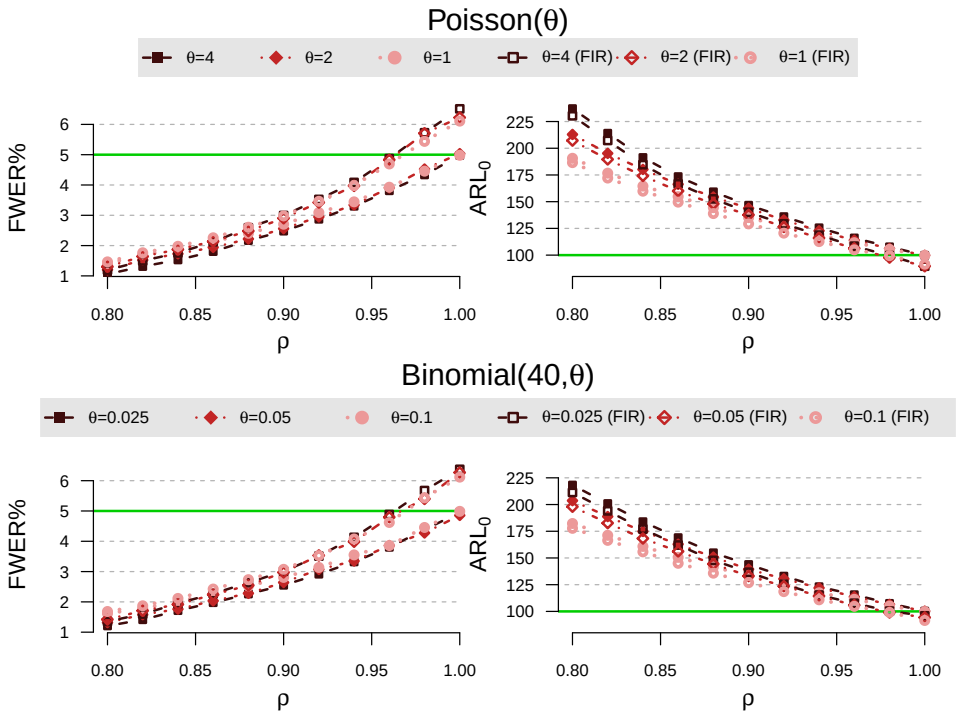


Figure 3.2.1: The h_m based achieved $FWER$ and ARL_0 metrics for different parameter values as function of ρ in a Poisson or a Binomial PRC process. The horizontal lines indicate the target value of $FWER = 5\%$ (with $N = 50$) and $ARL_0 = 100$, while for the FIR adjustment it was used $(f, d) = (1/2, 3/4)$.

tation of B_{10}^{n+1} and offered threshold values for decision making. Based on these guidelines, when $B_{10}^{n+1} > 100$, then the evidence against the null model is referred as "decisive", since the posterior probability of the alternative model will be at least 100 times greater than the corresponding of the null. Thus, we recommend to use $h_{BF} = log(100) \approx 4.605$ as an evidence based limit for S_{n+1}. In other words, if $S_{n+1} > h_{BF}$ (or $S_{n+1} < -h_{BF}$ for downward shifts), then we have a decisive cumulative evidence in favor of the OOC state.

The evidence based limits can be used for a few initial steps to monitor the process. At each step, as long as the posterior odds reveal that we are in the IC state, we can use the obtained data to update the prior setting (since the posterior at each time point acts as prior for the next observable) and examine whether it becomes informative (based on ρ) or not. Once we have an informative prior we move to

scenario (2), generating imaginary data from the marginal and deriving h_m, initiating a new PRC.

In Figure 3.2.2 we summarize all the proposed options for deriving a PRC's decision threshold. The threshold h, will depend on the likelihood of the data, the prior settings and the intervened vector $\boldsymbol{\tau}'_{\boldsymbol{n}}$, with the latter reflecting the discrepancy between the current (IC) and the intervened (OOC) distribution. In general, assuming that the deviation between IC and OOC state is considerably large, then if a change of smaller size occurs, PRC might absorb it. On the other hand, if the real shift is greater than the one we have set, then PRC probably will have a slightly delayed alarm, but is expected to react. Therefore, the choice of the OOC state must take into account the absorption risk, avoiding setting PRC for very large shifts (a similar discussion regarding SSC can be found in Zantek, 2006). This is an issue, closely related with West's (1986) CBF methodology, where the alternatives are set to be diffused, allowing potentially large shifts, a strategy that has a high risk in absorbing small shifts and not reacting on them (for more information refer to Subsection 3.3.1).

Since for scenarios 2 and 3 in determining the decision threshold, we will have h_m to be a more conservative estimate, resulting lower false alarms (from what we design), the use of FIR-PRC is motivated. Additionally, in some cases of handling big values of ARL_0, the FIR adjustment might be implemented for longer periods of data and not just for the very few first observations (Figure 3.2.1 visualizes the benefit of FIR-PRC).

3.2.2 PRC based inference

The control chart associated with PRC has the familiar form of a CUSUM, where the monitoring statistic S_{n+1} (from either (3.3.1) or (3.1.4)) is plotted versus time with a horizontal decision limit h, derived in Section 3.2, acting as an upper/lower control

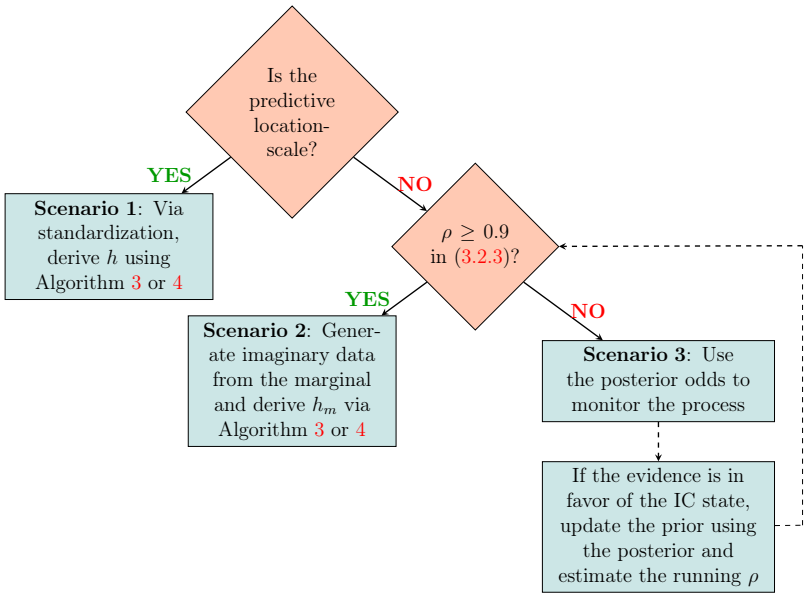


Figure 3.2.2: Determining the decision threshold h for a PRC scheme. A decision is represented by a rhombus and a rectangle corresponds to an operation after a decision making.

limit in detecting upward/downward shifts (graphical illustrations are available in Section 3.4). The area between the horizontal axis and h is considered as the IC region, so when S_{n+1} plots beyond the control limit h, then we raise an alarm and our suggestion is to stop the process and examine for an assignable cause, triggering a potential corrective action. From a root cause analysis point of view, a CUSUM alarm will indicate not only that the IC state has been rejected, but it will also offer an estimate of the time where the OOC state was initiated, which is simply the latest time for which we had $S_{n+1} = 0$. Once we correct the problem, then PRC is suggested to be reinitiated, using all past IC recordings as historical data in the power prior.

If we will not react to an alarm, then due to the dynamic update of PRC, OOC data will be involved in the learning process, affecting what is considered as IC state. As a result, the monitoring statistic will start moving back to the IC region. This is a well known issue for the self-starting methods, reported in the literature as "window of opportunity" for a control chart to alarm, before the running statistic stops to

alarm (in contrast to the fixed parameter CUSUM, where there is no updating and so an alarm will tend to persist). Thus, it is strongly recommended to act upon a PRC alarm.

As we mentioned earlier, PRC's monitoring can be considered as a sequential hypothesis testing regarding the unknown parameter. Furthermore, within the Bayesian decision theory framework, one can derive the point/interval estimate of the unknown parameter. Precisely, when the process is under the IC state, the posterior distribution of the unknown parameter(s) can be used to derive a Bayes point estimate (like the posterior mean under squared error loss) or the Highest Posterior Density (HPD) credible set. Such inference is also available via the predictive distribution when forecasting might be of interest.

Quite often in practice, we might need to employ more than a single PRC, like when we monitor the mean of a Normal distribution for either an upward or a downward shift. In such cases, we need to account for the multiple testing and so if we use the $FWER$ metric we simply need to adjust its value, using for example the Bonferroni's correction (Dunn, 1961). For the ARL metric, one can refer to Hawkins and Olwell (1998) among others, on how to combine the individual CUSUM $ARLs$, in getting a designed overall ARL.

Summarizing, all the possible options of PRC are provided next, both in pseudo-code in Algorithm 5 and in a flowchart in Figure 3.2.3.

3.3 Comparative study and sensitivity analysis

3.3.1 Competing methods

The first competing method of the study is SSC, which was demonstrated in chapter 7 in Hawknis and Olwell (1998) and we will present it for Normal, Poisson and Binomial data. The SSC charts use the Q statistics, which were presented in Subsection 2.3.2.

Algorithm 5 PRC algorithm

Determine the PRC model from the Table 1 and define the size of the shift k { *Model* }
Determine either the length of data N with $FWER$ or the ARL_0 { *FA tolerance* }
Is prior information available? { *initial prior* $\pi_0(\cdot)$ }
YES
  Determine the hyperparameters of the initial prior $\boldsymbol{\tau}$
NO
  Set the initial reference prior
Are prior data available? { *power prior* }
YES
  Provide the historical data $\boldsymbol{Y}$ and determine α_0
NO
  Set $\alpha_0 = 0$
Choose the appropriate threshold h based on Section 3.2 { *Decision Threshold* }
Is FIR-PRC of interest? { *FIR* }
YES
  Determine the parameters (f, d) in (13)
NO
  Set $f = 0$ in (13)
Once the data point x_n $(n \geq 1^{\bigstar})$ arrives, derive the predictive distribution of next observable $X_{n+1}|\,(\boldsymbol{X}_n, \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau})$
Obtain x_{n+1} and calculate $FIR_{adj} \cdot log(L_{n+1})$ in (7) and S_{n+1} in (8) { S_{n+1} }
if $S_{n+1} \leq h$ (or $S_{n+1} \geq h$ for downward shifts) **then** { *test* }
  $n \leftarrow n + 1$
  goto 19
else { *Stopping time alarm*}
  Raise an Alarm
  if you do not make a corrective action **then**
    goto 22
  else
    end PRC scheme
  endif
endif

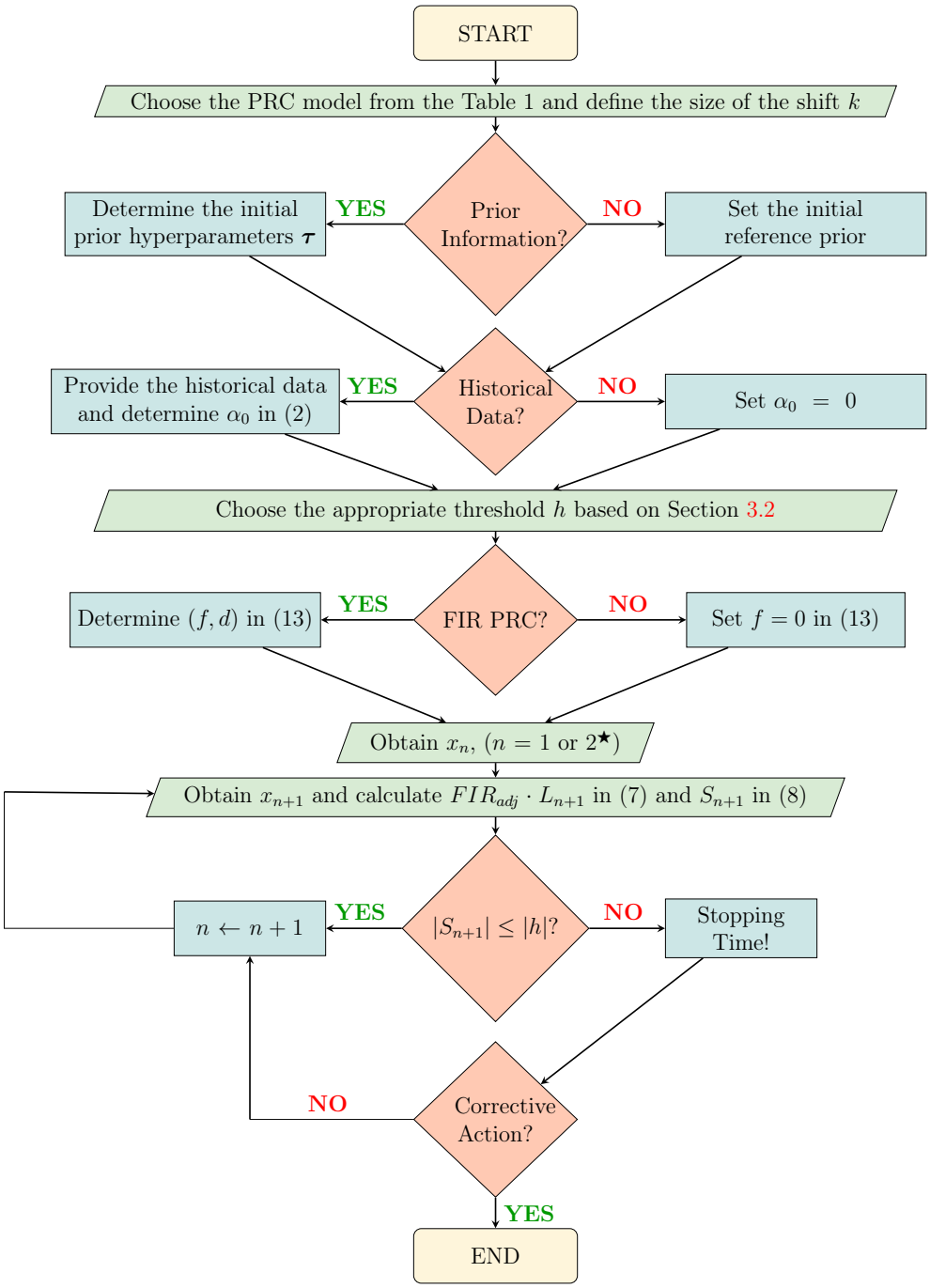


Figure 3.2.3: PRC flowchart. A parallelogram corresponds to an input/output information, a decision is represented by a rhombus and a rectangle denotes an operation after a decision making. In addition, the rounded rectangles indicate the beginning and end of the process.

*For the likelihoods with two unknown parameters and total prior ignorance (i.e. initial reference prior and $\alpha_0 = 0$ in the power prior) we need $n = 3$ to initiate PRC, while for all other cases, PRC starts right after x_1 becomes available.

After calculating the Q statistic at time $n+1$, then a CUSUM chart is constructed for it. Thus, the formula of one sided SSC is:

$$S_{n+1} = \max\{0, S_n + Q_{n+1} - k\} \quad \text{or} \quad S_{n+1} = \min\{0, S_n - Q_{n+1} + k\} \tag{3.3.1}$$

for upwards and downwards shifts respectively, where k is a reference value which determines the size of the shift for which the SSC is tuned. Hawknis and Olwell (1998) and Zantek (2006) discussed about the selection of k. For Normal data, if we are interested in detecting a shift for the variance, then we employ the square of the calculated statistic. In other words, a scale SSC is employed if in the formulas of S_{n+1} we replace Q_{n+1} by Q_{n+1}^2. We ring an alarm, if $|S_{n+1}| \geq |h_{SSC}|$, where h_{SSC} is a threshold appropriately chosen to respect to the false alarm criterion.

The second competing method is CBF by West (1986) and West and Harrison (1986). In the same philosophy with PRC, CBF uses the ratio of posterior predictive distributions. The main difference between these two procedure is that, while the predictive $f\left(X_{n+1}|\boldsymbol{X}_n\right)$ that represents the IC state is identical, the OOC predictives differ. The CBF's alternative predictive $f_A\left(X_{n+1}|\boldsymbol{X}_n\right)$ will be a diffused version of the IC predictive, with the same mean but a greater variance. Specifically, the variance of the neutral alternative will be:

$$Var\left(X_{n+1}|\boldsymbol{X}_n\right) = Var_A\left(X_{n+1}|\boldsymbol{X}_n\right)/r \tag{3.3.2}$$

where $0 < r < 1$ is the discount factor. The Bayes' factor at time $n+1$ will be:

$$H_{n+1} = f\left(X_{n+1}|\boldsymbol{X}_n\right)/f_A\left(X_{n+1}|\boldsymbol{X}_n\right) \tag{3.3.3}$$

Small values H_{n+1} indicate poor predictive performance of IC state. Then, the cumulative Bayes' factor of the most recent k observations is defined as:

$$W_{n+1}(k) = H_{n+1} \cdot H_n \cdot \ldots \cdot H_{n-k+2} = H_{n+1} \cdot W_n(k) \tag{3.3.4}$$

for $n+1 \geq 2$. As West and Harrison (1986) denoted, a single small $H_{n+1} = W_{n+1}(1)$ provides a warning of a potential outlier at time $n+1$ or the or the onset of change, while a small $W_{n+1}(k)$ for $k > 1$ suggests a possible change in the past. Finally, the cumulative statistic is given the formula:

$$V_{n+1}(k) = \min_{1 \leq t \leq n+1} W_{n+1}(k) = H_{n+1} \min\{1, V_n\}. \tag{3.3.5}$$

We ring an alarm, if $V_{n+1}(k)$ drops below a predetermined threshold h_{CBF} a threshold appropriately chosen to respect to the false alarm criterion.

3.3.2 Simulation study

In this subsection, we will evaluate the performance of PRC and compare it against SSC and CBF. The comparison will involve data from Normal, Poisson or Binomial, i.e. the most studied distributions in SPC/M. The goal will be to detect as soon as possible, step changes for the mean or inflation for the standard deviation in Normal data (when both parameters are unknown), rate increases in Poisson and increases in the odds of the success probability in Binomial data (all cases refer to typical process deterioration in SPC/M).

All competing methods, are aligned to have identical false alarm rate, while they are designed appropriately to detect the OOC scenario under study. Specifically, we tune the parameter k in PRC, the reference value of SSC, and the discount factor of CBF, to reflect on the size of the shift that we aim to detect. For the SSC with discrete distributions (i.e. Poisson and Binomial) we follow the suggestion (in chapter 7) of Hawkins and Olwell (1998), where the normal scores obtained based on the proposal of Quessenberry (1995) are winsorized by replacing, whenever necessary, the undefined $\Phi^{-1}(1)$ by $\Phi^{-1}(0.995)$.

To derive the decision limit of each method, we simulate 100,000 IC sequences of

size $N = 50$ observations from $N(\theta_1 = 0,\ \theta_2^2 = 1)$ (that will be used for both the mean and the variance charts), $P(\theta_3 = 1)$ and $Bin(40,\ \theta_4 = 0.025)$. In SPC/M we typically use Poisson or Binomial to model count or proportion of defects respectively and so small parameter values are more realistic. Furthermore, the Bayesian PRC and CBF methods require to define a prior distribution and so within this simulation we will take the opportunity to perform a sensitivity analysis, examining the effect of the presence/absence of prior information (reflecting the subjective/non-informative point of view). Therefore, for each scenario, we will compare the SSC against two versions for each of PRC and CBF (with/without prior knowledge). The initial priors $\pi_0(\cdot|\boldsymbol{\tau})$, considered are:

- Normal: reference (non-informative) prior $\pi_0\left(\theta_1, \theta_2^2\right) \propto 1/\theta_2^2 \equiv NIG(0, 0, -1/2, 0)$ or the moderately informative $NIG(0, 4, 2, 1.5)$.
- Poisson: reference (non-informative) prior $\pi_0(\theta_3) \propto 1/\sqrt{\theta_3} \equiv G(1/2, 0)$ or the moderately informative $G(4, 4)$.
- Binomial: reference (non-informative) prior $\pi_0(\theta_4) \propto 1/\sqrt{\theta_4(1-\theta_4)} \equiv Beta(1/2, 1/2)$ or the moderately informative $Beta(4, 156)$.

The OOC scenarios that will evaluate the detection power of the competing methods, come from the 100,000 IC sequences of length $N = 50$, where small or medium permanent parameter shifts (i.e. step changes) are introduced at one of the locations $\omega = \{11, 26 \text{ or } 41\}$. In other words, we have three scenarios for the unique change point location ω: either at the start, or in the middle, or near the end of the sample. For each location we will consider two shift sizes, which will be:

- Normal (mean): mean step change of size $\{1\theta_2 \text{ or } 1.5\theta_2\} = \{1 \text{ or } 1.5\}$, i.e. after the change point ω, the OOC data come from $N(1, 1)$ or $N(1.5, 1)$.
- Normal (standard deviation): sd inflation of size $\{50\% \text{ or } 100\%\}$, i.e. after the change point ω, the OOC data come from $N(0, 1.5^2)$ or $N(0, 2^2)$.
- Poisson (rate): parameter increase of size $\{50\% \text{ or } 100\%\}$, i.e. after the change

point ω, the OOC data come from $P(1.5)$ or $P(2)$.

- Binomial (probability of success): an increase of size {50% or 100%} for the odds of success, i.e. after the change point ω, OOC data come from $Bin(40, 0.037)$ or $Bin(40, 0.049)$.

Next, we provide the performance metrics used to evaluate the competing charts. First, we align all methods to have 5% Family Wise Error Rate (FWER) when we have IC data of length 50, i.e. $FWER(N) = P(T \leq N|\omega > N) = 0.05$, where T denotes the stopping time, ω is the time of the step change and $N = 50$ (length of the data in this study). Regarding OOC detection, the main goal of self-starting methods (especially in short runs), is to be able to ring an alarm before they absorb a change and also minimize the delay in ringing the alarm. The former, will be assessed, in the same philosophy as Frisen (1992), using the Probability of Successful Detection (PSD), where $PSD(\omega) = P(\omega \leq T \leq N)$ and the bigger $PSD(\omega)$, the better. For the latter, we estimate the delay of an alarm similar to Kenett and Pollak (2012), using the truncated Conditional Expected Delay, which is:

$$tCED(\omega) = E_\omega(T - \omega + 1|\omega \leq T \leq n) = \frac{E_\omega\left((T - \omega + 1) \cdot \mathbb{1}_{\{\omega \leq T \leq n\}}\right)}{P(\omega \leq T \leq n)} \qquad (3.3.6)$$

and it is the average delay of the stopping time T, given that this stopping time was after the change point occurrence and before the end of the sample (i.e. point of truncation) and the smaller the delay the better the performance. $tCED$ is in the same philosophy with Average Detection Delay $ADD(\tau) = E_\tau\left(T - \tau|T \geq \tau\right)$, which is minimized by the classical Shiryaev's process. In addition, other optimality properties have been investigated by Pollak and Tartakovsky (2009), however these problems are open when these distributions include unknown parameters. It is worth mentioning that the aforementioned metrics are more realistic compared with the rather restrictive Average Run Length (ARL, Lorden 1971), which in fact cannot be applied in self-starting procedures, especially when we have short runs.

The simulation results are summarized graphically in Figure 3.3.1 (and analytically in Table 3.3.1). Overall, the PRC outperforms both competing methods in all scenarios of jump sizes and change point locations, as it has steadily better performance in the detection ability and better or similar performance on the delay in signaling an alarm.

Initially, for the detection performance within each method, we observe (as it was expected) that the bigger the size of the shift, the higher the detection power. Regarding the effect of the location ω, we observe that in all cases the best performance appears when the change point is at the middle of the sequence ($\omega = 26$). The lower performance in the start ($\omega = 11$), is related to the fact that the learning process is not as mature as in the middle of the sequence. For the change near the end ($\omega = 41$) despite the fact that the learning has been significantly improved the performance decreases as there is not sufficiently long time to build up the evidence and ring an alarm (there exist only $50 - \omega + 1 = 10$ observations until we reach the end of the data sequence).

Comparing across methods via $PSD(\omega)$, we observe that the PRC achieves higher detection percentage than SSC for all distributions, shifts and locations. The PRC's outperformance against SSC is valid irrespectively of whether we have an informative or not prior distribution and their difference is greater at $\omega = 11$ (the earlier the shift the bigger the difference). The SSC's significantly lower performance versus PRC (even when a reference prior is in use) in the discrete distributions can be attributed to the fact that SSC is using an approximation to normality algorithm that in discrete data can be poor. The CBF, with one exception, is having the lowest performance of all competing methods. This is the price that CBF pays for aiming to be general and not specifying a target OOC distribution (it simply diffuses the predictive distribution keeping the same location). The exception is when we study shifts in the standard deviation of the normal data, where the CBF becomes informative, since the alternative (OOC) scenario involves the same location and

inflated variance. Thus, for this specific scenario, CBF coincides with PRC providing identical performance. Regarding $tCED(\omega)$, we observe that PRC is indifferent from SSC in Normal data (and better from CBF in normal mean PRC), while for the discrete distributions we have PRC to have comparable performance with CBF and a lot better (i.e. smaller delay) when compared to the SSC. Finally, the prior sensitivity indicates that even moderately informative prior information enhances the performance of PRC (and CBF). This is more intense at the early stages of the process ($\omega = 11$), when the volume of the data is very low.

3.3.3 PRC Robustness and FIR implementation

Like in Subsection 2.4, we will examine how robust is the PRC in misspecification of the setup and we will also perform a short simulation study for the evaluation of the performance of the FIR-PRC scheme. Namely, we will compare PRC, SSC and CBF, when we have:

(a) distributional violation (i.e. data are generated from a different distribution than the one assumed),

(b) jump misspecification (i.e. the real change in the process is different from the one that is tuned),

(c) misplaced prior distribution (i.e. the prior mean is significantly different than the real mean of the process),

(d) early shifts in a FIR-PRC scheme (i.e. shifts at the early start of the process).

For (a), we will examine the following scenarios:

- run the PRC process for the mean of Normal data, while the real data are generated from a Student t_5 distribution,

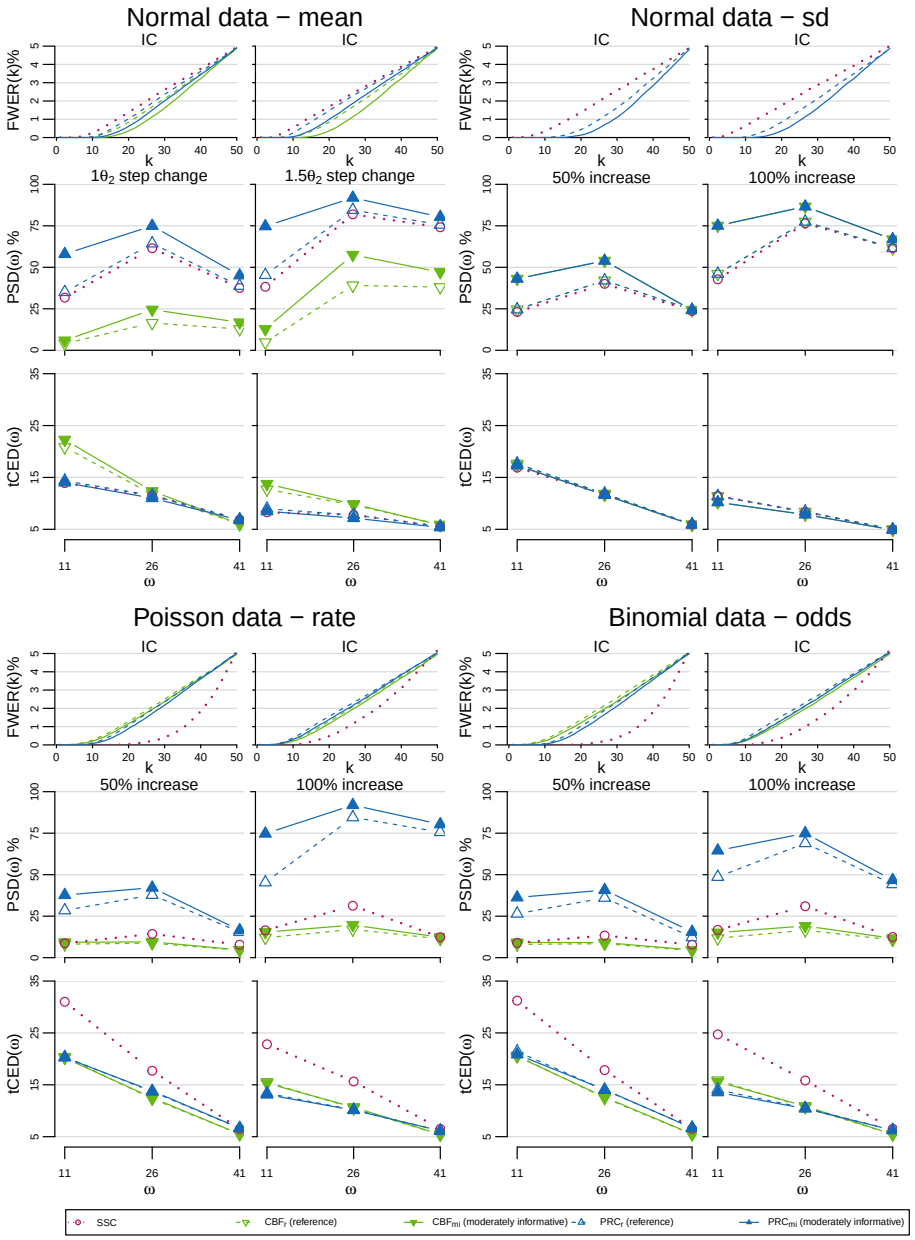


Figure 3.3.1: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 50$, the probability of successful detection, $PSD(\omega)$, (%) and the truncated conditional expected delay, $tCED(\omega)$ for shifts at locations $\omega = \{11, 26, 41\}$, of SSC, CBF and PRC, under a reference (CBF_r, PRC_r) or a moderately informative (CBF_{mi}, PRC_{mi}) prior. The results refer to Normal data with step changes for the mean of size $\{1\theta_2, 1.5\theta_2\}$, Normal data with inflated standard deviation of size $\{50\%, 100\%\}$, Poisson data with rate increase of size $\{50\%, 100\%\}$ and Binomial data with increases for the odds of size $\{50\%, 100\%\}$.

	Jump	ω	SSC $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$
Normal (mean)	$+1\sigma$	IC	4.973%	4.824%	4.871%	4.879%	4.917%
		11	31.724%	4.080%	5.913%	35.053%	58.009%
			13.908	20.838	22.275	14.355	14.013
			(8.738)	(10.522)	(9.441)	(8.562)	(8.234)
		26	61.487%	16.479%	24.513%	64.343%	74.954%
			11.509	11.722	12.357	11.534	11.004
			(5.720)	(6.594)	(6.436)	(5.657)	(5.552)
		41	37.571%	12.805%	16.832%	38.804%	45.088%
			6.885	5.985	6.034	6.926	6.778
			(2.198)	(2.701)	(2.680)	(2.171)	(2.200)
	$+1.5\sigma$	IC	4.968%	4.824%	4.871%	4.889%	4.926%
		11	38.256%	4.820%	12.768%	45.335%	74.709%
			8.278	12.691	13.750	8.958	8.476
			(6.641)	(8.906)	(8.034)	(6.411)	(6.004)
		26	81.899%	39.069%	57.505%	84.536%	91.916%
			7.798	9.684	9.867	7.765	7.158
			(4.818)	(5.890)	(5.825)	(4.690)	(4.399)
		41	74.124%	38.075%	47.120%	75.624%	80.342%
			5.525	5.897	5.810	5.557	5.372
			(2.250)	(2.584)	(2.581)	(2.210)	(2.203)
Normal (rate)	+50%	IC	5.175%	49.67%	4.951%	4.975%	5.021%
		11	8.681%	8.126%	9.271%	28.482%	37.723%
			31.006	20.369	20.178	20.335	20.262
			(7.190)	(11.433)	(11.260)	(9.778)	(9.714)
		26	14.299%	8.762%	9.531%	28.482%	37.782%
			17.710	12.318	12.493	13.676	13.819
			(5.788)	(7.130)	(7.101)	(6.272)	(6.215)
		41	7.803%	4.441%	4.677%	15.437%	16.528%
			6.491	5.495	5.508	6.636	6.653
			(2.698)	(2.841)	(2.839)	(2.538)	(2.528)
Poisson	+100%	IC	5.175%	4.967%	4.951%	5.004%	5.036%
		11	16.492%	12.015%	15.377%	49.866%	65.412%
			23.813	15.527	15.301	13.378	13.107
			(9.948)	(11.393)	(11.068)	(9.326)	(8.979)
		26	31.298%	17.045%	19.643%	70.240%	76.212%
			15.645	10.489	10.690	10.174	10.114
			(6.379)	(6.839)	(6.856)	(5.896)	(5.831)
		41	12.363%	11.307%	12.345%	46.594%	49.006%
			6.502	5.382	5.418	6.134	6.133
			(2.666)	(2.818)	(2.819)	(2.444)	(2.442)
Normal (sd)	+50%	IC	4.885%	4.894%	4.804%	4.894%	4.804%
		11	23.093%	24.591%	43.089%	24.591%	43.089%
			16.891	17.647	17.304	17.647	17.304
			(10.708)	(10.363)	(9.981)	(10.363)	(9.981)
		26	40.032%	42.066%	53.937%	42.066%	53.937%
			11.756	11.859	11.622	11.859	11.622
			(6.655)	(6.607)	(6.551)	(6.607)	(6.551)
		41	23.678%	24.372%	29.146%	24.372%	29.146%
			5.912	5.946	5.873	5.946	5.873
			(2.692)	(2.680)	(2.691)	(2.680)	(2.691)
	+100%	IC	5.001%	4.824%	4.871%	4.824%	4.871%
		11	42.764%	46.051%	75.060%	46.051%	75.060%
			11.443	11.381	10.204	11.381	10.204
			(9.579)	(9.229)	(8.411)	(9.229)	(8.411)
		26	76.465%	77.453%	86.604%	77.453%	86.604%
			8.355	8.392	7.840	8.392	7.840
			(6.015)	(5.971)	(5.738)	(5.971)	(5.738)
		41	61.593%	61.548%	66.863%	61.548%	66.863%
			4.957	5.004	4.897	5.004	4.897
			(2.678)	(2.671)	(2.667)	(2.671)	(2.667)
Binomial (odds)	+50%	IC	5.191%	5.069%	49.76%	5.030%	5.008%
		11	8.865%	8.025%	9.180%	26.472%	36.262%
			31.234	20.606	20.467	21.411	20.856
			(7.046)	(11.405)	(11.283)	(9.425)	(9.566)
		26	14.343%	8.399%	9.084%	36.033%	40.734%
			17.836	12.468	12.629	14.008	14.056
			(5.725)	(7.140)	(7.089)	(6.201)	(6.174)
		41	7.817%	4.311%	4.562%	14.279%	15.535%
			6.509	5.470	5.510	6.704	6.704
			(2.682)	(2.831)	(2.825)	(2.539)	(2.534)
	+100%	IC	5.200%	5.069%	4.976%	5.069%	5.046%
		11	16.731%	11.703%	15.107%	48.631%	64.529%
			24.719	15.867	15.852	14.042	13.552
			(9.692)	(11.385)	(11.172)	(9.247)	(8.969)
		26	30.966%	16.556%	19.111%	68.966%	74.943%
			15.834	10.691	10.885	10.502	10.412
			(6.326)	(6.880)	(6.859)	(5.890)	(5.836)
		41	12.431%	10.729%	11.658%	44.227%	46.715%
			6.511	5.429	5.466	6.279	6.254
			(2.671)	(2.811)	(2.803)	(2.414)	(2.416)

Table 3.3.1: The probability of successful detection, PSD, (%), the truncated conditional expected delay, $tCED$ and its corresponding standard deviation (in parenthesis) for shifts at locations $\omega = \{11, 26, 41\}$, of SSC, CBF and PRC, under a reference (CBF_r, PRC_r) or a moderately informative (CBF_{mi}, PRC_{mi}) prior. The results refer to Normal data with step changes for the mean of size $\{1\theta_2, 1.5\theta_2\}$, Normal data with inflated standard deviation of size $\{50\%, 100\%\}$, Poisson data with rate increase of size $\{50\%, 100\%\}$ and Binomial data with increases for the odds of size $\{50\%, 100\%\}$.

- run the PRC process for the mean of Normal data, while the real data are generated from a $Gumbel\,(\mu = -0.5, \beta = 0.8)$ distribution,
- run the PRC process for Poisson data, while the real data are generated from a $NBin\,(r = 4, p = 1/5)$ distribution.

In the distributional violation scenarios, the shifts will be of size $1sd$ of the actual distribution. Regarding (b), we will set all the methods for step changes in the mean of Normal data size of 1σ, while the real jump will be:

- a mean step change size of 0.5σ,
- a mean step change size of $1.5sd$,
- a sd inflation size of 100%.

The prior distributions (reference prior and moderately informative) for (a) and (b) will be identical to the ones used in Subsection 3.3.1. For the prior misspecification scenario (c), we will examine the performance of both the Bayesian methods, PRC and CBF, misplacing moderately informative priors. Precisely, we will set the processes for mean step changes size of 1σ with IC sequences from a standard Normal distribution. But, instead from the moderately informative $NIG(0, 4, 2, 1.5)$ with prior mean $\mu_0 = 0$, we will use the misplaced priors:

- $NIG(0.5, 4, 2, 1.5)$ with prior mean $\mu_0 = 0.5$,
- $NIG(-0.5, 4, 2, 1.5)$ with prior mean $\mu_0 = -0.5$.

The locations of the change points for all the above misspecification scenarios will be the same with those in Subsection 3.3.2, i.e. we will introduce contaminated data starting from $\omega = \{11, 26 \text{ or } 41\}$ until the end of the sample ($N = 50$). As regards the FIR-PRC implementation in (d), we will examine its performance using the somewhat conservative choice of $(f, d) = (1/2, 3/4)$. The setup will be the same as that of Subsection 3.3.1 with mean step changes of 1σ in Normal data, introducing at early locations $\omega = \{6 \text{ or } 11\}$.

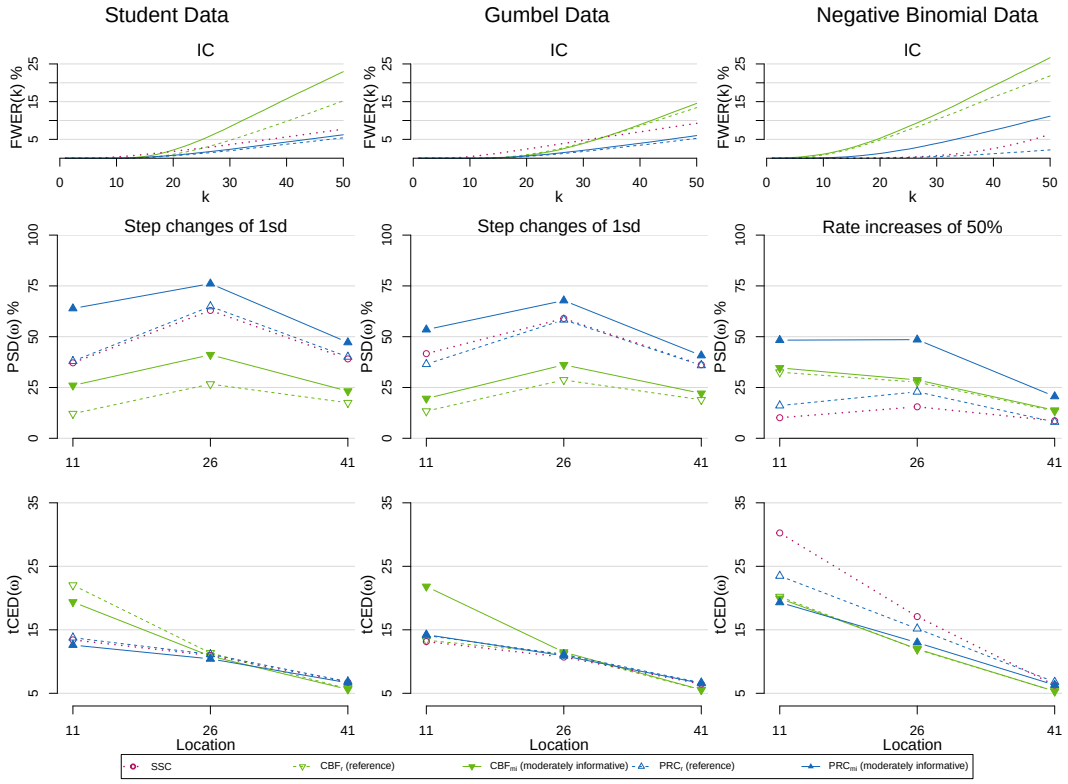


Figure 3.3.2: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 50$, the probability of successful detection, $PSD(\omega)$, (%) and the truncated conditional expected delay, $tCED(\omega)$ for shifts at locations $\omega = \{11, 26, 41\}$, of SSC, CBF and PRC, under a reference (CBF_r, PRC_r) or a moderately informative (CBF_{mi}, PRC_{mi}) prior for OOC scenarios with misspecified distributions. All the procedures are set for a mean step change size of 1σ in data from a standard Normal or an rate increase of 50% in Po(1) data.

Figures 3.3.2, 3.3.3, 3.3.4 and 3.3.5 provide a graphical representation of the results of Tables 3.3.2, 3.3.3 3.3.4 and 3.3.5, regarding the performance for misspecification in the distribution, the kind of the shift and the prior, along with the FIR-PRC respectively. As we see, PRC is less affected by the distributional violation in either the false alarms in a IC sequence or the detection power in the OOC scenarios. Specifically, the false alarms of SSC and CBF are unacceptable high in most cases, while those of PRC are close to the predetermined, especially with the reference prior, where PRC is almost unaffected. In addition, PRC has greater detection percentages, especially using the moderately informative prior.

	Actual Distribution	k'	SSC $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$
Misspecified distributions	IC		7.685%	6.217%	5.410%	22.942%	15.222%
	t_5	11	37.130%	63.885%	37.932%	26.015%	12.102%
			13.411	12.596	13.756	19.376	22.034
			(8.713)	(7.794)	(8.288)	(10.187)	(10.587)
		26	62.920%	76.107%	64.688%	41.083%	26.663%
			10.956	10.422	11.140	10.858	11.354
			(5.658)	(5.458)	(5.547)	(6.533)	(6.635)
		41	39.057%	47.175%	40.062%	23.203%	17.483%
			6.727	6.656	6.850	5.623	5.719
			(2.256)	(2.224)	(2.183)	(2.764)	(2.748)
	IC		9.247%	6.013%	5.272%	14.542%	13.421%
	$Gu(-0.5, 0.8)$	11	41.658%	53.488%	36.462%	19.560%	13.361%
			13.152	14.237	14.105	21.824	21.841
			(9.092)	(8.450)	(8.813)	(10.128)	(10.728)
		26	58.958%	67.790%	58.332%	36.149%	28.717%
			10.703	10.907	11.127	11.483	11.197
			(5.856)	(5.679)	(5.728)	(6.662)	(6.752)
		41	36.204%	40.736%	35.881%	22.105%	18.941%
			6.454	6.562	6.663	5.562	5.535
			(2.392)	(2.318)	(2.298)	(2.751)	(2.755)
	IC		6.379%	11.148%	2.204%	26.677%	21.836%
	$NB(4, 1/5)$	11	10.121%	48.294%	16.081%	34.639%	32.524%
			30.258	19.315	23.484	34.639	32.524
			(7.721)	(9.962)	(9.245)	(11.246)	(11.291)
		26	15.449%	48.547%	22.867%	28.720%	27.625%
			17.083	12.980	15.177	11.986	11.916
			(6.042)	(6.435)	(6.015)	(7.141)	(7.514)
		41	8.579%	20.601%	8.024%	13.801%	13.396%
			6.293	6.301	6.779	5.327	5.324
			(2.726)	(2.619)	(2.527)	(2.859)	(2.863)

Table 3.3.2: The percent probability of successful detection, the truncated conditional expected delay and the corresponding standard deviation (in parenthesis) for $\omega = \{11, 26, \text{ or } 41\}$, of SSC against PRC under a moderately informative prior (PRC_{mi}) or the reference prior (PRC_r) and CBF under a moderately informative prior (PRC_{mi}) or the reference prior (CBF_r) for OOC scenarios with misspecified distributions. All the procedures are set for a mean step change size of 1σ in data from a standard Normal or an rate increase of 50% in Po(1) data.

Regarding the misspecification in the shift, PRC is still robust, even when smaller changes from what we designed occur. Especially under the presence of IC prior information, PRC has significantly high detection percentages. Seemingly, the only exception is in the case of a change in variance. There, the CBF has a superior performance, but this is expected because, as we have mentioned, CBF is essentially a method that can successfully detect changes in variance. Moreover, it is identical to the PRC in this case, so the exact same performance would have achieved by a

PRC for the variance.

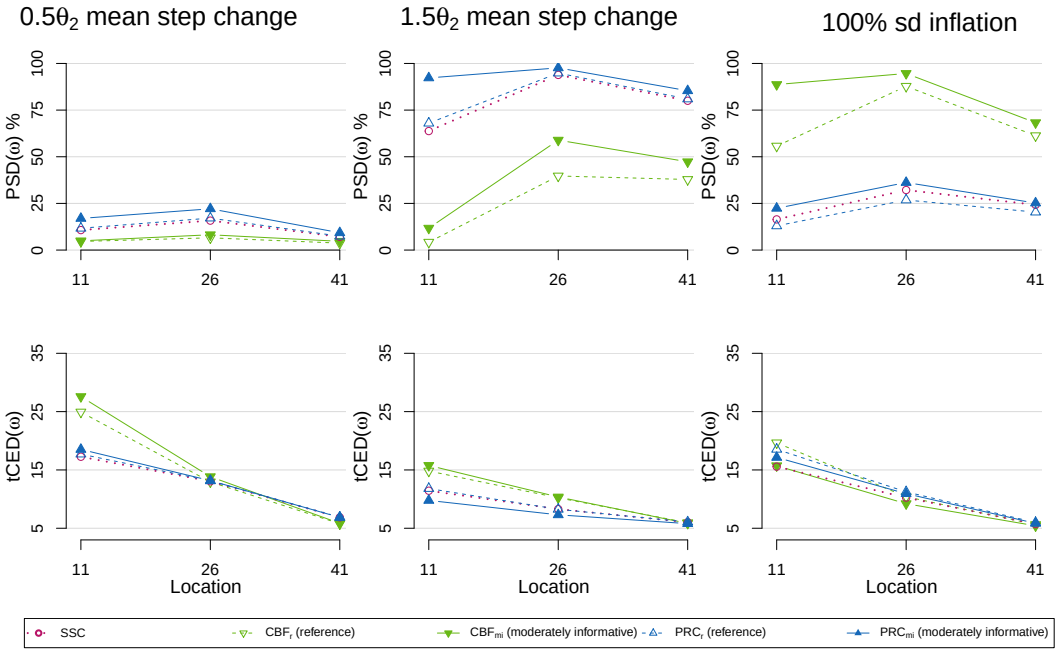


Figure 3.3.3: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 50$, the probability of successful detection, $PSD(\omega)$, (%) and the truncated conditional expected delay, $tCED(\omega)$ for shifts at locations $\omega = \{11, 26, 41\}$, of SSC, CBF and PRC, under a reference (CBF_r, PRC_r) or a moderately informative (CBF_{mi}, PRC_{mi}) prior for OOC scenarios with misspecified jumps. All the procedures are set for a mean step change size of 1σ in data from a standard Normal distribution.

	Jump	k'	SSC $PSD(\omega)$% $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_{mi} $PSD(\omega)$% $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_r $PSD(\omega)$% $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_{mi} $PSD(\omega)$% $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_r $PSD(\omega)$% $tCED(\omega)$ $(sd(tCED(\omega)))$
	IC		4.943%	4.917%	4.879%	4.871%	4.824%
Misspecified jumps	0.5σ	11	10.707%	16.954%	11.683%	4.943%	4.622%
			17.261	18.519	17.715	27.577	24.930
			(10.442)	(9.743)	(10.123)	(8.394)	(9.753)
		26	15.775%	22.115%	17.122%	8.199%	6.557%
			13.049	13.171	13.128	13.842	12.951
			(6.229)	(6.174)	(6.203)	(6.739)	(6.997)
		41	7.078%	9.279%	7.440%	4.693%	3.715%
			6.861	6.858	6.901	5.847	5.780
			(2.408)	(2.392)	(2.393)	(2.785)	(2.805)
	1.5σ	11	63.772%	92.249%	67.982%	11.736%	4.214%
			11.436	9.774	11.807	15.774	14.812
			(7.325)	(6.003)	(7.170)	(8.503)	(9.401)
		26	93.854%	97.632%	94.831%	58.903%	39.695%
			8.277	7.304	8.250	10.361	10.225
			(4.446)	(3.779)	(4.302)	(5.909)	(6.043)
		41	80.003%	85.372%	80.954%	47.309%	37.820%
			6.029	5.794	6.069	5.846	5.949
			(2.119)	(2.075)	(2.083)	(2.583)	(2.593)
	100%	11	16.435%	22.384%	19.958%	88.688%	55.722%
			15.566	17.148	18.534	15.721	19.648
			(11.139)	(10.535)	(11.027)	(8.585)	(9.531)
		26	32.170%	36.199%	26.820%	94.630%	87.762%
			10.253	10.932	11.242	9.209	10.409
			(6.446)	(6.433)	()6.487	(5.572)	(5.985)
		41	24.420%	25.208%	20.330%	68.253%	61.292%
			5.707	5.846	5.980	5.422	5.629
			(2.597)	(2.564)	(2.551)	(2.656)	(2.657)

Table 3.3.3: The percent probability of successful detection, the truncated conditional expected delay and the corresponding standard deviation (in parenthesis) for $\omega = \{11, 26, 41\}$, of SSC against PRC under a moderately informative prior (PRC_{mi}) or the reference prior (PRC_r) and CBF under a moderately informative prior (PRC_{mi}) or the reference prior (CBF_r) for OOC scenarios with misspecified jumps. All the procedures are set for a mean step change size of 1σ in data from a standard Normal distribution.

Continuing with the misplaced prior, PRC is very robust, even if the prior is on the side of the jump (PRC_+). Finally, the FIR-PRC with a conservative choice of (f, d) boosts the performance, especially at $\omega = 6$, while it has only a small effect on the false alarms. This is of great importance if we wish to detect changes at the beginning of a process.

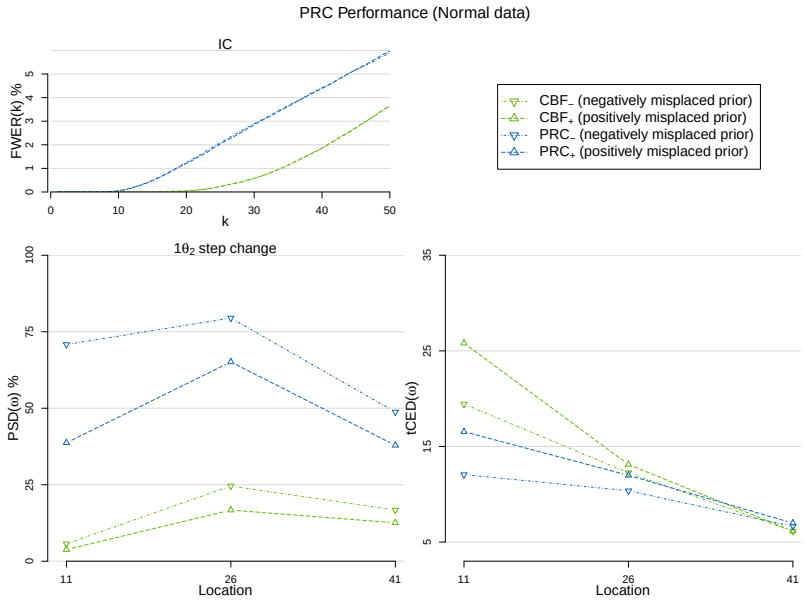


Figure 3.3.4: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 50$, the probability of successful detection, $PSD(\omega)$, (%) and the truncated conditional expected delay, $tCED(\omega)$ for shifts at locations $\omega = \{11, 26, 41\}$, of PRC under two misplaced moderately informative priors in the positive and the negative respectively (PRC_+ and PRC_+) along with the and CBF under the same priors (CBF_+ andCBF_-). All the procedures are set for a positive mean step change size of 1σ in Normal data.

	Jump	k'	PRC_+ $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_- $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_+ $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_- $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$
Misplaced priors	IC		5.962%	5.885%	3.648%	3.655%
	1σ	11	38.703%	70.832%	3.791%	5.700%
			16.549	12.059	25.806	19.439
			(8.709)	(7.738)	(8.897)	(9.502)
		26	65.132%	79.485%	16.704%	24.582%
			11.981	10.366	13.115	12.254
			(5.609)	(5.478)	(6.456)	(6.351)
		41	37.880%	48.775%	12.556%	16.708%
			6.982	6.643	6.149	6.135
			(2.128)	(2.243)	(2.685)	(2.650)

Table 3.3.4: The percent probability of successful detection, the truncated conditional expected delay and the corresponding standard deviation (in parenthesis) for $\omega = \{11, 26, 41\}$ of PRC under two misplaced moderately informative priors in the positive and the negative respectively (PRC_+ and PRC_+) along with the and CBF under the same priors (CBF_+ andCBF_-). All the procedures are set for a positive mean step change size of 1σ in Normal data.

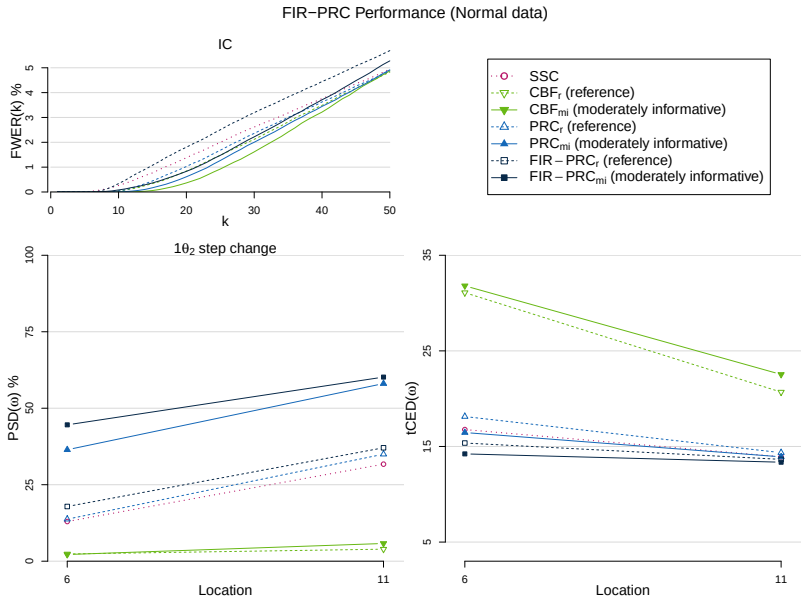


Figure 3.3.5: The $FWER(k)$ at each time point $k = 2, 3, \ldots, 50$, the probability of successful detection, $PSD(\omega)$, (%) and the truncated conditional expected delay, $tCED(\omega)$ for shifts at locations $\omega = \{11, 26, 41\}$, of SSC, CBF and PRC, under a reference (CBF_r, PRC_r) or a moderately informative (CBF_{mi}, PRC_{mi}) prior. Along with the standard version of PRC, the FIR-PRC ($FIR - PRC_{mi}$ and $FIR - PRC_r$) with $(f, d) = (1/2, 3/4)$ is employed. The results refer to Normal data with step change of 1σ for the mean.

	Jump	k'	SSC $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	PRC_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_{mi} $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	CBF_r $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	$FIR - PRC_{mi}$ $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$	$FIR - PRC_r$ $PSD(\omega)\%$ $tCED(\omega)$ $(sd(tCED(\omega)))$
F I R	IC		4.943%	4.917%	4.879%	4.871%	4.824%	5.278%	5.691%
	1σ		12.966%	36.430%	13.733%	2.180%	2.360%	44.568%	17.880%
		6	16.769	16.457	18.132	31.805	31.090	14.224	15.361
			(10.923)	(9.367)	(10.400)	(9.507)	(10.446)	(9.163)	(10.283)
			31.716%	58.059%	34.978%	5.790%	3.927%	60.181%	37.070%
		11	13.926	13.916	14.351	22.534	20.867	13.354	13.650
			(8.772)	(8.142)	(8.517)	(9.507)	(10.446)	(8.127)	(8.542)

Table 3.3.5: The percent probability of successful detection, the truncated conditional expected delay and the corresponding standard deviation (in parenthesis) for $\tau = \{6, 11\}$, of SSC against PRC under a moderately informative prior (PRC_{mi}) or the reference prior (PRC_r) and CBF under a moderately informative prior (PRC_{mi}) or the reference prior (CBF_r). Along with the standard version of PRC, the FIR-PRC ($FIR - PRC_{mi}$ and $FIR - PRC_r$) with $(f, d) = (1/2, 3/4)$ is employed. The results refer to Normal data with step change of 1σ for the mean.

3.4 PRC real data application

3.4.1 PRC application to Normal data

In this Section we will illustrate the use of PRC in real data, assuming normality. Regarding the continuous case, we will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory and specifically from the area of clinical hemostasis. We are interested in the variable "Factor V", measured in % regarding the international standards in Biology. Factor V is one of the serine protease enzymes of the procoagulant system, which interacts on a phospholipid surface to induce formation of stable clot of fibrin. Deficiencies of Factor V can induce bleeding disorders of varying severity. The normal range for factor V level is 61%−142% (for adults) and in this application we focus on pathological values (i.e. measurements below 60%, which Biologist's call abnormal values). In a medical lab, where control samples are used to monitor the quality of the process, it is known that a change of reagent batch might introduce a step change to the measurement of Factor V. This can occur at the early stages of the process and it is crucial to identify such a change point, when present, to avoid impacting clinically the patients care. We sequentially gathered 21 normally distributed IQC observations (X_i) from a medical lab (see Table 3.4.1), where $X_i|\left(\theta_1, {\theta_2}^2\right) \sim N\left(\theta_1, {\theta_2}^2\right)$.

$x_1 - x_{11}$	31.0	30.0	32.0	28.0	33.2	33.2	35.1	35.1	33.9	37.9	33.2
$x_{12} - x_{21}$	36.5	33.2	35.1	34.5	36.5	33.2	35.1	37.2	32.6	36.5	

Table 3.4.1: The Factor V (%) internal quality control observations of the current $\boldsymbol{X} = (x_1, x_2, \ldots, x_{21})$ data, reported during September 24, 2019 - October 8, 2019.

From the control sample manufacturer, we elicit the initial prior $\pi_0\left(\theta_1, {\theta_2}^2|\boldsymbol{\tau}\right) \sim NIG\left(31.8, 1/2, 2, 4.41\right)$. Furthermore, we have $n_0 = 37$ IC historical data (from a different reagent) available, with $\bar{\mathbf{y}} = 31.73$ and $var(\mathbf{y}) = 3.31$ and we set $\alpha_0 = 1/37$ in the power prior term, to convey the weight of a single data point to these. Combining the two sources of information within the power prior (2.1.3) we obtain:

$\pi\left(\theta_1, {\theta_2}^2 | \boldsymbol{Y}, \alpha_0, \boldsymbol{\tau}\right) \sim NIG\left(31.75, 3/2, 5/2, 6.02\right)$. The goal is to detect any small permanent positive or negative shift in the mean of the process, as this will have an impact on the reported patient results. In this setup, we choose the parameter $k = 1$, as at low levels of factor V the bleeding risk can hugely increase with small differences. Thus, we tune the PRC in detecting mean step changes, in either upward or downward direction, of one standard deviation size (i.e. $\pm\hat{\theta}_2$). The PRC control chart will plot two monitoring statistics: S^+_{n+1} (evolving in the nonnegative part) and S^-_{n+1} (evolving in the nonpositive numbers) that will test for upward and downward permanent mean shifts respectively. Furthermore, we will have two decision limits h^+ and h^-, which due to the normal distribution symmetry and the design of the same OOC step change shift ($\pm\hat{\theta}_2$) will be of the same magnitude (i.e. $|h^+| = |h^-|$). As the data are normally distributed, the standardized version of PRC is available and from scenario 1 of Section 3.2 we derive the decision limit $h^+ = 3.882$ ($h^- = -3.882$), to achieve $FWER = 5\%$ for 21 observations (since we run two tests we used Bonferroni's adjustment resulting $FWER = 2.5\%$ for each of the PRCs). As this study is offline, we will not interrupt the process after a PRC alarm (as we would have done when PRC runs online), but instead we will let it run until the end of the sample in order to perceive its behavior in the presence of contaminated data.

Figure 3.4.1, provides the two sided PRC chart along with the plot of the available data. The control chart rings an alarm at location eight indicating an upward mean shift, which seems to be initiated at location four (i.e. last time where $S^+_{n+1} = 0$ before the alarm), i.e. we have a delay of three observations in ringing the alarm. It worths noting also that the alarm persists till the end of the sample, indicating PRC's resistance in absorbing the change. We should also mention that due to the lack of knowledge of the actual parameter values one cannot provide a decision threshold to respect the required $FWER$ for either SSC or CBF, a huge obstacle for using them in everyday practice.

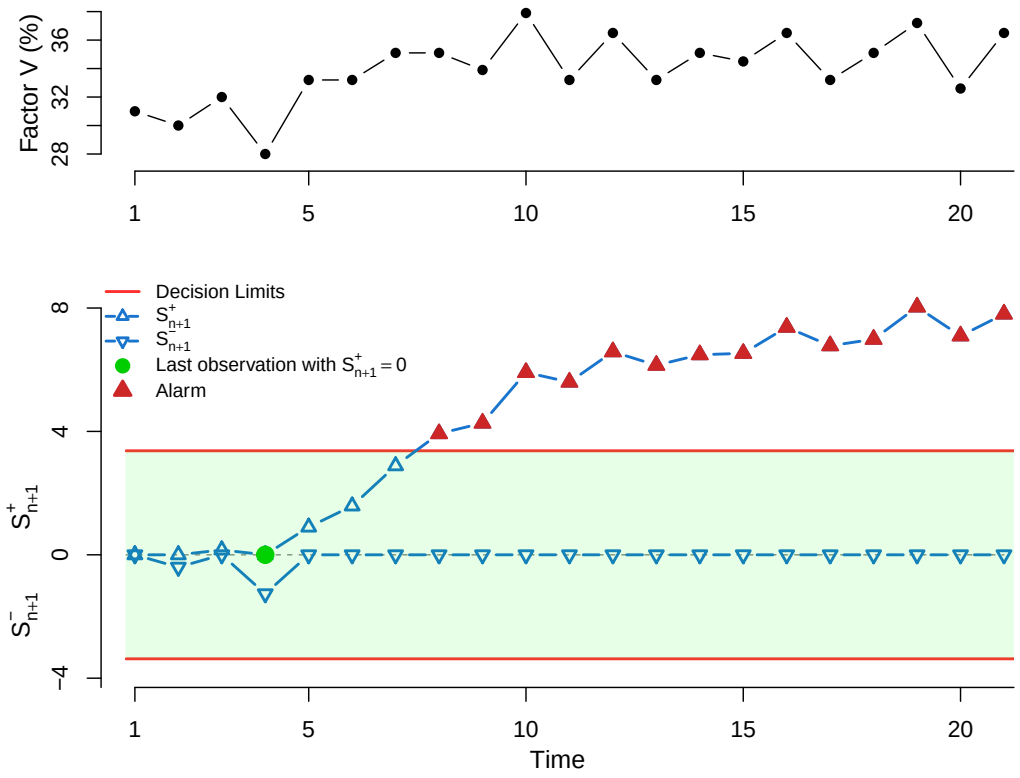


Figure 3.4.1: PRC for Normal data. At the top panel the data are plotted, while at the lower panel, we provide the PRC control chart, focused on detecting an upward or downward mean step change of one standard deviation size, when we aim a $FWER = 5\%$ for 21 observations.

3.4.2 PRC application to Poisson data

Now, we provide the PRC's illustration for discrete (Poisson) data, presented initially at Dong et al. (2008) and analyzed by Ryan and Woodall (2010) as well. They refer to counts of adverse events (x_i), per product exposure in millions (s_i), in a pharmaceutical company. We have 22 counts (see Table 3.4.2) arriving sequentially that we will model using the Poisson distribution with unknown rate parameter, i.e. $X_i|\theta \sim P(\theta \cdot s_i)$. In contrast to the previous application, neither prior information regarding the unknown parameter nor historical data exist. Therefore, we use the reference prior as initial prior for θ, i.e. $\pi_0(\theta|\boldsymbol{\tau}) \propto 1/\sqrt{\theta} \equiv G(1/2, 0)$ and we also set $\alpha_0 = 0$ for the power prior term.

Adverse events $(x_1 - x_{11})$	1	0	0	0	1	0	3	3	3	2	5
Product exposure $(s_1 - s_{11})$	0.206	0.313	0.368	0.678	0.974	0.927	0.814	0.696	0.659	0.775	0.731
Adverse events $(x_{12} - x_{22})$	5	2	4	4	3	4	3	8	3	2	2
Product exposure $(s_{12} - s_{22})$	0.710	0.705	0.754	0.682	0.686	0.763	0.833	0.738	0.741	0.843	0.792

Table 3.4.2: Counts of adverse events (x_i) and product exposure (s_i) per million ($i = 1, 2, \ldots, 22$), for each quarter reported during July 1, 1999 - December 31, 2004 (see Dong et al., 2008).

We tune PRC in detecting a 100% increase in the rate parameter and we also provide the FIR-PRC version, setting $(f, d) = (1/2, 3/4)$. As the predictive distribution is not a location/scale family and the prior is not informative, we fall under scenario 3 of Section 3.2 and so we will make use of the evidence based threshold $h_{BF} = log(100) \approx 4.605$. Just as we did in the previous application we will analyze all the data in an offline version and not interrupt the process after an alarm to record the alarm's persistence. In Figure 3.4.2 a plot of the data along with the two versions of PRC (with/without FIR) are provided.

The PRC provides the first alarm at observation 12, while the FIR-PRC gives an alarm at location 11, both indicating that we had a persistent rate increase, which appears to have started after location 6 (i.e. last time before the alarm, where the monitoring statistic was zero). Furthermore, the alarm persists until observation 21,

after which the monitoring statistics returns to the IC region. It is worth mentioning that, since we have a decisive evidence that the procedure is OOC, we maintained the evidence limit until the end of the sample, avoiding the option of elicitating h_m via the marginal distribution after the first few data, as described in scenario 3 of Section 3.2. We also note that both in Dong et al. (2008) and Ryan and Woodall (2010), where the aim was to have an IC Average Run Length $ARL_0 \approx 100$, their cumulative evidence monitoring approach, gave only a single alarm at location 19 (i.e. the alarm comes later compared to PRC and is absorbed instantly).

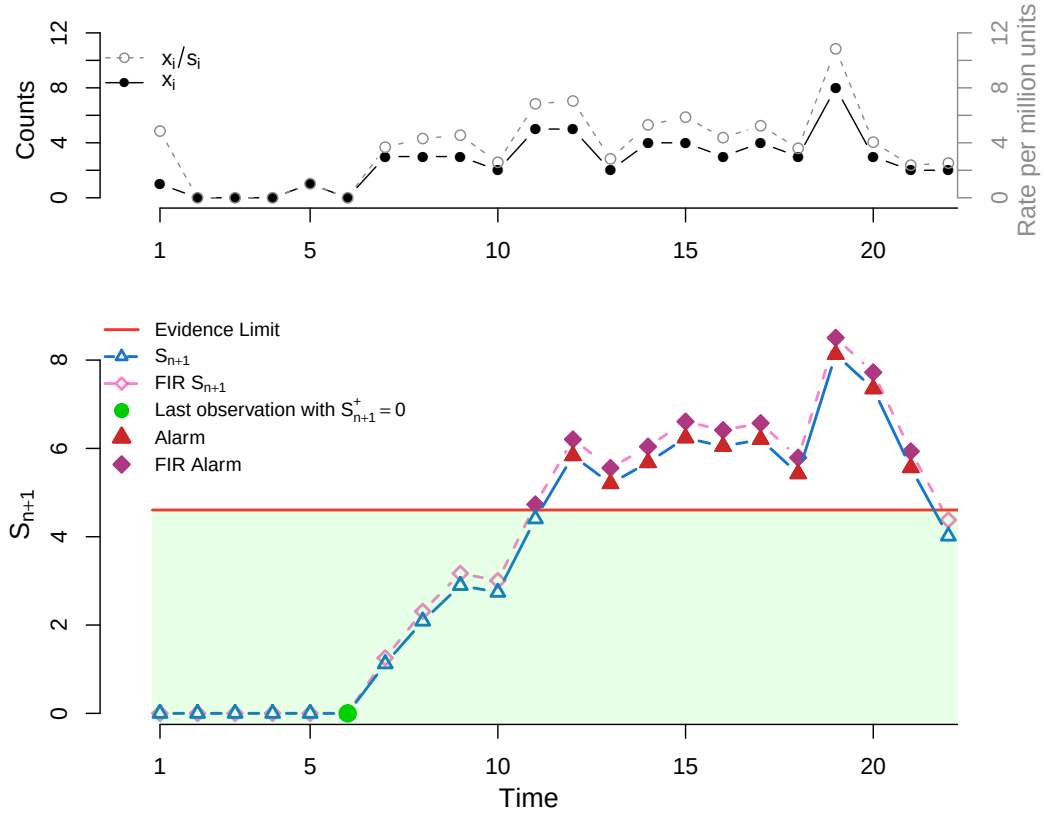


Figure 3.4.2: PRC for Poisson data. At the top panel we plot the counts of adverse events x_i (solid line) and the rate of adverse events per million units x_i/s_i (dashed line). At the lower panel, we provide the PRC control chart, focused on detecting 100% rate inflation and the evidence based limit of $h_{BF} = log(100) \approx 4.605$ is used. For the FIR-PRC (dashed line) the parameters $(f, d) = (1/2, 3/4)$ were used.

Chapter 4

Self-starting Shiryaev (3S)

4.1 3S Theoretical background

The efficient online detection of a shift in short horizon data and the reliable estimation of the unknown process parameters is not a trivial problem. In this chapter, we propose a Bayesian change point scheme, named Self-Starting Shiryaev (3S), providing all the assumptions and the methodological framework to handle either univariate (U3S) or multivariate (M3S) data. Especially for M3S, the high-dimensional version of 3S, we build it so that it can achieve the desired detection properties of a multivariate chart, i.e. to be able to identify when the process parameters experience directional invariance, anisotropic scaling or rotation, by using directional statistics. We assume that the observations are normally distributed, but potentially the proposed mechanism can be available for every distribution. The monitoring in the proposed methodology is based on the posterior marginal probability of a change point occurrence. 3S is a generalization of the Shiryaev's process (Shiryaev, 1963), relaxing the strict assumption of known parameters and offering a more flexible prior for the change point.

3S, as all the proposed methods in this dissertation, is a self-starting scheme and

therefore it does not require a phase I calibration and it is powerful in detecting process disturbances from the start of the process, utilizing prior information when available. Apart from the online testing and monitoring, our proposal provides posterior inference for all the unknown parameters of the IC and the OOC state and the change point. Additionally, we propose an adaptive decision limit, which is more realistic and better interpretable compared to a constant. An extensive simulation study evaluates our proposal against frequentist based and nonparametric alternatives, and performs a prior sensitivity analysis. Namely, we compare U3S against the Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998) and the Recursive Segmentation and Permutation (RS/P, Capizzi and Masarotto, 2013) for mean and scale shifts. In the multivariate case we compare M3S against the Self-Starting Multivariate EWMA SSMEWMA, (Hawkins and Maboudou-Tchao, 2007) and the Self-Starting CUSCORE (SSCUSC$^{(1)}$ Capizzi and Masarotto, 2010) for drifts in the mean vector. Further, we provide three applications in total to illustrate its use in practice. Two for univariate datasets, with a mean and a variance shift respectively, and one for a mean vector shift in multivariate data. All the technical details of the modelling and the applications are analytically provided in the provided in detail in Appendices.

4.1.1 3S methodological framework

Sequential change point methods in SPC/M aim to detect a change from the IC state of a distribution, as soon as it occurs, while keeping a predetermined tolerance in False Alarms (FA). Standard Shiryaev's process is a Bayesian sequential change point method, which is based on the posterior probability of a change point occurrence, given the dataset. More analytically, assume $\boldsymbol{x_n} = (x_1, x_2, ..., x_n)$ is a random sample of data, obtained sequentially. The known IC and OOC distributions are denoted by f_0 and f_1 respectively, not necessarily of the same parametric form. Regarding the unknown change point τ, it is assumed $\tau \sim G(p)$. Then, the likelihood combines

both the IC and OOC states, conditional on the time of a permanent shift. Namely, it is given by:

$$f\left(\boldsymbol{x_n}|\tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f_0\left(x_i\right) \prod_{i=\tau}^{n} f_1\left(x_i\right) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f_0\left(x_i\right) & \text{if } \tau > n \end{cases} \tag{4.1.1}$$

The stopping time $T(p^*)$ is the first observation for which the posterior probability of a change point occurrence exceeds a predetermined threshold that was chosen based on the desired FA tolerance, i.e.:

$$T(p^*) = inf\left\{n \geq 1 : p\left(\tau \leq n|\boldsymbol{x_n}\right) \geq p^*\right\} \tag{4.1.2}$$

Standard Shiryaev's process has certain optimality properties, as it minimizes the Average Detection Delay $ADD(\tau) = E_\tau\left(T - \tau|T \geq \tau\right)$ (Pollak, 1985), i.e. the average delay until the stopping time. Pollak and Tartakovsky (2009) investigated more optimality properties of the Shiryaev's process. It is worth mentioning that $ADD(\tau)$ is more realistic as a performance measure versus the widely used Average Run Length (ARL), which in fact cannot be applied in self-starting procedures.

Despite the fact that the ShiryaevâĂŹs process is known to be a very powerful method when compared to several alternatives, it has somewhat restrictive assumptions and thus there is room for improving the existing methodology by relaxing some these assumptions. Probably, the most important is that it is assumed that both of the IC (pre-change) density f_0 and the OOC (post-change) density f_1 are known. This is a very strict and rather unrealistic assumption. In practice, f_0 could be known under specific conditions, but it is extremely restrictive to assume that f_1 will be known in advance. However, only under the assumption of known distributions the Shiryaev's process will have the optimal properties, while the problem is open when the distribution will involve unknown parameters. In addition to this, the assumption of known OOC (post-change) density f_1 is rather unrealistic for multivariate data

processes. In the univariate case, f_1 simply yields a benchmark of the OOC state. For instance, if we are interested in detecting shifts for the mean, then the possible shifts (of predetermined magnitude) are either upward or downward. Therefore, we can easily build up two corresponding settings. But, as the data dimension increases, the number of the possible OOC scenarios for all components increases exponentially fast with the dimensionality. This is a plausible reason for the lack of a multivariate generalization of Shiryaev process, while there is a plethora of directionally invariant score based multivariate CUSUMs, like Crosier (1988) or Pignatiello and Runger (1990), and EWMAs, like Lowry et al. (1992). Another standard assumption of Shiryaev's process is that the prior distribution used for the (unknown) change point parameter is Geometric, i.e. $\tau \sim G(p)$. In other words, the prior probability on the location of the change point is constant over all data. However, in real world problems, the risk of failure in any process, is rather unrealistic to assume that it is constant over time.

We propose 3S, aiming to generalize the classical Shiryaev's approach in either one or more dimesions. Being self-starting, it will allow to test if a change is present and provide inference from the very early start of the data collection. Precisely, 3S will test if the procedure deviates from the IC state and simultaneously it will provide on-line estimates for all the unknown parameters. Thus, we will relax the strict assumptions of the standard Shiryaev's process enriching the methodology in four ways:

- allowing both the IC parameter(s) $\boldsymbol{\theta}$ and the OOC parameter(s) $\boldsymbol{\phi}$ to be unknown,
- allowing the OOC scenarios to fulfill certain desired properties in multivariate processes, namely: directional invariance, anisotropic scaling or rotation,
- offering a more flexible and general prior distribution for the change point τ and

- providing posterior inference for all the unknown parameters regarding the IC or the OOC scenario.

We will focus our attention on developing models for Normal data, but the description is general enough to cover any other (discrete or continuous) distribution. Namely, instead of the known IC and OOC distributions f_0 and f_1 respectively, we assume a distribution f, which reflects the IC state, allowing the IC parameters $\boldsymbol{\theta}$ (e.g. mean or variance) to be unknown. For the OOC state, we assume the OOC parameters $\boldsymbol{\phi}$ to be also unknown. The OOC parameters are suitably defined depending both on the dimension of the data and the type of the shift we are interested in detecting. For example, it may be the magnitude of a mean step change for univariate data, or the distance of a translocation for multivariate data. The IC and the OOC parameters are connected via the function $g(\boldsymbol{\theta}, \boldsymbol{\phi})$, which links the IC with the OOC scenario. In Subsections 4.1.2 and 4.1.3, we provide examples for the function g for univariate and multivariate models respectively. Introducing unknown parameters in the general model, we can build up a hierarchical prior setting, avoiding the strict assumption of known distributions. The general form of the likelihood combines both the IC and OOC scenarios, conditional on the change point occurrence τ. Specifically, denoting by $\boldsymbol{x_n}$, the random vector of the univariate or multivariate data up to time n, the likelihood will be:

$$f\left(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau\right) = \begin{cases} f\left(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n\right) & = \prod_{i=1}^{\tau-1} f\left(x_i|\boldsymbol{\theta}\right) \prod_{i=\tau}^{n} f\left(x_i|g(\boldsymbol{\theta}, \boldsymbol{\phi})\right) & \text{if } \tau \leq n \\ f\left(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n\right) & = \prod_{i=1}^{n} f\left(x_i|\boldsymbol{\theta}\right) & \text{if } \tau > n \end{cases} \tag{4.1.3}$$

Moreover, we will propose a general prior for the change point τ, relaxing the assumption of a constant probability over time. Differentiating from Shiryaev's approach, the prior distribution of the change point is a Discete Weibull (DW), i.e. $\tau \sim DW(p, \beta)$ a generalization of $G(p)$, where the parameter β represents the wear

out effect of a system, while p represents the probability of moving to the OOC state in a single observation. Thus, using $DW(p, \beta)$ we can control the prior probability of a change point, relaxing the assumption of constant probability. Precisely, for $\beta > 1$ the hazard function of the procedure increases, for $\beta < 1$ decreases, while for $\beta = 1$ it is constant, i.e. $DW(p, 1)$ coincides with $G(p)$. Therefore, the prior distribution for τ used in the standard Shiryaev's process is a special case of the corresponding distribution of 3S. In a change point model, the hazard function represents the risk that the next data point will be the first OOC observation given that the change point did not occur yet, i.e. all the previous observations are IC. The combination of p and β that we choose allows us to control the evolution of the hazard function and, consequently, to be more flexible. Specifically, we can enhance the detection ability of 3S when a shift detection is of utmost importance. For example, if the detection of shift at the very early stage is crucial, then a choice of large p and $\beta < 1$ is recommended. On the other hand, if the wear out effect of the process is slow and we are interested in detecting shifts at later stages, then a choice of small value for p and $\beta > 1$ is preferable. Moreover, if we lack prior knowledge regarding both the the location of the change point and the importance of the location of the detection, then a choice of a prior with constant hazard function seems plausible. Figure 4.1.1 provides a graphical representation of the hazard function for three different priors. Apparently, posterior inference for the location of the first OOC observation, i.e. the change point τ, will be available.

The prior information (if available), is of vital importance, especially for short runs, as it can help to boost the performance. Regarding the IC or OOC parameters, we recommend using the general class of power priors (Ibrahim and Chen, 2000). The structural advantage of power priors is that they can combine different sources of potentially available information. For more information regarding the power priors please refer to Section 2.1. In cases of total prior ignorance, we need to select a suitable non-informative prior that will balance the detection power and the false

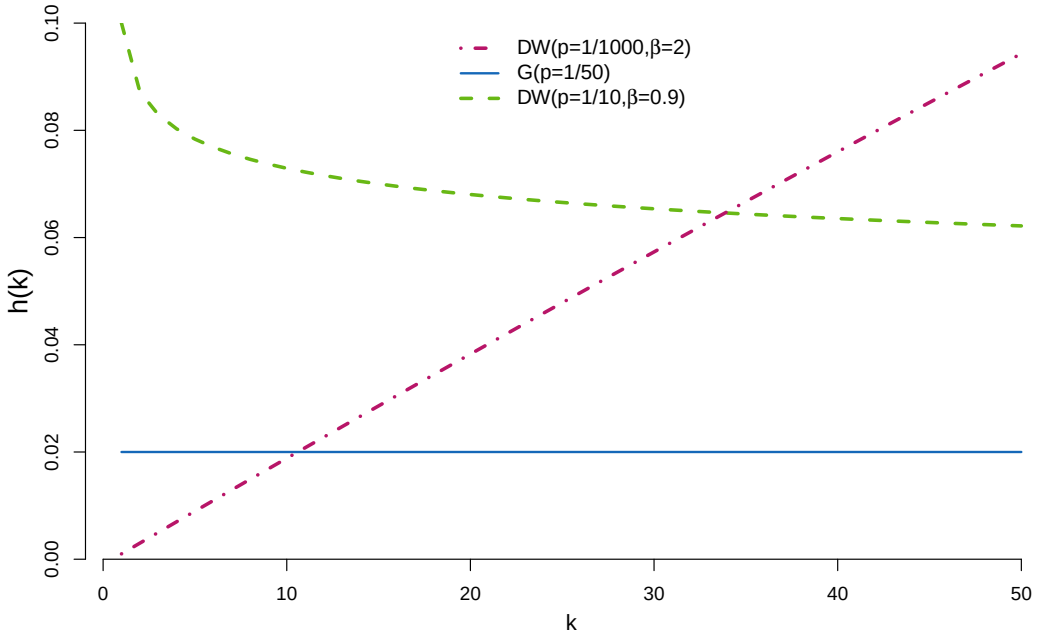


Figure 4.1.1: The hazard function for the priors $G(1/50)$, $DW(1/10, 9/10)$ and $DW(1/1000, 2)$, which are represented by the solid, the dashed and the dotted-dashed line respectively.

alarm tolerance, aiming to optimize the 3S performance. These cases are separately analyzed for univariate and multivariate processes in Subsections 4.1.2 and 4.1.3 respectively.

For the stopping time, it will be based on the posterior marginal probability of a change point occurrence, in an analogous manner to the classical Shiryaev process. Precisely:

$$p\left(\tau \leq n | \boldsymbol{x_n}\right) = \frac{f\left(\boldsymbol{x_n} | \tau \leq n\right) \pi(\tau \leq n)}{f\left(\boldsymbol{x_n} | \tau \leq n\right) \pi(\tau \leq n) + f\left(\boldsymbol{x_n} | \tau > n\right) \pi(\tau > n)} \tag{4.1.4}$$

In Section 4.2, we will provide all the details regarding the stopping rule options along with the respective decision limits. The marginal distributions involved in the computation of the probability (4.1.4), will be derived by integrating out the

unknown parameters. Precisely, for the IC and OOC scenarios we have respectively:

$$f(\boldsymbol{x_n}|\tau > n) = \int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \tau > n)\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{4.1.5}$$

$$f(\boldsymbol{x_n}|\tau \leq n) = \int_{\Phi}\int_{\Theta} f(\boldsymbol{x_n}|\boldsymbol{\theta}, \boldsymbol{\phi}, \tau \leq n)\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi} \tag{4.1.6}$$

However, these marginal will be undefined, when the prior $\pi(\boldsymbol{theta})$ is improper, for instance when a non-informative prior is adopted. In this case, we recommend to "sacrifice" the first s observations $\boldsymbol{x_s}$ necessary to make the posterior $p(\theta|\boldsymbol{x_s})$ proper and then use it instead of the initial prior $\pi(\theta)$ starting the process testing from the $s+1$ observation Thus, x_s will be used for the calibration and not for testing, but this is necessary to initiate a non-informative 3S scheme. When the process starts, inference for $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and τ becomes available by sampling from the corresponding full conditional posteriors. Therefore, we will not only be able to perform sequential testing if the procedure deviates from the IC state, and we will also be available to provide online estimates for the parameters of interest, with respect to the IC or the OOC scenario. The Directed Acyclic Graph (DAG) in Figure 4.1.2 synopsizes the general 3S scheme.

The general form of 3S allows its use for any type of parametric (i.e. distributional) setting, as long as the various marginal and posterior distributions indicated in the DAG of Figure 4.1.2 are computed (in closed form or numerically). Apart from its general form, the great advantage of 3S is its resistance in absorbing a change. In general, it is well known and documented in the literature, that the self-starting schemes face a big challenge. They have only a "small window of opportunity" to react in a change before they absorb it. Specifically, if a change from the IC state occurs and a self-starting process does not realize it "soon" after its occurrence, then the OOC data are involved in the calibration, contaminating the estimates of the unknown parameters. This affects the performance dramatically, as the chart absorbs

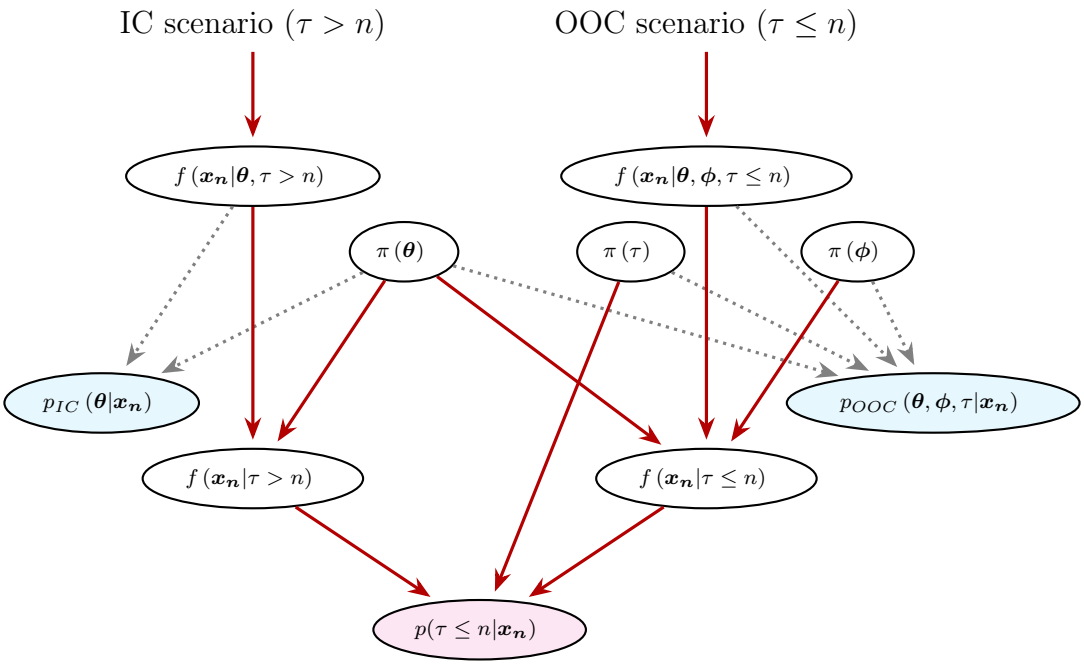


Figure 4.1.2: The DAG of the 3S process. The IC unknown parameters are denoted as $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ represents the OOC parameters, while τ is the change point. Combining the likelihoods and the priors, we obtain the corresponding marginals and consequently the posterior marginal of a change point occurrence $p(\tau \leq n|\boldsymbol{x_n})$. In addition, estimates for the unknown parameters are available by sampling from the corresponding posteriors of the IC or the OOC scenario respectively.

the change and considers the contaminated data as IC data. By construction, 3S is quite resistant in absorbing a change, as it splits the data appropriately, avoiding the involvement of the contaminated data in the IC estimates. This is a great advantage comparing with standard self-starting alternatives. Furthermore, if the change arrives at the early start, then self-starting processes have typically very poor performance, as the estimates for the unknown parameters are still vague. On the contrary, 3S can be efficient, even from the early start, thanks to its structure and the use of prior distributions.

4.1.2 U3S modelling

In this Subsection, we develop the specifics of the univariate models. U3S, is the univariate generalization of Shiryaev's (1963) process, offering a more flexible and realistic set up. It is worth mentioning that Shiryaev's change point model is a special case of U3S, when:

- the IC distribution f_0 and OOC f_1 are of the same parametric form (i.e. distribution),
- the IC parameters $\boldsymbol{\theta}$ and the OOC $\boldsymbol{\phi}$ are known, i.e. their U3S's respective prior distributions are set both to be point mass distributions, and
- the hazard function for change point τ is assumed to be constant, by setting the prior distribution to be DW with $\beta = 1$ (i.e. geometric).

In U3S, the unknown OOC parameters $\boldsymbol{\phi}$ are linked to the IC parameters θ via the function $g(\theta, \phi)$, which expresses the OOC scenario under study. For example, in the case of a Normal likelihood with $\boldsymbol{\theta} = (\theta_1, \theta_2^2)$ being the IC unknown parameters for the mean and the variance of the data respectively and assuming that we wish to guard against mean shifts, then we can have $\boldsymbol{\phi} = \delta$ being the magnitude of a mean shift and we can define $g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \theta_1 + \delta \cdot \theta_2$. For $\boldsymbol{\phi}$, we recommend the use of at least a weakly informative prior, which will express the size of the shift that we are more interested to detect and will play the role of a benchmark for the OOC state. Generally, an informative prior for $\boldsymbol{\phi}$ helps to "distinguish" the IC and OOC states, making U3S more robust and enhancing its detection ability, compared with the U3S scheme that adopts a totally non-informative prior. It is worth mentioning, that in the U3S model the OOC parameters are designed to be unitless. For instance a jump for the mean is expressed as a multiple of the standard deviation, even when the latter is unknown. This means we can set an informative prior for $\boldsymbol{\phi}$, even when we lack any information regarding $\boldsymbol{\theta}$. In the case that we are interested in detecting shifts of either direction, e.g. a positive or negative step change for the mean, or an inflation

or shrinkage for the variance, we recommend using a mixture of priors for $\boldsymbol{\phi}$. Worth's mentioning that if the priors in such a mixture modelling have significant overlap, e.g. the prior that represents a positive jump has significant probability in negative values, then some form of truncation is recommended, primarily for interpretation purposes.

In cases of total prior ignorance, we propose the adoption of the reference priors (Bernardo, 1979, Berger et al., 2009) for the IC parameters $\boldsymbol{\theta}$. Reference priors is a general class of non-informative priors that will coincide with Jeffreys priors (Jeffreys, 1961), when $\boldsymbol{\theta}$ is one-dimensional. Table 4.1.1 provides a synopsis of the four versions of U3S model for Normal data, for location or scale shifts, while in Appendix C, they are presented analytically. Precisely, we provide the assumptions about the likelihood and the priors and the resulting formulas about the posterior marginal probability of a change point occurrence and the posterior distributions under the IC and the OOC scenario.

4.1.3 M3S modelling

In this Subsection we will propose the multivariate version of 3S, i.e. the Multivariate Self-Starting Shiryaev (M3S). The focus will be on the detection of persistent shifts in the mean vector or the covariance matrix of short horizon multivariate Normal data. In a similar way with the univariate case, the stopping time will be based on the posterior marginal probability of a change point occurrence $p(\tau \leq n|\boldsymbol{x_n})$, remaining in the At Most One Change (AMOC) scenario. In the multivariate case, the change point τ will refer to the time that at least one component of the mean vector or the covariance matrix, shifts from its IC state. Therefore, from now on the change point will be the first observation when at least one component of the mean vector or the covariance matrix shifts from the IC to the OOC state. For instance, in the bivariate Normal distribution with both mean vector and covariance matrix

IC State $\boldsymbol{\theta}$ Prior	**OOC State** $\boldsymbol{\phi}$ Prior(s)	$\frac{f(\boldsymbol{x_n}\|\tau \leq n)}{f(\boldsymbol{x_n}\|\tau > n)}$
$\boldsymbol{N}(\theta, \sigma^2)$	$\boldsymbol{N}(\theta + \delta \cdot \sigma, \sigma^2)$	$\sum_{i=1}^{2} \frac{\pi_i \sigma}{\sqrt{e_{\delta i}}} \cdot exp\left\{ -\frac{\mu_{\delta i}^2}{2\sigma_{\delta i}^2} + \frac{\left(\sigma_{\delta i}^2 \left(X_{\tau:n} - n_\tau \mu_p\right) \sigma + \sigma^2 \mu_{\delta_i}\right)^2}{2\sigma_{\delta_i}^2 \sigma^2 e_{\delta_i}} \right\}$
θ	$\delta = \gamma \cdot \delta_1 + (1-\gamma) \cdot \delta_2$	
$\theta \sim N(\mu_0, \sigma_0^2)$	$\delta_i \sim N(\mu_{\delta i}, {\sigma_{\delta i}}^2)$, $\gamma \sim Ber(\pi)$	
$n_t = n - t + 1$, $X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i$, $\sigma_p^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}$, $\mu_p = \left(\frac{X_{1:n}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \cdot \sigma_p^2$, $e_{\delta i} = n_\tau \sigma_{\delta i}^2 \left(\sigma^2 - n_\tau \sigma_p^2\right) + \sigma^2$, $\pi_1 = \pi$, $\pi_2 = 1 - \pi$		
$\boldsymbol{N}(\mu, \theta^2)$	$\boldsymbol{N}(\mu, \kappa \cdot \theta^2)$	$\frac{\Gamma(a_p)}{b_n^{a_p}} \sum_{i=1}^{2} \frac{\pi_i d_i^{c_i} \Gamma(n_\tau/2 + c_i)}{\Gamma(c_i)} \int_{\theta^2} \left(\frac{1}{\theta^2}\right)^{a_p+1} \cdot \left(d_i + \frac{S_{\tau:n}^2}{2\theta^2}\right)^{-(n_\tau/2+c_i)} \cdot exp\left\{ -\frac{b_{\tau-1}}{\theta^2} \right\} d\theta^2$
θ^2	$\kappa = \gamma \cdot \kappa_1 + (1-\gamma) \cdot \kappa_2$	
$\theta^2 \sim IG(a, b)$	$\kappa_i \sim IG(c_i, d_i)$, $\gamma \sim Ber(\pi)$	
$n_t = n - t + 1$, $S_{t_1:t_2}^2 = \sum_{i=t_1}^{t_2} (x_i - \mu)^2$, $a_p = a + \frac{n}{2}$, $b_t = b + \frac{S_{1:t}^2}{2}$, $\pi_1 = \pi$, $\pi_2 = 1 - \pi$		

$\boldsymbol{N}(\theta_1, \theta_2^2)$	$\boldsymbol{N}(\theta_1 + \delta \cdot \theta_2, \theta_2^2)$	
(θ_1, θ_2^2)	$\delta = \gamma \cdot \delta_1 + (1-\gamma) \cdot \delta_2$	$\frac{b_p^{a_p}}{\Gamma(a_p)} \sum_{i=1}^{2} \frac{\pi_i}{\sigma_{\delta i}\sqrt{e_{\delta i}}} \cdot exp\left\{-\frac{\mu_{\delta i}^2}{2\sigma_{\delta i}^2}\right\} \int_{\theta_2^2} \left(\frac{1}{\theta_2^2}\right)^{a_p+1} \cdot exp\left\{-\frac{b_p}{\theta_2^2} + \frac{\left(\frac{X_{\tau:n} - n_\tau \mu_p}{\theta_2} + \frac{\mu_{\delta i}}{\sigma_{\delta i}^2}\right)^2}{2e_{\delta i}}\right\} d\theta_2^2$
$(\theta_1, \theta_2^2) \sim NIG(\mu_0, \lambda_0, a, b)$	$\delta_i \sim N(\mu_{\delta i}, {\sigma_{\delta i}}^2),\ \gamma \sim Ber(\pi)$	

$$n_t = n - t + 1,\ X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i,\ S^2_{t_1:t_2} = \sum_{i=t_1}^{t_2} (x_i - \theta_1)^2,\ \mu_p = \frac{\lambda\mu_0 + X_{1:n}}{\lambda + n},\ \lambda_p = \lambda + n,\ a_p = a + \frac{n}{2},\ b_p = b + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n}{\lambda_p}(\bar{x} - \mu_0)^2\right),$$
$$e_{\delta i} = n_\tau\left(1 - \frac{n_\tau}{\lambda_p}\right) + \frac{1}{\sigma_{\delta i}^2},\ \pi_1 = \pi,\ \pi_2 = 1 - \pi$$

$\boldsymbol{N}(\theta_1, \theta_2^2)$	$\boldsymbol{N}(\theta_1, \kappa \cdot \theta_2^2)$	
(θ_1, θ_2^2)	$\kappa = \gamma \cdot \kappa_1 + (1-\gamma) \cdot \kappa_2$	$b_p^{a_p}\sqrt{\lambda_p} \sum_{i=1}^{2} \frac{\pi_i d_i^{c_i}}{\Gamma(c_i)} \int_{\kappa_i} \frac{b_{\tau,\kappa_i}^{-a_p} \cdot \kappa_i^{-(n_\tau/2 + c_i + 1)}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} exp\left\{-\frac{d_i}{\kappa_i}\right\} d\kappa_i$
$(\theta_1, \theta_2^2) \sim NIG(\mu_0, \lambda_0, a, b)$	$\kappa_i \sim IG(c_i, d_i),\ \gamma \sim Ber(\pi)$	

$$n_t = n - t + 1,\ n_{t,k} = t - 1 + \frac{n_t}{k},\ X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i,\ X^2_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i^2,\ S^2_{t_1:t_2} = \sum_{i=t_1}^{t_2} (x_i - \theta_1)^2,\ \mu_p = \frac{\lambda\mu_0 + X_{1:n}}{\lambda + n},\ \lambda_p = \lambda + n,\ a_p = a + \frac{n}{2},$$
$$b_p = b + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n}{\lambda_p}(\bar{x} - \mu_0)^2\right),\ b_{t,k} = b + \frac{1}{2}\left(\lambda\mu_0^2 + X^2_{1:(t-1)} + \frac{X^2_{t:n}}{k} - \frac{\left(\lambda\mu_0 + X_{1:(t-1)} + \frac{X_{t:n}}{k}\right)^2}{\lambda + n_{t,k}}\right),\ \pi_1 = \pi,\ \pi_2 = 1 - \pi$$

Table 4.1.1: The U3S scheme for changes in the mean or the variance of Normal data, assuming a two-prior mixture for the OOC parameter $\boldsymbol{\phi}$.

being unknown, the change point τ will denote the time where at least one of the five involved parameter's components: $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ shifts.

Probably the most challenging part of the multivariate extension of 3S is the appropriate expression of the OOC scenarios in high dimensions. This is achieved by the beneficial use of the directional statistics (see Mardia and Jupp, 2009) and the selection of prior distributions for $\boldsymbol{\phi}$ that fulfill the desired properties. The models are carefully designed in order not only to allow the components to shift in their parameter space but to be interpretablle as well. Namely, the desired properties of M3S can be synopsized to:

- Directional Invariance, i.e. M3S can detect changes in any direction.
- Anisotropic scaling, i.e. M3S can detect different scale changes to each dimension or weight a shift by the variance of the corresponding component.
- Rotation, i.e. M3S can detect changes in the correlation between the variables, either positive or negative.

Figure 4.1.3 provides the graphical representation of these properties. M3S can detect changes in the means (translocation in any direction), the variances (shrinking or inflation) and in the correlation (angle drifting).

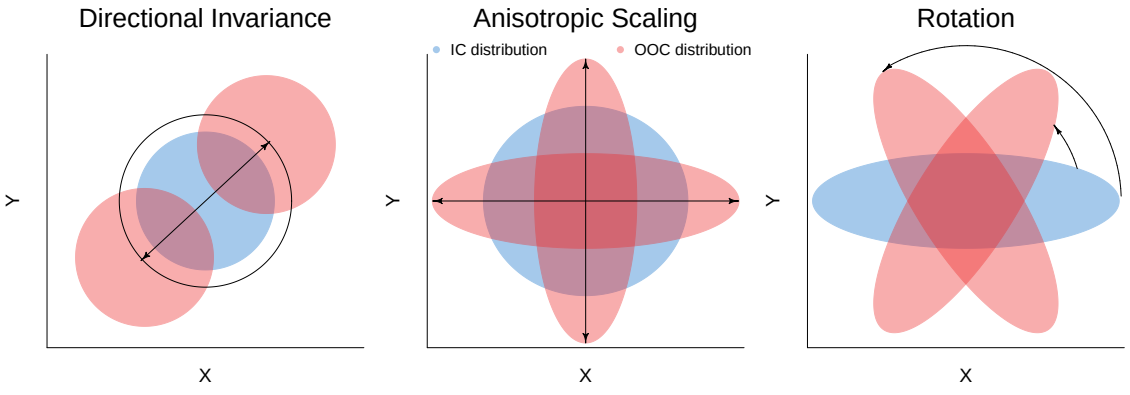


Figure 4.1.3: The graphical representation of the directional invariance (left panel), the anisotropic scaling (center panel) and the rotation (right panel) in two dimensions. The IC distribution is in blue, while the OOC distribution is in red.

As 3S is a general detection scheme, the M3S modelling is similar to the univariate case. This means that the process for the elicitation of $p(\tau \leq n|\boldsymbol{x_n})$ and the posteriors of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ remains the same. Similarly, the decision limits described in Subsection 4.2, and the corresponding stopping times $T(\cdot)$ are defined in an identical way. With respect to the IC parameters of the proposed M3Ss, just as before we have the mean vector and the covariance matrix of Normal data, i.e. $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. While the prior setting for the IC parameters is straightforward for the mean vector and the covariance matrix model, apparently this is not the case for $\boldsymbol{\phi}$, the parameters that represent the OOC scenarios. It is important to mention that in the multivariate case, the OOC parameters do not only play the role of the shifted parameters, but may also describe the association between the shifts of the components. For example, in M3S for the mean vector, $\boldsymbol{\phi}$ does not merely denote the distance of the IC and the OOC state, but the direction of the translocation as well, describing in this way the dependence between the individual drifts of the components of $\boldsymbol{\mu}$.

In case of prior ignorance, several types of objective priors can be proposed for the mean vector and the covariance matrix of a multivariate Normal distribution, based on different criteria. Considering that we do not have any available prior information, either from historical data or prior beliefs, we propose the use of the Jeffreys prior (1961) for M3S, which coincides with the prior introduced by Geisser and Cornfield (1963) for two-dimensional data. In D-dimensions, the Jeffreys prior is:

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(D+2)/2} \tag{4.1.7}$$

For further information about the non-informative priors in the multivariate case, one can refer to Sun and Berger (2007). All the details regarding the prior setting, the resulting posteriors and the calculation of the BF_{τ,n_+} of the referred M3S models are analytically provided in Appendix D, while Table 4.1.2 presents the models briefly.

IC State $\boldsymbol{\theta}$ Prior	**OOC State** $\boldsymbol{\phi}$ Prior(s)	$\frac{f(\boldsymbol{x_n}\|\tau \leq n)}{f(\boldsymbol{x_n}\|\tau > n)}$
$\boldsymbol{N_D(\mu,\Sigma)}$ $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ $(\boldsymbol{\mu},\boldsymbol{\Sigma}) \sim NIW(\boldsymbol{\mu_o},\lambda,\nu_0,\boldsymbol{\Psi})$	$\boldsymbol{N_D(\mu + rL^{1/2}T_\theta,\Sigma)}$ $r\boldsymbol{T_\theta}$ $r \sim NC_{\chi_D}(d,\sigma_d^2)$, $\boldsymbol{T_\theta} \sim vMF(\boldsymbol{\mu_\theta},\kappa)$	$\frac{\kappa^{D/2-1}\cdot d^{1-D/2}}{\sigma_d^2\cdot\sqrt{\pi}\cdot 2^{(\nu_p D+1)/2}}\cdot\left(\frac{\lambda}{\lambda_p}\right)^{D/2}\cdot\frac{\lvert\boldsymbol{\Psi_p}\rvert^{\nu_p/2}}{\Gamma_D\left(\frac{\nu_p}{2}\right)}\cdot\frac{I_{D/2-1}\left(\frac{rd}{\sigma_d^2}\right)}{I_{D/2}(\kappa)}\cdot exp\left\{-\frac{d^2}{2}\right\}\int_r\int_\theta\int_{\boldsymbol{\Sigma}}\frac{r^{3D/2-1}\prod_{j=0}^{D-2}[sin\theta_{j+1}]^{D-2-j}}{\lvert\boldsymbol{\Sigma}\rvert^{(\nu_p+D+1)/2}}$ $\times exp\left\{-\frac{1}{2}(\boldsymbol{A_\tau}+r^2)+\boldsymbol{B_\tau}+\kappa\boldsymbol{\mu_\theta}^T\boldsymbol{T_\theta}-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi}+\boldsymbol{x_i}\boldsymbol{x_i}^T+\lambda\boldsymbol{\mu_o}\boldsymbol{\mu_o}^T\right)+\frac{\lambda_p}{2}tr\boldsymbol{\Sigma}^{-1}\lambda_p\left(\boldsymbol{\mu_p}-\frac{n_\tau r\boldsymbol{L}^{1/2}\boldsymbol{T_\theta}}{\lambda_p}\right)\left(\boldsymbol{\mu_p}-\frac{n_\tau r\boldsymbol{L}^{1/2}\boldsymbol{T_\theta}}{\lambda_p}\right)^T\right\}d\boldsymbol{\Sigma}d\theta dr$
	$\boldsymbol{T_\theta}=\begin{pmatrix}cos\theta_1\\ \vdots\\ cos\theta_{D-1}\prod_{i<D-1}sin\theta_i\\ \prod_{i=1}^{D-1}sin\theta_i\end{pmatrix}$, $\boldsymbol{L}=diag(\boldsymbol{\Sigma})$, $\boldsymbol{C_n}=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_{\boldsymbol{n}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_{\boldsymbol{n}})^T$, $\boldsymbol{C_o}=(\bar{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_o})(\bar{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_o})^T$, $n_t=n-t+1$, $\boldsymbol{A_t}=n_t\boldsymbol{T_\theta}^T\boldsymbol{L}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{1/2}\boldsymbol{T_\theta}$, $\boldsymbol{B_t}=\sum_{i=t}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{1/2}\boldsymbol{T_\theta}$, $\boldsymbol{\mu_p}=\frac{n\bar{\boldsymbol{x}}_{\boldsymbol{n}}+\lambda\boldsymbol{\mu}_0}{n+\lambda}$, $\lambda_p=n+\lambda$, $\nu_p=n+\nu_0$, $\boldsymbol{\Psi_p}=\boldsymbol{C}+\boldsymbol{\Psi}+\frac{\lambda n}{\lambda+n}\boldsymbol{C_o}$	
$\boldsymbol{N_D(\mu,\Sigma)}$ $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ $(\boldsymbol{\mu},\boldsymbol{\Sigma}) \sim NIW(\boldsymbol{\mu_o},\lambda,\nu_0,\boldsymbol{\Psi})$	$\boldsymbol{N_D(\mu,T\Sigma T^T)}$ $\{diag(\boldsymbol{S}),\theta\}$ $\boldsymbol{S} \sim IW(\nu_s,\boldsymbol{D})$, $\theta_{\alpha,\beta} \sim vM(\boldsymbol{\mu}_{\alpha,\beta},\kappa_{\alpha,\beta})$ for $\alpha=1$ and $\beta>\alpha$ and $\theta_{\alpha,\beta} \sim \pi vM(\boldsymbol{\mu}_{\alpha,\beta},\kappa_{\alpha,\beta})$ otherwise	$\left(\frac{\lambda_p}{2^{\nu_p+\nu_s}}\right)^{D/2}\cdot\frac{\lvert\boldsymbol{\Psi_p}\rvert^{\nu_p/2}}{\Gamma_D\left(\frac{\nu_p}{2}\right)}\cdot\frac{\lvert\boldsymbol{D}\rvert^{\nu_s/2}}{\Gamma_D\left(\frac{\nu_s}{2}\right)}\cdot\frac{1}{\pi^{(D-1)(D-2)/2}}\prod_{\alpha=1,\beta=\alpha+1}^{D-1}\prod^{D}\frac{1}{I_0(\kappa_{\alpha,\beta})}\int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\frac{\lvert\boldsymbol{\Sigma_p}\rvert^{1/2}}{\lvert\boldsymbol{S}\rvert^{(n_\tau+\nu_s+D+1)/2}\cdot\lvert\boldsymbol{\Sigma}\rvert^{(\nu_p+D+1)/2}}$ $\times exp\left\{-\frac{1}{2}tr\boldsymbol{D}\boldsymbol{S}^{-1}+\kappa_{\alpha,\beta}\cdot cos\left(1+\mathbb{1}_{\{\alpha>1\}}\right)(\theta_{\alpha,\beta}-\kappa_{\alpha,\beta})-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi}+\boldsymbol{x_i}\boldsymbol{x_i}^T+\lambda\boldsymbol{\mu_o}\boldsymbol{\mu_o}^T-\lambda_p\boldsymbol{M_p}\boldsymbol{M_p}^T\right)\right\}d\boldsymbol{\Sigma}d\boldsymbol{T}$
	$\theta=(\theta_{1,1},...,\theta_{D-1,D})$, $\boldsymbol{T}=\boldsymbol{R}(\theta)\boldsymbol{S}^{1/2}$, $\boldsymbol{R}(\theta)=\prod_{\alpha=1}^{D-1}\prod_{\beta=\alpha+1}^{D}\boldsymbol{R_{\alpha,\beta}}(\theta_{\alpha,\beta})$, $\boldsymbol{R_{\alpha,\beta}}(\theta_{\alpha,\beta})=\left(\begin{array}{c:c}\multirow{3}{*}{$r_{i,j}$} & r_{\alpha,\alpha}=cos(\theta_{\alpha,\beta})\\ & r_{\beta,\beta}=cos(\theta_{\alpha,\beta})\\ & r_{\alpha,\beta}=-sin(\theta_{\alpha,\beta})\\ & r_{\beta,\alpha}=sin(\theta_{\alpha,\beta})\\ \hdashline r_{i,j}=1,\ i=j,\ i\neq\{\alpha,\beta\} & \\ r_{i,j}=0,\ elsewhere & \end{array}\right)$, $\boldsymbol{C_n}=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_{\boldsymbol{n}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_{\boldsymbol{n}})^T$, $\boldsymbol{C_o}=(\bar{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_o})(\bar{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_o})^T$, $\boldsymbol{X_{t_1:t_2}}=\sum_{i=t_1}^{t_2}\boldsymbol{x_i}$, $n_t=n-t+1$, $\boldsymbol{\mu_p}=\frac{n\bar{\boldsymbol{x}}_{\boldsymbol{n}}+\lambda\boldsymbol{\mu}_0}{n+\lambda}$, $\lambda_p=n+\lambda$, $\nu_p=n+\nu_0$, $\boldsymbol{\Psi_p}=\boldsymbol{C}+\boldsymbol{\Psi}+\frac{\lambda n}{\lambda+n}\boldsymbol{C_o}$, $\boldsymbol{M_p}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{X_{1:\tau-1}}+\lambda\boldsymbol{\mu_o})+\left(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\right)^{-1}\boldsymbol{X_{\tau:n}}$, $\boldsymbol{\Sigma_p}=\left(\left(\frac{\boldsymbol{\Sigma}}{\lambda_p-n_t}\right)^{-1}+\left(\frac{\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T}{n_t}\right)^{-1}\right)^{-1}$, $\boldsymbol{K_t}=\sum_{i=t}^{n}\boldsymbol{T}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\left(\boldsymbol{T}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right)^T$, $\boldsymbol{\Psi_{p,t}}=+\boldsymbol{\Psi}+\lambda(\boldsymbol{\mu}-\boldsymbol{\mu_o})(\boldsymbol{\mu}-\boldsymbol{\mu_o})^T+\boldsymbol{C_{t-1}}+\boldsymbol{K_t}$	

Table 4.1.2: The M3S scheme for changes in the mean vector or the covariance matrix of multivariate Normal data.

4.2 Decision making

In the sequential change point methods in SPC/M, the stopping time is defined by the stopping rule, which is essentially a stochastic decision on whether the process has experienced a change or not. Thus, the choice of the "appropriate" decision limit is crucial for every method. Bather (1967) discussed about the optimal stopping time $T(p^*)$ for the classical Shiryaev's process, where an alarm is raised when $p(\tau \leq n|\boldsymbol{x_n})$ exceeds a predetermined constant threshold p^*, a method that can be applied in the 3S scheme as well. The determination of a constant decision limit p^* for all the probabilities $p(\tau \leq n|\boldsymbol{x})$ is a plausible and simple strategy. However, this would be very conservative for the very first tests, reducing significantly the detection power for the early stages of the process. This is a major issue for self-starting methods in short runs, where we aim to have efficient performance from the start of the process.

As the prior $\pi(\tau)$ is involved in forming the posterior $p(\tau \leq n|\boldsymbol{x_n})$, it must be considered in the determination of the decision limit. Apart from 3S model novelty, we will propose a new adaptive decision limit p_n^*, which takes into account the effect of the prior at time n. The prior-adjusted threshold p_n^*will reduce the delay of an alarm, especially at the early start of the process, while respecting the required false alarm tolerance. Its derivation relies on the property that $p(\tau \leq n|\boldsymbol{x_n})$ can be written as function of prior weighted Bayes Factors (BF). More specifically, $p(\tau \leq n|\boldsymbol{x_n})$ can be written as:

$$\begin{aligned} p(\tau \leq n|\boldsymbol{x}) &= \frac{f\left(\boldsymbol{x_n}|\tau \leq n\right)\pi(\tau \leq n)}{f\left(\boldsymbol{x_n}|\tau \leq n\right)\pi(\tau \leq n) + f\left(\boldsymbol{x_n}|\tau > n\right)\pi(\tau > n)} \\ &= \frac{\frac{f\left(\boldsymbol{x_n}|\tau \leq n\right)\pi(\tau \leq n)}{f\left(\boldsymbol{x_n}|\tau > n\right)\pi(\tau > n)}}{\frac{f\left(\boldsymbol{x_n}|\tau \leq n\right)\pi(\tau \leq n)}{f\left(\boldsymbol{x_n}|\tau > n\right)\pi(\tau > n)} + 1} \end{aligned}$$

$$= \frac{\sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} \cdot BF_{k,n+}}{\sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} \cdot BF_{k,n+} + 1} \tag{4.2.1}$$

where $BF_{k,n+} = \frac{f(\boldsymbol{x_n}|\tau = k)}{f(\boldsymbol{x_n}|\tau > n)}$ compares the evidence the $k^{th} \leq n$ observation to be the change point against the evidence all the observations to be IC. We establish a decision limit, which will be less conservative for the first tests and consider the prior adjustment. As $p(\tau \leq n|\boldsymbol{x_n})$ is function of n Bayes factors, if we will replace $BF_{k,n_+} = A$, then the equation (4.2.1) will become:

$$p(\tau \leq n|\boldsymbol{x}) = \frac{A \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)}}{A \cdot \sum_{k=1}^{n} \frac{\pi(\tau = k)}{\pi(\tau > n)} + 1} \tag{4.2.2}$$

Thus, the adaptive stopping time $T\left(p_n^*\right)$ will be:

$$T\left(p_n^*\right) = inf\{n > 1 : p\left(\tau \leq n|\boldsymbol{x_n}\right) \geq p_n^*\} \tag{4.2.3}$$

By controlling A, we can control the FA tolerance via a predetermined metric, like the Probability of False Alarm, $PFA(n) = P(T \leq n|\tau > n) = \alpha$, the IC Average Run Length (ARL_0) etc. Apart from improving the overall performance, the choice of p_n^* can be interpretable according to the evidence scale that Kass and Raftery (1995) proposed, controlling the posterior FA evidence as it was defined in Section 3.2.

If we will use the $PFA(n)$ metric for a short horizon of data, then the adaptive p_n^* will be significantly more sensitive for the first observations, but a little more conservative for the last few, compared to the constant threshold p^*. The prior weights of τ play an important role in the final form of p_n^*. Additionally, the testing using p_n^* reduces

considerably the truncated Conditional Expected Delay, $tCED(\tau) = E_\tau(T - \tau + 1|\tau \leq T \leq n)$, especially when the shift occurs at the early start of a process. This is very important, as the main disadvantage of the self starting charts is the low performance and the slow reaction when the change occurs at early stages. Apart from the performance, the adaptive limit distributes in a bigger range the false alarms compared with the constant limit. Figure 4.2.1 provides a graphical representation of the evolution of p_n^* and p^* and the histograms of the corresponding false alarms. The plots refer to certain OOC scenarios for univariate processes of Subsection 4.3.1.2, using the non-informative prior setting, named *vs*, assuming $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$, $\delta|\gamma \sim \gamma \cdot N(1, 0.5^2) + (1-\gamma) \cdot N(-1, 0.5^2)$ and $\tau \sim DW(1/50, 1)$. As we can see, the stopping time of $T(p^*)$ are mostly concentrated at the end of the sample, while $T(p_n^*)$ are almost uniformly distributed after the 10^{th} observation.

We recommend the use of the adaptive decision limit, when the failure risk is non increasing, i.e. when the DW parameter $\beta \leq 1$. In this case, we will gain the beneficial properties of p_n^* at the start of a process with a negligible influence at the end. Contra wise, if $\beta \gg 1$, then the classical decision limit p^* seems to be better choice on average.

4.3 Competing methods and sensitivity analysis

In this chapter, we will compare the performance of 3S against well established alternative methods. We will provide the simulation results both for univariate and multivariate settings.

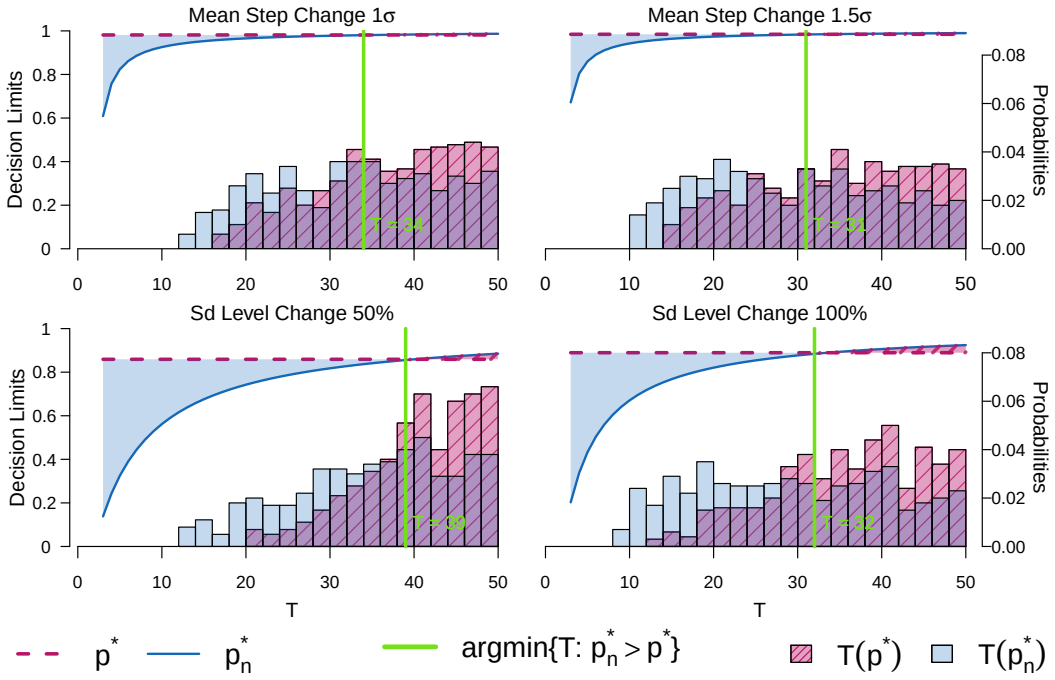


Figure 4.2.1: A graphical comparison of the evolution of the adaptive decision limit p_n^* (solid line) against the constant p^* (dashed line) for various OOC scenarios for univariate processes assuming $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$, $\delta|\gamma \sim \gamma \cdot N(1, 0.5^2) + (1-\gamma) \cdot N(-1, 0.5^2)$ and $\tau \sim DW(1/50, 1)$. The shaded region denotes where p_n^* is more sensitive, while the dashed denotes where it is more conservative. Further, the histograms of the corresponding stopping times $T(\cdot)$, i.e. the locations of the FAs, are provided. Finally, the times where the two decision thresholds are crossing are illustrated with vertical green segments.

4.3.1 Simulation study for U3S

4.3.1.1 Competing methods for U3S

In this Subsection, we will present the competing methods whose performance will be measured against the proposed U3S scheme in detecting changes in the mean or the variance of univariate Normal data. The first competing method is the frequentist Self-Starting CUSUM (SSC, Hawkins and Olwell, 1998) for location or scale shifts, which was described in Subsection 3.3.1.

A second competing method that we will use is the non-parametric scheme introduced by Capizzi and Masarotto (2013) named Recursive Segmentation and Permu-

tation (denoted as RS/P from now on). RS/P is designed to detect single or multiple mean and/or scale shifts, of individual or subgrouped observations. Let K be the maximum number of change points that we wish to detect, of m subgroups with n data points each. The control statistic for location shifts is:

$$T_0 = \max_{i=1,...,m} |\bar{x}_i - \bar{\bar{x}}| \tag{4.3.1}$$

where $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ and $\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$. Regarding the statistics T_1, ..., T_K of possible change points $0 < \tau_1 < ... < \tau_k < m$, they are computed using a simple forward recursive segmentation approach. The algorithm starts with $k = 0$ and proceeds in K successive stages. At the beginning of stage k, the interval $[1, m]$ is partitioned into k subintervals, each having a length greater or equal to l_{MIN}, which is a user-controllable constant giving the minimum number of subgroups allowed between two change points. At stage k, one of these subintervals is split, adding a new potential change point. Every new change point is estimated by maximizing the quantity:

$$\sum_{i=1}^{k+1} \left(\hat{\tau}_i - \hat{\tau}_{i-1}\right) \left(\bar{x}\left(\hat{\tau}_{i-1}, \hat{\tau}_i\right) - \bar{\bar{x}}\right)^2 \tag{4.3.2}$$

conditionally on the results of the previous stages, where $\bar{x}(a, b) = \frac{1}{b-a}\sum_{i=a+1}^{b} \bar{x}_i$. The statistic T_k is the maximum value of equation (4.3.2). Now, we continue with the permutation step to calculate the p-value (p), to test the null hypothesis that the process is IC. Assume L random permutations of the pooled sample. Let e $\tilde{T}_{kl}$ be the value of the k^{th} statistic obtained from the l^{th} permutation, where $k = 0, ..., K$ and $l = 1, ..., L$. Then, the overall control statistics are:

$$W = \max_{k=0,...,K} \frac{T_k - u_k}{\nu_k} \tag{4.3.3}$$

and

$$\tilde{W}_l = \max_{k=0,\dots,K} \frac{\tilde{T}_{kl} - u_k}{\nu_k} \tag{4.3.4}$$

where $u_k = \frac{1}{L}\sum_{l=1}^{L}\tilde{T}_{kl}$ and $\nu_k = \frac{1}{L-1}\sum_{l=1}^{L}\left(\tilde{T}_{kl} - u_k\right)^2$. The empirical p-value that indicates if the process is IC or OOC is $p = \frac{1}{L}\mathbb{1}_{\tilde{W}_l \geq W}$, where $\mathbb{1}$ is the indicator function. We raise an alarm if $p < \alpha$, with α to be the probability that controls a false alarm metric. The RS/P procedure is slightly modified for detecting scale changes. Now, T_0 is defined as:

$$T_0 = \max_{i=1,\dots,m} |s_i^2 - s^2| \tag{4.3.5}$$

where $s_i^2 = \frac{1}{n}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2$ and $s^2 = \frac{1}{m}\sum_{i=1}^{m}s_i^2$. Furthermore, we maximize the quantity:

$$\sum_{i=1}^{k+1}(\hat{\tau}_i - \hat{\tau}_{i-1})\, log\left(\frac{s^2}{s^2\left(\hat{\tau}_{i-1}, \hat{\tau}_i\right)}\right) \tag{4.3.6}$$

where $s^2(a,b) = \frac{1}{b-a}\sum_{i=a+1}^{b}s_i^2$, replacing the equation (4.3.2).

4.3.1.2 Simulation results for U3S

In this Subsection, our goal is to evaluate the performance of U3S and compare it against the SSC and RS/P alternative methods. Precisely, we examine the effective detection and the speed of reaction for step changes of the mean or inflations of the variance. Both of the IC and the OOC data are from a Normal distribution, assuming all parameters unknown. For IC data, we simulate 10,000 sequences size of N=50 observations from a typical Normal distribution, used to obtain the decision limits of each of the competing methods. Concerning the decision limits of U3S, we employ both the constant limit p^* and the adaptive p_n^* (noted by superscript n in the upcoming tables and graphs), so that one can compare their performance. In

addition, we start the testing from the third observation, as SSC needs the first two observations to estimate the unknown parameters and also we need to sacrifice the same number of data points in order to have a proper marginal distribution for U3S, when the reference prior for the IC parameters is in use. Thus, we perform $N-2$ hypothesis tests for each method and the decision limits are elicited such that all the competing methods to have $PFA(48) = 5\%$ for an IC sample. For the OOC scenarios, we introduce change points expressing contaminated data with small or medium persistent shifts. The change points are introduced at one of the locations, $\tau = \{11, 26 \text{ or } 41\}$, i.e. near the start, the middle or near the end of each OOC data sequence. The shifts are step changes of size 1 or 1.5 standard deviations for the mean, i.e. the OOC states are $N(1,1)$ or $N(1.5,1)$ and variance inflations of size 50% and 100% for the standard deviation, i.e. the OOC states are $N(0,1.5^2)$ or $N(0,2^2)$. Regarding the Monte Carlo efficiency, we also generate $L = 10,000$ random permutations to calculate the p-value of the test statistic of RS/P and we generate the same number of observations to estimate the marginals for $BF_{\tau,n+}$ and consequently to estimate the $p(\tau \leq n|\boldsymbol{x_n})$ for each data point in U3S decision limit derivation. In addition, SSC is optimally tuned in detecting the predetermined shifts. For the mean, we set a two sided SSC, for a positive and a negative shift, while for the variance we set it only for inflation detection, as this scenario is of major importance in practice.

For the priors used in U3S, we have a standard prior setting assuming (almost) total prior ignorance for $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and τ. Precisely, in the standard prior setting, we assume the non informative reference prior for the IC parameters, i.e. $\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2 \equiv NIG(0,0,-1/2,0)$. For the change point we assume $\tau \sim DW(1/N,1) \equiv G(1/N)$, i.e. an non informative prior about the location of the first OOC observation (constant hazard function), while the probability of moving to the OOC state for each data point is the reciprocal of the sample size. In order to have comparable assumptions with SSC, which is optimally tuned for the existed shifts, we assume

informative priors for the OOC parameter $\boldsymbol{\phi}$. Precisely, for the mean step changes of size 1 standard deviation, we assume $\delta|\gamma \sim \gamma \cdot N(1, 0.25^2) + (1-\gamma) \cdot N(-1, 0.25^2)$ and $\gamma \sim Ber(1/2)$, while for 1.5 standard deviations jumps the prior is $\delta|\gamma \sim \gamma \cdot N(1.5, 0.25^2) + (1-\gamma) \cdot N(-1.5, 0.25^2)$. This mixture of priors coincides with a two sided detection scheme for U3S, detecting for positive or negative shifts. For the variance inflation and specifically for the 50% standard deviation increase, we assume $\kappa \sim IG(50, 112.5)$, while $\kappa \sim IG(50, 200)$ is employed for the scenario of the 100% standard deviation increase. In addition, we apply a sensitivity analysis changing only one prior at a time, in order to perceive its effect. This reference prior setting corresponds to *r,c* indicator in the upcoming tables and graphs. The indicator *wi,c* corresponds to the weakly informative prior for $\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$, while the rest prior setting is the same with the standard. Regarding the change point, we employ a prior with an increasing hazard function (indicator *r,i*), where $\tau \sim DW(0.001, 2)$ and with a decreasing hazard function (indicator *r,d*), where $\tau \sim DW(0.1, 0.9)$. Further, we complete the sensitivity analysis, employing a more vague prior for the shift (indicator *vs*). We select weakly informative priors by increasing (doubling) the standard deviation of the standard setting priors. Thus, the priors for the mean are $\delta|\gamma \sim \gamma \cdot N(1, 0.5^2) + (1-\gamma) \cdot N(-1, 0.5^2)$ and $\delta|\gamma \sim \gamma \cdot N(1.5, 0.5^2) + (1-\gamma) \cdot N(-1.5, 0.5^2)$, while for the variance $\kappa \sim IG(14, 31.5)$ and $\kappa \sim IG(14, 52)$ respectively. Table 4.3.1 synopsizes all the different prior setting used, where $\theta_1^{OOC} = \{1 \text{ or } 1.5\}$ and $\theta_2^{OOC} = \{1.5 \text{ or } 2\}$ the true OOC values for the mean and the standard deviation respectively,

Setting	Priors
r, c	$\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ $\delta\|\gamma \sim \sum_{i=0}^{1} \gamma^{1-i}(1-\gamma)^i N((-1)^i \cdot \theta_1^{OOC}, 1/4^2)$ (mean model) $\gamma \sim Ber(1/2)$ (mean model) $\kappa \sim IG(50, 50 \cdot (\theta_2^{OOC})^2)$ (variance model) $\tau \sim G(1/N)$
wi, c	$\boldsymbol{\theta} \sim NIG(0, 5, 2.5, 2)$ $\delta\|\gamma \sim \sum_{i=0}^{1} \gamma^{1-i}(1-\gamma)^i N((-1)^i \cdot \theta_1^{OOC}, 1/4^2)$ (mean model) $\gamma \sim Ber(1/2)$ (mean model) $\kappa \sim IG(50, 50 \cdot (\theta_2^{OOC})^2)$ (variance model) $\tau \sim G(1/N)$
r, i	$\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ $\delta\|\gamma \sim \sum_{i=0}^{1} \gamma^{1-i}(1-\gamma)^i N((-1)^i \cdot \theta_1^{OOC}, 1/4^2)$ (mean model) $\gamma \sim Ber(1/2)$ (mean model) $\kappa \sim IG(50, 50 \cdot (\theta_2^{OOC})^2)$ (variance model) $\tau \sim DW(10^{-3}, 2)$
r, d	$\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ $\delta\|\gamma \sim \sum_{i=0}^{1} \gamma^{1-i}(1-\gamma)^i N((-1)^i \cdot \theta_1^{OOC}, 1/4^2)$ (mean model) $\gamma \sim Ber(1/2)$ (mean model) $\kappa \sim IG(50, 50 \cdot (\theta_2^{OOC})^2)$ (variance model) $\tau \sim DW(0.1, 0.9)$
vs	$\pi(\boldsymbol{\theta}) \propto 1/\theta_2^2$ $\delta\|\gamma \sim \sum_{i=0}^{1} \gamma^{1-i}(1-\gamma)^i N((-1)^i \cdot \theta_1^{OOC}, 1/2^2)$ (mean model) $\gamma \sim Ber(1/2)$ (mean model) $\kappa \sim IG(14, 14 \cdot (\theta_2^{OOC})^2)$ (variance model) $\tau \sim G(1/N)$

Table 4.3.1: The prior settings of U3S for the simulation study.

Concerning the performance, we estimate the metrics introduced in the Subsection 3.3.2 and precisely the $PSD(\tau)$, along with the mean and the sd of the $tCED(\tau)$ for all the OOC scenarios. Furthermore, in the graphs we provide the running $PFA(n)$, $n = 3, ..., N$, in order to perceive the FA behavior of the competing methods. In Tables 4.3.2 and 4.3.3, we provide the results of the simulations, while their graphical representation is in Figures 4.3.1, 4.3.2, 4.3.3 and 4.3.4. As demonstrated, the results, regarding the location and size of the shift, are corresponding to those in simulation study for PRC in Subsection 3.3.2. all methods have their best performance when the change point τ is at the middle of the sequence $\tau = 26$. In addition, they improve their performance for medium shifts ($1.5\theta_2$ mean step change and $+100\%$ sd (θ_2) inflation) compared with the small ($1\theta_2$ mean step change and $+50\%$ sd (θ_2) inflation).

Comparing the methods, U3S achieves greater performance compared with SSC and RS/P beyond the shadow of any doubt. More specifically, U3S achieves greater detection percentages than SSC, especially when the change point is at the start of the process. This is true as U3S is more resistant in absorbing a change, which is of vital importance for the self-starting methods. The $tCED(\tau)$ behavior is similar, apart from the cases when U3S has much higher detection percentages than the other methods. In these cases, the much greater $PSD(\tau)$ denotes that has larger "window of opportunity" to detect a change. Thus, it has the chance to react later in a change and this inflates the $tCED(\tau)$. It is worth mentioning that RS/P does not seem to be competitive at all for the variance inflations, as it has much worse performance than the other methods. This is the price the nonparametric approach pays in being general enough, as the other methods are flexible in aid of the detection power, as they are tuned for one-sided shifts.

Regarding the prior setting, it is clear that the performance using a weakly informative prior for $\boldsymbol{\theta}$ (prior setting wi, c) is significantly better for all the OOC scenarios, but mostly the when change points occurs at the start, i.e. the IC information from

the data is low. The results also denote the role that plays the hazard function in the detection power. Using a decreasing hazard function (prior setting r, i), the detection power is better at the early stages and vice versa, when using a increasing hazard function (r, d), the detection power improves at later stages. Regarding the vague priors for the shifts (prior setting vs), we observe that the detection percentages are slightly decreased, which was expected, but U3S is still robust. This is very important, as we considerably relax the assumption of known jumps.

Further, comparing the constant and the adaptive decision limit for U3S we can see that, as the adaptive (superscript n) is more conservative at the later stages, then it has reduced performance. But, when the change point is at the start, then it has the similar detection percentages and considerably less delay for an alarm, than using the constant one. Furthermore, via Figures 4.3.1, 4.3.2, 4.3.3 and 4.3.4, we realize that using p_n^*, then the FAs are more spread in the range of a sample, while using p^*, they are more concentrated at the end.

âĂČ

Mean

step change	τ	SSC	RS/P	$U3S^n_{r,c}$	$U3S_{r,c}$	$U3S^n_{wi,c}$	$U3S_{wi,c}$
		$PSD(\tau)\%$	$PSD(\tau)\%$	$PSD(\tau)\%$	$PSD(\tau)\%$	$PSD(\tau)\%$	$PSD(\tau)\%$
		$tCED(\tau)$	$tCED(\tau)$	$tCED(\tau)$	$tCED(\tau)$	$tCED(\tau)$	$tCED(\tau)$
		$(sd(tCED(\tau)))$	$(sd(tCED(\tau)))$	$(sd(tCED(\tau)))$	$(sd(tCED(\tau)))$	$(sd(tCED(\tau)))$	$(sd(tCED(\tau)))$
$+1\theta_2$	11	31.320%	43.190%	56.810%	57.740%	83.340%	84.050%
		13.837	14.929	19.404	22.607	15.950	18.787
		(8.737)	(9.706)	(9.049)	(8.502)	(8.717)	(8.432)
	26	61.790%	60.500%	77.360%	80.400%	85.600%	88.430%
		11.391	12.535	12.454	12.942	11.604	11.844
		(5.640)	(6.004)	(5.617)	(5.485)	(5.466)	(5.332)
	41	38.060%	29.240%	38.290%	43.910%	44.080%	50.520%
		6.873	6.996	7.157	7.078	7.038	6.880
		(2.217)	(2.105)	(2.150)	(2.146)	(2.150)	(2.185)
$+1.5\theta_2$	11	38.130%	82.860%	88.110%	89.240%	98.830%	99.040%
		8.632	12.254	12.786	14.919	8.633	10.157
		(6.658)	(8.389)	(7.800)	(7.808)	(5.720)	(5.786)
	26	81.230%	93.570%	95.260%	96.830%	97.460%	98.590%
		7.747	8.995	8.395	8.436	7.208	7.128
		(4.785)	(4.893)	(4.854)	(4.427)	(4.000)	(3.803)
	41	74.060%	72.620%	76.350%	80.560%	81.590%	85.630%
		5.516	6.341	6.088	5.850	5.749	5.470
		(2.250)	(1.983)	(2.154)	(2.146)	(2.154)	(2.150)

	mean step change	τ	$U3S^n_{r,i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{r,i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{r,d}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{r,d}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{vs}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{vs}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$
Mean	$+1\theta_2$	11	47.970%	52.060%	66.420%	67.240%	54.510%	56.620%
			24.030	28.434	22.827	25.062	18.522	21.705
			(9.179)	(7.418)	(8.834)	(8.073)	(9.544)	(9.067)
		26	80.420%	83.930%	77.590%	78.940%	73.990%	78.590%
			13.148	14.127	13.875	14.455	12.440	12.854
			(5.596)	(5.236)	(5.451)	(5.284)	(5.782)	(5.693)
		41	44.340%	52.200%	31.880%	34.710%	36.470%	43.080%
			7.162	7.087	7.416	7.428	7.058	6.927
			(2.123)	(2.124)	(2.125)	(2.123)	(2.196)	(2.205)
	$+1.5\theta_2$	11	86.180%	87.860%	92.480%	92.830%	87.720%	88.970%
			14.994	19.498	14.389	16.489	12.885	15.078
			(8.913)	(8.242)	(7.895)	(7.743)	(8.143)	(8.157)
		26	96.920%	98.530%	96.180%	97.190%	95.270%	97.020%
			8.198	8.697	9.209	9.571	8.538	8.583
			(4.484)	(4.321)	(4.596)	(4.545)	(4.712)	(4.579)
		41	81.100%	86.670%	72.070%	75.060%	75.420%	79.880%
			5.845	5.595	6.487	6.436	6.071	5.841
			(2.147)	(2.140)	(2.091)	(2.089)	(2.206)	(2.200)

Table 4.3.2: The percent probability of successful detection, the censored conditional expected delay and the corresponding standard deviation (in parenthesis) for change points occurred at $\tau = \{11, 26 \text{ or } 41\}$, of SSC and RS/P against U3S under the reference prior and a constant hazard function (indicator *r,c*), increasing (indicator *r,i*) or decreasing (indicator *r,d*). For the IC parameters, a weakly informative is denoted by indicator *wi,c*. The indicator *vs* corresponds to a vague prior for the parameter shift. Further, we use the adaptive decision limit (superscript n) or the constant. The results refer to step changes for the mean.

Variance

sd increase	τ	SSC $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	RS/P $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{r,c}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{r,c}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{wi,c}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{wi,c}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$
$+50\%\theta_2$	11	23.520%	9.580%	24.650%	23.480%	53.480%	49.260%
		16.636	15.189	22.111	28.814	18.613	25.555
		(10.577)	(10.624)	(9.863)	(7.564)	(10.522)	(8.650)
	26	40.470%	11.010%	48.760%	51.980%	59.870%	63.150%
		13.797	14.312	12.509	14.069	12.247	13.467
		(6.643)	(6.333)	(6.658)	(6.345)	(6.601)	(6.343)
	41	23.750%	4.370%	25.840%	30.650%	29.150%	34.470%
		5.976	6.822	6.024	6.063	6.024	6.062
		(2.701)	(2.475)	(2.670)	(2.649)	(2.680)	(2.674)
$+100\%\theta_2$	11	43.590%	27.200%	59.830%	57.810%	91.650%	91.960%
		11.192	14.471	13.019	19.323	9.609	12.774
		(9.512)	(9.930)	(9.742)	(9.945)	(8.079)	(8.688)
	26	76.740%	44.450%	82.820%	86.720%	90.820%	93.730%
		8.303	13.121	8.808	9.146	7.996	8.125
		(5.987)	(6.175)	(6.090)	(6.043)	(5.691)	(5.609)
	41	61.410%	22.600%	58.380%	64.580%	64.250%	70.100%
		4.935	6.687	5.251	5.159	5.153	4.993
		(2.662)	(2.179)	(2.680)	(2.675)	(2.676)	(2.656)

	sd increase	τ	$U3S^n_{r,i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{r,i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{r,d}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{r,d}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S^n_{vs}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$U3S_{vs}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$
Variance	$+50\%\theta_2$	11	20.060% 28.342 (8.930)	20.00% 34.325 (4.874)	30.420% 31.254 (6.953)	30.710% 33.651 (5.274)	25.230% 21.139 (10.210)	23.390% 28.206 (7.951)
		26	52.910% 14.297 (6.582)	53.010% 17.364 (5.415)	47.260% 15.627 (6.079)	46.900% 17.196 (5.419)	47.830% 12.382 (6.689)	50.890% 13.947 (6.420)
		41	31.950% 6.195 (2.646)	34.650% 6.536 (2.565)	24.410% 6.399 (2.654)	24.970% 6.644 (2.557)	25.220% 6.027 (2.680)	30.020% 6.059 (2.648)
	$+100\%\theta_2$	11	51.670% 15.793 (11.275)	50.730% 28.030 (8.556)	66.570% 20.192 (10.485)	66.820% 25.252 (8.722)	59.620% 13.062 (9.807)	57.480% 19.411 (10.002)
		26	86.510% 9.074 (6.182)	89.430% 10.401 (6.171)	84.060% 10.114 (6.287)	84.330% 11.092 (6.253)	83.090% 8.867 (6.116)	87.070% 9.252 (6.090)
		41	64.330% 5.200 (2.676)	70.580% 5.189 (2.662)	55.860% 5.458 (2.687)	57.400% 5.536 (2.684)	58.510% 5.253 (2.684)	64.860% 5.164 (2.677)

Table 4.3.3: The percent probability of successful detection, the censored conditional expected delay and the corresponding standard deviation (in parenthesis) for change points occurred at $\tau = \{11, 26 \text{ or } 41\}$, of SSC and RS/P against U3S under the reference prior and a constant hazard function (indicator *r,c*), increasing (indicator *r,i*) or decreasing (indicator *r,d*). For the IC parameters, a weakly informative is denoted by indicator *wi,c*. The indicator *vs* corresponds to a vague prior for the parameter shift. Further, we use the adaptive decision limit (superscript n) or the constant. The results refer to inflations of the standard deviation.

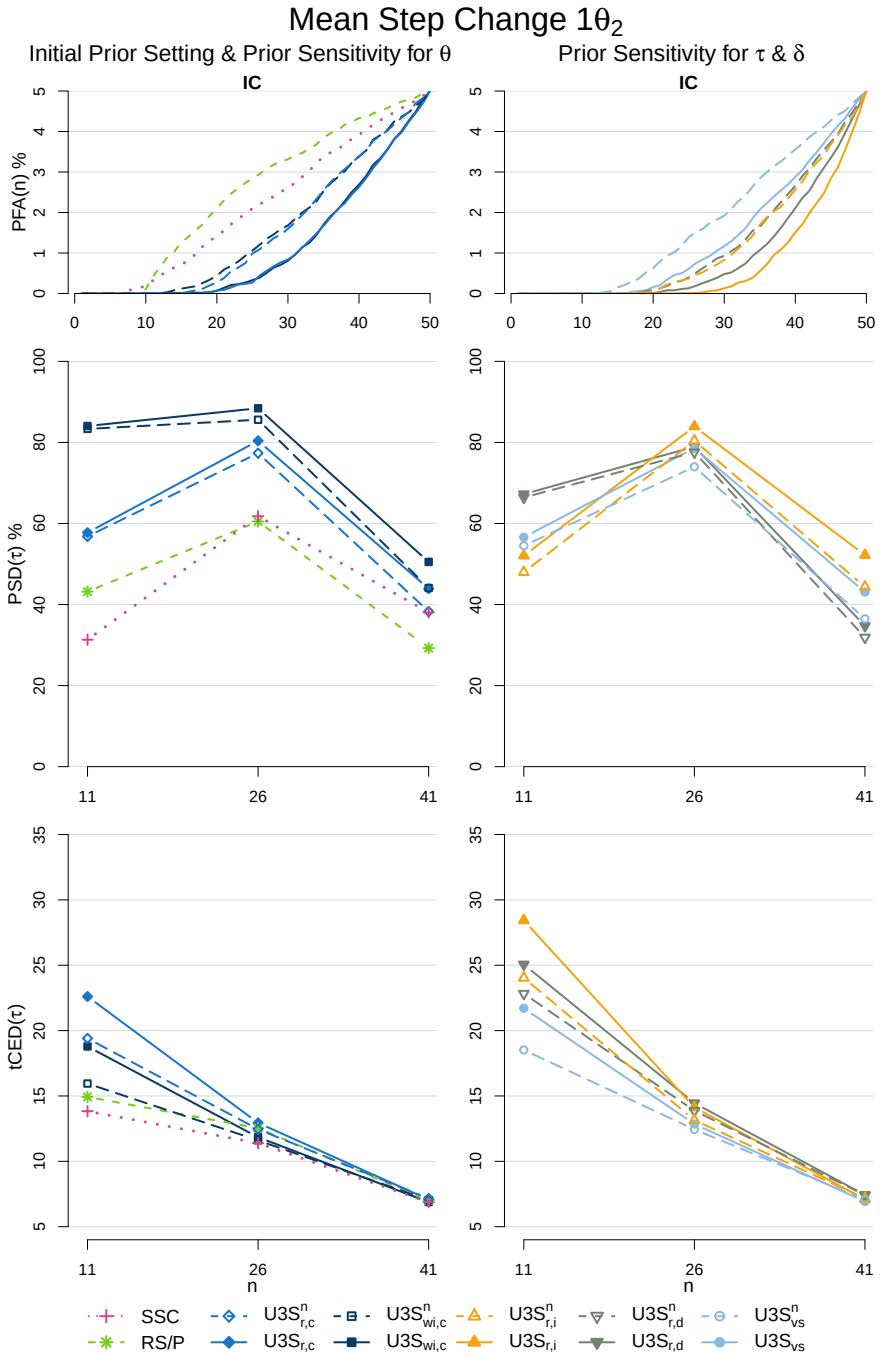


Figure 4.3.1: The $PFA(n)$ at each time point $n = 3, 4, \ldots, 50$ (top row), the $PSD(\tau)$ (middle row) and the $tCED(\tau)$ at $\tau = 11, 26$ or 41, of the U3S with all the prior settings against SSC and RS/P, when we have step changes for the mean size of 1 standard deviation.

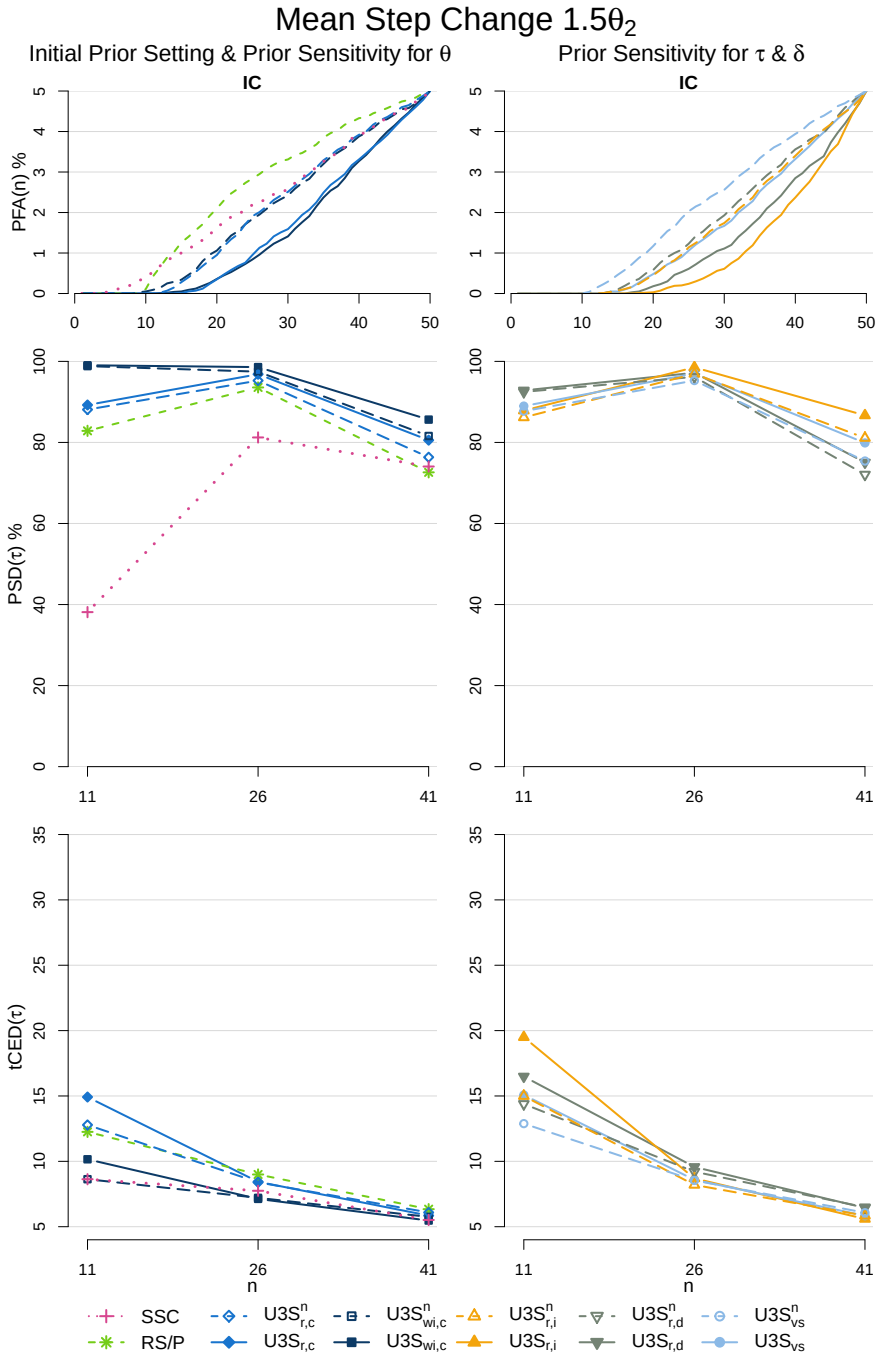


Figure 4.3.2: The $PFA(n)$ at each time point $n = 3, 4, \ldots, 50$ (top row), the $PSD(\tau)$ (middle row) and the $tCED(\tau)$ at $\tau = 11, 26$ or 41, of the U3S with all the prior settings against SSC and RS/P, when we have step changes for the mean size of 1.5 standard deviations.

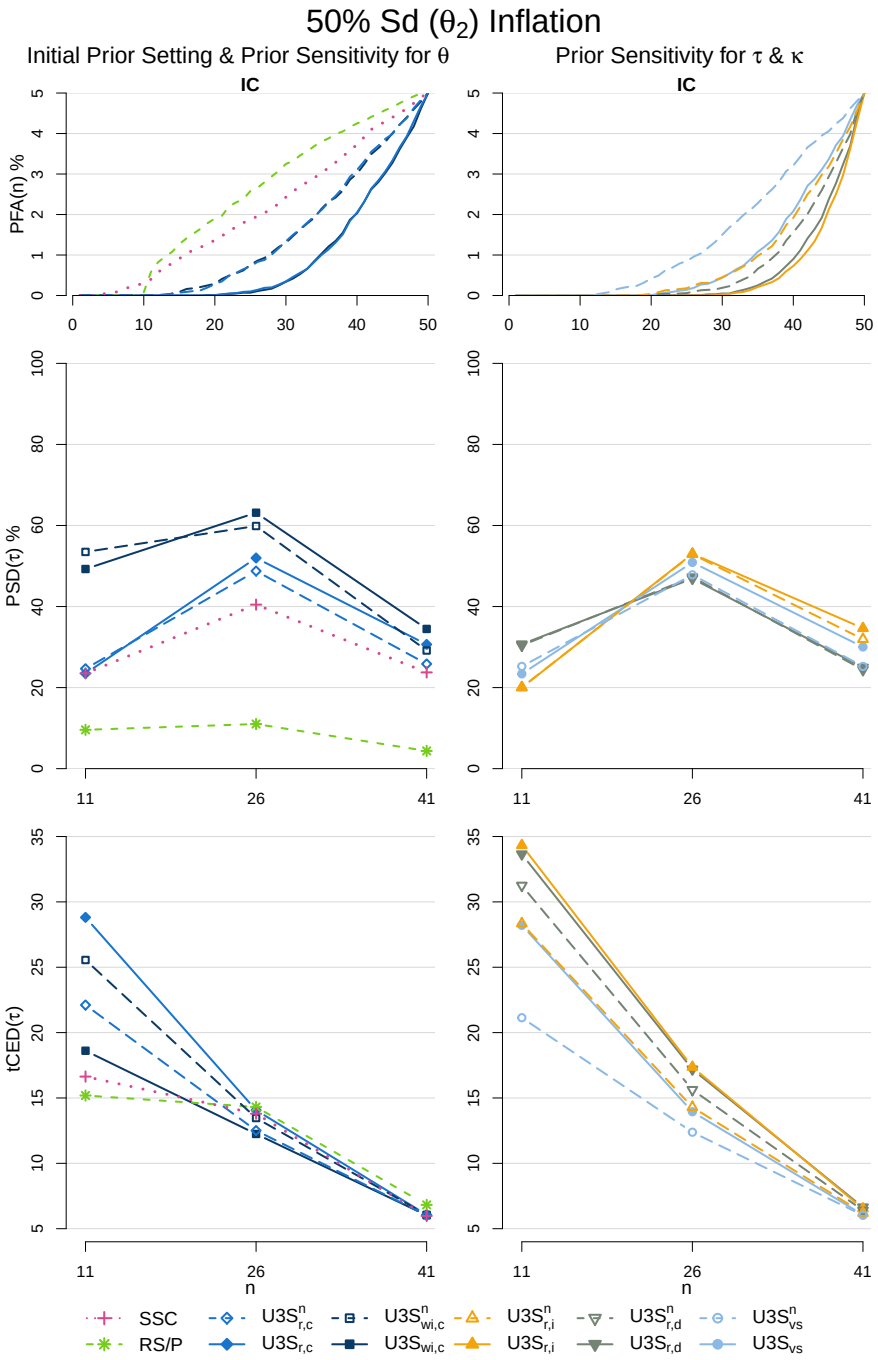


Figure 4.3.3: The $PFA(n)$ at each time point $n = 3, 4, \ldots, 50$ (top row), the $PSD(\tau)$ (middle row) and the $tCED(\tau)$ at $\tau = 11, 26$ or 41, of the U3S with all the prior settings against SSC and RS/P, when we have inflations for the standard deviation size of 50%.

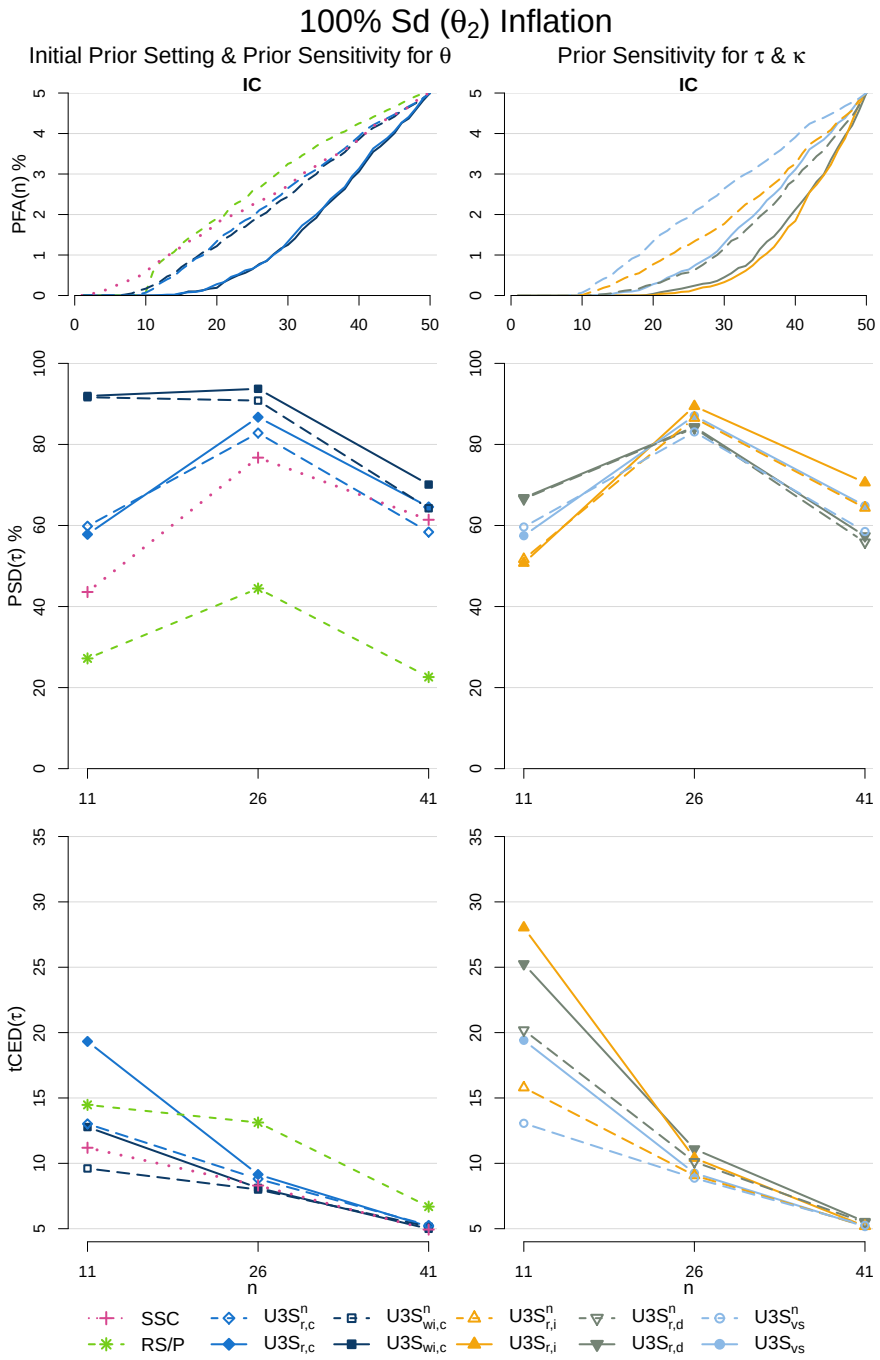


Figure 4.3.4: The $PFA(n)$ at each time point $n = 3, 4, \ldots, 50$ (top row), the $PSD(\tau)$ (middle row) and the $tCED(\tau)$ at $\tau = 11, 26$ or 41, of the U3S with all the prior settings against SSC and RS/P, when we have inflations for the standard deviation size of 100%.

4.3.2 Simulation study for M3S

4.3.2.1 Competing methods for M3S

In this Subsection, we will present the two methods that we will use in comparing the performance of the suggested M3S for several OOC scenarios, in sequences of multivariate Normal data. These are the Self-Starting Multivariate EWMA (SSMEWMA, Hawkins and Maboudou-Tchao, 2007) and the Self-Starting CUSCORE (SSCUSC$^{(1)}$, Capizzi and Masarotto, 2010) for drifts in the mean vector. Both competing methods are based on the recursive residuals, a regression methodology, introduced by Brown et al. (1975). Starting from $t = D + 2$, where D the dimensions of the data, then for $t > j + 1$, the sequence of recursive residuals for the t^{th} observation on the j^{th} variable for $x_{t,j}, ..., x_{t,j-1}$ are obtained by repeatedly using the regression coefficients estimated from vectors $\boldsymbol{x_1}, ..., \boldsymbol{x_{t-1}}$. Let $r_{t,j} = x_{t,j} - \hat{x}_{t,j}$ denoting the difference between the observed and the predicted value of the j^{th} variable, respectively. The standardized recursive residual will be:

$$e_{t,j} = \frac{r_{t,j}}{\sqrt{\sum_{i=j+1}^{t-1} r_{t,j}^2/(t-j-1)}} \tag{4.3.7}$$

Then a standard transformation is applied to $e_{t,j}$ to obtain independent standard normal statistics $Q_{t,j}$:

$$Q_{t,j} = \Phi^{-1}\left\{G_{t-j-1}\left(e_{t,j}\right)\right\} \tag{4.3.8}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard Normal CDF, $G_\nu(\cdot)$ the Student-t CDF with ν degrees of freedom. For SSMEWMA, the process statistic is given by:

$$Z_{t,j} = \lambda Q_{t,j} + (1-\lambda) Z_{t-1,j} \tag{4.3.9}$$

where $0 \leq \lambda 1$ is a constant that controls the depth of memory of the chart, $Z_{0,j} = 0$ and $j = 1, ..., D$. Defining, the constant $h > 0$ to achieve a predetermined IC false alarm tolerance, the SSMEWMA gives an out-of-control signal whenever:

$$\sum_{j=1}^{D} Z_{t,j}^2 > h \frac{\lambda \left[1 - (1-\lambda)^{2(t-p-1)}\right]}{2 - \lambda} \tag{4.3.10}$$

As concerns to SSCUSC$^{(1)}$, the $Q_{t,j}$ are used again, but this time through a CUSUM type formula. Starting from $t = D + 2$ and initializing the parameters $C_{D+1,j}^{L} = C_{D+1,j}^{U} = 0$ and $\tau_{D+1,j}^{L} = \tau_{D+1,j}^{U} = D + 2$, the downwards and upwards CUSCORE statistics will be respectively:

$$C_{t,j}^{L} = \min \left\{ 0, C_{t-1,j}^{L} + f_t \left(\tau_{t-1,j}^{L} \right) \cdot \left[Q_{t,j} + \frac{1}{2} f_t \left(\tau_{t-1,j}^{L} \right) \right] \right\} \tag{4.3.11}$$

$$C_{t,j}^{U} = \max \left\{ 0, C_{t-1,j}^{U} + f_t \left(\tau_{t-1,j}^{U} \right) \cdot \left[Q_{t,j} - \frac{1}{2} f_t \left(\tau_{t-1,j}^{U} \right) \right] \right\} \tag{4.3.12}$$

for $j = 1, ..., D$, where the involved parameters $\tau_{t,j}^{L}$, $\tau_{t,j}^{U}$ and the function $f_t(\tau)$ are respectively:

$$\tau_{t,j}^{L} = \begin{cases} t+1 & \text{if } C_{t,j}^{L} = 0 \\ \tau_{t-1,j}^{L} & \text{if } C_{t,j}^{L} < 0 \end{cases} \tag{4.3.13}$$

$$\tau_{t,j}^{U} = \begin{cases} t+1 & \text{if } C_{t,j}^{U} = 0 \\ \tau_{t-1,j}^{L} & \text{if } C_{t,j}^{U} > 0 \end{cases} \tag{4.3.14}$$

$$f_t(\tau) = m \cdot \max \left\{ c, \frac{\tau - 1}{\sqrt{t(t-1)}} \right\} \tag{4.3.15}$$

Regarding the design constants $m > 0$ and $0 < c < 1$, the former is related to a mean vector shift size of particular importance and the latter is introduced to prevent the CUSCORE statistics from trapping into a constant value. Defining a threshold h

with respect to the false alarm metric, an alarm is raised if:

$$SSCUSC_t^{(1)} = \sum_{j=1}^{D} \max\left\{-C_{t,j}^{L}, C_{t,j}^{U}\right\} > h \tag{4.3.16}$$

4.3.2.2 Simulation results for M3S

A simulation study, along with a sensitivity analysis, will compare the mean vector model M3S against the methods SSMEWMA and CUSCORE that we described in Subsection (4.3.2.1). Analogously to the univariate simulation study, we will examine here the performance in detecting permanent shifts. For IC data, we simulate 1,000 sequences size of N=50 observations each from a bivariate ($D = 2$) Normal distribution, needed in deriving the decision limits. As we wish to test the performance for different structures of the covariance matrix, we consider two scenarios of IC data; from a standard bivariate Normal (i.e. with zero mean vector and identity covariance matrix) and a bivariate Normal with mean zero vector and a covariance matrix with the i, j element to be $c_{i,j} = 0.6^{|i-j|}$. For M3S, we employ both decision thresholds p^* and p_n^* with the latter being denoted by superscript n in the upcoming tables and graphs. We use the first three observations to initiate the charts, i.e. to calculate the statistics of SSMEWMA and CUSCORE and to estimate the marginal of M3S. Thus, the decision limit for each method has been set so that we have $PFA(47) = 5\%$ for an IC sample, while for the OOC scenarios, we introduce contaminated data with a shift $\boldsymbol{\delta}$ to the mean vector, starting at one of the locations $\tau = \{11, 26 \text{ or } 41\}$ until the end of the sample. Specifically, the OOC shifts are:

- $\boldsymbol{\delta} = (+0.5 \ +0.5)^T$ or $\boldsymbol{\delta} = (+1 \ +1)^T$, while $\boldsymbol{\Sigma} = \boldsymbol{I}_2$
- $\boldsymbol{\delta} = (+0.5 \ +0.5)^T$ or $\boldsymbol{\delta} = (+1 \ +1)^T$ or $\boldsymbol{\delta} = (+0.5 \ -0.5)^T$ or $\boldsymbol{\delta} = (+1 \ -1)^T$, while $\boldsymbol{\Sigma}$ with the i, j element to be $c_{i,j} = 0.6^{|i-j|}$

This means that we have six OOC states with three different τ locations, i.e. 18 different OOC scenarios in total. Regarding the priors, we assume two settings for

M3S, an objective and an informative. For the objective setting (indicator n-i), we assume the Jeffreys' prior for the IC parameters, equal density for every possible shift direction and a constant hazard function for the change point. For to the informative setting (indicator i), we use the expected values of mean vector and the sum of pairwise deviation products of five bivariate IC imaginary data to set the prior parameters for $\boldsymbol{\theta}$. We choose the number of the IC imaginary data to be equal with the number of the unknown IC parameters. For the angle of the shift, we set the prior parameters so that over than 3/4 of the prior distribution to lie in the corresponding quadrant with the actual change, while for the change point we aim the hazard function to be approximately double at the end of the sample. The prior of the radius is suitably located for small or medium size shifts and in a similar manner we set $\lambda = \{0.05, 0.1\}$ for SSMEWMA and $m = \{0.25, 0.5\}$ for CUSCORE, in case of small or medium size shifts respectively. All the prior distributional assumptions are provided in Table 4.3.4, while Figures 4.3.5 roughly visualizes the settings for uncorrelated data.

Non-informative setting	**Informative setting**
$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \lvert\boldsymbol{\Sigma}\rvert^{-2}$	$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim NIW\left((0\ 0)^T, 5, 5, \boldsymbol{\Psi}_{i,j} = 4 \cdot 0.6^{\lvert i-j\rvert}\right)$
$r \sim NC_{\chi_D}(\lvert\lvert\boldsymbol{\delta}\rvert\rvert_2, 1/4)$	$r \sim NC_{\chi_D}(\lvert\lvert\boldsymbol{\delta}\rvert\rvert_2, 1/4)$
$\theta \sim U(0, 2\pi)$	$\boldsymbol{T}_\theta \sim vM(\pi/4 + \pi \cdot \mathbb{1}_{\delta_l \neq \delta_2}, 4)$
$\tau \sim G(1/N)$	$\tau \sim DW(1/N, 3/2)$

Table 4.3.4: The prior settings of M3S for the simulation study.

Regarding the performance metrics, similarly to the univariate case we estimate the Probability of Successful Detection, $PSD(\tau)$, the mean and the sd of the truncated Conditional Expected Delay, $tCED(\tau)$ for all the OOC scenarios and we plot the running Probability of False Alarm, $PFA(n)$, $n = 4, ..., N$. The results of the comparisons along with their graphical representations are provided in Table 4.3.5 and Figures 4.3.6, 4.3.7 and 4.3.8. As we can see, all the methods achieve the highest per-

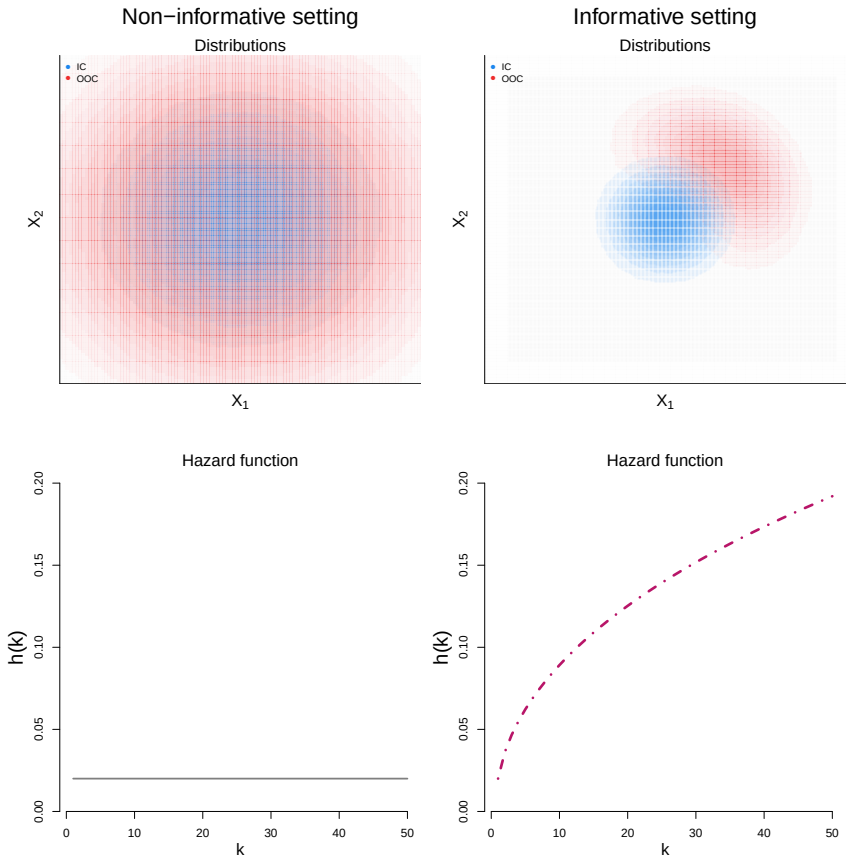


Figure 4.3.5: A graphical representation of the non-informative prior setting (left) and the informative prior setting (right) for uncorrelated data.

formance when the change point occurs at location 26, as they have both sufficient IC information available and a large enough "window of opportunity" to detect the change. Comparing the methods, M3S with the informative setting outperforms all the others, achieving significantly higher detection percentages and reacting faster to a change. This is the result of the beneficial use of the prior information. However, M3S is competitive even in the non-informative prior setup. Specifically, it has better performance than the competitors for small shifts in uncorrelated data, but it loses some detection power for medium shifts or correlated data. This is probably caused partly by the objective prior for the IC parameters and partly by the prior for radius r. Regarding the Jeffreys prior that we used, Sun and Berger (2007) noted that it seems to be quite bad for correlations, even if it achieves frequentist

matching for means and variances. A variety of other objective priors have been proposed for the bivariate case by Berger and Sun (2008). This is a challenging open problem (especially for high dimensions) that requires further investigation. For the prior of r, the location parameter d yields a benchmark for the size of the shift that we wish to detect. In many papers in the literature (e.g. Zantek, 2006 or Capizzi and Masarotto, 2010) it is suggested to avoid setting a self-starting for large shifts, especially when the IC history is not so long. Thus, a prior sensitivity regarding d could be employed. Finally, in comparing the constant and adaptive decision limits, the results are in par with the findings in the univariate case.

4.4 3S pplications to real data

4.4.1 U3S illustration

The application of U3S to real datasets is of interest in this Subsection. As demonstrated in Subsection 4.3.1.2, U3S has superior performance than the competitors in detecting shifts for the mean or the variance in short runs. We will apply the U3S methodology in two real data sets, where it appears that the first one experiences a mean step change and the second one a variance shift. The first dataset analyzed by Hawkins (1987) and it refers to a chemical laboratory that carries out routine indirect (instrumental) assays for precious metals of batches of a feedstock. As a control measure, a sample of a standard reference material is assayed along with each batch of unknowns. Due to confidentiality issues, the standardized data were presented in the paper of Hawkins (1987), subtracting the mean and dividing by the standard deviation. The standardized dataset is in Table 4.4.1.

	$\boldsymbol{\mu}^T$	τ	$SSMEWMAC$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	SSCUSC$^{(1)}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$M3S^n_{n-i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$M3S_{n-i}$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$M3S^n_i$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$M3S_i$ $PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$
$\rho = 0$	$(0.5, 0.5)$	11	13.9%	15.0%	18.2%	19.5%	77.9%	77.5%
			18.532	30.307	26.698	22.882	32.583	30.905
			(10.680)	(6.821)	(8.138)	(9.555)	(5.150)	(6.311)
		26	20.8%	26.5%	30.2%	27.6%	79.3%	79.2%
			15.591	17.826	16.096	14.931	18.845	17.931
			(6.232)	(5.191)	(5.597)	(5.978)	(4.016)	(4.460)
		41	4.7%	10.2%	11.1%	8.9%	31.1%	30.4%
			7.596	6.784	6.964	6.910	8.032	7.918
			(2.490)	(2.460)	(2.343)	(2.415)	(2.024)	(2.059)
	$(1, 1)$	11	34.7%	55.9%	68.9%	66.6%	99.7%	99.4%
			14.343	24.113	16.538	14.824	12.606	9.551
			(8.999)	(8.247)	(8.819)	(9.077)	(6.213)	(6.098)
		26	73.7%	91.8%	80.7%	76.4%	99.7%	98.9%
			13.168	14.192	10.869	10.975	7.497	6.970
			(5.571)	(5.137)	(5.641)	(5.591)	(3.509)	(3.596)
		41	51.8%	56.9%	45.0%	53.9%	89.2%	85.8%
			7.062	7.276	6.831	6.699	5.517	5.5661
			(2.142)	(2.065)	(2.228)	(2.280)	(2.197)	(2.237)
$\rho = 0.6$	$(0.5, -0.5)$	11	11.2%	10.8%	9.1%	7.5%	53.6%	51.9%
			18.234	30.852	25.154	20.533	35.386	34.497
			(10.483)	(6.905)	(9.403)	(11.413)	(3.904)	(4.652)
		26	13.4%	21.4%	12.0%	7.0%	56.5%	54.2%
			15.731	18.220	16.717	16.486	20.773	20.168
			(6.289)	(5.044)	(5.969)	(6.363)	(3.315)	(3.857)
		41	2.7%	7.6%	3.4%	1.2%	20.0%	17.3%
			7.148	7.211	6.618	5.000	8.020	7.780
			(2.699)	(2.340)	(2.374)	(1.414)	(2.187)	(2.357)

	$\boldsymbol{\mu}^T$	τ	$SSMEWMAC$	SSCUSC$^{(1)}$	$M3S^n_{n-i}$	$M3S_{n-i}$	$M3S^n_i$	$M3S_i$
			$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$	$PSD(\tau)\%$ $tCED(\tau)$ $(sd(tCED(\tau)))$
$\rho = 0.9$	$(1,1)$	11	27.6% 15.641 (9.498)	42.3% 25.858 (7.925)	37.1% 21.135 (9.144)	29.3% 18.997 (9.290)	97.8% 20.740 (7.428)	97.3% 17.731 (8.226)
		26	53.2% 15.160 (5.771)	72.0% 16.477 (5.552)	55.1% 13.744 (5.668)	45.0% 14.500 (5.714)	97.1% 11.456 (4.734)	96.5% 10.802 (4.967)
		41	30.5% 7.230 (2.166)	33.1% 7.263 (2.176)	22.2% 7.239 (2.215)	14.3% 7.476 (2.048)	66.5% 6.869 (2.112)	62.3% 6.907 (2.088)
	$(0.5,0.5)$	11	32.7% 18.104 (10.159)	39.6% 28.098 (7.128)	49.5% 20.141 (10.141)	41.8% 17.746 (10.516)	98.2% 21.554 (7.385)	98.1% 18.674 (8.297)
		26	55.0% 14.614 (5.927)	70.6% 16.194 (5.215)	67.5% 13.184 (5.549)	59.0% 13.590 (5.632)	99.3% 12.411 (4.506)	99.2% 11.550 (4.785)
		41	18.7% 7.465 (2.103)	26.9% 7.483 (2.154)	30.4% 7.454 (1.991)	19.7% 7.741 (2.058)	72.7% 7.003 (2.120)	70.3% 6.974 (2.178)
	$(1,-1)$	11	66.0% 11.397 (7.508)	92.9% 18.483 (7.308)	95.5% 8.356 (5.492)	93.9% 7.843 (5.593)	100% 4.957 (2.287)	99.9% 3.820 (2.003)
		26	96.8% 7.239 (3.621)	99.9% 7.979 (2.968)	94.3% 5.445 (3.041)	93.1% 5.793 (3.267)	99.4% 3.465 (1.520)	98.5% 3.396 (1.527)
		41	90.8% 5.752 (2.010)	93.2% 5.948 (1.947)	88.2% 4.706 (1.984)	85.3% 5.097 (2.003)	97.4% 3.098 (1.292)	96.7% 3.364 (1.425)

Table 4.3.5: The percent probability of successful detection, the truncated conditional expected delay and the corresponding standard deviation (in parenthesis) for change points occurred at $\tau = \{11, 26 \text{ or } 41\}$, of SSMEWMA and RS/P against M3S under the non-informative setting (indicator n-i) or the informative (indicator i). Furthermore, we use the adaptive decision limit (superscript n) or the constant. The results refer to step changes for the mean vector.

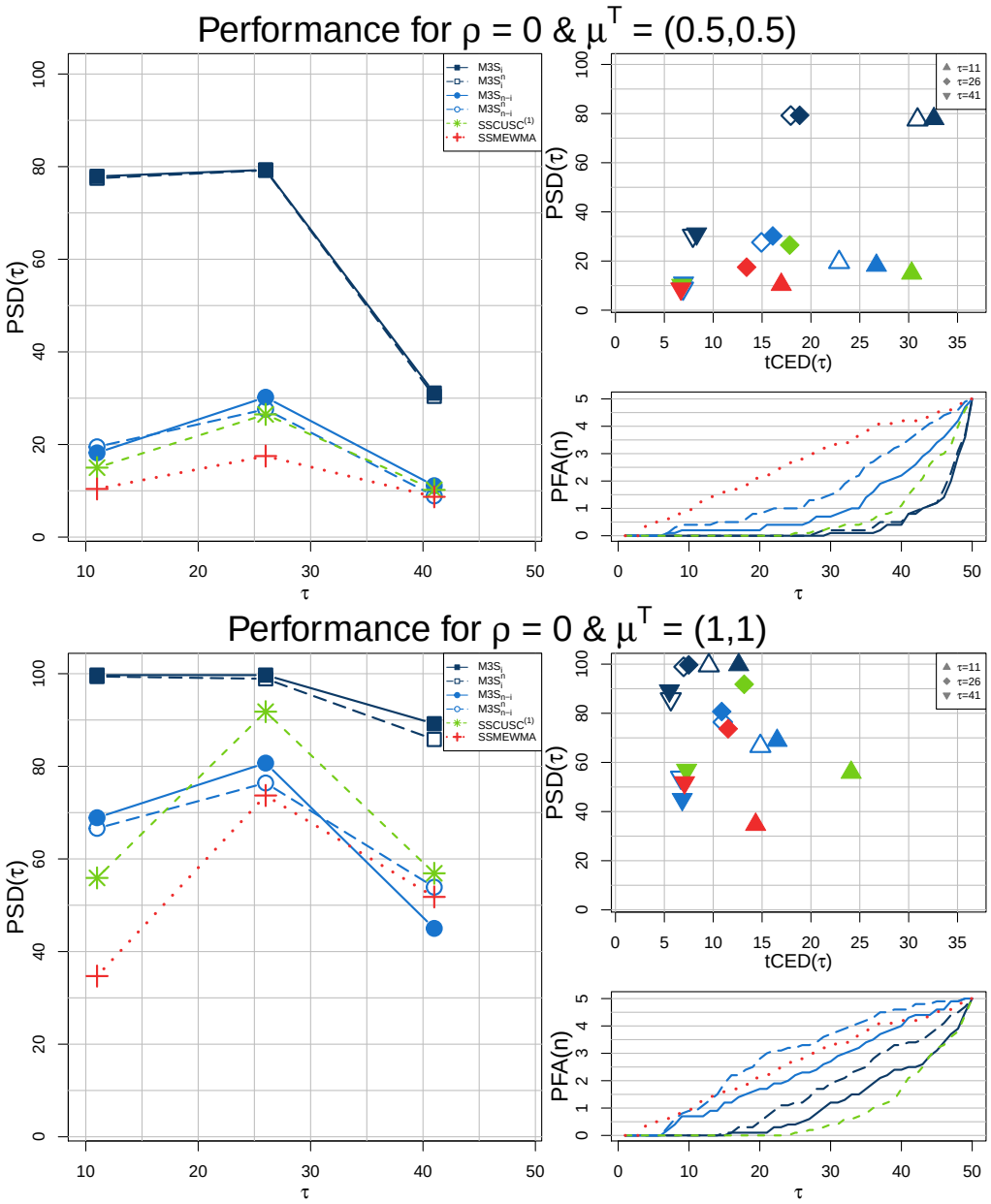


Figure 4.3.6: The $PSD(\tau)$ (left), the $tCED(\tau)$ vs the $PSD(\tau)$ (top right) for shift vectors $\boldsymbol{\delta}^T = (+0.5 \ +0.5)$ or $\boldsymbol{\delta}^T = (1 \ 1)$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_2$ starting at $\tau = \{11, 26 \text{ or } 41\}$ and the $PFA(n)$ at each time point $n = 4, 5, \ldots, 50$ (bottom right) of the M3S with all the prior settings against SSCUSC$^{(1)}$ and SSMEWMA, when we have step changes for the mean vector.

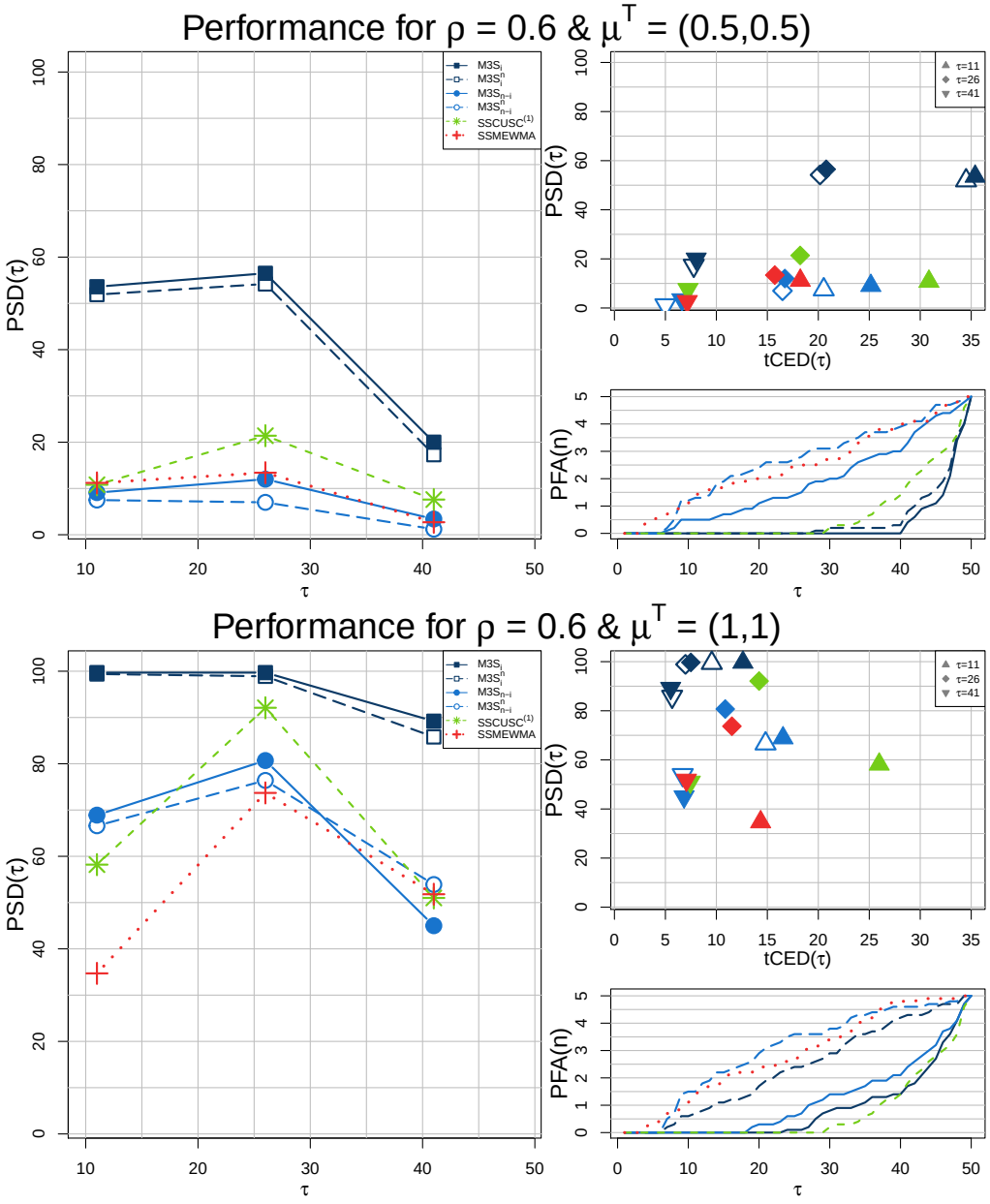


Figure 4.3.7: The $PSD(\tau)$ (left), the $tCED(\tau)$ vs the $PSD(\tau)$ (top right) for shift vectors $\boldsymbol{\delta}^T = (0.5\ 0.5)$ or $\boldsymbol{\delta}^T = (1\ 1)$ and $\boldsymbol{\Sigma}$ with the i, j element to be $c_{i,j} = 0.6^{|i-j|}$ starting at $\tau = \{11, 26 \text{ or } 41\}$ and the $PFA(n)$ at each time point $n = 4, 5, \ldots, 50$ (bottom right) of the M3S with all the prior settings against SSCUSC$^{(1)}$ and SSMEWMA, when we have step changes for the mean vector.

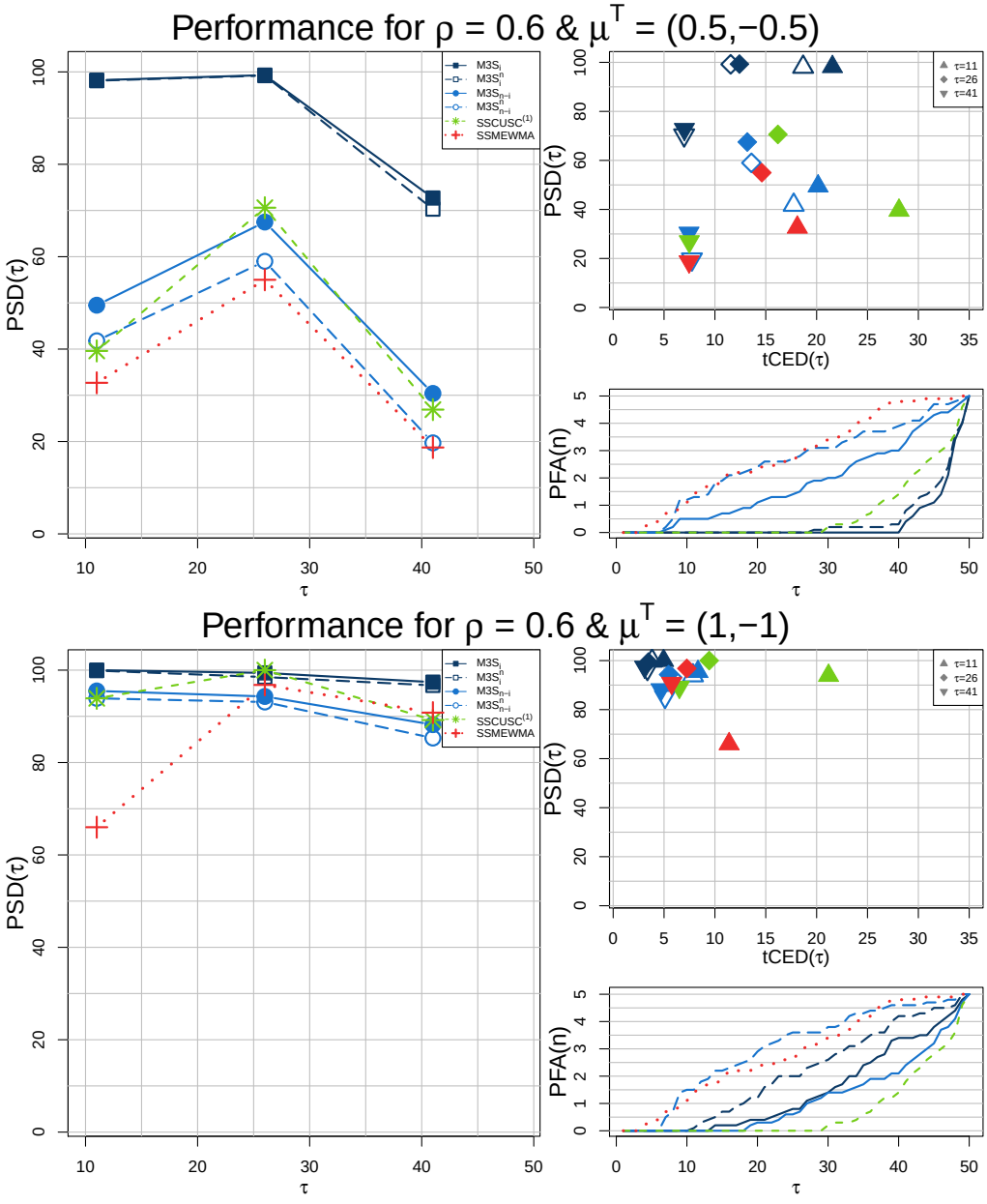


Figure 4.3.8: The $PSD(\tau)$ (left), the $tCED(\tau)$ vs the $PSD(\tau)$ (top right) for shift vectors $\boldsymbol{\delta}^T = (0.5\ -0.5)$ or $\boldsymbol{\delta}^T = (1\ -1)$ and $\boldsymbol{\Sigma}$ with the i, j element to be $c_{i,j} = 0.6^{|i-j|}$ starting at $\tau = \{11, 26 \text{ or } 41\}$ and the $PFA(n)$ at each time point $n = 4, 5, \ldots, 50$ (bottom right) of the M3S with all the prior settings against SSCUSC$^{(1)}$ and SSMEWMA, when we have step changes for the mean vector.

$x_1 - x_{11}$	0.82	0.40	-2.02	-0.02	-2.18	-0.64	-0.39	-0.51	1.17	0.49	-1.77
$x_{12} - x_{22}$	-0.64	-2.30	-1.55	-0.90	0.03	0.50	0.60	-0.65	0.19	-0.38	-0.72
$x_{23} - x_{33}$	-0.21	-0.50	0.95	1.59	0.68	-0.34	0.30	2.23	-0.75	1.39	1.01
$x_{34} - x_{44}$	-0.80	0.15	1.37	-1.39	0.86	0.64	-0.21	-0.51	-0.21	0.51	0.12
$x_{45} - x_{55}$	-0.33	1.010	-1.34	1.01	-0.04	1.67	1.26	-0.01	0.06	-0.82	0.12

Table 4.4.1: The sequence of the standardized data $\boldsymbol{x_n} = (x_1, x_2, \ldots, x_{55})$ from the first laboratory carrying out routine indirect (instrumental) assays for precious metals.

The prior distributions are the same with the standard prior setting for step changes of 1 standard deviation in Subection 4.3.1.2. In order to avoid being much more conservative compared to Hawkins (1987), who choose the decision limit of SSC to control $ARL_0 = 100$, we control the $PFA(55) = 20\%$, using the adaptive decision limit p_n^*. However, U3S would have similar reaction using the constant p^* as well. As we observe in the Figure 4.4.1, U3S realizes a step change and raises the first alarm at time $T = 33$, while the first alarm of SSC was at time $T = 30$. A plausible strategy is to stop the process after the first alarm, but we continue until the end of the sample in order to perceive its behavior in a permanent shift. Clearly, U3S resists in absorbing the change and it raises consecutive (14 in total) alarms, until the end of the data sequence. Apart from testing, U3S also provides posterior inference for the parameters of interest. The inference in Figure 4.4.1 is based on the whole sample, i.e. n = 55, but we can have online posterior inference as each data point becomes available. Analytical details for the sampling are provided in Appendix E. Regarding the posterior of the change point τ, the location 16 is the most probable for the first OOC observation, while there is a second mode at location 25. For the IC mean (pre change) and variance, the posterior means are $\hat{\theta}_1 = -0.69$ and $\hat{\theta}_2^2 = 0.91$ respectively, while the posterior mean for the size of the shift is $\hat{\delta} = 0.98$ times the standard deviation of the process. In other words, the point estimate for the post change mean is $\hat{\theta}_1 + \hat{\delta} \cdot \hat{\theta}_2 = 0.24$. Further, the lines which are plotted with the data (upper left panel) do not represent a control chart, but they provide visualization of the IC and the OOC state. Specifically, the dashed lines represent the pre and post change means, while the solid lines are positioned at $\pm 2 \cdot \hat{\theta}_2$ distance from each mean

respectively.

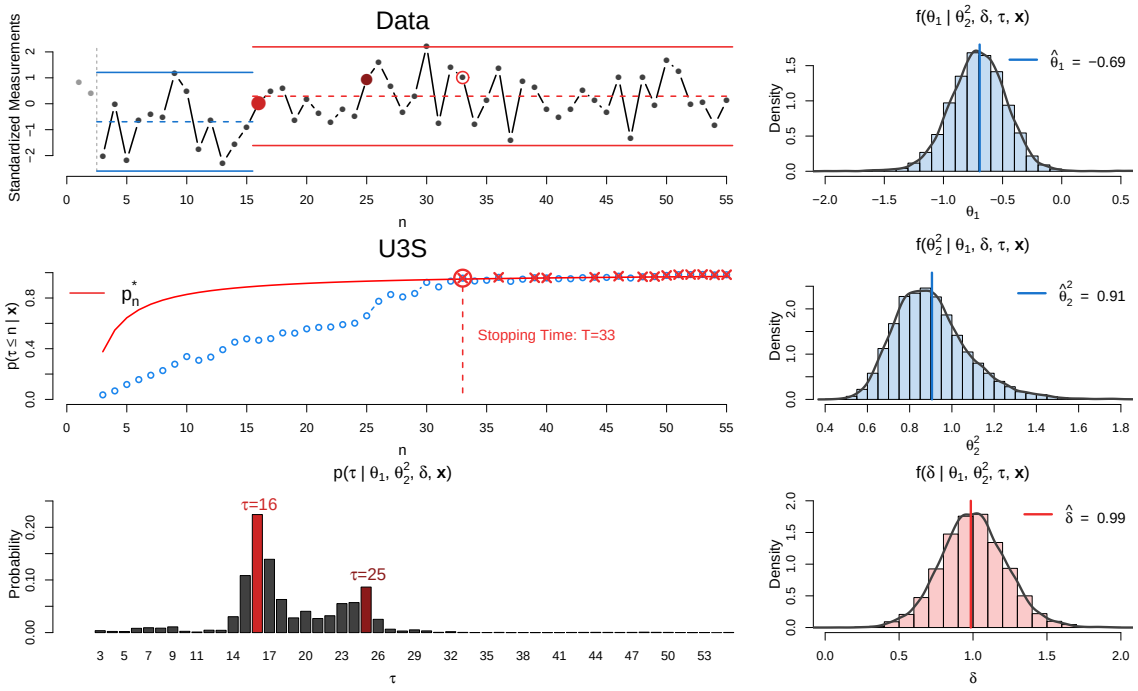


Figure 4.4.1: The U3S application to a mean step change. At the left panel (upper) we provide the data, the U3S process (middle) and the barplot of the full conditional for τ, while at the right panel we have the histograms for the IC parameters θ_1 and θ_2^2 and the size of the OOC shift δ. The adaptive decision p_n^* limit controls the $PFA(55) = 20\%$.

The second dataset for analysis comes from Villanueva-Guerra et al. (2017) and it refers to 60 monthly increments in the S&P 500, which is an American stock market index. Specifically, these are the first differences in S&P 500 values, or the amount that it grows or decays in a given month. The dataset is provided in Table 4.4.2.

$x_1 - x_{12}$	-6.36	31.40	13.19	77.65	92.64	-18.84	-79.67	-8.63	17.88	23.47	-40.65	70.46
$x_{13} - x_{24}$	9.79	77.32	78.20	-4.45	100.79	-70.48	-50.55	56.49	24.12	14.55	-7.03	-68.74
$x_{25} - x_{36}$	-46.67	-10.91	-15.81	-17.70	12.05	-54.75	-7.49	-53.71	-19.21	-13.81	18.73	18.91
$x_{37} - x_{48}$	-13.38	-20.15	39.12	-2.52	-10.35	-15.62	-43.63	-38.09	30.65	-22.33	23.01	23.74
$x_{49} - x_{60}$	-34.65	0.17	-42.85	13.85	-8.48	21.81	-42.47	1.19	-31.79	-0.58	-14.16	-15.78

Table 4.4.2: The monthly increments $\boldsymbol{x_n} = (x_1, x_2, \ldots, x_{60})$ in the S&P 500, reported during July 2004 - June 2009.

The prior distributions are the same with the standard prior setting in Section 4.3.1.2, apart from the prior of the size of the shift κ. Now, we are interested in a two-sided U3S, either for a inflation or a shrinkage of the variance. Thus, we will adopt a mixture $\kappa = \gamma \cdot IG(50, 200) + (1 - \gamma) \cdot IG(50, 12.5)$, where $\gamma \sim Ber(1/2)$. The components of the prior are centred in an increase of 100% or a decrease 50% for the sd. That is to say, we are more interested if the sd gets double or it becomes half. We use again the adaptive decision limit p_n^*, controlling the $PFA(60) = 10\%$. The graphical representation of the application is visually summarized in the Figure 4.4.2.

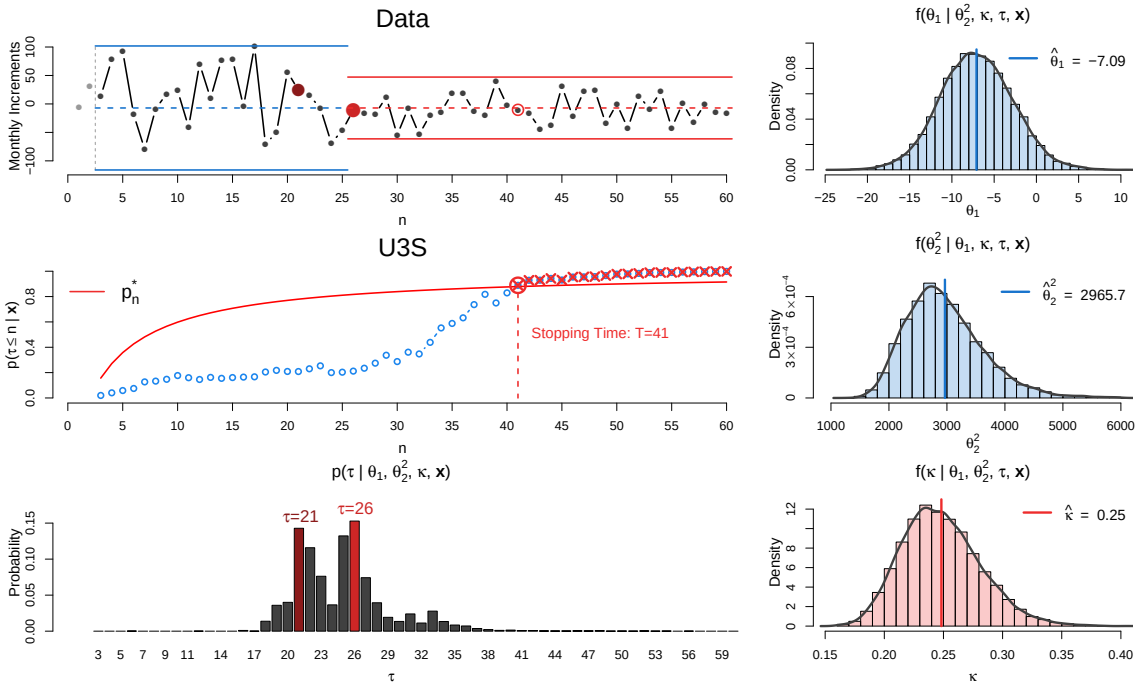


Figure 4.4.2: The U3S application to a variance level change. At the left panel (upper) we provide the data, the U3S process (middle) and the barplot of the full conditional for τ, while at the right panel we have the histograms for the IC parameters θ_1 and θ_2^2 and the size of the OOC shift κ. The adaptive decision p_n^* limit controls the $PFA(60) = 10\%$.

U3S realizes a level change for the variance and raises the first alarm at time $T =$

41. Similarly to the first application, we allow the process to run until the end of the sample and the posterior inference is based on the whole sample (Appendix E provides all the details regarding the MCMC sampling). U3S avoids again to absorb the change, raising 20 consecutive alarms. For the posterior inference of the change point, τ, the distribution is bimodal, with the modes being at locations 26 and 21. The posterior mean estimates for the IC mean and variance (pre change) of the process are $\hat{\theta}_1 = -7.09$ and $\hat{\theta}_2^2 = 2965.7$ respectively, while the $\hat{\kappa} = 0.25$, denoting that the post change variance is four times smaller, i.e. $\hat{\kappa} \cdot \hat{\theta}_2^2 = 736.24$. At the upper left panel, the dashed line denotes the mean of the process, while the solid lines are positioned from the mean distance of $\pm 2 \cdot \hat{\theta}_2$ pre change and $\pm 2 \cdot \hat{\kappa} \cdot \hat{\theta}_2$ post change, respectively. Regarding the analysis of Villanueva-Guerra et al. (2017), where the first 10 observations were used to initiate the monitoring, the chart reacted later to the change (under comparable false alarm tolerance). Additionally, the inference about the change point τ was misleading, as the point estimate was the location 30, where there is no any graphical evidence to start any change in the data.

4.4.2 M3S illustration

In this Subsection, we will illustrate the M3S for the mean vector changes to a real bivariate Normal dataset that appears to experience some type of disorder. The dataset is well documented and given by Holmes and Mergen (1983). It has been further analyzed via the self starting methods Self-Starting Multivariate EWMA (SS-MEWMA) by Sullivan and Jones (2002) and Self-Starting CUSCORE (SSCUSC$^{(1)}$) by Capizzi and Masarotto (2010). The dataset consists of $n = 56$ bivariate observation from a European plant producing gravel:

$$\boldsymbol{x} = \left(\begin{pmatrix} x_{1,1} \\ x_{2,1} \end{pmatrix}, \begin{pmatrix} x_{1,2} \\ x_{2,2} \end{pmatrix}, ..., \begin{pmatrix} x_{1,56} \\ x_{2,56} \end{pmatrix} \right)$$

where $x_{1,i}$ and $x_{2,i}$ represent the percentage of particles (by weight) of large and

medium sizes respectively. The dataset is provided in Table 4.4.3.

Large % $(x_{1,1} - x_{1,14})$	5.4	3.2	5.2	3.5	2.9	4.6	4.4	5.0	8.4	4.2	3.8	4.3	3.7	3.8
Medium % $(x_{2,1} - x_{2,14})$	93.6	92.6	91.7	86.9	90.4	92.1	91.5	90.3	85.1	89.7	92.5	91.8	91.7	90.3
Large % $(x_{1,15} - x_{1,28})$	2.6	2.7	7.9	6.6	4.0	2.5	3.8	2.8	2.9	3.3	7.2	7.3	7.0	6.0
Medium % $(x_{2,15} - x_{2,28})$	94.5	94.5	88.7	84.6	90.7	90.2	92.7	91.5	91.8	90.6	87.3	79.0	82.6	83.5
Large % $(x_{1,29} - x_{1,42})$	7.4	6.8	6.3	6.1	6.6	6.2	6.5	6.0	4.8	4.9	5.8	7.2	5.6	6.9
Medium % $(x_{2,29} - x_{2,42})$	83.6	84.8	87.1	87.2	87.3	84.8	87.4	86.8	88.8	89.8	86.9	83.8	89.2	84.5
Large % $(x_{1,43} - x_{1,56})$	7.4	8.9	10.9	8.2	6.7	5.9	8.7	6.4	8.4	9.6	5.1	5.0	5.0	5.9
Medium % $(x_{2,43} - x_{2,56})$	84.4	84.3	82.2	89.8	90.4	90.1	83.6	88.0	84.7	80.6	93.0	91.4	86.2	87.2

Table 4.4.3: Percentage of particles (by weight) of Large % $(x_{1,i})$ and Medium % $(x_{2,i})$ sizes respectively per time point $(i = 1, 2, \ldots, 56)$, in a European plant producing gravel.

The observations are tested sequentially, assuming $\boldsymbol{X}_i | (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For the unknown parameters, we have the non-informative prior setting, used in Subsection 4.3.2.2, defining the location parameter of radius $d = \sqrt{2}$. Further, as the prior is improper, we sacrifice the first three initial data points for the calculation of the marginal distribution, as described in Section 4.1.1. Regarding the decision limits, we use both of the constant p^* and the adaptive decision limit p_n^*, to control the $PFA(56) = 10\%$ for each limit. As we can see in the Figure 4.4.3, M3S with the adaptive decision limit is quite sensitive and raises an alarm from the first test, i.e. the fourth data point of the process (as the first three were used in the calibration). We could say it is a plausible alarm, as there is a discrepancy between the fourth data point and the initial three (in light blue) that involved in the marginal. We take it into account only as a warning message, letting the process continue. But in any case, this is quite important, as a big issue of the self-starting procedures is that they are incapable to react and detect a shift when it occurs in the very beginning of a process, because of the lack of IC history. A plausible stopping time would be $T = 28$, where we have the first of 29 consecutive alarms, until the end of the sample. Like U3S, M3S is also very persistent in raising alarms, when a disorder is detected. This is also a very important property, as it reduces the risk of absorption of a change. It is worth mentioning, that both of the SSMEWMA and the SSCUSC$^{(1)}$ reacted later and raised the first alarm at the 29^{th} and the 30^{th} observation respectively.

M3S is not merely superior in the early detection of the shift, but also provides analytical posterior inference regarding the unknown IC and OOC states. Based on the whole sample, the most probable location for the change point is 25, while it is quite probable a second change point to have been experienced around the location 43. In the scatterplot of Figure 4.4.3, the points 1-24 are in blue, while the rest are in red, noting stopping time $T = 28$. The posterior means for the IC parameters (pre change) are $\hat{\mu}_1 = 4.64$, $\hat{\mu}_2 = 89.16$, $\hat{\sigma}_1 = 2.08$, $\hat{\sigma}_2 = 6.23$ and $\hat{\rho} = -0.69$, while for the OOC parameters we have $\hat{r} = 2.19$ and $\hat{\theta} = 334.37^o$. Thus the point estimates for components of the OOC mean vector are $\hat{\mu}_1' = \hat{\mu}_1 + \hat{r} \cdot cos\hat{\theta} \cdot \hat{\sigma}_1 = 7.46$ and $\hat{\mu}_2' = \hat{\mu}_2 + \hat{r} \cdot cos\hat{\theta} \cdot \hat{\sigma}_2 = 86.23$. Appendix F provides the trace and the ACF plots for all the posteriors, along with the details of the MCMC sampling.

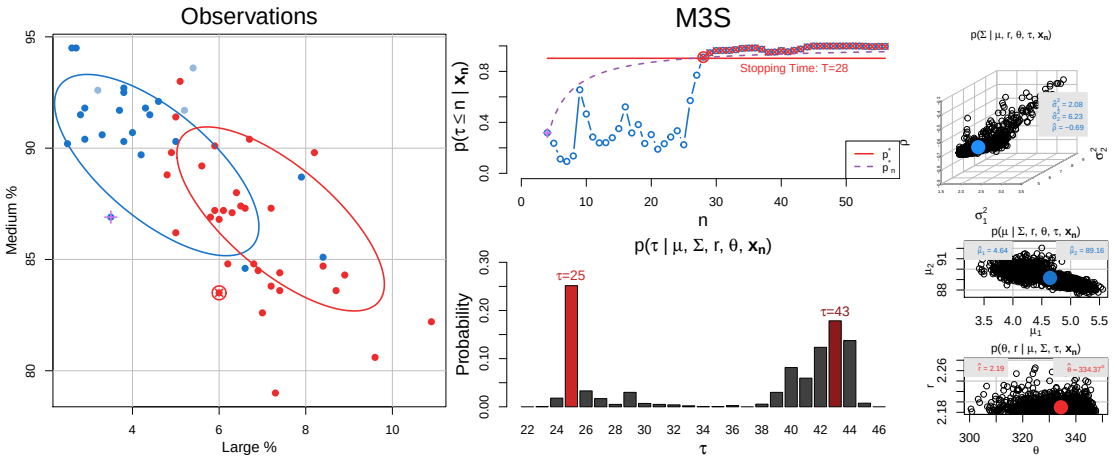


Figure 4.4.3: The M3S application to a mean vector step change change. At the left panel we provide the data, at the center panel the U3S process (upper) and the barplot of the full conditional for τ (lower), while at the right panel we provide the histograms for the IC parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ and the OOC parameters r and θ. The constant and the adaptive decision limit (p^* and p_n^*) control the $PFA(56) = 10\%$.

Chapter 5

Conclusions and Discussion

5.1 Conclusions

Statistical Process Control and Monitoring (SPC/M) is a widely used area of Statistics with a plethora of applications. The standard approach calls for a phase I calibration followed by a phase II testing phase, relying on strict assumptions and may under certain conditions be problematic. Namely, it may not be applicable in case where low volume data are available and the online analysis is of interest, or when there is no prior information about the process. Self-starting control charts aim to alleviate such issues. The Bayesian approach in SPC/M is underdeveloped, despite the advantages offered in handling the uncertainty of the unknown parameters utilizing prior knowledge about the process.

In this dissertation, we started with an extensive literature review on self-starting methods that we classified in distinct domains, based on certain criteria (like dimensionality, type of approach etc.). In this way we demarcate the area of Bayesian SPC/M, where the developed research methodology will be devoted. Specifically, the goal was the development of innovative Bayesian self-starting methods for individual observations in short runs. All things considered, we successfully dealt with

the detection of transient or permanent shifts, for location and/or scale parameters of either univariate or multivariate data, developing the appropriate methods. In addition to the successful detection of assignable causes of variation and the process monitoring, we also provided reliable inference about the unknown parameters, which is of major importance, when such methods are used in practice.

Namely, we expanded the work of Bourazas (2014) and Kiagias (2014) for Predictive Control Charts (PCC), which is a new general Bayesian method that permits online process monitoring for various types of univariate data, as long as their distribution belongs to the regular exponential family. We introduced the use of power priors, which, along with the initial priors, offer the flexibility to incorporate historical data and/or subjective knowledge in the decision making scheme allowing valid online inference, from the very early start of the process, aborting the need of a preliminary calibration phase. It is the use of prior distribution that provides a structural advantage over the non-parametric and self-starting frequentist based methods, especially in short runs and phase I data, where only brief IC information is available from the current data. The effect of the prior settings (as long as we avoid extremely informative priors), will decay soon, as more data become available. Furthermore, for users that might not be accustomed to the Bayesian approach, the choice of non-informative (reference or Jeffeys) prior, allows direct PCC implementation, using only the incoming data (and historical data if available). Also, we provided guidelines for the prior elicitation, and, generally, we developed the axiomatic framework, where the PCC process is applied, including the definition of HPrD/M region, a FIR scheme and a discussion on decision making.

PCC puts emphasis in online outlier detection of short production runs and it does not require a phase I/II split. Traditional phase I studies, where online inference regarding the presence of large transient shifts is of interest, are ideal settings for PCC. Furthermore, it is feasible for a user to switch from standard phase I/II monitoring methods to PCC, as it will not only provide online outlier detection monitoring dur-

ing the phase I segment, but thanks to its sequentially updated nature, it will allow incorporation of the phase II data into the monitoring mechanism (something that is not done with typical Frequentist methods). Thanks to the Bayesian posterior distribution, we are also able to perform inference regarding each of the unknown parameters. Lemma 2.3.1 and an extended simulation study, shows that PCC has better performance compared to the frequentist based alternative,Q Chart, achieving greater power in outlier detection and being robust for all the parameter values, even those that are too close to the boundary of their support set. Additionally, PCC is quite robust in model misspecifications, like violation in the independence or the distribution of the data.

PCC seems to be ideal for everyone that deals with either short runs or applications that require online monitoring during phase I. However, practitioners that employ a traditional phase I/II protocol in their routine, can benefit from the use of PCC during their phase I. Precisely, they will not only be able to monitor the process online while in phase I, but also obtain the posterior point estimates of the unknown parameters at the end of phase I, that will be necessary to build traditional phase II control charts. The benefits are significant in short runs, where most of the existing methods are unable to have robust performance and reliable estimates of the unknown parameter(s).

Next, we developed Predictive Ratio CUSUMs (PRC), a Bayesian change point model, able to accommodate any univariate data generating distribution that belongs to the regular exponential family, much like PCC. PRC is an enhanced Bayesian version of the frequentist Self-Starting CUSUM (SSC). In addition, PRC utilizes the fact that the alternative (competing) models (OOC in the SPC/M framework) are known, providing a method that boosts significantly the West's (1986) Cumulative Bayes Factors (CBF) approach. Most importantly though, PRC comes with detailed guidelines in deriving the decision thresholds (something missing to a large degree from both SSC and CBF) and a FIR scheme topics that are desperately needed from

a practitioner in employing such schemes in real life practice. An extensive simulation, evaluating the detection (both in power and alarm delay) of persistent shifts, shows that PRC outperforms SSC even when non-informative prior is used and it is also more powerful from CBF, except the special case where we look for shifts in the variance of a location scale family distribution, where CBF becomes a special case of PRC.

The PRC methodology was developed as a self starting quality monitoring scheme within the SPC/M area, but it can be used at any other field, where we are interested in online detection of persistent parameter shifts, especially when only low volume of data is available (short runs). Apart from the change point detection aspect of PRC (i.e. alarm a shift and provide an estimate of when this shift was originated), thanks to the Bayesian framework, at each time we can have a point/interval estimate of the unknown parameter, which will be sequentially updated. Finally, the detailed description of the methodology (in closed form) and the associated decision limits (typically absent in standard competing methods), based on the false alarm policy that one wishes to have, allows its straightforward implementation in either short (using $FWER$) or long (via ARL_0) sequences of data.

Next research work in this dissertation was the development of the Self-Starting Shiryaev (3S) methodology, a general detection scheme focusing on the efficient detection of permanent shifts in short runs, under the absence of phase I. Precisely, 3S is a family of innovative Bayesian online change point models under the At Most One Change (AMOC) scenario, which is general enough to be employed in any distribution, either continuous or discrete. It is a generalization of classical Shiryaev's approach in two main ways. First, by relaxing the strict assumption of known process parameters. This is true by allowing the parameters that describe the IC state or reflect an OOC shift to be all unknown. Secondly, by offering a more general prior (Discrete Weibull) for the change point. The selected prior is more flexible in the management of the hazard function for the occurrence of a change point, which is of

major importance in real life applications. We set the general 3S framework for any distributional setting, but we studied in detail the Univariate (U3S) and Multivariate (M3S) schemes for Normal data, providing the assumptions, the development, the evaluation and illustration to real data.

An extensive simulation study showed U3S to be more effective in detecting persistent shifts for the mean or the variance in short horizon Normal datasets compared to the standard frequentist or nonparametric alternatives, like SSC and Recursive Segmentation and Permutation (RS/P). Thanks to the prior distribution, U3S can utilize any available source of prior information in aid of the detection power, but it is still robust and effective under total prior ignorance. Concerning M3S, we define the models appropriately, in order to achieve the desired properties of the directional invariance, anisotropic scaling and rotation, which are beneficial for its applicability in real world problems. The simulation study, where we compare M3S for the mean vector against SSMEWMA and CUSCORE in the presence of step changes in the mean vector, showed that M3S outperforms the competing methods, in case of available prior information, while it is competitive under total prior ignorance.

The enhanced detection performance in both univariate and multivariate 3S schemes can be attributed to the proposed model structure, which splits appropriately the data and summarizes the evidence of a change point occurrence, using the whole sample. In this manner, it avoids including contaminated data in the IC estimates and consequently is more resistant in absorbing a shift. Aside from the superiority in the change point detection, U3S offers an online posterior inference for the unknown parameters, regarding the IC or the OOC scenario, including the change point. This is a great advantage for an online monitoring and a successful root cause analysis after a change point occurrence. Furthermore, we proposed a more realistic adaptive decision limit, which takes into account the prior of the change point and we provided guidelines for its successful use. Summarizing, 3S is an excellent choice for the online detection of a change point, especially in short runs, offering also posterior inference

for the unknown process parameters and the shift.

Despite the innovative methodology, which introduced and developed in this dissertation, there is still room for further improvement. Future research will be focused on the simulation study of M3S for the detection of rotations or scale shifts in the covariance matrix or the detection of mean vector drifts in higher dimensions. A further investigation will concern the non-informative prior setting in the multivariate processes along with some topics on the robustness. Finally, an `R` package with the existed methodology is planned to be developed, allowing the free and direct use of Bayesian SPC/M (BSPC/M) methods in the community.

Chapter 6

Appendices

Appendix A: Technical details regarding the derivation of the log ratio of the predictive OOC over IC models, $log(L_{n+1})$, for all PRC scenarios presented in Table 3.1.1.

A1: PRC for the rate of a Poisson likelihood.

Assume $X_i|\theta \sim P(\theta \cdot s_i)$, where s_i is the known number of events for the i^{th} observation, while for the rate (per event) unknown parameter we assume $\theta \sim G(c, d)$. Then, the resulting IC posterior is $\theta|\boldsymbol{\tau}_n \sim G\left(\hat{c}_n, \hat{d}_n\right)$, while the corresponding predictive is $f(X_{n+1}|\boldsymbol{X}_n) = NBin\left(\hat{c}_n, s_{n+1}/\left(\hat{d}_n + s_{n+1}\right)\right)$, where $\hat{c}_n = c + \sum_{j=1}^{N_D} w_j d_j$ and $\hat{d}_n = d + \sum_{j=1}^{N_D} w_j s_j$. Thus, the vector of IC posterior parameters, the predictive's sufficient statistic and $K(\boldsymbol{\tau_n})$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\hat{d}_n, \hat{c}_n - 1\right), \quad \boldsymbol{t}_f(X_{n+1}) = (s_{n+1}, x_{n+1}) \quad \text{and} \quad K(\boldsymbol{\tau_n}) = \frac{\Gamma(\tau_{n,1}+1)}{(\tau_{n,0})^{\tau_{n,1}+1}}$$

For the OOC scenario we introduce the shift to the unknown rate parameter θ by multiplying it by k (i.e. the OOC parameter is $k \cdot \theta$), which corresponds to a $(k-1)\cdot 100\%$ rate increase if $k > 1$, or to a $(1-k)\cdot 100\%$ decrease when $k < 1$. Since Gamma is

a scale family it follows that the OOC posterior will be $\theta|\boldsymbol{\tau}_n' \sim G\left(\hat{c}_n, \hat{d}_n/k\right)$, resulting the predictive $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = NBin\left(\hat{c}_n, s_{n+1}/\left(\hat{d}_n/k + s_{n+1}\right)\right)$. Therefore, the vector of intervened posterior parameters will be $\boldsymbol{\tau}_n' = \left(\hat{d}_n/k, \hat{c}_n - 1\right)$. Finally, the score function $log\left(L_{n+1}\right)$ will be given by

$$\begin{aligned} log\left(L_{n+1}\right) &= log\frac{f'\left(X_{n+1}|\boldsymbol{X}_n\right)}{f\left(X_{n+1}|\boldsymbol{X}_n\right)} = log\frac{K\left(\boldsymbol{\tau}_n' + \boldsymbol{t}_f(X_{n+1})\right)\cdot K\left(\boldsymbol{\tau}_n\right)}{K\left(\boldsymbol{\tau}_n + \boldsymbol{t}_f(X_{n+1})\right)\cdot K\left(\boldsymbol{\tau}_n'\right)} \\ &= log\frac{\dfrac{\Gamma\left(\hat{c}_n + x_{n+1}\right)}{\left(\hat{d}_n/k + s_{n+1}\right)^{\hat{c}_n + x_{n+1}}}\cdot\dfrac{\Gamma\left(\hat{c}_n\right)}{\hat{d}_n^{\hat{c}_n}}}{\dfrac{\Gamma\left(\hat{c}_n + x_{n+1}\right)}{\left(\hat{d}_n + s_{n+1}\right)^{\hat{c}_n + x_{n+1}}}\cdot\dfrac{\Gamma\left(\hat{c}_n\right)}{\left(\hat{d}_n/k\right)^{\hat{c}_n}}} \\ &= \left(\hat{c}_n + x_{n+1}\right) log\frac{\hat{d}_n + s_{n+1}}{\hat{d}_n/k + s_{n+1}} + x_{n+1}\cdot logk \end{aligned}$$

A2: PRC for the probability of success of a Binomial likelihood.

Let $X_i|\theta \sim Bin(N_i, \theta)$, where N_i is the known number of Bernoulli trials of the i^{th} observation and for the unknown success probability we assume $\theta \sim Beta\left(a, b\right)$. The IC posterior is $\theta|\boldsymbol{\tau}_n \sim Beta\left(\hat{a}_n, \hat{b}_n\right)$, while the predictive is $f\left(X_{n+1}|\boldsymbol{X}_n\right) = BetaBin\left(\hat{a}_n,\ \hat{b}_n,\ N_{n+1}\right)$, where $\hat{a}_n = a + \sum_{j=1}^{N_D} w_j d_j$ and $\hat{b}_n = b + \sum_{j=1}^{N_D} w_j N_j - \sum_{j=1}^{N_D} w_j d_j$. Thus, the vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\frac{\hat{a}_n + \hat{b}_n - 2}{\sum_{i=1}^{n} N_i}, \hat{a}_n - 1\right), \quad \boldsymbol{t}_f(x_{n+1}) = (N_{n+1}, x_{n+1}) \quad \text{and}$$

$$K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma(\tau_{n,1}+1)\Gamma(N_i\cdot\tau_{n,0} - \tau_{n,1} + 1)}{\Gamma\left(\sum_{i=1}^{n} N_i\cdot\tau_{n,0} + 2\right)}$$

For the OOC scenario we multiply the expected odds of θ by k, i.e. the OOC

shift is $k \cdot E\left(\frac{\theta}{1-\theta}\right)$. This shift corresponds to a $(k-1)\cdot 100\%$ expected odds increase if $k > 1$, or to a $(1-k)\cdot 100\%$ decrease when $k < 1$. The OOC posterior will be $\theta|\boldsymbol{\tau}'_n \sim Beta\left(k\cdot\hat{a}_n, \hat{b}_n\right)$ and the corresponding predictive $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = BetaBin\Big(k\cdot\hat{a}_n, \hat{b}_n, N_{n+1}\Big)$. Therefore, the vector of the intervened posterior parameters will be $\boldsymbol{\tau}'_n = \left(\frac{k\cdot\hat{a}_n+\hat{b}_n-2}{\sum_{i=1}^{n} N_i}, k\cdot\hat{a}_n - 1\right)$. The score function $log\left(L_{n+1}\right)$ will be

$$
\begin{aligned}
log\left(L_{n+1}\right) &= log\frac{\dfrac{\Gamma\left(k\cdot\hat{a}_n+x_{n+1}\right)\Gamma\left(\hat{b}_n+N_{n+1}-x_{n+1}\right)}{\Gamma\left(k\cdot\hat{a}_n+\hat{b}_n+N_{n+1}\right)}\cdot\dfrac{\Gamma\left(\hat{a}_n\right)\Gamma\left(\hat{b}_n\right)}{\Gamma\left(\hat{a}_n+\hat{b}_n\right)}}{\dfrac{\Gamma\left(\hat{a}_n+x_{n+1}\right)\Gamma\left(\hat{b}_n+N_{n+1}-x_{n+1}\right)}{\Gamma\left(\hat{a}_n+\hat{b}_n+N_{n+1}\right)}\cdot\dfrac{\Gamma\left(k\cdot\hat{a}_n\right)\Gamma\left(\hat{b}_n\right)}{\Gamma\left(k\cdot\hat{a}_n+\hat{b}_n\right)}} \\
&= log\frac{B\left(k\cdot\hat{a}_n+\hat{b}_n, N_{n+1}\right)\cdot B\left(\hat{a}_n, x_{n+1}\right)}{B\left(\hat{a}_n+\hat{b}_n, N_{n+1}\right)\cdot B\left(k\cdot\hat{a}_n, x_{n+1}\right)}
\end{aligned}
$$

A3: PRC for the probability of success of a Negative Binomial likelihood. Let $X_i|\theta \sim NBin(r,\theta)$, where r represents the known number of failures until the experiment stops and for the unknown probability of success we assume $\theta \sim Beta\left(c,d\right)$. The IC posterior and predictive will be $\theta|\boldsymbol{\tau}_n \sim Beta\left(\hat{a}_n, \hat{b}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = NBetaBin\Big(\hat{a}_n, \hat{b}_n, r\Big)$ respectively, where $\hat{a}_n = a + r\sum_{j=1}^{N_D} w_j$ and $\hat{b}_n = b + \sum_{j=1}^{N_D} w_j d_j$. Thus, the vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$
\boldsymbol{\tau}_n = \left(\frac{\hat{a}_n-1}{r}, \hat{b}_n - 1\right), \quad \boldsymbol{t}_f(x_{n+1}) = (1, x_{n+1}) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma(\tau_{n,1}+1)\Gamma(r\cdot\tau_{n,0}+1)}{\Gamma(r\cdot\tau_{n,0}+\tau_{n,1}+2)}
$$

As in the Binomial case, for the OOC scenario we multiply the expected odds of θ by k. This shift represents a $(k-1)\cdot 100\%$ expected odds increase if $k > 1$, or a

$(1-k)\cdot 100\%$ decrease when $k<1$. The OOC posterior is $\theta|\boldsymbol{\tau}_n' \sim Beta\left(k\cdot\hat{a}_n, \hat{b}_n\right)$ and the corresponding predictive $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = NBetaBin\Big(k\cdot\hat{a}_n, \hat{b}_n, r\Big)$. The intervened posterior parameters are $\boldsymbol{\tau}_n' = \left(\dfrac{k\cdot\hat{a}_n - 1}{r}, \hat{b}_n - 1\right)$. Finally, the score function $log\left(L_{n+1}\right)$ will be given by

$$log\left(L_{n+1}\right) \;=\; log\frac{B\left(k\cdot\hat{a}_n+\hat{b}_n, r+x_{n+1}\right)\cdot B\left(\hat{a}_n, r\right)}{B\left(\hat{a}_n+\hat{b}_n, r+x_{n+1}\right)\cdot B\left(k\cdot\hat{a}_n, r\right)}$$

A4: PRC for the mean of a Normal likelihood with known variance.

Let $X_i|\theta \sim N\left(\theta,\sigma^2\right)$, where σ^2 is the known variance, and for the unknown mean parameter we assume $\theta \sim N\left(\mu_0,\sigma_0^2\right)$. The IC posterior and predictive will be $\theta|\boldsymbol{\tau}_n \sim N\left(\hat{\mu}_n,\hat{\sigma}_n^2\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = N\left(\hat{\mu}_n,\hat{\sigma}_n^2+\sigma^2\right)$ respectively, where $\hat{\mu}_n = \left(\sigma^2\mu_0 + {\sigma_0}^2\sum_{j=1}^{N_D} w_j d_j\right) \Big/ \left(\sigma^2 + {\sigma_0}^2\sum_{j=1}^{N_D} w_j\right)$ and $\hat{\sigma}_n^2 = {\sigma_0}^2\sigma^2 \Big/ \left(\sigma^2 + {\sigma_0}^2\sum_{j=1}^{N_D} w_j\right)$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\frac{\sigma^2}{\hat{\sigma}_n^2}, \frac{\hat{\mu}_n}{\hat{\sigma}_n^2}\right),\;\; \boldsymbol{t}_1\left(x_{n+1}\right) = \left(1, \frac{x_{n+1}}{\sigma^2}\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \sqrt{\frac{2\pi\sigma^2}{\tau_{n,0}}} exp\left\{\frac{\sigma^2\tau_{n,1}^2}{2\tau_{n,0}}\right\}$$

For the OOC shift, we introduce a step change of size $k\cdot\sigma$ on the mean, i.e. the OOC mean is $\theta + k\cdot\sigma$ and the mean shift is upward or downward, depending on whether $k>0$ or $k<0$ respectively. Since Normal is a location family, the OOC posterior is $\theta|\boldsymbol{\tau}_n' \sim N\left(\hat{\mu}_n + k\cdot\sigma, \hat{\sigma}_n^2\right)$ and the corresponding OOC predictive will be $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = N\left(\hat{\mu}_n + k\cdot\sigma, \hat{\sigma}_n^2+\sigma^2\right)$. The vector of the intervened posterior parameters is $\boldsymbol{\tau}_n' = \left(\dfrac{\sigma^2}{\hat{\sigma}_n^2}, \dfrac{\hat{\mu}_n + k\cdot\sigma}{\hat{\sigma}_n^2}\right)$. If we will standardize the future observable, setting $Z_{n+1} = \left(X_{n+1}-\hat{\mu}_n\right)/\sqrt{\hat{\sigma}_n^2+\sigma^2}$, then the standardized predictives will be $f\left(Z_{n+1}|\boldsymbol{X}_n\right) = N\left(0,1\right)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right) = N\left(k\cdot\sigma/\sqrt{\hat{\sigma}_n^2+\sigma^2}, 1\right)$. The score func-

tion $log\left(L_{n+1}\right)$ will be given by

$$\begin{aligned} log\left(L_{n+1}\right) &= log\frac{f'\left(X_{n+1}|\boldsymbol{X}_n\right)}{f\left(X_{n+1}|\boldsymbol{X}_n\right)} = log\frac{f'\left(Z_{n+1}|\boldsymbol{X}_n\right)}{f\left(Z_{n+1}|\boldsymbol{X}_n\right)} \\ &= \left(z_{n+1} - \frac{k}{2}\cdot\frac{\sigma}{\sqrt{\hat{\sigma}_n^2+\sigma^2}}\right)\cdot\frac{\sigma}{\sqrt{\hat{\sigma}_n^2+\sigma^2}} \end{aligned}$$

A5: PRC for the variance of a Normal likelihood with known mean.
Let $X_i|\theta^2 \sim N\left(\mu,\theta^2\right)$, where μ is the known mean, and for the unknown variance parameter we assume $\theta^2 \sim IG\left(a,b\right)$. The IC posterior and predictive distributions will be $\theta^2|\boldsymbol{\tau}_n \sim IG\left(\hat{a}_n,\hat{b}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(\mu,\hat{b}_n/\hat{a}_n\right)$ respectively, where $\hat{a}_n = a + \sum_{j=1}^{N_D} w_j/2$ and $\hat{b}_n = b + \sum_{j=1}^{N_D} w_j\left(d_j-\mu\right)^2/2$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(2(\hat{a}_n+1), 2\hat{b}_n\right),\ \ \boldsymbol{t}_1\left(x_{n+1}\right) = \left(1,\left(x_{n+1}-\mu\right)^2\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma\left(\frac{\tau_{n,0}}{2}-1\right)}{\left(\frac{\tau_{n,1}}{2}\right)^{\frac{\tau_{n,0}}{2}-1}}$$

For the OOC shift, we multiply the variance by k, i.e. the OOC parameter is $k\cdot\theta^2$ and this shift corresponds to a $(k-1)\cdot 100\%$ variance increase if $k>1$ or a $(1-k)\cdot 100\%$ decrease if $k<1$. Since the Inverse Gamma is a scale family, the OOC posterior will be $\theta|\boldsymbol{\tau}'_n \sim IG\left(\hat{a}_n, k\cdot\hat{b}_n\right)$ with the corresponding predictive being $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(\mu, k\cdot\hat{b}_n/\hat{a}_n\right)$. Thus, the intervened parameters are given by $\boldsymbol{\tau}'_n = \left(2(\hat{a}_n+1), k\cdot 2\hat{b}_n\right)$. Standardizing the future observable we have $Z_{n+1} = \left(X_{n+1}-\hat{\mu}_n\right)/\sqrt{\hat{b}_n/\hat{a}_n}$, resulting the IC and OOC predictive distributions to be $f\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(0,1\right)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(0,k\right)$ respectively. Finally, the score function will be

$$log\left(L_{n+1}\right) = \left(\hat{a}_n+1/2\right)\cdot log\frac{2\hat{a}_n+z_{n+1}^2}{2\hat{a}_n+z_{n+1}^2/k} - log\sqrt{k}$$

A6: PRC for the mean of a Normal likelihood with both parameters unknown.

Let $X_i|\,(\theta_1,\theta_2^2) \sim N(\theta_1,\theta_2^2)$ with both parameters being unknown and assumed $(\theta_1,\theta_2^2) \sim NIG(\mu_0,\lambda,a,b)$. The IC posterior and predictive distributions will be given by

$(\theta_1,\theta_2^2)\,|\boldsymbol{\tau}_n \sim NIG\left(\hat{\mu}_n,\hat{\lambda}_n,\hat{a}_n,\hat{b}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(\hat{\mu}_n,(\lambda_n+1)\cdot\hat{b}_n/\left(\lambda_n\cdot\hat{a}_n\right)\right)$

respectively, where $\hat{\mu}_n = \left(\lambda\mu_0+\sum_{j=1}^{N_D}w_jd_j\right)\Big/\left(\lambda+\sum_{j=1}^{N_D}w_j\right)$, $\hat{\lambda}_n = \lambda+\sum_{j=1}^{N_D}w_j$, $\hat{a}_n = a+\sum_{j=1}^{N_D}w_j/2$ and $\hat{b}_n = b+\left(\lambda\mu_0^2+\sum_{j=1}^{N_D}w_jd_j^2\right)\Big/2-\left(\lambda\mu_0+\sum_{j=1}^{N_D}w_jd_j\right)^2\Big/\left(2\left(\lambda+\sum_{j=1}^{N_D}w_j\right)\right)$.

The vector of IC posterior parameters, the predictive's sufficient statistic and $K(\boldsymbol{\tau_n})$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(2(\hat{a}_n+1),2\hat{b}_n+\hat{\lambda}_n\hat{\mu}_n^2,\hat{\lambda}_n\hat{\mu}_n,\hat{\lambda}_n\right),\ \ \boldsymbol{t}_1(x_{n+1}) = \left(1,x_{n+1}^2,x_{n+1},1\right)\quad\text{and}$$

$$K(\boldsymbol{\tau_n}) = \frac{\sqrt{2\pi}}{\tau_{n,3}}\cdot\frac{\Gamma\left(\dfrac{\tau_{n,0}-3}{2}\right)}{\left(\dfrac{\tau_{n,1}}{2}-\dfrac{\tau_{n,2}^2}{2\tau_{n,3}}\right)^{\frac{\tau_{n,0}-3}{2}}}$$

For the OOC shift, we introduce a step change of size $k\cdot\hat{\theta}_2$ to the mean (i.e. the OOC parameter will be $\theta_1+k\hat{\theta}_2$), where $\hat{\theta}_2 = \sqrt{\hat{b}_n/\hat{a}_n}$ (the shift will be upward or downward, depending on whether $k>0$ or $k<0$ respectively). The parameter $\hat{\theta}_2$ is the mean of the posterior marginal for the standard deviation θ_2. This choice preserves the conjugacy and expresses the shift in terms of the estimated standard deviation. Furthermore, it is always well defined when the predictive is available and it allows the pivotal statistic to depend only on $\hat{\lambda}_n$. Given that the posterior marginal Student t is a location family, the OOC posterior is $(\theta_1,\theta_2^2)\,|\boldsymbol{\tau}_n' \sim NIG\left(\hat{\mu}_n+k\cdot\hat{\theta}_2,\hat{\lambda}_n,\hat{a}_n,\hat{b}_n\right)$, while the corresponding predictive and the intervened posterior parameters are

$f'\left(X_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(\hat{\mu}_n+k\cdot\hat{\theta}_2,\left(\hat{\lambda}_n+1\right)\cdot\hat{b}_n/\left(\hat{\lambda}_n\cdot\hat{a}_n\right)\right)$ and

$\boldsymbol{\tau}'_n = \left(2(\hat{a}_n+1), 2\hat{b}_n + \hat{\lambda}_n\hat{\mu}_n^2, \hat{\lambda}_n(\hat{\mu}_n + k\cdot\hat{\theta}_2), \hat{\lambda}_n\right)$ respectively. Standardizing the future observable (using the IC parameters) we get $Z_{n+1} = (X_{n+1} - \hat{\mu}_n)/\sqrt{\left(\hat{\lambda}_n+1\right)\cdot\hat{b}_n/\left(\hat{\lambda}_n\cdot\hat{a}_n\right)}$. Then the IC and OOC predictive will be $f(Z_{n+1}|\boldsymbol{X}_n) = t_{2\hat{a}_n}(0,1)$ and $f'(Z_{n+1}|\boldsymbol{X}_n) = t_{2\hat{a}_n}\left(k\cdot\sqrt{\hat{\lambda}_n/\left(\hat{\lambda}_n+1\right)}, 1\right)$ respectively. The score function $log(L_{n+1})$ will be given by

$$log(L_{n+1}) = (\hat{a}_n + 1/2)\cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + \left(z_{n+1} - k\cdot\hat{\lambda}_n/(\hat{\lambda}_n+1)\right)^2}$$

A7: PRC for the variance of a Normal likelihood with both parameters unknown.

In this scenario, the likelihood, prior and the IC posterior distributions are identical to the ones of scenario **A6**. However, here we consider the PRC for the variance term and so for the OOC shift, we multiply the variance by k, i.e. $k\cdot\theta_2^2$. The shift corresponds to a $(k-1)\cdot 100\%$ variance increase if $k > 1$ or a $(1-k)\cdot 100\%$ decrease when $k < 1$. Furthermore, as the posterior marginal of θ_2^2 is Inverse Gamma, i.e. a scale family, the OOC posterior will be given by $(\theta_1, \theta_2^2)|\boldsymbol{\tau}'_n \sim NIG\left(\hat{\mu}_n, \hat{\lambda}_n, \hat{a}_n, k\cdot\hat{b}_n\right)$, while the corresponding predictive will be $f'(X_{n+1}|\boldsymbol{X}_n) = t_{2\hat{a}_n}\left(\hat{\mu}_n, k\cdot\left(\hat{\lambda}_n+1\right)\cdot\hat{b}_n/\left(\hat{\lambda}_n\cdot\hat{a}_n\right)\right)$. Thus the vector of the intervened posterior parameters will be $\boldsymbol{\tau}'_n = \left(2(\hat{a}_n+1), 2k\cdot\hat{b}_n + \hat{\lambda}_n\hat{\mu}_n^2, \hat{\lambda}_n\hat{\mu}_n, \hat{\lambda}_n\right)$. Standardizing the future observable (just as in **A6**) we get the standardized IC and OOC predictive distributions to be $f(Z_{n+1}|\boldsymbol{X}_n) = t_{2\hat{a}_n}(0,1)$ and $f'(Z_{n+1}|\boldsymbol{X}_n) = t_{2\hat{a}_n}(0,k)$ respectively. Finally, the score function $log(L_{n+1})$ will be

$$log(L_{n+1}) = (\hat{a}_n + 1/2)\cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$$

A8: PRC for the rate of a Gamma likelihood.

Let $X_i|\theta \sim G(\alpha,\theta)$, where α is the known shape parameter, and for the unknown rate parameter we assume that $\theta \sim G(c,d)$. Then, the resulting IC posterior and

predictive will be $\theta|\boldsymbol{\tau}_n \sim G\left(\hat{c}_n, \hat{d}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = CompG\left(\alpha, \hat{c}_n, \hat{d}_n\right)$ (i.e. Compound Gamma) respectively, where $\hat{c}_n = c + \alpha \sum_{j=1}^{N_D} w_j$ and $\hat{d}_n = d + \sum_{j=1}^{N_D} w_j d_j$. Therefore, the vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\frac{\hat{c}_n - 1}{\alpha}, \hat{d}_n\right), \quad \boldsymbol{t}_f(x_{n+1}) = (1, x_{n+1}) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma(\alpha\tau_{n,0} + 1)}{\tau_{n,1}^{\alpha\tau_{n,0}+1}}$$

Just as in the Poisson case, the OOC scenario is introduced as a shift to the rate θ parameter, by multiplying it by k, representing a $(k-1)\cdot 100\%$ rate increase if $k > 1$ or a $(1-k)\cdot 100\%$ decrease when $k < 1$. As Gamma is a scale family it follows that the OOC posterior will be $\theta|\boldsymbol{\tau}'_n \sim G\left(\hat{c}_n, \hat{d}_n/k\right)$, and the corresponding predictive will be $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = CompG\left(\alpha, \hat{c}_n, \hat{d}_n/k\right)$. Therefore, the vector of intervened posterior parameters will be $\boldsymbol{\tau}'_n = \left(\frac{\hat{c}_n - 1}{\alpha}, \frac{\hat{d}_n}{k}\right)$. Finally, the score function $log\left(L_{n+1}\right)$ will be given by

$$log\left(L_{n+1}\right) \;\; = \;\; (\hat{c}_n + \alpha) \cdot log\frac{\hat{d}_n + x_{n+1}}{\hat{d}_n + k \cdot x_{n+1}} + \alpha \cdot logk$$

A9: PRC for the scale of a Weibull likelihood.

If $X_i|\theta \sim W(\theta, \kappa)$, where κ is the known shape parameter, and for the unknown scale parameter we assume $\theta \sim IG\left(a, b\right)$. The IC posterior and predictive distributions will be $\theta|\boldsymbol{\tau}_n \sim IG\left(\hat{a}_n, \hat{b}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = Burr\left(\kappa, \hat{a}_n, \hat{b}_n^{1/\kappa}\right)$ respectively, where $\hat{a}_n = a + \sum_{j=1}^{N_D} w_j$ and $\hat{b}_n = b + \sum_{j=1}^{N_D} w_j d_j^{\kappa}$. Thus, the vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\hat{a}_n + 1, \hat{b}_n\right), \quad \boldsymbol{t}_f(x_{n+1}) = \left(1, x_{n+1}^{\kappa}\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma(\tau_{n,0} - 1)}{\tau_{n,1}^{\tau_{n,0}-1}}$$

Similarly to scenario **A5**, we introduce the OOC shift by multiplying the scale parameter θ^{κ} by k. The shift corresponds to a $(k-1)\cdot 100\%$ scale increase if $k > 1$

or a $(1-k)\cdot 100\%$ decrease when $k<1$. The Inverse Gamma is a scale family, thus the OOC posterior will be $\theta^{\kappa}|\boldsymbol{\tau}_n' \sim IG\left(\hat{a}_n, k\cdot\hat{b}_n\right)$ and the corresponding OOC predictive will be given by $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = Burr\left(\kappa, \hat{a}_n, \left(k\cdot\hat{b}_n\right)^{1/\kappa}\right)$. Finally, the vector of the intervened posterior parameters is $\boldsymbol{\tau}_n' = \left(\hat{a}_n+1, k\cdot\hat{b}_n\right)$, while the score function becomes

$$log\left(L_{n+1}\right) \;=\; (\hat{a}_n+1)\cdot log\frac{\hat{b}_n + x_{n+1}^{\kappa}}{\hat{b}_n + x_{n+1}^{\kappa}/k} - logk$$

A10: PRC for the scale of an Inverse Gamma likelihood.

Let $X_i|\theta \sim IG(\alpha,\theta)$, where α is the known shape parameter while for the unknown scale parameter we assume $\theta \sim G\left(c,d\right)$. The IC posterior is $\theta|\boldsymbol{\tau}_n \sim G\left(\hat{c}_n, \hat{d}_n\right)$, while the resulting predictive is $f\left(X_{n+1}|\boldsymbol{X}_n\right) = GB2\left(-1, 1/\hat{d}_n, \alpha, \hat{c}_n\right)$ (i.e. Generalized Beta of the second kind), where $\hat{c}_n = c+\alpha\sum_{j=1}^{N_D} w_j$ and $\hat{d}_n = d+\sum_{j=1}^{N_D} w_j/d_j$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\frac{\hat{c}_n-1}{\alpha}, \hat{d}_n\right),\;\; \boldsymbol{t}_f(x_{n+1}) = \left(1, \frac{1}{x_{n+1}}\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma(\alpha\tau_{n,0}+1)}{\left(\tau_{n,1}\right)^{\alpha\tau_{n,0}+1}}$$

Similarly to earlier scenarios, where Gamma was the prior, we introduce the shift to the shape θ by multiplying it by k, which represents a $(k-1)\cdot 100\%$ scale increase if $k>1$ or a $(1-k)\cdot 100\%$ decrease if $k<1$. Gamma is a scale family, thus the OOC posterior will be $\theta|\boldsymbol{\tau}_n' \sim G\left(\hat{c}_n, \hat{d}_n/k\right)$, and the corresponding predictive will be $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = GB2\left(-1, k/\hat{d}_n, \alpha, \hat{c}_n\right)$. The intervened posterior parameters will be $\boldsymbol{\tau}_n' = \left(\frac{\hat{c}_n-1}{\alpha}, \hat{d}_n/k\right)$ and the score function $log\left(L_{n+1}\right)$ will be given by

$$log\left(L_{n+1}\right) \;=\; (\hat{c}_n+\alpha)\cdot log\frac{\hat{d}_n\cdot x_{n+1}+1}{\hat{d}_n\cdot x_{n+1}+k} + \alpha\cdot logk$$

A11: PRC for the shape of Pareto likelihood.

Let $X_i|\theta \sim Pa(\theta, m)$, where m is the known minimum parameter, and for the shape parameter we assume $\theta \sim G(c, d)$. The IC posterior and predictive distribution are $\theta|\boldsymbol{\tau}_n \sim G\left(\hat{c}_n, \hat{d}_n\right)$ and $f(X_{n+1}|\boldsymbol{X}_n) = expGPD\left(\hat{d}_n/(m \cdot \hat{c}_n), \hat{c}_n^{-1}\right)$ (i.e. exponentiated Generalized Pareto Distribution) respectively, where $\hat{c}_n = c + \sum_{j=1}^{N_D} w_j$ and $\hat{d}_n = d + \sum_{j=1}^{N_D} w_j log(d_j/m)$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K(\boldsymbol{\tau_n})$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(\hat{c}_n - 1, \hat{d}_n\right), \ \boldsymbol{t}_f(x_{n+1}) = (1, log(x_{n+1}/m)) \quad \text{and} \quad K(\boldsymbol{\tau_n}) = \frac{\Gamma(\tau_{n,0}+1)}{\tau_{n,1}^{\tau_{n,0}+1}}$$

Just as it was done in the earlier cases where Gamma was involved as prior, we multiply the shape θ by k, which represents a $(k-1)\cdot 100\%$ shape increase if $k > 1$ or to a $(1-k)\cdot 100\%$ decrease when $k < 1$. As Gamma is a scale family, the OOC posterior $\theta|\boldsymbol{\tau}'_n \sim G\left(\hat{c}_n, \hat{d}_n/k\right)$, and the OOC predictive: $f'(X_{n+1}|\boldsymbol{X}_n) = expGPD\left(\hat{d}_n/(k \cdot m \cdot \hat{c}_n), \hat{c}_n^{-1}\right)$. The intervened posterior parameters will be $\boldsymbol{\tau}'_n = \left(\frac{\hat{c}_n - 1}{\alpha}, \hat{d}_n/k\right)$ and the score function $log(L_{n+1})$ will be given by

$$log(L_{n+1}) \ = \ (\hat{c}_n + 1) \cdot log\frac{\hat{d}_n + log(x_{n+1}/m)}{\hat{d}_n + k \cdot log(x_{n+1}/m)} + logk$$

A12: PRC for the scale of Lognormal likelihood with known shape parameter.

Let $X_i|\theta \sim LogN(\theta, \sigma^2)$, where σ^2 is the known shape parameter, and for the scale parameter we assume $\theta \sim N(\mu_0, \sigma_0^2)$. Similarly to the corresponding Normal case (scenario **A4**) we have that the IC posterior and predictive distributions to be $\theta|\boldsymbol{\tau}_n \sim N(\hat{\mu}_n, \ \hat{\sigma}_n^2)$ and $f(X_{n+1}|\boldsymbol{X}_n) = LogN(\hat{\mu}_n, \ \hat{\sigma}_n^2 + \sigma^2)$ respectively, where $\hat{\mu}_n = \left(\sigma^2\mu_0 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j log(d_j)\right) \Big/ \left(\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j\right)$ and $\hat{\sigma}_n^2 = {\sigma_0}^2\sigma^2 \Big/ \left(\sigma^2 + {\sigma_0}^2 \sum_{j=1}^{N_D} w_j\right)$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K(\boldsymbol{\tau_n})$,

needed in PRC are

$$\boldsymbol{\tau}_n = \left(\frac{\sigma^2}{\hat{\sigma}_n^2}, \frac{\hat{\mu}_n}{\hat{\sigma}_n^2}\right), \ \boldsymbol{t}_1\left(x_{n+1}\right) = \left(1, \frac{log(x_{n+1})}{\sigma^2}\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \sqrt{\frac{2\pi\sigma^2}{\tau_{n,0}}} exp\left\{\frac{\sigma^2\tau_{n,1}^2}{2\tau_{n,0}}\right\}$$

For the OOC shift, we introduce a step change of size of $k \cdot \sigma$ for θ, i.e. the OOC parameter is $\theta + k \cdot \sigma$ with the shift being upwards or downwards depending if $k > 0$ or $k < 0$ respectively. Since the Normal is a location family, the OOC posterior will be $\theta|\boldsymbol{\tau}_n' \sim N\left(\hat{\mu}_n + k \cdot \sigma,\ \hat{\sigma}_n^2\right)$ with the corresponding predictive $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = LogN\left(\hat{\mu}_n + k \cdot \sigma,\ \hat{\sigma}_n^2 + \sigma^2\right)$. The vector of the intervened posterior parameters will be $\boldsymbol{\tau}_n' = \left(\frac{\sigma^2}{\hat{\sigma}_n^2}, \frac{\hat{\mu}_n + k \cdot \sigma}{\hat{\sigma}_n^2}\right)$. If we will standardize the log-transformed future observable, setting $Z_{n+1} = \left(log(X_{n+1}) - \hat{\mu}_n\right) / \sqrt{\hat{\sigma}_n^2 + \sigma^2}$, then the standardized predictives will be $f\left(Z_{n+1}|\boldsymbol{X}_n\right) = N\left(0,1\right)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right) = N\left(k \cdot \sigma / \sqrt{\hat{\sigma}_n^2 + \sigma^2}, 1\right)$. The score function $log\left(L_{n+1}\right)$ will be given by:

$$log\left(L_{n+1}\right) \ = \ \left(z_{n+1} - \frac{k}{2} \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}\right) \cdot \frac{\sigma}{\sqrt{\hat{\sigma}_n^2 + \sigma^2}}$$

A13: PRC for the shape of Lognormal likelihood with known scale parameter.

Let $X_i|\theta^2 \sim LogN\left(\mu, \theta^2\right)$, where μ is the known scale, and for the shape parameter we assume $\theta^2 \sim IG\left(a, b\right)$. Similarly to the corresponding Normal case (scenario **A5**) we have that the IC posterior and predictive distributions to be $\theta^2|\boldsymbol{\tau}_n \sim IG\left(\hat{a}_n,\ \hat{b}_n\right)$, and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = Logt_{2\hat{a}_n}\left(\mu,\ \hat{b}_n/\hat{a}_n\right)$ respectively, where $\hat{a}_n = a + \sum_{j=1}^{N_D} w_j/2$ and $\hat{b}_n = b + \sum_{j=1}^{N_D} w_j \left(log(d_j) - \mu\right)^2/2$. The vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$$\boldsymbol{\tau}_n = \left(2(\hat{a}_n + 1), 2\hat{b}_n\right), \ \boldsymbol{t}_1\left(x_{n+1}\right) = \left(1, \left(log(x_{n+1}) - \mu\right)^2\right) \quad \text{and} \quad K\left(\boldsymbol{\tau_n}\right) = \frac{\Gamma\left(\frac{\tau_{n,0}}{2} - 1\right)}{\left(\frac{\tau_{n,1}}{2}\right)^{\frac{\tau_{n,0}}{2}-1}}$$

For the OOC shift, we multiply the shape parameter by k, i.e. $k \cdot \theta^2$. The shift corresponds to a $(k-1)\cdot 100\%$ increase if $k > 1$ or to a $(1-k)\cdot 100\%$ decrease when $k < 1$. Since, Inverse Gamma is a scale family, the OOC posterior and predictive will be $\theta|\boldsymbol{\tau}'_n \sim IG\left(\hat{a}_n,\ k\cdot\hat{b}_n\right)$ and $f'\left(X_{n+1}|\boldsymbol{X}_n\right) = Logt_{2\hat{a}_n}\left(\mu,\ k\cdot\hat{b}_n/\hat{a}_n\right)$. The vector of the intervened posterior parameters will be $\boldsymbol{\tau}'_n = \left(2(\hat{a}_n+1),\ k\cdot 2\hat{b}_n\right)$. Standardizing the log-transformed future observable we have $Z_{n+1} = \left(log(X_{n+1})-\mu\right)/\sqrt{\hat{b}_n/\hat{a}_n}$, resulting the IC and OOC predictive distributions to be $f\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}(0,1)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}(0,k)$ respectively. Finally, the score function will be

$$log\left(L_{n+1}\right) \quad = \quad (\hat{a}_n + 1/2)\cdot log\frac{2\hat{a}_n + z^2_{n+1}}{2\hat{a}_n + z^2_{n+1}/k} - log\sqrt{k}$$

A14: PRC for the scale of Lognormal likelihood with both parameters unknown.

Let $X_i|\left(\theta_1,\theta_2^2\right) \sim LogN\left(\theta_1,\theta_2^2\right)$, where both parameters are being unknown and we assume $\left(\theta_1,\theta_2^2\right) \sim NIG\left(\mu_0,\lambda,a,b\right)$. Similarly to the corresponding Normal case (scenario **A6**) we have that the IC posterior and predictive distributions will be $\left(\theta_1,\theta_2^2\right)|\boldsymbol{\tau}_n \sim NIG\left(\hat{\mu}_n,\hat{\lambda}_n,\hat{a}_n,\hat{b}_n\right)$ and $f\left(X_{n+1}|\boldsymbol{X}_n\right) = Logt_{2\hat{a}_n}\left(\hat{\mu}_n,\ \left(\lambda_n+1\right)\cdot\hat{b}_n/\left(\lambda_n\cdot\hat{a}_n\right)\right)$ respectively, where
$\hat{\mu}_n = \left(\lambda\mu_0 + \sum_{j=1}^{N_D} w_j log(d_j)\right) \Big/ \left(\lambda + \sum_{j=1}^{N_D} w_j\right)$, $\hat{\lambda}_n = \lambda + \sum_{j=1}^{N_D} w_j$, $\hat{a}_n = a + \sum_{j=1}^{N_D} w_j/2$ and
$\hat{b}_n = b + \left(\lambda\mu_0^2 + \sum_{j=1}^{N_D} w_j(log(d_j))^2\right) \Big/ 2 - \left(\lambda\mu_0 + \sum_{j=1}^{N_D} w_j log(d_j)\right)^2 \Big/ \left(2\left(\lambda + \sum_{j=1}^{N_D} w_j\right)\right)$.
The vector of IC posterior parameters, the predictive's sufficient statistic and $K\left(\boldsymbol{\tau_n}\right)$, needed in PRC are

$\boldsymbol{\tau}_n = \left(2(\hat{a}_n+1), 2\hat{b}_n + \hat{\lambda}_n\hat{\mu}_n^2, \hat{\lambda}_n\hat{\mu}_n, \hat{\lambda}_n\right)$, $\boldsymbol{t}_1\left(x_{n+1}\right) = \left(1, (log(x_{n+1}))^2, log(x_{n+1}), 1\right)$ and

$$K\left(\boldsymbol{\tau_n}\right)=\frac{\sqrt{2\pi}}{\tau_{n,3}}\cdot\frac{\Gamma\left(\frac{\tau_{n,0}-3}{2}\right)}{\left(\frac{\tau_{n,1}}{2}-\frac{\tau_{n,2}^2}{2\tau_{n,3}}\right)^{\frac{\tau_{n,0}-3}{2}}}$$

For the OOC shift, we introduce a step change of size of $k\cdot\hat{\theta}_2$ to the mean, where $\hat{\theta}_2=\sqrt{\hat{b}_n/\hat{a}_n}$, (i.e the expected value of the posterior marginal for the θ_2) and so the OOC parameter will be $\theta_1+k\hat{\theta}_2$. The shift is upward or downward depending on whether $k>0$ or $k<0$ respectively. As the posterior marginal Student t is a location family, the OOC posterior is $\left(\theta_1,\theta_2^2\right)|\boldsymbol{\tau}_n'\sim NIG\left(\hat{\mu}_n+k\cdot\hat{\theta}_2,\hat{\lambda}_n,\hat{a}_n,\hat{b}_n\right)$, while the corresponding predictive is $f'\left(X_{n+1}|\boldsymbol{X}_n\right)=Logt_{2\hat{a}_n}\left(\hat{\mu}_n+k\cdot\hat{\theta}_2,\ \left(\lambda_n+1\right)\cdot\hat{b}_n/\left(\lambda_n\cdot\hat{a}_n\right)\right)$. The vector of intervened posterior parameters is $\boldsymbol{\tau}_n'=\left(2(\hat{a}_n+1),2\hat{b}_n+\hat{\lambda}_n\hat{\mu}_n^2,\hat{\lambda}_n(\hat{\mu}_n+k\cdot\hat{\theta}_2),\hat{\lambda}_n\right)$. Standardizing the the log-transformed future observable (using the IC parameters) we get
$Z_{n+1}=\left(log(X_{n+1})-\hat{\mu}_n\right)\Big/\sqrt{\left(\hat{\lambda}_n+1\right)\cdot\hat{b}_n/\left(\hat{\lambda}_n\cdot\hat{a}_n\right)}$. Then the IC and OOC predictive will be $f\left(Z_{n+1}|\boldsymbol{X}_n\right)=t_{2\hat{a}_n}\left(0,1\right)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right)=t_{2\hat{a}_n}\left(k\cdot\sqrt{\hat{\lambda}_n/\left(\hat{\lambda}_n+1\right)},1\right)$ respectively. The score function $log\left(L_{n+1}\right)$ will be given by

$$log\left(L_{n+1}\right)\ =\ \left(\hat{a}_n+1/2\right)\cdot log\frac{2\hat{a}_n+z_{n+1}^2}{2\hat{a}_n+\left(z_{n+1}-k\cdot\hat{\lambda}_n/(\hat{\lambda}_n+1)\right)^2}$$

A15: PRC for the shape of Lognormal likelihood with both parameters unknown.

The likelihood and the IC distributions and parameters are identical with the ones presented in scenario **A14**, but for the OOC shift, we multiply the shape parameter θ^2 by k, referring to a $(k-1)\cdot100\%$ increase if $k>1$ or $(1-k)\cdot100\%$ decrease when $k<1$. Furthermore, as the posterior marginal (Inverse Gamma) is a scale family, the OOC posterior and predictive will be $\left(\theta_1,\theta_2^2\right)|\boldsymbol{\tau}_n'\sim NIG\left(\hat{\mu}_n,\hat{\lambda}_n,\hat{a}_n,k\cdot\hat{b}_n\right)$ and $f'\left(X_{n+1}|\boldsymbol{X}_n\right)=Logt_{2\hat{a}_n}\left(\hat{\mu}_n,\ k\cdot\left(\lambda_n+1\right)\cdot\hat{b}_n/\left(\lambda_n\cdot\hat{a}_n\right)\right)$ respectively. The intervened posterior parameters will be $\boldsymbol{\tau}_n'=\left(2(\hat{a}_n+1),2k\cdot\hat{b}_n+\hat{\lambda}_n\hat{\mu}_n^2,\hat{\lambda}_n\hat{\mu}_n,\hat{\lambda}_n\right)$.

Standardizing the the log-transformed future observable (just as in **A14**) we get the standardized IC and OOC predictive distributions to be $f\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(0,1\right)$ and $f'\left(Z_{n+1}|\boldsymbol{X}_n\right) = t_{2\hat{a}_n}\left(0,k\right)$ respectively. Finally, the score function $log\left(L_{n+1}\right)$ will be

$$log\left(L_{n+1}\right) \quad = \quad \left(\hat{a}_n + 1/2\right) \cdot log\frac{2\hat{a}_n + z_{n+1}^2}{2\hat{a}_n + z_{n+1}^2/k} - log\sqrt{k}$$

Appendix B: Decision thresholds h for PRC Models

Table 6.0.1 provides the decision threshold h for the Normal (or equivalently the logarithm based transformation of a Lognormal) likelihood for scenario 1 in Section 3.2. Specifically, we derive the h values for different choices of $(FWER, N)$ or ARL_0 and specific size of OOC parameter shift k, when we make use of a reference prior and no historical data are available. The models for which the h decision limit is calculated refer to the following scenarios:

I. PRC for mean shift of a Normal Likelihood with known variance.

II. PRC for variance shift of a Normal Likelihood with known mean.

III. PRC for mean shift of a Normal Likelihood with both mean and variance unknown.

IV. PRC for variance shift of a Normal Likelihood with both mean and variance unknown.

		$(k,\ model)$															
		$(0.5, m_1)$	$(1, m_1)$	$(1.5, m_1)$	$(2, m_1)$	$(0.5, m_2)$	$(1, m_2)$	$(1.5, m_2)$	$(2, m_2)$	$(0.5, m_3)$	$(1, m_3)$	$(1.5, m_3)$	$(2, m_3)$	$(0.5, m_4)$	$(1, m_4)$	$(1.5, m_4)$	$(2, m_4)$
$\boldsymbol{ARL_0}$	**50**	3.244	2.177	1.574	1.173	2.305	1.624	0.974	1.320	1.564	2.144	2.378	2.436	2.290	1.614	0.970	1.311
	100	4.355	2.823	2.025	1.526	2.970	2.195	1.406	1.839	2.132	2.799	3.041	3.103	2.958	2.186	1.404	1.836
	150	5.050	3.208	2.287	1.727	3.358	2.552	1.691	2.171	2.493	3.192	3.434	3.492	3.353	2.544	1.689	2.169
	200	5.556	3.487	2.472	1.869	3.644	2.807	1.906	2.415	2.749	3.477	3.717	3.768	3.636	2.807	1.902	2.415
	250	5.956	3.703	2.617	1.981	3.862	3.013	2.077	2.610	2.957	3.694	3.935	3.984	3.856	3.006	2.080	2.612
	300	6.294	3.881	2.738	2.069	4.042	3.184	2.225	2.772	3.128	3.873	4.112	4.154	4.035	3.181	2.227	2.773
	370	6.685	4.088	2.877	2.173	4.250	3.377	2.398	2.963	3.323	4.078	4.317	4.359	4.242	3.379	2.396	2.957
	500	7.245	4.383	3.075	2.320	4.540	3.664	2.657	3.238	3.610	4.377	4.615	4.656	4.544	3.656	2.660	3.241
	600	7.591	4.563	3.197	2.411	4.724	3.836	2.816	3.408	3.788	4.558	4.801	4.836	4.727	3.836	2.816	3.407
	750	8.017	4.782	3.343	2.521	4.952	4.053	3.013	3.623	3.999	4.773	5.020	5.055	4.947	4.054	3.014	3.618
	1000	8.569	5.067	3.535	2.663	5.235	4.334	3.270	3.893	4.277	5.057	5.301	5.332	5.232	4.332	3.275	3.896
	1500	9.358	5.470	3.806	2.866	5.642	4.728	3.648	4.285	4.674	5.467	5.706	5.735	5.642	4.730	3.650	4.287
$\boldsymbol{(FWER\%,\ N)}$	**(1%, 20)**	7.362	5.178	3.810	2.929	4.794	3.058	2.465	3.565	2.617	4.178	5.006	5.458	4.726	3.013	2.454	3.533
	(5%, 20)	5.509	3.799	2.772	2.138	3.608	2.395	1.797	2.554	1.994	3.118	3.724	4.040	3.562	2.360	1.781	2.532
	(10%, 20)	4.621	3.182	2.333	1.801	3.082	2.063	1.483	2.099	1.692	2.623	3.126	3.413	3.043	2.031	1.468	2.079
	(1%, 30)	8.673	5.739	4.138	3.170	5.532	3.823	3.034	4.200	3.241	4.882	5.653	6.021	5.507	3.798	3.020	4.185
	(5%, 30)	6.511	4.293	3.096	2.364	4.175	2.958	2.209	3.024	2.451	3.656	4.238	4.535	4.149	2.933	2.198	3.012
	(10%, 30)	5.496	3.646	2.635	2.011	3.593	2.537	1.830	2.506	2.087	3.098	3.613	3.876	3.567	2.513	1.819	2.493
	(1%, 40)	9.547	6.123	4.343	3.321	6.024	4.379	3.444	4.647	3.731	5.402	6.123	6.453	6.004	4.343	3.435	4.634
	(5%, 40)	7.177	4.628	3.304	2.514	4.582	3.350	2.511	3.358	2.800	4.023	4.588	4.866	4.562	3.335	2.505	3.351
	(10%, 40)	6.126	3.959	2.836	2.165	3.940	2.886	2.090	2.805	2.381	3.431	3.930	4.205	3.924	2.869	2.084	2.794
	(1%, 50)	10.290	6.422	4.522	3.445	6.371	4.789	3.758	4.961	4.098	5.758	6.447	6.727	6.352	4.773	3.750	4.949
	(5%, 50)	7.748	4.880	3.470	2.636	4.860	3.658	2.751	3.614	3.073	4.302	4.847	5.114	4.848	3.645	2.744	3.603
	(10%, 50)	6.630	4.206	3.004	2.279	4.217	3.158	2.308	3.029	2.620	3.684	4.180	4.439	4.204	3.144	2.302	3.022
	(1%, 75)	11.538	6.915	4.841	3.669	6.922	5.492	4.330	5.458	4.741	6.286	6.880	7.152	6.912	5.477	4.318	5.450
	(5%, 75)	8.742	5.325	3.754	2.849	5.355	4.210	3.214	4.064	3.563	4.789	5.313	5.544	5.349	4.201	3.209	4.059
	(10%, 75)	7.498	4.640	3.288	2.489	4.689	3.633	2.698	3.437	3.050	4.136	4.615	4.855	4.682	3.625	2.695	3.433
	(1%, 100)	12.332	7.283	5.042	3.811	7.273	5.948	4.740	5.844	5.166	6.640	7.205	7.440	7.263	5.941	4.735	5.843
	(5%, 100)	9.375	5.620	3.963	3.002	5.697	4.596	3.529	4.379	3.919	5.122	5.613	5.846	5.692	4.590	3.527	4.375
	(10%, 100)	8.117	4.933	3.488	2.639	5.018	3.978	2.987	3.736	3.356	4.453	4.924	5.152	5.012	3.973	2.984	3.730

Table 6.0.1: The threshold h for different choices of ARL_0 or $(FWER\%, N)$, size of shift k and *model* for Normal data. Specifically, m_1, m_2, m_3 and m_4 represent the $N-N$, the $N-IG$, the mean-$N-NIG$ and the variance-$N-NIG$ model respectively.

Appendix C: List of U3S Models

C1: Normal - Normal Model for the mean $\boldsymbol{\theta}$

We provide the U3S model, assuming the distribution of the IC state be a Normal distribution with only the mean θ be unknown, wanting to investigate a potential shift for it, either upwards or downwards. In other words, $X_i|\theta \sim N(\theta, \sigma^2)$, $i = 1, ..., n$ are sequentially arrived observations. Regarding the IC mean we assume $\theta \sim N(\mu_0, \sigma_0^2)$. As OOC parameter, we set $\boldsymbol{\phi} = \delta$, which represents the magnitude of a shift in terms of the standard deviation and we define $g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \theta + \delta \cdot \sigma$ as the link function for the OOC state, replacing the mean. Regarding δ we select a mixture of of Normal distributions. This setting gives a two-sided U3S for an upward and a downward shift respectively. Alternatively, instead of a mixture of priors for a two-sided U3S, we could implement two one-sided U3S, using simple priors for δ. For the mixture, we assume $\delta = \gamma \cdot \delta_1 + (1 - \gamma) \cdot \delta_2$, where $\delta_i \sim N(\mu_{\delta i}, {\sigma_{\delta i}}^2)$ and $\gamma \sim Ber(\pi)$. The probability π is the prior probability of the shift δ_1 in the mixture and the choice of $\pi = 1/2$ corresponds to the same FA tolerance, upwards or downwards. Finally, for the location of the change point, we assume $\tau \sim DW(p, \beta)$. The likelihood is:

$$f(\boldsymbol{x}|\theta, \delta, \tau) = \begin{cases} \prod_{i=1}^{\tau-1} f(x_i|\theta) \prod_{i=\tau}^{n} f(x_i|\theta + \delta \cdot \sigma) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f(x_i|\theta) & \text{if } \tau > n \end{cases}$$

Setting $n_t = n - t + 1$, $X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i$, $\sigma_p^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}$, $\mu_p = \left(\frac{X_{1:n}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \cdot \sigma_p^2$, $e_{\delta i} = n_\tau \sigma_{\delta i}^2 \left(\sigma^2 - n_\tau \sigma_p^2\right) + \sigma^2$, $\pi_1 = \pi$ and $\pi_2 = 1 - \pi$, then the Bayes Factor will be:

$$BF_{\tau, n+} = \frac{\sum_{i=1}^{2} \pi_i f(\boldsymbol{x}|\tau \leq n)}{f(\boldsymbol{x}|\tau > n)} \tag{6.0.1}$$

where the IC and the OOC marginal will be respectively:

$$
\begin{aligned}
f(\boldsymbol{x}|\tau > n) &= \int_{\theta} f(\boldsymbol{x}|\theta, \tau > n)\pi(\theta)d\theta \\
&= \int_{\theta} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n} \cdot exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right\} \cdot \frac{1}{\sigma_0\sqrt{2\pi}} \cdot exp\left\{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\} d\theta \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n} \cdot \frac{1}{\sigma_0\sqrt{2\pi}} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2}\right)\right\} \times \\
&\quad \int_{\theta} exp\left\{-\frac{-2(\sigma_0^2 X_{1:n} + \sigma^2\mu_0)\theta + (n\sigma_0^2 + \sigma^2)\theta^2}{2\sigma^2\sigma_0^2}\right\} d\theta \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^{n} \cdot \frac{\sigma_p}{\sigma_0\sigma^n} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_p^2}{\sigma_p^2}\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{x}|\tau \leq n) &= \sum_{i=1}^{2} \pi_i \int_{\delta_i}\int_{\theta} f(\boldsymbol{x}|\theta, \delta_i, \tau \leq n)\pi(\theta)\pi(\delta_i)d\theta d\delta_i \\
&= \sum_{i=1}^{2} \pi_i \int_{\delta_i}\int_{\theta} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n} \cdot exp\left\{-\frac{\sum_{i=1}^{\tau-1}(x_i-\theta)^2 + \sum_{i=\tau}^{n}(x_i-(\theta+\delta_i\sigma))^2}{2\sigma^2}\right\} \times \\
&\quad \times \frac{1}{\sigma_0\sqrt{2\pi}} \cdot exp\left\{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\} d\theta\pi(\delta_i)d\delta_i
\end{aligned}
$$

$$= \sum_{i=1}^{2} \pi_i \int_{\delta_i} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \frac{1}{\sigma_0\sqrt{2\pi}} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{2X_{\tau:n}\delta_i}{\sigma} + n_\tau\delta_i^2\right)\right\} \times$$

$$\times \int_{\theta} exp\left\{-\frac{-2(\sigma_0^2 X_{1:n} + \sigma^2\mu_0 - \sigma_0^2 n_\tau\delta_i\sigma)\theta + (n\sigma_0^2 + \sigma^2)\theta^2}{2\sigma^2\sigma_0^2}\right\} d\theta\pi(\delta_i)d\delta_i$$

$$= \sum_{i=1}^{2} \pi_i \int_{\delta_i} \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\sigma_p}{\sigma_0 \cdot \sigma^n} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{2X_{\tau:n}\delta_i}{\sigma} + n_\tau\delta_i^2\right)\right\} \times$$

$$\times exp\left\{+\frac{\left(\mu_p - n_\tau\delta_i\sigma_p^2/\sigma\right)^2}{2\sigma_p^2}\right\} \cdot \frac{1}{\sigma_{\delta_i}\sqrt{2\pi}} \cdot exp\left\{-\frac{(\delta_i - \mu_{\delta_i})^2}{2\sigma_{\delta_i}^2}\right\} d\delta_i$$

$$= \sum_{i=1}^{2} \pi_i \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\sigma_p}{\sigma_0\sigma^n} \cdot \frac{1}{\sigma_{\delta_i}\sqrt{2\pi}} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_p^2}{\sigma_p^2} + \frac{\mu_{\delta_i}^2}{\sigma_{\delta_i}^2}\right)\right\} \times$$

$$\times \int_{\delta_i} exp\left\{-\frac{-2(\sigma_{\delta_i}^2(X_{\tau:n} - n_\tau\mu_p)\sigma + \sigma^2\mu_{\delta_i})\delta + e_{\delta_i}\delta_i^2}{2\sigma^2\sigma_{\delta_i}^2}\right\} d\delta_i =$$

$$= \sum_{i=1}^{2} \pi_i \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\sigma_p}{\sigma_0\sigma^{n-1}\sqrt{e_{\delta_i}}} \cdot exp\left\{-\frac{1}{2}\left(\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_p^2}{\sigma_p^2} + \frac{\mu_{\delta_i}^2}{\sigma_{\delta_i}^2}\right)\right\} \times$$

$$\times exp\left\{+\frac{(\sigma_{\delta_i}^2(X_{\tau:n} - n_\tau\mu_p)\sigma + \sigma^2\mu_{\delta_i})^2}{2\sigma_{\delta_i}^2\sigma^2 e_{\delta_i}}\right\}$$

Substituting in (6.0.1) we get

$$BF_{\tau,n_+} = \sum_{i=1}^{2} \frac{\pi_i \sigma}{\sqrt{e_{\delta i}}} \cdot exp\left\{ -\frac{\mu_{\delta i}^2}{2\sigma_{\delta i}^2} + \frac{\left(\sigma_{\delta i}^2 \left(X_{\tau:n} - n_\tau \mu_p\right) \sigma + \sigma^2 \mu_{\delta_i}\right)^2}{2\sigma_{\delta_i}^2 \sigma^2 e_{\delta_i}} \right\}$$

Regarding the full conditional posteriors we have:

- $\theta|(\tau > n, \boldsymbol{x})$

$$\begin{aligned} p(\theta|\tau > n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta, \tau > n)\pi(\theta) \\ &\propto exp\left\{ -\frac{-2(\sigma_0^2 X_{1:n} + \sigma^2 \mu_0)\theta + (n\sigma_0^2 + \sigma^2)\theta^2}{2\sigma^2 \sigma_0^2} \right\} \end{aligned}$$

Thus $\theta|(\tau > n, \boldsymbol{x}) \sim N\left(\mu_p, \sigma_p^2\right)$

- $\theta|(\delta, \tau \leq n, \boldsymbol{x})$

$$\begin{aligned} p(\theta|\delta, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta, \delta, \tau \leq n)\pi(\theta) \\ &\propto exp\left\{ -\frac{-2(\sigma_0^2 X_{1:n} + \sigma^2 \mu_0 - \sigma_0^2 n_\tau \delta \sigma)\theta + (n\sigma_0^2 + \sigma^2)\theta^2}{2\sigma^2 \sigma_0^2} \right\} \end{aligned}$$

Thus $\theta|(\delta, \tau \leq n, \boldsymbol{x}) \sim N\left(\mu_p - \sigma_p^2 n_\tau \delta/\sigma, \sigma_p^2\right)$

- $\delta_i|\left(\theta, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} p(\delta_i|\theta, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta, \delta, \tau \leq n)\pi(\delta_i) \\ &\propto exp\left\{ -\frac{-2\left(\mu_{\delta i} + \sigma_{\delta i}^2 \left(X_{\tau:n} - n_\tau \theta\right)/\sigma\right)\delta_i + \left(1 + n_\tau \sigma_{\delta i}^2\right)\delta_i^2}{2\sigma_{\delta i}^2} \right\} \end{aligned}$$

Thus $\delta_i|\left(\theta, \tau \leq n, \boldsymbol{x}\right) \sim N\left(\frac{\mu_{\delta i} + \sigma_{\delta i}^2 \left(X_{\tau:n} - n_\tau \theta\right)/\sigma}{1 + n_\tau \sigma_{\delta i}^2}, \frac{\sigma_{\delta i}^2}{1 + n_\tau \sigma_{\delta i}^2}\right)$

- $\gamma|\left(\theta, \delta_i, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned}\pi^* &= \frac{\pi \cdot \int_{\delta_1} f\left(\boldsymbol{x}|\theta, \delta_1, \tau \leq n\right) \pi(\delta_1) d\delta_1}{\pi \cdot \int_{\delta_1} f\left(\boldsymbol{x}|\theta, \delta_1, \tau \leq n\right) \pi(\delta_1) d\delta_1 + (1-\pi) \cdot \int_{\delta_2} f\left(\boldsymbol{x}|\theta, \delta_2, \tau \leq n\right) \pi(\delta_2) d\delta_2} \\ &= \frac{\pi}{\pi + (1-\pi) \cdot exp\left\{\frac{\mu_{\delta p2}^2}{2\sigma_{\delta p2}^2} - \frac{\mu_{\delta p1}^2}{2\sigma_{\delta p1}^2}\right\} \frac{\sigma_{\delta p2}}{\sigma_{\delta p1}}}\end{aligned}$$

$$\text{Thus } \gamma|\left(\theta, \delta_i, \tau \leq n, \boldsymbol{x}\right) \sim Ber\left(\frac{\pi}{\pi + (1-\pi) \cdot exp\left\{\frac{\mu_{\delta p2}^2}{2\sigma_{\delta p2}^2} - \frac{\mu_{\delta p1}^2}{2\sigma_{\delta p1}^2}\right\} \frac{\sigma_{\delta p2}}{\sigma_{\delta p1}}}\right)$$

- $\tau|\left(\theta, \delta, \boldsymbol{x}\right)$

$$\begin{aligned}p\left(\tau = k|\theta, \delta, \boldsymbol{x}\right) &= \frac{f\left(\boldsymbol{x}|\theta, \delta, \tau\right) \pi(\tau = k)}{\sum_{j=1}^{n} f\left(\boldsymbol{x}|\theta, \delta, \tau\right) \pi(\tau = j)} \\ &= \frac{exp\left\{\frac{\delta\left(X_{k:n} - n_k\theta\right)}{\sigma} - \frac{n_k\delta^2}{2}\right\}\left((1-p)^{(k-1)^\beta} - (1-p)^{k^\beta}\right)}{\sum_{j=1}^{n} exp\left\{\frac{\delta\left(X_{j:n} - n_j\theta\right)}{\sigma} - \frac{n_j\delta^2}{2}\right\}\left((1-p)^{(j-1)^\beta} - (1-p)^{j^\beta}\right)}\end{aligned}$$

C2: Normal - Inverse Gamma Model for the variance $\boldsymbol{\theta^2}$

Let the distribution of the IC state be a Normal distribution with unknown variance θ^2, while the mean μ is known, i.e. $X_i|\theta \sim N\left(\mu, \theta^2\right)$, $i = 1, ..., n$. Regarding the detection scheme, we want to investigate a level change for the variance. Regarding the IC parameter, we assume $\theta^2 \sim IG\left(a, b\right)$. We set $\boldsymbol{\phi} = \kappa$, which represents the magnitude of an inflation and for the link function for the OOC state, we replace the variance with $g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \kappa \cdot \theta^2$. For κ, we assume a mixture of Inverse Gamma distributions, for a potential inflation or shrinkage of the variance respectively. Specifically, $\kappa = \gamma \cdot \kappa_1 + (1-\gamma) \cdot \kappa_2$, where $\kappa_i \sim IG\left(c_i, d_i\right)$ and $\gamma \sim Ber(\pi)$. In

this manner, we implement again a two-sided U3S. For the change point, we assume $\tau \sim DW(p, \beta)$. The likelihood is given by:

$$f\left(\boldsymbol{x}|\theta^2, \kappa, \tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f\left(x_i|\theta^2\right) \prod_{i=\tau}^{n} f\left(x_i|\kappa \cdot \theta^2\right) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f\left(x_i|\theta^2\right) & \text{if } \tau > n \end{cases}$$

Setting $n_t = n - t + 1$, $S^2_{t_1:t_2} = \sum_{i=t_1}^{t_2}(x_i - \mu)^2$, $a_p = a + \frac{n}{2}$, $b_t = b + \frac{S^2_{1:t}}{2}$, $\pi_1 = \pi$ and $\pi_2 = 1 - \pi$, then the IC and the OOC marginal will be respectively:

$$\begin{aligned} f(\boldsymbol{x}|\tau > n) &= \int_{\theta^2} f(\boldsymbol{x}|\theta^2, \tau > n)\pi(\theta^2)d\theta^2 \\ &= \int_{\theta^2} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot exp\left\{-\frac{S^2_{1:n}}{2\theta^2}\right\} \cdot \frac{b^a}{\Gamma(a)} \cdot \left(\frac{1}{\theta^2}\right)^{a+1} \cdot exp\left\{-\frac{b}{\theta^2}\right\} d\theta^2 \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{b^a}{\Gamma(a)} \int_{\theta^2} \left(\frac{1}{\theta^2}\right)^{a_p+1} \cdot exp\left\{-\frac{b_n}{\theta^2}\right\} d\theta^2 \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a_p)}{b_n^{a_p}} \end{aligned}$$

$$\begin{aligned} f(\boldsymbol{x}|\tau \leq n) &= \sum_{i=1}^{2} \pi_i \int_{\theta^2} \int_{\kappa_i} f(\boldsymbol{x}|\theta^2, \kappa_i, \tau \leq n)\pi(\theta^2)\pi(\kappa_i)d\kappa_i d\theta^2 \\ &= \sum_{i=1}^{2} \pi_i \int_{\theta^2} \int_{\kappa_i} \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot exp\left\{-\frac{S^2_{1:\tau-1}}{2\theta^2} - \frac{S^2_{\tau:n}}{2\kappa_i\theta^2}\right\} \times \\ &\quad \times \frac{d_i^{c_i}}{\Gamma(c_i)} \cdot \left(\frac{1}{\kappa_i}\right)^{c_i+1} \cdot exp\left\{-\frac{d_i}{\kappa_i}\right\} d\kappa_i \pi(\theta^2)d\theta^2 \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{2} \pi_i \int_{\theta^2} \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n \cdot \frac{d_i^{c_i}}{\Gamma(c_i)} \cdot exp\left\{-\frac{S^2_{1:\tau-1}}{2\theta^2}\right\} \times \\
&\quad \times \int_{\kappa_i} \left(\frac{1}{\kappa_i}\right)^{n_\tau/2+c_i+1} \cdot exp\left\{-\frac{d_i + \frac{S^2_{\tau:n}}{2\theta^2}}{\kappa_i}\right\} d\kappa_i \pi(\theta^2) d\theta^2 \\
&= \sum_{i=1}^{2} \pi_i \int_{\theta^2} \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n \cdot \frac{d_i^{c_i}}{\Gamma(c_i)} \cdot exp\left\{-\frac{S^2_{1:\tau-1}}{2\theta^2}\right\} \times \\
&\quad \times \frac{\Gamma(n_\tau/2 + c_i)}{\left(d_i + \frac{S^2_{\tau:n}}{2\theta^2}\right)^{n_\tau + c_i}} \cdot \frac{d_i^{c_i}}{\Gamma(c_i)} \cdot \left(\frac{1}{\theta^2}\right)^{a+1} \cdot exp\left\{-\frac{b}{\theta^2}\right\} d\theta^2 \\
&= \sum_{i=1}^{2} \pi_i \frac{d_i^{c_i}\Gamma(n_\tau/2 + c_i)}{(2\pi)^{n/2}\Gamma(c_i)} \int_{\theta^2} \frac{1}{\theta^n} \cdot \left(d_i + \frac{S^2_{\tau:n}}{2\theta^2}\right)^{-(n_\tau/2+c_i)} \cdot exp\left\{-\frac{S^2_{1:\tau-1}}{2\theta^2}\right\} d\theta^2
\end{aligned}
$$

The Bayes' Factor will be:

$$
BF_{\tau,n_+} = \frac{\Gamma(a_p)}{b_n^{a_p}} \sum_{i=1}^{2} \frac{\pi_i d_i^{c_i}\Gamma(n_\tau/2 + c_i)}{\Gamma(c_i)} \int_{\theta^2} \left(\frac{1}{\theta^2}\right)^{a_p+1} \cdot \left(d_i + \frac{S^2_{\tau:n}}{2\theta^2}\right)^{-(n_\tau/2+c_i)} \cdot exp\left\{-\frac{b_{\tau-1}}{\theta^2}\right\} d\theta^2
$$

Regarding the full conditional posteriors we have:

- $\theta^2|(\tau > n, \boldsymbol{x})$

$$
\begin{aligned}
p(\theta^2|\tau > n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta^2, \tau > n)\pi(\theta) \\
&\propto \left(\frac{1}{\theta^2}\right)^{a_p+1} \cdot exp\left\{-\frac{b_n}{\theta^2}\right\}
\end{aligned}
$$

Thus $\theta^2|(\tau > n, \boldsymbol{x}) \sim IG\,(a_p, b_n)$

- $\theta^2|(\kappa, \tau \leq n, \boldsymbol{x}, \boldsymbol{x})$

$$\begin{aligned} p(\theta^2|\kappa, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta^2, \kappa, \tau \leq n)\pi(\theta^2) \\ &\propto \left(\frac{1}{\theta^2}\right)^{a_p+1} \cdot exp\left\{-\frac{b_{\tau-1} + \frac{S^2_{\tau:n}}{2\theta^2}}{\theta^2}\right\} \end{aligned}$$

Thus $\theta^2|(\kappa, \tau \leq n, \boldsymbol{x}) \sim IG\left(a_p, b_{\tau-1} + \frac{S^2_{\tau:n}}{2\kappa}\right)$

- $\kappa_i|\left(\theta^2, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} p(\kappa|\theta^2, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\kappa, \theta^2, \tau \leq n)\pi(\theta) \\ &\propto \left(\frac{1}{\kappa_i}\right)^{n_\tau/2+c_i} \cdot exp\left\{-\frac{d_i + \frac{S^2_{\tau:n}}{2\theta^2}}{\kappa_i}\right\} \end{aligned}$$

Thus $\kappa_i|\left(\theta^2, \tau \leq n, \boldsymbol{x}\right) \sim IG\left(\frac{n_\tau}{2} + c_i, d_i + \frac{S^2_{\tau:n}}{2\theta^2}\right)$

- $\gamma|\left(\theta^2, \kappa_i, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} \pi^* &= \frac{\pi \cdot \int_{\kappa_1} f\left(\boldsymbol{x}|\theta^2, \kappa_1, \tau \leq n\right)\pi(\kappa_1)d\kappa_1}{\pi \cdot \int_{\kappa_1} f\left(\boldsymbol{x}|\theta^2, \kappa_1, \tau \leq n\right)\pi(\kappa_1)d\kappa_1 + (1-\pi) \cdot \int_{\kappa_2} f\left(\boldsymbol{x}|\theta^2, \kappa_2, \tau \leq n\right)\pi(\kappa_2)d\kappa_2} \\ &= \frac{\pi}{\pi + (1-\pi) \cdot \frac{\Gamma\left(c_{p2}\right)}{\Gamma\left(c_{p1}\right)} \cdot d_{p1}^{c_{p1}} \cdot d_{p2}^{-c_{p2}}} \end{aligned}$$

Thus $\gamma|\left(\theta^2, \kappa_i, \tau \leq n, \boldsymbol{x}\right) \sim Ber\left(\frac{\pi}{\pi + (1-\pi) \cdot \frac{\Gamma\left(c_{p2}\right)}{\Gamma\left(c_{p1}\right)} \cdot d_{p1}^{c_{p1}} \cdot d_{p2}^{-c_{p2}}}\right)$

- $\tau|\left(\theta^2, \kappa, \boldsymbol{x}\right)$

$$p\left(\tau=k|\theta^2,\kappa,\boldsymbol{x}\right) = \frac{f\left(\boldsymbol{x}|\theta^2,\kappa,\tau\right)\pi(\tau=k)}{\sum_{j=1}^{n} f\left(\boldsymbol{x}|\theta^2,\kappa,\tau\right)\pi(\tau=j)}$$

$$= \frac{\left(\frac{1}{\kappa}\right)^{\frac{n_k}{2}} exp\left\{-\frac{S^2_{1:(k-1)}}{2\theta^2}-\frac{S^2_{k:n}}{2\kappa\theta^2}\right\}\left((1-p)^{(k-1)^\beta}-(1-p)^{k^\beta}\right)}{\sum_{j=1}^{n}\left(\frac{1}{\kappa}\right)^{\frac{n_j}{2}} exp\left\{-\frac{S^2_{1:(j-1)}}{2\theta^2}-\frac{S^2_{j:n}}{2\kappa\theta^2}\right\}\left((1-p)^{(j-1)^\beta}-(1-p)^{j^\beta}\right)}$$

C3: Normal - Normal Inverse Gamma Model for the mean $\boldsymbol{\theta_1}$

In this Subsection, we provide the U3S process, assuming that the distribution which represents the IC state is a Normal distribution with both the mean θ_1 and the variance θ_2^2 unknown. We want to investigate a potential shift for the mean, either upwards or downwards. More specifically, $X_i|\theta \sim N\left(\theta_1,\theta_2^2\right)$, $i=1,...,n$ are sequentially arrived observations. Now, the IC parameters are two, assuming $(\theta_1,\theta_2^2) \sim NIG\left(\mu_0,\lambda,a,b\right)$. For the OOC shift, we set $\boldsymbol{\phi}=\delta$, which represents the magnitude of a shift and we replace the mean with $g(\boldsymbol{\theta},\boldsymbol{\phi})=\theta_1+\delta\cdot\theta_2$ as the link function which reflects the OOC state. Regarding δ we select a mixture of of Normal distributions. For the mixture, we assume $\delta=\gamma\cdot\delta_1+(1-\gamma)\cdot\delta_2$, where $\delta_i \sim N\left(\mu_{\delta i},{\sigma_{\delta i}}^2\right)$ and $\gamma \sim Ber(\pi)$. Finally, for the location of the change point, we assume $\tau \sim DW(p,\beta)$. The likelihood is:

$$f\left(\boldsymbol{x}|\theta_1,\theta_2^2,\delta,\tau\right)=\begin{cases}\prod_{i=1}^{\tau-1} f\left(x_i|\theta_1,\theta_2^2\right)\prod_{i=\tau}^{n} f\left(x_i|\theta_1+\delta\cdot\theta_2,\theta_2^2\right) & \text{if } \tau\le n\\ \prod_{i=1}^{n} f\left(x_i|\theta_1,\theta_2^2\right) & \text{if } \tau>n\end{cases}$$

Setting $n_t=n-t+1$, $X_{t_1:t_2}=\sum_{i=t_1}^{t_2}x_i$, $S^2_{t_1:t_2}=\sum_{i=t_1}^{t_2}(x_i-\theta_1)^2$, $\mu_p=\frac{\lambda\mu_0+X_{1:n}}{\lambda+n}$, $\lambda_p=\lambda+n$, $a_p=a+\frac{n}{2}$, $b_p=b+\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{\lambda n}{\lambda_p}(\bar{x}-\mu_0)^2\right)$, $e_{\delta i}=n_\tau\left(1-\frac{n_\tau}{\lambda_p}\right)+\frac{1}{\sigma^2_{\delta i}}$, $\pi_1=\pi$ and $\pi_2=1-\pi$ then the IC and the OOC marginal will be respectively:

$$
\begin{aligned}
f(\boldsymbol{x}|\tau > n) &= \int_{\theta_2^2}\int_{\theta_1} f(\boldsymbol{x}|\theta_1,\theta_2^2,\tau>n)\pi(\theta_1|\theta_2^2)\pi(\theta_2^2)d\theta_1 d\theta_2^2 \\
&= \int_{\theta_2^2}\int_{\theta_1} \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \cdot exp\left\{-\frac{S_{1:n}^2}{2\theta_2^2}\right\}\cdot\frac{\sqrt{\lambda}}{\theta_2\sqrt{2\pi}} \\
&\quad \times exp\left\{-\frac{\lambda(\theta_1-\mu_0)^2}{2\theta_2^2}\right\} d\theta_1\pi(\theta_2^2)d\theta_2^2 \\
&= \int_{\theta_2^2} \sqrt{\lambda}\cdot\left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^{n+1}\cdot exp\left\{-\frac{\sum_{i=1}^n x_i^2+\lambda\mu_0^2}{2\theta_2^2}\right\}\times \\
&\quad \times\int_{\theta_1} exp\left\{-\frac{-2(X_{1:n}+\lambda\mu_0)\theta_1+\lambda_p\theta_1^2}{2\theta_2^2}\right\} d\theta_1\pi(\theta_2^2)d\theta_2^2 \\
&= \int_{\theta_2^2}\frac{\sqrt{\lambda}}{\sqrt{\lambda_p}}\cdot\left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n\cdot exp\left\{-\frac{\sum_{i=1}^n x_i^2+\lambda\mu_0^2-\mu_p^2}{2\theta_2^2}\right\} \\
&\quad \times\frac{b^a}{\Gamma(a)}\cdot\left(\frac{1}{\theta_2^2}\right)^{a+1}\cdot exp\left\{-\frac{b}{\theta_2^2}\right\}d\theta_2^2 \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n\cdot\frac{\sqrt{\lambda}}{\sqrt{\lambda_p}}\cdot\frac{b^a}{\Gamma(a)}\cdot exp\left\{-\frac{\sum_{i=1}^n x_i^2+\lambda\mu_0^2-\mu_p^2}{2\theta_2^2}\right\} \\
&\quad \times\int_{\theta_p^2}\left(\frac{1}{\theta^2}\right)^{a_p+1}\cdot exp\left\{-\frac{b_p}{\theta_2^2}\right\}d\theta^2
\end{aligned}
$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\sqrt{\lambda}}{\sqrt{\lambda_p}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a_p)}{b_n^{a_p}} \cdot exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 + \lambda\mu_0^2 - \mu_p^2}{2\theta_2^2} \right\}$$

$$\begin{aligned}
f(\boldsymbol{x}|\tau \leq n) &= \sum_{i=1}^{2} \pi_i \int_{\delta_i} \int_{\theta_2^2} \int_{\theta_1} f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta_i, \tau \leq n)\pi(\theta_1|\theta_2^2)\pi(\theta_2^2)\pi(\delta_i) d\theta_1 d\theta_2^2 d\delta_i \\
&= \sum_{i=1}^{2} \pi_i \int_{\delta_i} \int_{\theta_2^2} \int_{\theta_1} \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \cdot exp\left\{ -\frac{S_{1:\tau-1}^2 + \sum_{i=\tau}^{n}(x_i - (\theta_1 + \delta_i\theta_2))^2}{2\theta_2^2} \right\} \times \\
&\quad \times \frac{\sqrt{\lambda}}{\theta_2\sqrt{2\pi}} \cdot exp\left\{ -\frac{\lambda(\theta_1 - \mu_0)^2}{2\theta_2^2} \right\} d\theta_1 \pi(\theta_2^2) d\theta_2^2 \pi(\delta_i) d\delta_i \\
&= \sum_{i=1}^{2} \pi_i \int_{\delta_i} \int_{\theta_2^2} \sqrt{\lambda} \cdot \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^{n+1} \cdot exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 + \lambda\mu_0^2 - 2X_{\tau:n}\delta_i\theta_2}{2\theta_2^2} - n_\tau\delta_i^2 \right\} \times \\
&\quad \times \int_{\theta_1} exp\left\{ -\frac{-2(X_{1:n} + \lambda\mu_0 - n_\tau\delta_i\theta_2)\theta_1 + \lambda_p\theta_1^2}{2\theta_2^2} \right\} d\theta_1 \pi(\theta_2^2) d\theta_2^2 \pi(\delta_i) d\delta_i \\
&= \sum_{i=1}^{2} \pi_i \int_{\theta_2^2} \int_{\delta_i} \frac{\sqrt{\lambda}}{\sqrt{\lambda_p}} \cdot \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \cdot exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 + \lambda\mu_0^2 - 2X_{\tau:n}\delta_i\theta_2}{2\theta_2^2} - n_\tau\delta_i^2 \right\} \times \\
&\quad \times exp\left\{ +\frac{\lambda_p\left(\mu_p - n_\tau\delta_i\theta_2/\lambda_p\right)^2}{2\theta_2^2} \right\} \cdot exp\left\{ -\frac{(\delta_i - \mu_{\delta_i})^2}{2\sigma_{\delta_i}^2} \right\} d\delta_i \pi(\theta_2^2) d\theta_2^2 \\
&= \sum_{i=1}^{2} \pi_i \int_{\theta_2^2} \frac{\sqrt{\lambda}}{\sqrt{\lambda_p}} \cdot \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \cdot \frac{1}{\sigma_{\delta_i}\sqrt{2\pi}} \cdot exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 + \lambda\mu_0^2 - \lambda_p\mu_p^2}{2\theta_2^2} - \frac{\mu_{\delta_i}^2}{2\sigma_{\delta_i}^2} \right\} \times
\end{aligned}$$

$$\times \int_{\delta_i} exp\left\{-\frac{-2\left(\sigma^2_{\delta_i}(X_{\tau:n}-n_\tau\mu_p)/\theta_2+\mu_{\delta_i}\right)\delta_i+\left(\sigma^2_{\delta_i}(n_\tau-n^2_\tau/\lambda_p)+1\right)\delta_i^2}{2\sigma^2_{\delta_i}}\right\} d\delta_i \pi(\theta_2^2) d\theta_2^2$$

$$\begin{aligned}
&= \sum_{i=1}^{2}\pi_i \int_{\theta_2^2}\frac{\sqrt{\lambda}}{\sqrt{\lambda_p}}\cdot\left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n\cdot\frac{1}{\sigma_{\delta_i}\sqrt{e_{\delta_i}}}\cdot exp\left\{-\frac{\sum_{i=1}^{n}x_i^2+\lambda\mu_0^2-\lambda_p\mu_p^2}{2\theta_2^2}-\frac{\mu^2_{\delta_i}}{2\sigma^2_{\delta_i}}\right\}\times\\
&\quad\times exp\left\{+\frac{\left(\frac{X_{\tau:n}-n_\tau\mu_p}{\theta_2}+\frac{\mu_{\delta i}}{\sigma^2_{\delta i}}\right)^2}{2e_{\delta i}}\right\}\cdot\frac{b^a}{\Gamma(a)}\cdot\left(\frac{1}{\theta_2^2}\right)^{a+1}\cdot exp\left\{-\frac{b}{\theta_2^2}\right\}d\theta_2^2\\
&= \sum_{i=1}^{2}\pi_i\frac{\sqrt{\lambda}}{\sqrt{\lambda_p}}\cdot\left(\frac{1}{\sqrt{2\pi}}\right)^n\cdot\frac{1}{\sigma_{\delta_i}\sqrt{e_{\delta_i}}}\cdot\frac{b^a}{\Gamma(a)}\cdot exp\left\{-\frac{\mu^2_{\delta_i}}{2\sigma^2_{\delta_i}}\right\}\times\\
&\quad\times\int_{\theta_2^2}\left(\frac{1}{\theta_2^2}\right)^{a_p+1}\cdot exp\left\{-\frac{b_p}{\theta_2^2}+\frac{\left(\frac{X_{\tau:n}-n_\tau\mu_p}{\theta_2}+\frac{\mu_{\delta i}}{\sigma^2_{\delta i}}\right)^2}{2e_{\delta i}}\right\}d\theta_2^2
\end{aligned}$$

The Bayes' Factor will be:

$$BF_{\tau,n_+}=\frac{b_p^{a_p}}{\Gamma(a_p)}\sum_{i=1}^{2}\frac{\pi_i}{\sigma_{\delta i}\sqrt{e_{\delta i}}}\cdot exp\left\{-\frac{\mu^2_{\delta i}}{2\sigma^2_{\delta i}}\right\}\int_{\theta_2^2}\left(\frac{1}{\theta_2^2}\right)^{a_p+1}\cdot exp\left\{-\frac{b_p}{\theta_2^2}+\frac{\left(\frac{X_{\tau:n}-n_\tau\mu_p}{\theta_2}+\frac{\mu_{\delta i}}{\sigma^2_{\delta i}}\right)^2}{2e_{\delta i}}\right\}d\theta_2^2$$

The full conditional posteriors will be:

- $(\theta_1,\theta_2^2)|(\tau>n,\boldsymbol{x})$

$$\begin{aligned}
p(\theta_1,\theta_2^2|\tau>n,\boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1,\theta_2^2,\tau>n)\pi(\theta_1,\theta_2^2)\\
&\propto \left(\frac{1}{\theta_2^2}\right)^{a_p+3/2}\cdot exp\left\{-\frac{2b_p+\lambda_p(\theta_1-\mu_p)^2}{2\theta_2^2}\right\}
\end{aligned}$$

Thus $(\theta_1,\theta_2^2)|(\tau>n,\boldsymbol{x})\sim NIG\left(\mu_p,\lambda_p,a_p,b_p\right)$

- $\theta_1|(\theta_2^2, \delta, \tau \leq n, \boldsymbol{x})$

$$\begin{aligned} p(\theta_1|\theta_2^2, \delta, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta, \tau \leq n)\pi(\theta_1|\theta_2^2) \\ &\propto exp\left\{-\frac{-2(X_{1:n} + \lambda_0\mu_0 - n_\tau\delta\theta_2)\theta_1 + \lambda_p\theta_1^2}{2\theta_2^2}\right\} \end{aligned}$$

Thus $\theta_1|(\theta_2^2, \delta, \tau \leq n, \boldsymbol{x}) \sim N\left(\mu_p - \frac{n_\tau\delta\theta_2}{\lambda_p}, \frac{\theta_2^2}{\lambda_p}\right)$

- $\theta_2^2|(\theta_1, \delta, \tau \leq n, \boldsymbol{x})$

$$\begin{aligned} p\left(\theta_2^2|\theta_1, \delta, \tau \leq n, \boldsymbol{x}\right) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta, \tau \leq n)\pi(\theta_2^2) \\ &\propto \left(\frac{1}{\theta_2^2}\right)^{\alpha_p+3/2} exp\left\{-\frac{2b + S_{1:n}^2 + \lambda\left(\theta_1 - \mu_0\right)^2}{2\theta_2^2} - \frac{\left(X_{\tau:n} - n_\tau\theta_1\right)\delta}{\theta_2}\right\} \end{aligned}$$

- $\delta_i|\left(\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} p(\delta_i|\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta_i, \tau \leq n)\pi(\delta_i) \\ &\propto exp\left\{-\frac{\left(\mu_{\delta i} + \sigma_{\delta i}^2\left(X_{\tau:n} - n_\tau\theta_1\right)/\theta_2\right)\delta_i + \left(1 + n_\tau\sigma_{\delta i}^2\right)\delta_i^2}{2\sigma_{\delta i}^2}\right\} \end{aligned}$$

Thus $\delta_i|\left(\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x}\right) \sim N\left(\frac{\mu_{\delta i} + \sigma_{\delta i}^2\left(X_{\tau:n} - n_\tau\theta_1\right)/\theta_2}{1 + n_\tau\sigma_{\delta i}^2}, \frac{\sigma_{\delta i}^2}{1 + n_\tau\sigma_{\delta i}^2}\right)$

- $\gamma|\left(\theta_1, \theta_2^2, \delta_i, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} \pi^* &= \frac{\pi \cdot \int_{\delta_1} f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta_1, \tau \leq n)\pi(\delta_1)d\delta_1}{\pi \cdot \int_{\delta_1} f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta_1, \tau \leq n)\pi(\delta_1)d\delta_1 + (1-\pi) \cdot \int_{\delta_2} f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta_2, \tau \leq n)\pi(\delta_2)d\delta_2} \\ &= \frac{\pi}{\pi + (1-\pi) \cdot exp\left\{\frac{\mu_{\delta p2}^2}{2\sigma_{\delta p2}^2} - \frac{\mu_{\delta p1}^2}{2\sigma_{\delta p1}^2}\right\}\frac{\sigma_{\delta p2}}{\sigma_{\delta p1}}} \end{aligned}$$

Thus $\gamma|\left(\theta_1, \theta_2^2, \delta_i, \tau \leq n, \boldsymbol{x}\right) \sim Ber\left(\dfrac{\pi}{\pi + (1-\pi)\cdot exp\left\{\dfrac{\mu_{\delta p2}^2}{2\sigma_{\delta p2}^2} - \dfrac{\mu_{\delta p1}^2}{2\sigma_{\delta p1}^2}\right\}\dfrac{\sigma_{\delta p2}}{\sigma_{\delta p1}}}\right)$

- $\tau|\left(\theta_1, \theta_2^2, \delta, \boldsymbol{x}\right)$

$$
\begin{aligned}
p\left(\tau = k|\theta_1, \theta_2^2, \delta, \boldsymbol{x}\right) &= \frac{f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta, \tau \leq n)\pi(\tau = k)}{\sum_{j=1}^{n} f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta, \tau \leq n)\pi(\tau = j)} \\
&= \frac{exp\left\{\dfrac{\delta\left(X_{k:n} - n_k\theta_1\right)}{\theta_2} - \dfrac{n_k\delta^2}{2}\right\}\left((1-p)^{(k-1)^\beta} - (1-p)^{k^\beta}\right)}{\sum_{j=1}^{n} exp\left\{\dfrac{\delta\left(X_{j:n} - n_j\theta_1\right)}{\theta_2} - \dfrac{n_j\delta^2}{2}\right\}\left((1-p)^{(j-1)^\beta} - (1-p)^{j^\beta}\right)}
\end{aligned}
$$

C4: Normal - Normal Inverse Gamma Model for the variance $\boldsymbol{\theta_2^2}$

Assume that the distribution of the IC state is a Normal distribution with both the mean θ_1 and the variance θ_2^2 to be unknown and we wish to investigate a level change for the variance. The sequentially arrived observations are $X_i|\theta \sim N\left(\theta, \sigma^2\right)$, $i = 1, ..., n$. Regarding the IC parameters, we assume $\left(\theta_1, \theta_2^2\right) \sim NIG\left(\mu_0, \lambda, a, b\right)$. For the OOC shift, we set $\boldsymbol{\phi} = \kappa$, which represents the magnitude of a shift and for the link function, we define $g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \kappa \cdot \theta_2^2$, replacin gthe variance. The prior for κ is a mixture of Inverse Gammma distributions, for a potential inflation and shrinkage of the variance respectively. Specifically, $\kappa = \gamma \cdot \kappa_1 + (1-\gamma)\cdot\kappa_2$, where $\kappa_i \sim IG\left(c_i, d_i\right)$ and $\gamma \sim Ber(\pi)$. For the first OOC observation we assume $\tau \sim DW(p, \beta)$, while the likelihood is given by:

$$
f\left(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa, \tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f\left(x_i|\theta_1, \theta_2^2\right)\prod_{i=\tau}^{n} f\left(x_i|\theta_1, \kappa\cdot\theta_2^2\right) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f\left(x_i|\theta_1, \theta_2^2\right) & \text{if } \tau > n \end{cases}
$$

Setting $n_t = n-t+1$, $n_{t,k} = t-1+\frac{n_t}{k}$, $X_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i$, $X^2_{t_1:t_2} = \sum_{i=t_1}^{t_2} x_i^2$, $S^2_{t_1:t_2} = \sum_{i=t_1}^{t_2}(x_i-\theta_1)^2$, $\mu_p = \frac{\lambda\mu_0 + X_{1:n}}{\lambda+n}$, $\lambda_p = \lambda+n$, $a_p = a+\frac{n}{2}$, $b_p = b+\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + \frac{\lambda n}{\lambda_p}(\bar{x}-\mu_0)^2\right)$, $b_{t,k} = b+\frac{1}{2}\left(\lambda\mu_0^2 + X^2_{1:(t-1)} + \frac{X^2_{t:n}}{k} - \frac{\left(\lambda\mu_0 + X_{1:(t-1)} + \frac{X_{t:n}}{k}\right)^2}{\lambda+n_{t,k}}\right)$, $\pi_1 = \pi$ and $\pi_2 = 1-\pi$ then the IC and the OOC marginal will be respectively:

$$
\begin{aligned}
f(\boldsymbol{x}|\tau > n) &= \int_{\theta_2^2}\int_{\theta_1} f(\boldsymbol{x}|\theta_1,\theta_2^2,\tau>n)\pi(\theta_1|\theta_2^2)\pi(\theta_2^2)d\theta_1 d\theta_2^2 \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\sqrt{\lambda}}{\sqrt{\lambda_p}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a_p)}{b_n^{a_p}} \cdot exp\left\{-\frac{\sum_{i=1}^{n} x_i^2 + \lambda\mu_0^2 - \mu_p^2}{2\theta_2^2}\right\}
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{x}|\tau \le n) &= \sum_{i=1}^{2}\pi_i \int_{\kappa_i}\int_{\theta_2^2}\int_{\theta_1} f(\boldsymbol{x}|\theta_1,\theta_2^2,\delta_i,\tau\le n)\pi(\theta_1|\theta_2^2)\pi(\theta_2^2)\pi(\kappa_i)d\theta_1 d\theta_2^2 d\kappa_i \\
&= \sum_{i=1}^{2}\pi_i \int_{\kappa_i}\int_{\theta_2^2}\int_{\theta_1} \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot exp\left\{-\frac{S^2_{1:\tau-1}}{2\theta_2^2} - \frac{S^2_{\tau:n}}{2\kappa_i\theta_2^2}\right\} \times \\
&\quad \times \frac{\sqrt{\lambda}}{\theta_2\sqrt{2\pi}} \cdot exp\left\{-\frac{\lambda(\theta_1-\mu_0)^2}{2\theta_2^2}\right\} d\theta_1\pi(\theta_2^2)d\theta_2^2\pi(\kappa_i)d\kappa_i \\
&= \sum_{i=1}^{2}\pi_i \int_{\kappa_i}\int_{\theta_2^2} \sqrt{\lambda}\left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^{n+1} \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot exp\left\{-\frac{X^2_{1:\tau-1} + X^2_{\tau:n}/\kappa_i + \lambda\mu_0}{2\theta_2^2}\right\} \times \\
&\quad \times \int_{\theta_1} exp\left\{-\frac{-2\left(\lambda\mu_0 + X_{1:\tau-1} + X_{\tau:n}/\kappa_i\right)\theta_1 + \left(\lambda + n_{\tau,\kappa_i}\right)\theta_1^2}{2\theta_2^2}\right\} d\theta_1\pi(\theta_2^2)d\theta_2^2\pi(\kappa_i)d\kappa_i
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{2} \pi_i \int_{\kappa_i} \int_{\theta_2^2} \frac{\sqrt{\lambda}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} \left(\frac{1}{\theta_2\sqrt{2\pi}}\right)^n \times \\
&\quad \times \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot exp\left\{-\frac{X_{1:\tau-1}^2 + X_{\tau:n}^2/\kappa_i + \lambda\mu_0}{2\theta_2^2}\right\} \times \\
&\quad \times exp\left\{+\frac{(\lambda\mu_0 + X_{1:\tau-1} + X_{\tau:n}/\kappa_i)^2}{2\left(\lambda + n_{\tau,\kappa_i}\right)\theta_2^2}\right\} \cdot \frac{b^a}{\Gamma(a)} \cdot \left(\frac{1}{\theta_2^2}\right)^{a+1} \cdot exp\left\{-\frac{b}{\theta_2^2}\right\} d\theta_2^2 \pi(\kappa_i) d\kappa_i \\
&= \sum_{i=1}^{2} \pi_i \int_{\kappa_i} \frac{\sqrt{\lambda}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot \frac{b^a}{\Gamma(a)} \times \\
&\quad \times \int_{\theta_2^2} \left(\frac{1}{\theta_2^2}\right)^{a_p+1} \cdot exp\left\{-\frac{b_{\tau,\kappa_i}}{\theta_2^2}\right\} d\theta_2^2 \pi(\kappa_i) d\kappa_i \\
&= \sum_{i=1}^{2} \pi_i \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\Gamma(a_p)}{\Gamma(a)} \cdot \frac{\sqrt{\lambda}}{b^{-a}} \times \\
&\quad \times \int_{\kappa_i} \frac{b_{\tau,\kappa_i}^{-a_p}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2} \cdot \frac{d_i^{c_i}}{\Gamma(c_i)} \cdot \left(\frac{1}{\kappa_i}\right)^{c_i+1} \cdot exp\left\{-\frac{d_i}{\kappa_i}\right\} d\kappa_i \\
&= \sum_{i=1}^{2} \pi_i \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{\Gamma(a_p)}{\Gamma(a)} \cdot \frac{\sqrt{\lambda}}{b^{-a}} \cdot \frac{d_i^{c_i}}{\Gamma(c_i)} \times \\
&\quad \times \int_{\kappa_i} \frac{b_{\tau,\kappa_i}^{-a_p}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} \cdot \left(\frac{1}{\kappa_i}\right)^{n_\tau/2+c_i+1} \cdot exp\left\{-\frac{d_i}{\kappa_i}\right\} d\kappa_i
\end{aligned}
$$

The Bayes' Factor will result:

$$
BF_{\tau,n_+} = b_p^{a_p}\sqrt{\lambda_p} \sum_{i=1}^{2} \frac{\pi_i d_i^{c_i}}{\Gamma(c_i)} \int_{\kappa_i} \frac{b_{\tau,\kappa_i}^{-a_p} \cdot \kappa_i^{-(n_\tau/2+c_i+1)}}{\sqrt{\lambda + n_{\tau,\kappa_i}}} exp\left\{-\frac{d_i}{\kappa_i}\right\} d\kappa_i
$$

The full conditional posteriors will be:

- $(\theta_1, \theta_2^2)|(\tau > n, \boldsymbol{x})$

$$\begin{aligned} p(\theta_1, \theta_2^2 | \tau > n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \tau > n)\pi(\theta_1, \theta_2^2) \\ &\propto \left(\frac{1}{\theta_2^2}\right)^{a_p+3/2} \cdot exp\left\{-\frac{2b_p + \lambda_p(\theta_1 - \mu_p)^2}{2\theta_2^2}\right\} \end{aligned}$$

Thus $(\theta_1, \theta_2^2)|(\tau > n, \boldsymbol{x}) \sim NIG\left(\mu_p, \lambda_p, a_p, b_p\right)$

- $\theta_1|(\theta_2^2, \delta, \tau \leq n, \boldsymbol{x})$

$$\begin{aligned} p(\theta_1|\theta_2^2, \kappa, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \delta, \tau \leq n)\pi(\theta_1|\theta_2^2) \\ &\propto exp\left\{-\frac{-2(\lambda\mu_0 + X_{1:(\tau-1)} + X_{\tau:n}/\kappa)\theta_1 + (\lambda + n_{\tau,\kappa})\,\theta_1^2}{2\theta_2^2}\right\} \end{aligned}$$

Thus $\theta_1|(\theta_2^2, \kappa, \tau, \boldsymbol{x}) \sim N\left(\frac{\lambda\mu_0 + X_{1:(\tau-1)} + X_{\tau:n}/\kappa}{\lambda + n_{\tau,\kappa}}, \frac{\theta_2^2}{\lambda + n_{\tau,\kappa}}\right)$

- $\theta_2^2|(\theta_1, \kappa, \tau \leq n, \boldsymbol{x})$

$$\begin{aligned} p\left(\theta_2^2|\theta_1, \kappa, \tau \leq n, \boldsymbol{x}\right) &\propto f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa, \tau \leq n)\pi(\theta_2^2) \\ &\propto \left(\frac{1}{\theta_2^2}\right)^{a_p+1} exp\left\{-\frac{b_{\tau,\kappa_i}}{\theta_2^2}\right\} \end{aligned}$$

Thus $\theta_2^2|\left(\theta_1, \kappa, \tau, \boldsymbol{x}\right) \sim IG\left(a_p, b_{\tau,\kappa_i}\right)$

- $\kappa_i|\left(\theta^2, \tau \leq n, \boldsymbol{x}\right)$

$$\begin{aligned} p(\kappa_i|\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta_1, \theta^2, \kappa_i, \tau \leq n)\pi(\kappa_i) \\ &\propto \left(\frac{1}{\kappa_i}\right)^{n_\tau/2+c_i} \cdot exp\left\{-\frac{d_i + \frac{S_{\tau:n}^2}{2\theta^2}}{\kappa_i}\right\} \end{aligned}$$

Thus $\kappa_i|\left(\theta_1, \theta_2^2, \tau \leq n, \boldsymbol{x}\right) \sim IG\left(c_i + \frac{n_\tau}{2}, d_i + \frac{S_{\tau:n}^2}{2\theta^2}\right)$

- $\gamma|\left(\theta_1, \theta_2^2, \kappa_i, \tau \leq n, \boldsymbol{x}\right)$

$$\pi^* = \frac{\pi \cdot \int_{\kappa_1} f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa_1, \tau \leq n)\pi(\kappa_1)d\kappa_1}{\pi \cdot \int_{\kappa_1} f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa_1, \tau \leq n)\pi(\kappa_1)d\kappa_1 + (1-\pi) \cdot \int_{\kappa_2} f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa_2, \tau \leq n)\pi(\kappa_2)d\kappa_2}$$

$$= \frac{\pi}{\pi + (1-\pi) \cdot \frac{\Gamma(c_{p2})}{\Gamma(c_{p1})} \cdot d_{p1}^{c_{p1}} \cdot d_{p2}^{-c_{p2}}}$$

Thus $\gamma|\left(\theta_1, \theta_2^2, \kappa_i, \tau \leq n, \boldsymbol{x}\right) \sim Ber\left(\frac{\pi}{\pi + (1-\pi) \cdot \frac{\Gamma(c_{p2})}{\Gamma(c_{p1})} \cdot d_{p1}^{c_{p1}} \cdot d_{p2}^{-c_{p2}}}\right)$

- $\tau|\left(\theta_1, \theta_2^2, \kappa, \boldsymbol{x}\right)$

$$p\left(\tau = k|\theta_1, \theta_2^2, \kappa, \boldsymbol{x}\right) = \frac{f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa, \tau \leq n)\pi(\tau = k)}{\sum_{j=1}^{n} f(\boldsymbol{x}|\theta_1, \theta_2^2, \kappa, \tau \leq n)\pi(\tau = j)}$$

$$= \frac{\left(\frac{1}{\kappa}\right)^{\frac{n_k}{2}} exp\left\{-\frac{S_{1:(k-1)}^2}{2\theta_2^2} - \frac{S_{k:n}^2}{2\kappa\theta_2^2}\right\}\left((1-p)^{(k-1)^\beta} - (1-p)^{k^\beta}\right)}{\sum_{j=1}^{n}\left(\frac{1}{\kappa}\right)^{\frac{n_j}{2}} exp\left\{-\frac{S_{1:(j-1)}^2}{2\theta_2^2} - \frac{S_{j:n}^2}{2\kappa\theta_2^2}\right\}\left((1-p)^{(j-1)^\beta} - (1-p)^{j^\beta}\right)}$$

Appendix D: List of M3S Models

D1: Normal Model for the mean vector $\boldsymbol{\mu}$ in D dimensions

In this Subsection, we provide the M3S model for detecting shifts in the mean vector in D dimensions, obtaining the data sequentially. We assume that the distribution of the IC state is a Normal with unknown the mean vector and the covariance matrix, i.e. $X_i|\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right) \sim N_D\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$, $i = 1, ..., n$. For the IC parameters $\boldsymbol{\theta}$, we assume $\boldsymbol{\theta} = \left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right) \sim NIW\left(\boldsymbol{\mu_o}, \lambda, \nu_0, \boldsymbol{\Psi}\right)$. In case of available historical IC data $\boldsymbol{Y}$, we can use the power prior (Ibrahim and Chen, 2000), assuming the NIW to be the initial term. Then, $\pi\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right) \propto L\left(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{Y}\right)^{\alpha_0}\pi_0\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$, where $0 \leq \alpha_0 \leq 1$ fixed and $\pi_0\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = NIW$ the initial prior. For the OOC shift, we set $\boldsymbol{\phi} = \boldsymbol{\delta}$, where $\boldsymbol{\delta} = \left(\delta_1\ \delta_2\ ...\ \delta_D\right)^T$ is the vector with the magnitude of the shift for each component of the mean vector. For the shifted mean of the OOC state, we replace $\boldsymbol{\mu}$ with $g(\boldsymbol{\theta},\boldsymbol{\phi}) = \boldsymbol{\mu} + \boldsymbol{L}^{1/2}\boldsymbol{\delta}$, where $\boldsymbol{L} = diag(\boldsymbol{\Sigma})$, i.e. a diagonal matrix with the same main diagonal with the covariance matrix. Multiplying by $\boldsymbol{L}^{1/2}$, we achieve the desired property of the anisotropic scaling, or in other words the jumps are weighted by the variance of the corresponding component. Regarding the prior distributions, we assume $\delta_i \sim N\left(\mu_{di}, {\sigma_d}^2\right)$ and for the change point $\tau \sim DW(p,\beta)$, as in the univariate models. The likelihood is:

$$f\left(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\delta},\tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f\left(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}\right)\prod_{i=\tau}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}+\boldsymbol{L}^{1/2}\boldsymbol{\delta},\boldsymbol{\Sigma}\right) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}\right) & \text{if } \tau > n \end{cases}$$

Although we achieve the anisotropic scaling with the above set, we do not achieve the directional invariance, i.e. $\boldsymbol{\delta}$ cannot translocate in any direction. For this reason, we transform the shift vector $\boldsymbol{\delta}$ into D-sphere coordinates. Thus, assume r be the radius and $\theta = (\theta_1, \theta_2, ..., \theta_{D-1})$ be the $D-1$ angular components, where $r > 0$, $(\theta_1, \theta_2, ..., \theta_{D-2}) \in [0,\pi)^{D-2}$ and $\theta_{D-1} \in [0, 2\pi)$. Note that $\boldsymbol{\theta}$ refer to the IC parameter, while nonbold θ is the vector of the angles. We apply the transformation $\boldsymbol{\delta} = r\boldsymbol{T}_\theta$, such that $r\boldsymbol{T}_\theta$ to belong to a $(D-1)$-sphere of radius r, or $r\boldsymbol{T}_\theta \in \mathbb{S}^{D-1}(r)$.

$\boldsymbol{T}_\theta$ is a unit vector, i.e. it is a vector of length 1 and specifically:

$$\boldsymbol{T}_\theta = \begin{pmatrix} cos\theta_1 \\ \vdots \\ cos\theta_{D-1} \prod_{i<D-1} sin\theta_i \\ \prod_{i=1}^{D-1} sin\theta_i \end{pmatrix}$$

if $D \geq 3$ or $\boldsymbol{T}_\theta = (cos\theta, sin\theta)$ if $D = 2$. After the transformation, the directional invariant likelihood will be:

$$f\left(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, r, \theta, \tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \prod_{i=\tau}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu} + r\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta, \boldsymbol{\Sigma}\right)|J| & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) & \text{if } \tau > n \end{cases}$$

where $|J| = r^{D-1}\prod_{j=0}^{D-2}[sin\theta_{j+1}]^{D-2-j}$ is the Jacobian determinant of the transformation. Note that the radius r denotes the size of the jumps, while the angular components θ represents the association between the jump for each variable of the process. Applying standard transformation properties of the Normal distribution, the radius r follows a two parameter Noncentral Chi distribution, or $r \sim NC_{\chi_D}(d, \sigma_d^2)$ with $d = \left(\sum_{i=1}^{D}\mu_{di}^2\right)^{1/2}$. The Rice distribution is a special case on $NC_{\chi_D}(d, \sigma_d^2)$, when $D = 2$. For the unit vector $\boldsymbol{T}_\theta$, we assume a von Mises-Fisher distribution, or $\boldsymbol{T}_\theta \sim vMF(\boldsymbol{\mu}_\theta, \kappa)$, which is the analogue of the multivariate Normal over the unit sphere. When $D = 2$, then it is von Mises distribution over the unit circle, while it reduces to the U_{D-1} (Uniform in $D-1$ dimensions), when the concentration parameter $\kappa = 0$. For the next expressions, we set $\boldsymbol{C_n} = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{\boldsymbol{n}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{\boldsymbol{n}})^T$, $\boldsymbol{C_o} = (\bar{\boldsymbol{x}}_{\boldsymbol{n}} - \boldsymbol{\mu_o})(\bar{\boldsymbol{x}}_{\boldsymbol{n}} - \boldsymbol{\mu_o})^T$, $n_t = n - t + 1$, $\boldsymbol{A_t} = n_t\boldsymbol{T}_\theta^T\boldsymbol{L}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta$, $\boldsymbol{B_t} = \sum_{i=t}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta$, $\boldsymbol{\mu_p} = \frac{n\bar{\boldsymbol{x}}_{\boldsymbol{n}} + \lambda\boldsymbol{\mu}_0}{n+\lambda}$, $\lambda_p = n + \lambda$, $\nu_p = n + \nu_0$ and $\boldsymbol{\Psi_p} = \boldsymbol{C} + \boldsymbol{\Psi} + \frac{\lambda n}{\lambda + n}\boldsymbol{C_o}$, then the IC and the

OOC marginal will be respectively:

$$
\begin{aligned}
f(\boldsymbol{x_n}|\tau > n) &= \int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},\tau>n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})d\boldsymbol{\mu}d\boldsymbol{\Sigma} \\
&= \int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} \left(\frac{1}{2\pi}\right)^{nD/2}\cdot|\boldsymbol{\Sigma}|^{-n/2}\cdot exp\left\{-\frac{1}{2}\left(tr\boldsymbol{C_n}\boldsymbol{\Sigma}^{-1}+n(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_{\boldsymbol{n}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_{\boldsymbol{n}})\right)\right\} \\
&\quad\times\left(\frac{\lambda}{2\pi}\right)^{D/2}\cdot|\boldsymbol{\Sigma}|^{-1/2}\cdot exp\left\{-\frac{\lambda}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_o})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_o})\right\}\pi(\boldsymbol{\Sigma})d\boldsymbol{\mu}d\boldsymbol{\Sigma} \\
&= \int_{\boldsymbol{\Sigma}}\left(\frac{1}{2\pi}\right)^{nD/2}\cdot\left(\frac{\lambda}{\lambda_p}\right)^{D/2}\cdot|\boldsymbol{\Sigma}|^{-n/2}\cdot exp\left\{-\frac{1}{2}\left(tr\boldsymbol{C_n}\boldsymbol{\Sigma}^{-1}+n\bar{\boldsymbol{x}}_{\boldsymbol{n}}^T\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}_{\boldsymbol{n}}\right)\right\} \\
&\quad\times exp\left\{-\frac{1}{2}\left(\lambda\boldsymbol{\mu_o}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu_o}-\lambda_p\boldsymbol{\mu_p}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu_p}\right)\right\} \\
&\quad\times\frac{|\boldsymbol{\Psi}|^{\nu_0/2}}{2^{\nu_0 D/2}\Gamma_D\left(\frac{\nu_0}{2}\right)}\cdot|\boldsymbol{\Sigma}|^{-(\nu_0+D+1)/2}\cdot exp\left\{-\frac{1}{2}tr\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right\}d\boldsymbol{\Sigma} \\
&= \frac{1}{\pi^{nD/2}}\cdot\left(\frac{\lambda}{\lambda_p}\right)^{D/2}\cdot\frac{\Gamma_D\left(\nu_p/2\right)}{\Gamma_D\left(\nu_0/2\right)}\cdot\frac{|\boldsymbol{\Psi}|^{\nu_0/2}}{|\boldsymbol{\Psi_p}|^{\nu_p/2}}
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{x_n}|\tau \le n) &= \int_r\int_\theta\int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\tau\le n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})\pi(r)\pi(\theta)d\boldsymbol{\mu}d\boldsymbol{\Sigma}d\theta dr \\
&= \int_r\int_\theta\int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} \frac{r^{D-1}\prod_{j=0}^{D-2}[sin\theta_{j+1}]^{D-2-j}}{(2\pi)^{nD/2}\cdot|\boldsymbol{\Sigma}|^{n/2}}\cdot exp\left\{-\frac{1}{2}\sum_{i=1}^{\tau-1}(\boldsymbol{x_i}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\right\} \\
&\quad\times exp\left\{-\frac{1}{2}\sum_{i=\tau}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu}-r\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu}-r\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta)\right\} \\
&\quad\times\left(\frac{\lambda}{2\pi}\right)^{D/2}\cdot|\boldsymbol{\Sigma}|^{-1/2}\cdot exp\left\{-\frac{\lambda}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_o})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_o})\right\}\pi(\boldsymbol{\Sigma})\pi(r)\pi(\theta)d\boldsymbol{\mu}d\boldsymbol{\Sigma}d\theta dr
\end{aligned}
$$

$$
\begin{aligned}
&= \int_r \int_\theta \int_{\boldsymbol{\Sigma}} \frac{r^{D-1} \prod_{j=0}^{D-2} [sin\theta_{j+1}]^{D-2-j}}{(2\pi)^{nD/2} \cdot |\boldsymbol{\Sigma}|^{n/2}} \cdot \left(\frac{\lambda}{\lambda_p}\right)^{D/2} \\
&\quad \times exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n} \boldsymbol{x_i}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x_i} + \lambda \boldsymbol{\mu_o}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_o} + \boldsymbol{A_\tau}\right) + \boldsymbol{B_\tau}\right\} \\
&\quad \times exp\left\{+\frac{\lambda_p}{2}\left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)\right\} \pi(\boldsymbol{\Sigma}) \pi(r) \pi(\theta) d\boldsymbol{\Sigma} d\theta dr \\
&= \frac{\kappa^{D/2-1} \cdot d^{1-D/2}}{\sigma_d^2 \cdot (2\pi)^{(nD+1)/2}} \cdot \left(\frac{\lambda}{\lambda_p}\right)^{D/2} \cdot \frac{|\boldsymbol{\Psi}|^{\nu_0/2}}{2^{\nu_0 D/2} \Gamma_D\left(\frac{\nu_0}{2}\right)} \cdot \frac{I_{D/2-1}\left(\frac{rd}{\sigma_d^2}\right)}{I_{D/2}(\kappa)} \cdot exp\left\{-\frac{d^2}{2}\right\} \\
&\quad \times \int_r \int_\theta \int_{\boldsymbol{\Sigma}} \frac{r^{3D/2-1} \prod_{j=0}^{D-2} [sin\theta_{j+1}]^{D-2-j}}{|\boldsymbol{\Sigma}|^{(\nu_p+D+1)/2}} \cdot exp\left\{-\frac{1}{2}\left(\boldsymbol{A_\tau} + r^2\right) + \boldsymbol{B_\tau} + \kappa \boldsymbol{\mu_\theta}^T \boldsymbol{T}_\theta\right\} \\
&\quad \times exp\left\{-\frac{1}{2} tr \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi} + \boldsymbol{x_i} \boldsymbol{x_i}^T + \lambda \boldsymbol{\mu_o} \boldsymbol{\mu_o}^T\right)\right\} \\
&\quad \times exp\left\{+\frac{\lambda_p}{2} tr \boldsymbol{\Sigma}^{-1} \lambda_p \left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)\left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)^T\right\} d\boldsymbol{\Sigma} d\theta dr
\end{aligned}
$$

The Bayes' Factor will be:

$$
\begin{aligned}
BF_{\tau,n_+} &= \frac{\kappa^{D/2-1} \cdot d^{1-D/2}}{\sigma_d^2 \cdot \sqrt{\pi} \cdot 2^{(\nu_p D+1)/2}} \cdot \left(\frac{\lambda}{\lambda_p}\right)^{D/2} \cdot \frac{|\boldsymbol{\Psi_p}|^{\nu_p/2}}{\Gamma_D\left(\frac{\nu_p}{2}\right)} \cdot \frac{I_{D/2-1}\left(\frac{rd}{\sigma_d^2}\right)}{I_{D/2}(\kappa)} \cdot exp\left\{-\frac{d^2}{2}\right\} \\
&\quad \times \int_r \int_\theta \int_{\boldsymbol{\Sigma}} \frac{r^{3D/2-1} \prod_{j=0}^{D-2} [sin\theta_{j+1}]^{D-2-j}}{|\boldsymbol{\Sigma}|^{(\nu_p+D+1)/2}} \cdot exp\left\{-\frac{1}{2}\left(\boldsymbol{A_\tau} + r^2\right) + \boldsymbol{B_\tau} + \kappa \boldsymbol{\mu_\theta}^T \boldsymbol{T}_\theta\right\} \\
&\quad \times exp\left\{-\frac{1}{2} tr \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi} + \boldsymbol{x_i} \boldsymbol{x_i}^T + \lambda \boldsymbol{\mu_o} \boldsymbol{\mu_o}^T\right)\right\} \\
&\quad \times exp\left\{+\frac{\lambda_p}{2} tr \boldsymbol{\Sigma}^{-1} \lambda_p \left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)\left(\boldsymbol{\mu_p} - \frac{n_\tau r \boldsymbol{L}^{1/2} \boldsymbol{T}_\theta}{\lambda_p}\right)^T\right\} d\boldsymbol{\Sigma} d\theta dr
\end{aligned}
$$

The full conditional posteriors under the IC or the OOC scenario will be:

- $(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,|(\tau > n, \boldsymbol{x_n})$

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\tau > n, \boldsymbol{x_n}) &\propto f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau > n)\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{(\nu_p+D+2)/2} \cdot exp\left\{-\frac{1}{2}\left(tr\boldsymbol{\Psi_p}\boldsymbol{\Sigma}^{-1} + \lambda_p(\boldsymbol{\mu}-\boldsymbol{\mu_p})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_p})\right)\right\} \end{aligned}$$

Thus $\boldsymbol{\mu}, \boldsymbol{\Sigma}|\tau > n, \boldsymbol{x_n}) \sim NIW\,(\boldsymbol{\mu_p}, \lambda_p, \nu_p, \boldsymbol{\Psi_p})$

- $\boldsymbol{\mu}|(\boldsymbol{\Sigma}, r, \theta, \tau \leq n, \boldsymbol{x_n})$

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{\Sigma}, r, \theta, \tau \leq n, \boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, r, \theta, \tau \leq n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \\ &\propto exp\left\{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}(n\bar{\boldsymbol{x}}_{\boldsymbol{n}} - n_\tau r\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta) - \frac{1}{2}\lambda_p\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\} \end{aligned}$$

Thus $\boldsymbol{\mu}|(\boldsymbol{\Sigma}, r, \theta, \tau \leq n, \boldsymbol{x_n}) \sim N_D\left(\boldsymbol{\mu_p} - \frac{n_\tau r\boldsymbol{L}^{1/2}\boldsymbol{T}_\theta}{\lambda_p}, \frac{\boldsymbol{\Sigma}}{\lambda_p}\right)$

- $\boldsymbol{\Sigma}|(\boldsymbol{\mu}, r, \theta, \tau \leq n, \boldsymbol{x_n})$

$$\begin{aligned} p\left(\boldsymbol{\Sigma}|\boldsymbol{\mu}, r, \theta, \tau \leq n, \boldsymbol{x_n}\right) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, r, \theta, \tau \leq n)\pi(\boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-(n+\nu_0+D+1)/2} \cdot exp\left\{-\frac{1}{2}\left(tr\boldsymbol{\Psi_p}\boldsymbol{\Sigma}^{-1} + r^2\boldsymbol{A_\tau}) + r\boldsymbol{B_\tau}\right)\right\} \end{aligned}$$

- $r|(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \theta, \tau \leq n, \boldsymbol{x_n})$

$$\begin{aligned} p(r|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \theta, \tau \leq n, \boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, r, \theta, \tau \leq n)\pi(r) \\ &\propto r^{3D/2-1} \cdot exp\left\{-\frac{r^2\left(\boldsymbol{A_\tau} + 1/{\sigma_d}^2\right)}{2} + r\boldsymbol{B_\tau}\right\} \cdot I_{D/2-1}\left(\frac{rd}{\sigma_d^2}\right) \end{aligned}$$

where $I_{D/2-1}(\cdot)$ is the modified Bessel function of the first kind with order $D/2-1$.

- $\theta|\,(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r, \tau \leq n, \boldsymbol{x_n})$

$$
\begin{aligned}
p(\theta|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\tau\leq n,\boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\tau\leq n)\pi(\theta)\\
&\propto \prod_{j=0}^{D-2}\left(sin\theta_{j+1}\right)^{D-2-j}\cdot exp\left\{-\frac{r^2\boldsymbol{A_\tau}}{2}+r\boldsymbol{B_\tau}+\kappa\boldsymbol{\mu_\theta}\boldsymbol{T_\theta}\right\}
\end{aligned}
$$

- $\tau|\left(\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\boldsymbol{x_n}\right)$

$$
\begin{aligned}
p\left(\tau=k|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\boldsymbol{x_n}\right) &= \frac{f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\tau\leq n)\pi(\tau=k)}{\sum_{j=1}^{n}f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\tau\leq n)\pi(\tau=j)}\\
&= \frac{exp\left\{r\boldsymbol{B_k}-\frac{r^2\boldsymbol{A_k}}{2}\right\}\left((1-p)^{(k-1)^\beta}-(1-p)^{k^\beta}\right)}{\sum_{j=1}^{n}exp\left\{r\boldsymbol{B_j}-\frac{r^2\boldsymbol{A_j}}{2}\right\}\left((1-p)^{(j-1)^\beta}-(1-p)^{j^\beta}\right)}
\end{aligned}
$$

D2: Normal Model for the covariance matrix $\boldsymbol{\Sigma}$

Now, the interest is placed on detecting scale or rotation shifts of sequentially gathered multivariate data. Specifically, we develop M3S model for detecting shifts in the the covariance matrix in D dimensions, obtaining the data sequentially. The IC state, including the likelihood, the IC parameters $\boldsymbol{\theta}$, the prior and the marginal, is the same with the M3S for the mean vector. For the expression of a scale shift, we assume a positive definite and diagonal matrix $\boldsymbol{S}\in diag\left(\mathbb{R}_{>0}^{D\times D}\right)$:

$$
\boldsymbol{S}=\begin{pmatrix}\kappa_1 & & 0\\ & \ddots & \\ 0 & & \kappa_D\end{pmatrix}
$$

The most challenging and demanding part of the model is the appropriate expression of a rotation shift in high dimensions. Geometrically, every rotation of a point in D dimensions takes places in a 2 dimensional plane, while all the other $D-2$ dimensions are fixed (free). This means that the rotated point maintains a constant distance to the invariant pivot element, which is not necessarily a 0-dimensional point. Specifically, only in 2 dimensions, the invariant element will be a point, while

in 3 dimensions will be a line, in 4 a plane etc. In other words, the rotation is around a point, a line and a plane respectively. For a plane $\alpha\beta$, which is formed by the axes α and β, the rotation matrix for an angle $\theta_{\alpha,\beta}$ is:

$$\boldsymbol{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\theta_{\alpha,\beta}) = \left(\begin{array}{c|c} & r_{\alpha,\alpha} = cos(\theta_{\alpha,\beta}) \\ & r_{\beta,\beta} = cos(\theta_{\alpha,\beta}) \\ & r_{\alpha,\beta} = -sin(\theta_{\alpha,\beta}) \\ r_{i,j} & r_{\beta,\alpha} = sin(\theta_{\alpha,\beta}) \\ & r_{i,j} = 1,\ i = j,\ i \neq \{\alpha, \beta\} \\ & r_{i,j} = 0,\ elsewhere \end{array} \right)$$

All the possible rotations in a D-dimensional space equals to the number of all the pairs of planes formed, i.e. the combination $\binom{D}{2} = \frac{D(D-1)}{2}$. The matrix $\boldsymbol{R}(\theta)$, which allows all the possible rotations in D dimensions, is the product of all the possible $\boldsymbol{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, i.e.:

$$\boldsymbol{R}(\theta) = \prod_{\alpha=1}^{D-1} \prod_{\beta=\alpha+1}^{D} \boldsymbol{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\theta_{\alpha,\beta})$$

where $\theta = (\theta_{1,1}, ..., \theta_{D-1,D})$ and $\theta \in [0, 2\pi)^{D-1} \times [0, \pi)^{(D-1)(D-2)/2}$. $\boldsymbol{R}(\theta)$ is a real and orthogonal matrix with $|\boldsymbol{R}(\theta)| = 1$, belonging to the special orthogonal group, i.e. $\boldsymbol{R}(\theta) \in SO(D)$. It is worth to note that $\boldsymbol{R}(\theta)$ consists of all the possible rotation in D dimensions, modelling all the correlations $\rho_{\alpha,\beta}$ of the covariance matrix. However, if not all the correlations are of interest, then we can more flexible by setting the corresponding $\boldsymbol{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\theta_{\alpha,\beta}) = \boldsymbol{I}_D$. In this way, we set the angle of the corresponding rotation $\theta_{\alpha,\beta} = 0$, not allowing a drift for $\rho_{\alpha,\beta}$.

For the development of an appropriate M3S model, which will be able to detect a potential shift in all the components of the covariance matrix, we define the rotation and scaling matrix $\boldsymbol{T} = \boldsymbol{R}(\theta)\boldsymbol{S}^{1/2}$. $\boldsymbol{T}$ which will provide the properties of anisotropic scaling and rotation to the model. We set $\boldsymbol{\phi} = \{diag(\boldsymbol{S}), \theta\}$, i.e. the diagonal elements of $\boldsymbol{S}$ and all the rotation angles $\theta_{\alpha,\beta}$. For the OOC state, we replace $\boldsymbol{\Sigma}$

with $g(\boldsymbol{\theta}, \boldsymbol{\phi}) = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T$. Thus, the likelihood will be:

$$f\left(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \tau\right) = \begin{cases} \prod_{i=1}^{\tau-1} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \prod_{i=\tau}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\right) & \text{if } \tau \leq n \\ \prod_{i=1}^{n} f\left(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) & \text{if } \tau > n \end{cases}$$

Regarding the priors, for the change point we assume $\tau \sim DW(p, \beta)$, as in the previous 3S models. For the scaling matrix $\boldsymbol{S}$, we assume an Inverse Wishart, or $\boldsymbol{S} \sim IW(\nu_s, \boldsymbol{D})$. For the components of the rotation matrix $\boldsymbol{R}(\theta)$, we assume a von-Mises distribution for the first $D-1$ components and a p-periodic von-Mises for the rest $\frac{(D-1)(D-2)}{2}$. Specifically, $\theta_{\alpha,\beta} \sim vM(\boldsymbol{\mu}_{\alpha,\beta}, \kappa_{\alpha,\beta})$ for $\alpha = 1$ and $\beta > \alpha$ and $\theta_{\alpha,\beta} \sim \pi vM(\boldsymbol{\mu}_{\alpha,\beta}, \kappa_{\alpha,\beta})$ otherwise. For the reduction of the next mathematical expressions, we set $\boldsymbol{C_n} = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{\boldsymbol{n}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{\boldsymbol{n}})^T$, $\boldsymbol{C_o} = (\bar{\boldsymbol{x}}_{\boldsymbol{n}} - \boldsymbol{\mu_o})(\bar{\boldsymbol{x}}_{\boldsymbol{n}} - \boldsymbol{\mu_o})^T$, $\boldsymbol{X_{t_1:t_2}} = \sum_{i=t_1}^{t_2}\boldsymbol{x_i}$, $n_t = n - t + 1$, $\boldsymbol{\mu_p} = \frac{n\bar{\boldsymbol{x}}_{\boldsymbol{n}} + \lambda\boldsymbol{\mu}_0}{n+\lambda}$, $\lambda_p = n + \lambda$, $\nu_p = n + \nu_0$, $\boldsymbol{\Psi_p} = \boldsymbol{C} + \boldsymbol{\Psi} + \frac{\lambda n}{\lambda + n}\boldsymbol{C_o}$, $\boldsymbol{M_p} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X_{1:\tau-1}} + \lambda\boldsymbol{\mu_o}\right) + \left(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\right)^{-1}\boldsymbol{X_{\tau:n}}$, $\boldsymbol{\Sigma_p} = \left(\left(\frac{\boldsymbol{\Sigma}}{\lambda_p - n_t}\right)^{-1} + \left(\frac{\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T}{n_t}\right)^{-1}\right)^{-1}$, $\boldsymbol{K_t} = \sum_{i=t}^{n}\boldsymbol{T}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\left(\boldsymbol{T}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right)^T$, $\boldsymbol{\Psi_{p,t}} = +\boldsymbol{\Psi} + \lambda(\boldsymbol{\mu} - \boldsymbol{\mu_o})(\boldsymbol{\mu} - \boldsymbol{\mu_o})^T + \boldsymbol{C_{t-1}} + \boldsymbol{K_t}$ then the IC and the OOC marginal will be respectively:

$$\begin{aligned} f(\boldsymbol{x_n}|\tau > n) &= \int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau > n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})d\boldsymbol{\mu}d\boldsymbol{\Sigma} \\ &= \frac{1}{\pi^{nD/2}} \cdot \left(\frac{\lambda}{\lambda_p}\right)^{D/2} \cdot \frac{\Gamma_D\left(\nu_p/2\right)}{\Gamma_D\left(\nu_0/2\right)} \cdot \frac{|\boldsymbol{\Psi}|^{\nu_0/2}}{|\boldsymbol{\Psi_p}|^{\nu_p/2}} \end{aligned}$$

$$f(\boldsymbol{x_n}|\tau \leq n) = \int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \tau \leq n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{T})d\boldsymbol{\mu}d\boldsymbol{\Sigma}d\boldsymbol{T}$$

$$
\begin{aligned}
&= \int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\int_{\boldsymbol{\mu}} \frac{|\boldsymbol{S}|^{-n_\tau/2}\cdot|\boldsymbol{\Sigma}|^{-n/2}}{(2\pi)^{nD/2}}\cdot exp\left\{-\frac{1}{2}\sum_{i=1}^{\tau-1}(\boldsymbol{x_i}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\right\}\\
&\quad\times exp\left\{-\frac{1}{2}\sum_{i=\tau}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu})^T\left(\boldsymbol{T}^T\boldsymbol{\Sigma}\boldsymbol{T}\right)^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\right\}\\
&\quad\times\left(\frac{\lambda}{2\pi}\right)^{D/2}\cdot|\boldsymbol{\Sigma}|^{-1/2}\cdot exp\left\{-\frac{\lambda}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_o})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_o})\right\}\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{T})d\boldsymbol{\mu}d\boldsymbol{\Sigma}d\boldsymbol{T}\\
&= \int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\left(\frac{\lambda}{(2\pi)^n}\right)^{D/2}\cdot\frac{|\boldsymbol{\Sigma_p}|^{1/2}}{|\boldsymbol{S}|^{n_\tau/2}\cdot|\boldsymbol{\Sigma}|^{(n+1)/2}}\\
&\quad\times exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{\tau-1}\boldsymbol{x_i}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x_i}+\sum_{i=\tau}^{n}\boldsymbol{x_i}^T\left(\boldsymbol{T}^T\boldsymbol{\Sigma}\boldsymbol{T}\right)^{-1}\boldsymbol{x_i}+\lambda\boldsymbol{\mu_o}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu_o}\right)\right\}\\
&\quad\times exp\left\{+\frac{1}{2}\boldsymbol{M_p}^T\boldsymbol{\Sigma_p}^{-1}\boldsymbol{M_p}\right\}\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{T})d\boldsymbol{\mu}d\boldsymbol{\Sigma}d\boldsymbol{T}\\
&= \left(\frac{\lambda}{(2\pi)^n}\right)^{D/2}\cdot\frac{|\boldsymbol{\Psi}|^{\nu_0/2}}{2^{\nu_0 D/2}\Gamma_D\left(\frac{\nu_0}{2}\right)}\cdot\frac{|\boldsymbol{D}|^{\nu_s/2}}{2^{\nu_s D/2}\Gamma_D\left(\frac{\nu_s}{2}\right)}\cdot\frac{1}{\pi^{(D-1)(D-2)/2}}\\
&\quad\times\prod_{\alpha=1}^{D-1}\prod_{\beta=\alpha+1}^{D}\frac{1}{I_0(\kappa_{\alpha,\beta})}\int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\frac{|\boldsymbol{\Sigma_p}|^{1/2}}{|\boldsymbol{S}|^{(n_\tau+\nu_s+D+1)/2}\cdot|\boldsymbol{\Sigma}|^{(\nu_p+D+1)/2}}\\
&\quad\times exp\left\{-\frac{1}{2}tr\boldsymbol{D}\boldsymbol{S}^{-1}+\kappa_{\alpha,\beta}\cdot cos\left(1+\mathbb{1}_{\{\alpha>1\}}\right)(\theta_{\alpha,\beta}-\kappa_{\alpha,\beta})\right\}\\
&\quad\times exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi}+\boldsymbol{x_i}\boldsymbol{x_i}^T+\lambda\boldsymbol{\mu_o}\boldsymbol{\mu_o}^T-\lambda_p\boldsymbol{M_p}\boldsymbol{M_p}^T\right)\right\}d\boldsymbol{\Sigma}d\boldsymbol{T}
\end{aligned}
$$

The Bayes' Factor will be:

$$
\begin{aligned}
BF_{\tau,n_+} &= \left(\frac{\lambda_p}{2^{\nu_p+\nu_s}}\right)^{D/2}\cdot\frac{|\boldsymbol{\Psi_p}|^{\nu_p/2}}{\Gamma_D\left(\frac{\nu_p}{2}\right)}\cdot\frac{|\boldsymbol{D}|^{\nu_s/2}}{\Gamma_D\left(\frac{\nu_s}{2}\right)}\cdot\frac{1}{\pi^{(D-1)(D-2)/2}}\\
&\quad\times\prod_{\alpha=1}^{D-1}\prod_{\beta=\alpha+1}^{D}\frac{1}{I_0(\kappa_{\alpha,\beta})}\int_{\boldsymbol{T}}\int_{\boldsymbol{\Sigma}}\frac{|\boldsymbol{\Sigma_p}|^{1/2}}{|\boldsymbol{S}|^{(n_\tau+\nu_s+D+1)/2}\cdot|\boldsymbol{\Sigma}|^{(\nu_p+D+1)/2}}\\
&\quad\times exp\left\{-\frac{1}{2}tr\boldsymbol{D}\boldsymbol{S}^{-1}+\kappa_{\alpha,\beta}\cdot cos\left(1+\mathbb{1}_{\{\alpha>1\}}\right)(\theta_{\alpha,\beta}-\kappa_{\alpha,\beta})\right\}
\end{aligned}
$$

$$\times exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi}+\boldsymbol{x_i}\boldsymbol{x_i}^T+\lambda\boldsymbol{\mu_o}\boldsymbol{\mu_o}^T-\lambda_p\boldsymbol{M_p}\boldsymbol{M_p}^T\right)\right\}d\boldsymbol{\Sigma}d\boldsymbol{T}$$

The full conditional posteriors under the IC or the OOC scenario will be:

- $(\boldsymbol{\mu},\boldsymbol{\Sigma})\,|(\tau>n,\boldsymbol{x_n})$

$$\begin{aligned}p(\boldsymbol{\mu},\boldsymbol{\Sigma}|\tau>n,\boldsymbol{x_n}) &\propto f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\tau>n)\pi(\boldsymbol{\mu},\boldsymbol{\Sigma})\\ &\propto |\boldsymbol{\Sigma}|^{(\nu_p+D+2)/2}\cdot exp\left\{-\frac{1}{2}\left(tr\boldsymbol{\Psi_p}\boldsymbol{\Sigma}^{-1}+\lambda_p(\boldsymbol{\mu}-\boldsymbol{\mu_p})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_p})\right)\right\}\end{aligned}$$

Thus $\boldsymbol{\mu},\boldsymbol{\Sigma}|\tau>n,\boldsymbol{x_n})\sim NIW\left(\boldsymbol{\mu_p},\lambda_p,\nu_p,\boldsymbol{\Psi_p}\right)$

- $\boldsymbol{\mu}|(\boldsymbol{\Sigma},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n})$

$$\begin{aligned}p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},r,\theta,\tau\leq n)\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\\ &\propto exp\left\{\boldsymbol{\mu}^T\boldsymbol{M_p}-\frac{1}{2}\lambda_p\boldsymbol{\mu}^T\boldsymbol{\Sigma_p}^{-1}\boldsymbol{\mu}\right\}\end{aligned}$$

Thus $\boldsymbol{\mu}|(\boldsymbol{\Sigma},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n})\sim N_D\left(\boldsymbol{\Sigma_p}\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X_{1:\tau-1}}+\lambda\boldsymbol{\mu_o}\right)+\left(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\right)^{-1}\boldsymbol{X_{\tau:n}}\right),\boldsymbol{\Sigma_p}\right)$

- $\boldsymbol{\Sigma}|(\boldsymbol{\mu},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n})$

$$\begin{aligned}p\left(\boldsymbol{\Sigma}|\boldsymbol{\mu},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n}\right) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{T},\tau\leq n)\pi(\boldsymbol{\Sigma})\\ &\propto |\boldsymbol{\Sigma}|^{-(\nu_p+D+1)/2}\cdot exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Psi_p}+\boldsymbol{C_{\tau-1}}+\boldsymbol{K_\tau}\right)\right\}\end{aligned}$$

Thus $\left(\boldsymbol{\Sigma}|\boldsymbol{\mu},\boldsymbol{T},\tau\leq n,\boldsymbol{x_n}\right)\sim IW_D\left(\nu_p,\boldsymbol{\Psi_p}+\boldsymbol{C_{\tau-1}}+\boldsymbol{K_\tau}\right)$

- $\boldsymbol{S}|(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{R}(\theta),\tau\leq n,\boldsymbol{x_n})$

$$\begin{aligned}p(\boldsymbol{S}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{R}(\theta),\tau\leq n,\boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{T},\tau\leq n)\pi(\boldsymbol{S})\\ &\propto |\boldsymbol{S}|^{-(n_\tau+\nu_s+D+1)/2}\cdot exp\left\{-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{K_\tau}+\boldsymbol{S}^{-1}\boldsymbol{D}\right)\right\}\end{aligned}$$

- $\theta_{a,b}|\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{S}, \tau \leq n, \boldsymbol{x_n}\right)$

$$\begin{aligned} p(\theta_{a,b}|(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{S}, \tau \leq n, \boldsymbol{x_n}) &\propto f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \tau \leq n)\pi(\theta_{a,b}) \\ &\propto exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\boldsymbol{K_\tau} + \kappa_{a,b} \cdot cos(\theta_{a,b} - \kappa_{a,b})\right\} \end{aligned}$$

- $\tau|\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \boldsymbol{x_n}\right)$

$$\begin{aligned} p\left(\tau = k|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \boldsymbol{x_n}\right) &= \frac{f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \tau \leq n)\pi(\tau = k)}{\sum_{j=1}^{n} f(\boldsymbol{x_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{T}, \tau \leq n)\pi(\tau = j)} \\ &= \frac{|\boldsymbol{S}|^{-nk/2}exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{C_{k-1}} + \boldsymbol{K_k}\right)\right\}\left((1-p)^{(k-1)^\beta} - (1-p)^{k^\beta}\right)}{\sum_{j=1}^{n}|\boldsymbol{S}|^{-nj/2}exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{C_{j-1}} + \boldsymbol{K_j}\right)\right\}\left((1-p)^{(j-1)^\beta} - (1-p)^{j^\beta}\right)} \end{aligned}$$

Appendix E: Convergence Plots for U3S Applications

Regarding the Markov Chain Monte Carlo (MCMC) sampling of the first application (precious metals dataset), we applied the Metropolis within Gibbs algorithm to obtain an independent posterior sample size of 10,000. The burn in period had length of 20,000 iterations, while the thinning was 500. It is worth mentioning that the reason of the large thinning that posterior of τ was bimodal with the modes to be far from each other, which created correlated batches in the Markov Chain. Figure 6.0.1 provides the trace plots and the autocorrelation function (ACF) plots of the full conditional posteriors for θ_1, θ_2^2, δ and τ. For the MCMC sampling of the second application (monthly increment dataset), we applied the Gibbs algorithm to generate 10,000 independent data points from the posterior distributions. Now, the burn in period had length of 2,000 iterations and the thinning was 50. The Figure 6.0.2 summarizes all the trace ACF plots of the full conditional posteriors for θ_1, θ_2^2, κ and τ.

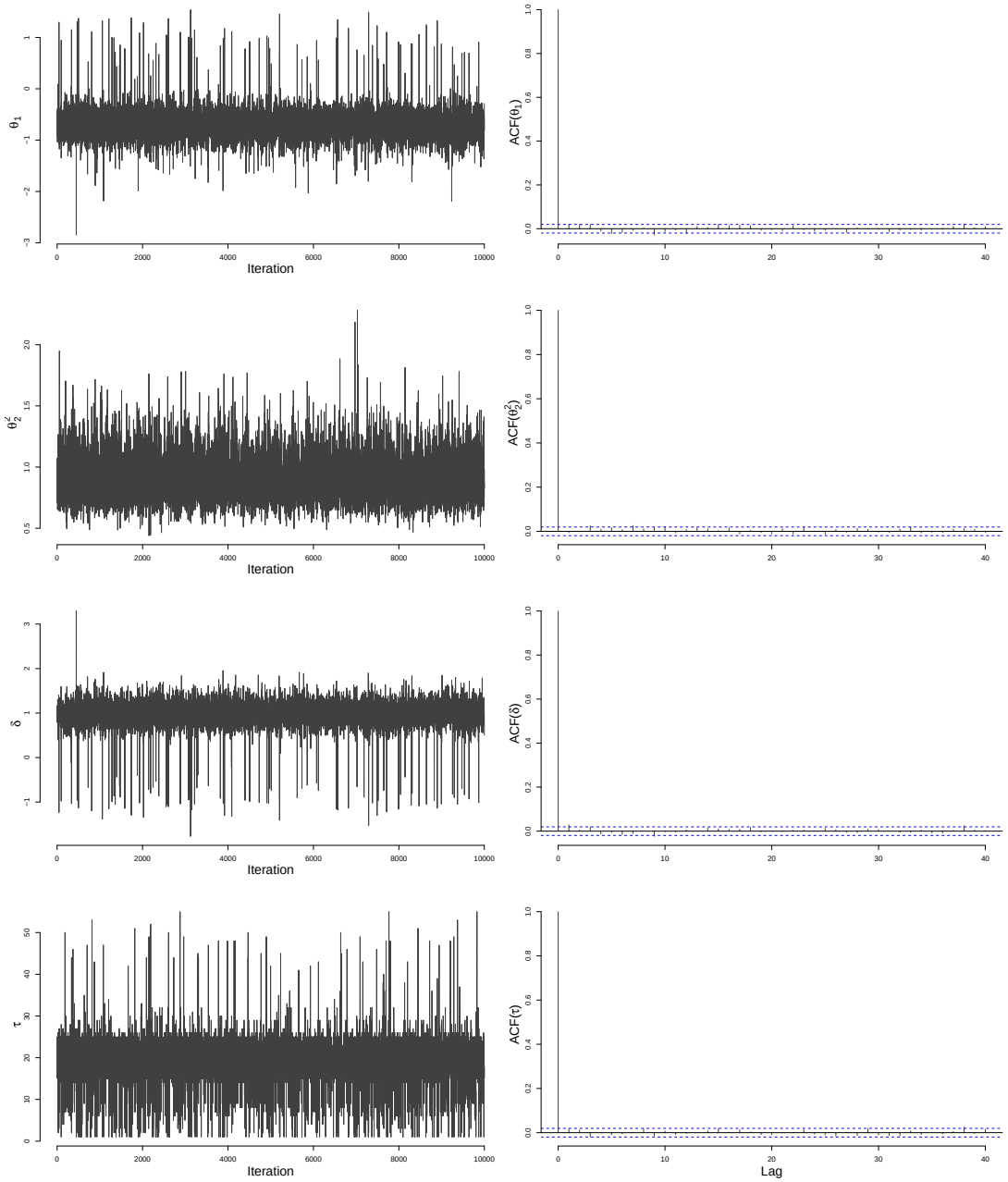


Figure 6.0.1: The trace plots and the ACF plots of the posterior samples for θ_1, θ_2^2, δ and τ for the application to the precious metals dataset

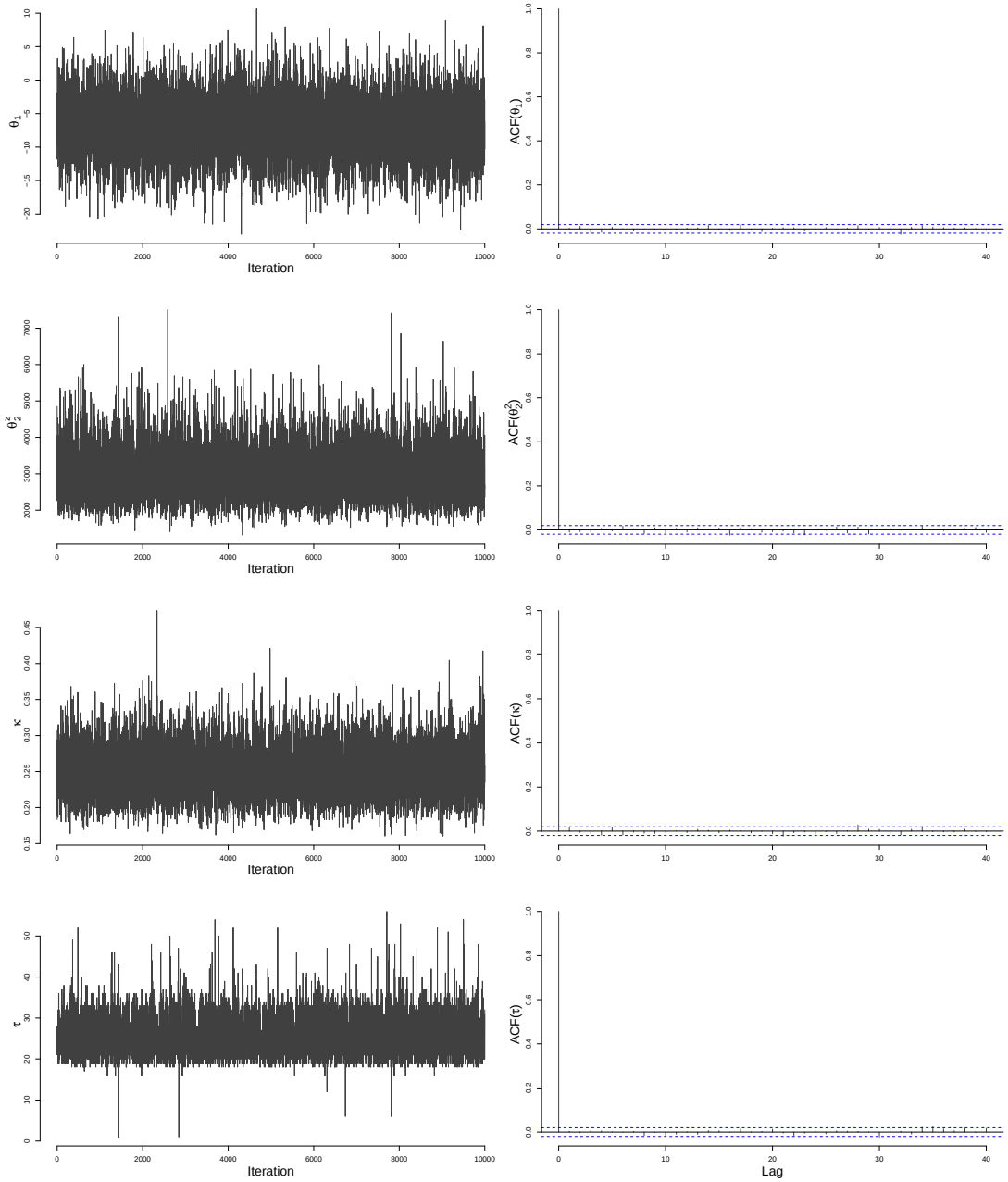
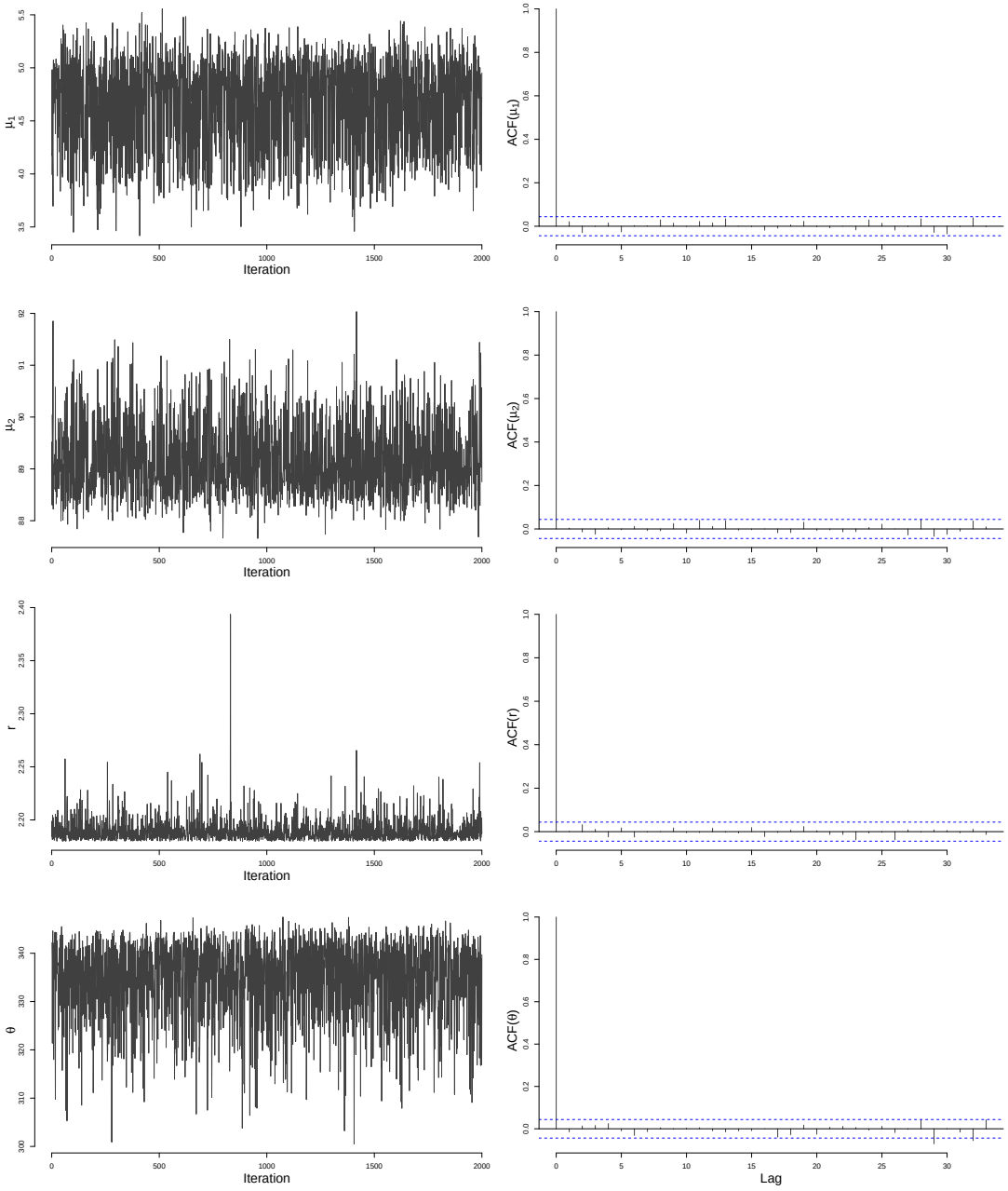


Figure 6.0.2: The trace plots and the ACF plots of the posterior samples for θ_1, θ_2^2, κ and τ for the application to the monthly increments dataset

Appendix F: Convergence Plots for M3S Application

Regarding the Markov Chain Monte Carlo (MCMC) sampling of the M3S applica-

tion to gravel data, we applied the Metropolis within Gibbs algorithm to obtain an independent posterior sample size of 2,000. The burn in period had length of 10,000 iterations, while the thinning was 100. The Figure 6.0.2 summarizes all the trace ACF plots of the full conditional posteriors for μ_1, μ_2, r, θ, σ_1^2, σ_2^2, ρ and τ.


μ1
Iteration
ACF(μ1)
μ2
Iteration
ACF(μ2)
r
Iteration
ACF(r)
θ
Iteration
ACF(θ)
Lag

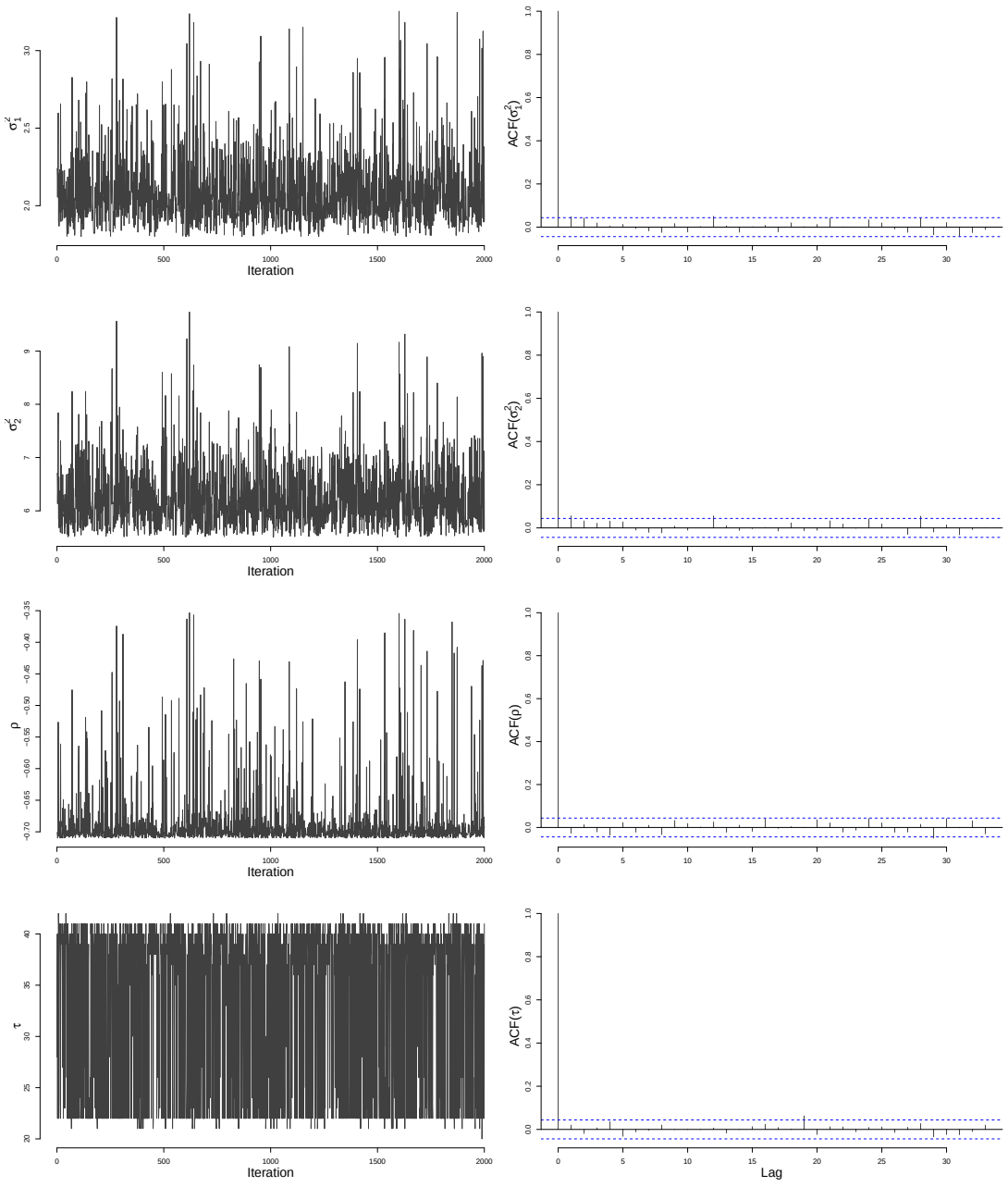


Figure 6.0.3: The trace plots and the ACF plots of the posterior samples for μ_1, μ_2, r, θ, σ_1^2, σ_2^2, ρ and τ for the gravel dataset

Bibliography

[1] Adams, R. P. and MacKay, D. J. (2007), "Bayesian online changepoint detection", *arXiv preprint arXiv:0710.3742.*

[2] Ali, S. (2020), "A predictive Bayesian approach to sequential time-between-events monitoring", *Quality and Reliability Engineering International*, 36, 1, pp. 365-387.

[3] Ali, S., and Riaz, M. (2020), "On designing a new Bayesian dispersion chart for process monitoring", *Arabian Journal for Science and Engineering*, 45, 3, pp. 2093-2111.

[4] Alloway Jr, J. A., and Raghavachari, M. (1991), "Control chart based on the Hodges-Lehmann estimator", *Journal of Quality Technology*, 23, 4, pp. 336-347.

[5] Amin, R. W., and Searcy, A. J. (1991), "A nonparametric exponentially weighted moving average control scheme", *Communications in Statistics-Simulation and Computation*, 20, 4, pp. 1049-1072.

[6] Aminnayeri, M., and Sogandi, F. (2016), "A risk adjusted self-starting Bernoulli CUSUM control chart with dynamic probability control limits", *AUT Journal of Modelling and Simulation*, 48, 2, pp. 103-110.

[7] Amirkhani, F., Amiri, A., and Tighkhorshid, E. (2018), "A Self-starting Risk-adjusted AFT based Control Chart for Monitoring the Survival Time of Pa-

tients", *Proceedings of the 2nd IEOM European conference on Industrial Engineering and Operations management.*

[8] Amiri, A., Ghashghaei, R., and Khosravi, P. (2016), "A self-starting control chart for simultaneous monitoring of mean and variance of autocorrelated simple linear profile", *In 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 209-213.

[9] Atalay, M., Caner Testik, M., Duran, S. and Weiß, C. H. (2020), "Guidelines for Automating Phase I of Control Charts by Considering Effects on Phase-II Performance of Individuals Control Chart", *Quality Engineering*, 32, 2, pp. 223-243.

[10] Atwi, A., Savla, K., and Dahleh, M. A. (2011), "A case study in robust quickest detection for hidden markov models", *In Proceedings of the 2011 American Control Conference*, IEEE, pp. 780-785.

[11] Bakir, S. T., and Reynolds, M. R. (1979), "A nonparametric procedure for process control based on within-group ranking", *Technometrics*, 21, 2, pp. 175-183.

[12] Baron, M. I. (2001), "Bayes stopping rules in a change-point model with a random hazard rate", *Sequential Analysis*, 20, 3, pp. 147-163.

[13] Baron, M. I. (2004), "Early detection of epidemics as a sequential change-point problem", *Longevity, aging and degradation models in reliability, public health, medicine and biology, LAD*, 2, pp. 31-43.

[14] Bather, J. A. (1967), "On a quickest detection problem", *The Annals of Mathematical Statistics*, 38, 3, 711-724.

[15] Bayarri, M. J. and García-Donato, G. (2005), "A Bayesian Sequential Look at u-Control Charts", *Technometrics*, 47, 2, pp. 142-151.

[16] Berger, J. O., Bernardo, J. M., and Sun, D. (2009), "The Formal Definition of Reference Priors", *Annals of Statistics*, 37, pp. 905-938.

[17] Bernardo, J. M. (1979), "Reference Posterior Distributions for Bayesian Inference", *Journal of the Royal Statistical Society* Series B (Methodological), 41, pp. 113-147.

[18] Bernardo, J. M., and Smith, A. F. M. (2000), *Bayesian Theory*, First edition, Wiley, New York.

[19] Bersimis, S., Psarakis, S., and Panaretos, J. (2007), "Multivariate statistical process control charts: an overview", *Quality and Reliability engineering international*, 23, 5, pp. 517-543.

[20] Bhattacharya, P. K., and Frierson Jr, D. (1981), "A nonparametric control chart for detecting small disorders", *The Annals of Statistics*, 9, 3, pp. 544-554.

[21] Bourazas, K. (2014), "Bayesian Statistical Process Control: Predictive Control Charts for Continuous Distributions in the Regular Exponential Family", *M.Sc. Thesis*, Department of Statistics, Athens University of Economics and Business, http://www.pyxida.aueb.gr/index.php?op=view_object&object_id=7217.

[22] Bourazas, K., Kiagias, D., and Tsiamyrtzis, P. (2021), "Predictive Control Charts (PCC): A Bayesian approach in online monitoring of short runs", *Journal of Quality Technology*, pp. 1-35.

[23] Brodsky, B. (2010), "Sequential detection and estimation of change-points", *Sequential Analysis*, 29, 2, pp. 217-233.

[24] Brook, D. A. E. D. and Evans, D. (1972), "An approach to the probability distribution of CUSUM run length" *Biometrika*, 59, 3, pp. 539-549.

[25] Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for testing the constancy of regression relationships over time", *Journal of the Royal Statistical Society: Series B (Methodological)*, 37, 2, pp. 149-163.

[26] Capizzi, G. and Masarotto, G. (2010), "Self-starting CUSCORE control charts for individual multivariate observations", *Journal of Quality Technology*, 42, 2, pp. 136-151.

[27] Capizzi, G. and Masarotto, G. (2012), "An enhanced control chart for start-up processes and short runs", *Quality Technology & Quantitative Management*, 9, 2, pp. 189-202.

[28] Capizzi, G. and Masarotto, G. (2013), "Phase I distribution-free analysis of univariate data", *Journal of Quality Technology*, 45, pp. 273-284.

[29] Capizzi, G., and Masarotto, G. (2019), "Guaranteed in-control control chart performance with cautious parameter learning", *Journal of Quality Technology*, pp. 1-19.

[30] Carlin, B.P. and Louis, T. A. (2009), *Bayesian Methods for Data Analysis*, Chapman & Hall, London.

[31] 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, pp. 1353-1366.

[32] Castillo, E. D., Grayson, J. M., Montgomery, D. C., and Runger, G. C. (1996), "A review of statistical process control techniques for short run manufacturing systems", *Communications in Statistics–Theory and Methods*, 25, 11, pp. 2723-2737.

[33] Castillo, E. D., and Montgomery, D. C. (1994), "Short-run statistical process control Q-chart enhancements and alternative methods", *Quality and Reliability Engineering International*, 10, 2, pp. 87-97.

[34] Castillo, E. D., and Montgomery, D. C. (1995), "A Kalman filtering process control scheme with an application in semiconductor short run manufacturing", *Quality and Reliability Engineering International*, 11, 2, pp. 101-105.

[35] Celano, G., Castagliola, P., Fichera, S., and Nenes, G. (2013), "Performance of t control charts in short runs with unknown shift sizes", *Computers & Industrial Engineering*, 64, 1, pp. 56-68.

[36] Celano, G., Castagliola, P., and Trovato, E. (2012), "The economic performance of a CUSUM t control chart for monitoring short production runs", *Quality Technology & Quantitative Management*, 9, 4, pp. 329-354.

[37] Celano, G., Castagliola, P., Trovato, E., and Fichera, S. (2012), "The economic performance of the Shewhart t chart", *Quality and Reliability Engineering International*, 28, 2, pp. 159-180.

[38] Chakraborti, S., and Graham, M. A. (2019), "Nonparametric (distribution-free) control charts: An updated overview and some results", *Quality Engineering*, 31, 4, pp. 523-544.

[39] Chakraborti, S., Human, S. W. and Graham, M. A. (2008), "Phase I Statistical Process Control Charts: An Overview and Some Results", *Quality Engineering*, 21, pp. 52-62.

[40] Chang, Z., and Sun, J. (2016), "AEWMA t control chart for short production runs", *Journal of Systems Science and Information*, 4, 5, pp. 444-459.

[41] Colosimo, B. M., and Castillo, E. D. (2006), *Bayesian process monitoring, control and optimization*, CRC Press, Chapman & Hall.

[42] Conover, W. J., Tercero-Gómez, V. G. and Cordero-Franco, A. E. (2017), "The sequential normal scores transformation", *Sequential Analysis*, 36, 3, pp. 397-414.

[43] Conover, W. J., Tercero-Gómez, V. G. and Cordero-Franco, A. E. (2018), "A Look at Sequential Normal Scores and How They Apply to Financial Data Analysis", *Journal of Applied Mathematics and Physics*, 6, pp. 787-816.

[44] Conover, W. J., Tercero-Gómez, V. G. and Cordero-Franco, A. E. (2019), "An Approach to Statistical Process Control that is New, Nonparametric, Simple, and Powerful", *arXiv*:1901.04443.

[45] Cornélissen, G., Halberg, F., Hawkins, D., Otsuka, K., and Henke, W. (1997), "Individual assessment of antihypertensive response by self-starting cumulative sums", *Journal of medical engineering & technology*, 21, 3-4, pp. 111-120.

[46] Crosier, R. B. (1988), "Multivariate generalizations of cumulative sum quality-control schemes", *Technometrics*, 30, 3, pp. 291-303.

[47] Crowder, S. V., and Eshleman, L. (2001), "Small sample properties of an adaptive filter applied to low volume SPC", *Journal of Quality Technology*, 33, 1, pp. 29-46.

[48] Dasdemir, E., Weiß, C., Testik, M. C. and Knoth, S. (2016), "Evaluation of Phase I Analysis Scenarios on Phase II Performance of Control Charts for Autocorrelated Observations", *Quality Engineering* , 28, 3, pp. 293-304.

[49] Deming, W. E. (1986), *Out of Crisis*, The MIT Press.

[50] Dessein, A., and Cont, A. (2013), "Online change detection in exponential families with unknown parameters", *In International Conference on Geometric Science of Information*, Springer, Berlin, Heidelberg, pp. 633-640.

[51] Dogu, E., and Kim, M. J. (2020), "Self-starting single control charts for multivariate processes: a comparison of methods", *Production*, 30.

[52] Dogu, E., and Noor-ul-Amin, M. (2021), "Monitoring exponentially distributed time between events data: self-starting perspective", *Communications in Statistics-Simulation and Computation*, 50, 1, pp. 1-13.

[53] Dong, Y., Hedayat, A. S. and Sinha, B. K. (2008), "Surveillance Strategies for Detecting Changepoint in Incidence Rate Based on Exponentially Weighted

Moving Average Methods", *Journal of the American Statistical Association*, 103, 482, pp. 843-853.

[54] Dunn, O. J. (1961), "Multiple Comparisons Among Means", *Journal of the American Statistical Association*, 56, pp. 52-64.

[55] Feltz, C. J., and Shiau, J. J. H. (2001), "Statistical process monitoring using an empirical Bayes multivariate process control chart", *Quality and Reliability Engineering International*, 17, 2, pp. 119-124.

[56] Fetter, G., and Rakes, T. R. (2011), "A self-balancing CUSUM approach for the efficient allocation of resources during post-disaster debris disposal operations", *Operations Management Research*, 4, 1-2, pp. 51-60.

[57] Frisen M. (1992). "Evaluations of Methods for Statistical Surveillance" *Statistics in Medicine*, 11, pp. 1489-1502.

[58] Geisser, S. (1993), *Predictive Inference: An Introduction*, Chapman & Hall, London.

[59] Geisser, S., and Cornfield, J. (1963), "Posterior distributions for multivariate normal parameters". *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 368-376.

[60] Geweke, J., and Amisano, G. (2010), "Comparing and evaluating Bayesian predictive distributions of asset returns". *International Journal of Forecasting*, 26, 2, pp. 216-230.

[61] Goedhart, R., Schoonhoven, M., Does, R. J. (2017), "Guaranteed in-control performance for the Shewhart X and $\overline{X}$ control charts, *Journal of Quality Technology*, 49, 2, pp. 155-171.

[62] Gombay, E. (2003), "Sequential change-point detection and estimation". *Sequential analysis*, 22, 3, pp. 203-222.

[63] Gombay, E. (2004), "U-statistics in sequential tests and change detection". *Sequential analysis*, 23, 2, pp. 257-274.

[64] Gu, K., Jia, X., You, H., and Zhang, S. (2014), "A t-chart for monitoring multivariety and small batch production run". *Quality and Reliability Engineering International*, 30, 2, pp. 287-299.

[65] Haldane, J. B. S. (1932), "A Note on Inverse Probability", *Mathematical Proceedings of the Cambridge Philosophical Society*, 28(1), pp.55-61.

[66] Hansen, B., and Ghare, P. (1987), *Quality Control and Application*, Prentice-Hall, Englewood Cliffs, NJ.

[67] Harrison, P. J. (1999), "Statistical process control and model monitoring", *Journal of Applied Statistics*, 26, 2, pp. 273-292.

[68] Harrison, P. J., and Veerapen, P. P. (1994), "A Bayesian decision approach to model monitoring and Cusums", *Journal of Forecasting*, 13, 1, pp. 29-36.

[69] Harrison, J., and West, M. (1991), "Dynamic linear model diagnostics", *Biometrika*, 78, 4, pp. 797-808.

[70] Harrison, J., and West, M. (1987), "Practical bayesian forecasting", *Journal of the Royal Statistical Society: Series D (The Statistician)*, 36, 2-3, pp. 115-125.

[71] Hawkins, D. M. (1977), "Testing a sequence of observations for a shift in location", *Journal of the American Statistical Association*, 72, 357, pp. 180-186.

[72] Hawkins, D. M. (1987), "SelfâĂŘstarting CUSUM charts for location and scale", *Journal of the Royal Statistical Society: Series D (The Statistician)*, 36, 4, 299-316.

[73] Hawkins, D. M., and Maboudou-Tchao, E. M. (2007), "Self-starting multivariate exponentially weighted moving average control charting", *Technometrics*, 49, 2, pp. 199-209.

[74] Hawkins, D. M., and Olwell, D. H. (1998), *Statistics for engineering and physical science-cumulative sum charts and charting for quality improvement*, Springer-Verlag, New York.

[75] Hawkins, D. M., Qiu, P., and Kang, C. W. (2003), "The changepoint model for statistical process control", *Journal of quality technology*, 35, 4, pp. 355-366.

[76] Hawkins, D. M. and Zamba, K. D. (2005), "A change-point model for a shift in variance", *Technometrics*, 37, 1, pp. 21-31.

[77] Hawkins, D. M. and Zamba, K. D. (2005), "Statistical process control for shifts in mean or variance using a changepoint formulation", *Technometrics*, 47, 2, pp. 164-173.

[78] He, F., Jiang, W., and Shu, L. (2008), "Improved self-starting control charts for short runs", *Quality Technology & Quantitative Management*, 5, 3, pp. 289-308.

[79] Holmes, D. S. and Mergen, A. (1983). "Improving the Performance of the T^2 Control Chart", *Quality Engineering*, Vol. 5, pp. 619-625.

[80] Howley, P., Hancock, S. and Ford, M. (2009), "Monitoring clinical indicators". *3rd Annual ASEARC Research Conference.*

[81] Hou, Y., He, B., Zhang, X., Chen, Y., and Yang, Q. (2020), "A new Bayesian scheme for self-starting process mean monitoring". *Quality Technology & Quantitative Management*, 17, 6, pp. 661-684.

[82] Hou, S. and Yu, K. (2020), "A non-parametric CUSUM control chart for process distribution change detection and change type diagnosis". *International Journal of Production Research*, 1-21.

[83] Ibrahim, J., and Chen, M. (2000), "Power Prior Distributions for Regression Models", *Statistical Science*, 15, pp. 46-60.

[84] Ibrahim, J., Chen, M., and Sinha, D. (2003), "On Optimality Properties of the Power Prior". *Journal of the American Statistical Association*, 98, pp. 204-213.

[85] Jeffreys, H. (1961), *Theory of Probability*, third edition, Oxford University Press.

[86] 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, pp. 349-364

[87] 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, pp. 265-280.

[88] Kass, R. E. and Raftery A. E. (1995), "Bayes factors", *Journal of the American Statistical Association*, 90, 430, pp. 773-795.

[89] Kawamura, H., Nishina, K., Higashide, M., and Suzuki, T. (2013), "Application of Q charts for short run autocorrelated data". *International Journal of Innovative Computing, Information and Control*, 9, 9, pp. 3667-3676.

[90] Keefe, M. J., Woodall, W. H. and Jones-Farmer, L. A. (2015), "The Conditional In-Control Performance of Self-Starting Control Charts", *Quality Engineering* , 27, 4, pp. pp. 488-499.

[91] Kenett, R. S. and Pollak, M. (2012). "On Assessing the Performance of Sequential Procedures for Detecting a Change" *Quality and Reliability Engineering International*, 28, 5, pp. 500-507.

[92] Kenett, R. S., and Zacks, S. (2021), *Modern industrial statistics: With applications in R, MINITAB, and JMP*, Third edition, John Wiley & Sons, New York.

[93] Kerman, J. (2011), "Neutral Noninformative and Informative Conjugate Beta and Gamma Prior Distributions". *Electronic Journal of Statistics*, 5, pp. 1450-1470.

[94] Khoo, M. B., and Quah, S. H. (2002), "Proposed short runs multivariate control charts for the process mean", *Quality Engineering* , 14, 4, pp. pp. 603-621.

[95] Khoo, M. B., Quah, S. H., Low, H. C., and Ch'ng, C. K. (2005), "Short runs multivariate control chart for process dispersion", *International Journal of Reliability, Quality and Safety Engineering* , 12, 2, pp. 127-147.

[96] Khosravi, P., and Amiri, A. (2019), "Self-Starting control charts for monitoring logistic regression profiles", *Communications in Statistics-Simulation and Computation* , 48, 6, pp. 1860-1871.

[97] Kiagias, D. (2014), "Bayesian Statistical Process Control: Predictive Control Charts for Discrete Distributions in the Regular Exponential Family", *M.Sc. Thesis*, Department of Statistics, Athens University of Economics and Business, http://www.pyxida.aueb.gr/index.php?op=view_object&object_id=7218.

[98] Korzenowski, A. L., Vidor, G., Vaccaro, G. L. R., and Ten Caten, C. S. (2015), "Control charts for flexible and multi-variety production systems", *Computers & Industrial Engineering* , 88, pp. 284-292.

[99] Kumar, N., Chakraborti, S. (2017), "Bayesian Monitoring of Times Between Events: The Shewhart t_r-Chart.", *Journal of Quality Technology*, 49, 2, pp. 136-154.

[100] Lai, T. L., and Xing, H. (2010), "Sequential change-point detection when the pre-and post-change parameters are unknown", *Sequential analysis*, 29, 2, pp. 162-175.

[101] Lampreia, S., and Requeijo, J. (2012), "Analysis of an equipment condition by Q & multivariate Q control charts", *In Proceedings of the 14th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems (MAMECTIS'12)*, pp. 1-3.

[102] Lang, M. (2019), "Control Limits for an Adaptive Self-Starting Distribution-Free CUSUM Based on Sequential Ranks", *Technologies*, 7, 4, pp. 71.

[103] Lauro, C., Antoch, J., Vinzi, V. E., and Saporta, G. (Eds.). (2012), *Multivariate Total Quality Control: foundation and recent advances*, Springer Science & Business Media, Berlin.

[104] Lee, J., Wang, N., Xu, L., Schuh, A. and Woodall, W. H. (2013), "The Effect of Parameter Estimation on Upper-Sided Bernoulli Cumulative Sum Charts", *Quality and Reliability Engineering International*, 29, 5, 639-651.

[105] Li, J. (2021), "Nonparametric adaptive CUSUM chart for detecting arbitrary distributional changes", *Quality and Reliability Engineering International*, 53, 2, 154-172.

[106] Li, Y., Liu, Y., Zou, C., and Jiang, W. (2014), "A self-starting control chart for high-dimensional short-run processes", *International Journal of Production Research*, 52, 2, pp. 445-461.

[107] Li, Z., Luo, Y., and Wang, Z. (2010), "Cusum of Q chart with variable sampling intervals for monitoring the process mean", *International Journal of Production Research*, 48, 16, pp. 4861-4876.

[108] Li, W., Pu, X., Tsung, F., and Xiang, D. (2017), "A robust self-starting spatial rank multivariate EWMA chart based on forward variable selection", *Computers & Industrial Engineering*, 103, pp. 116-130.

[109] Li, Z., and Wang, Z. (2010), "Adaptive CUSUM of the Q chart", *International Journal of Production Research*, 48, 5, pp. 1287-1301.

[110] 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, pp. 4537-4553.

[111] Liu, L., Tsung, F., and Zhang, J. (2014), "Adaptive nonparametric CUSUM scheme for detecting unknown shifts in location", *International Journal of Production Research*, 52, 6, pp. 1592-1606.

[112] Liu, L., Zhang, J., and Zi, X. (2015), "Dual nonparametric CUSUM control chart based on ranks", *Communications in Statistics-Simulation and Computation*, 44, 3, pp. 756-772.

[113] Liu, L., Zi, X., Zhang, J., and Wang, Z. (2013), "A sequential rank-based nonparametric adaptive EWMA control chart", *Communications in Statistics-Simulation and Computation*, 42, 4, pp. 841-859.

[114] Lombard, F., and van Zyl, C. (2018), "Signed sequential rank CUSUMs", *Computational Statistics & Data Analysis*, 118, pp. 30-39.

[115] Lorden, G. (1971), "Procedures for reacting to a change in distribution", *The Annals of Mathematical Statistics*, 42, 6, 1897-1908.

[116] Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992), "A multivariate exponentially weighted moving average control chart", *Technometrics*, 34, 1, pp. 46-53.

[117] Lucas, J. M. and Crosier, R. B. (1982), "Fast Initial Response for CUSUM Quality-Control Schemes: Give your CUSUM a Head Start" *Technometrics*, 24, 3, pp. 199-205.

[118] Maboudou-Tchao, E. M., and Hawkins, D. M. (2011), "Self-starting multivariate control charts for location and scale", *Journal of Quality Technology*, 43, 2, pp. 113-126.

[119] Madrid Padilla, O. H., Athey, A., Reinhart, A. and Scott, J. G. (2019), "Sequential nonparametric tests for a change in distribution: an application to detecting radiological anomalies", *Journal of the American Statistical Association*, 114, 256, pp. 514-528.

[120] Makis, V. (2008), "Multivariate Bayesian control chart", *Operations Research*, 56, 2, pp. 487-496.

[121] Makis, V. (2009), "Multivariate Bayesian process control for a finite production run", *European Journal of Operational Research*, 194, 3, pp. 795-806.

[122] Mardia, K. V., and Jupp, P. E. (2009). "Directional statistics", *John Wiley & Sons*, New York.

[123] Marques, P. A., Cardeira, C. B., Paranhos, P., Ribeiro, S., and Gouveia, H. (2015), "Selection of the most suitable statistical process control approach for short production runs: a decision-model", *International Journal of Information and Education Technology*, 5, 4, pp. 303.

[124] Mason, R. L., and Young, J. C. (2002), "Multivariate statistical process control with industrial applications", *Society for Industrial and Applied Mathematics*.

[125] McDonald, D. (1990), "A CUSUM procedure based on sequential ranks", *Naval Research Logistics (NRL)*, 37, 5, pp. 627-646.

[126] Mei, Y. (2006), "Sequential change-point detection when unknown parameters are present in the pre-change distribution", *The Annals of Statistics*, 34, 1, pp. 92-122.

[127] Montgomery, D. C. (2020), *Introduction to Statistical Quality Control*, Eighth edition, Wiley, New York.

[128] Moustakides, G. V. (1986). "Optimal Stopping Times for Detecting Changes in Distributions", *Annals of Statistics*, 14, 4, pp. 1379-1387.

[129] Noor, S., Noor-ul-Amin, M., and Abbasi, S. A. (2020), "Bayesian EWMA control charts based on Exponential and transformed Exponential distributions", *Quality and Reliability Engineering International*.

[130] Noor-ul-Amin, M., and Noor, S. (2021), "An adaptive EWMA control chart for monitoring the process mean in Bayesian theory under different loss functions", *Quality and Reliability Engineering International*, 37, 2, pp. 804-819.

[131] Oakland, J. S. (2019), *Statistical process control*, Seventh edition, Routledge, London.

[132] Olwell, D. H. (1996), "Predictive Quality Control Charts", *US Army Conference on Applied Statistics*, pp. 67.

[133] Olwell, D. H. (1997), "Predictive Statistical Process Control for the Military", *US Army Conference on Applied Statistics*, pp. 70.

[134] Page, E. S. (1954), "Continuous inspection schemes", *Biometrika*, 41, 1/2, pp. 100-115.

[135] Paynabar, K., Zou, C., and Qiu, P. (2016), "A change-point approach for phase-I analysis in multivariate profile monitoring and diagnosis" *Technometrics*, 58, 2, pp. 191-204.

[136] Pazhayamadom, D. G., Kelly, C. J., Rogan, E., and Codling, E. A. (2016), "Self-starting cumulative sum harvest control rule (SS-CUSUM-HCR) for status-quo management of data-limited fisheries" *Canadian Journal of Fisheries and Aquatic Sciences*, 73, 3, pp. 366-381.

[137] Pazhayamadom, D. G., Kelly, C. J., Rogan, E., and Codling, E. A. (2013), "Self-starting CUSUM approach for monitoring data poor fisheries" *Fisheries research*, 145, pp. 114-127.

[138] Pignatiello Jr, J. J., and Runger, G. C. (1990). "Comparisons of multivariate CUSUM charts", *Journal of quality technology*, 22, 3, 173-186.

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

[140] Pollak, M., and Siegmund, D. (1991), "Sequential detection of a change in a normal mean when the initial value is unknown" *The Annals of Statistics*, 19, 1, pp. 394-416.

[141] Pollak M. and Tartakovsky A. (2009). "Optimality properties of the Shiryaev-Roberts procedure", *Statistical Sinica*, 8, 4, 1729-1739.

[142] Qiu, P. (2014), *Introduction to Statistical Process Control*, CRC Press, Chapman & Hall.

[143] Qiu, P., and Li, Z. (2011), "Distribution-free monitoring of univariate processes", *Statistics & probability letters*, 81, 12, pp. 1833-1840.

[144] Quesenberry, C. P. (1991a), "SPC Q Charts for Start-Up Processes and Short or Long Runs" *Journal of Quality Technology*, 23, 3, pp. 213-224.

[145] Quesenberry, C. P. (1991b), "SPC Q Charts for a Binomial Parameter p: Short or Long Runs" *Journal of Quality Technology*, 23, 3, pp. 239-246.

[146] Quesenberry, C. P. (1991c), "SPC Q Charts for a Poisson Parameter: Short or Long Runs" *Journal of Quality Technology*, 23, 4, pp. 296-303.

[147] Quesenberry, C. P. (1995a), "On properties of Q charts for variables" *Journal of Quality Technology*, 27, 3, pp. 184-203.

[148] Quesenberry, C. P. (1995b), "On properties of Binomial Q charts for attributes" *Journal of Quality Technology*, 27, 3, pp. 204-213.

[149] Quesenberry, C. P. (1995c), "On properties of Poisson Q charts for attributes" *Journal of Quality Technology*, 27, 4, pp. 293-303.

[150] Quesenberry, C. P. (1995d), "Geometric Q charts for high quality processes" *Journal of Quality Technology*, 27, 4, pp. 304-315.

[151] Quesenberry, C. P. (1997), *SPC methods for quality improvement*, Wiley, New York.

[152] Quesenberry, C. P. (2000), "Statistical process control geometric Q-chart for nosocomial infection surveillance" *American Journal of Infection Control*, 28, 4, pp. 314-320.

[153] Quesenberry, C. P. (2001), "The multivariate short-run snapshot Q chart" *Quality Engineering*, 13, 4, pp. 679-683.

[154] Ravichandran, J. (2019), "Self-starting X-bar control chart based on Six Sigma quality and sometimes pooling procedure" *Journal of Statistical Computation and Simulation*, 89, 2, pp. 362-377.

[155] Reynolds Jr, M. R. (1975), "A sequential signed-rank test for symmetry", *The Annals of Statistics*, 3, 2, pp. 382-400.

[156] Ritov, Y. A. (1990). "Decision theoretic optimality of the CUSUM procedure", *The Annals of Statistics*, pp. 1464-1469.

[157] Roberts, S. W. (1959), "Control chart tests based on geometric moving averages", *Technometrics*, 1, 3, 239-250.

[158] Roberts, S. W. (1966), "A Comparison of Some Control Chart Procedures", *Technometrics*, 8, 3, 411-430.

[159] Roes, K. C., Does, R. J., and Jonkers, B. S. (1999), "Effective application of $Q(R)$ charts in low-volume manufacturing", *Quality and Reliability Engineering International*, 15, 3, pp. 175-190.

[160] Ross, G. J., Tasoulis, D. K., and Adams, N. M. (2011), "Nonparametric monitoring of data streams for changes in location and scale", *Technometrics*, 53, 4, pp. 379-389.

[161] Ryan, A. G. and Woodall, W. H. (2010), "Control Charts for Poisson Count Data with Varying Sample Sizes", *Journal of Quality Technology*, 42, 3, pp. 260-275.

[162] Shen, X., Tsui, K. L., Zou, C. and Woodall, W. H. (2016), "Self-Starting Monitoring Scheme for Poisson Count Data with Varying Population Sizes", *Technometrics*, 58, 4, pp. 460-471.

[163] Shewhart, W. A. (1926), "Quality control charts", *The Bell System Technical Journal*, 5, 4, pp. 593-603.

[164] Shiryaev, A. (1963), "On optimum methods in quickest detection problems", *Theory of probability and its applications*, 8, 1, 22-46.

[165] Šidák, Z. K. (1967), "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions", *Journal of the American Statistical Association*, 62, pp. 626-633.

[166] Siegmund, D., and Venkatraman, E. S. (1995), "Using the generalized likelihood ratio statistic for sequential detection of a change-point", *The Annals of Statistics*, 23, 1, pp. 255-271.

[167] Smith, A. F. M. (1975), "A Bayesian approach to inference about a change-point in a sequence of random variables", *Biometrika*, 62, 2, pp. 407-416.

[168] Snoussi, A., Ghourabi, M. E., and Limam, M. (2005), "On SPC for short run autocorrelated data", *Communications in Statistics-Simulation and Computation*, 34, 1, pp. 219-234.

[169] Sobas, F., Jousselme, E., Bourazas, K., Geay-Baillat, M. O., Beghin, M., Nougier, C., and Tsiamyrtzis, P. (2020), "Estimation of uncertainty in measurement interest of short-term Bayesian model as a complement to the conventional approach", *Blood Coagulation & Fibrinolysis*, 14, 7, pp. 492-495.

[170] Sobas, F., Tsiamyrtzis, P., Benattar, N., Lienhart, A., and Négrier, C. (2014), "A comparison of the 1_{2s} rule and Bayesian approach for quality control: application to one-stage clotting factor VIII assay", *Blood Coagulation & Fibrinolysis*, 25, 6, pp. 634-643.

[171] Song, H., Xu, R., and Wang, C. (2020), "Research on statistical process control method for multi-variety and small batch production mode", *In 2020 Chinese Control And Decision Conference (CCDC)*, IEEE, pp. 2377-2381.

[172] Srivastava, M. S., and Worsley, K. J. (1986), "Likelihood ratio tests for a change in the multivariate normal mean", *Journal of the American Statistical Association*, 81, 393, pp. 199-204.

[173] Steiner, S. H. (1999), "EWMA Control Charts with Time-Varying Control Limits and Fast Initial Response" *Journal of Quality Technology*, 31, 1, pp. 75-86.

[174] Sullivan, J. H., and Jones, L. A. (2002), "A self-starting control chart for multivariate individual observations", *Technometrics*, 44, 1, pp. 24-33.

[175] Sun, D., and Berger, J. O. (2007), "Objective Bayesian analysis for the multivariate normal model" *Bayesian Statistics*, 8, pp. 525-562.

[176] Sun, D., and Berger, J. O. (2008), "Objective priors for the bivariate normal model" *Annals of Statistics*, 36, 6, pp. 963-982.

[177] Taleb, H., and Arfa, H. B. (2012), "Self-Starting Multivariate Control Chart", *Third Meeting on Statistics and Data Mining*, 103-108.

[178] Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014), *Sequential analysis: Hypothesis testing and changepoint detection*, CRC Press.

[179] Tercero-Gómez, V. G., Cordero-Franco, A., Pérez-Blanco, A., and Hernández-Luna, A. (2014), "A Self-Starting CUSUM Chart Combined with a Maximum Likelihood Estimator for the Time of a Detected Shift in the Process Mean". *Quality and Reliability Engineering International*, 30, 4, pp. 591-599.

[180] Theroux, E., Galarneau, Y., and Chen, M. (2014), "Control Charts for Short Production Runs in Aerospace Manufacturing". *SAE International Journal of Materials and Manufacturing*, 7, 1, pp. 65-72.

[181] Toubia-Stucky, G., Liao, H., and Twomey, J. (2012), "A sequential Bayesian cumulative conformance count approach to deterioration detection in high yield processes". *Quality and Reliability Engineering International*, 28, 2, pp. 203-214.

[182] Triantafyllopoulos, K. (2006), "Multivariate control charts based on Bayesian state space models", *Quality and Reliability Engineering International*, 22, 6, 693-707.

[183] Tsiamyrtzis, P. and Hawkins, D. M. (2005), "A Bayesian scheme to detect changes in the mean of a short run process", *Technometrics*, 47, 4, pp. 446-456.

[184] 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, pp. 721-735.

[185] Tsiamyrtzis, P. and Hawkins, D. M. (2010), "Bayesian start up phase mean monitoring of an autocorrelated process that is subject to random sized jumps", *Technometrics*, 52, 4, pp. 438-452.

[186] Tsiamyrtzis, P. and Hawkins, D. M. (2019), "Bayesian Statistical process control for Phase I count type data", *Applied Stochastic Models in Business and Industry*, 35, pp. 766-787.

[187] Tsiamyrtzis, P., Sobas, F., and Négrier, C. (2015), "Use of prior manufacturer specifications with Bayesian logic eludes preliminary phase issues in quality control: an example in a hemostasis laboratory", *Blood Coagulation & Fibrinolysis*, 26, 5, pp. 590-596.

[188] Villanueva-Guerra, E. C., Tercero-Gómez, V. G., Cordero-Franco, A. E. and Conover, W. J. (2017), "A control chart for variance based on squared ranks", *Journal of Statistical Computation and Simulation*, 87, 18, pp. 3537-3562.

[189] Wang, Y., Hu, X., Zhou, X., Qiao, Y., and Wu, S. (2020), "New One-Sided EWMA t Charts without and with Variable Sampling Intervals for Monitoring

the Process Mean", *Mathematical Problems in Engineering*, 2020.

[190] Wang, X., Nott, D. J., Drovandi, C. C., Mengersen, K. and Evans, M. (2018), "Using History Matching for Prior Choice", *Technometrics*, 60, 4, pp. 445-460.

[191] Wasserman, G. S. (1995), "An adaptation of the EWMA chart for short run SPC", *International journal of production research*, 33, 10, pp. 2821-2833.

[192] Wasserman, G. S. (1994), "Short run SPC using dynamic control chart", *Computers & industrial engineering*, 27, 1-4, pp. 353-356.

[193] Wasserman, G. S., and Sudjianto, A. (1993), "Short run SPC based upon the second order dynamic linear model for trend detection", *Communications in Statistics-Simulation and Computation*, 22, 4, pp. 1011-1036.

[194] West, M. (1986), "Bayesian model monitoring", *Journal of the Royal Statistical Society: Series B (Methodological)*, 48, 1, pp. 70-78.

[195] West, M. and Harrison, P. J. (1996), "Bayesian forecasting", *Encyclopedia of Statistical Sciences*.

[196] West, M. and Harrison, P. J. (1986), "Monitoring and adaptation in Bayesian forecasting models", *Journal of the American Statistical Association*, 81, 395, pp. 741-750.

[197] Woodall, W. H., and Montgomery, D. C. (1999), "Research issues and ideas in statistical process control", *Journal of Quality Technology*, 31, 4, pp. 376-386.

[198] Worsley, K. J. (1979), "On the likelihood ratio test for a shift in location of normal populations", *Journal of the American Statistical Association*, 74, 366a, pp. 365-367.

[199] Worsley, K. J. (1986), "Confidence regions and tests for a change-point in a sequence of exponential family random variables", *Biometrika*, 73, 1, pp. 91-104.

[200] Xia, Z., and Tsung, F. (2019), "A computationally efficient self-starting scheme to monitor general linear profiles with abrupt changes", *Quality Technology & Quantitative Management*, 16, 3, pp. 278-296.

[201] Xue, L., and Qiu, P. (2020), "A nonparametric CUSUM chart for monitoring multivariate serially correlated processes", *Journal of Quality Technology*, pp. 1-14.

[202] Yang, K., and Qiu, P. (2021), "Design variable-sampling control charts using covariate information", *IISE Transactions*, pp. 11-53.

[203] Yu, J., Kim, S. B., Bai, J., and Han, S. W. (2020), "Comparative Study on Exponentially Weighted Moving Average Approaches for the Self-Starting Forecasting", *Applied Sciences*, 10, 20, pp. 7351

[204] Zamba, K. D., and Hawkins, D. M. (2006), "A multivariate change-point model for statistical process control", *Technometrics*, 48, 4, pp. 539-549.

[205] Zamba, K. D., Tsiamyrtzis, P., and Hawkins, D. M. (2013), "A three-state recursive sequential Bayesian algorithm for biosurveillance", *Computational Statistics & Data Analysis*, 58, pp. 82-97.

[206] Zantek, P. F. (2005), "Run-length distributions of Q-chart schemes", *IIE transactions*, 37, 11, pp. 1037-1045.

[207] Zantek, P. F. (2006), "Design of cumulative sum schemes for start-up processes and short runs", *Journal of Quality Technology*, 28, 4, pp. 365-375.

[208] Zantek, P. F. (2008), "A Markov-chain method for computing the run-length distribution of the self-starting cumulative sum scheme", *Journal of Statistical Computation and Simulation*, 78, 5, pp. 463-473.

[209] Zantek, P. F., and Nestler, S. T. (2009), "Performance and properties of Q-Statistic monitoring schemes", *Naval Research Logistics (NRL)*, 56, 3, pp. 279-292.

[210] Zantek, P. F., Wright, G. P., and Plante, R. D. (2006), "A self-starting procedure for monitoring process quality in multistage manufacturing systems", *Iie Transactions*, 38, 4, pp. 293-308.

[211] Zellner, A. (1988), "Optimal Information Processing and Bayes's Theorem", *The American Statistician*, 42, 4, pp. 278-280.

[212] Zeng, L., and Zhou, S. (2011), "A Bayesian approach to risk-adjusted outcome monitoring in healthcare", *Statistics in medicine*, 30, 29, pp. 3431-3446.

[213] Zhang, M., Megahed, F. M. and Woodall, W. H. (2014), "Exponential CUSUM charts with estimated control limits", *Quality and Reliability Engineering International*, 30, 2, 275-286.

[214] Zhang, M., Peng, Y., Schuh, A., Megahed, F. M. and Woodall, W. H. (2013), "Geometric charts with estimated control limits", *Quality and Reliability Engineering International*, 29, 2, 209-223.

[215] Zhang, C. W., Xie, M., and Jin, T. (2012), "An improved self-starting cumulative count of conforming chart for monitoring high-quality processes under group inspection", *International Journal of Production Research*, 50, 23, pp. 7026-7043.

[216] Zhang, C. W., Ye, Z., and Xie, M. (2017), "Monitoring the shape parameter of a Weibull renewal process", *IISE Transactions*, 49, 8, pp. 800-813.

[217] Zou, C., Jiang, W., and Tsung, F. (2011), "A lasso-based diagnostic framework for multivariate statistical process control", *Technometrics*, 55, 3, pp. 297-309.

[218] Zou, C., Zhou, C., Wang, Z., and Tsung, F. (2007), "A self-starting control chart for linear profiles", *Journal of Quality Technology*, 39, 4, pp. 364-375.