State Space Models in Medical Time Series

Mohamed M. Shakandli

School of Mathematics and Statistics

University of Sheffield

A thesis submitted for the degree of

Doctor of Philosophy

Supervisor: Dr. Kostas Triantafyllopoulos

Co-supervisor: Prof. Michael Campbell

January 2018

Acknowledgements

Firstly, I am extremely grateful to my supervisors Dr. Kostas Triantafyllopoulos and Professor Michael J Campbell for their guidance, encouragement, advice and support throughout my PhD study. I would also like to thank deeply all the postgraduate students and the staff in the Department of Mathematics and Statistics at University of Sheffield for making the four years of the period of study one of the most interesting experiences of my life. Special praise must be reserved for my family, especially my wife, my father, my brother and sisters for their personal support.

Abstract

This thesis concerns the set-up and application of a state space model to medical time series. Considering medical count time series (such as number of asthma patients or a number of sudden infant death syndrome recorded over time), we discuss and propose non-linear and non-Gaussian state space models, in particular dynamic generalized linear models (DGLMs). Sequential Monte Carlo methods, also known as particle filters are employed for tracking a posterior state distribution. We assess the proposed methodology by way of an extensive simulation experiment. In the first simulation study, we found that the results from the Liu and West particle filter algorithm have shown better performance over the Storvik particle filter algorithm in terms of precision of the estimation of hyper-parameters and accuracy of forecasting. Beside, the obtained results from the Liu and West particle filter algorithm are quite similar from the ones that were obtained by the MCMC. In addition, in the second simulation study, we found the Liu and West particle filter algorithm with the Poisson model still does better than the other proposed models, even if it is incorrectly specified. The Smith (1985) method for model diagnostics is used. The results obtained from simulation studies showed that this methodology was successful. Finally, we developed a Bayesian monitoring model for evaluating the performance of the fitted model in a sequential way by using the Bayes factors and nonparametric binomial control chart with proposed runs rules. The novelty of this approach is to exploit the results obtained from the PSR or INTPSR for the model diagnostics to calculate the values of the Bayes factors.

We found that the proposed control procedure provided an effective way of detecting out-of-control signals of the process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem statement and motivation

Time series analysis is considered as an important statistical method which can be used to explain behaviour occurring during a specific period of time, where the time series data of this phenomenon are collected over time. This analysis aims to obtain an accurate description of the time series data through interpreting the dependence structure between the successive observations of the phenomenon. A suitable model can be built based on this dependence structure for explaining the behaviour of the phenomenon. This model can then be used to forecast what changes will be happening to the underlying phenomenon behaviour in the future. The process that is used to generate the time series under investigation might satisfy the stationarity requirement: this means that the features of the process, such as a mean and variance, do not change over time. However, in practice, especially in medical applications, the time series under consideration may not satisfy this requirement. Therefore, the difference method on the original time

series is used in order to delete the effects of the trend and seasonality components from the time series. Thus, the time series process will be stationary with a constant mean and variance over time.

In medical research, there are many time series methods that focus on Gaussian models whose observations are continuous and follow a normal distribution. However, in practice, there are several medical time series data under study comprising counts. These count data are discrete and they follow one of the discrete distributions – for instance, the count data may be the daily hospital contacts for patients who suffer from particular diseases or the occurrence of rare infections over time. Data analysis in medical research is often required whether by regression methods, such as multiple regression models or generalised linear models, or by classical time series models. However, not much work has taken place in medical research when model parameters change over time as the following point of view "The class as so far defined is basically static but there are many applications that require a dynamic formulation, in which parameters change over time, and covariance structures characteristic of time-series are needed. We have as yet no practical experience of fitting these dynamic Hierachical Generalized Linear models (HGLMs), and much work remains to be done; the prospects, however, are exciting" (Nelder, 1998, p.2752). For these reasons, the purpose of this thesis is to apply the Bayesian dynamic models with their recursive estimation methods for unknown parameters to medical time series count data.

The state space model (SSM) has become one of the standard parametric modelling forms with parameters changing over time in the Bayesian time series framework (West and Harrison, 1997). It is defined as a class of probabilistic graphical models that describe dynamic relationships between the observed variable and the unobserved (latent, state or hidden) variable using two parallel linked

equations. The first equation is called the observational equation, and describes the relationship between a sequence of time series data and states at specific time points. The second equation is a state equation, which describes an evolution in the latent states between current and previous times. The structure of the transition equation is based on the Markovian property, where the vectors of states change linearly over time. The SSM provides a unified and flexible framework for describing, modelling and forecasting a wide array of time series data whose parameters change over time. Moreover, given the state at a specific time point, the observation at this specific time point is independent of all other observations at different times. In other words, the state at a specific time point carries all the known information about the future of the observation at the same time point. Therefore, it is not necessary to describe the whole of the model from the start until the end for all time; rather, only the conditional distribution of observations given the state at a specific time moment is needed in order to make a forecast. This is a very attractive reason for using the SSM.

In terms of modelling the data by the Gaussian SSM, a recursive Bayesian method, based on a Kalman filter algorithm by Kalman (1960), provides an optimal solution of the state inference when a new observation is available. This means that the expression of the posterior distribution of the hidden variable can be obtained in a closed form. However, in some situations when the process of the data is neither linear nor Gaussian, there is no exact analytic expression of the state posterior distribution. In this case, other Bayesian algorithms can be used to construct an approximate solution of the conditional posterior distribution of the state. West et al. (1985) introduce the dynamic generalised linear model (DGLM) as an extension of the Gaussian SSM where the probabilistic distribution of the observed variable is a member of the exponential family. The sequential Monte Carlo (SMC) methods, also known as particle filters (PF), are

an alternative class of recursive Bayesian algorithms that can be used to infer the states and hyper-parameters in the non-linear and non-Gaussian SSMs (Durbin and Koopman, 2000; Doucet et al., 2000; Kitagawa, 1996; Liu and Chen, 1998). They are defined as simulation-based methods. The key idea behind the particle filter is to recursively over time provide an approximate solution of the posterior distribution of the current value of the state when a new observation arrives. This can be achieved by using a set of particles with associated weights that are simulated from a convenient importance density. The crucial issue when implementing the PF is to choose the importance density that plays a key role in the prediction efficiency. Gordon et al. (1993) suggested using the transition distribution as the suboptimal choice of the importance density in order to generate the particles. This choice often leads to poor performance because it does not take into consideration the current value of the observations. The conditional distribution of the current value of the state given the previous state and the current observation is defined as the optimal choice of the proposed density. The optimal choice is designed for minimising the variance of importance weights (Doucet et al., 2000). Thus, the effective sample size, which is used as a measure to monitor the problem of particle degeneracy in the algorithm, will be maximised.
This thesis is concerned with recursive Bayesian estimation methods of the parameters of time series for count data in medical research.

1.2 Aim and objective of the thesis

The preliminary aim of this thesis is to apply the DGLM to fit a univariate counting medical time series data related to asthmatic children in order to produce one-step-ahead forecasting. In addition, the DGLM is applied to investigate

the effect of environmental temperature on the sudden deaths of infants. The following objectives have been determined in order to achieve the required aims:

- The Liu and West (2001) and Storvik (2002) particle filter algorithms are used as recursive Bayesian methods for estimating the marginal posterior distributions of the latent state and static parameters for the DGLM.
- A model diagnostic with Smith (1985) method including the P-scores residuals and the inverse transformed P-scores residuals which are calculated based on an one-step-ahead predictive distribution for the fitted model is applied to evaluate the proposed fitted model.
- We will develop the Bayesian monitoring model in order to assess the accuracy of the fitted model in a recursive manner. Bayes factors and statistical process control are used as tools to achieve this target.

1.3 Outline of the thesis

The remaining five chapters of the thesis are organised as follows:

- **Chapter 2** provides the literature review relevant to the work in this thesis in two main sections. In the first section the background literature on classical time series analysis, SSMs and DGLMs are presented. It also presents recursive Bayesian methods for estimating the value of the state, including the Kalman filter and extended Kalman filter. In the second section, the background theory related to basic sequential Monte Carlo methods that can be used for approximating the posterior distribution of the state and the hyper-parameters is provided. A brief review of the Markov

chain Monte Carlo (MCMC) methods with the Metropolis-Hasting and the Gibbs sampler algorithms are also given in this chapter.

- **Chapter 3** presents necessary aspects of the principal concepts concerning online Bayesian approaches which are based on the Liu and West PF and the Storvik PF, to learn about the joint posterior distribution of the state and the hyper-parameters in the DGLMs. In the rest of this chapter, we apply these algorithms to simulated data and two medical datasets related to medical contacts for school children who suffer from asthma and sudden infant death syndrome in the UK.

- **Chapter 4** focuses on a methodology of the diagnostic checks of the non-Gaussian time series models. It presents the details of the Smith method by using the p.score residuals and the inverse transformed P-scores residuals as a tool for the model diagnostic analysis. The applications of the model diagnostics of the proposed fitted models on both the simulated data and the medical data are provided.

- **Chapter 5** provides a background and literature review for the related fields of the Bayesian monitoring model, including statistical process control, control charts and Bayes factor (BF). The applications of automatic diagnostic methods on both the simulated data and the medical data are offered in this chapter.

- **Chapter 6** presents a summary and concludes the main results of the thesis. Additionally, some possible directions for future work regarding the topics assessed in this thesis are suggested.

Chapter 2

Background and relevant literature review

2.1 Introduction

This chapter aims to review the literature relevant to the thesis in two main sections. This first section provides background literature on classical time series analysis and the state space models. It also includes inference about parameters, using an approximate Bayesian method, the Kalman filter, and extended Kalman filter. In the second section, the literature review focuses on estimating the latent variables (states) in the non-Gaussian state space models by using sequential Monte Carlo methods. A brief review of the Markov chain Monte Carlo (MCMC) methods with the Metropolis-Hasting and the Gibbs sampler algorithms are also provided in this chapter.

2.2 Classical time series analysis

Time series analysis is considered an important technique in statistics. It is used to study the behaviour of a particular phenomenon over specific time periods. The objective of time series analysis is to obtain an accurate description of process features that can be used to generate the time series under study. This leads to a model used to explain the behaviour of the time series. Hence the created model can be used to predict the series behaviour in the future based on a previous time series of the observed data. Additionally, the time series analysis can be used to control the process, by examining what happens when certain parameters in the model are changed and then generating the series. In order to achieve this, a thorough analytical study is required of time series models based on statistical and mathematical methods.

A time series is defined as a collection of observations describing a certain phenomenon recorded over time. Any type of time series may be described in terms of three components: trend, seasonal component, and irregular fluctuation. There are two kinds of time series: continuous and discrete according to the determination of time. In order to implement the analysis of classical time series by using Box and Jenkins (1976), a stationarity condition of the time series is required. This means that probabilistic properties of the process, such as the mean and the variance, do not change over time and the covariance depends only on lag. This kind of stationarity condition is called stationarity of order two, or weak stationarity. The time series is said to be a weak stationary if its mean and variance do not change over time. A stationarity assumption is a basic requirement for applying a time series analysis. In practice, however many time series are not stationary, particularly with respect to economic and medical data.

In this case, taking a difference of an original time series may remove the effect of trend and the impact of the seasonal variations. As a result, the first two moments of time series data, the mean and the variance, become constant over time. The time series model is defined as a function that connects the value of the time series with its previous values and its error. The classical time series analysis consists of four sequential phases, beginning with the identification of the model, which is the most important stage. The second step is an estimate of the parameters of the model selected in the first step. Diagnosing the checking of the appropriateness of the fitted model, by using analysis of residuals, is the third step. The generated residuals by the fitted model must be achieved under the following assumptions: independence, not correlated, and normally distributed. The final stage, the main target of applying the time series analysis, is a prediction of future values of the phenomenon under study. For more details, the following informative sources should be consulted: Box and Jenkins (1976), Anderson (1976), Priestley (1981), and Montgomery et al. (1990).

2.3 Time series applications in medicine

2.3.1 Asthma and related work

Asthma is a chronic lung disorder that leads to inflammation in the airways. This disease is characterised by bouts of difficult breathing. These bouts can be more difficult at night, and they are often associated with a cough. Asthma attacks are common when there is a viral disease in the respiratory system as well as in particular seasons, especially autumn, and for some patients during the spring time. A family history is considered to be the main factor for developing asthma;

and it is the most common chronic disease for school-aged children. One child out of every ten experiences asthma during childhood and there continues to be an increase in cases of asthma throughout the world. However, most cases of childhood asthma improve and are rectified as the child grows older. According to statistics from the World Health Organization, globally there are 300 million people currently suffering from asthma. As a result, asthma is classified as one of the most serious public health problems. Several studies have described the seasonal pattern of admission to hospital for children with asthma, and have also explained the relationship between unscheduled medical contacts and the end of the summer holidays (Campbell et al., 1997; Storr and Lenney, 1989; Kimes et al., 2004; Grech et al., 2004; Johnston et al., 2005; Lincoln et al., 2006; Julious et al., 2007, 2011). Lincoln et al. (2006) used time-series analysis to describe the seasonal pattern of admissions to hospital for school-aged children who suffered from asthma in Sydney, Australia over a seven-year period between 1994 and 2000. They also verified whether there was a relationship between the end of the summer break and hospital admissions. The authors found that there was a strong relationship between the end of summer break and the return of children to their schools and the increase of hospital admissions for children with asthma.

Julious et al. (2007) illustrated peaks in the daily hospital admissions for school-aged children with asthma in two UK cities (Aberdeen and Doncaster) in early autumn. They provided strong evidence that there is a relationship between an entry of children to the hospital for treatment and their return to school after the summer holiday.

Julious et al. (2011) used a linear regression model to investigate whether there is a relationship between the daily unscheduled medical contacts for children, who either suffer or do not suffer from asthma, after their return to school after their

summer holiday in England and Scotland. In addition, they used the Lowess curve for both datasets in order to identify the trends. They discovered that the unexpected medical contacts for both data sets are increased when students come back from summer holiday to school in both England and Scotland. This is probably due to a decline in the use of preventive medicine such as inhaled corticosteroids over the summer period (Julious et al., 2011).

2.3.2 Medical time series analysis by state space model

The state space models (SSMs), which fall under the Bayesian time series framework, have been widely used to describe a dynamic mechanism in systems. They can be efficiently used as an attractive statistical tool in terms of modelling and prediction for phenomena that exhibit a dynamic behaviour. The SSMs were initially designed in research related to aerospace science. Their applications were then subsequently developed to include several disciplines such as engineering, economics, and finance. With respect to medical studies, there is a lack of current research focusing on the Bayesian time series and their estimation methods (Nelder, 1998). However, the state space approaches with Kalman recursive estimation methods have been adopted as a powerful tool to model the different medical data. Several medical studies, using this approach in the statistical analysis, are summarised below.

Hulin (2000) suggests the use of a state space framework to model the HIV epidemic. The purpose of his research was to use data on the homosexual population in San Francisco to estimate the number of incidences of AIDS cases. The number of people with infection at different phases was taken into consideration. The recursive Bayesian estimation technique based on the Kalman smoothing ap-

proach was used. This aims to obtain the optimal estimate of the latent variable (state) at each time point. Hulin (2000) concluded that the SSM can successfully provide an accurate estimation of HIV incidences. This means that the predicted case numbers of HIV are very close to the actual case numbers. Another finding was that the estimated numbers of infected homosexuals with the AIDS virus at different stages was a high percentage in San Francisco.

Wonga et al. (2006) adopted state space methodology to model a time series process related to electroencephalogram (EEG) data. A linear autoregressive model was employed to represent a non-stationary EEG time series within the framework of the state space. The objective of their study was to describe a transition stage of a clinical case of the electroencephalogram for patients from a state of consciousness to an anaesthesia state. To achieve this, they used the same EEG data as in John (2002). A maximum likelihood method and the Kalman recursion approach were applied to determine hyper-parameters and optimal filtering of the state of the dynamic linear model. Researchers concluded that a real-time monitoring of the anesthesia status of patients during surgery can be accurately determined by using the dynamic linear models.

Myers et al. (2007) proposed using the SSM to correct corruptions of optical imaging data. In their study, the data, collected by Seong-Gi Kim's research group at the University of Pittsburgh's Magnetic Resonance Research Center in the Department of Radiology, is concerned with the signal of the activations of the brain. The corruption of the data is caused by artifacts such as the difficulty of interpreting a signal of brain activity. The cycles of respiration and heartbeat are considered the main causes of signal corruption in the optical imaging of the brain. The unknown parameters of the dynamic linear model are estimated by the maximum likelihood method. Additionally, the recursive Kalman filter approach

is employed to track the brain activity level in each pixel of the video data. This is referred to as the unobserved variable. In terms of assessing the performance of the fitted model, the analysis of residuals is applied. The resulting residuals of the pixel data suggest that there is no evidence that respiration and heartbeat cause corruption in video data of the brain. Myers et al. (2007) concluded that the dynamic linear model with the Kalman filter are beneficial and powerful tools that help to correct data corruption caused by extraneous physiological processes such as respiration and heartbeat.

The dynamic generalised linear model (DGLM) is adopted by Christensen et al. (2012) to model the rate of hospitalisation for patients who suffered from strokes in a Danish hospital between 1977 and 2011. The aim of this study was to describe any variation of seasonality in stroke occurrences. The hospitalisation rates data are assumed to follow a Poisson distribution. In this case, the iterated extended Kalman smoothing, based on the first two derivatives of the observation process, was used in order to linearise the DGLM, as suggested by Durbin and Koopman (2012). Therefore the Kalman recursion approach can be used to estimate the state variables. The expectation maximisation algorithm (Dempster et al., 1977) is used to estimate the model hyper-parameters. The authors concluded that the DGLM is a useful technique for investigating the seasonal variations of the hospitalisation rates of patients who suffer from strokes. They found that the hospital admissions rate for patients increases in the winter and summer of each year.

Ullo et al. (2015) used an SSM to model the structure of the electrode connectivity of microscopy images. A novel data encoding regarding brain tractography was used. The particle filter approach was used to track the path of the unknown neurite on the image of brain fibres (neuron). In the context of the state space

framework it is referred to as the latent variable or the state. The approximate posterior distributions of the neurite paths on the image can be estimated. In terms of addressing the different problems in the bioimage analysis, Myatt et al. (2006), Pontabry and Rousseau (2011), Yang et al. (2013), Yap et al. (2011), and Yuan et al. (2012) are adopted, i.e. the concept of the particle filter for tracking the hidden variable of interest was used. Researchers concluded that the SSM with the particle filter algorithm was able to capture the neural network of the brain fibre in order to construct the electrode connectivity of the microscopy images.

2.4 State space model

The state space model (SSM) is an important development of time series analysis. It is considered to be a flexible and attractive methodology in a statistical framework for integrating a dynamic model and data for modelling dynamic time series processes. This model is given by two parallel linked processes that evolve over time and are represented as conditional distributions. It is useful to describe the dynamic system in the time series data. The term 'dynamic' is defined as a change which occurs in a process due to changes in time. The SSM was initially designed in research related to aerospace science, and subsequently developed its applications to include several areas such as engineering, economics, finance and medicine.

The SSM, in general, can be written in term of three probability distributions as follows:

$$x_t|x_{t-1} \sim p(x_t|x_{t-1}) \tag{2.4.1}$$

$$y_t|x_t \sim p(y_t|x_t) \tag{2.4.2}$$

$$x_0 \sim p(x_0) \tag{2.4.3}$$

Equation (2.4.1) is known as the state (system, evaluation, transition) equation that describes an evolution in the states between time t and previous time $t-1$. Additionally, the structure of the system equation is based on a Markovian property. This means that the state x_t at time t depends only on the previous state at time $t-1$. In other words, given x_{t-1}, x_t is conditionally independent of $x_{t-2}, x_{t-3}, \ldots$. That is the current state x_t given the previous state x_{t-1} is independent of the state histories, $p(x_t|x_{1:t-1} = \{x_1, ..., x_{t-1}\}) = p(x_t|x_{t-1})$. The second equation (2.4.2) is an observational equation that describes the relationship between a sequence of time series data $\{y_t\}$ and states $\{x_t\}$ at the same time t. The prior distribution of the initial state is given by Equation (2.4.3). The SSM is completely defined through specifying the prior distribution of this initial state, x_0. In addition, a probabilistic approach is used to describe both the available data and the evolution of the states. A graphical form of the SSM is given in Figure 2.1. It shows the dependence between the observations and the states in the SSM.

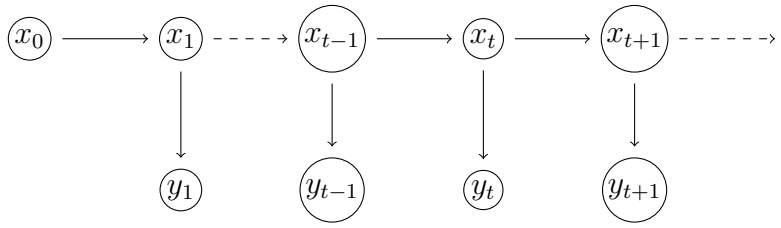

Figure 2.1: *Graphical representation of the state space model.*

Figure 2.1 provides a graphical view of the dependence relationship between the unobserved state and observed data in the SSM. It shows the conditional construction between the state and observation over time – i.e. that observation y_t at time t depends only on the state x_t at the same time. As a result, given

the current state x_t, the current observation y_t is conditionally independent of the rest of the observations and the history of the states. In other words, the state x_t at time t conveys all the required information concerning the future of the observation y_t. In addition, the Markovian structure of the states and the conditional independence structure of the observations can be easily noted from the graphical model. A popular subclass of model (2.4.1)–(2.4.3) is the following:

$$x_t = G(x_{t-1}) + \omega_t \tag{2.4.4}$$

$$y_t = F(x_t) + \nu_t, \tag{2.4.5}$$

where $F(.)$ and $G(.)$ are defined as state and observation functions respectively, ω_t is known as the state noise process and ν_t is defined as the observation noise process. For more details on this subject, refer to Harvey (1989); West and Harrison (1997); Shumway and Stoffer (2006); Durbin and Koopman (2012); Kitagawa and Gersch (1996).

2.4.1 Gaussian state space model

The Gaussian SSM is also referred to as the dynamic linear model (DLM). It is a class of SSM, a type of Bayesian time series model whose parameters vary with time. It can be obtained from the general SSM when the functions $f(.)$ and $g(.)$ are linear, and both noise processes ω_t and ν_t are Gaussian. In other words, when the relationship between observation and states, and between states, are linear. The class of Gaussian SSMs is deemed to be helpful in regression and time series analysis. They are able to supply a flexible and unified framework to describe and model a wide range of time series and other types of longitudinal data in different disciplines. According to Migon et al. (2005) the DLM can be seen as a generalisation of regression models, which enable changes in parameter

values throughout time by the introduction of an equation governing the temporal evolution of regression coefficients. The form of the Gaussian SSM is defined by the following set of equations:

$$y_t = F_t' x_t + \nu_t, \qquad \nu_t \sim N(0, V_t), \qquad t = 1, \ldots, T \tag{2.4.6}$$

$$x_t = G_t x_{t-1} + \omega_t \qquad \omega_t \sim N(0, W_t) \tag{2.4.7}$$

$$x_0 \sim N_p(m_0, C_0), \tag{2.4.8}$$

where

- y_t is a scalar (univariate) sequence of observed variables at time t,
- x_t is a p x 1 vector consisting of unobserved variables (states) at time t,
- F_t is a p x 1 design vector consisting of covariates at time t,
- G_t is a p x p transition matrix which describes the evolution in the states at time t,
- ω_t is the term of evolution error. It describes the stochastic changes in the unobserved variable (state); and follows a normal distribution with zero mean and covariance variance matrix W_t, and
- ν_t is a term of observational noise. It represents a measurement and sampling error corrupting the observation of y_t, assumed normally distributed with zero mean and variance V_t.

The error terms v_t and w_t at any time point t are defined as the white noise process: the two terms are mutually independent and are assumed not to be correlated with the state vector x_t . The initial information of the state x_0 provides the expert prior beliefs about state x_0: this is assumed to have a Normal distribution with m_0 mean and variance C_0. Thus from this preliminary information a modeller can determine the prior distribution of the initial state. Harrison and Stevens (1971) and Harrison and Stevens (1976) define the Gaussian SSM as in Equations (2.4.7)-(2.4.8). The four quantities $\{F_t, G_t, V_t, W_t\}$ are used for completely determining the DLM. The algorithm of the recursive Kalman filter is used for analytically inferring the states when the value of these quantities is known. West and Harrison (1997) describe this procedure in greater detail.

The analytical approach within DLMs is not available when the hyperparameters $\{V_t, W_t\}$ are not known. This issue can be solved by using several proposals for performing approximate inference in DLMs using Bayesian methods based on simulation approaches, such as Markov chain Monte Carlo (MCMC) methods and sequential Monte Carlo (SMC) methods, also known as particle filters (Gordon et al., 1993). The DLMs can be employed to model univariate time series. In addition, the ARIMA process, defined within the linear time series models by Box and Jenkins (1976), can be represented as a special case of the SSM when assuming the quantities $\{F_t, G_t, V_t, W_t\}$ do not change with time. The SSM is called a Hidden Markov model when the variables of state are discrete (Rabiner, 1989).

2.4.2 The dynamic generalised linear model

West et al. (1985) introduce the dynamic generalised linear model (DGLM) as an

extension case of the dynamic linear model where observations in a time series follow a parametric non-Gaussian distribution, but the probabilistic distribution of the observations is a member of the exponential family. In addition, the DGLM is defined as an extension of the generalised linear model (GLM), which is proposed by Nelder and Wedderburn (1972), when its parameters are changed over time. The equation of the state transition retains the features of the linear Markov model of order one. Conversely, it can be assumed that the DGLM is an extension of the generalised linear model that assumes the parameters vary stochastically over time. However, the state posterior distribution of the DGLM is not analytically available. Hence West et al. (1985) use an approximate Bayesian method with conjugate prior in order to obtain the recursive estimation of the posterior distribution of state. Simulation-based approaches such as MCMC and particle filter are also used for approximating the state posterior distribution. Gamerman (1998) employed MCMC for inference. Fahrmeir (1992) also defines the exponential family of the dynamic linear model, using the approximation of conditional modes based on the observations for estimating the unobserved states.

Suppose that the equation of observation of the univariate time series $\{y_t\}$, which is generated from a probability distribution belonging to the exponential family, is defined as follows:

$$f(y_t|\gamma_t, \phi_t) = \exp\{\frac{d(y_t)\gamma_t - b(\gamma_t)}{a(\phi_t)}\}c(y_t, \phi), \tag{2.4.9}$$

where γ_t is a natural parameter, and $c(y_t, \phi_t)$ is a known function of the observation y_t and the dispersion parameter ϕ_t. The function $d(.)$ is either represented as a simple linear function of the response y_t or, in many cases, as the identity function. The function $b(\gamma_t)$ is considered to be known and twice differentiable.

According to McCullagh and Nelder (1989), the expectation and the variance of $d(y_t)$ given the natural parameter γ_t are defined as the first and the second derivative of the function $b(\gamma_t)$:

$$\mu_t = E\{d(y_t)|\gamma_t\} = \partial b(\gamma_t)/\partial\gamma_t \tag{2.4.10}$$

$$Var\{d(y_t)|\gamma_t\} = a(\phi_t)\partial^2 b(\gamma_t)/\partial\gamma_t^2. \tag{2.4.11}$$

The distribution of the observations is known as the standard or canonical system when $d(y_t) = y_t$. The link function $g(.)$ which is used to connect the expectation of y_t and the linear predictor η_t is known as the continuous monotonic function. In addition, the link function is called the canonical link when $g(\mu_t) = \gamma_t$.

The linear predictor η_t in the GLM is considered as a linear model, while in the DGLM it is defined as an SSM, i.e.

$$g(\mu_t) = \eta_t = F_t' x_t \tag{2.4.12}$$

$$x_t = G_t x_{t-1} + \omega_t, \qquad \omega_t \sim (0, W_t), \tag{2.4.13}$$

where F_t is a $p \times 1$ design vector of covariates, G_t is a $p \times p$ transition matrix and w_t is a vector of innovation which follows a distribution that has zero mean and known variance W_t. The W_t is defined as the state variance.

The linear Bayesian approach is one of the first estimation methods used in the inference problem of the DGLM. It is considered as an approximate method for estimating the first and the second moments of the posterior state distribution in a sequential manner (West and Harrison, 1997). Triantafyllopoulos (2008) and Triantafyllopoulos (2009) presented diverse applications of the problems estimation and forecasting for several continuous and discrete response distributions

such as Poisson, Weibull, Gamma, and negative binomial in the DGLM framework.

2.5 Parameter estimation approaches

Statistical estimation methods are used in statistical inference to estimate unknown parameters x in the statistical models. There are two different kinds of estimation techniques, namely, a frequentist and a Bayesian approach. In the maximum likelihood estimation (MLE) method, considered as a classical frequentist estimation technique, the parameter x is assumed to be a constant. The concept of the MLE method is to estimate the unknown parameter x by maximising the joint probability density function of the observations with respect to x:

$$\hat{x} = \arg\max \left(\prod_{t=1}^{T} p(y_t|x)\right), \qquad t = 1, \ldots, T \tag{2.5.1}$$

In the Bayesian estimation method, however, the parameter is assumed to be a random variable with a specific probability distribution. The aim of the Bayesian estimation method is to infer the uncertainty of the parameter. This uncertainty can be formulated in terms of the posterior probability distribution of the parameter by using Bayes' theorem as follows:

$$p(x|y_1, \ldots, y_t) = \frac{p(y_1, \ldots, y_t|x)p(x)}{p(y_1, \ldots, y_t)}, \tag{2.5.2}$$

where $p(x|y_{1:t})$ is known as a posterior probability density function of the parameter of the interest, $p(y_t, \ldots, y_t|x)$ is called the likelihood function, and $p(x)$ is defined as the prior probability density function of the parameter x which represents the initial information or the prior beliefs about the parameter before seeing the data. The term in the denominator of Equation (2.5.2) is called a normalising

term. It does not depend on the parameter x . Thus it is often neglected in the calculation of the posterior of interest. In other words, the posterior density of interest is only known up to a normalising term. Therefore, in the Bayesian approach, the posterior distribution of the parameter can be formulated as follows:

$$p(x|y_1, \ldots, y_t) \propto p(y_1, \ldots, y_t|x)p(x). \tag{2.5.3}$$

In the next section, we provide a methodology of the recursive Bayesian method for calculating the posterior distribution of the state in the SSM.

2.5.1 Recursive Bayesian inference procedure in parameter estimation problem of the state space model

As mentioned above, the recursive Bayesian method can be involved in the methodology of the SSM in terms of the state estimation problem. It is considered as an online estimation method when the observations are available sequentially over time. The purpose of the Bayesian method within the SSM is to construct recursively in time the posterior distribution of the states given the observations up to time t, $p(x_{0:t}|y_{1:t})$. On the whole, the full posterior distribution of the states is given by

$$p(x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{p(y_{1:t})}, \tag{2.5.4}$$

where $x_{0:t} = \{x_0, \ldots, x_t\}$ and $y_{1:t} = \{y_1, \ldots, y_t\}$, and the other parameters are defined as above. In this thesis, interest lies in estimating the conditional marginal posterior distribution of the state $p(x_t|y_{1:t})$ at a time instant t in a recursive technique. This estimation method is referred to as Bayesian filtering, and consists of two basic steps, the prediction and update steps (Ristic et al., 2004). At the time

instant $t-1$, the posterior distribution of the state $P(x_{t-1}|y_{1:t-1})$ is assumed to be available. In this step, the knowledge of the state at time step $t-1$ is used to estimate the same state at the next time t. Additionally, the transition equation $p(x_t|x_{t-1})$ is involved in the prediction step for obtaining the prior (predictive) probability density function of the state, as follows:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}. \tag{2.5.5}$$

After acquiring a new state by using the prediction step, the next step is to correct and update this prediction when a new observation arrives. Bayes' theorem is used to do so by estimating the posterior probability density function of the state as follows:

$$p(x_t|y_{1:t}) = \frac{p(y_{1:t}|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}, \tag{2.5.6}$$

where the denominator represents the normalising term and is calculated as

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t. \tag{2.5.7}$$

As noted above, the posterior distribution of the state can be estimated only up to the normalising term, and therefore it can be calculated as follows:

$$p(x_t|y_{1:t}) \propto p(y_{1:t}|x_t)\, p(x_t|y_{1:t-1}). \tag{2.5.8}$$

After acquiring the marginal posterior distribution, the next task is to calculate the optimal estimation of the state. They can be formulated respectively as follows:

$$\hat{x_t} = \mathbb{E}(x_t|y_{1:t}) = \int x_t\, p(x_t|y_{1:t})dx_t \tag{2.5.9}$$

$$\hat{x_t} = \arg\max_{x_t}\, p(x_t|y_{1:t}), \tag{2.5.10}$$

where $\mathbb{E}(x_t|y_{1:t})$ is referred to as the conditional mean of the current state based on the observations $y_{1:t}$, and $\hat{x_t}$ in Equation (2.5.10) is that x_t which maximizes the posterior density, $p(x_t|y_{1:t})$.

The main task in dealing with SSMs is an inference problem relating to the state and hyper-parameters. The objective is to estimate the posterior probability density function of the state given the available observations. In practice, the sequence of the observations $y_{1:t}$ is known, whereas the sequence of the states $x_{1:t}$ and the hyper-parameters are unknown. There are several recursive Bayesian filtering algorithms that can be used to compute the posterior mean and variance of the state, such as the Kalman, the extended Kalman and a particle filter. In the following section, the focus is on background knowledge regarding recursive Bayesian estimation methods used in inferring the posterior distribution of the states for linear SSMs, and includes the Kalman filter and an extended Kalman filter.

2.5.2 Kalman filter

The Kalman filter is one of a number of recursive Bayesian estimation methods used in the problem of tracking the latent variables (states) of linear dynamic models. This is achieved by calculating the posterior distribution of the state. The Kalman filter, introduced by Kalman (1960) and Kalman and Bucy (1961), is a recursive and sequential algorithm consisting of a set of mathematical equations. It is used to infer the posterior distribution of the unobserved state given the observed information of the response variable. The Kalman filter is considered as an optimal filter when both equations in the SSM are linear and Gaussian. The posterior distribution obtained by the Kalman filter is to be Gaussian be-

cause of the assumptions of the dynamic linear model and the Gaussian prior distribution of the initial state x_0. Therefore, it is parameterised by the mean and the covariance of state. It aims to provide an optimal estimate of the state by minimising the mean square error (MMSE) of the estimate. The Kalman filter is widely employed in several fields such as statistics, economics and engineering. The reason for this is its simplicity in application, its robustness and the optimality of its estimate. There are two kinds of methodology in terms of solving the problem of estimation in the SSM: filtering and smoothing (Kitagawa and Gersch, 1996). The Kalman filter and Kalman smoother have different ways of dealing with an observation when performing the estimation. The Kalman filter takes into account for implementing the previous and the current observations, whereas the Kalman smoother also takes into account the future observations. West and Harrison (1997), (in Chapter 4), defined the Kalman filter as a filter for updating the states in the Gaussian state space model.

At each time point, the Kalman filtering algorithm consists of two steps, namely a prediction and an update step. In the prediction step, the mean $\hat{x}_{t|t-1} = \mathbb{E}(x_t|y_{1:t-1})$ and the variance $C_{t|t-1} = \mathbb{V}ar(x_t|y_{1:t-1})$ of the prior distribution of state $p(x_t|y_{1:t-1})$ are calculated. In the second (update) step, the observation y_t is used to correct the prior mean given in the first step and then estimate the first two moments of the posterior distribution of the current state $p(x_t|y_{1:t})$. These moments can be expressed in the form of the mean $\hat{x}_{t|t} = \mathbb{E}(x_t|y_{1:t})$ and the variance $C_{t|t} = \mathbb{V}ar(x_t|y_{1:t})$.

Under the observation Equation (2.4.6), the system Equation (2.4.7), and the initial information of the prior state (2.4.8), the recursive sequential Kalman filter equations with prediction and updating steps are described as follows.

(a) It is assumed that the posterior distribution of the vector of state x_{t-1} at time $t-1$ based on the set of information $y_{1:t-1} = \{y_1, ...y_{t-1}\}$ follows the Normal distribution with some mean m_{t-1} and C_{t-1} as the variance matrix,

$$x_{t-1}|y_{1:t-1} \sim \mathrm{N}(x_{t-1}; m_{t-1}, C_{t-1}). \tag{2.5.11}$$

(b) The distribution of prior (predictive distribution) for the state vector x_t at time t follows a Normal distribution with a_t and R_t as the mean vector and the variance matrix respectively,

$$x_t|y_{1:t-1} \sim \mathrm{N}(x_t; a_t, R_t), \tag{2.5.12}$$

where

$$a_t = G_t\, m_{t-1} \tag{2.5.13}$$

$$R_t = G_t C_{t-1} G_t' + W_t. \tag{2.5.14}$$

(c) Given the set of information $y_{1:t-1}$, the distribution of the one step forecast (predictive distribution) also follows the Normal distribution:

$$y_t|y_{1:t-1} \sim \mathrm{N}(f_t, Q_t), \tag{2.5.15}$$

where

$$f_t = F_t' a_t \tag{2.5.16}$$

$$Q_t = F_t' R_t F_t + V_t. \tag{2.5.17}$$

(d) After acquiring a new observation, the next step is to update the prior distribution in order to obtain the posterior distribution for state x. The posterior distribution for state x at time t, given the data $y_{1:t} = \{y_1, ..., y_t\}$ can be obtained by using Bayes' theorem through combining the likelihood

of the data and the prior of the state. Therefore, the posterior for the state x_t based on the set of information $y_{1:t}$ is

$$x_t|y_{1:t} \sim \mathrm{N}(m_t, C_t),$$

with

$$m_t = a_t + A_t e_t \tag{2.5.18}$$

$$C_t = R_t - A_t Q_t A_t', \tag{2.5.19}$$

where

$$A_t = R_t F_t Q_t^{-1} \tag{2.5.20}$$

$$e_t = y_t - f_t. \tag{2.5.21}$$

Here $A_t; (0 < A_t < 1)$ is defined as the Kalman gain or adaptive coefficient.

As noted in Equation (2.5.18), the mean of the posterior distribution of a state is calculated by the previous one and corrected by the predictive error. A closed-form solution to the posterior distributions of state is an attractive feature extracted from the Kalman filter algorithm. This means that the posterior distribution of the state is exactly Gaussian without the need for numerical approximations, due to the assumptions of the linear Gaussian model. Therefore the Kalman filter is considered the natural choice to provide the optimal solution of estimation through online inference in a Bayesian framework. For more details on a sequential inference for the state and parameters of the online Bayesian method with the Kalman filter, Harrison and Stevens (1971), Harrison and Stevens (1976), West and Harrison (1997), Petris et al. (2009), Triantafyllopoulos and Montana (2011), and Durbin and Koopman (2012) are useful sources. Table 2.1 provides

a summary of a sequential updating by the Kalman filter algorithm of the state in the dynamic linear model.

Table 2.1: Kalman filter algorithm

At time point t-1:

Assign the posterior state mean vector m_{t-1}, and covariance matrix C_{t-1}.

For time instant $t = 1, \ldots, T$:

Prediction:

Calculate the prior state mean vector a_t and covariance matrix R_t,

$$a_t = G_t\, m_{t-1},$$
$$R_t = G_t C_{t-1} G_t' + W_t$$

Update:

Compute posterior state mean vector m_t, and covariance matrix C_t when a new observation is available.

$$f_t = F_t'\, G_t\, m_{t-1},$$
$$Q_t = F_t' R_t F_t + V_t,$$
$$A_t = R_t F_t Q_t^{-1},$$
$$m_t = a_t + A_t(y_t - f_t),$$
$$C_t = R_t - A_t Q_t A_t'.$$

The question here is whether one or both functions $f(.), g(.)$ in Equations (2.4.4)– (2.4.5) are nonlinear. If one of them is nonlinear, the Kalman filter may fail to give a reasonable estimation of the posterior distribution to the states. For this reason, it is necessary to resort to using another filter to overcome this

problem. The extended Kalman filter and the particle filter have been adopted to achieve this aim.

2.5.3 The extended Kalman filter

As described in the previous section, the basic Kalman filter with linearity of the observations and evolution equations, and under the assumption of the Gaussian model, can be applied to obtain the optimal estimate of the state associated with the dynamic linear model. However, in most applications of interest, either the observation equation or the system process is nonlinear. In this case, to overcome these problems, the common alternative method to the Kalman filter is an extended Kalman filter (EKF). It is employed in order to overcome the problems of nonlinear functions through using numerical approximation methods.

The SSM with nonlinear equations can be written as:

$$x_t = f(x_{t-1}) + \omega_t, \qquad \omega_t \sim N(0, W_t) \tag{2.5.22}$$

$$y_t = g(x_t) + \nu_t, \qquad \nu_t \sim N(0, V_t), \tag{2.5.23}$$

where x_t and y_t are the vectors of the state and the observations respectively, $f(.)$ and $g(.)$ are nonlinear functions, and v_t and ω_t are the error terms with the zero mean white noise processes for the observation and state equations. The EKF is employed to avoid the linearity assumption of the model. However, it is considered as a sub-optimal approach for tracking the states in order to obtain an approximate estimate of the posterior probability density function. The analytic solution to the filtering problem in the nonlinear and non-Gaussian state space is not, however, available. The first order of the Taylor approximation around the current state mean is used to linearize the nonlinear model. The Jacobian of the nonlinear equations is used to implement this approximation. Therefore

the equations of the basic Kalman filter can be adopted for sequential updating. With respect to linearising the nonlinear functions $f(.)$, $g(.)$, the first-order Taylor expansion can be defined as follows:

$$f(x_{t-1}) \simeq f(\hat{x}_{t-1}) + \mathbb{G}_t(x_{t-1} - \hat{x}_{t-1}) \tag{2.5.24}$$

$$g(x_t) \simeq g(\hat{x}_t) + \mathbb{F}_t(x_t - \hat{x}_t), \tag{2.5.25}$$

where

$$\mathbb{G}_t = \frac{\partial f(x_{t-1})}{\partial x_{t-1}} \mid_{x_{t-1}=\hat{x}_{t-1}}, \qquad \mathbb{F}_t = \frac{\partial g(x_t)}{\partial x_t} \mid_{x_t=\hat{x}_t}, \tag{2.5.26}$$

$\hat{x}_{t-1} = \mathbb{E}(x_{t-1}|y_{1:t-1})$, and $\hat{x}_t = \mathbb{E}(x_t|y_{1:t-1})$.

The EKF is considered as a minimum mean square error (MMSE) estimator of the state of the nonlinear model. To be able to implement the EKF, the observation and state equations are required to be differentiable. This means that the first and second derivatives of both equations in the model are available. In this way, the required Jacobian matrices can be calculated. The EKF ensures that the posterior distribution of the state, $p(x_t|y_{1:t})$, will be approximated by Gaussian, $x_t|y_{1:t} \approx \mathrm{N}(x_t|m_t, \hat{C}_t)$. In this case, due to this approximation, the EKF does not provide the closed-form solutions to the posterior distribution of the states. This is why it is known as a sub-optimal filter. If the equations in the model are highly nonlinear and the assumption of linearity breaks down, the approximation to Gaussian which has been made will lead to large errors in the posterior mean and variance of the state. This may result in divergence or a slow convergence of the EKF. (For further details see (Jazwinski, 1973; Anderson and Moore, 1979; Ristic et al., 2004). A summary of the EKF algorithm is offered in Table 2.2

Table 2.2: Extended Kalman filter algorithm

At time point t-1:

Assign the posterior state mean vector m_{t-1}, and covariance matrix C_{t-1}.

For time instant $t = 1, \ldots, T$:

Prediction:

1. Calculate the Jacobian matrix $\mathbb{G}_t$, and evaluate it

   at $x_{t-1} = \hat{x}_{t-1} = \mathbb{E}(x_{t-1}|y_{1:t-1})$

2. Calculate the prior state mean vector a_t and covariance matrix R_t,

   $$a_t = f(\hat{x}_{t-1}),$$

   $$R_t = \mathbb{G}_t C_{t-1} \mathbb{G}_t + W_t$$

Update:

1. Calculate the Jacobian matrix $\mathbb{F}_t$, and evaluate it

   at $x_t = \hat{x}_t = \mathbb{E}(x_t|y_{1:t-1})$

2. Calculate the posterior state mean vector m_t, and covariance

   matrix C_t when a new observation is available.

   $$Q_t = \mathbb{F}_t R_t \mathbb{F}_t + V_t,$$

   $$A_t = R_t \mathbb{F}_t Q^{-1},$$

   $$m_t = a_t + A_t(y_t - g(\hat{x}_t)),$$

   $$C_t = R_t - A_t Q_t A_t'.$$

2.6 Particle filter

2.6.1 Introduction

The Kalman filter approach provides an estimation framework to determine the posterior distribution of a state and its features recursively when the observation and system equations in the SSM are linear. However, it is no longer a suitable method if the assumptions fail to hold. This means the closed-form of the distribution of interest may be impossible to calculate. The EKF algorithm is a modification to the Kalman filter that allows nonlinear and non-Gaussian models. It provides an approximate inference of the mean and the covariance of the states. Therefore, it fails to give reasonable estimates when the models are far from linear and Gaussian assumptions.

A sequential Monte Carlo method (SMC), also known as particle filter, is an alternative to the recursive Bayesian filter approach for inference of the states and unknown parameters in the SSMs when the assumptions of the Kalman filter do not hold. The particle filter and Kalman filter are Bayesian estimation methods. The particle filter consists of simulation-based techniques that can be used to approximate the posterior distribution of the states based on observations in nonlinear and non-Gaussian SSMs. This approximation is obtained by drawing a set of random samples with associated weights from a target distribution $p(x_t|y_{1:t})$. These random samples are termed 'particles'. In other words, the key point of the SMC method is to use Monte Carlo simulation to approximate the posterior filtering distribution by generating a set of weighted samples. The estimation results using the particle filter are more accurate than those produced by other approximate filters. When the number of generated samples is large,

the SMC method provides the equivalent representation to the true posterior distribution. The SMC algorithms are designed to estimate the states sequentially when new observations are available.

The first use of Monte Carlo methods in nonlinear filtering in SSMs was by Handschin and Mayne (1969) and Handschin and Mayne (1970), in which only the posterior mean and covariance of the state were estimated by Monte Carlo methods. Several different particle filters are used for online estimation in DGLMs. The standard and easiest is called sequential importance resampling (SIR), and was first proposed by Gordon et al. (1993). Pitt and Shephard (1999) proposed the auxiliary particle filter for improving the performance of the standard particle filter. Here the auxiliary variable is used instead of the importance function to generate the particles sequentially. Liu and West (2001) and Storvik (2002) suggested particle filter algorithms for online estimation of the state and the hyper-parameters simultaneously.

Kitagawa (1996), Liu and Chen (1998), Liu (2001), Doucet et al. (2001), Andrieu and Doucet (2002), Fearnhead and Clifford (2003) and DelMoral et al. (2006) present an extensive amount of literature regarding the methodology of online and offline filtering by using the particle filter.

2.6.2 The basis of a particle filter

The particle filter represents a popular tool for obtaining, through Monte Carlo simulations, the approximate solution of the recursive Bayesian estimation problem in DGLMs. It aims to infer recursively at each time point the required joint posterior distribution of the states $p(x_{0:t}|y_{1:t})$, or the required marginal distribution of the state at the current time given the historical observations $p(x_t|y_{1:t})$,

by a set of weighted random samples. It then calculates their features such as the expectation based on these weighted samples (Doucet et al., 2000).

Bayes' theorem is used to find the filtering posterior distribution of state $p(x_t|y_{1:t})$ as follows:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \tag{2.6.1}$$

$$\begin{aligned} p(x_t|y_{1:t}) &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} \\ p(x_t|y_{1:t}) &\propto p(y_t|x_t)\int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}, \end{aligned} \tag{2.6.2}$$

where x_t is referred to as the current state and $y_{1:t}$ denotes the observation up to time t. Due to difficulties in calculating the intractable integrals of Equations (2.6.1) and (2.6.2), the particle filter can be used to solve these intractable integrals numerically. To achieve this, importance sampling is used sequentially in order to approximate the intractable integrals. Using the core idea of the SMC method, the approximation of the marginal posterior distribution of the state in Equation (2.6.2) can be written as

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)}\delta(x_t - x_t^{(i)}), \qquad t = 1, \ldots, T \tag{2.6.3}$$

where $w_t^{(i)}$ denotes the weight of the i^{th} particle at time t, $\delta(.)$ is known as a Dirac delta function and $x_t^{(i)}$ is the i^{th} particle at time t. The problem in the implementation of this approximation lies in the difficulty of drawing the samples from the desired posterior probability density function $p(x_t|y_{1:t})$. This issue is treated by resorting to another convenient probability distribution $q(.)$ so that the samples can be taken from it. As a result, these samples are used to estimate the true posterior probability distribution. The proposed convenient

function is defined as the importance function (IF) and the sample is called the importance sampling (IS), which is widely used in sampling techniques. This concept is similar to the prediction step in the Bayesian estimation method.

2.6.3 Importance sampling

The idea behind importance sampling (IS) is that when there is a difficulty in sampling directly from a target (true) distribution $p(x_t|y_{1:t})$, a proposal distribution $q(x_t|y_{1:t})$ can be used instead for generating identically and independently distributed samples (x_i, $i = 1, ..., N$). In other words, IS is referred to as the estimation process about the underlying distribution of interest using the set of observations that are generated from a different distribution. This proposal density is chosen so that it should have the same support as the target density, i.e. $p(x_t|y_{1:t}) > 0$ whenever $q(x_t|y_{1:t}) > 0$. Generally speaking consider the posterior distribution $q(x_t|y_{1:t})$ as the IF.

The expectation of any function $h(.)$ of the state posterior distribution can be calculated as

$$\begin{aligned} I &= \mathbb{E}[h(x_t)] = \int h(x_t)p(x_t|y_{1:t})dx_t \\ &= \int h(x_t)\frac{p(x_t|y_{1:t})}{q(x_t|y_{1:t})}q(x_t|y_{1:t})dx_t, \end{aligned} \tag{2.6.4}$$

where the first part is the expectation of $h(x)$ from the true (target) state posterior distribution, $p(x_t|y_{1:t})$, and the second part is the expectation with respect to the IF, $q(x_t|y_{1:t})$. In the second part of Equation 2.6.4, the true state posterior probability density function is replaced by the IF. Consequently, we are able to use it for generating sequentially independent and identically distributed samples. The second integral in the expectation formula can be approximated by the Monte

Carlo method and the particles as follows:

$$\hat{I} \approx \frac{1}{N}\sum_{i=1}^{N} \tilde{w}_t^{(i)} h(x_t^{(i)}), \qquad t = 1, \ldots, T \tag{2.6.5}$$

where $\tilde{w}_t^{(i)} = \frac{p(x_t^{(i)}|y_{1:t})}{q(x_t^{(i)}|y_{1:t})}$ are called the importance weights, representing the ratio between the distributions. These importance weights are employed to re-adjust the error that is produced by drawing the samples from the IF; $\sum_{i=1}^{N} w_t^{(i)} = 1$, and $x_t^{(i)}$ is sampled from the target density $q(x_t|y_{1:t})$. According to the central limit theorem the expectation based on the IF converges to the expectation of the actual posterior distribution of the state when increasing the number of particles (as$N \to \infty$); (see (Geweke, 1989)). This convergence is an attractive feature of Monte Carlo methods for obtaining asymptotically consistent estimates of the true posterior state distribution (Doucet and Johansen, 2009). If the normalising factor of the target distribution is not known, we can evaluate the importance weights up to a normalising constant; hence $\tilde{w}_t^{(i)} \propto \frac{p(x_t^{(i)}|y_{1:t})}{q(x_t^{(i)}|y_{1:t})}$. Thus the expectation in Equation (2.6.5) can be rewritten as:

$$\begin{aligned} \hat{I} &= \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_t^{(i)} h(x_t^{(i)})}{\frac{1}{N}\sum_{j=1}^{N} \tilde{w}_t^{(j)}} \\ \hat{I} &\equiv \sum_{i=1}^{N} w_t^{(i)} h(x_t^{(i)}) \end{aligned} \tag{2.6.6}$$

where $w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$ are called the normalised importance weights. The advantage of IS is that it reduces the variance of Monte Carlo estimators. In addition, the generated samples from the proposal density $q(x_t|y_{1:t})$ with the associated weights $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^{N}$ can be considered as an approximation of the target density $p(x_t|y_{1:t})$.

2.6.4 Sequential importance sampling

Sequential importance sampling (SIS) is defined as a repeat of applications of IS at each time instant. It maintains the posterior density of interest in the form of the weighted samples $\{x_t^{(i)}, w_t^{(i)} : i = 1, \ldots, N\}$ for approximating the state posterior distribution recursively. Therefore this approximate distribution can be employed to compute the expectation of an arbitrary function $h(x_t)$ as a weighted sample mean as follows:

$$E(h(x_t)|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} h(x_t^{(i)}). \tag{2.6.7}$$

In other words, SIS uses the principal concept of IS to construct sequentially the posterior distribution of interest (Doucet et al., 2000). The joint states posterior distribution, $p(x_{0:t}|y_{1:t})$, is considered to derive the algorithm of the SIS filter. To evaluate an operation for computing recursively the particle weights, Bayes' theorem is used to calculate in a recursive manner the desired posterior distribution of states in the approximation form as follows:

$$\begin{aligned} p(x_{0:t}|y_{1:t}) &\propto p(y_t|x_{0:t}, y_{1:t-1}) p(x_{0:t}|y_{1:t-1}) \\ &= p(y_t|x_{0:t}, y_{1:t-1}) \{p(x_t|x_{0:t-1}, y_{1:t-1}) p(x_{0:t-1}|y_{1:t-1})\}. \end{aligned} \tag{2.6.8}$$

By using the Markovian assumption, the posterior distribution in Equation (2.6.8) can be rewritten as follows:

$$p(x_{0:t}|y_{1:t}) = p(y_t|x_t) p(x_t|x_{t-1}) p(x_{0:t-1}|y_{1:t-1}). \tag{2.6.9}$$

If the importance density is chosen so that it can be decomposed as a recursion form as follows:

$$q(x_{0:t}|y_{1:t}) = q(x_t|x_{0:t-1}, y_{1:t}) q(x_{0:t-1}|y_{1:t-1}), \tag{2.6.10}$$

then the un-normalised importance weights of the particles can be updated recursively with the following expression:

$$\begin{aligned}
\tilde{w}_t^{(i)} &= \frac{p(x_{0:t}^{(i)}|y_{1:t})}{q(x_{0:t}^{(i)}|y_{1:t})} \\
&= \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})p(x_{0:t-1}^{(i)}|y_{1:t-1})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})q(x_{0:t-1}^{(i)}|y_{1:t-1})} \\
\tilde{w}_t^{(i)} &= w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})}, \qquad (2.6.11)
\end{aligned}$$

where $w_{t-1}^{(i)}$ is referred to as the normalised importance weights at previous times, and $\frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})}$ is known as the incremental weight. The normalised importance weights can be then computed as

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}. \qquad (2.6.12)$$

Thus, the expression in Equation (2.6.11) allows us, based on the importance function, to update the importance weights of the particles in a sequential way. Note that it is important to select a convenient form of the importance function so that it satisfies the Markovian assumption, $q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t}) = q(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:t})$,

$$\tilde{w}_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:t})} \qquad (2.6.13)$$

Therefore, to apply the algorithm of the filter it is only necessary to store the current state instead of the history of all the states. Table 2.3 shows the SIS algorithm used to obtain the recursive form for updating the importance weights.

Table 2.3: Sampling importance sampling filter(SIS) algorithm

---

Initialisation: At t=0, for i=1,...,N

Sample particles from the prior density $x_0^{(i)} \sim p(x_0)$.

Assign the initial importance weight as $w_0^{(i)} = N^{-1}$.

For $t \geq 1$ and i=1,...,N

Prediction:

Sample particles from the importance density $x_t^{(i)} \sim q(x_t|x_{t-1}^{(i)}, y_t)$.

Update:

Compute the unnormalised importance weights according to,

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}.$$

Normalise the importance weights, $w_t^{(i)} = \dfrac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$

2.6.4.1 The degeneracy issue of the SIS algorithm

The main problem facing the use of SIS is known as particle degeneracy. This problem occurs when the variance of the importance weights increases as time increases (Doucet et al., 2000). As a result, the distribution of the importance weights will have a large variance. In other words, after a few iterations few particles have high weight, whereas most of the particles have low weight, tending to zero. So when the Monte Carlo simulation is used to calculate the average of the posterior distribution of interest, it will be an inefficient estimate. This is because this calculation is based only on the particles with high weights and the effects of the particles with small weights are ignored. Therefore the SIS

algorithm fails to provide an adequate description of the posterior distribution of the states.

To avoid this problem, a resampling step can be added to the particle filter algorithm to reduce the effect of particle degeneracy. The main point of the concept of the resampling procedure is to focus on the particles with large weights and exclude the particles with small weights. In other words, in the resampling step, the particles with low weights will be discarded and replaced with replicating particles with high weights and set the weights equal. This means that, after doing the resampling step for each time, the weights of the new set of particles become uniform. Therefore this unweighted sample can be employed in the next step of the particle filter algorithm to approximate the posterior distribution of interest. Various types of resampling approaches are used to resolve the problem of particle degeneracy – for example multinomial resampling, stratified resampling, systematic resampling, and residuals resampling. Since there is no significant difference in the impacts that the different resampling methods have on the performance of the particle filter algorithm (Hol et al., 2006), the multinomial resampling (Gordon et al., 1993) technique is used in this thesis. The difference between the methods lies only in their computational complexity (Hol et al., 2006). In multinomial resampling, the discrete approximation distribution of $p(x_t = x_t^{(i)}) = w_t^{(i)}$ is used to generate with replacement a sample of N independent random variables, and the weights $w_t^{(i)}$ are set as equal, i.e. $w_t^{(i)} = \frac{1}{N}$, $i = 1, \ldots, N$. Therefore the unweighted sample is still used as an approximation of the posterior distribution of the state, $p(x_t|y_{1:t})$.

The effective sample size (ESS) introduced by Kong et al. (1994) is used as a measure to monitor particle degeneracy. It is defined as a function of the variation

coefficient of the importance weights as follows:

$$ESS = \big(\sum_{i=1}^{N}(w_t^{(i)})^2\big)^{-1}. \tag{2.6.14}$$

The ESS is the number of available effective particles from the target distribution that are used to approximate the posterior distribution of interest using the Monte Carlo approach. In other words, the ESS is a measurement to determine the algorithm efficiency which is calculated based on the weight variable. The range of the ESS is between 1 and N. Particle degeneracy is severe when the value of the ESS is small and vice versa. Therefore the resampling step is needed when the value of the ESS is lower than a predefined threshold value.

2.6.4.2 Choosing the importance function

The selection of the importance function is a crucial task in the particle filter methodology (Doucet et al., 2000). The choice of the IF has a significant effect on the performance of the particle filter algorithm for obtaining an accurate estimate of the state posterior distribution, $p(x_t|y_{1:t})$. There are two options in terms of selecting the IF, namely, a suboptimal choice and an optimal choice. In the suboptimal choice, the prior distribution of the state in the SSM is used as the importance function: $q(x_t|x_{0:t-1}, y_{1:t}) = p(x_t|x_{t-1})$. As a result the weight updating equation referred in Equation (2.6.13) can be simplified as follows:

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)}\, p(y_t|x_t^{(i)}). \tag{2.6.15}$$

Although the suboptimal importance density is known as the simplest one to calculate, it suffers from two disadvantages. First, the particle filter algorithm will provide a poor estimation of parameters of interest. This is because the construction of the IF does not take into account the knowledge of the current

value of an observation. Therefore the particles used to calculate the posterior distribution of interest will be obsolete. The second drawback is that, as time increases, the importance weights variance will be quite large.

The optimal choice of the importance function is expressed as the conditional distribution of the state given the state at a previous time and the current observation. The IF can be written as $q(x_t|x_{0:t-1}, y_{1:t}) = p(x_t|x_{t-1}, y_t)$ (Arulampalam et al., 2002). The optimal choice method is designed to reduce the importance weights variance. Therefore the ESS will be maximised. The expression of the updating of the unnormalised importance weights in Equation (2.6.13) can be expressed with the optimal choice as follows:

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{p(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} = w_{t-1}^{(i)}\, p(y_t|x_{t-1}^{(i)}). \tag{2.6.16}$$

The importance weight is obtained from the previous equation based on the following relationship of the optimal importance function:

$$p(x_t|x_{t-1}, y_t) = \frac{p(y_t|x_t)p(x_t|x_{t-1})}{p(y_t|x_{t-1})}. \tag{2.6.17}$$

Note that the calculation of the updating of the weight $\tilde{w}_t^{(i)}$ relies only on the state at previous time x_{t-1}, not on the current time x_t, and this is a characteristic feature of this selection.

2.6.5 Sampling importance resampling filter

The sampling importance resampling filter (SIR), also called the Bootstrap filter, was introduced by Gordon et al. (1993). It is considered as the simplest sequential Monte Carlo method for propagating and updating the particles in order to obtain the posterior filtering distribution of interest. This is because the prior density of the state is selected as the importance function, i.e. $q(x_t|x_{t-1}^{(i)}, y_{1:t}) = p(x_t|x_{t-1}^{(i)})$.

Therefore at time point t, the sequential updating of the importance weights, ω_t, can be done in a simple form. This updating is based on only the previous importance weight ω_{t-1} and the likelihood of a new observation, $p(y_t|x_t(i))$, as follows:

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} p(y_t|x_t^{(i)}). \tag{2.6.18}$$

The difference between the SIS and SIR algorithms lies in the implementation of the resampling step. In the SIR (Bootstrap) filter algorithm, the resampling step is performed at each time instant regardless of the value of the ESS, whereas in the SIS filter algorithm it is only needed at a time point when the value of the ESS is less than a threshold. The SIR algorithm is summarised in Table 2.4.

Table 2.4: Sampling importance resampling filter(SIS) algorithm

---

Initialisation: At t=0, for i=1,...,N

Sample particles from the prior density $x_0^{(i)} \sim p(x_0)$.

Assign the initial importance weight as $w_0^{(i)} = N^{-1}$.

For $t \geq 1$ and i=1,...,N

Prediction: Sample particles from the importance density $x_t^{(i)} \sim p(x_t|x_{t-1}^{(i)})$.

Update: Compute the unnormalised importance weights according to,

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} p(y_t|x_t^{(i)}).$$

Normalise the importance weights, $w_t^{(i)} = \dfrac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$

2.4 Resampling step

2.4.1 Simulate a new sample sets $\{\tilde{x}_t^{(i)}\}$ with size N from the sets $\{x_t^{(i)}\}$ according to the importance weights $w_{t-1}^{(i)}$.

2.4.2 Set $w_{t-1}^{(i)} = \dfrac{1}{N}$, for $i = 1, \ldots, N$

2.7 Auxiliary particle filter (APF)

As mentioned in Section 2.6.4.2, the performance of the particle filter depends on the selection of the importance function. In the recursive Bayesian estimation framework by the particle filters in the SSMs, obtaining an efficient importance function is considered difficult. However, an inappropriate choice of the importance function (IF) produces a problem of severe degeneration of the weights, and this leads to a poor performance of the algorithm. Pitt and Shephard (1999) introduced an extension of the generic particle filter, called an auxiliary particle filter (APF), to overcome this issue. They incorporate a discrete auxiliary variable (k), called the latent variable, to the SIR particle filter to help the simulation process. This also allows the particle filter to be adapted in a more efficient way. The key idea behind the APF is using an auxiliary variable as an extra dimension in the importance function, $q(x_t, k|y_{1:t})$, so that the particles $\{x_t^{(i)}, k^{(i)}\}_{i=1}^N$ can be generated easily.

Assuming the set of particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^N$ is available from the state filtering distribution $p(x_{t-1}|y_{1:t-1})$ at time point $t-1$, then the Monte Carlo approximation for the state prior density is obtained as follows:

$$\hat{p}(x_t|y_{1:t-1}) = \sum_{i=1}^{N} p(x_t|x_{t-1}^{(i)}) w_{t-1}^{(i)}. \tag{2.7.1}$$

Our target is to calculate the state posterior distribution $p(x_t|y_{1:t})$ by updating $p(x_t|y_{1:t-1})$ when a new observation arrives. Using Bayes' theorem, the approximation of the state posterior distribution at time instant t can be calculated by combining the likelihood of the observations $p(y_t|x_t)$ with the prior

approximation of the state in Equation (2.7.1) as follows:

$$\begin{aligned}\hat{p}(x_t|y_{1:t}) &\propto p(y_t|x_t)\,\hat{p}(x_t|y_{1:t-1}) \\ &= p(y_t|x_t)\sum_{i=1}^{N} p(x_t|x_{t-1}^{(i)})w_{t-1}^{(i)} \\ &= \sum_{i=1}^{N} p(y_t|x_t)\,p(x_t|x_{t-1}^{(i)})w_{t-1}^{(i)}. \end{aligned} \tag{2.7.2}$$

Pitt and Shephard (1999) suggested introducing the auxiliary variable k, defined as the natural number $k \in \{1,\ldots,N\}$, to the approximation in Equation (2.7.2). Therefore the expression of the posterior approximation in Equation (2.7.2) becomes the approximation of the joint posterior distribution of the state and the auxiliary variable and it can be written in the following form:

$$\begin{aligned}\hat{p}(x_t, k=i|y_{1:t}) &\propto p(y_t|x_t)\,p(x_t, k=i|y_{1:t-1}) \\ &= p(y_t|x_t)\,p(x_t|k=i, y_{1:t-1})\,p(k=i|y_{1:t-1}) \\ &= p(y_t|x_t)\,p(x_t|x_{1:t-1}^{(i)})\,w_{t-1}^{(i)} \\ &= \sum_{i=1}^{N} p(y_t|x_t^{(i)})\,p(x_t|x_{t-1}^{(i)})\,w_{t-1}^{(i)}, \end{aligned} \tag{2.7.3}$$

where the auxiliary variable k is the latent variable representing the particles index at time point $t-1$ so that $p(k=i) = w_{t-1}^{(i)}$, and for deriving the importance function. The joint importance function of the state and index variable proposed by Pitt and Shephard (1999) is defined as the factorisation formula up to the normalised factor as below:

$$\begin{aligned}q(x_t, k=i|y_{1:t}) &\propto q(x_t|i, y_{1:t})q(i|y_{1:t}) \\ &= p(y_t|\hat{x}_t^{(i)})\,p(x_t|x_{t-1}^{(i)})\,w_{t-1}^{(i)}, \end{aligned} \tag{2.7.4}$$

where $\hat{x}_t^{(i)}$ is a representative feature of the predictive density of the state, such as the mean or the mode. The mean is the most common selection: $\hat{x}_t^{(i)} =$

$\mathbb{E}(x_t|x_{t-1}^{(i)})$. Thus the marginal posterior distribution of the auxiliary variable k_j, where $(j = 1, \ldots, N)$, can be obtained from Equation (2.7.4) as follows:

$$\begin{aligned} q(k_j = i|y_{1:t}) &= \int q(x_t, k = i|y_{1:t})dx_t \\ q(k_j = i|y_{1:t}) &= \int p(y_t|\hat{x}_t^{(i)})\, p(x_t|x_{t-1}^{(i)})\, w_{t-1}^{(i)} dx_t \\ &\propto p(y_t|\hat{x}_t^{(i)}) w_{t-1}^{(i)}. \qquad i = 1, \ldots, N \end{aligned} \tag{2.7.5}$$

Consequently, the auxiliary variable k_j is selected based on the density in Equation (2.7.5) and then this auxiliary variable is used to simulate the state $x_t^{(j)}$ from the transition density, $x_t^{(j)} \sim p(x_t|x_{t-1}^{(i)})$, and set $x_t^{(j)} = (x_{0:t-1}^{(i)}, x_t^{(j)})$.

Using Equations 2.7.3 and 2.7.5, the second stage of the unnormalised importance weights of the j^{th} draw can be updated recursively as follows:

$$\begin{aligned} \tilde{w}_t^{(j)} &= \frac{p(x_t, k = i|y_{1:t})}{q(x_t, k = i|y_{1:t})} \\ &= \frac{\tilde{w}_{t-1}^{(j)} p(y_t|x_t^{(j)}) p(x_t^{(j)}|x_{t-1}^{(j)})}{\tilde{w}_{t-1}^{(j)} p(y_t|\hat{x}_t^{(k_j)}) p(x_t^{(j)}|x_{t-1}^{(j)})} \\ &= \frac{p(y_t|x_t^{(j)})}{p(y_t|\hat{x}_t^{(kj)})} \end{aligned} \tag{2.7.6}$$

Finally, the resampling step is implemented when the effective sample size (ESS) falls below a specified threshold as with the SIR particle filter algorithm in Table 2.4.

The main feature of the APF compared to the generic particle filters (as mentioned in Chapter 2) is that it allows the use of the predictive (prior) distribution of the state to generate the particles without losing much efficiency. The construction of the APF algorithm is provided in Table 2.5.

Table 2.5: The auxiliary particle filter algorithm

1. **Initialisation:** At t=0, for i=1,...,N

   Sample particles from the prior density $x_0^{(i)} \sim p(x_0)$.

   Assign the initial importance weight as $w_0^{(i)} = N^{-1}$.

2. For $t \geq 1$ and i=1,...,N

   2.1 For j=1,...,N

      2.1.1 Calculate $\hat{x}_t^{(i)} = \mathbb{E}(x_t|x_{t-1}^{(i)})$.

      2.1.2 Sample an auxiliary variable k_j with probability

      $$p(k_j = i) \propto p(y_t|\hat{x}_t^{(i)}) w_{t-1}^{(i)}$$

      2.1.3 Based on k_j generate new particle $x_t^{(j)}$ from $p(x_t|x_{t-1}^{(k_j)})$

      and set $x_t^{(j)} = (x_{0:t-1}^{(i)}, x_t^{(j)})$.

      2.1.4 Computing the unnormalised importance weights

      $$\tilde{w_t}^{(j)} = \frac{p(y_t|x_t^{(j)})}{p(y_t|\hat{x}_t^{(kj)})}$$

   2.3 Normalise importance weights

   $$w_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}$$

   2.4 Resampling step is needed if $N_{ESS} < N_{thre}$

      2.4.1 Simulate a sample set with size N from the discrete distribution

      $$p(x_{0:t}) = x_{0:t}^{(i)}$$

      and relabel this sample $x_{0:t}^{(1)}, \ldots, x_{0:t}^{(N)}$

   2.5 Set $\hat{p}(x_{0:t}|y_{1:t}) = \sum_{i=1}^{N} w_t^{(i)} p(x_{0:t}|x_{0:t}^{(i)})$

2.8 Markov Chain Monte Carlo Methods (MCMC)

In some situations, the posterior distribution of interest is hard to characterise analytically. In such a case, sampling from that distribution is needed in order to make inferences about the unknown parameters and latent variables. Thus, numerical methods are needed in order to deal with complex Bayesian inference. Markov chain Monte Carlo (MCMC) methods are Bayesian estimation techniques which can be used for estimating the posterior distribution of interest. In other words, MCMC is a class of algorithms for generating samples from a posterior probability distribution based on constructing a Markov chain whose stationary distribution is the desired probability distribution. Simulation from this chain with a large number of iterations is then used as a sample to approximate the posterior distribution of interest. Summaries of the sample, such as its mode, median, histogram etc, can be used in order to explore the posterior distribution, which is approximated by the Markov chain. For a detailed relevant to MCMC methods the reader is referred to, see Gamerman and Lopes (2006),Robert and Casella (2010), Robert and Casella (2013) and Gilks et al. (1993). There are different MCMC algorithms which can be used to generate the Markov chain. The Metropolis-Hastings and the Gibbs Sampler are common algorithms for obtaining the Markov chain. Below we provide a brief discussion of the Metropolis-Hastings algorithm and the Gibbs Sampler.

2.8.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm was first introduced by Metropolis et al. (1953) and then this algorithm was extended later by Hastings (1970). The idea of the MH algorithm is to generate a sequence of samples from a suitable proposal density $q(\theta|\theta^{(t)})$, where $\theta^{(t)}$ is the current state of the chain. These samples are then used to approximate the posterior distributions of unknown model parameters $p(\theta|x)$,where x is data. Because the mechanism of the MH algorithm is to construct a Markov Chain, the current sample of the sequence is only dependent on the previous sample value. Thus, the previously accepted sample value is exploited in order to construct the Markov chain. After a number of iterations, this Markov chain will converge to a desired probability distribution. The MH algorithm is used when the marginal distributions of the unknown parameters cannot be obtained analytically. However, we can easily draw samples from them. For using the MH algorithm, a normalising constant is not needed. The proposal density is used to generate a new candidate state, $(\theta^* \sim q(\theta|\theta^{(t)}))$ and a decision should be taken whether to accept it or not. The decision to accept or reject the candidate state for next state in the MH chain is made based on a value of an acceptance rate. This acceptance rate is defined as a ratio of importance weights as follows:

$$\boldsymbol{\alpha}(\theta^{(t)}, \theta^*) = \min\left(1, \frac{p(\theta^*)q(\theta^{(t)}|\theta^*)}{p(\theta^t)q(\theta^*|\theta^{(t)})}\right)$$

. Thus, if the candidate state is accepted with the acceptance rate $\boldsymbol{\alpha}(\theta^{(t)}, \theta^*)$, then the M-H chain will move and set $\theta^{(t+1)} = \theta^*$, otherwise the chain remains at the current state and we set $\theta^{(t+1)} = \theta^{(t)}$. Table 2.6 provides the pseudo-code for the Metropolis-Hastings algorithm to show how to generate a chain.

Table 2.6: The Metropolis-Hastings Algorithm

1. Generate a starting point from the initial distribution, $\theta^0 \sim p_0(.)$
2. For $t = 1, \ldots, T$
   - 2.1 Generate a candidate point θ^* given θ^{t-1} from a proposal distribution $q(\theta^*|\theta^{(t-1)})$
   - 2.2 Calculate an acceptance rate $\boldsymbol{\alpha} = \min\left\{1, \frac{p(\theta^*)q(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)})q(\theta^*|\theta^{(t-1)})}\right\}$
   - 2.3 Generate a value $\mathbb{U} \sim U(0,1)$, a uniform distribution on $[0,1]$
   - 2.4 If $\mathbb{U} \leq \alpha$, set $\theta^{(t)} = \theta^*$; otherwise set $\theta^{(t)} = \theta^{(t-1)}$
   - 2.5 Take the sequence $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}\}$

The choice of the proposal density and its relationship with the target density play a key role in influencing the performance of the MH algorithm for estimating the posterior distribution of interest. A poor proposal density leads to a poor exploration of the posterior distribution. In such a case, a large number of samples in the Markov chain are required in order to reach the convergence which results in obtaining a good approximation of the posterior distribution. The proposal distribution is chosen based on the point that the irreducibility and aperiodicity conditions are satisfied. This is achieved when the proposal is a positive density and has the same domain as the target density. A part of the Markov chain generated from the first state until the chain reaches its equilibrium is called burn-in and in any inference, burn-in should be removed from the chain, and the remaining should be used. To reduce the correlation between the successive states in Markov chain, thinning which means taking every m^{th} $(m \geq 2)$ state is used.

2.8.2 The Gibbs Sampler

The Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm for generating a Markov chain from the full conditional distributions of parameter when this full conditional is in a known distribution form. Namely, the Gibbs sampling scheme is used only when the full conditional distributions of interest are known distributions and we can easily sample from them. The idea behind the Gibbs sampling is that given parameters of interest $\theta = (\theta_1, \ldots, \theta_q)$, conditional distributions $p(\theta_i|\theta_{-i}) = p(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_q)$ are used for getting sample from the target distribution. In other words, the conditional distribution of each parameter $p(\theta_i|\theta_{-i}), i = 1, \ldots, q$ is used to update the parameter state given the current values of the rest of the parameters. The Gibbs sampler was first introduced by Geman and Geman (1984) where it has been used in the field of Bayesian image processing. The Gibbs sampler is considered as a special case of the Metropolis-Hastings algorithm. In the Gibbs sampler, the proposal densities are the full conditional distributions and the acceptance rate is 1. As a result, all proposed samples are always accepted. Table 2.7 provide the Gibbs sampling algorithm.

Table 2.7: The Gibbs sampler Algorithm

1. Generate a starting point from the initial distribution, $\boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_q^0)$
2. For $t = 1, \ldots, T$

   Generate $\theta_1^{(t)}$ from $p(\theta_1 | \theta_2^{(t-1)}, \ldots, \theta_q^{(t-1)})$

   Generate $\theta_2^{(t)}$ from $p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_q^{(t-1)})$

   $\vdots$

   Generate $\theta_q^{(t)}$ from $p(\theta_q | \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{q-1}^{(t)})$

   Set $\boldsymbol{\theta}^{(t)} = \{\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_q^{(t)}\}$

2.9 Summary

This chapter has aimed to present the literature review in terms of the classical and Bayesian time series methodology used throughout this thesis. The classical time series and state space model applications in various disciplines in medicine have also been provided in this chapter. With respect to the methodology of the Bayesian time series models, linear and nonlinear SSMs have been discussed in detail. The online Bayesian techniques used to track the dynamic state process in the SSMs, such as the Kalman filter and the extended Kalman filter (EKF), have been presented in a brief review.

In the latter part of the chapter, the literature review focused on an approximation solution of the estimation problem for the SSMs based on sequential Monte Carlo methods. These simulation based methods, which are also called particle filters, use the principle of sequential importance sampling (SIS). The

aim is to track the latent variables (states) in non-Gaussian SSMs. Therefore the approximation of the posterior distribution of the state can be estimated numerically using different particle filter algorithms. The degeneracy problem of the SIS algorithm and how it can be addressed, and choosing the importance density, have also been discussed in this chapter. A brief review of the Markov chain Monte Carlo (MCMC) methods with the Metropolis-Hastings and the Gibbs sampler algorithms was also offered in this chapter.

Chapter 3

Particle filter-based estimation

3.1 Introduction

In Chapter 2, background material for the state space models (SSMs) and recursive Bayesian estimation methods based on the particle filter was provided. The main purpose of using sequential Monte Carlo methods is to estimate sequentially the posterior filtering distribution of the current state when a new observation is collected. Generic particle filter algorithms such as sequential importance sampling (SIS), sequential importance resampling (SIR),and auxiliary particle filter (APF) were also discussed in Chapter 2. This chapter provides a survey of the principal concepts concerning different Bayesian estimation techniques. The aim is to address shortcomings in the inference of hyper-parameters in the generic particle filter algorithms. We discuss background preliminaries of the Liu and West (2001) and Storvik (2002) algorithms used for estimating the states and learning about the hyper-parameters simultaneously. With respect to evaluating the performance accuracy of estimating the hyper-parameters and forecasting, a

simulation study using the Liu and West and Storvik particle filter algorithms with the Poisson model will be implemented. Moreover, the estimated posterior distributions of the hyper-parameters for both algorithms at the last time point will be compared with those obtained from the MCMC run using a huge number of iterations. The second simulation study investigates what happens when the proposed fitted model is misspecified. To perform this study, the Liu and West (2001) particle filter algorithm with Poisson, negative binomial and mixture Poisson models will be applied. In addition, after investigating the proposed models in the simulation studies, they are applied to the different medical datasets.

3.2 Sequential hyper-parameters estimation in state space models

3.2.1 Overview

In a statistical modelling framework, the statistical models are defined through their parameters. In the methodology of the SSM, when the vector of the hyper-parameter θ is known, then the sequential Monte Carlo algorithms are considered effective approaches to estimate sequentially the posterior distribution of the states. In this case, the optimal solution of the filtering problem can be obtained by using the particle filter methods. If the vector of the hyper-parameters is also time-varying, it can be included in the same vector of the state x_t. Thus the sequential Monte Carlo algorithms can be applied effectively to estimate simultaneously both elements via the joint posterior distribution of the state x_t and hyper-parameters θ_t giving the observations up to time $t, p(x_t, \theta_t|y_{1:t})$. However, the particle filter algorithms do not provide an efficient solution for the

estimation of the joint posterior probability density function of the state and hyper-parameters when the hyper-parameter vector is an invariant. The cause of this issue lies in the lack of the dynamic availability of the hyper-parameter, which leads to the problem of the degeneracy of the particle filter algorithm. As a result, the particle filter methods provide a poor performance of the estimation. In practical time series applications with SSMs, the hyper-parameters are always unknown. Therefore it is necessary to make an inference of the unknown parameter vector before performing the sequential Monte Carlo method, see Andrieu et al. (2005).

In the literature on off-line statistical Bayesian estimation methods, the Markov Chain Monte Carlo (MCMC) approach is thought to be a successful technique to estimate the static parameters of the model when the data is available beforehand (Liu, 2001).

In the methodology of the DGLMs, there are several ways of estimating the static parameters in the model based on the implementation of the particle filter. Liu and West (2001) introduced a way that treats the shortcomings of the unknown static parameters estimation by adding a small noise to the fixed parameters so that they become time-varying. They have quoted this idea from Gordon et al. (1993). The new hyper-parameters can be added to the state vector. Thus, the joint posterior distribution of both can be calculated simultaneously in the same particle. Another technique for tackling the estimation problem of unknown static parameters has been suggested by Storvik (2002). He proposed a particle filter algorithm to simulate the latent state variables along with the static parameters in a sequential manner based on a low dimensional vector of sufficient statistics.

In the next two sections, we introduce the principal idea of estimating the

hyper-parameters by using the artificial evolution method by Gordon et al. (1993) as well as the kernel smoothing method by West (1993), which are considered the cornerstones of the construction of the Liu and West (2001) algorithm.

3.2.2 Artificial evolution method of static parameters

To counter the problem of the degeneracy of the state particles in the recursive Bayesian estimation method for DGLMs, Gordon et al. (1993) proposed making a modification to the state particles by adding small random perturbations to these particles of the state over time. This idea has been extended by Liu and West (2001) to accommodate the problem estimation of the static parameter model. The hyper-parameters θ are indexed by time and replaced by θ_t. Thus the static parameters with the artificial dynamic can be included in the augmented state vector (x_t, θ_t). The resulting evolution model of the static parameter of interest with an additional artificial dynamic can be expressed according to a random walk as follows:

$$\theta_t = \theta_{t-1} + \xi_t, \tag{3.2.1}$$

where ξ_t is a random disturbance which follows a normal distribution, $\xi_t \sim N(0, \zeta_t)$, with mean zero and variance ζ_t. The principal motivation of the idea of the artificial evolution method lies in providing a mechanism to generate new values of the static parameters at each time instant. With the artificial dynamic of the static parameters, the particle filter algorithm can be implemented to obtain the approximation of the joint posterior distribution of the state and static parameters $p(x_t, \theta|y_{1:t})$. The resulting marginal filtering distribution of the hyper-parameters, $p(\theta|y_{1:t})$, will be very diffuse. This is because of the loss of information. In other words, the use of the artificial dynamic for the static parameters leads to a distortion of the resulting posterior distribution of the es-

timated hyper-parameters compared to the marginal filtering distribution of the true static parameters in theory.

3.2.3 Kernel density estimation method

In statistical literature, a kernel density estimation method (KDEM) is defined as a non-parametric way of estimating the distribution of a random variable. In addition, for the SSMs, West (1993) suggested a different technique to solve the filtering problem of the static parameters $p(\theta|y_{1:t})$ based on the concept of kernel smoothing, with the idea of a neat shrinkage.

In SSMs, suppose the particles of the static parameters with their associated weights $\{\theta_{t-1}^{(i)}, \omega_{t-1}^{(i)}\}_{i=1}^{N}$ at time instant $t-1$ are available so they can be used to obtain the approximation of the posterior distribution of the static parameter as follows:

$$\hat{p}(\theta|y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta(\theta_{t-1} - \theta_{t-1}^{(i)}). \tag{3.2.2}$$

Note that the subscript t of the parameter θ_{t-1} does not denote time-varying, but is only used for indicating the posterior distribution that is employed to generate the samples at this time.

West (1993) proposed modifying Equation 3.2.2 by replacing the dirac delta function $\delta(\theta_{t-1} - \theta_{t-1}^{(i)})$ with the kernel smoothing density. To calculate the filtering density function of the hyper-parameter, the following applies:

$$\tilde{p}(\theta|y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta|m_{t-1}^{(i)}, h^2 \mathbf{V}_{t-1}), \tag{3.2.3}$$

where $N(.|m, \Sigma)$ is defined as a multivariate normal distributions with mean m and variance Σ. This multivariate Normal distribution is weighted by the sample weights $w_{t-1}^{(i)}$. Therefore the discrete approximation of the posterior of the static

parameter θ becomes a continuous distribution. $m_{t-1}^{(i)}$ are kernel locations of the i^{th} mixture Normal component and $h \in [0,1]$ is known as the smoothing parameter which is used to control the over-dispersion of the resulting filtering distribution of θ. This is based on the standard kernel method via the multivariate Normal distribution, as we will discuss later. Additionally, $\mathbf{V}_{t-1}$ is the Monte Carlo posterior variance of $\hat{p}(\theta|y_{1:t-1})$ which is given by:

$$\mathbf{V}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} (\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})^T. \tag{3.2.4}$$

In the standard kernel method, the kernel locations $m_{t-1}^{(i)}$ are determined as existing sample values of the static parameter $\theta_{t-1}^{(i)}$, i.e. $m_{t-1}^{(i)} = \theta_{t-1}^{(i)}$. However, in this context, the resulting posterior variance from this kernel density $(1+h^2)V_{t-1}$ will always be larger than the Monte Carlo posterior variance, V_{t-1}. This leads to the loss of information. This means that the approximate posterior density function of the static parameter θ, which is calculated based on the standard kernel density via the mixture Normal, will be away from the desired posterior probability density function of the true fixed parameter (Liu and West, 2001).

The issue of over-dispersion for the posterior density resulting from the standard kernel density will be exacerbated as the approximation operation of the filtering distribution is repeated over time. In other words, the approximation of the posterior $p(\theta|y_{1:t+1})$ distribution at time $t+1$ will be affected by the over-dispersed approximation of the filtering $p(\theta|y_{1:t})$ density at time t (Liu and West, 2001). This problem is considered a serious blemish in implementing the sequential simulation. For this reason, West (1993) suggested a solution to this issue of variance increasing by using the shrinkage concept for specifying the kernel locations $m_{t-1}^{(i)}$ as follows:

$$m_{t-1}^{(i)} = c\,\theta_{t-1}^{(i)} + (1-c)\bar{\theta}_{t-1}, \tag{3.2.5}$$

where $c = \sqrt{(1-h^2)}$ and $c \in (0,1)$ is a shrinkage (tuning) parameter which is defined as the efficiency particle measurement. The smoothing parameter h is determined using a discount factor as follows:

$$h^2 = 1 - \left[\frac{3\delta - 1}{2\delta}\right]^2, \tag{3.2.6}$$

where the discount factor δ range is typically specified as around (0.95–0.99) (Doucet et al., 2001). This will lead to the range of the shrinkage parameter c being (0.974–0.995). The specification of these kernel locations by the shrinkage rule ensures the stability of the resulting mean $\bar{\theta}_{t-1}$ is retained and makes a correction of the resulting variance from the mixture normal so that it becomes the same as the Monte Carlo posterior variance, $\mathbf{V}_{t-1}$.

3.3 Liu and West particle filter (LWPF)

The estimation of the static parameters in SSMs is represented as a challenging task. Different algorithms based on simulation within the recursive Bayesian framework exist for online learning of the hyper-parameters. The principle of the artificial evolution of the static parameters by Gordon et al. (1993) has been adopted here to make parameters time-varying. Thus, when using an artificial dynamic for the static parameters, they can be added to the state vector and estimated simultaneously in a sequential way. In this context, the joint posterior distribution of the state and static parameters, $p(x_t, \theta|y_{1:t})$, can be approximated sequentially by using the sequential Monte Carlo algorithms when the new observation y_t at time point t becomes available. Liu and West (2001) provided their particle filtering algorithm to achieve this target.

The key idea behind the Liu and West (2001) approach is to construct the

joint target filtering distribution of the state and static parameters by making an extension to the APF of Pitt and Shephard (1999) which was introduced in Section 2.7. Incorporating the APF and the kernel smoothing approximation by West (1993) leads to a better estimation of the posterior distribution of the fixed parameters posterior probability density function, $p(\theta|y_{1:t})$, using mixture Normal distributions. In addition, the shrinkage concept is included in this algorithm in an effort to counter the problem of over-dispersion in the resulting filtering distribution created by pretending that the parameter is time-varying.

At time point $t-1$, suppose the approximation of the joint filtering distribution $p(x_{t-1}, \theta|y_{1:t-1})$ is available, and it is in the form of

$$\hat{p}(x_{t-1}, \theta|y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{(x_{t-1}^{(i)}, \theta^{(i)})}(x_{t-1}, \theta). \tag{3.3.1}$$

Then the discrete approximation of the marginal distribution of the static parameter θ can be obtained from the joint distribution in Equation (3.3.1) as follows:

$$\begin{aligned} \hat{p}(\theta|y_{1:t-1}) &= \int \hat{p}(x_{t-1}, \theta|y_{1:t-1}) dx_{t-1} \\ &\approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta(\theta - \theta_{t-1}^{(i)}). \end{aligned} \tag{3.3.2}$$

The Monte Carlo estimation of the mean vector $\bar{\theta}$ and the variance matrix of the discrete approximation $\hat{p}(\theta|y_{1:t-1})$ are given by

$$\bar{\theta}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} \theta_{t-1}^{(i)} \tag{3.3.3}$$

$$\mathbf{V}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} (\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})'. \tag{3.3.4}$$

As mentioned above, the subscript $t-1$ of the parameter θ_{t-1} does not denote time-varying, but is only used to indicate the posterior distribution employed to generate the samples at this time. By replacing each point mass of the dirac

delta function $\delta(\theta - \theta_{t-1}^{(i)})$ in Equation (3.3.2) with the kernel estimation via a Gaussian density as Liu and West (2001) proposed, the resulting approximation of the static parameter distribution $\tilde{p}(\theta|y_{1:t-1})$ will be a continuous distribution in the following form:

$$\tilde{p}(\theta|y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta|m_{t-1}^{(i)}, h^2\mathbf{V}_{t-1}), \tag{3.3.5}$$

where $m_{t-1}^{(i)}$ are kernel locations of the i^{th} mixture Normal component, and $h \in [0,1]$ is known as the smoothing parameter. As mentioned before in Section 3.2.3, setting the kernel locations $m_{t-1}^{(i)}$ as existing particles $\theta_{t-1}^{(i)}$ ensures the preservation of the mean $\bar{\theta}_{t-1}$ that is calculated by both of the approximations distributions. However, the resulting variance of $\tilde{p}(\theta|y_{1:t-1})$ is always bigger than the variance of $\hat{p}(\theta|y_{1:t-1})$. This leads to the degeneracy problem of the particles used to approximate the filtering distribution of the static parameter of interest.

To see the resulting case of this issue, let us calculate the mean and the variance of the static parameter θ under the approximation distribution $\tilde{p}(\theta|y_{1:t-1})$ in Equation (3.3.5). The mixture Normal components in this approximation are indexed by the auxiliary latent variable k. So we have

$$\begin{aligned} \mathbb{E}(\theta) &= \mathbb{E}(\mathbb{E}(\theta|k)) = \mathbb{E}(\theta^{(k)}) \\ &= \sum_{i=1}^{N} w_{t-1}^{(i)} \theta_{t-1}^{(i)} = \bar{\theta}_{t-1}. \end{aligned} \tag{3.3.6}$$

$$\begin{aligned} \mathbb{V}ar(\theta) &= \mathbb{E}(\mathbb{V}ar(\theta|k)) + \mathbb{V}ar(\mathbb{E}(\theta|k)) \\ &= \mathbb{E}(h^2\mathbf{V}_{t-1}) + \mathbb{V}ar(\theta^{(k)}) \\ &= h^2\mathbf{V}_{t-1} + \mathbf{V}_{t-1} = (1+h^2)\mathbf{V}_{t-1} > \mathbf{V}_{t-1}. \end{aligned} \tag{3.3.7}$$

To avoid the problem of loss of information and to overcome the increase in the resulting variance, West (1993) suggested the idea of shrinkage to specify the

kernel locations. In this context, the kernel locations are given by

$$m_{t-1}^{(i)} = c\,\theta_{t-1}^{(i)} + (1-c)\bar{\theta}_{t-1}, \tag{3.3.8}$$

where $c = \sqrt{(1-h^2)}$, $c \in (0,1)$ is a shrinkage (tuning) parameter and $c^2 + h^2 = 1$. As a result, the definition of $\tilde{p}(\theta|y_{1:t-1})$ in Equation (3.3.5) will be altered to

$$\tilde{p}(\theta|y_{1:t-1}) \approx \sum_{i=1}^{N} wt - 1^{(i)} N(\theta|c\,\theta_{t-1}^{(i)} + (1-c)\bar{\theta}_{t-1}, h^2\mathbf{V}_{t-1}). \tag{3.3.9}$$

Thus with this setting of the kernel locations as in Equation (3.3.8), the resulting variance of $\tilde{p}(\theta|y_{1:t-1})$ has been corrected so that it becomes the same as the resulting variance of $\hat{p}(\theta|y_{1:t-1})$, and retains the same mean:

$$\begin{aligned} \mathbb{E}(\theta) &= \mathbb{E}(\mathbb{E}(\theta|k)) = \mathbb{E}(c\,\theta^{(k)} + (1-c)\bar{\theta}) \\ &= c\,\bar{\theta} + (1-c)\bar{\theta} = \bar{\theta} \end{aligned} \tag{3.3.10}$$

and the variance is

$$\begin{aligned} \mathbb{V}ar(\theta) &= \mathbb{E}(\mathbb{V}ar(\theta|k)) + \mathbb{V}ar(\mathbb{E}(\theta|k)) \\ &= \mathbb{E}(h^2 V_{t-1}) + \mathbb{V}ar(c\,\theta^{(k)} + (1-c)\bar{\theta}) \\ &= h^2\mathbf{V}_{t-1} + c^2\,\mathbb{V}ar(\theta^{(k)}) = h^2\mathbf{V}_{t-1} + c^2\,\mathbf{V}_{t-1} \\ &= (h^2 + c^2)\mathbf{V}_{t-1} = \mathbf{V}_{t-1}. \end{aligned} \tag{3.3.11}$$

The importance of the choice of the shrinkage parameter, c, lies in deriving the smoothing parameter, h, which is included in the Normal mixture approximation of the static parameter distribution. As a result, the performance of the Liu and West algorithm is affected by this choice. In practice, the shrinkage parameter is determined based on the discount factor, as Liu and West (2001) proposed, as $c = (3\delta - 1)/(2\delta)$, so that the smoothing parameter can be obtained as

$$h^2 = 1 - \left[\frac{3\delta - 1}{2\delta}\right]^2. \tag{3.3.12}$$

Generally, the range of the discount factor δ is specified as (0.95–0.99), which results in the shrinkage parameter range c of (0.974–0.995).

The approximation form of the joint filtering distribution of the state and the static parameter at time instant $t-1$ in Equation (3.3.1) can be expressed based on $\tilde{p}(x_{t-1}, \theta|y_{1:t-1})$ as follows:

$$\tilde{p}(x_{t-1}, \theta|y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(c\,\theta^{(i)} + (1-c)\bar{\theta}, h^2 \mathbf{V}_{t-1}).\delta(x_{t-1} - x_{t-1}^{(i)}). \quad (3.3.13)$$

Hence, performing the Bayesian approach in two stages, prediction and update, the approximation of the target joint distribution of interest, $p(x_t, \theta|y_{1:t})$, can be acquired when a new observation y_t is made, as follows:

$$\begin{aligned}
p(x_t, \theta|y_{1:t}) &\propto p(x_t, \theta, y_t|y_{1:t-1}) \\
&= p(y_t|x_t, \theta, y_{1:t-1}) p(x_t, \theta|y_{1:t-1}) \\
&= p(y_t|x_t, \theta) p(x_t|x_{t-1}, \theta, y_{1:t-1}) p(x_{t-1}, \theta|y_{1:t-1}) \\
&\approx p(y_t|x_t, \theta) p(x_t|x_{t-1}, \theta) \tilde{p}(x_{t-1}, \theta|y_{1:t-1}) \\
&= p(y_t|x_t, \theta) p(x_t|x_{t-1}, \theta) \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta|m^{(i)}, (1-c^2)\mathbf{V}_{t-1}) \delta(x_{t-1} - x_{t-1}^{(i)}) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)} p(y_t|x_t, \theta) p(x_t|x_{t-1}^{(i)}, \theta) N(\theta|m^{(i)}, (1-c^2)\mathbf{V}_{t-1}) \delta(x_{t-1} - x_{t-1}^{(i)})
\end{aligned}$$

By introducing the auxiliary variable as defined in Section 2.7, the target joint distribution of the state and the static parameter based on this auxiliary variable can be formulated in the following form:

$$\begin{aligned}
p(x_t, \theta, k = i|y_{1:t}) \approx\ & w_{t-1}^{(i)} p(y_t|x_t, \theta) p(x_t|x_{t-1}^{(i)}, \theta) \\
& N(\theta|m^{(i)}, (1-c^2)\mathbf{V}_{t-1}) \delta(x_{t-1} - x_{t-1}^{(i)}). \quad (3.3.14)
\end{aligned}$$

A suitable choice of the importance function proposed by Liu and West (2001) with the introduction of the auxiliary variable can be expressed in the following

form:

$$q(x_t, \theta, k = i|y_{1:t}) \propto w_{t-1}^{(i)} p(y_t|x_t = \hat{x}_t, \theta = m^{(i)}) p(x_t|x_{t-1}^{(i)}, \theta) \\ N(\theta|m^{(i)}, h^2\mathbf{V}_{t-1})\delta(x_{t-1} - x_{t-1}^{(i)}), \qquad (3.3.15)$$

where $\hat{x}_t^{(i)}$ is determined as the mean or the mode of the predictive density of the state given the static parameter via the kernel location, $p(x_t|x_{t-1} = x_{t-1}^{(i)}, \theta = m^{(i)})$. To obtain the sample with size N of the pair $(x_t^{(i)}, \theta^{(i)})$ from the auxiliary joint importance $q(x_t, \theta, k = i|y_{1:t})$ density in Equation (3.3.15), the following three stages will be iterated for $j = 1, \ldots, N$ at each time point.

(a) Draw an auxiliary variable k_j with

$$p(k_j = i) \propto w_{t-1}^{(i)} p(y_t|x_t = \hat{x}_t^{(i)}, \theta = m^{(i)}), \qquad i =, 1 \ldots, N.$$

(b) Based on $k_j = i$, sample the static parameter from the Gaussian density, $\theta^{(j)} \sim N(\theta|m^{(i)}, (1 - c^2)\mathbf{V}_{t-1})$, and set $\theta^{(j)} = \theta$.

(c) Based on $k_j = i$ and $\theta^{(j)} = \theta$, sample $x_t^{(j)} \sim p(x_t|x_{t-1} = x_{t-1}^{(i)}, \theta = \theta^{(j)})$.

Using Equations (3.3.14) and (3.3.15), the unnormalised importance weights can be updated recursively as follows:

$$\begin{aligned} w_t^{(k_j)} &= \frac{p(x_t, \theta, k = i|y_{1:t})}{q(x_t, \theta, k = i|y_{1:t})} \\ &= \frac{w_{t-1}^{(k_j)} p(y_t|x_t = x_t^{(j)}, \theta = \theta^{(j)}) p(x_t^{(j)}|x_{t-1}^{(j)}, \theta^{(j)}) N(\theta|m^{(k_j)}, (1 - c^2)\mathbf{V}_{t-1})}{w_{t-1}^{(k_j)} p(y_t|x_t = \hat{x}_t^{(k_j)}, \theta = m^{(k_j)}) p(x_t^{(j)}|x_{t-1}^{(j)}, \theta^{(j)}) N(\theta|m^{(k_j)}, (1 - c^2)\mathbf{V}_{t-1})} \\ &= \frac{p(y_t|x_t = x_t^{(j)}, \theta = \theta^{(j)})}{p(y_t|x_t = \hat{x}_t^{(k_j)}, \theta = m^{(k_j)})}. \qquad (3.3.16) \end{aligned}$$

Thus at time point t, the approximation of the joint filtering distribution of the state and static parameter can be obtained after a normalising step of the

importance weights as follows:

$$\hat{p}(x_t, \theta | y_{1:t}) \approx \sum_{i=1}^{N} \omega_t^{(i)} \delta_{(x_t^{(i)}, \theta^{(i)})}(x_t, \theta).$$

Similar to the case of the generic particle filter algorithm described in Chapter 2, the resampling step is performed based on the multinomial approach when the number of efficient particles falls below a specified threshold. The ESS measurement is used to determine how many particles with a high frequency are used for obtaining the approximation distribution of interest. According to Liu and West (2001), the approximation posterior distribution of the static parameter of interest must be supported by an appropriate mixture distribution, so that the expression of the static parameter is then consistent with the selected mixture distribution. In other words, the distributions of the mixture components and the unknown static parameter must have the same support of the entire real line. A summary of the Liu and West particle filter algorithm is provided in Table 3.1.

Table 3.1: Liu and West's particle filter algorithm

1. **Initialisation:** Suppose, at time $t-1$, for i=1,...,N

   Sample pair particles $(x_{t-1}^{(i)}, \theta^{(i)})$ independently from

   the prior densities $x_0^{(i)} \sim p(x_{t-1})$ and $\theta^{(i)} \sim p(\theta)$

   Assign the initial importance weight as $w_{t-1}^{(i)} = N^{-1}$, and

   $$\hat{p}(x_{t-1}, \theta | y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{(x_{t-1}^{(i)}, \theta^{(i)})}(x_{t-1}, \theta)$$

2. For $t \geq 1$ and i=1,...,N

   2.1 calculate $\bar{\theta}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} \theta_{t-1}^{(i)}$, and

   $$\mathbf{V}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} (\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})^T,$$
   $$m_{t-1}^{(i)} = c\,\theta_{t-1}^{(i)} + (1-c)\bar{\theta}_{t-1},$$
   $$\hat{x}_t^{(i)} = \mathbb{E}(x_t | x_{t-1} = x_{t-1}^{(i)}, \theta = m_{t-1}^{(i)}).$$

   2.2 For j=1,...,N

      2.1.1 Sample an auxiliary variable k_j with probability

      $$p(k_j = i) \propto p(y_t | \hat{x}_t^{(i)}, \theta = m_{t-1}^{(i)}) w_{t-1}^{(i)}.$$

      2.1.2 Based on k_j sample $\theta^{(j)} \sim N(\theta | m^{(i)}, (1-c^2)\mathbf{V}_{t-1})$.

      2.1.3 Based on k_j and $\theta^{(j)} = \theta$, sample $x_t^{(j)} \sim p(x_t | x_{t-1} = x_{t-1}^{(i)}, \theta = \theta^{(j)})$.

      2.1.3 Compute the unnormalised importance weights

      $$\tilde{w}_t^{(j)} = \frac{p(y_t | x_t^{(j)}, \theta = \theta^{(j)})}{q(y_t | \hat{x}_t^{(kj)}, \theta = m^{(k_j)})}$$

   2.3 Normalise importance weights $\quad w_t^{(i)} = \dfrac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$

   2.4 Resampling step in needed if $N_{ESS} < N_{thre}$

      2.4.1 Simulate a sample set with size N from the discrete joint distribution

      $$p(x_t, \theta = x_t^{(i)}, \theta(i)) = w_t^{(i)}$$

      and relabel this sample $(x_t^{(1)}, \theta^{(1)}), \ldots, (x_t^{(N)}, \theta^{(N)})$

      2.4.2 For i=1,...,N reset the weights $w_t^{(i)} = N(-1)$

   2.5 Set $\hat{p}(x_{0:t} | y_{1:t}) = \sum_{i=1}^{N} w_t^{(i)} p(x_{0:t} | x_{0:t}^{(i)})$

3.4 Storvik particle filter (SKPF)

Incorporating a vector of static parameters as a part of the state vector, in the context of the recursive Bayesian estimation method and the DGLMs, is considered as a usual approach for dealing with unknown static parameters. However, for a special class of DGLM, Storvik (2002) suggests an alternative particle-based technique to determine the hyper-parameters and the latent state variables. The concept of the Storvik particle filter algorithm is to use low dimensional sufficient statistics, denoted by s, rather than kernel smoothing densities as described in Liu and West (2001). Storvik (2002) employs these sufficient statistics to simulate sequentially the unknown static parameters and the state variables. He utilises the fact that the vector of the sufficient statistics can be updated sequentially when a new state, x_t, is generated from its posterior density and a new observation, y_t, is created via mapping $s_t = S(s_{t-1}, x_t, y_t)$. Therefore, for the static parameters, θ_t, is also tracked by the sufficient statistics in a sequential manner. For the Gaussian system process, in particular, the particle set of the sufficient statistics at time instant $t-1$ can be updated sequentially to a new particle set at time point t by using a class of equations of the Kalman filter as introduced by Storvik (2002). Thus, the marginal posterior distribution of the static parameter, $p(\theta_t|x_t, y_{1:t})$, can be defined conditionally on sufficient statistics in the form

$$p(\theta_t|x_t, y_{1:t}) = p(\theta_t|s_t). \tag{3.4.1}$$

Furthermore, in Storvik's algorithm, only the particle of sufficient statistics, $\{s_t^{(i)}\}_{i=1}^N$, needs to be stored, instead of the state particles ,$\{x_t^{(i)}\}_{i=1}^N$, at each iteration.

At time point $t-1$, the assumed approximation of the posterior of the state distribution $p(x_{t-1}|y_{1:t-1})$ is available. In this case, using Bayes' theory, the joint

filtering distribution of the state and the static parameters given the observation up to time t according to the Storvik (2002) approach is given as:

$$\begin{aligned} p(x_t, \theta | y_{1:t}) &\propto p(x_t, \theta, y_t | y_{1:t-1}) \\ &= p(y_t | x_t, y_{1:t-1}, \theta)\, p(x_t | x_{t-1}, y_{1:t-1}, \theta)\, p(\theta | x_{t-1}, y_{1:t-1}) p(x_{t-1} | y_{1:t-1}) \\ &= p(y_t | x_t, \theta)\, p(x_t | x_{t-1}, \theta)\, p(\theta | s_{t-1}) p(x_{t-1} | y_{1:t-1}). \end{aligned} \tag{3.4.2}$$

In Storvik's algorithm at each time step, conditional on the existing hyper-parameters, the proposal density function is used to generate new particles of states, and is given in the following manner:

$$x_t^{(i)} \sim\ q(x_t | x_{t-1}^{(i)}, y_t, \theta). \tag{3.4.3}$$

New hyper-parameter particles are also simulated from their importance function conditional on sufficient statistics as follows:

$$\theta \sim\ q(\theta | x_{t-1}^{(i)}, y_{1:t}). \tag{3.4.4}$$

Thus the un-normalised importance weights can be updated recursively proportional to $p(y_t, x_t, s_t, \theta)$ and both proposal densities as follows:

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)}\, \frac{p(\theta | s_{t-1}^{(i)})\, p(x_t^{(i)} | x_{t-1}^{(i)}, \theta)\, p(y_t | x_t^{(i)}, \theta)}{q(\theta | x_{t-1}^{(i)}, y_{1:t})\, q(x_t | x_{t-1}^{(i)}, y_t, \theta)}. \tag{3.4.5}$$

For a special case of dynamic linear models, called the Gaussian system process, Storvik (2002) proposed a technique for updating sufficient statistics in a recursive way according to a special class of the Kalman filter updating equations. In this model, assuming an observation process is arbitrary, a transition process is assumed to be Gaussian and satisfies

$$x_t = Z_t' \beta + \varepsilon_t, \qquad\qquad \varepsilon \sim\ N(0, \sigma^2 \varphi), \tag{3.4.6}$$

where $Z_t' = Z(x_{t-1})$ is a matrix of elements that might be nonlinear functions of x_{t-1},and $\beta\,, \sigma^2$ are unknown parameters with prior distributions $\beta \sim\ N(\mathbf{m}_0, \sigma^2\mathcal{C}_0)$ and $\sigma^2 \sim\ IG(v_0/2, d_0/2)$, and φ can be determined as a known value. In line with the standard theory by West and Harrison (1997), the unknown parameters can be simulated based on sufficient statistics vector $s_t = (\mathbf{m}_t, C_t, v_t, d_t)$ as follows:

$$\sigma_t^2|x_t, y_{1:t} \sim IG\left(\frac{v_t}{2}, \frac{d_t}{2}\right) \tag{3.4.7}$$

$$\beta_t|x_t, y_{1:t}, \sigma^2 \sim N(\mathbf{m}_t, \sigma^2\mathcal{C}_t). \tag{3.4.8}$$

The vector of sufficient statistics $s_t = (\mathbf{m}_t, \mathcal{C}_t, v_t, d_t)$ can be updated by using a special class of Kalman filter equations as follows:

$$D_t = Z_t'\mathcal{C}_{t-1}Z_t + \varphi \tag{3.4.9}$$

$$\mathcal{C}_t = \mathcal{C}_{t-1} - \mathcal{C}_{t-1}Z_tD_t^{-1}Z_t'\mathcal{C}_{t-1} \tag{3.4.10}$$

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathcal{C}_{t-1}Z_tD_t^{-1}Z_t'\left(x_t - Z_t'\mathbf{m}_{t-1}\right) \tag{3.4.11}$$

$$d_t = d_{t-1} - \left[x_t - Z_t'\mathbf{m}_{t-1}\right]' D_t^{-1}(x_t - Z_t'\mathbf{m}_{t-1}) \tag{3.4.12}$$

$$v_t = v_{t-1} + b, \tag{3.4.13}$$

where b is the dimension of x_t. A summary of the Storvik (2002) particle filter algorithm is provided in Table 3.2. The main benefit of Storvik's algorithm is that, in a simulation procedure at each iteration, the current simulated particles of the hyper-parameters, $\{\theta_t^{(i)}\}_{i=1}^N$, do not depend on the simulated particles in a previous time, $\{\theta_{t-1}^{(i)}\}_{i=1}^N$. Consequently, the problem of the impoverishment of the hyper-parameters particles is avoided.

Table 3.2: Storvik's particle filter algorithm

---

1. **Initialisation:** Suppose, at time $t-1$, for $i = 1, \ldots, N$

   Sample state particles $(x_{t-1}^{(i)})$ from the prior density $x_0^{(i)} \sim p(x_{t-1})$

   Assign the initial vector of sufficient statistics as $s_{t-1}^{(i)}$

   Assign the initial importance weight as $w_{t-1}^{(i)} = N^{-1}$

2. For $t = 1, \ldots, T$ and $i = 1, \ldots, N$

   2.1 Simulate $\theta^{(i)}$ and $(x_t^{(i)})$ from their importance functions as

   $$\theta^{(i)} \sim q(\theta | x_{t-1}^{(i)}, y_t),$$
   $$x_t^{(i)} \sim q(x_t | x_{t-1}^{(i)}, y_t, \theta)$$

   2.2 Evaluate the unnormalised importance weights

   $$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\theta | s_{t-1}^{(i)})\, p(x_t^{(i)} | x_{t-1}^{(i)}, \theta)\, p(y_t | x_t^{(i)}, \theta)}{q(\theta | x_{t-1}^{(i)}, y_{1:t})\, q(x_t | x_{t-1}^{(i)}, y_t, \theta)}$$

   2.3 Normalise importance weights $\quad w_t^{(i)} = \dfrac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$

   2.4 Resampling step in needed if $N_{ESS} < N_{thre}$

   2.4.1 Simulate a new sample sets $\{\tilde{x}_t^{(i)}, \tilde{\theta}^{(i)}\}$ with size N from the sets $\{x_t^{(i)}, \theta^{(i)}\}$ according to the importance weights $w_t^{(i)}$.

   2.4.2 Set $w_t^{(i)} = \dfrac{1}{N}$, for $i = 1, \ldots, N$

3.5 Dynamic models for count time series

This section discusses some dynamic models for time series count data used in this study. These dynamic models are dynamic Poisson model, dynamic negative binomial model, and dynamic mixture Poisson model.

3.5.1 Dynamic Poisson Model

The dynamic Poisson model is defined as an extension of the Poisson regression model when its parameter is time-varying. Assume $\{y_t,\, t = 1, \ldots, T\}$ is the time series for count data that follow a Poisson distribution. Then the DGLMs, with an observation process and a transition process, can be expressed as follows:

- Observation model

$$y_t|\lambda_t \sim Pois(\lambda_t) \equiv Pois(\exp(x_t))$$

- Link function

$$\log(\lambda_t) = \eta_t = x_t$$

- State model

$$x_t - \mu = \phi(x_{t-1} - \mu) + \omega_t, \qquad \omega_t \sim N(0, W),$$

where ϕ is defined as first-order autoregressive coefficient and ω_t is an evolution error that follows Normal distribution with mean zero and variance W. Note that the kernel smoothing densities with multivariate Normal and multivariate gamma are used in implementing the LWPF's algorithm as a supporting distribution for the autoregressive coefficient and state error respectively. In addition,

when applying Storvik's algorithm with the dynamic Poisson model, the EKF, described in Section 2.5.3, was used to calculate the approximation of the importance function. Therefore the proposal distribution will be approximated to Gaussian with mean $\mathfrak{m}_t$ and variance $\mathbb{C}_t$ in the following form:

$$\mathfrak{m}_t = \mu(1-\phi)+\phi x_{t-1}+\frac{\exp(\mu(1-\phi)+\phi x_{t-1})\, W\, [y_t - \exp(\mu(1-\phi)+\phi x_{t-1})]}{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, W + \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t)}$$

Where $\omega_t \sim N(0, W)$

$$\mathbb{C}_t = W - \frac{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, [W]^2}{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, W + \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t)}.$$

The derivation of the calculation of the Gaussian approximation of the importance function by using the EKF is given in Appendix A.

3.5.2 Dynamic negative binomial model

Let $\{y_t,\, t = 1, \ldots, T\}$ be the negative binomial observation process of the time series count data. The observation process and the state process, in the non-Gaussian SSM, can therefore be represented as:

- Observation model

$$p(y_t|\tau, \pi_t) \sim NB(\tau, \pi_t) \equiv NB\left(\tau, \frac{1}{1+\exp(x_t)}\right)$$

- Link function

$$\log\left(\frac{1-\pi_t}{\pi_t}\right) = \eta_t = x_t$$

- State model

$$x_t - \mu = \phi(x_{t-1} - \mu) + \omega_t, \qquad \omega_t \sim N(0, W),$$

where τ denotes the dispersion parameter, ϕ is the first-order autoregressive coefficient and ω_t is an evolution error that follows Normal distribution with mean zero and variance W_t.

In the LWPF algorithm, at each time point, the kernel smoothing densities with multivariate Normal are used for simulating autoregressive coefficient particles. The kernel multivariate gamma is used to generate state error and dispersion parameter particles.

3.5.3 Two-component dynamic mixture Poisson model

Suppose $\{y_{1t},\, t = 1, \ldots, T\}$ and $\{y_{2t},\, t = 1, \ldots, T\}$ denote two different time series for count data where each follows a Poisson distribution with parameters λ_{1t}, λ_{2t}. A two-component dynamic mixture Poisson model can be obtained through mixing two single dynamic Poisson models. In particular, for the SSM, the observation and the state processes of the two-component dynamic mixture Poisson model can be represented as follows:

- Observation model

$$\begin{aligned} p(y_t|\lambda_{1t}, \lambda_{2t}) &= \vartheta p(y_{1t}|\lambda_{1t}) + (1-\vartheta)p(y_{2t}|\lambda_{2t}) \\ &= \vartheta Pois(y_{1t}|\lambda_{1t}) + (1-\vartheta)Pois(y_{2t}|\lambda_{2t}) \end{aligned}$$

- State models

$$x_{1t} - \mu_1 = \phi_1(x_{1,t-1} - \mu_1) + \omega_{1t} \qquad \omega_{1t} \sim N(0, W_1)$$

$$x_{2t} - \mu_2 = \phi_2(x_{2,t-1} - \mu_2) + \omega_{2t} \qquad \omega_{2t} \sim N(0, W_2),$$

where (ϕ_1, ϕ_2) are defined as first-order autoregressive coefficients for each state model, ω_{1t}, ω_{1t} are evolution errors of the two state models that follow a Normal distribution with mean zero and variances W_1, W_2 respectively, and ϑ is a mixing parameter which represents the probability of an observation belonging to the first distribution (component). In this study, only the two-component dynamic mixture Poisson model is used with the LWPF algorithm, for both the second simulation and asthma data. A distribution which has support (0,1) is used to sample the mixing parameter, ϑ. The multivariate Beta distributions are assumed for that.

3.5.4 Numerical implementation by MCMC

In order to conduct the MCMC method for estimating the posterior distributions of the Poisson model parameters in the first simulation study, the following independent prior distributions of the autoregressive coefficient ϕ and the process mean μ are employed: $\phi \sim N(\phi|\phi_0, \sigma_\phi^2)$, $\mu \sim N(\mu|\mu_0, \sigma_\mu^2)$. In terms of the state variance W, we use an independent Inverse Gamma as the prior distribution. Beside, in the MCMC applications, we work with the precision parameter ($\psi = W^{-1}$) rather than the state variance. A Gamma distribution will be used as the prior with the shape and rate parameters respectively as follows: $\psi \sim \Gamma(\alpha, \beta)$. After working with the precision parameter, we used its reciprocal to obtain the variance ($W = \psi^{-1}$). In updating the state of the unknown model parameters, the Bayes formula is used for computing the full posterior density for each parameter. The Gibbs sampler and Metropolis-Hastings Algorithms are used for generating a Markov chain based on whether the full conditional distribution of each parameter is in a known form or not.

With respect to the DGLM with generated Poisson data, the joint conditional posterior distribution of the states and the unknown parameters up to proportionality can be written as follows:

$$p(\phi, \mu, \psi, x_{1:t}|y_{1:t}) \propto \prod_{t=1}^{T} p(y_t|x_t, \phi, \mu, \psi) \prod_{t=2}^{T} p(x_t|x_{t-1}, \phi, \mu, \psi)p(\phi)p(\mu)p(\psi)p(x_1) \tag{3.5.1}$$

Then the full conditional posterior distributions for (ϕ and μ) given the states $(x_1, \ldots, x_T)$ are presented as follows:

$$\begin{aligned} p(\phi|x_{1:t}) &\propto p(\phi) \prod_{t=2}^{T} p(x_t|x_{t-1}) \\ &= N(\phi|\phi_0, \sigma_\phi^2) \prod_{t=2}^{T} N(x_t|\mu(1-\phi) + \phi x_{t-1}, \psi) \end{aligned} \tag{3.5.2}$$

$$\begin{aligned} p(\mu|x_{1:t}) &\propto p(\mu) \prod_{t=2}^{T} p(x_t|x_{t-1}) \\ &= N(\mu|\mu_0, \sigma_\mu^2) \prod_{t=2}^{T} N(x_t|\mu(1-\phi) + \phi x_{t-1}, \psi) \end{aligned} \tag{3.5.3}$$

Similarly, we can see that the full conditional posterior distribution of ψ, given the states $x_{1:t}$ is given by

$$\begin{aligned} p(\psi|x_{1:t}) &\propto p(\psi) \prod_{t=2}^{T} p(x_t|x_{t-1}) \\ &\propto \Gamma(\psi|\alpha, \beta) \prod_{t=2}^{T} N(x_t|\mu(1-\phi) + \phi x_{t-1}, \psi) \\ &\propto \Gamma(\psi|\frac{T}{2} + \alpha,\ \beta + \frac{1}{2}\sum_{t=2}^{T}(x_t - \mu(1-\phi) - \phi x_{t-1})^2) \end{aligned} \tag{3.5.4}$$

With respect to Equations (3.5.2)–(3.5.4), the full conditional distribution for the precision parameter ψ is in a known form which is proportional to the

Gamma distribution whereas, the full conditional distributions for the autoregressive coefficient ϕ and the process mean μ are not. Therefore, we use the Gibbs sampler algorithm for generating a Markov chain of the precision parameter ψ and the Metropolis-Hastings Algorithm is used to generate a Markov chain for both the autoregressive coefficient ϕ and the constant mean μ. In terms of applying the M-H algorithm, candidate values for the autoregressive coefficient ϕ and the constant mean μ are sampled using a re-scaled beta and normal proposal distributions respectively.

3.6 Simulation experiments

A simulation study is an important technique of statistical analysis. It enables researchers to evaluate the performance of the proposed methods in achieving the desired results. This section has two simulation studies. In the first simulation study, we will first investigate whether the Liu and West (2001) particle filter algorithm or the Storvik (2002) algorithm with a Poisson proposal model provides more accurate estimates of unknown hyper-parameters. It will also provide robust forecasts: these should be closer to the true values of the simulated data. In addition, the approximations to the posterior distribution of the hyper-parameters will be compared with a good approximation to the true posterior distributions. This posterior distribution can be obtained by running the MCMC algorithm using a huge number of iterations. A purpose of the second simulation study is to see what happens when the fitted model is misspecified. In this study, the Liu and West (2001) algorithm will be applied. We consider three cases, depending on whether the data have been simulated from the Poisson, negative binomial, or mixture Poisson distributions.

3.6.1 Simulation description

3.6.1.1 Simulation study when the fitted model is correct

For this simulation scenario, the dynamic Poisson model, including the observation and state process with the link function, as described in Section 3.5, is applied. A sample of 100 observations was generated from true models with response observations following the Poisson distribution. The dynamic Poisson model with two proposal particle filter algorithms described in Sections 3.3 and 3.4 will be applied to determine how they deal with both state and sequential parameter learning efficiently with the simulated data. We will compare the estimated posterior distributions of hyper-parameters by two proposal particle filter algorithms with the results obtained by the MCMC using a large number of iterations.

In terms of the dynamic Poisson model, the simulated data are generated by using the true parameters $\phi = 0.75$, $W = 0.135$ and $\mu = 0.85$. For two algorithms, the Liu and West and Storvik particle filters, we assume the initial values of hyper-parameters having the same prior. The prior distributions are represented as follows: $W_0 \sim IG(v_0/2, d_0/2), \phi_0 \sim N(\mathbf{m}_0, W_0\mathcal{C}_0)$, and $\mu_0 \sim N(\mathbf{m}_0, W_0\mathcal{C}_0)$. The sufficient statistics are set up as follows: $v_0 = 2$, $d_0 = 1$, $m_0 = 0$, $C_0 = 10$, $D_0 = 1$. In addition, the prior distribution of the initial value of state in both algorithms is $x_0 \sim N(0, 4)$. Regarding the use of the Liu and West algorithm, the value of the shrinkage parameter is determined as $c = 0.975$. In Storvik's algorithm, In order to calculate the importance function by using the EKF, the following setup for the state prior mean and variances was used: $\mathfrak{m}_0 = 0$, $\mathbb{C}_0 = 100$.

A Monte Carlo study with 100 runs is performed to obtain a new set of obser-

vations at each run. The proposed algorithms are demonstrated by the simulated data from a Poisson distribution in order to observe the algorithm performance in tracking the latent variable (state). Also observed is how accurately the parameters are estimated. A multinomial resampling step is used if needed for each particle of the state and hyper-parameters at each time point. A square root of mean square error (RMSE) is used to distinguish between the performance of the proposed algorithms for obtaining an accurate forecast. In two proposed models, a sample of 1000 particles (N = 1000) is used at each time point.

Table 3.3: Summary of the performance of the LW and Storvik algorithms with Poisson model

Filter		
	LW.Poiss	STK.Poiss
Mean/SD ($\hat{\phi}$)	0.6293/0.0933	0.5638/0.1868
Mean/SD ($\hat{W}$)	0.1229/0.1174	0.4121/0.0998
Mean/SD ($\hat{\mu}$)	0.7487/0.1406	0.6271/0.1312
Mean RMES Pred	1.4913	2.988
Mean SD Pred	1.373	2.8951

Table 3.3 provides a summary comparison of the performance accuracy of the Liu and West particle filter with a Poisson model (LW.Poiss) and the Storvik particle filter with a Poisson model (STK.Poiss). According to the estimation of the hyper-parameters, we can see the average Monte Carlo of the posterior mean for the autoregressive coefficient by the LW.Poiss algorithm ($\hat{\phi} = 0.6293$) is closer to the true value than that of the STK.Poiss algorithm ($\hat{\phi} = 0.5638$). In addition, the LW.Poiss algorithm provides a good estimation of the average Monte Carlo of

the posterior mean for the state variance ($\hat{W} = 0.1229$), which is closer to the true value compared to the value generated by the STK.Poiss algorithm ($\hat{W} = 0.4121$). In terms of one-step-ahead forecasting, the average Monte Carlo of the root mean squared error (RMSE) of prediction by the LW.Poiss algorithm (RMSE= 1.4913) is smaller than that of the STK.Poiss algorithm (RMSE=2.988).In addition, the LW.Poiss algorithm provides an average Monte Carlo standard deviation of prediction error of (1.373) which is smaller than the STK.Poiss algorithm equivalent (2.8951). Therefore the Liu and West particle filter offers the best result for forecasting accuracy, compared with the Storvik particle filter. From this comparison part of the simulation study, we can conclude that the LW algorithm with a Poisson model outperforms the Storvik particle filter algorithm.

With respect to the MCMC implementation, we used the same prior distributions of the hyper-parameters as in the particle filters. These prior distributions are suggested by Storvik (2002). The prior distributions are represented as follows: $\phi_0 \sim N(\mathbf{m}_0, W_0\mathcal{C}_0)$, and $\mu_0 \sim N(\mathbf{m}_0, W_0\mathcal{C}_0)$, where $\mathbf{m}_0 = 0$ and $\mathcal{C}_0$=100. We used the gamma prior with shape parameter $v_0/2$ and rate parameter $d_0/2$ for the precision parameter $\psi = W^{-1}$, where $v_0 = 2$ and $d_0 = 1$. We set the total number of iterations to 50000 and the first 10000 samples are set for burn-in. The trace plots of the model parameters is shown in Figure B.1, which is provided in the appendix B.1. These plots provide evidence of convergence of the chain, although the trace plot of the mean parameter μ includes some autocorrelation.

Figure 3.1 shows histograms of the samples of the approximation of the posterior distributions of the hyper-parameters at the last point of time (T=100) obtained from the two proposal particle filter algorithms and the MCMC. The histograms displayed in the top panels of Figure 3.1 provide the approximate posterior distributions of the hyper-parameters obtained by MCMC. The histograms

of the posterior distributions for the hyper-parameters obtained by the LW.Poiss particle filter are provided in the middle panels. In the bottom panel, the histograms of the posterior distributions for the hyper-parameters obtained by the STK.Poiss particle filter are exhibited.

Now, we are going to compare the posterior distributions of each hyper-parameters obtained using the two online methods (LW.Poiss and STK.Poiss particle filters) and one offline method (MCMC). A visual inspection of the graphs for the parameter ϕ indicates that the shapes of the posterior distributions in both MCMC and LW.Poiss particle filter have some similarity whereas STK.Poiss particle filter produced a posterior distribution which appears to be further apart from that of MCMC. Similar comments apply for the estimation of the parameters W and μ. Furthermore, we note that for μ both LW.Poiss and STK.Poiss particle filters provide distributions with smaller tail probability compared to the MCMC. In conclusion, Figure 3.1 suggests that both LW.Poiss and STK.Poiss particle filters appear to approximate reasonably the shape of the posterior distribution produced by MCMC.

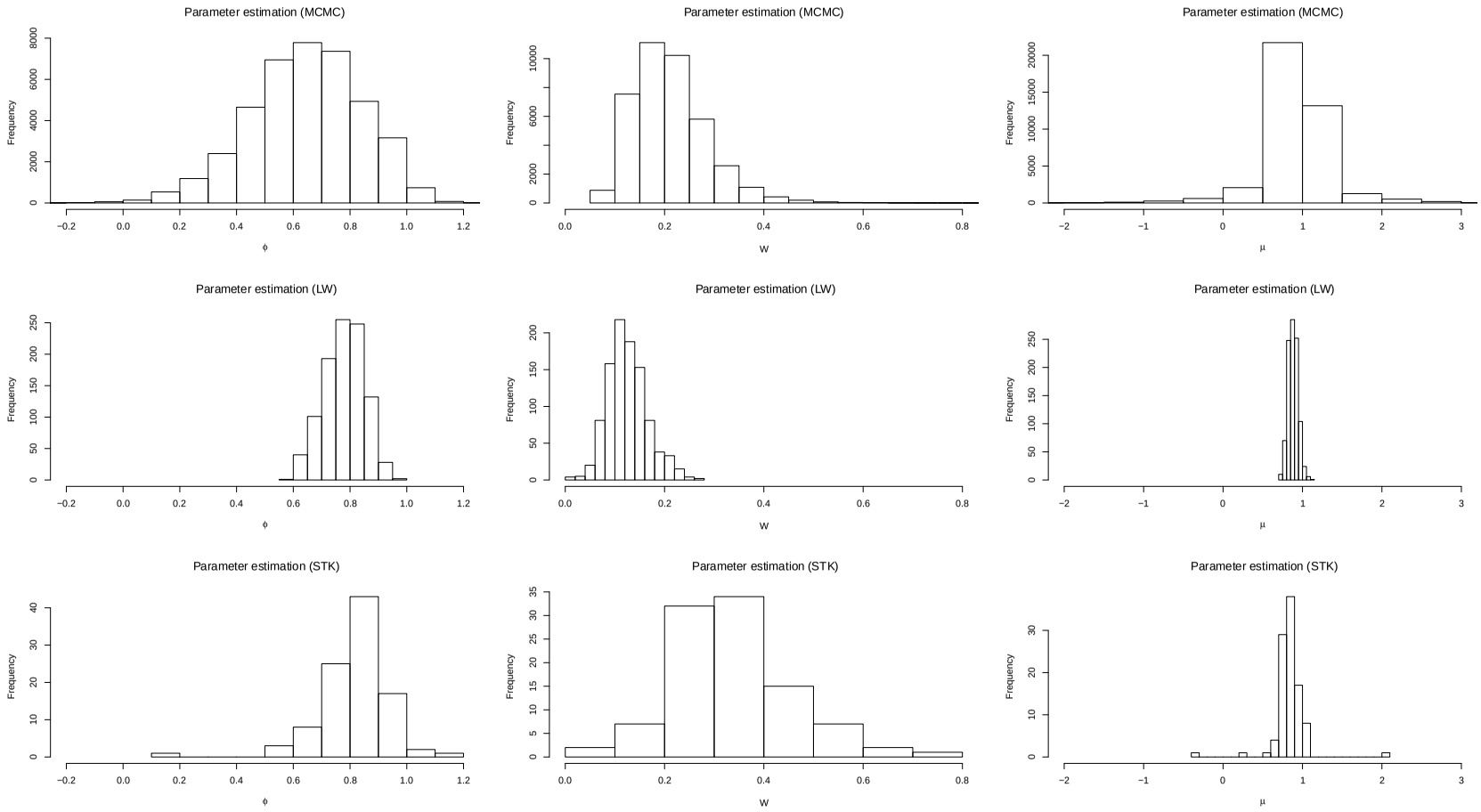

Figure 3.1: *Histograms of the approximation posterior distributions of the Poisson model parameters from the MCMC and two proposal particle filter algorithms for Poisson simulated data*

3.6.1.2 Simulation study when the fitted model is misspesified

The second simulation study investigate what happens when the fitted model is misspecified. In this study, the Liu and West (2001) algorithm will be applied. Three cases are considered, depending on whether the data have been simulated from the Poisson, negative binomial, or mixture Poisson distributions. With respect to compare the performance of the Liu and West algorithm with a Poisson model, different distributions for a response variable such as a negative binomial and mixture Poisson are proposed. In the simulation study using the Poisson and negative binomial model, the true parameters in the simulation study are defined as follows: $\phi = 0.75$, $W = 0.135$ and $\mu = 0.85$. In additional, We set up $\tau = 1$ for a true value of a dispersion parameter in negative binomial model. Moreover, when applying the LW algorithm with Poisson and negative binomial model, the prior distributions of the initial value of the hyper-parameters are specified as follows: $\phi \sim N(0, 0.25)$, $W \sim Unif(0, 2)$, $\mu \sim N(0, 0.25)$. The prior distribution of the initial value of state is $x_0 \sim N(0, 4)$, and the value of the shrinkage parameter is determined as $c = 0.975$. In addition, the prior distribution of the dispersion parameter is determined as $\tau \sim Unif(0, 2)$. In terms of the mixture Poisson model, the data is simulated with the true value as follows: $\phi_1 = 0.75$, $\phi_2 = 0.65$, $W_1 = 0.135$, $W_2 = 0.235$, $\mu_1 = 0.85$, $\mu_2 = 0.75$, $\vartheta = 0.6$. Furthermore, the prior distribution for the static parameters and the initial states are set up as follows: $\phi_1 \sim N(0, 0.25)$, $\phi_2 \sim N(0, 0.5)$, $W_1 \sim Unif(0, 2)$, $W_2 \sim Unif(0, 1)$, $\mu_1 \sim N(0, 0.25)$, $\mu_2 \sim N(0, 0.5)$, $\vartheta \sim Beta(2, 1)$, $x_{01} \sim N(0, 4)$, $x_{02} \sim N(0, 2)$.

In the second simulation study, the accuracy of the estimation of hyper-parameters as well as the one-step-ahead forecasting, when the fitted model is misspecified, are investigated. In other words, if there is simulation data from a Poisson distribution and the wrong model, such as the negative binomial or mix-

ture Poisson is applied, what then is the precision of the parameter estimation and prediction?. As in the first simulation study, the different 100 datasets from each proposed distribution are simulated. The misspecified models are applied to fit the simulated datasets to estimate the state and the hyper-parameters. The output results from using the wrong models are compared with the estimated parameters and the one-step-ahead prediction by the correct model. Tables 3.4–3.6 provide a summary comparison for an accuracy performance of the Liu and West algorithm with a correct model and the misspecified models. As can be observed from Table 3.4, the LW.Poiss algorithm appears to provide a reasonable inference mechanism for estimating the posterior mean of all unknown hyper-parameters, even if it is considered as the wrong model. In other words, the estimation of the posterior mean of the hyper-parameters is quite close to the true parameters $\phi = 0.75$, $W = 0.135$ and $\mu = 0.85$. Furthermore, in terms of prediction, the average Monte Carlo of the RMSEs of one-step-ahead forecasting by applying the LW.Poiss algorithm are small, whether is it the correct model or the wrong model (1.1561,1.3752,1.2086) and quite similar.

Table 3.4: Summary comparison of a Monte Carlo study of LW.Pois algorithm for different simulated data

Simulated Data			
	Sim.Poiss	Sim.NB	Sim.MPoiss
Mean/SD ($\hat{\phi}$)	0.6447/0.0199	0.5941/0.0714	0.5622/0.0415
Mean/SD ($\hat{W}$)	0.1213/0.0563	0.1137/0.0318	0.1071/0.0492
Mean/SD ($\hat{\mu}$)	0.7411/0.0537	0.7103/0.0711	0.7056/0.0783
Mean RMES Pred	1.1561	1.3752	1.2086
Mean SD Pred	1.1320	1.3645	1.2008

This means that the LW.Poiss algorithm offers a more precise estimation and efficient predictions, regardless of whether under the correct or wrong models. Further, the average Monte Carlo standard deviation of residuals calculated by LW.Poiss is small. With respect to forecasting, Figure 3.2 shows a comparison between Poisson simulated data versus one-step-ahead prediction by applying LW.Poiss as a current model and LW.Neg.bin and LW.MPois as wrong models in one run.

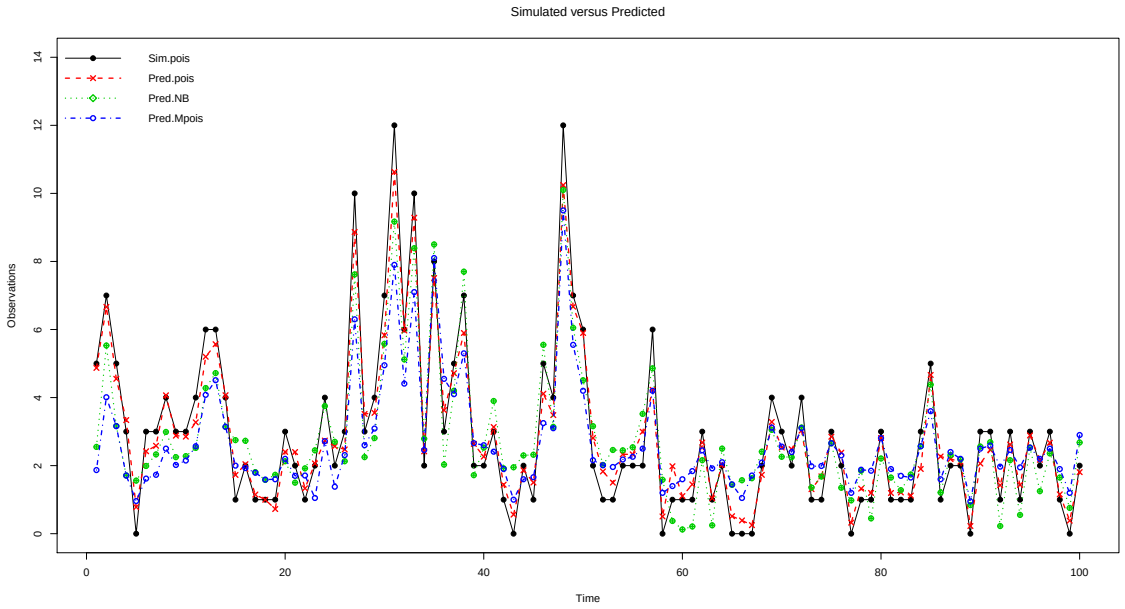

Figure 3.2: *Comparison between Poisson simulated data versus forecasting by correct and wrong models in one run*

A summary comparison of the posterior means of applying the LW.NB, both as a correct or wrong model, is provided in Table 3.5. The resulting estimates indicate that the estimated posterior means for the autoregressive coefficient (ϕ), state variance (W), and location parameter (μ) are reasonable estimates. In addition, the algorithm offers a reasonable estimate for the dispersion parameter (τ). This means that they varied considerably from the true values. Moreover,

the average of Monte Carlo for RMSE and the standard deviation of residuals are quite similar.

Table 3.5: Summary comparison of a Monte Carlo study of LW.NB algorithm for different simulated data

Simulated Data			
	Sim.Poiss	Sim.NB	Sim.MPoiss
Mean/SD ($\hat{\phi}$)	0.5322/0.0247	0.5117/0.0716	0.5156/0.0513
Mean/SD ($\hat{W}$)	0.1054/0.0431	0.0996/0.0535	0.1008/0.0622
Mean/SD ($\hat{\mu}$)	0.6311/0.0471	0.5922/0.0421	0.5956/0.0539
Mean/SD ($\hat{\tau}$)	0.8125/0.0752	0.8713/0.0205	0.7803/0.0332
Mean RMES Pred	2.176	2.033	2.243
Mean SD Pred	2.055	2.004	2.125

Figure 3.3 shows a comparison between negative binomial simulated data versus one-step-ahead prediction by applying LW.NB as a right model and LW.Poiss and LW.MPois as misspecified models in one run. It can be seen that when the LW.Poiss filter is used as a misspecified model, it provides the predicted values close to the real values.

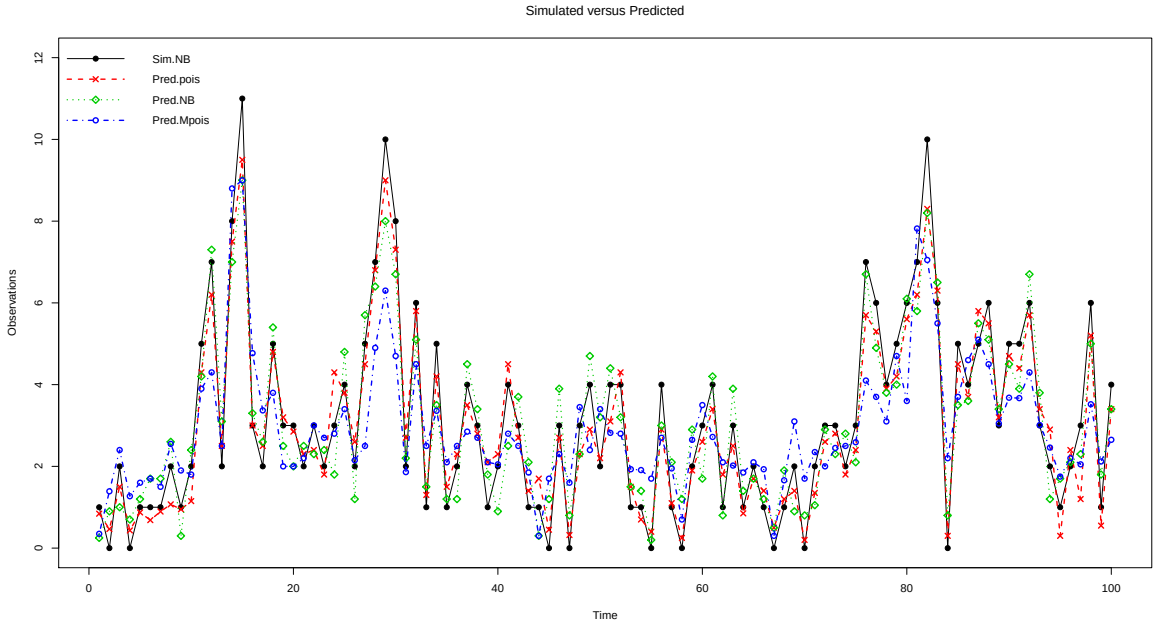

Figure 3.3: *Comparison between negative binomial simulated data versus forecasting by correct and wrong models in one run*

Table 3.6 presents a summary comparison for parameter estimation and forecasting accuracy by applying the LW.MPoiss, both as a 'correct' or a 'wrong' model. The resulting estimates indicate that the algorithm provides a reasonable estimate of the posterior mean for the autoregressive coefficients, state variances, the location parameters, and the mixing parameter. Moreover, the average of Monte Carlo for RMSE and the standard deviation of residuals are large and quite similar. Figure 3.4 shows a comparison between mixture Poisson simulated data versus one-step-ahead forecasting by applying LW.Poiss and LW.NB as wrong models and LW.MPois as a current model in one run.

Table 3.6: Summary comparison of a Monte Carlo study of LW.MPois algorithm for different simulated data

Simulated Data			
	Sim.Poiss	Sim.NB	Sim.MPoiss
Mean/SD ($\hat{\phi_1}$)	0.5211/0.0527	0.4002/0.0326	0.535/0.0192
Mean/SD ($\hat{\phi_2}$)	0.4735/0.0512	0.4822/0.0221	0.4656/0.0335
Mean/SD ($\hat{W_1}$)	0.1301/0.0425	0.1299/0.0854	0.1382/0.0452
Mean/SD ($\hat{W_2}$)	0.1921/0.0259	0.2001/0.0432	0.1775/0.0277
Mean/SD ($\hat{\mu_1}$)	0.5922/0.0442	5236/0.0155	0.6008/0.0585
Mean/SD ($\hat{\mu_2}$)	0.4778/0.0512	0.4412/0.0559	0.5301/0.0228
Mean/SD ($\hat{\vartheta}$)	0.7094/0.0854	0.5014/0.0649	0.5288/0.0845
Mean RMES Pred	2.823	2.682	2.551
Mean SD Pred	2.651	2.484	2.436

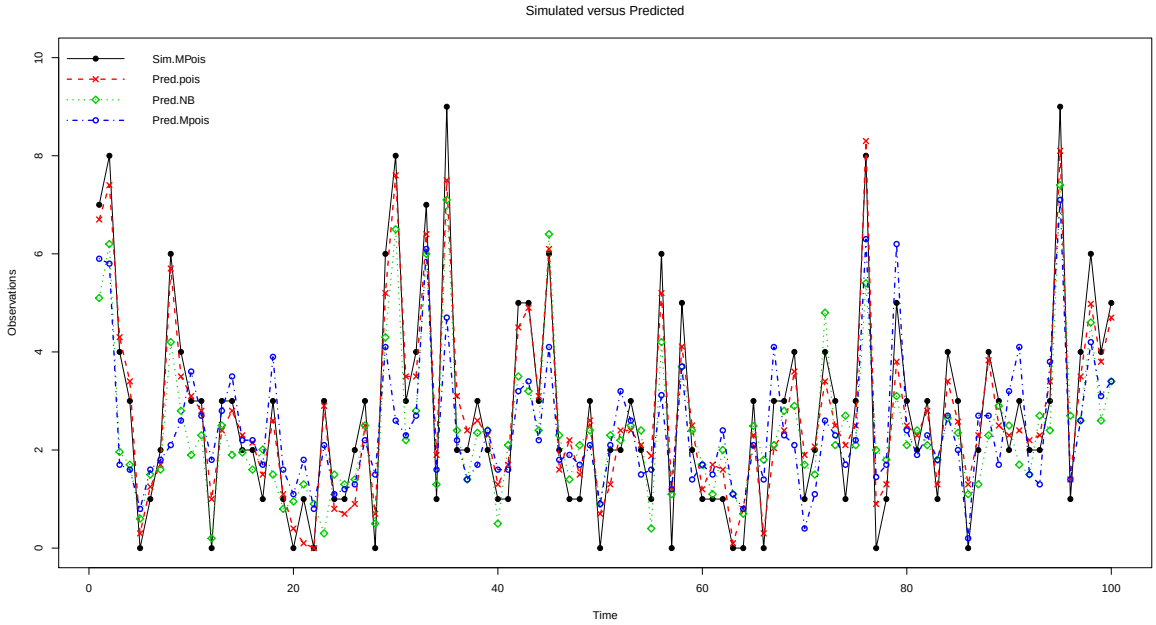

Figure 3.4: *Comparison between mixture Poisson simulated data versus forecasting by correct and wrong models in one run*

Overall, based on this simulation study, the sample output results from the proposed models illustrate that the Liu and West particle filter with Poisson model (LW.Poiss) provides a more precise estimation of unknown parameters and provides efficient one-step-ahead forecasting.

3.7 Analysis of asthma data

After testing the performance of the LW particle filter with different proposed algorithms and the STK.Poiss algorithm on the simulated data, we want to apply these different proposed algorithms to the medical data. The dataset in this study originates from the General Practice Research Database (GPRD) in England. This medical dataset is the same as in Julious et al. (2011) and it consists of the

daily medical contacts for schoolchildren aged between 5 and 16 years old who suffered from asthma over a seven-year period between 1999 and 2005 in England.

Figure 3.5 shows the daily and weekly total medical contacts for asthmatic children. Since the preliminary goal of this study is forecasting, weekly totals of medical contacts were used instead of daily medical contacts; since weekly forecasts are sufficient. The goal of this study is a one step ahead forecasting. So this is a reason why we dealt with the weekly data instead of daily data. In other words, in daily data it is quite difficult to see what happened and give an advice for a short time (one day). However, in weekly data, there is enough time (one week) for people to do something for increasing of medical contacts. It can be seen from the Figure 3.5 that the weekly data does not seem to be stationary. Therefore, adding a constant mean (μ) would not be appropriate. Thus, the state model for all proposed models will not contain the location parameter (μ). However, we will keep the autoregressive coefficient as it is in order to verify this effect.

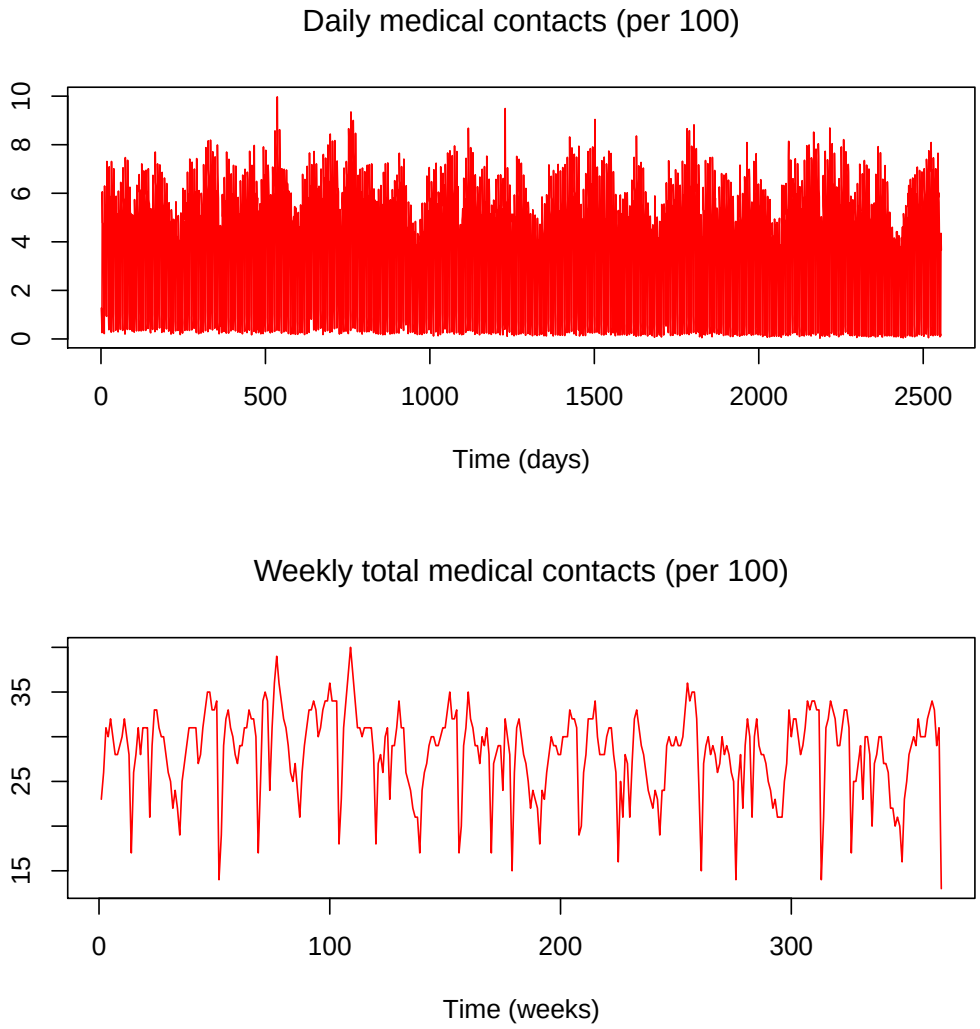

Figure 3.5: *Daily and weekly total medical contacts for asthmatic children*

In this section the Liu and West and Storvik particle filter algorithms are applied to asthma time series data with the proposed models as in the simulation study without including the process mean within the proposal models. We use the number of particles as N=1000. In terms of the Poisson and Negative binomial models, the prior distributions of the hyper-parameters and the initial value of the state are specified as follows: $\phi \sim N(0, 0.25)$, $W \sim Unif(0, 2)$

and $x_0 \sim N(0, 4)$. Beside,the prior distribution of the dispersion parameter is determined as $\tau \sim Unif(0, 2)$. With respect of the mixture Poisson model,the prior distribution for the static parameters and the initial states are set up as follows: $\phi_1 \sim N(0, 0.25)$, $\phi_2 \sim N(0, 0.5)$, $W_1 \sim Unif(0, 2)$, $W_2 \sim Unif(0, 1)$, $\vartheta \sim Beta(2, 1)$, $x_{01} \sim N(0, 4)$, and $x_{02} \sim N(0, 2)$. The value of the shrinkage parameter is determined as $c = 0.975$ for all proposal models. Figure 3.6 exhibits the final 105 observations of the real data together with one-week-ahead forecasting by using the Liu and West algorithm with all different proposed models. As shown by this plot, the LW with Poisson model gives a high number of forecasts closer to the real data than does the negative binomial. This means the Poisson model can successfully be used to capture dynamic variations in the data.

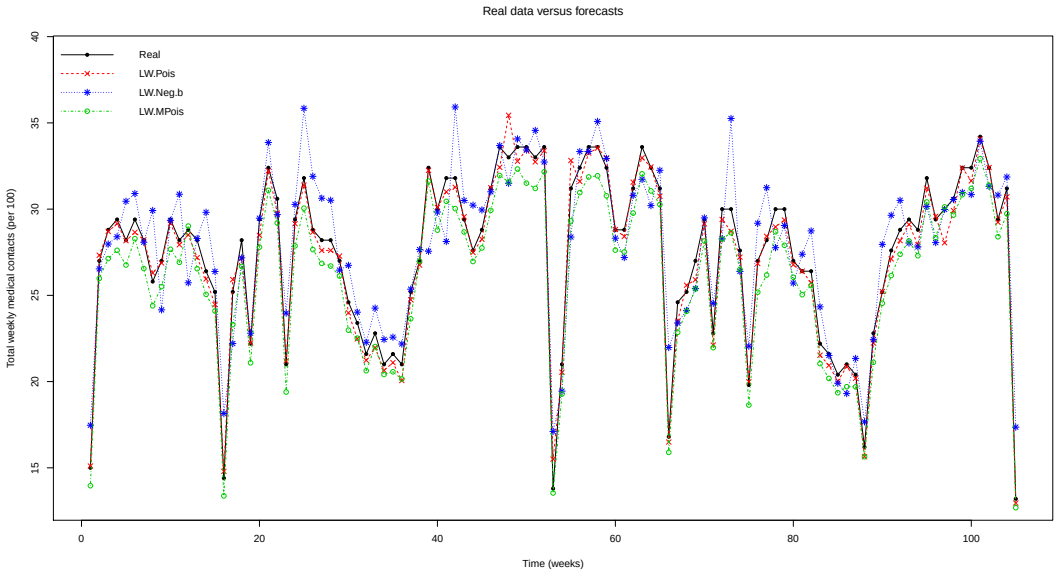

Figure 3.6: *Real data (black solid line with bullet) against one step ahead forecasts by LW.poisson model (red dashed line with cross), by LW. negative binomial model (blue dotted line with circle), and LW.poisson model (green dashed line with star)*

Since the objective of this study is to make a forecast, the performance of

the different proposed algorithms is evaluated based on a RMSE value of the one-step-ahead prediction error. The residuals sequence is used to calculate the measurement. This is defined as a difference between a real and predicted value at each time. Table 3.7 provides a summary comparison for the accuracy of these proposed models. The mean absolute deviation (MAD) and the standard deviation of residuals are also shown. According to these figures, the LW.Poiss algorithm exhibits the smallest RMSE (1.082) followed by the LW.MPoiss algorithm (2.226), and the LW.Neg.bin algorithm (3.340). However, the error prediction by Storvik's algorithm with Poisson model (STK.Poiss) is the highest (6.440). In terms of the standard deviation of the residuals, the LW.MPoiss algorithm has the lowest value compared to other models. Moreover, the standard deviation of the residuals by the STK.Poiss algorithm again is the highest. Furthermore, performances with respect to the MAD of all the models shows a similar behaviour as the RMSE. From this comparison, it is proposed that the LW.Poiss algorithm outperforms the other models. Therefore, it may be considered as an appropriate model for capturing the dynamic behaviour in asthma time series data or describing the dependency structure in asthma time series data well.

Table 3.7: Performance summary of all algorithms

Filter	RMSE	SD.resduals	MAD
LW.Poiss	1.082	0.972	0.822
LW.Neg.bin	3.340	3.270	2.659
LW.MPoiss	2.226	0.727	2.116
STK.poiss	6.440	6.141	5.108

The output results of the estimated marginal posterior means of the static parameters by the LW.Poiss, LW.Neg.bin, and LW.MPoiss particle filter algorithms

are shown in Figures 3.7–3.9. A dynamism in the time series of the estimated posterior mean of the state and the autoregressive parameter ϕ and the state variance W by the LW.Poiss model are shown in Figure 3.7. The histograms of the posterior distributions of the autoregressive parameter ϕ and the state variance W at the last time point (T=365)are also provided in Figure 3.7. As shown by the plot in the lift panels, the sequence of the posterior means for both parameters estimation converge rapidly to the fixed value after 20 weeks. This means that the system only required twenty weeks of training data. As a result, the estimated sequence of the hyper-parameters has become constant over time. The histograms of the posterior distributions of the autoregressive parameter ϕ and the state variance W at the last time point (T=365) are shown in the right panels.

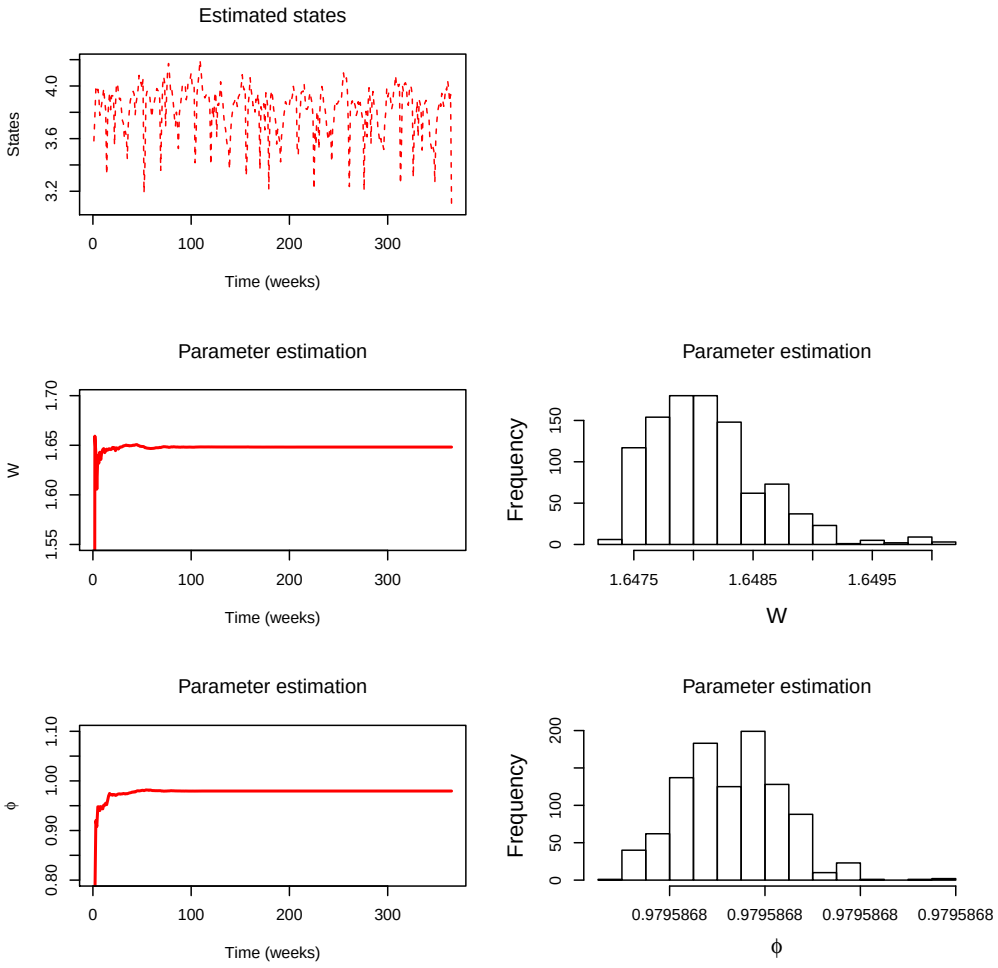

Figure 3.7: *The estimated mean of the state process and hyper-parameters and the posterior histograms of hyper-parameters at the last time point (t=365) by LW.Poiss model*

The estimated posterior mean of the states and all model parameters by the LW.Neg.bin particle filter are exhibited in the left panels in Figure 3.8. The histograms of the posterior distributions of the model parameters at the last time point (T=365) are also shown in the right panels in Figure 3.8. It can be seen from the plot that he estimated posterior mean of the autoregressive parameter ϕ and the state variance W and the dispersion parameter τ are converging to

the fixed values. As can be observed from Figures 3.7–3.8, the fluctuations in the estimated states by applying the LW.Poiss and LW.Neg.bin particle filter algorithms are quite similar, oscillating between 3.2 and 4.2.

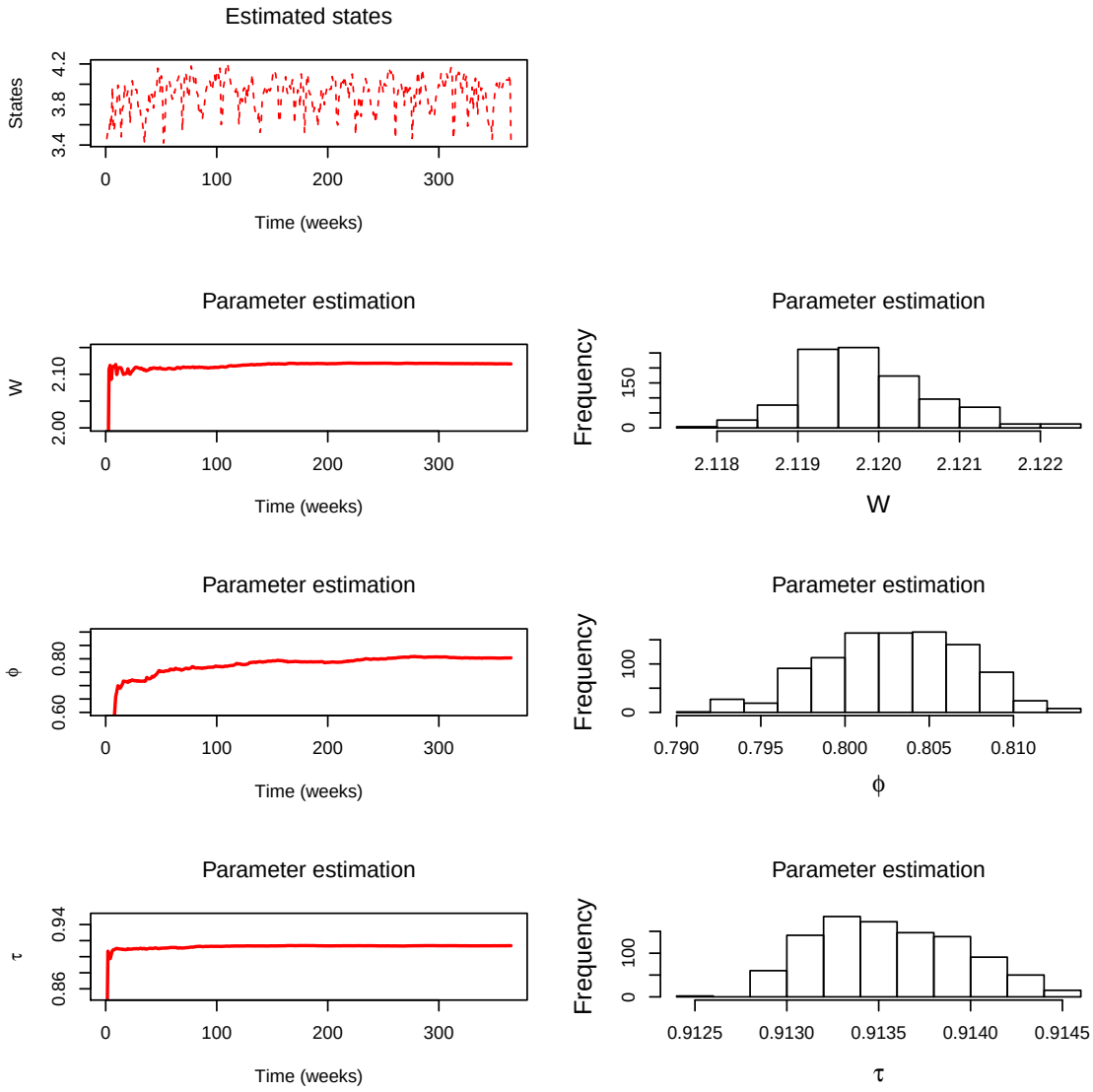

Figure 3.8: *The estimated posterior mean of the state process and hyper-parameters and the posterior histograms of hyper-parameters at the last time point (t=365) by LW.Neg.bin model*

Figure 3.9 displays the output results of the sequential parameter learning by applying the LAW.MPoiss particle filter algorithm to the asthma data. The posterior means of the vector of the model parameters $\theta = (\phi_1, \phi_2, W_1, W_2, \vartheta)$ are presented in the left panels in Figure 3.9. As can be seen from the graph, there are obvious fluctuations in all sequences in the beginning, except for ϕ_2, before converging to the fixed value. Moreover, they required more training approximately 100 weeks more than other algorithms to achieve a convergence. In terms of the mixing parameter we can see from Figure 3.9 that the mean posterior of the mixing parameter ϑ is very small which is close to zero (0.035). In addition, the posterior mean of ϕ_2 is large, whereas, the posterior mean of ϕ_1 is small (-0.11). These indicate that the second component has significant influence, dominating the mixture model. Therefore, in this case the mixture model is almost reduced to one component model. The histograms of the posterior distributions of all model parameters at the last time point (T=365) are also provided in the right panels in Figure 3.9.

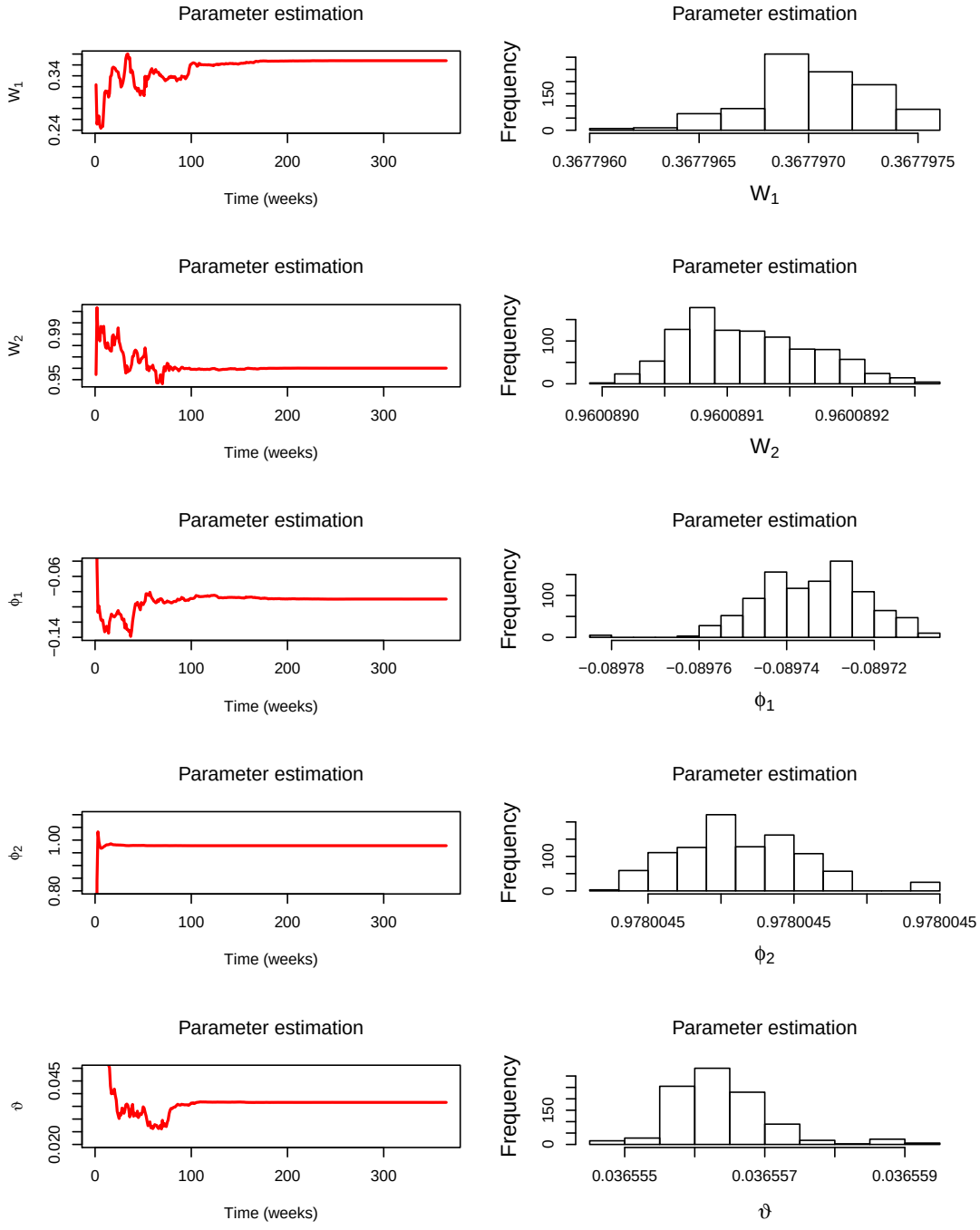

Figure 3.9: *The estimated posterior mean of the hyper-parameters and the posterior histograms of hyper-parameters at the last time point (t=365) by LW.MPoiss model*

Figure 3.10 presents the comparison of the histograms of the predictive particles from both models against the real values at different time points. As can be observed from this figure, the histograms from three proposed algorithms show that the actual value lies in the highest probability of the predictive particles. In other words the majority of the predicted values are closer to the maximum of frequency with a high bar probability.

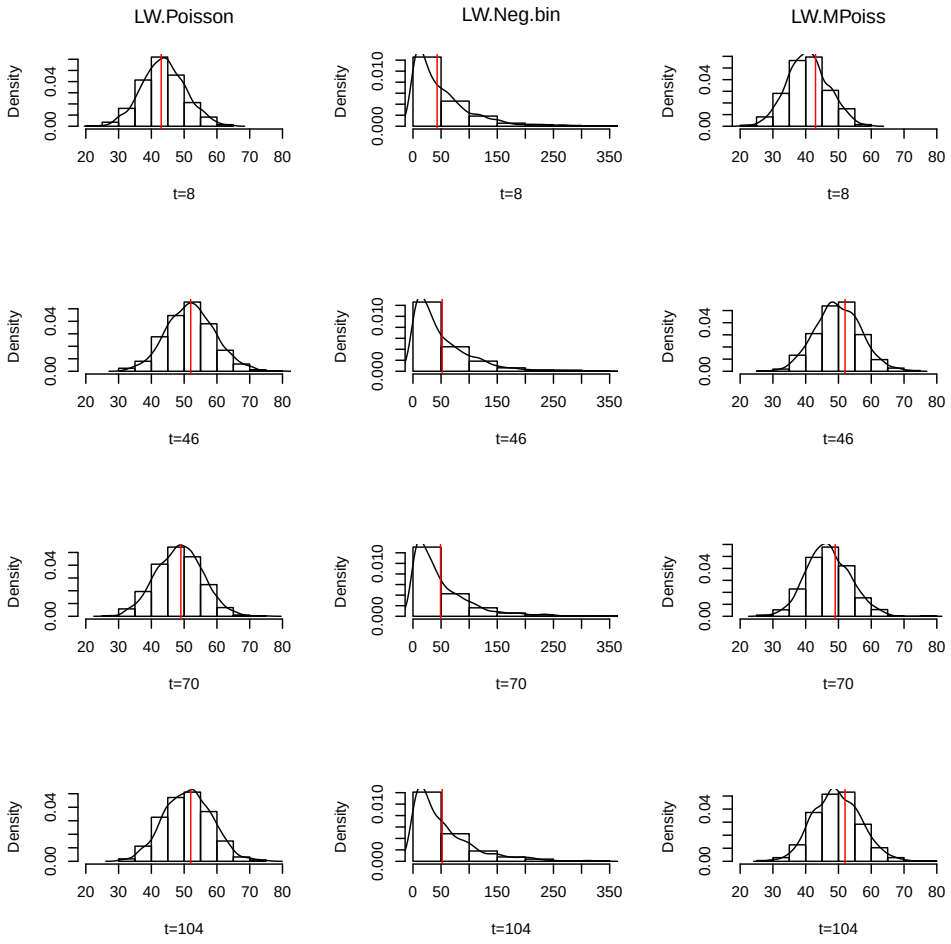

Figure 3.10: *Histograms of the predictive particles generated by different models*

In the model for the asthma data, we used the autoregressive model of order one for the state equation as follows: $x_t = \phi x_{t-1} + \omega_t$. Because the asthma data is not stationary, we expected the estimate ϕ will be very high which is closer to 1. Alternatively, we can take the first order differences in order to make the process as stationary and then we can apply the model $(x_t - x_{t-1}) = \phi(x_{t-1} - x_{t-2}) + \omega_t$. In that case, we expect the estimate ϕ will be small. Below we will show that these two approaches provide a similar result for the estimate ϕ.

With respect to clarify the nature of the results obtained for asthma data, we will provide some comments through experimental study with simulated data . To illustrate this, we will simulate some nonstationary time series data from a random walk model. Then we will take the first differences of the data in order to have a stationary time series as follows: $x_t - x_{t-1} = \phi(x_{t-1} - x_{t-2}) + \omega_t$. The second step is to fit this stationary data and estimate the autoregressive parameter ϕ.

On the other hand, we can rearrange the stationary model with the first differences as follows: $x_t = (\phi + 1)x_{t-1} - \phi x_{t-2} + \omega_t$, which the second part ϕx_{t-2} can be neglected (assuming ϕ is small). Therefore, the above model will become $x_t = \phi^* x_{t-1} + \omega_t$, where $\phi* = (\phi + 1)$. We will fit the nonstationary data and estimate the autoregressive parameter $\phi*$. This estimated parameter $\phi*$ will be equal to the estimated parameter $\phi + 1$ by using $(x_t - x_{t-1}) = \phi(x_{t-1} - x_{t-2}) + \omega_t$. From here we can conclude that two estimated parameters $\phi*$ and ϕ will be similar.

In order to illustrate this we generated a nonstationary time series data with size 100 from the random walk and we fitted the data by using $x_t = \phi * x_{t-1} + \omega_t$. The obtained result for the estimated parameter was $\phi* = 0.978$. On the other hand we used the stationary model with the first differences to fit the data and

we found the estimated parameter $\phi = -0.0037$ which is small. By adding 1 to the estimated parameter ϕ, (-0.0037 + 1= 0.996) we found that the two estimates ϕ and ϕ^* are similar.

3.8 Analysis of Sudden Infant Deaths Syndrome data (SIDS)

The purpose of this section is to apply the methodology of SSMs and a recursive Bayesian estimation method based on the Liu and West (2001) particle filter algorithm with a Poisson model. The aim is to identify the impact of environmental temperature on the sudden deaths of infants. To achieve this purpose, the same medical dataset used by Campbell (1994) was considered. This dataset is sourced from the Office of Population Censuses and Surveys (OPCS) in London. It gives the daily SIDS figures in England and Wales over a three-year period from 1981 to 1983. Daily temperature records for the same period were obtained from the London Weather Centre.

In this section of the study, an extension to the previous model is made by adding a set of covariates within the SSM. This set of covariates includes the environmental temperature z_t, trend and seasonal components. The linear predictor can therefore be defined as an aggregation of all elements within this set of covariates. A time series of environmental temperature was considered as covariate, since exposure to this environmental temperature can increase the risk of SIDS (Campbell, 1994). The daily and weekly total of sudden deaths in the infant data and an average daily and weekly average temperature are plotted in Figure 3.11. The figure shows that there was a negative relationship between the

environmental temperature and the sudden death of infants. This suggests that increasing numbers of infant deaths were due to a decrease in average temperatures in the previous three days, and vice versa. To examine the dynamic in the data: the graph shows that the dataset fluctuates around a mean with a clear evidence of evolution over time.

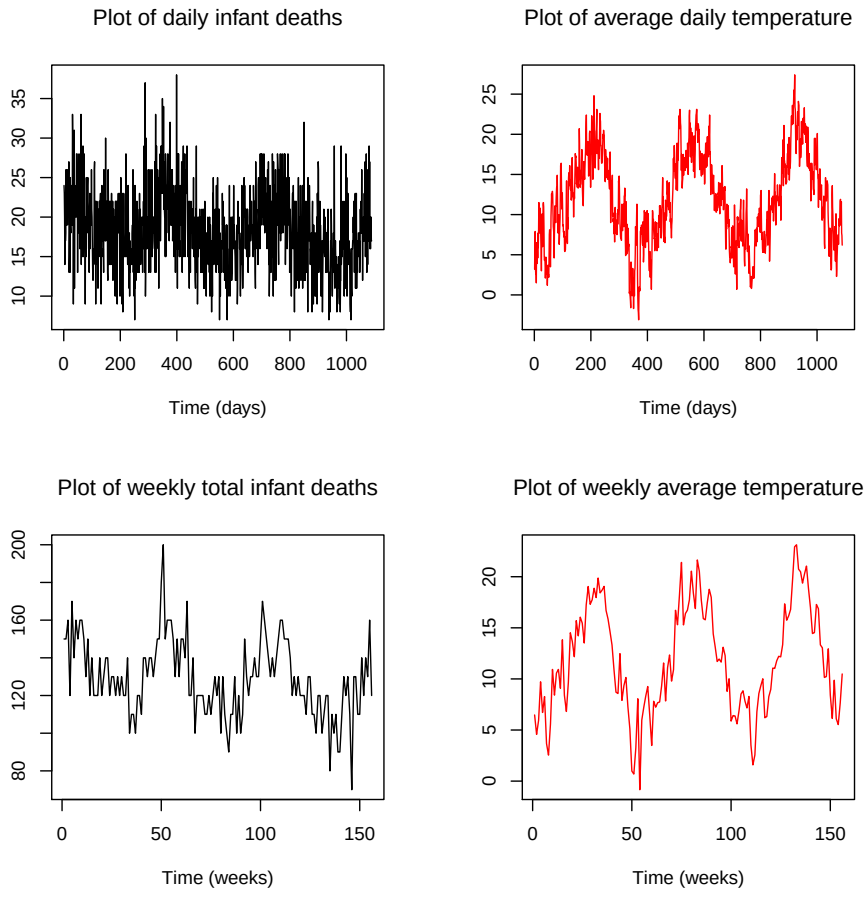

Figure 3.11: *Data description of sudden infant deaths syndrome*

Figure 3.11 shows that the SIDS data exhibit seasonality. There may also be a dependence on temperature in addition to the seasonal effects. Another interesting finding is that the weekly total sudden deaths of infants in each year during three years exhibited a similar pattern of seasonality, but the level differed.

Therefore, the trend-seasonal DGLM with a Poisson model can be adopted to deal with the SIDS time series data. Thus, in this case, the model will consist of a component of trend, a seasonal component, and covariate as environmental temperature.

The seasonal components, which are responsible for fluctuations in time series data, depend on the number of their harmonic components. These harmonics are used to describe a variation in the SSM. In the weekly total data, the seasonal cycle is considered to be equal to 52 weeks. To describe the seasonality in the SIDS data, a reduced model with five harmonic components is used. This is similar to the Triantafyllopoulos (2009) example for tourist arrival data, where only five rather than all harmonic components are used. This results in an SSM with a 13-dimensional state process. It is difficult to implement the full seasonal form DGLM, which needs 52 harmonic components. Furthermore, a transition matrix, which is responsible for an evolution of the states corresponding to the trend, the seasonal components, and the covariate, will have a large dimension of 107×107. It is therefore difficult to apply. In any case, the first harmonic has the most information required to obtain the desired results.

As noted in the previous section, the Liu and West particle filter Poisson outperforms the negative binomial observations for providing accuracy of predictions and precise estimations of the hyper-parameters. Subsequently, the DGLM with a Poisson distribution for a response variable was used in order to achieve the target. The formulation of the extension of the SSM with a 13-dimensional latent process, including a temperature, seasonality components, and a first-order polynomial trend is:

- Observation model

$$y_t|\lambda_t = y_t|\exp(\eta_t) \sim Pois(\exp(\eta_t)),$$

where the linear predictor η is given by

$$\eta_t \;=\; F_t'\mathbf{x}_t = [z_t, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0][x_{1,t}, x_{2,t}, \ldots, x_{13,t}]'$$

- State model

$$x_t = \begin{pmatrix} \phi & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & J_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & J_2(r\pi/26) & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 & J_2(r\pi/26) \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \\ \vdots \\ x_{13,t-1} \end{pmatrix} + \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \vdots \\ \omega_{13,t} \end{pmatrix}$$

where J_1, responsible for the variation of trend, is referred as a Jordan block. Furthermore, $J_t(\varphi)$ denotes the component of the harmonic which is responsible for the seasonal variation in the time series data. The general form of the Jordan block and harmonic component are given in the following:

$$J_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \textit{and} \qquad J_2(\varphi) = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

where $\varphi = 2r\pi/\mathfrak{c}$, $\varphi \in (0, \pi)$ is known as a frequency of the time series data, and r^{th}, $r = 1, \ldots, 5$ is the number of harmonic components in the model. Moreover, $\mathfrak{c}$ is a cycle period of the data. The class of the SSM described above is referred to as the trend-seasonal model for seasonal variation.

The Liu and West (2001) particle filter algorithm was applied to estimate the states and determine the hyper-parameters, including the covariate coefficient and the state variances. To implement this algorithm, the prior distribution of

the initial value of the temperature coefficient is given by $\phi \sim N(0, 0.25)$, and all initial values of the states errors are determined with the same prior as a uniform distribution as follows: $\omega_k \sim Unif(0, 2)$, $(k = 1, \ldots, 13)$. Additionally, the prior distribution of the vector of the initial states, $\mathbf{x_0}$, is specified with a multivariate normal distribution with the mean vector, $\mathbf{m}_0$, and covariance matrix, $\mathbf{C_0}$, as follows: $\mathbf{x} \sim MVN(\mathbf{m}_0, \mathbf{C_0})$ where $\mathbf{m}_0 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$, and $\mathbf{C_0} = 1000 * \text{diag}(13)$. The number of particles at each time is specified as N=1000, and the shrinkage parameter is set as $c = 0.975$. In this analysis, we suggested a change in temperature affects the sudden mortality rate for infants after 3 days. In this case, time-lag with 3 days has been assumed between a change in temperature and a sudden mortality rate for infants.

The performance of the Liu and West algorithm is demonstrated in Figures 3.12–3.13. The time series trajectories of the mean of the approximation of the marginal posterior distributions of the environmental temperature coefficient and the state error are described in Figure 3.12 and Figure 3.13. As can be observed from these plots, the estimated posterior mean of the temperature coefficient converges to the negative value ($\hat{\phi} = -0.0354$). This negative value for the estimation of covariate coefficient indicates the presence of the negative relationship between temperature and the sudden mortality rate for infants. This suggests that increases in the rate of infant mortality are caused by temperatures decreases and this often happens in the winter. This result for the coefficient of the average of the temperature is close to the results obtained by Campbell (1994) using the method from Zeger (1988) which gave ($\hat{\phi} = -0.041$), and the generalised Poisson model by Nelder and Wedderburn (1972), which resulted in ($\hat{\phi} = -0.045$).

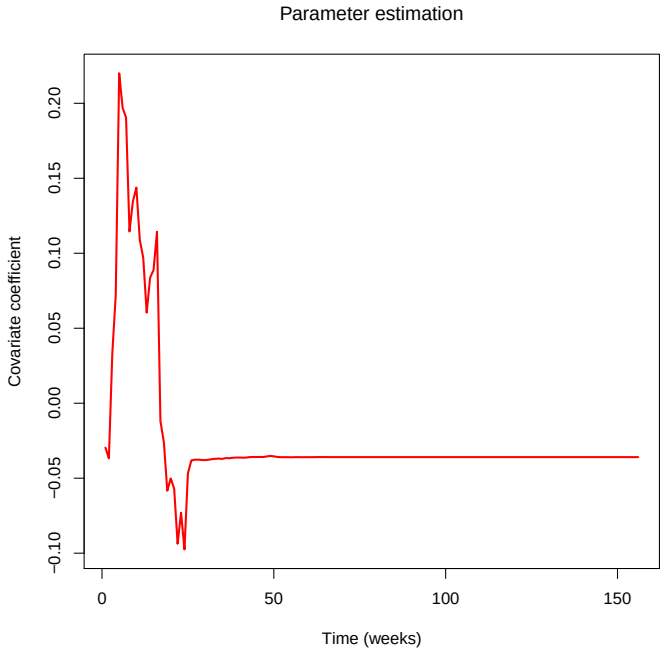

Figure 3.12: *Estimation of covariate (temperature) coefficient*

In terms of the estimation result for the temperature coefficient ϕ, there are two features that should be observed from the plot. The first finding is that the LWPF algorithm requires only the first forty observation from the dataset to reach this estimation. However, the whole dataset has been used in other methods by Campbell (1994). Another finding is that, as the model parameter is assumed to be a constant, the plot shows the estimated coefficient of temperature is changing a lot in the first period of time before becoming stable at a constant value, perhaps at time point (t=40). This implies that the posterior mean of the coefficient changes over time. Nevertheless, the other methods only offer an estimate without knowing whether the coefficient was constant or time varying.

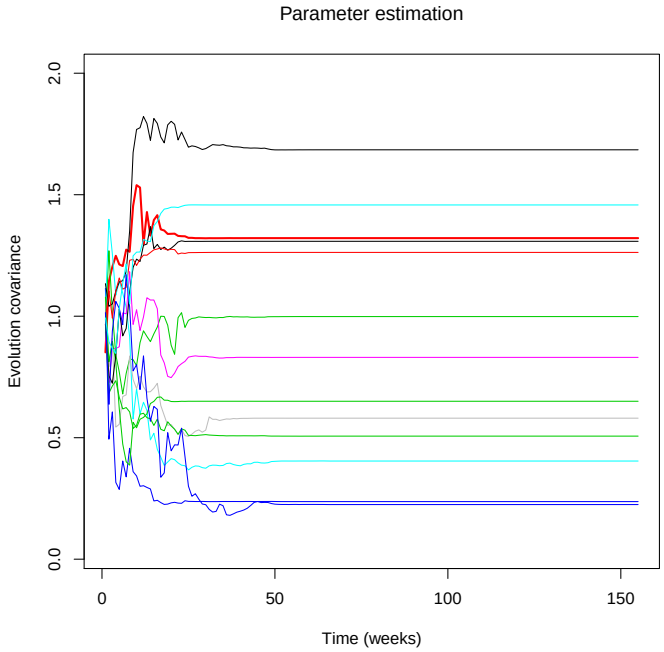

Figure 3.13: *Estimation of state variances*

To conclude this result with respect to the medical aspect, "The observed time-lag of 2–6 days implies that the fall in temperature was not a direct cause of an infant's death, i.e. the infants were not dying of hypothermia. The most obvious explanation is that the temperature decrease rendered the infants susceptible to other agents that subsequently may have caused death, and an obvious agent is a virus" (Campbell, 1994, p. 205). As can be noted from Figure 3.13, all the means of the state errors posterior distribution approximation are also converging to their fixed values at around the same time point (t=30). In addition, they are seen to have the same behaviour as the temperature coefficient.

3.9 Summary

In this chapter, background material on the recursive Bayesian methodology for online estimating of the state space models (SSMs) has been provided. Two different techniques based on the concept of the particle filter were proposed. The goal of these methods is to tackle the problem of deficiencies in generic particle filters, such as the SIS and the SIR algorithms, in the estimation of the static parameters of the SSM. The key point of the Liu and West algorithm is to track the posterior distribution of the hyper-parameters by combining the auxiliary particle filter (APF) and the kernel smoothing approximation with the concept of the shrinkage method. The principal idea of the artificial evolution of the static parameters is adapted by the Liu and West algorithm in order to make parameters time-varying. As a result, it can be added to the dynamic state vector and can be estimated simultaneously in a sequential manner. In the Storvik particle filter algorithm, low dimensional sufficient statistics are used to simulate sequentially the hyper-parameters and the state variables. In addition, a class of the Kalman filter equations is used to update the sufficient statistics vector sequentially.

The simulation study described in this chapter was designed to evaluate the performance of all algorithms. Different distributions of the observations were proposed. Regarding the first simulation experiment, the Liu and West and Storvik particle filter algorithms were used to fit the simulated Poisson data. In order to test the performance of both particle filter algorithms, the MCMC posterior approximation with a large number of iterations is used as a benchmark of the model parameters estimation. The results obtained by the Liu and West particle filter algorithm with the Poisson model outperformed the Storvik algo-

rithm in terms of estimating the unknown parameters and accuracy of forecasting. Moreover, the LW particle filter and MCMC provided a better estimation of the unknown model parameters. In the second simulation study, we investigated the performance of the Liu and West algorithm with regards to the parameter estimation and the accuracy of forecasting, when the fitted model is misspecified. Based on the results from the Monte Carlo simulation, the Liu and West algorithm with Poisson model still does better than the proposed models, even if it is specified as a wrong model.

We applied these suggested particle filter algorithms to two medical datasets. The first dataset comprised the weekly total medical contacts for school children in England who suffer from asthma. The preliminary objective of this study is a forecasting. The LW particle filter algorithm with different proposal models and the STK particle filter algorithm with a Poisson model were applied to fit asthmatic data. The results obtained by applying the proposed models showed that the Liu and West algorithm with Poisson model (LW.Pois) offered accurate predictions when compared to actual weekly total medical contacts for asthma patients. Moreover, it has the smallest value of the RMSE compared with other proposed fitted models. This result is similar to the results in the simulation study.

The second dataset comprised the weekly total of sudden infant death syndrome (SIDS) in England and Wales between 1981 and 1983. The purpose of this study was to examine any impact that environmental temperature may have on increasing the risk of SIDS. A set of covariates including the temperature and trend and seasonal components was added in the construction of the SSM. The Liu and West particle filter algorithm with Poisson model was applied to estimate the posterior mean of the temperature coefficient. The estimated posterior mean

by our model (-0.0354) is quite close to the result by Campbell (1994) that was obtained using different methods.

Chapter 4

Model diagnostic and model checking

4.1 Introduction

Diagnostic checking, using generalised linear models and linear and nonlinear time series models, is considered a significant phase in statistical model building. Its purpose is to help researchers assess the adequacy of the chosen model. In other words, model diagnostic analysis is defined as a measurement of the goodness of fit used to investigate whether the selected model adequately describes all features of the data. In linear time series methodology several diagnostic tests can be conducted to determine how the data series is modelled. If the data series is modelled well, the fitted model can then be used to predict future values of the phenomenon under study. In the statistics literature several diagnostic techniques are described that can be used to examine the quality of the chosen model in terms of fitting the data. In the context of linear time series and generalised linear

models, the sequence of residuals play a vital role in the model diagnostic for the fitted model. In terms of the linear time series analysis, if the chosen model is appropriate for fitting the data, then the residuals from this model should satisfy the assumptions of a white noise process. During the model diagnostic testing of the GLM, the fitted model can be evaluated by a special class of residuals called the Pearson residuals and deviance residuals. However, in the case of linear time series models, the model criticism is based on the residuals. There have been few studies based on a residual analysis in the model diagnostic literature on the nonlinear time series models. Therefore, a class of residuals called P-score residuals proposed by Smith (1985) is used here as diagnostic tools for adequacy of the fitted model.

The organisation of this chapter is as follows. Section 4.2 provides a methodology of the diagnostic checks of the DLMs. The model diagnostic analysis of DGLMs is offered in Section 4.3. The applications of the model diagnostics of DGLMs on both the simulated data and the medical data are offered in Section 4.4. A brief summary of the chapter is presented in the final section.

4.2 Model diagnostics for dynamic linear models

A main objective in time series analysis is forecasting of the future values of the phenomenon under study. Therefore finding an appropriate model which adequately describes the correlation structure between successive observation over time is deemed a desirable end by researchers. When the fitted model of the time series data is created, the next step is to examine whether the chosen model

satisfies the linear model assumptions. To do this, the model diagnostic can be applied. The diagnostic analysis is considered a crucial step in assessing the adequacy of the estimated DLM being used to describe the data under study. A method proposed by Box and Jenkins (1976) is to implement a model diagnostic for an autoregressive integrated moving average model, ARIMA(p,d,q) model. In the context of DLMs, several standard quantitative and graphical tools are used as in classical time series models. The raw residuals from a chosen model, which are defined as the differences between the real and predicted data, are used as a general tool for implementing the diagnostic testing. This aims to evaluate the adequacy of the estimated model for describing the correlation structure of the time series data. According to Box and Jenkins (1976), if a chosen model is correctly estimated, the sequence of residuals generated by the estimated model should satisfy the assumptions of the white noise process. A summary of some standard techniques of the model diagnostic for dynamic linear and standard time series models is given below.

With respect to the model diagnostic, residuals analysis is referred to as a general tool for checking the adequacy of the fitted model of time series data. There are two methods based on the residuals which can be adopted to investigate whether the fitted model adequately represents the data under study. They are visual inspection and nonparametric test of the residuals. A scatter plot of the generated residuals against time is used to examine whether there are any outliers, and to determine if the residuals are randomly distributed. To investigate whether the residuals series follows a normal distribution, the histogram and the normal probability plot can be employed. The Kolmogorov-Smirnov (K-S) test and the Shapiro-Wilk (S-W) test, which are known as nonparametric tests, can be used as alternative numerical methods for normality testing. In order to determine whether the residuals are uncorrelated, a simple autocorrelation func-

tion (ACF) of the residuals can be used. In terms of testing the significance of the autocorrelation of the residuals series, there are several numerical diagnostic tools which can be used instead of the visual inspection. The Ljung and Box (1978) test, which is the modified version of the Box and Pierce (1970) test, is referred to as the Portmanteau test and is the most common to be applied. Model diagnostic testing for time series analysis, which follows the generalised linear models, is implemented with respect to a class of residuals, namely the Pearson and deviance residuals. Kedem and Fokianos (2002) discuss in more detail the diagnostic analysis for the GLM based on the Pearson and deviance residuals.

4.3 Model diagnostic analysis for dynamic generalised linear models

As mentioned in Section 4.2, the raw residuals play a key role in the implementation of the model diagnostic checks in the framework of the dynamic linear and standard time series models. However, literature on model diagnostics indicates only limited use of a standard residual analysis on the nonlinear time series models. Smith (1985) suggests a measure of goodness of fit for nonlinear and non-Gaussian time series models. He proposes a class of recursive residuals for continuous time series data, known as P-score residuals (PSR), which are denoted by U_t, $t = 1, \ldots, T$. They are used to assess the adequacy of the fitted model in terms of the distributional assumptions of the conditional one-step-ahead predictive cumulative distribution. In other words, the PSR are applied to assess how well the predictive distribution is able to clarify a dependence construction in the time series data. The PSR are also known in some statistics literature as the probability integral transformation (PIT)(Dawid, 1984),(Czado et al., 2009),(Diebold

et al., 1998). With respect to the continuous distributions, the PSR are defined as random variables. They are calculated by using a continuous predictive cumulative density function (CDF) of the particular model, which is evaluated at each value of the actual time series data. The one-step-ahead predictive CDF of the continuous value y_t given the observations up to time point $t-1$ is defined in the following manner:

$$F(y_t) = Pr(Y_t \leq y_t | y_{1:t-1}), \qquad t = 1, \ldots, T \tag{4.3.1}$$

Using the form of the one-step-ahead predictive CDF in Equation (4.3.1), the PSR(U_t) and the inverse transformed P-scores residuals (INTPSR) (B_t) at each time point are defined as in Smith (1985), and Frühwirth-Schnatter (1996) in the following forms:

$$U_t = F(y_t) \tag{4.3.2}$$

$$B_t = \Phi^{-1}(U_t), \tag{4.3.3}$$

where $\Phi(.)$ is the CDF of the standard normal distribution. When the time series data follow a continuous distribution, the predictive density function is also continuous. If the fitted model is well estimated, it means it adequately explains the dependence structure of the time series data. According to the theory, this leads to the sequence of the PSR being independent and identically distributed (i.i.d) following a uniform distribution on the interval [0,1](Smith, 1985), (Frühwirth-Schnatter, 1996). Moreover, the transformed PSR B_t are i.i.d random variables following a standard normal distribution (Rosenblatt, 1952). The power of the INTPSR by Smith (1985), is that instead of having a test for a uniform distribution, we have a test for a normal distribution. Therefore, the Shapiro-Wilk test can be used. The objective of this kind of diagnostic check is to examine whether the distribution of a sequence of the PSR satisfies the assumptions of being independently and uniformly distributed on the interval

[0,1]. To find out whether this is so there are several graphical and quantitative diagnostic tools for time series data available. The empirical cumulative density function (ECDF) of the PSR can be employed in a model diagnostic. The plot of the ECDF of the PSR can be compared with the CDF of a uniform distribution. In addition, the histogram plot and the Kolmogorov-Smirnov test can be used to examine the behaviour of the PSR. To investigate whether the PSR are correlated, a simple ACF can be used as a visual check of sample independence.

In the context of a time series for count data, the predictive CDF is a discrete distribution. As a result, under the null hypothesis of ideal predictions, the distribution of a sequence of the PSR, which has been directly calculated from Equation (4.3.2), is not exactly a continuous uniform distribution. In addition, the PSR does not satisfy the statistical properties of i.i.d of the uniform distribution (Smith, 1985). However, the resulting PSR will have a discrete distribution on the interval [0,1] (Frühwirth-Schnatter, 1996). With respect to clarify this concept, the cumulative distribution function (CDF) of a continuous distribution includes all values between 0 and 1, and any value in (0,1) corresponds to a value of the random variable. Therefore, the CDF follows the uniform distribution on the interval (0,1). However, the CDF of a discrete distribution does not include all values between 0 and 1 as there is a gap between point steps. In other words we can find a value in (0,1) which does not correspond to a value of the random variable. Hence, the CDF can not be a uniform distribution. For this reason, some amendments to the main method used to the calculation of the PSR are required. To tackle this issue, Smith (1985), Frühwirth-Schnatter (1996) and several others adopt a modified method for computing the PSR from the one-step-ahead predictive density function via randomisation. Using the predictive distribution, the modification of the PSR, proposed by Smith (1985), can be obtained in the

following manner:

$$\tilde{U}_t = (1 - z_t)Pr(Y_t \leq y_t - 1|y_{1:t-1}) + z_t Pr(Y_t \leq y_t|y_{1:t-1}), \tag{4.3.4}$$

where z_t, $t = 1, \ldots, T$ is sequence of i.i.d random draw from a uniform distribution on the interval [0,1]. Furthermore, $Pr(Y_t \leq y_t|y_{1:t-1})$ is a one-step-ahead predictive cumulative distribution. As in a continuous distribution, if the fitted model is well estimated, the sequence of $\tilde{U}_t$ is i.i.d and has a uniform distribution on the interval [0,1].

In the framework of the DGLMs, the predictive CDF can be approximated as follows:

$$Pr(Y_t \leq y_t|y_{1:t-1}) = \sum_{i=1}^{N} Pr(Y_t \leq y_t|x_t^{(i)})\omega_t^{(i)} \tag{4.3.5}$$

where $x_t^{(i)}$ is a particle of estimated states and $\omega_t^{(i)}$ are associated particle weights at time t.

Typically, the diagnostic test is based on both the uniform distribution of the PSR and a transformation of a uniform to the normal distribution INTPSR. In this case, a test for a uniform and a normal distributions, such as the K-S and S-W test, can be used.

4.4 Application of diagnostic models to simulated and medical data

In this section, diagnostic checking based on the PSR and INTPSR as discussed in the previous section is applied to the simulated and medical data. The aim of the model diagnostics is to assess the accuracy of a fitted model in terms of the predictive distribution.

4.4.1 Simulation experiments

The simulation experiments aim to use the PSR and INTPSR as diagnostic tools to determine whether the fitted model is perfectly estimated. Furthermore, it aims to demonstrate the ability of the model diagnostics to detect a misspecification of the proposed fitted models. The purpose of the simulation study is to understand the behaviour of the PSR when the fitted model is correctly specified. To do this, we will generate different datasets from different proposed distributions as in the simulation study in Section 3.6.1.2. These simulated datasets will then be fitted by the correct model and misspecified models. The Liu and West algorithm is used with different proposed models. The purpose of this study is to determine the expected behavior of the PSR and INTPSR when the fitted model is correctly estimated. It is also interested in examining their behaviours when the selected model is misspecified. For the simulation studies, in this section, we use the same simulated datasets with the same true parameters and the prior distribution of the initial value of the state and hyper-parameters as described in Section 3.6.1.2. In addition, the Monte Carlo study is used to investigate the consistency of the output results.

In the first experiment, the dataset with 100 observations is simulated from a Poisson distribution. In this case, the Liu and West algorithm is applied to fit this simulated data by an LW.Pois algorithm as a correct model and by LW.NB algorithm and M.Pois algorithm as misspecified models. The summary comparison of results, based on a Monte Carlo study, is given in Table 4.1. As can be observed in Table 4.1, the Monte Carlo average of the means of the P-scores residuals (AMPSR) by the correct model and wrong models are quite similar (0.509, 0.478, 0.545). In addition, they are close to the mean of a uniform distribution

on the interval [0,1], which is equal 0.5. The small value of the standard deviation over the Monte Carlo study for all models indicates that the AMPSR is more accurate. The Monte Carlo average of the variances of the P-score residuals (AVPSR) by both the correct model and misspecified models are quite similar (0.077, 0.71, 0.062), and they are close to the theoretical variance of the uniform distribution, which is equal to 0.08. To judge the distributional assumptions for the PSR and the INTPSR, the K-S test for a uniform assumptions and the S-W for a normality test are used. The average of P-value of the K-S test (APVPSR) over Monte Carlo runs by correct and wrong models were (0.469, 0.281, 0.2106). This confirms that the PSR for all models has a uniform distribution. The figures of the average of P-value (APVINTPSR) over Monte Carlo samples obtained from the S-W test(0.557, 0.312, 0.257) confirm that the BMT for all models has a standard normal distribution. From the results reported in Table 4.1 it can be concluded that there are no significant differences in terms of applying the correctly and incorrectly specified models.

Table 4.1: Summary comparison of a Monte Carlo study of different proposed LWPF algorithms for simulated Poisson data.

Different proposed LW algorithms			
	LW.Pois	LW.NB	LW.MPois
AMPSR/SD	0.5092/0.024	0.478/0.056	0.545/0.087
AVPSR/SD	0.077/0.008	0.071/0.006	0.062/0.074
APVPSR/SD	0.469/0.304	0.281/0.211	0.2106/0.207
APVINTPSR/SD	0.557/0.225	0.312/0.271	0.257/0.232

With respect to model diagnostics, Figures 4.1 and 4.2 show a graphical comparison between the LW.poiss algorithm as a correct model and LW.NB and

LW.MPois algorithms as misspecified models in one run. Different diagnostic tools based on histograms and the ECDF are shown in the Figure 4.1. The histograms displayed in the top panels of Figure 4.1 indicate that the PSR calculated by the proposed models seem to be close to the histogram of the uniform distribution. This result can be confirmed by the ECDF and the K-S test: the p-value by the LW.Pois filter as the correct model is (0.6994) and by the misspecified models are (0.2106,0.2106). The resulting histograms, the S-W test and the ECDF plot, which are provided in the bottom panels, indicate that the INTPSR by both a correct model and by misspecified models have a standard normal distribution. In other words, under the correct and incorrect model, therefore, the resulting histograms of the INTPSR indicate acceptance of the fitted model. In addition, the resulting S-W test with p- values (0.5536, 0.1998, 0.375) support that the INTPSR have the normal distribution. In terms of the independence test, the ACF for the diagnostic tools are provided in Figure 4.2. As can be seen from the ACF of the PSR and INTPSR for the LW.Pois and LW.MPois models, all the autocorrelation coefficients lie inside the standard bounds $\pm 1.96/\sqrt{n}$. However, for the LW.NB, there are some small peaks at different lags which lie outside, and hence they can be ignored.

Finally, from this simulation study we can deduce that even if the fitted model is misspecified, the diagnostic tools based on the PSR and INTPSR calculated by using the predictive distribution have the same behaviour as the correct model.

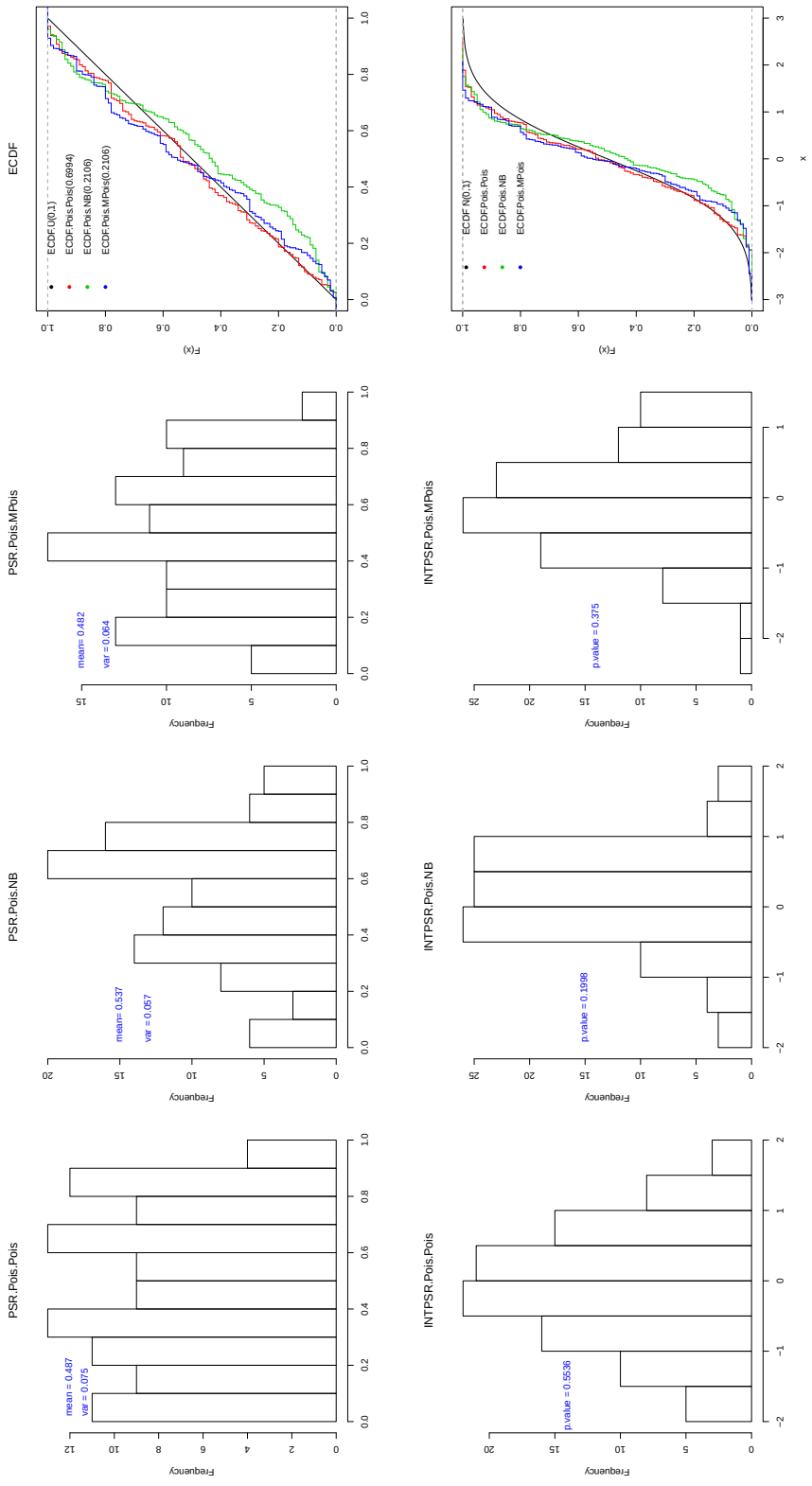

Figure 4.1: *Histograms and ECDFs of the PSR and INTPSR from different proposed algorithms for Poisson simulated data*

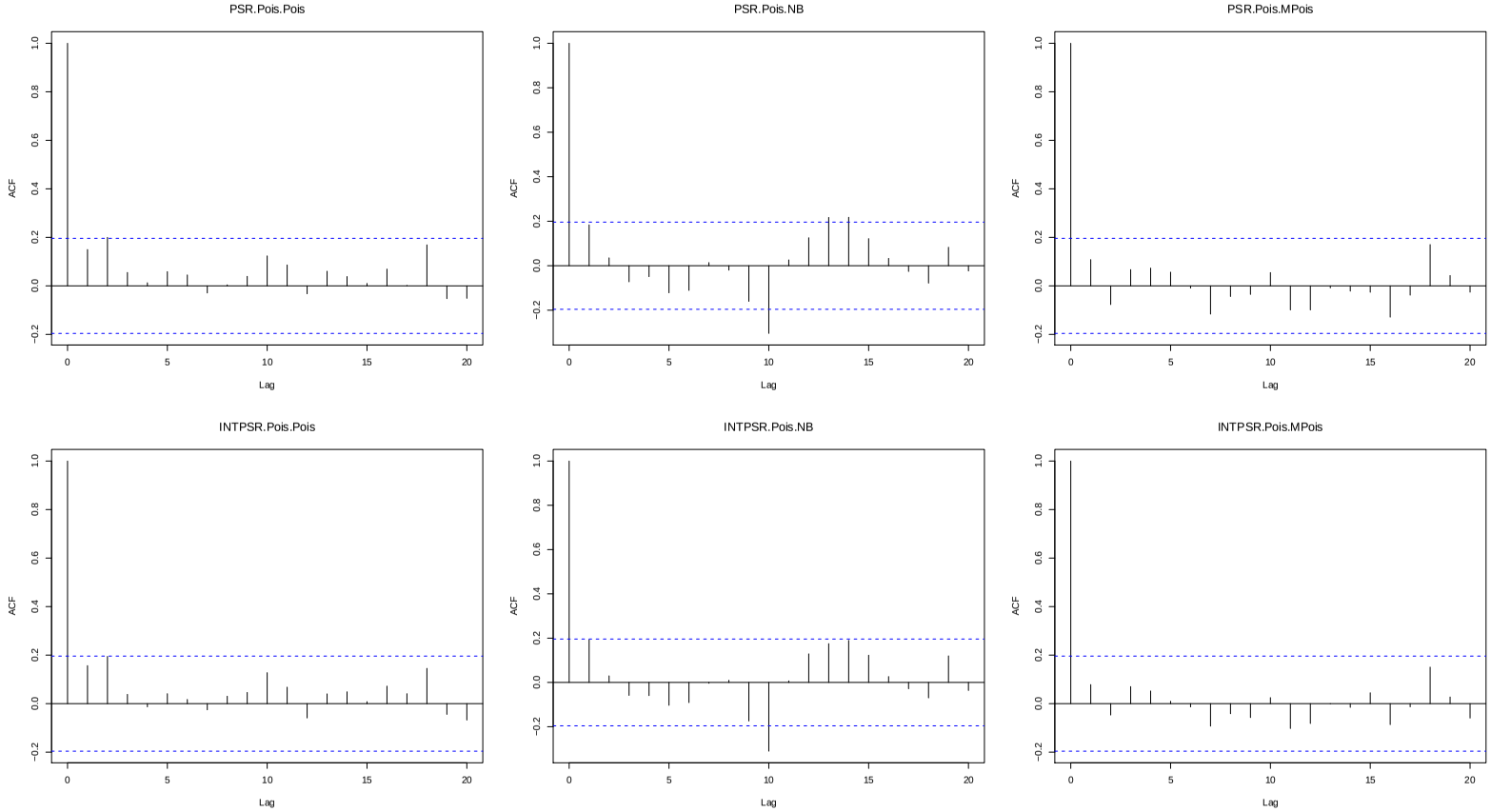

Figure 4.2: *ACFs of the PSR and INTPSR from different proposed algorithms for Poisson simulated data*

The goal of the second simulation study is to generate a dataset from a negative binomial model and then see what happens when it is fitted by using the LW.NB algorithm as a correct model and by using LW.Pois and LW.MPois algorithms as wrong models. Table 4.2 provides a summary of a Monte Carlo study for the PSR and INTPSR resulting from fitting all proposed models. As can be observed from the Table 4.2, the output results based on the PSR and INTPSR for different proposed models appear to be quite similar. The average means (AMPSR) and variances (AVPSR) of the Monte Carlo study of the PSR, which have been calculated from both correct and misspecified models, are closer as in a uniform distribution on the interval [0,1]. This result can be supported by the K-S test with p-values (APVPSR) of (0.5086, 0.467, 0.2507) by the correct model and misspecified models respectively.

Regarding the model diagnostic, based on the INTPSR, the S-W test for all models does not indicate rejection of the null hypothesis of the normal distributional assumption: the average of the p-values (APVINTPSR) of the LW.NB filter as a correct model is 0.359 and p-values of of the LW.Pois and LW.MPois as misspecified models are 0.3358 and 0.112.

Table 4.2: Summary comparison of a Monte Carlo study of different proposed LWPF algorithms for simulated Negative binomial data.

Different proposed LW algorithms			
	LW.Pois	LW.NB	LW.MPois
AMPSR/SD	0.4845/0.0259	0.515/0.0421	0.462/0.037
AVPSR/SD	0.074/0.0075	0.082/0.009	0.087/0.0112
APVPSR/SD	0.5086/0.379	0.467/0.312	0.2507/0.217
APVINTPSR/SD	0.3358/0.283	0.359/0.284	0.112/0.135

Overall, the results provided in Table 4.2 suggest that, even if data generated by the negative binomial model is analysed by misspecified models, i.e, LW.Pois or LW.MPois algorithms, the behaviour of the PSR and INTPSR is similar to the behaviour of the PSR and INTPSR when using a correct model (LW.NB).

Figures 4.3 and 4.4 show a comparison between the LW.NB algorithm and the LW.Pois and LW.Mpois algorithm in terms of various diagnostic tools in one simulation run. Histograms of the PSR and the INTPSR and the ECDF plot are provided in Figure 4.3. As can be seen from the top panels in the figure, the histograms of the PSR for all models satisfy the uniform distributional assumption. In addition, the means and variances for all models are closer to a uniform distribution on the interval [0,1]. However, it is clear from the PSR histograms that the performance of the LW.Pois algorithm is much better compared with the performance the LW.NB and LW.MPois algorithms. This result can be supported by the ECDF and the K-S test, where the p-value by the LW.Pois model is (0.713) and by the LW.NB and LW,MPois models are respectively (0.381, 0.367). The graphical diagnostics, based on the INTPSR for both a correct and misspecified models, are depicted in the bottom panels of Figure 4.3. The resulting histograms and the ECDF chart with p-values of the S-W test (0.3164, 0.517, 0.141) provide evidence that the INTPSR for all models have a standard Normal distribution. To test the dependence structure for the PSR and INTPSR, the ACF plots are shown in Figure 4.4. The top and bottom panels of the figure show that neither ACF of the PSR nor the ACF of the INTPSR, for the LW.Pois and LW.NB models display any evidence of correlation structure. This means that all the autocorrelation coefficients exist inside the standard bounds $\pm 1.96/\sqrt{n}$. However, for the LW.MPois, there is one small peak at one lag which lies outside, and hence it can be ignored. This leads to the sequences of the PSR and INTPSR for all models seeming to be uncorrelated.

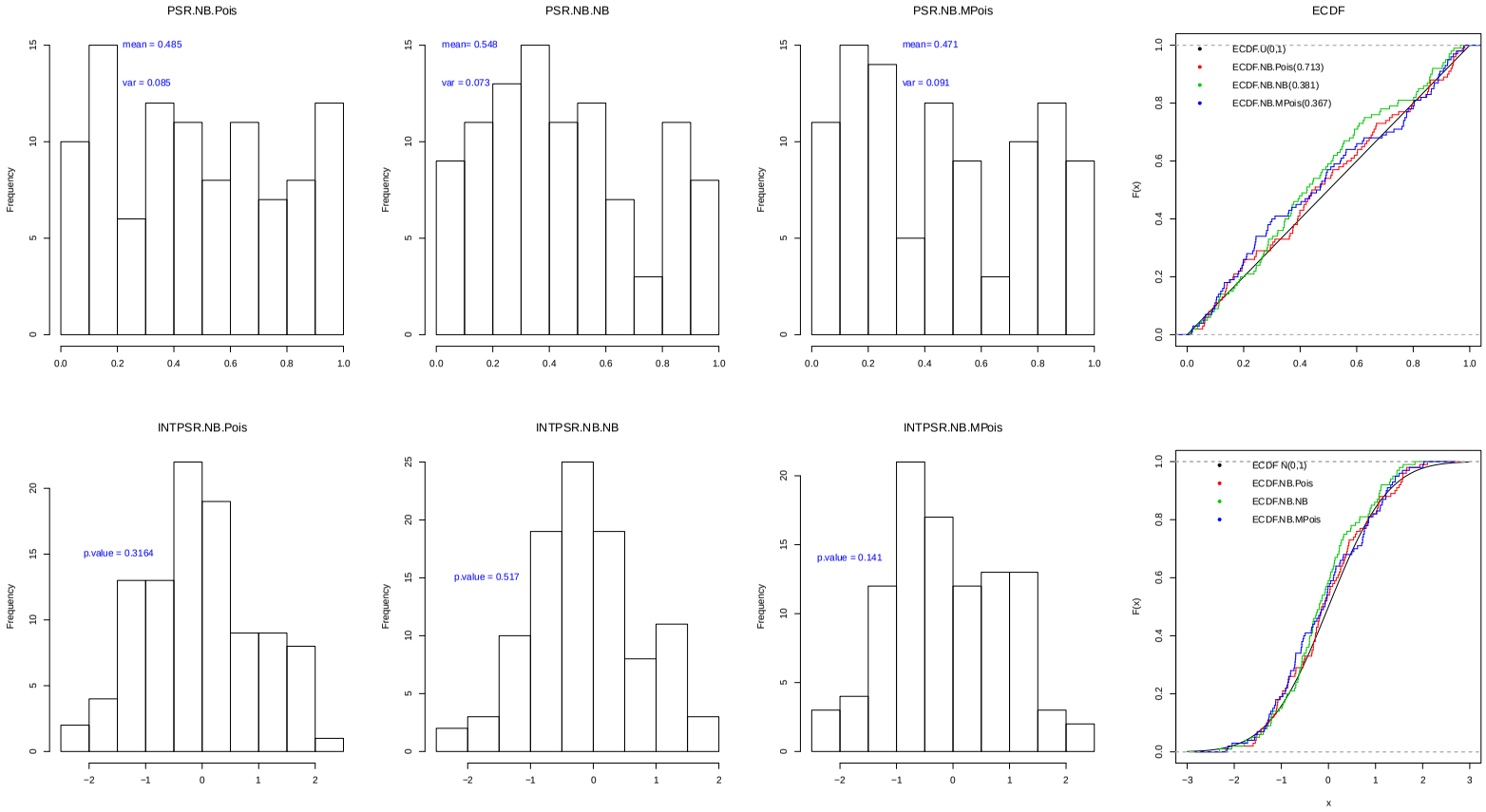

Figure 4.3: *Histograms and ECDFs of the PSR and INTPSR from different proposed algorithms for negative binomial simulated data*

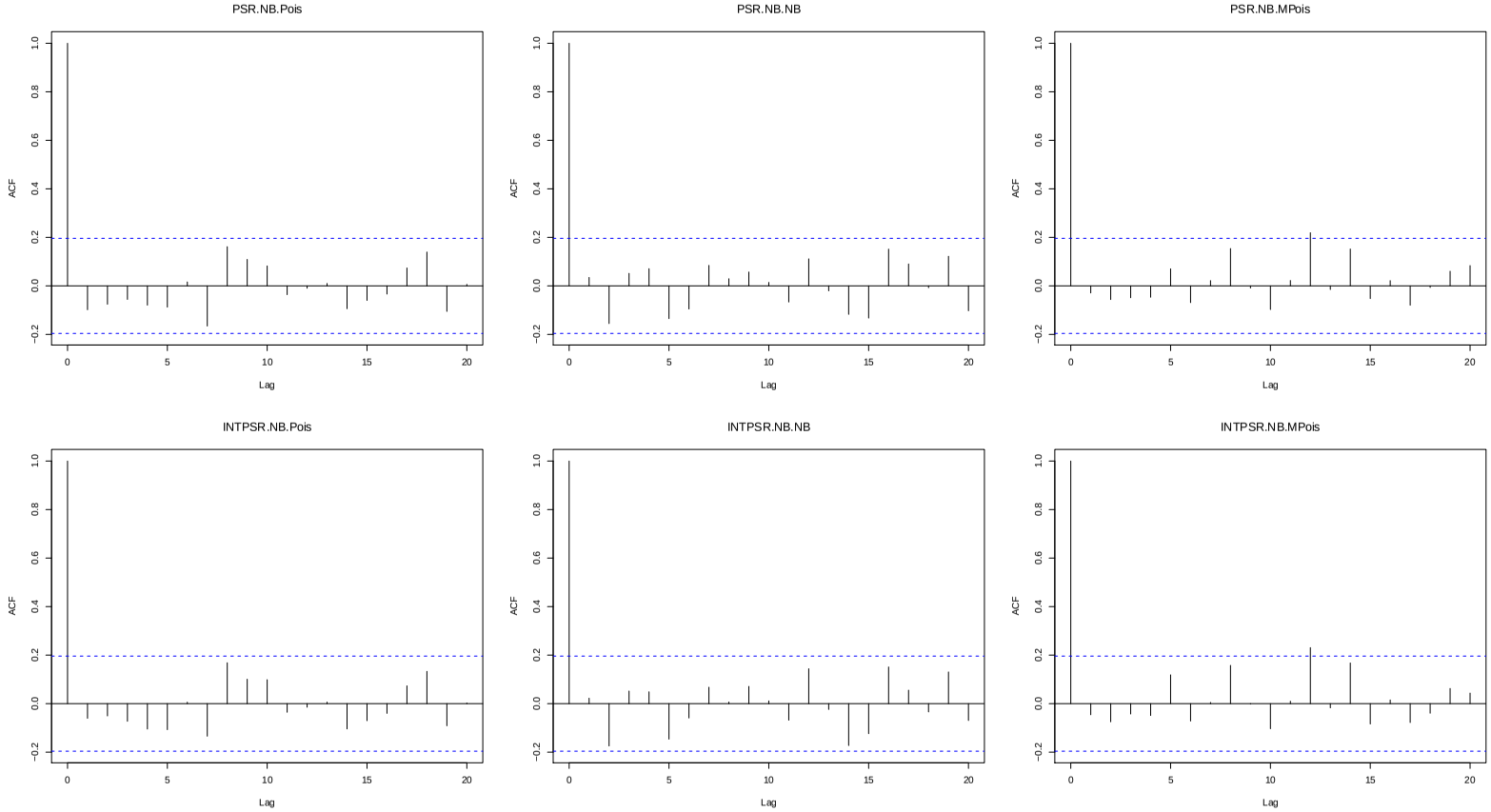

Figure 4.4: *ACFs of the PSR and INTPSR from different proposed algorithms for negative binomial simulated data*

Overall, we conclude from the results from diagnostic devices that the LW.NB algorithm (correct model) and LW.Pois and LW.Mpois algorithms (misspecified models) provide the same behavior in terms of a distributional assumption.

In the third experiment, we simulate the dataset from a mixture Poisson distribution. The different Liu and West algorithms are applied to fit this simulated data. In terms of fitting, the LW.MPois algorithm is applied as a correct model and M.Pois and LW.NB algorithms are used as models misspecified. A comparison summary for all models, based on a Monte Carlo study, is offered in Table 4.3. In addition, different diagnostic tools, based on histograms, the ECDF and the ACF of the PSR and INTPSR for one run, are shown in Figures 4.5 and 4.6. This simulation experiment provides the same conclusion obtained from the Monte Carlo study in the previous simulation experiments with respect to the diagnostic model.

Table 4.3: Summary comparison of a Monte Carlo study of different proposed LWPF algorithms for simulated mixture Poisson data.

Different proposed LW algorithms			
	LW.Pois	LW.NB	LW.MPois
AMPSR/SD	0.5358/0.0239	0.485/0.0324	0.473/0.0224
AVPSR/SD	0.076/0.0068	0.069/0.0081	0.071/0.0095
APVPSR/SD	0.3287/0.2008	0.221/0.204	0.2619/0.2011
APVINTPSR/SD	0.3687/0.284	0.211/0.197	0.281/0.225

Overall, in terms of the above simulation experiments, it is concluded that the obtained results of diagnostic checking tools by using misspecified models demonstrate similar behavior to the obtained results as in a correct model. In addition, the ACF charts of the PRS and INTPSR values do not show any ev-

idence that they are correlated. In this case the K-S and the S-W test can be used to investigate whether the PRS and INTPSR agree with the distributional assumption of the uniform and normal distribution respectively.

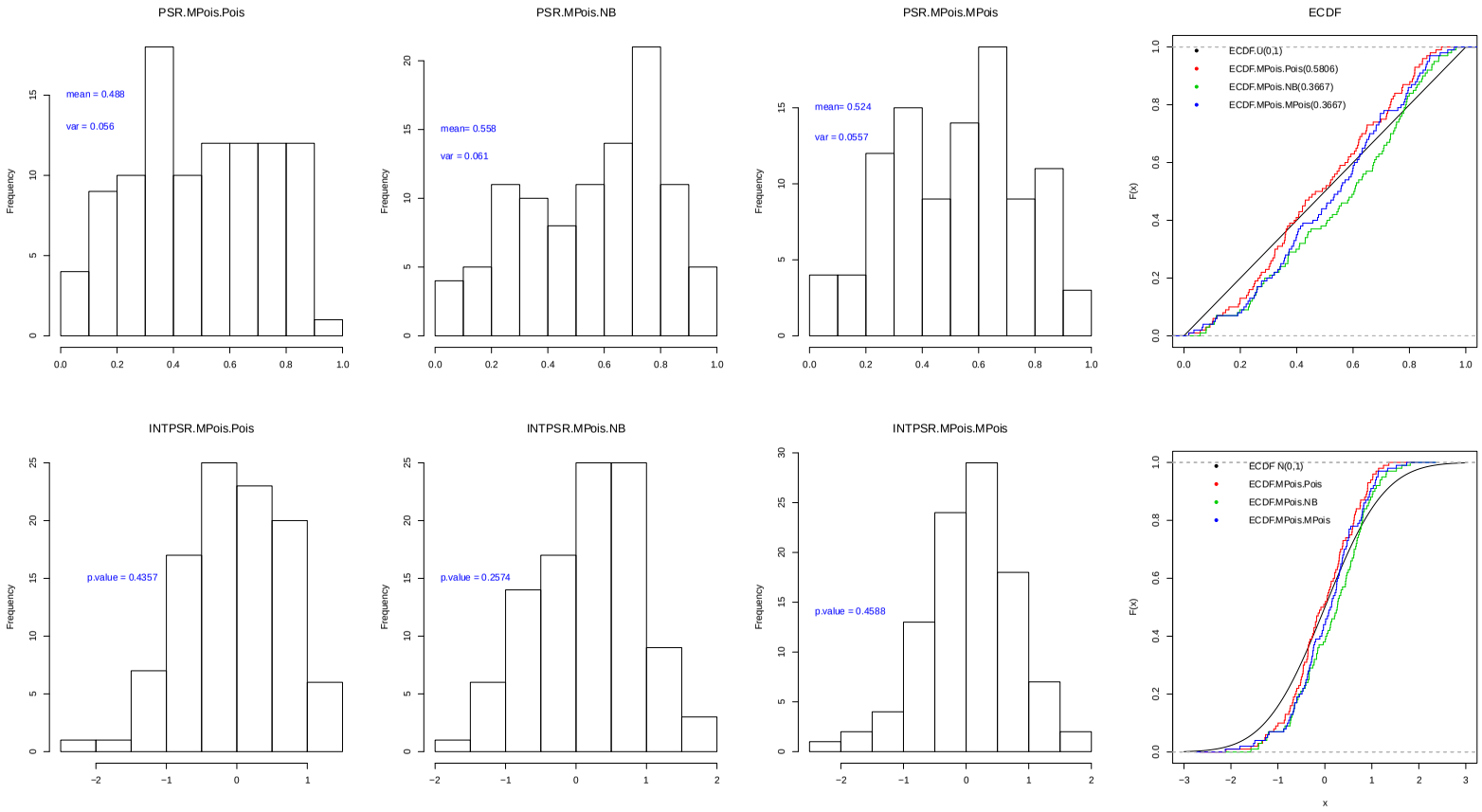

Figure 4.5: *Histograms and ECDFs of the PSR and INTPSR from different proposed algorithms for mixture Poisson simulated data*

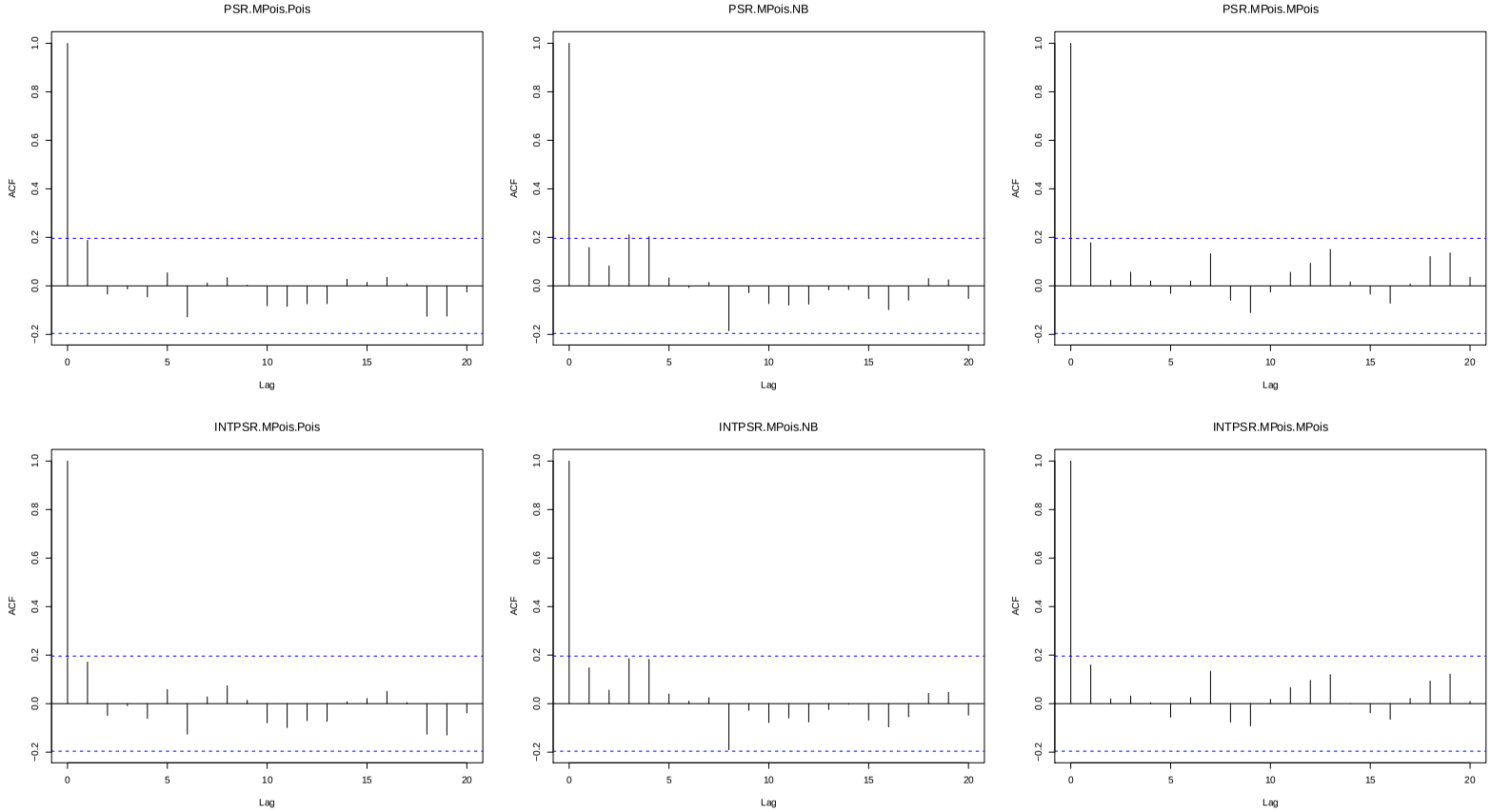

Figure 4.6: *ACFs of the PSR and INTPSR from different proposed algorithms for mixture Poisson simulated data*

4.4.2 Analysis of Asthma data

The simulation study in the previous section aimed to understand the expected behaviour of the PSR and INTPSR when the fitted model to the simulated data is correctly or incorrectly determined. In this section, we will apply the diagnostic checking technique to the asthma patients' data. The objective of the model diagnostics is to evaluate the performance of the proposed fitted model in terms of a one-step-ahead predicted distribution. According to the theory, when the model is perfectly specified, the distribution of the PSR and the INTPSR should be standard uniform and normal distribution respectively. The Liu and West particle filter algorithm with the different models as in Section 3.7 is used. The proposed fitted models are LW.Pois, LW.NB and LW.MPois. In this case, a one-step-ahead cumulative predictive distribution, which is evaluated at the actual date and the estimated parameters, is used to compute the PSR values.

The results of the model diagnostics for all proposed models of the last 105 observations of asthma data are shown in Figures 4.7–4.9. Figure 4.7 shows the histogram, ACF, and the ECDF chart of the PSR and INTPSR for the LW.Poiss particle filter algorithm. As can be observed from plots of the PSR in the top row panels, the K-S test with a p.value of 0.0002 indicates a mis-specification of the distributional assumption of a uniform distribution. However, the mean and variance of the PSR is close to the mean of a uniform distribution, (0.404, 0.056). In addition, the ACF does not show any evidence of correlation structure: all of the autocorrelation coefficients lie inside the standard bounds $\pm 1.96/\sqrt{n}$. This means that the sequence of the PSR is independent.

To help judge the distributional assumption of the INTPSR, the bottom row panels in Figure 4.7 show the results from running diagnostic tools on the INTPSR

by a LW.Poiss particle filter algorithm for the asthma data. The histogram and the ECDF indicate that there is a departure from a standard normal distribution. This is confirmed by the S-W test, which has a p-value of $(1.67e^{-05})$. However, the ACF of the INTPSR shows that they seem to be uncorrelated.

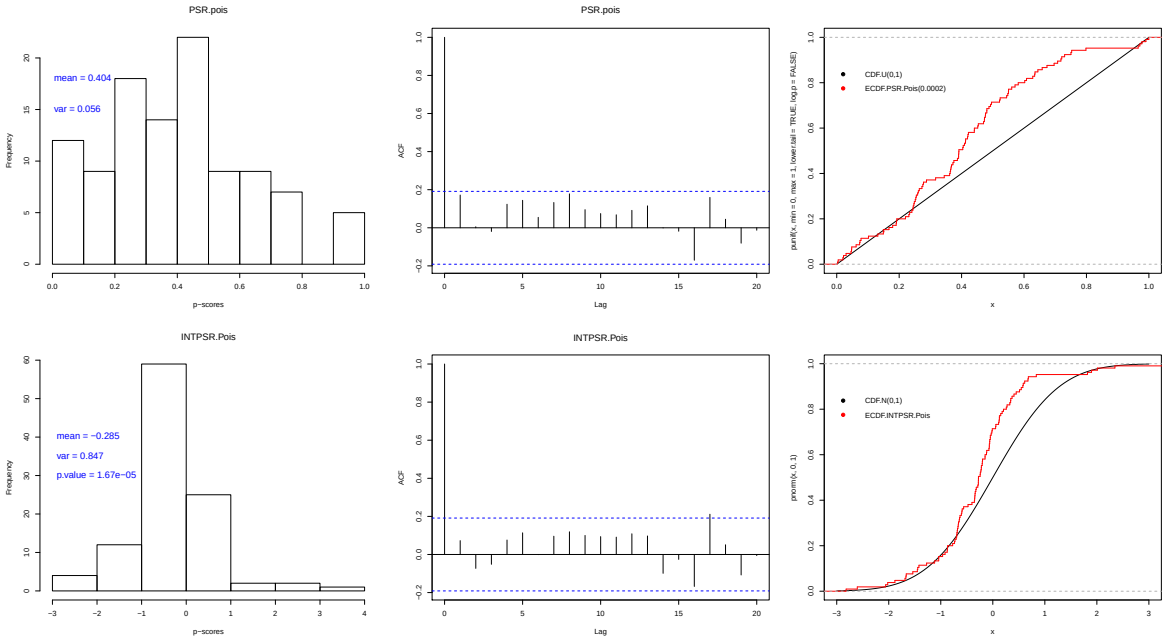

Figure 4.7: *Histograms, ACFs and ECDFs of the PSR and INTPSR from LW.Pois algorithm for asthma data*

We deduce from this diagnostic analysis that the LW.Poiss algorithm with a Poisson model provides a poor performance in terms of applying different diagnostic tools. However, according to the RMSE value (1.082), of one-step-ahead forecast, as reported in Section 3.7, the LW.Pois algorithm offers a good performance regarding the prediction of future values of an asthma patients data. Additionally, the estimation of the mean and variance of the PSR for the fitted model was quite close to the mean of the uniform distribution on the interval [0,1].

The obtained results of diagnostic analysis using LW.NB and LW.MPois particle filter algorithms on the last 105 observations of asthma data are shown in Figures 4.8 and 4.9. Similar to the LW.Pois particle filter algorithm, we observe from the graphs that the resulting histograms of different diagnostic tools of the PSR and INTPSR calculated by both particle filter algorithms have significant deviations from the shape of a uniform and normal distribution. In addition, the ECDF and the S-W test with p-values $(9.135\, e^{-11}, 0.00066)$ provide evidence indicating departure from the distributional assumptions of the diagnostic tools. However, the mean of the PSR for both particle filter algorithms (0.6438, 0.4558) is close to 0.5. But, the variance of the PSR of the LW.NB particle filter algorithm is very small (0.003) compared to the variance of the uniform distribution (0.08).

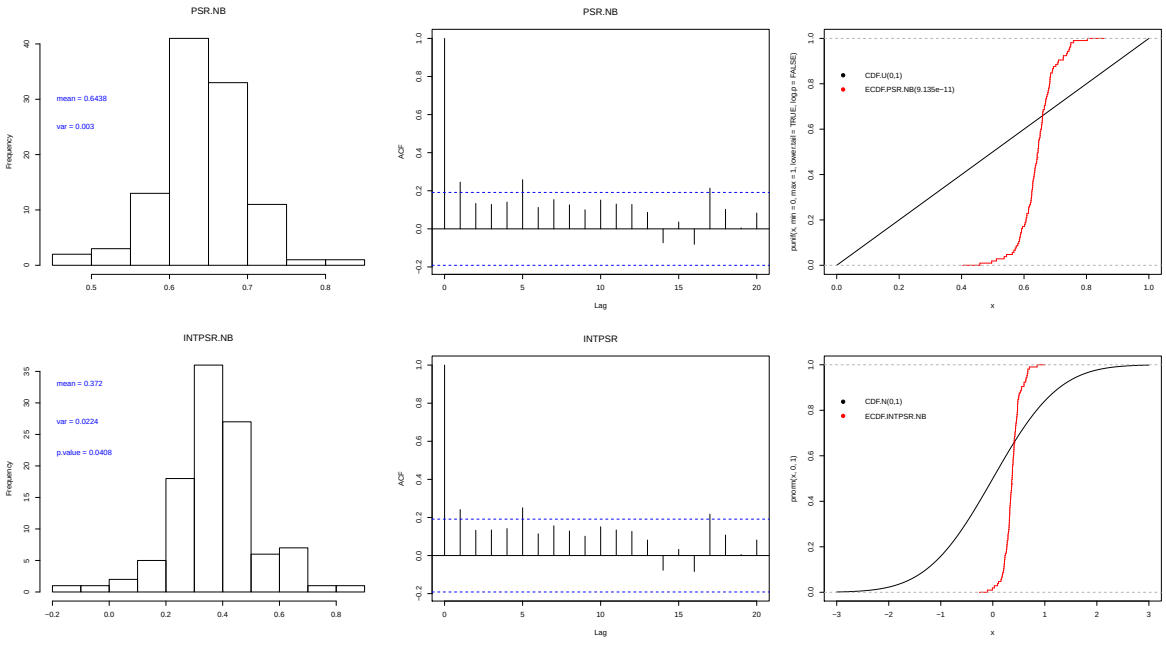

Figure 4.8: *Histograms, ACFs and ECDFs of the PSR and INTPSR from LW.NB algorithm for asthma data*

Regarding the ACF, it can be noted that there is no evidence of the correlation in sequence of the PSR or the INTPSR for the LW.MPois particle filter algorithm. But, the ACF for the LW.NB particle filter algorithm shows some correlations at different lags.

Overall, based on the output results of the diagnostic checks by the PSR and INTPS, the overall performance of all the proposed models is not good. With respect to this result, there is a possibility that the proposed model did not fit asthma data well. As can be seen from simulation studies, however, the diagnostic tools still work even if the model is misspecified. In addition, the means of the PSR for all proposed models of asthma data (0.404,0.6438,0.456) are quite close to 0.5. This accurate estimation of the means of the PSR is reflected by the values of the RMSE of one-step-ahead forecasting for all proposed fitted models, as provided in Table 3.7, which were satisfactory (1.082,3.340,2.226).

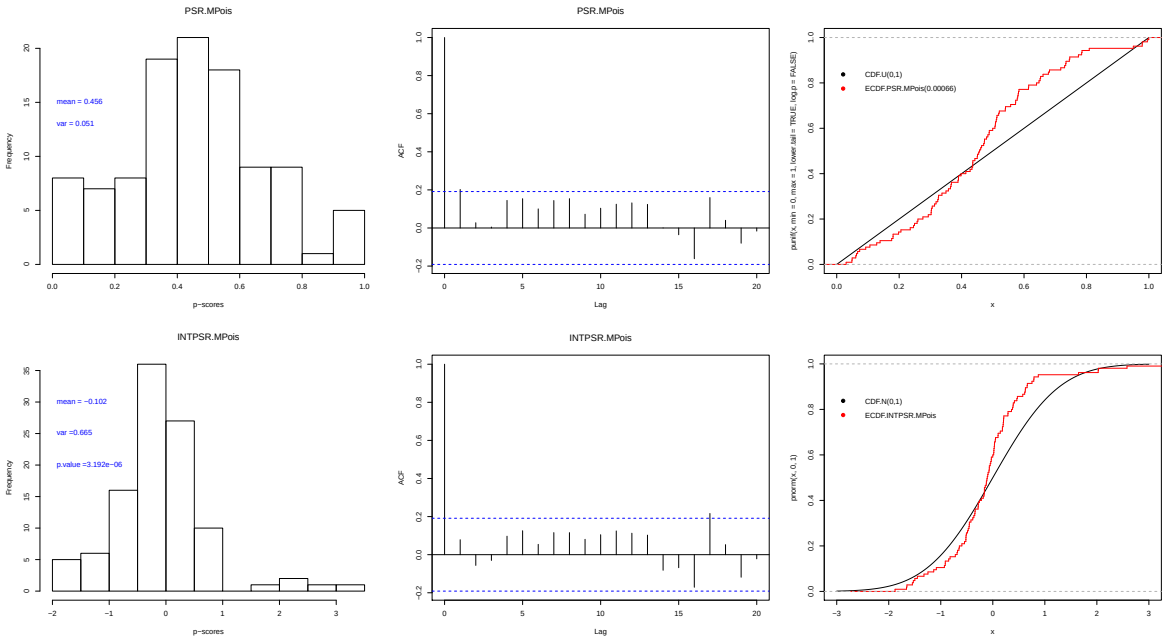

Figure 4.9: *Histograms, ACFs and ECDFs of the PSR and INTPSR from LW.MPois algorithm for asthma data*

With respect to the results of the RMSE, we can speculate that even if the distribution of the data is mis-specified, the forecast will be adapted. This is because this misspecified distribution is used to simulate the predicted observations based on the estimated parameters. Therefore, from time to time, even if the distribution is wrong, the proposed fitted model will correct itself. Figure 3.6 shows how well the predicted values by the proposed fitted models with LW particle filter algorithms are close to the actual values of asthma data.

To understand why the PSR and INTPSR provide poor results of a model diagnostic for all proposed models for the asthma data, we provided a clarification below with some further justifications. Assuming we have iid data from a classical generalized linear model (GLM) with a Poisson distribution with parameter λ as following: $y_i \sim Poiss(\lambda), i = 1, \ldots, n$ with a linear predictor, $\eta = log(\lambda) = X\beta$. Because the data are iid, the parameter λ is believed to be the same for all data. Therefore, we can look at the histogram for whole data in order to see whether this histogram is similar as in the Poisson distribution. Suppose now, we have some time series data from a dynamic generalized linear model (DGLM) with a Poisson distribution as follows: $y_t \sim Poiss(\lambda_t), t = 1, \ldots, T$, with a linear predictor, $\eta_t = log(\lambda_t) = X_t\beta_t$. Because the parameter λ_t here is time varying, at each time point we have the same form of the Poisson distribution but with different value of the parameter λ_t. This means that, at each time point there is possibly a different shape of the Poisson distribution. From this point on, if we plot the histogram for the whole of data as in the GLM that could be invalid.

Turning now to asthma data, there might be some parts of the time series that the proposed model fits well and there might be other parts the model does not fits well. Therefore, we need some kind of time varying application for monitoring of the uniform distribution on the interval (0,1). One possibility is to use control

charts based on the normal distribution of the residuals.

With respect to asthma data we divided the time series into 36 groups. Figure 4.10 provides histograms of 36 groups with their means and variances. As can be seen from this graph, the shape of the distribution of the groups changed over time: some are negatively skewed, some are positively skewed, some are symmetrical, and some are approximately uniformly distributed. There is not a specific shape of the distribution that can be adopted for all groups.

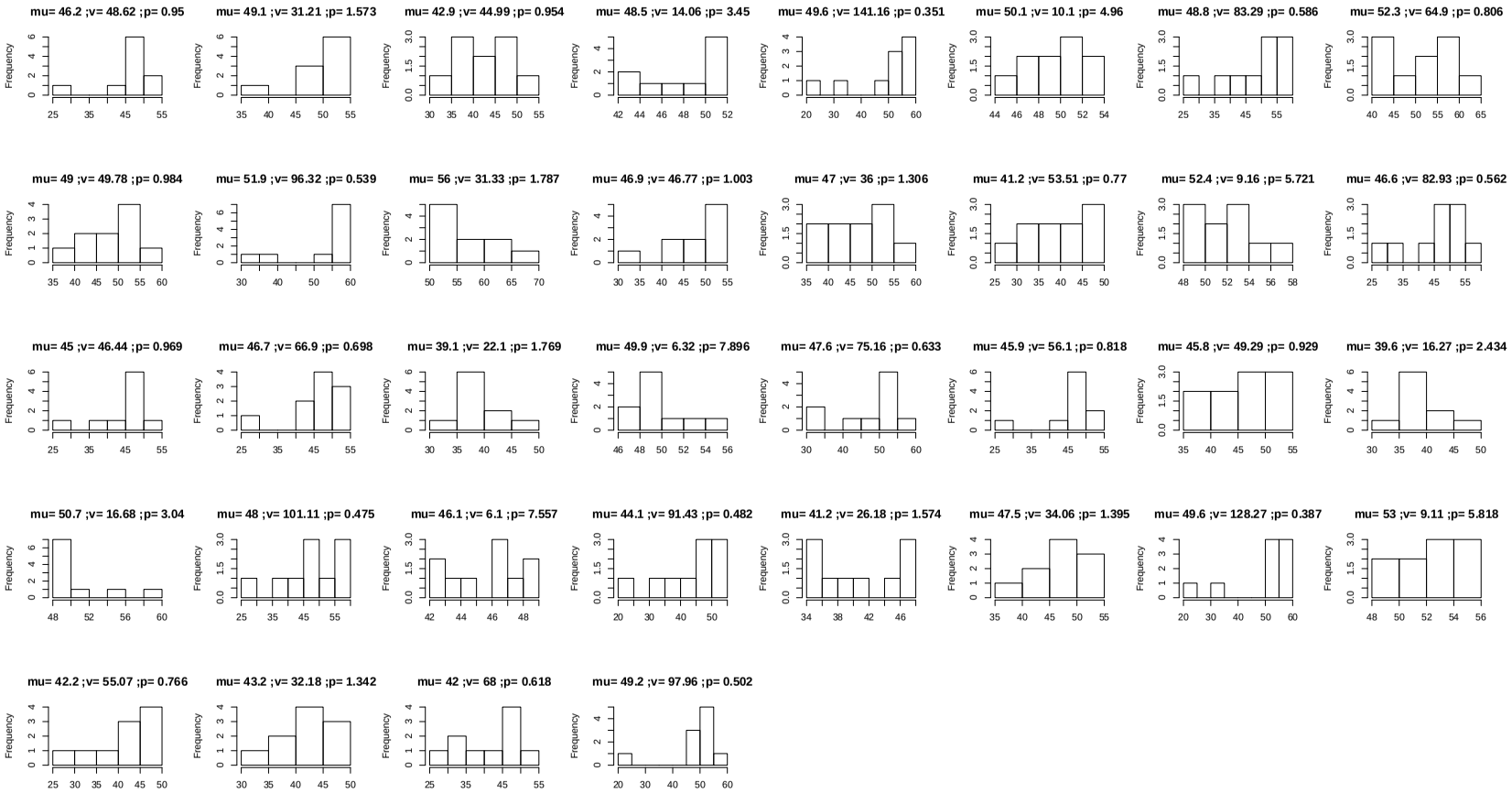

Figure 4.10: *Output graphical results of dividing asthma data into several groups*

4.4.3 Analysis of SIDS data

In this section, the model diagnostics with different tools are applied to the SIDS data. In Section 3.8, this dataset was represented in the form of a DGLM. A set of covariates (such as the environmental temperature, trend component, and seasonal component) were included in this model. The reduced model with five harmonic components was used to describe the seasonality in the SIDS data. In addition, the LW.Pois particle filter algorithm was applied to estimate the joint posterior distribution of the state and hyper-parameters. The aim of this study focused on identifying the impact of environmental temperature on the sudden deaths of infants. Figure 4.11 shows three rows of graphs associated with different diagnostic tools, based on the PSR and INTPSR, for the LW.Pois particle filter algorithm applied to the SIDS data. As can be seen from the graphs, the obtained results of all diagnostic tools when applied to the SIDS data have the same behaviour as with the asthma data. In other words, we observe from Figure 4.11 that the resulting histograms of different diagnostic tools of the PSR and INTPSR calculating by the one-step-ahead predictive distribution of the LW.Pois particle filter algorithms have remarkable departures from the shape of a uniform and normal distribution. This issue occurred because the variance of the PSR is very small (0.0005), which reflects negatively on the shape of the PSR histogram. Thus, it departs from the histogram shape of a uniform distribution on the interval [0,1]. In addition, the ECDF and the S-W test with p-values (0.0288) for the INTPSR provide evidence indicating departure from the distributional assumptions of the diagnostic tools. However, the mean of the PSR is close to the mean of a uniform distribution, (0.512). In terms of the correlation, the ACF of the sequence of the PSR or the INTPSR shows that they seem to be uncorrelated. As a result, the overall explanation concerning the diagnostic

tools for data asthma as discussed in the previous section can be generalised to the SIDS data.

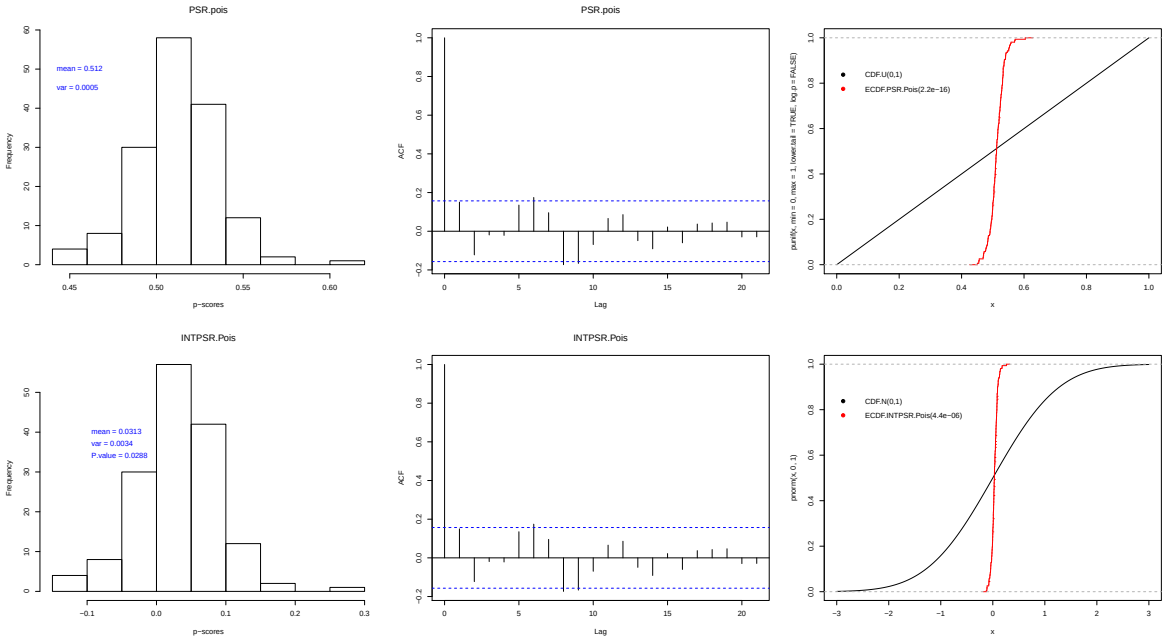

Figure 4.11: *Histograms, ACFs and ECDFs of the PSR and INTPSR from LW.Pois algorithm for SIDS data*

4.5 Summary

In this chapter, background material on model diagnostics with different tools for classical time series models and the DLMs was described. In addition, a diagnostic test based on the PSR and INTPSR as proposed by Smith (1985) for the non-Gaussian time series models was presented. The aim of using the diagnostic tests was to assess the adequacy of the proposed fitted model with respect to the predictive distribution.

The simulation experiments described in this chapter were designed to apply a

diagnostic analysis to determine the expected behaviour of the PSR and INTPSR when the proposed model is correctly or incorrectly specified. Based on Monte Carlo studies, it was concluded that the obtained results of the diagnostic checking tools, by using mis-specified models, exhibit similar behaviours to the obtained results as in a correct model.

The suggested procedure of the model diagnostics was applied to the asthma and sudden infant death syndrome data. The obtained results of the model diagnostic indicated that all the proposed models for both medical datasets provided a poor performance regarding the shape of the PSR and INTPSR histograms. In addition, The deviation of a distributional assumption of the PSR and the INTPSR were confirmed with the ECDF chart and the K-S test with a p-value. These results, however, were based on the shape of the histograms of the whole of the PSR and INTPSR over time. In terms of the DGLM, because the model parameters are time varying, at each time point there is possibly a different shape of the desired distribution. From this point on, if we plot the histogram for the whole of data as in the GLM that could be invalid. In this case, it is therefore necessary to investigate the distributional assumption of the PSR and INTPSR sequentially at each time point. Therefore, we need some kind of time varying application for monitoring of the uniform distribution on the interval (0,1). The control charts based on the normal distribution of the residuals are one possibility to use.

Chapter 5

Monitoring and Adaptation

5.1 Introduction and Motivation

In Chapter 4, model diagnostics were used to assess the performance of the proposed fitted model. The diagnostic analysis is a crucial step towards deciding whether the fitted model describes the dependence structure of the time series for count data successfully. The principle idea of the model diagnostics, in Chapter 4, is based on the one-step-ahead cumulative predictive distribution. The objective of using the model diagnostics is to evaluate the accuracy of the forecast calculated by using the fitted model. The PSR and INTPSR, proposed by Smith (1985), are used as diagnostic tools. If the fitted model adequately explains a dependence structure of the time series data, then the PSR, being i.i.d. random variables, follows a uniform distribution. In addition, the INTPSR is i.i.d. random variables having a standard normal distribution with mean zero and variance one.

This diagnostic approach is an off-line method. However, there is a problem

in using this model diagnostic when data are collected sequentially over time. In this case, in order to do model diagnostics an estimation of unknown parameters, one-step-ahead forecasting, and diagnostic tools, are required to be calculated at each time point when a new observation is available. Therefore, this diagnostic approach is not effective as time goes on. The approach can go further if, at a given time point, the fitted model is good and, at another time point, the fitted model is not. In this case it is necessary to have an effective way of checking continuously whether the fitted model is good or not. Bayesian monitoring methods based on the Bayes factors (BFs) and statistical process control (SPC) can be used to carry out sequential monitoring. These methods are considered as online approaches for model diagnostics.

In this chapter a background and literature review for the related fields of the Bayesian monitoring, including a statistical process control, control charts and BFs are provided. The applications of automatic diagnostic methods on both the simulated data and the medical data are offered in this chapter. A brief summary of the chapter is presented in the final section.

5.2 Bayesian monitoring model

The monitoring technique is considered as a crucial step in the applications of time series analysis. It is used as a diagnostic tool to evaluate the performance of the predictive ability of the proposed fitted model to the data. The model monitoring is concerned with detecting and assessing discrepancies between actual data and their predictions calculated from the proposed fitted time series model. The occurrence of outliers and structural change in the series are major sources of the

deterioration of such time series models and may prevent their forecasting (West, 1986). In Bayesian methodology for Gaussian time series models, West (1986) proposed a general monitoring approach to assess an adequacy of the model in a sequential manner. The recursive monitoring scheme is based on the Bayes factors (BF), where the calculation of the BF is based on the standardized one-step ahead forecasting error. The details of the methodology of the BF will be provided later. (West and Harrison, 1997, Chapter 11) developed the monitoring algorithm by using the cumulative BF. The cumulative BF can be calculated by replacing the predictive density with a joint distribution of the forecasting error in the standard formulation of the BF. Gargallo and Salvador (2003) introduced the sequential monitoring approach in order to detect deteriorations in the performance of the one-step ahead forecast of a DLM due to the presence of residual autocorrelations. Relevant literature on Bayesian monitoring for Gaussian time series models can be found in West and Harrison (1997)

5.3 Statistical process control

Statistical process control (SPC) is a common application of statistics in manufacturing and industry. It is a widely used in statistical practice for process monitoring and change detection in the production processes. The objective of SPC is to detect the presence of defective products and to make sure that the final product is of a high level of quality. In the framework of SPC, control charts are considered essential for monitoring. They are defined as a graphical representation used to monitor the behaviour of the processes under investigation in a sequential manner. They can also be used to follow up the progress of the process over time to identify any aberrant change in the process. Therefore, an early in-

tervention can be made to correct the process and help us to determine the cause of the change. The first control chart was proposed by Shewhart (1931). He suggested an X-bar chart as a common tool in an application of SPC. The goal of using this chart is to monitor an averages process. Page (1954) and Roberts (1959) introduced two effective alternative control charts to detect a small shift instead of the Shewhart charts. They are called the cumulative sum (CUSUM) control chart and the exponentially weighted moving average (EWMA) control chart respectively. The advantage of using these control charts instead of the Shewhart chart is that the decision about the process at each time is based on all information in the sequence of the observations.

A violation of these assumptions (i.e. that the observations of the process under consideration should be independent and normally distributed) may lead to the wrong calculation of the control chart limits and thus the adoption of a wrong decision about the process. Satisfaction of the assumptions of the independence and normal distribution of the data is necessary for the monitoring process to make a correct decision and avoid a false alarm caused by a wrong decision (Alwan and Roberts, 1988; Zhang, 1997; Stoumbos and Reynolds, 2000; Bisgaard and Kulahci, 2005; Castagliola and Tsung, 2005). In the case of autocorrelated data, Alwan and Roberts (1988),Montgomery and Mastranglo (1991) and Singh and Prajapati (2011) reported that the presence of the autocorrelation within the process data will affect the performance of the standard control charts. However, these traditional control charts can be used as monitoring tools for detecting the process change with autocorrelated data by making some modifications to them. In the applications of time series analysis, Alwan and Roberts (1988) proposed a control chart for the autocorrelated data. The design of this control chart is based on the residuals, which are calculated from an adequate fitted time series model, instead of using the original autocorrelated data.

To evaluate the performance of the standard control charts, such as the CUSUM and EWMA charts, the average run length, denoted by ARL, can be used. The in-control ARL is defined as the expected number of the plotted observations within the lower and upper limits of the control chart until a false alarm occurs. The quality of the control charts is measured by how quickly the change of the process is detected. When the process is in the in-control state, where all plotted signals lie between the control limits, the value of the ARL will be very large. The meaning of the ARL is to be infinity is that, in the in-control state, the alternative density is exactly the same as the null density. Therefore, the value of the BF is always equal to 1. Thus, there is no chance to issue an out-of-control signal. This is of theoretical interest only, whereas in practice the alternative density will always be different from the null.

5.4 Bayes factor

The Bayes factor (BF) is a mathematical tool that can be used within Bayesian model monitoring to assess the performance of the fitted model. In addition, it can also be used to compare the performance of two models, based on their predictive performance, in a recursive manner. In Bayesian framework, the application of the BF can play a similar role as that in hypothesis testing in frequentist statistics. Within a Bayesian framework, Jeffreys (1935,1961) developed a methodology of hypothesis testing, based on the BF, to quantify the evidence, by the observed data, in favour of two competing scientific theories.

Suppose we have two competing statistical models, M_1 and M_2, that have been used to fit the same observed data $y_{1:t}$ under study. These two fitted models have the same mathematical structure but they have different values of model

parameters. After estimating the unknown parameters in each fitted model, the predictive density for each model can be calculated by integrating out the parameters θ_i as follows:

$$p(y_{1:t}|M_i) = \int p(y_{1:t}|M_i, \theta_i)p(\theta_i|M_i)d\theta_i, \qquad i = 1, 2, \tag{5.4.1}$$

where $p(y_{1:t}|M_i, \theta_i)$ represents the marginal likelihood function of a vector of parameters θ_i under the model M_i and $p(\theta_i|M_i)$ is the prior distribution of the parameters vector θ_i. Using Bayes theorem and given observations up to time t, $y_{1:t}$, the formulation of the BF for M_1 versus M_2 is given by:

$$BF_t = \frac{p(y_{1:t}|M_1)}{p(y_{1:t}|M_2)}, \qquad t = 1, \ldots, T \tag{5.4.2}$$

Note that the BF in Equation (5.4.2) is defined as the ratio of the integrated likelihood functions for two models. In the context of a Bayesian approach, Jeffreys (1961) proposed a guidelines for interpreting the value of the BF to evaluate the strength of evidence for one model relative to another. This interpretation of the BF is presented in Table 5.1. Kass and Raftery (1995) suggested some modification on a scale proposed by Jeffreys (1961) for interpreting the value the BF.

BF_t	Evidence against M_2
1 - 3.2	Not worth more than a bare mention
3.2 - 10	Substantial
10 - 100	Strong
> 100	Very strong

Table 5.1: Evidence categories for the interpretation of the Bayes factor proposed by Jeffreys (1961)

In this thesis, alternative models (M_A) of the INTPSR are used. These alter-

native models can be expressed in terms of distributions of the INTPSR deviating from N(0,1). This is found to be more convenient than specifying alternatives in terms of distributions for the PSR deviating from U(0,1). The alternative distribution on the INTPSR scale could, of course, be transformed to a distribution on the PSR scale and this would result in the same Bayes factor. From this definition, given the observation up to time $t-1$, the BF for comparing M_0 against M_A at time point t is given by:

$$BF_t = \frac{p(INTPSR_t|y_{1:t-1}, M_A)}{p(INTPSR_t|y_{1:t-1}, M_0)}, \qquad t = 1, \ldots, T \tag{5.4.3}$$

where $p(INTPSR_t|y_{1:t-1}, M_A)$ is the one-step-ahead predictive density based on the alternative model and $p(INTPSR_t|y_{1:t-1}, M_0)$ is the one-step-ahead predictive density by the fitted model. In the Equation (5.4.3), the formulation of the Bayes factor is designed to compare between the null model and alternative models with some deviations while in the Equation (5.4.2) it compares between two different models. A Bayes factor greater than 1 is evidence in favour of M_A over M_0.

5.5 Non-parametric control chart

As mentioned above, in the context of the SPC the standard control charts are used in a recursive way to monitor the behaviour of the process. The most common assumption in the literature of the SPC is that the distribution of the process under consideration is normal. In this case, when the normality assumption is satisfied, the traditional control charts provide effective performance monitoring. However, in many applications, the process under investigation does not follow a normal distribution. As a result, the statistical properties of standard control charts, such as control limits, false alarm rate, and the in-control average

run length can be highly affected. To remedy the shortcomings of the statistical properties of charts, nonparametric or distribution-free control charts can be used for achieving this purpose. The idea behind the nonparametric control chart is that there is no particular assumption on the distribution for the underlying process, such as the normal distribution. In addition, the in-control probability calculations and associated conclusions remain valid for any continuous distribution (Chakraborti et al., 2004). An extensive overview of the literature on nonparametric control charts has been provided by Chakraborti et al. (2001). They listed several practical reasons for using distribution-free control charts. For more detail on the nonparametric control charts literature Amin et al. (1995), Bakir (2012), Chakraborti and Graham (2007), and Chakraborti et al. (2011) are useful resources.

In this thesis, the proposed nonparametric control chart is based on the classification of the value of the BFs. These values are used, in Phase II monitoring, for constructing the in-control situation or out-of-control state of the process. The reason for using the nonparametric chart here is that we do not know the distribution of the BFs. A binomial control chart using the threshold rules of the BFs with the runs-rules is adopted as a proposed procedure to monitor the process under consideration at each time point.

The binomial control chart, also known as a P-chart, is considered a valuable tool used to detect the proportion of defective units in a sample. The proportion of the nonconforming units is defined as the ratio of the number of defective units to the sample size. Since the construction of the binomial chart is based on the binomial distribution, the assumptions of the binomial distribution are required.

Nonparametric tests are tests that do not assume a specific distribution of the data, so they are called distribution-free tests. In other words, nonpara-

metric tests can be used when there are no assumptions about the population distribution. This means that, if the assumption of normality of the data under study is available, then parametric tests such as T and F tests can be implemented, but if this condition is not available, we should resort to nonparametric tests. In the literature the most frequently used nonparametric tests are given as the Chi-square test, run test, Wilcoxon signed rank test, Mann-Whitney test, and the Kruskal-Wallis test.

Using the interpretation rules of the BF by Jeffreys (1935,1961) in Table 5.1, the proposed control chart is constructed into three categories in terms of the classification of the value of the BF. The statistical formula of the BF is based on the INTPSR as in Equation (5.4.3), where the process described by the null model M_0 is considered to be in control state, whereas the alternative model M_A describes the process out of control. The proposed control with three classification categories is provided in Table 5.2.

Table 5.2: The proposed procedure of the control process based on the Bayes factor

Category	BF_t	Interpretation
0	0 - 3.2	In control state
1	3.2 - 10	Substantial evidence that the process is out of control state
2	> 10	Strong evidence that the process is out of control state

As can be seen from this table, when the BF at a specific time point is less than 3.2, the process lies in category 0. Therefore, the process is considered as being in an in-control state. In the case of the out-of-control state, in this thesis, the following composite rules are proposed: the process is considered as being in an out-of-control state when the value of the BF leads to category 2, or two

consecutive values of the BF in a window of length k = 4 are assigned to category 1. The window length is defined as the control procedure parameter, which can be specified by the modeller in terms of a desired sensitivity set-up in the procedure. A small sensitivity of the procedure can be obtained by using a smaller value of the window length, while larger values of the length of the window lead to an increase in the sensitivity of the control procedure (Triantafyllopoulos and Bersimis, 2016).

5.6 An illustration for the interpretation of the ARL

As noted above, the average run length (ARL) is considered as a measurement to assess the quality of the control chart through how quickly the system detects an out-of-control signal in the process under investigation. In the context of the SPC, the ARL is defined as the expected number of the plotted points within the lower and upper limits of the control chart before the first out-of-control signal is detected. In this section, we perform a simulation study in order to discuss the behaviour of the ARL. We consider the in control and out of control ARL. The calculation of the ARL is based on the values of the BF with different scenarios of deviations from the mean and variance for the alternative model. The aim of this experiment is to see what effect taking a small or a big shift from the mean or variance for the alternative model has on the value of the ARL. In this simulation experiment, we will use the model,M_S which is used to simulate the data. This M_S model could be either the null model, M_0, in the case of the in control ARL, or some other models which measure shifts from the M_0 in the case of the out of control ARL. The alternative model, which is called M_A, will be

used in order to calculate the Bayes factors as described in Equation (5.4.3). In each experiment with different scenarios, we will use a different M_S while the M_A will be the same. In addition, the null model for all simulation experiments will be the same, $M_0 \sim N(0,1)$. To do this a Monte Carlo simulation with 100 runs of size 1000 observations is used.

State	M_S	M_A	M_0	ARL
In-control state	N(0,1)	N(1,1)	N(0,1)	21.82
Mean shifts from in-control state	N(0.5,1)	N(1,1)	N(0,1)	6.76
	N(1,1)	N(1,1)	N(0,1)	2.92
	N(1.5,1)	N(1,1)	N(0,1)	1.29
	N(2,1)	N(1,1)	N(0,1)	0.39
	N(2.5,1)	N(1,1)	N(0,1)	0.17
	N(3,1)	N(1,1)	N(0,1)	0.08
	N(3.5,1)	N(1,1)	N(0,1)	0.02

Table 5.3: ARL when the process are affected by mean shifts

Table 5.3 shows the different proposed models M_S, the same proposed alternative model M_A, and the null model. The values of the ARL, which were corresponding to a different M_S are also presented in Table 5.3. The different simulated data have been generated by taking different means-shifts for the model M_S. The Bayes factors have been calculated based on the same alternative model $M_A \sim N(1,1)$ in every experiment. The table shows in the first row that the value of the ARL under the in-control state which is defined as the numbers of points in the control state before facing the first out of control point, is large (ARL=21.82)

compared to the rest of the ARL values. Another finding from the Table 5.3 is that, when the mean shifts of the model M_S depart increasingly from the mean of the null model, $N \sim N(0, 1)$, then the process enters an out-of-control situation, and the ARL diminishes steeply from higher to smaller values. In addition, when the model M_S has a large mean shift, e.g. 3, the proposed procedure was very fast for identifying a shift from the null model (ARL=0.08).

Table 5.4 exhibits the values of the ARL in terms of taking different deviations of the variance shifts from the in-control state of the proposed models M_S. We used the same alternative model $M_A \sim N(0, 2)$ for calculating the Bayes factors. With respect of the in control state, the table shows that the value of the ARL (40.15) was large compared to the ARL values with variance shifts from the in control state. This table also provides that the ARL values decrease as the variance shifts for the model M_S increase (for variance shifts of 1.5 or 9 the ARL is equal to 7.01 or 0.36 respectively). As we can also see, the proposed procedure was very fast (ARL=0.36) for identifying a shift from the null model when the alternative model M_S has a large variance shift, e.g. 9.

State	M_S	M_A	M_0	ARL
In-control state	N(0,1)	N(0,2)	N(0,1)	40.15
variance shifts from in-control state	N(0,1.5)	N(0,2)	N(0,1)	7.01
	N(0,2)	N(0,2)	N(0,1)	3.05
	N(0,2.5)	N(0,2)	N(0,1)	1.69
	N(0,3)	N(0,2)	N(0,1)	1.37
	N(0,4)	N(0,2)	N(0,1)	0.81
	N(0,6)	N(0,2)	N(0,1)	0.44
	N(0,9)	N(0,2)	N(0,1)	0.36

Table 5.4: ARL when the process are affected by variance shifts

The ARL values of the in-control state and both mean and variance shifts from the in control are shown in Table 5.6. The different models M_S and the same an alternative model $M_A \sim N(2, 2)$ are also shown in Table 5.6. From this table, we can conclude that the ARL values are small compared to the ARL calculated through only shifts from the mean or variance. Moreover, the behaviour of the ARL is identical as in the mean or variance shifts, where the ARL value decreases as increasing both shifts from the mean and variance.

State	M_S	M_A	M_0	ARL
In-control state	N(0,1)	N(2,2)	N(0,1)	39.09
Both Mean and variance shifts from in-control state	N(1,2)	N(2,2)	N(0,1)	2.23
	N(2,2)	N(2,2)	N(0,1)	0.83
	N(3,2)	N(2,2)	N(0,1)	0.26
	N(1,3)	N(2,2)	N(0,1)	1.41
	N(2,3)	N(2,2)	N(0,1)	0.79
	N(3,3)	N(2,2)	N(0,1)	0.34
	N(1,6)	N(2,2)	N(0,1)	0.48
	N(2,6)	N(2,2)	N(0,1)	0.36
	N(3,6)	N(2,2)	N(0,1)	0.29
	N(1,9)	N(2,2)	N(0,1)	0.36
	N(2,9)	N(2,2)	N(0,1)	0.28
	N(3,9)	N(2,2)	N(0,1)	0.26

Table 5.5: ARL when the process are affected by both mean and variance shifts

The conclusion from the above results of the simulation study is that the proposed scheme performs satisfactorily at identifying the deterioration in the fit of the process.

5.7 Application of Bayesian monitoring to simulated and medical data

In this section, diagnostic checking based on Bayesian monitoring as discussed in the previous sections is applied to the simulated and medical data. The objective

of the model diagnostics is to assess the accuracy of a fitted model in a sequential way. The BF is used as a proposed tool of a model comparison to implement this task of continuous diagnostics. In addition the nonparametric control chart, based on the values of the BF and the runs-rules, is also used to examine whether the process is in an in-control or out-of-control state. The one-step-ahead cumulative predictive distribution and the INTPSR from the fitted model are used in order to calculate the BF. The idea of doing the model comparison based on the BF is to compare and contrast the performance of the fitted model under the null hypothesis and some alternative proposed models. These alternative models can be obtained by shifting the values of the mean, variance, or both of the INTPSR generated by the fitted model.

5.7.1 Simulation experiments

The goal of this section is to illustrate the mechanism of a Bayesian monitoring of the simulated data carried out to assess the performance of the proposed fitted model. The diagnostic procedure consists of two phases. In Phase I, when the fitted model describes the time series well in terms of its estimated parameters and its forecasts then the process is assured to be in the in-control state. In Phase II, we will use a recursive Bayesian monitoring approach, which is based on the BF as a diagnostic tool, in order to see whether the process is in the in-control state or not. In Phase I, the fitted model for the simulated data was optimised and the state variables and the hyper-parameters were estimated from the historical observed data by using the Liu and West particle filter (PF) algorithm. To evaluate the performance of the fitted model with respect to the cumulative predictive distribution, the PSR and the INTPSR were used as diagnostic tools for a goodness of fit. In Section 4.4.1, with respect to the simulation experiments,

we obtained a good results for the PSR and the INTPSR. This means that, in Phase I, the process of the time series is considered as being in the in control state. In Phase II, the BF is used as a diagnostic tool as in a methodology of a recursive Bayesian monitoring.

In the first step of the monitoring study, we will adopt to use the uniform distribution as the null model. Therefore, the values of the Bayes factors (BF) can be calculated based on the PS residuals of the fitted model and the proposed alternative models. From Chapter 4, where the proposed model fitted the data well, then the PS residuals are i.i.d random variables following a uniform distribution over the interval (0,1). The alternative model could be like some distribution measuring the deviation from the null model. The alternative model must be in some way similar to the null model. Thus, one possibility is that the alternative model can also be chosen as a uniform, but with different parameters than the null model parameters; i.e: $U(a_A, b_A)$.

Below we express the uniform parameters, a, b as a function of the mean μ_U and variance σ_U^2 of the uniform distribution. This allows us to consider shifts of a, b in the alternative distribution by considering shifts in the mean and variance. It can also provide information on the departure of the alternative distribution in terms of location (mean) and dispersion (variance) which are more commonly discussed in the monitoring literature.

The mean μ_U and variance σ_U^2 of the uniform distribution with parameters (a, b) are defined as follows:

$$\mu_U = \frac{a+b}{2}, \quad \text{and} \qquad \sigma_U^2 = \frac{(b-a)^2}{12} \tag{5.7.1}$$

Therefore, the parameters of the uniform distribution (a, b) can be obtained by

solving the mean and variance is Equation 5.7.1 for (a, b) as follows:

$$a = \mu_U - \frac{\sqrt{12}}{2}\,\sigma_U, \quad \text{and} \qquad b = \mu_U + \frac{\sqrt{12}}{2}\,\sigma_U \tag{5.7.2}$$

Under the null model, the PS residuals have the uniform distribution over the interval (0, 1) with the mean and variance equal to $\frac{1}{2}$ and $\frac{1}{12}$ respectively. In terms of calculating the BF, the alternative model can be selected by taking the shift from the null model within three scenarios as follows:

- **Mean shift**: The first scenario to define the alternative model is to have the shift (s) from the mean of the null model and keeping the variance constant as follows:

  $$\mu_A = \frac{1}{2} + s, \quad \text{and} \qquad \sigma_A^2 = \frac{1}{12} \tag{5.7.3}$$

  Then, the alternative model parameters (a_A, b_A) can be calculated by compensating the values of the alternative model mean and variance in Equation 5.7.3 in the formulas of (a, b) in Equations 5.7.2 as follows:

  $$a_A = \frac{1}{2} + s - \frac{\sqrt{12}}{2}.\frac{1}{\sqrt{12}} = s, \quad \text{and} \quad b_A = \frac{1}{2} + s + \frac{\sqrt{12}}{2}.\frac{1}{\sqrt{12}} = s + 1 \tag{5.7.4}$$

  Therefore, the alternative model is defined as a uniform on the interval $(s, s + 1)$.

- **Variance shift**: The second scenario for defining the alternative model is to have the shift (k) from the variance of the null model and keeping the mean constant as follows:

  $$\mu_A = \frac{1}{2}, \quad \text{and} \qquad \sigma_A^2 = \frac{1}{12}k \tag{5.7.5}$$

  Then, the alternative model parameters (a_A, b_A) can be calculated by compensating the values of the alternative model mean and variance in Equa-

tions 5.7.5 in the formulas of (a, b) in Equations 5.7.2 as follows:

$$a_A = \frac{1-\sqrt{k}}{2}, \quad \text{and} \quad b_A = \frac{1+\sqrt{k}}{2} \tag{5.7.6}$$

Therefore, the alternative model is defined as a uniform on the interval $(\frac{1-\sqrt{k}}{2}, \frac{1+\sqrt{k}}{2})$.

- **Both shift**: The third scenario for defining the alternative model is to have both shifts $(s), (k)$ from the mean and variance of the null model as follows:

$$\mu_A = \frac{1}{2} + s, \quad \text{and} \qquad \sigma_A^2 = \frac{1}{12}k \tag{5.7.7}$$

Then, the alternative model parameters (a_A, b_A) can be calculated by compensating the values of the alternative model mean and variance in Equations 5.7.7 in the formulas of (a, b) in Equations 5.7.2 as follows:

$$a_A = s + \frac{1-\sqrt{k}}{2}, \quad \text{and} \qquad b_A = s + \frac{1+\sqrt{k}}{2} \tag{5.7.8}$$

Therefore, the alternative model is defined as a uniform on the the interval $(s + \frac{1-\sqrt{k}}{2}, s + \frac{1+\sqrt{k}}{2})$.

With respect to the simulation studies in this section, the implementation of the sequential model monitoring in Phase II is applied on one run of the same simulated data as in Section 4.4.1. In Phase I, the LW particle filter algorithm was applied to this simulated data by using a LW.Pois model as a correct model and by LW.NB and LW.MPois models as wrong models (see subsection 3.6.1.2). The values of the BF have been calculated based on the predictive distribution of the PSR. The obtained results of the PSR and INTPSR for all proposed fitted models were provided in Figures 4.1 and 4.2. In Phase I, the results of the PSR and INTPSR for all models assured that the process of the simulated data is in an

in-control state. Regarding the first scenario, Figure 5.1 shows the behaviour of the resulting BFs with several proposed shifts on the mean of the null model PSR $\sim U(0,1)$ for the LW.Pois particle filter algorithm, which is defined as a correct model, with a Poisson simulated data. Figures B.2–B.4 show the behaviour of the BFs based on three different scenarios of the alternative model of the PSR the LW.Pois particle filter algorithm. These figures are provided in the appendix B.1.

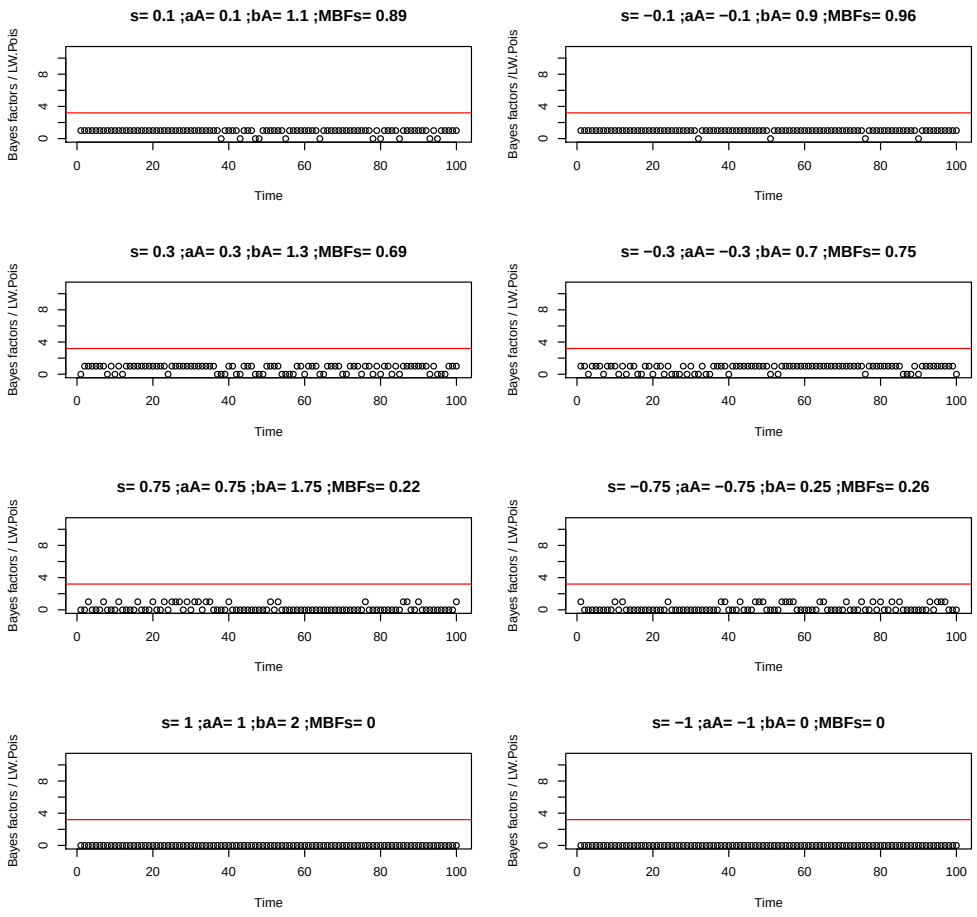

Figure 5.1: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on some proposed alternative models through the first scenario with mean shifts.*

The different shifts of the mean from the null model of the PSR are provided along with the alternative model parameters and an average of BFs for the first scenario in Figure 5.1. As can be observed from this figure, all values of the BFs with different shifts from the mean of the null model, which will lead to shifts on the alternative model parameters, are less than the threshold value 3.2. This result actually does not provide support on the null model. The reason of this is that the threshold 3.2 is not appropriate for us to make a decision. This is because the value of the BF at each time point is either equal to 1 when the PSR is within the range of the alternative model parameters or 0 if not. Therefore, the alternative model will be accepted when the BF is equal 1, whereas it will be rejected when the BF is equal 0. In terms of using the uniform distribution for the null model, we observed that the behaviour of the BFs for three different scenarios and all proposed fitted models, which are correctly or incorrectly specified, gave identical results as here in the first scenario. However, there is a problem in using the uniform distribution as a null model for calculating the BFs. This is because; the domain of the density function of the uniform distribution depends on its parameter range. This means that, the value of the BF at any time point equals either the difference between the alternative parameters if the value of the PSR is within the range of the alternative parameters or zero if not. In addition to that, using this specification, the shift of the variance is not independent of the mean.

The difficulty caused by the support of a $U(a_A, b_A)$ not being the same as the of $U(0, 1)$ can be avoided if we use a Beta(a_A, b_A) distribution as an alternative instead. We can be reparameterise this by setting $a_A = m_A p_A$ and $b_A = (1 - m_A)p_A$. Then the mean is simply m_A and the variance is $m_A(1 - m_A)/(p_A + 1)$. We can then look at the effect of varying m_A and p_A.

As an alternative to using a beta distribution for the PSR under M_A we can use a probit-normal distribution. That is, if R_t is the PSR at time t, then we give $Q_t = \Phi^{-1}(R_t)$ a normal N(m, k^2) distribution where $\Phi()$ is the standard normal distribution function. Thus, under M_0, Q_t ∼N(0,1). Under M_A we set $m = m_A$ and $k = k_A$, which are chosen appropriately. Of course, this is equivalent to using the INTPSR with its standard normal distribution under M_0 and a N(m_A, k_A^2) distribution under M_A.

Although the uniform and beta distributions can be used for the calculation of the Bayes factor (as discussed above) there is still some advantage in using the normal distribution instead. This approach seems to be adopted in the literature, see Smith (1985)and Thode (2002). There are available informal and formal tests for the normal distribution; for a discussion related to our models sees Smith (1985) and more generally Thode (2002). In Gaussian time series literature the theory of residuals-based model selection and monitoring is well established. By using the normal distribution in the context of the models considered in this thesis, the theory of residuals-based monitoring is extended to the non-Gaussian case.

In terms of the implementation of the monitoring in Phase II, the calculation of the BF is based on the predictive distribution of the INTPSR. Under the null hypothesis, when the process is in control in Phase I, the INTPSR is considered to be following a standard normal distribution with mean zero and variance one. The alternative models are suggested with the deviations of the null model N (0,1) in three different scenarios as before. Therefore, the proposed alternative models will follow a normal distribution, $N(m_A, k_A^2)$ where m_A, k_A^2 can be determined by the modeller. Note that if the deviations are selected so that they are close to the mean and variance of the standard normal distribution N(0,1), then it will be

difficult to separate the two models. On the other hand the monitoring process may fail to detect small changes in the process when large deviations are selected.

Figure 5.2 shows the behaviour of the resulting BFs with several proposed shifts on the mean or the variance of the null model for the LW. Pois particle filter algorithm, which is defined as a correct model. The different positive shifts of the mean and two proposed shifts on the variance from the null model of the INTPSR are provided along with the alternative model parameters and an average of BFs in Figure 5.2. The results of the BFs with all proposed shifts of the mean and variance are given in Figures B.5 and B.6 which are shown in the appendix B.1. As can be seen from Figure 5.2, the average of the BFs with different shifts of the alternative model for the correct model is less than 3.2, (0.8242–1.0796). This result provides evidence through the INTPSR that supports the null model rather than the alternative. Therefore, the process is in-control state. Besides, with a small shift of the mean, we can see that all values of the BFs are located under the threshold value 3.2. This is because a small shift is quite close to the mean of the null model. Therefore, the two densities for the null and the alternative are similar. However, when the mean shift is large, several values of the BF are larger than the threshold. That means that the distribution of these outliers given the alternative model has more weights. In this case, the model is defined as out of control state at these time points. In terms of the variance shifts, when we have a small shift ($k_A = 0.1$), there are some values of the BFs which exceed the threshold. In addition, with variance shift of 0.2, the last plot in the bottom panels indicates the same behaviour of the BF but with less magnitude. Based on the results obtained, the mean shifts provide the similar behaviour of the BF, but with variance shifts are not. Regarding to the behaviour of the BFs for all simulation experiments, the mean shifts and variance shifts are suggested to be quite small. This is because as we observed from the plots of the BFs with a

normal distribution, the outliers were detected when the shifts are quite small. In addition, with the big shift the model does not have the probability. Therefore, it will always be accepted.

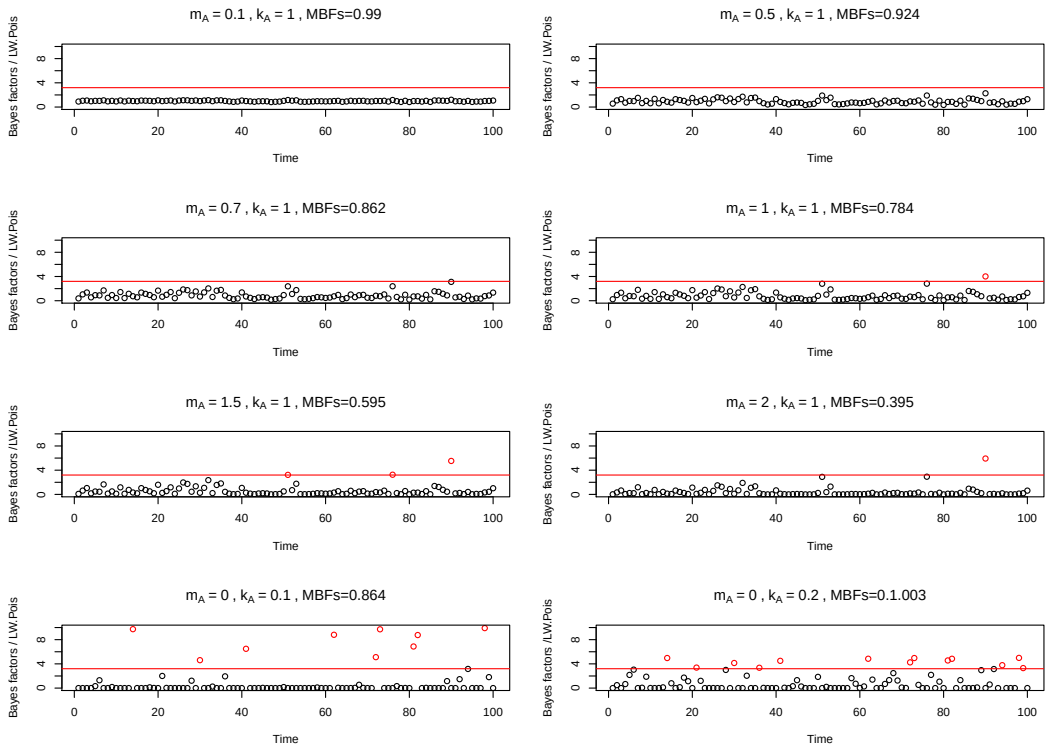

Figure 5.2: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on some proposed alternative models with different shifts from the mean and the variance of the INTPSR N(0,1).*

Table 5.6 exhibits the average of the BFs with different shifts of the mean and the variance of the null model, and the percentage of the BFs which is more than the threshold value. As can be seen from Table 5.6, for both different shifts, the percentage of the BF, which is larger than the threshold, is varying according to our choice of alternative models.

State	Mean	Variance	MBFs	% > Threshold
mean shifts	0.1	1.00	0.99	0.0 %
	0.3	1.00	0.963	0.0 %
	0.5	1.00	0.924	0.0 %
	0.75	1.00	0.862	0.0 %
	1.00	1.00	0.784	1 %
	1.5	1.00	0.595	3 %
	2.00	1.00	0.395	1 %
	2.5	1.00	0.224	1 %
variance shift	0.00	0.1	0.864	9 %
	0.00	0.2	1.003	13 %
	0.00	0.3	1.054	5 %
	0.00	0.5	1.069	0.0 %

Table 5.6: The output results of the BFs for the LW.Pois particle filter algorithm to a Poisson simulated data based on the different shifts of the mean or variance of the null model.

Figure 5.3 presents the behaviour of the BFs with the alternative model with the mean shift, INTPSR N (1.5, 1) in the top panel. The forecasting plot of the same simulated data is also provided in the bottom panel. As can be seen from the BF plot, there are three values which exceed the threshold 3.2. In addition, one of them has an extreme value which can be indicated as an outlier. Therefore, the model is out of control state at these time points. Besides, the model might be overall acceptable, but sometimes it might be not. This is caused by the existence of the outliers, and the BF helps us to detect these outliers.

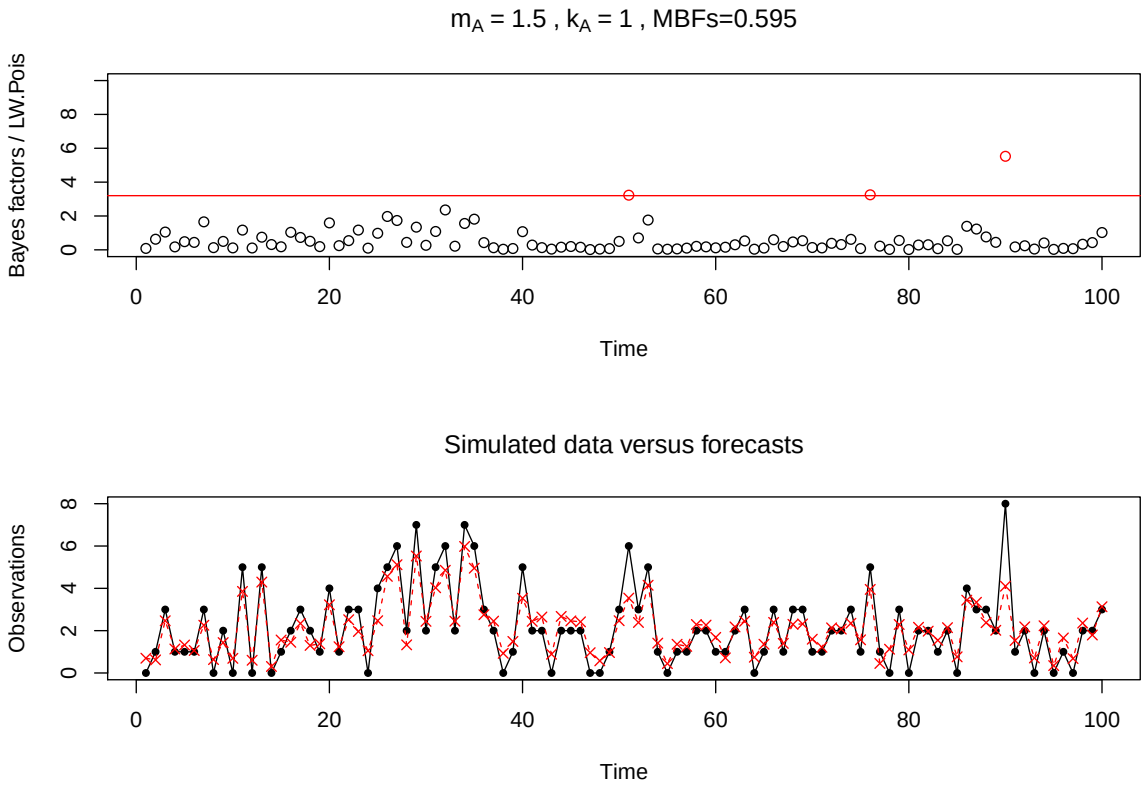

Figure 5.3: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on the proposed alternative model with mean shift; INTPSR∼ N(1.5,1).*

To determine whether the process is in an in-control or out-of-control state, Bayesian monitoring by a nonparametric control procedure based on binomial-type and runs-rules is used. This control procedure is based on the BF. In addition, the proposed control procedure is constructed using the classification of the three categories. These categories are based on the rules interpretation by Jeffreys (1935,1961) of the BFs. If the resulting value of the BF at a specific time point is less than 3.2, this leads to the Bayes factor lying in category 0. Therefore, the process is considered as being in an in-control state. The process is consid-

ered as being in an out-of control state when the value of the BF is assigned to category 2, or two consecutive values of the Bayes factor in a window of length four are assigned to category 1.

Based on the Jeffreys suggestion for the control process provided in Table 5.2, the results of the control chart of the BF with the alternative model INTPSR $\sim N(1.5, 1)$ indicate that all signals are in control state. However, if we want to have a last conservative point of view, then we can lower the threshold to 5 instead of 10. Therefore, the process is considered as being in an out of control state to category 1 when the value of the BF is between 3.2–5, or to category 2 when the BF is more than 5. Thus, the value of the BF in category 2 can be expressed as outliers. The obtained results of the proposed control procedure of the BFs with alternative model INTPSR $\sim N(1.5, 1)$ for the LW.Pois particle filter algorithm with the Poisson simulated data is provided in Tables 5.7. The result from Table 5.7 illustrates that the process is considered as being in an out-of-control state to category 2 at time point 90. For all other time points the process is considered as being in an in-control state.

In the second and third simulation experiments, the behaviour of the BFs based on the uniform and normal distributions and the proposed control procedure provided the quite similar results as here in the first simulation experiment when the LW.Pois particle filter algorithm was used as a correct model.

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	43	44	45	46	47	48	49	50	51	52	53	54	55	56
Category	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Time	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Category	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Time	85	86	87	88	89	90	91	92	93	94	95	96	97	98
Category	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Time	99	100												
Category	0	0												

Table 5.7: Control classification for Phase II of the INTPSR of LW.Poiss algorithm and a Poisson simulated data using the Bayes factor with alternative model; INTPSR $\sim N(1.5, 1)$. In-control signals are referenced by 0 and out-of-control are referenced by 1.

5.7.2 Analysis of asthma data

The aim of the simulation study in the previous section was to apply the Bayesian monitoring model using a BF to assess the performance of the fitted model in a recursive way and, in addition, to understand the expected behaviour of the BF,

which is calculated based on whether the PSR or INTPSR, when the proposed fitted model is correctly or incorrectly determined. In this section, the diagnostic checking, based on the BF and the proposed control procedure, is applied to the asthma patients' data. The objective of the model diagnostics is to assess the performance of the proposed fitted model in terms of a one-step-ahead predicted distribution. According to the theory, when the model is perfectly specified, the distribution of the PSR should be uniform over the interval (0,1) or the distribution of the INTPSR should be a standard normal distribution. The BFs for all proposed fitted models are calculated based on Equation (5.4.3), where the INTPSR under the null model follows $N(0,1)$ and the alternative model can be selected based on the shape of the histogram of the INTPSR from each fitted model of the asthma data.

In terms of calculating the BFs with the uniform distribution, Figures B.7–B.9 show the behaviour of the BFs based on the last 104 observations of the asthma data through three different scenarios (mean shifts, variance shifts or both shifts). These figures are provided in the appendix B.1. As can be seen from these figures all values of the BF for three different scenarios are less than the threshold 3.2. This result does not provide an evidence in favour of the null model against the alternative model and this is what we have clarified before in the simulation study.

On the other hand, Figure 5.4 presents the behaviour of the BFs when using the normal distribution as the null model and proposed alternative models of the INTPSR with different mean shifts. The values of the BF with all proposed shifts for the first and the second scenarios are shown in Figures B.10–B.11 which are provided in the appendix B.1. As can be seen from Figure 5.4 with small shifts on the mean, all values of the BF are less than the threshold 3.2. This means

that the null model will not be rejected. However, when the mean shifts are larger, the BF plot indicates some values exceed the threshold. These values can be expressed as outliers and the null model will be rejected at these time points. This variable behavior of the BF indicates that the proposed fitted model may be suitable for asthma data at some time points, and is inappropriate at other time points. This recursive test for checking the fitted model cannot be implemented by other tests such as in a Smith (1985) approach.

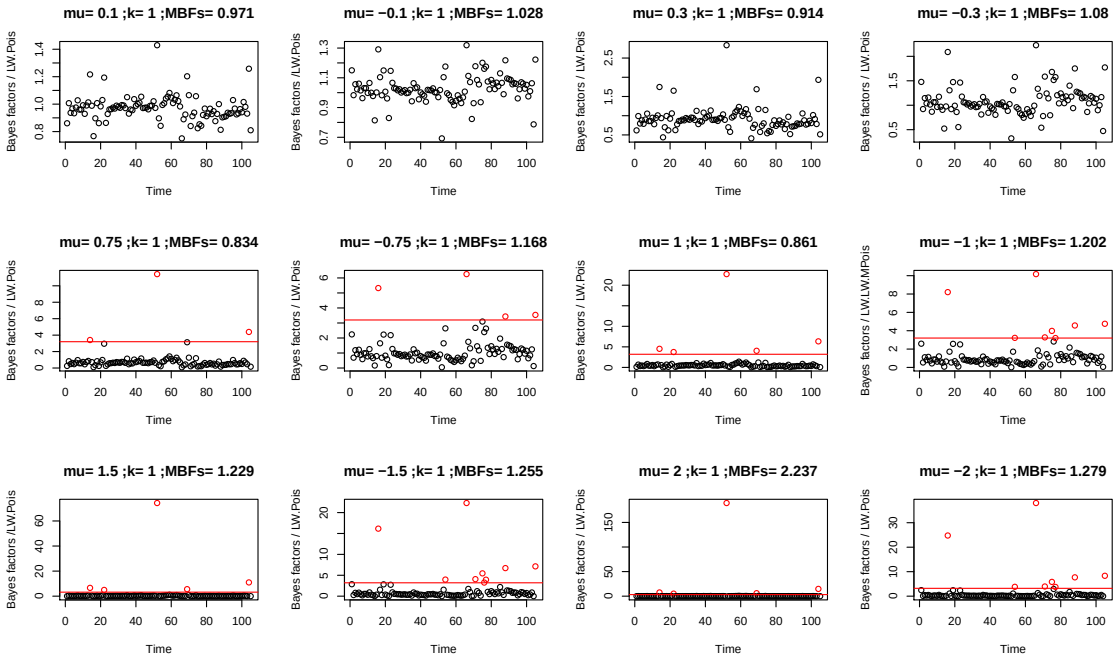

Figure 5.4: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a asthma data based on some proposed alternative model of the INTPSR with mean shift*

Figure 5.5 shows the behaviour of the BFs for the last 105 observations with the alternative model with the mean shift, INTPSR N (0.75, 1) in the top panel. The forecasting plot of the real asthma data against forecasts for the last 105 observations by applying the LW.pois particle filter algorithm is also provided in

the bottom panel. As can be observed from the BF plot, three values exceed the threshold 3.2, and one of them has an extreme value which can be indicated as an outlier. Therefore, the null model at these time points will be rejected.

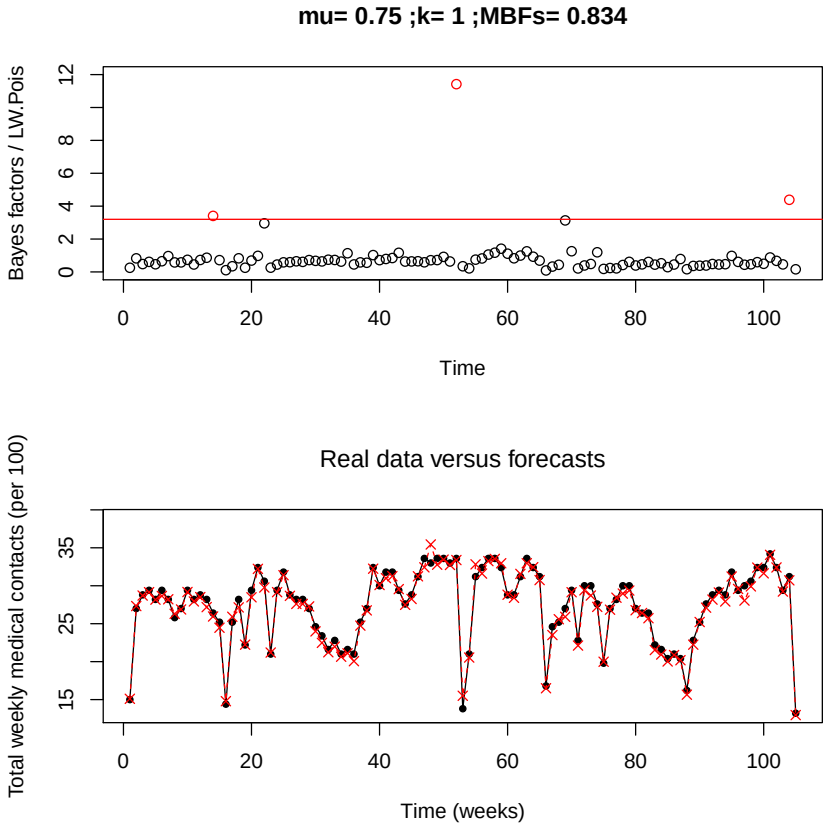

Figure 5.5: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to the last 105 observations of a asthma data based on the proposed alternative model with mean shift; INTPSR∼ N(0.75,1) and the forecasting .*

On the other hand, based on the shape of the histogram of the INTPSR, which is shown in Figure 4.7 in Chapter 4, the proposed fitted model was rejected. However, using the BF, the proposed fitted model was accepted. In addition, the forecast plot, which is presented in the bottom panel of Figure 5.5 , provides a good predication for the real asthma data. This result leads us to conclude that

the proposed model fitted the asthma data well.

To investigate whether the process is in an in-control or out-of-control state, the proposed control procedure based on binomial-type statistics with the runs-rules as in the simulation experiments is used, where the control procedure has been based on guidelines for the BF. The control procedure of the classification of the BF values with proposed alternative model and mean shift, INTPSR $\sim$N(0.75,1) is provided in Table 5.8. As can be seen from Table 5.8 1 out of 105 signals is out-of-control signals and it is allocated to category 2 whereas the rest of the signals are in control state. As a result, the process is considered as being in an in-control state at these time points while it is out-of-control at time point 52.

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Time	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	43	44	45	46	47	48	49	50	51	52	53	54	55	56
Category	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Time	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	85	86	87	88	89	90	91	92	93	94	95	96	97	98
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	99	100	101	102	103	104	105							
Category	0	0	0	0	0	1	0							

Table 5.8: Control classification for Phase II of the INTPSR of LW.Poiss algorithm and a Poisson simulated data using the Bayes factor with alternative model; INTPSR $\sim N(0.75, 1)$. In-control signals are referenced by 0 and out-of-control are referenced by 2.

With respect to applying the LW.NB particle filter algorithm on the asthma data, the behaviour of the BFs of the last 105 observations of the asthma data with the uniform distribution through three different scenarios are presented in Figures B.12–B.14. These figures are shown in the appendix B.1. For three scenarios, all values of the BF do not exceed the threshold 3.2. This result actually does not

provide support on the null model as we clarified before in the simulation study.

In terms of using the normal distribution as the null model, Figure 5.6 shows the behaviour of the BF with some proposed alternative model of the INTPSR with different mean shifts. The values of the BF with all proposed shifts for whether the mean or the variance are shown in Figures B.15–B.16 which are provided in the B.1. As can be seen from Figure 5.6, all values of the BF are less than the threshold 3.2. This is because the null and alternative models are unlikely to follow the normal distribution and this is what we observed from the shape of the histogram of the INTPSR in Figure 4.8 in Chapter 4. Therefore, the probability density function of that normal distribution will be low. Thus, when we obtain the alternative model from the null model, it will shift from the low distribution. So, the value of the BF does not change much. This means that both the probability density functions of the null and alternative models have comparable values. Therefore, for this result of the BF, the null model will not be rejected. With respect to investigating whether the process is in an in-control or out-of-control state, the control procedure based on the BF with different proposed alternative models of the INTPSR with mean shifts provides that the process is in-control state at all time points.

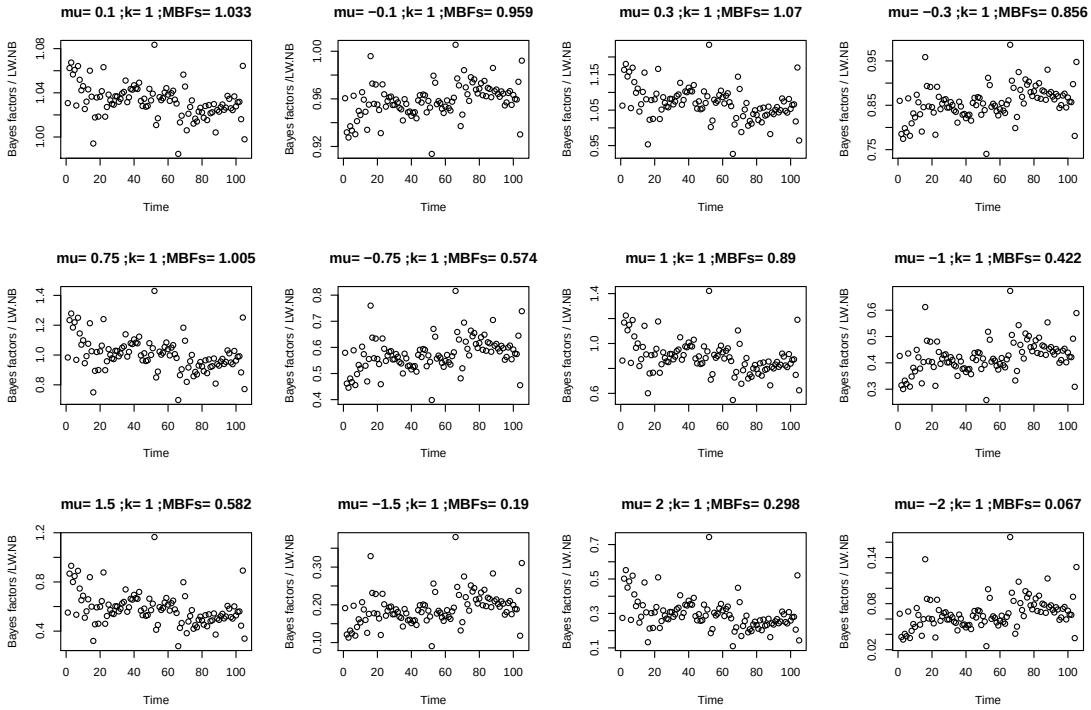

Figure 5.6: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to a asthma data based on some proposed alternative model of the INTPSR with different mean shift*

Based on applying the Liu and West particle filter algorithm with the mixture Poisson model (LW.MPois) on the asthma time series data, Figures B.17–B.19 present the behaviour of the BFs of the last 105 observations of the asthma data with three scenarios of the alternative model of the PSR. These figures are shown in the appendix B.1. As can be noted from these figures that all values of the BF for three different scenarios do not exceed the threshold 3.2. This result actually does not provide support on the null model as we clarified before in the simulation study.

In contrast, Figure 5.7 presents the behaviour of the BFs when using the normal distribution as the null model and proposed alternative models of the

INTPSR with different mean shifts. The values of the BF with all proposed shifts for the first and the second scenarios are shown in Figures B.20–B.21 which are provided in the appendix B.1. The conclusion to be drawn from Figure 5.7 is that with small shifts on the mean, all values of the BF are less than the threshold 3.2. However, when the mean shifts are larger, the BF plot indicates some values exceed the threshold. These values can be expressed as outliers and the null model will not be accepted at these time points.

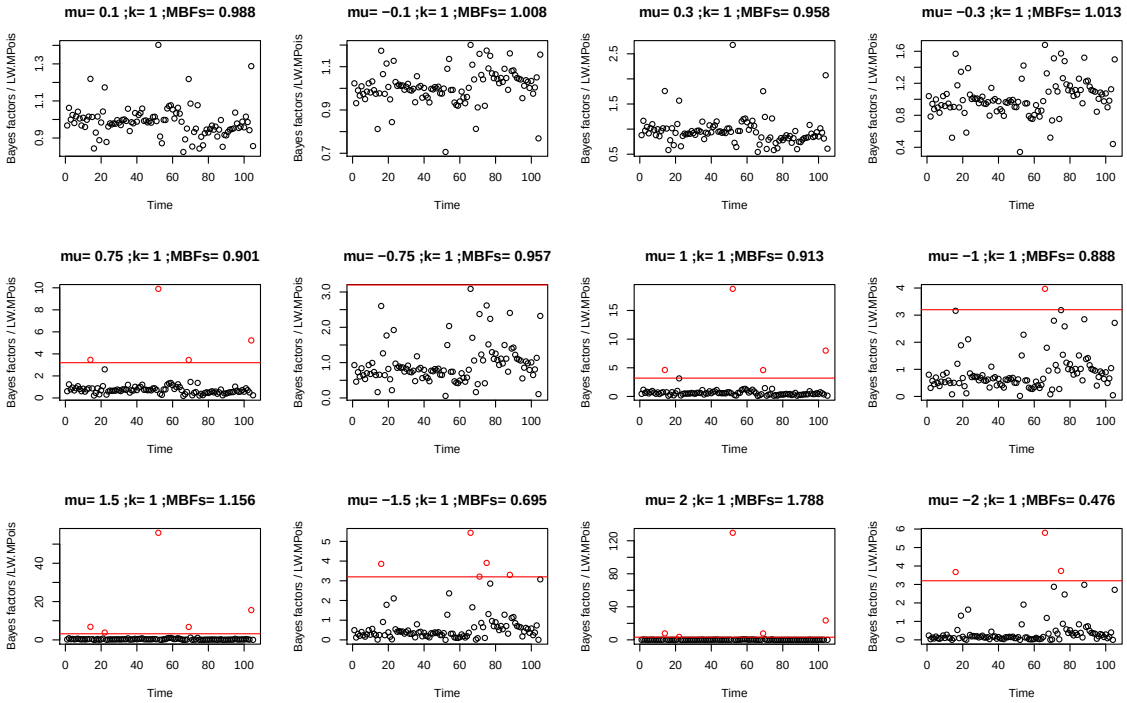

Figure 5.7: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to a asthma data based on some proposed alternative model of the INTPSR with mean shift*

Figure 5.8 shows the behaviour of the BFs for the last 105 observations with the alternative model with the mean shift, INTPSR N (0.75, 1) in the top panel. The forecasting plot of the real asthma data against forecasts for the last 105

observations by applying the LW.pois particle filter algorithm is also provided in the bottom panel. It can be seen through the BF plot, there are four values exceed the threshold 3.2, and one of them has an extreme value which can be indicated as an outlier. Therefore, the null model at these time points will be rejected.

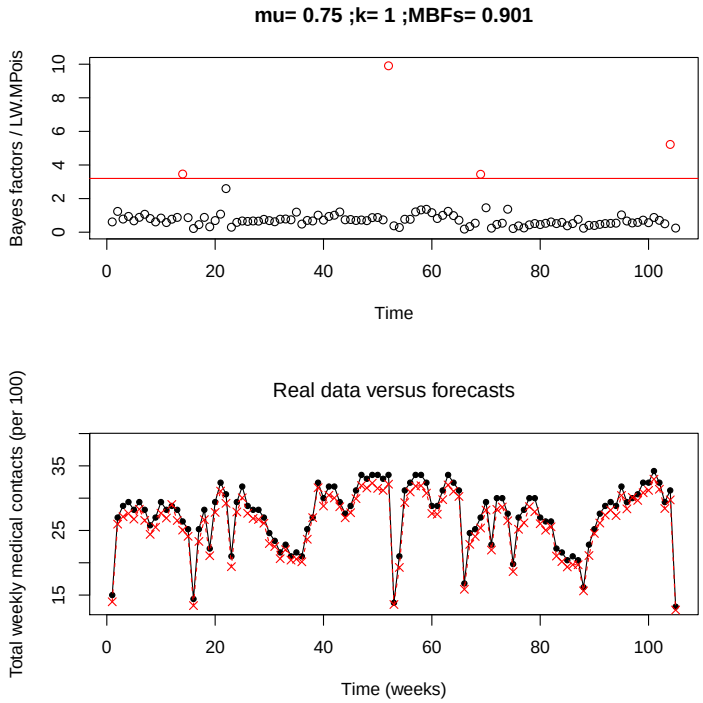

Figure 5.8: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to the last 105 observations of a asthma data based on the proposed alternative model with mean shift; INTPSR∼ N(0.75,1) and the forecasting .*

With respect to examining whether the process of BFs is in an in-control or out-of-control state, Table 5.9 offers the results of the proposed control scheme based on the classification of the BFs via the proposed alternative model, INTPSR ∼N(0.75,1). From this table we observe that 1 out of 105 signals is an out-of-control signal and it is allocated to category 2 whereas the rest of the signals

are in the in-control state. As a result, the process is considered as being in an in-control state at these time points while it is out-of-control at time point 52.

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Time	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	43	44	45	46	47	48	49	50	51	52	53	54	55	56
Category	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Time	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Category	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Time	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	85	86	87	88	89	90	91	92	93	94	95	96	97	98
Category	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	99	100	101	102	103	104	105							
Category	0	0	0	0	0	1	0							

Table 5.9: Control classification for Phase II of the INTPSR of LW.Poiss algorithm and a Poisson simulated data using the Bayes factor with alternative model; INTPSR $\sim N(0.75, 1)$. In-control signals are referenced by 0 and out-of-control are referenced by 2.

The overall conclusion, according to the above results of the Bayesian monitoring model using the BFs in a sequential way, is that the null model of the PSR and INTPSR is accepted for all proposed fitted models to the asthma data. To

investigate whether the process is in-control or not, the proposed control procedure for the LW.Pois and LW.MPois particle filter algorithm illustrated that the process of the BFs via the proposed alternative model of the INTPSR $\sim$N(0.75,1) are overall in an in-control state over time. However, for all proposed fitted models, the process of the BFs is out-of-control at some time points. In addition, based on the BF for the LW.NB particle filter algorithm, the model was accepted as in the in-control state for all proposed alternative models with mean shifts.

5.7.3 Analysis of SIDS data

In this section we will use the same diagnostic approach as we did for the asthma data to test the performance of the proposed fitted model for the SIDS data. The BF and the control chart were adopted as the Bayesian tools to achieve this purpose. The aim of this examination is to monitor the behaviour of the BFs of the process of the SIDS data in a recursive manner. In Phase I, in order to fit the SIDS data, the LW.MPois algorithm was used to estimate the posterior distribution of the hidden variables and the hyper-parameters of the DGLM. With respect to the implementation of the Bayesian monitoring model for the proposed fitted model, Figures B.22–B.24 provide the behaviour of the BFs via three different scenarios of the alternative model of the PSR. These figures are presented in the appendix B.1. As can be observed from all figures, all values of the BF are less than the threshold value 3.2. This result does not provide an evidence in favour of the null model against the alternative model and this is what we have clarified before in the simulation study.

Based on using the normal distribution as the null model, Figure 5.9 shows the behaviour of the BF with some proposed alternative model of the INTPSR

with different mean shifts. The values of the BF with all proposed shifts for the first and second scenarios are shown in Figures B.25–B.26 which are provided in the appendix B.1. As can be seen from Figure 5.9, all values of the BF are very small and less than the threshold 3.2. This is because the null and alternative models are unlikely to follow the normal distribution and this is what we observed from the shape of the histogram of the INTPSR in Figure 4.11 in Chapter 4. Consequently, for this result of the BF, the null model will not be rejected at all time points. In order to investigate whether the process is in an in-control or out-of-control state, the control procedure based on the BF with different proposed alternative models of the INTPSR with mean shifts provides that the process is in-control state at all time points.

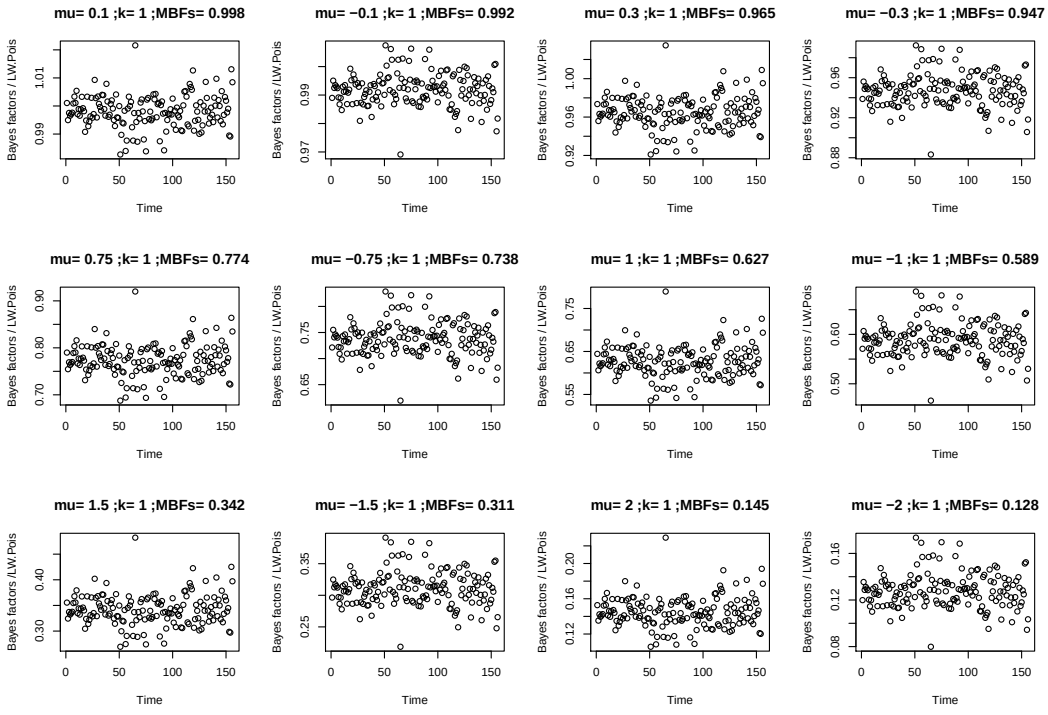

Figure 5.9: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a SIDS data based on some proposed alternative model of the INTPSR with different mean shift*

5.8 Summary

This chapter has introduced and given the motivation for using Bayesian monitoring. In addition, the background material on the Bayesian model diagnostics including a statistical process control, control charts and a Bayes factor (BF) have been provided. The purpose of the simulation experiments in this chapter was to apply a methodology of the Bayesian diagnostic analysis to determine the expected behaviour of the BFs of the process of the simulated data over time, when the proposed fitted model is correctly or incorrectly specified. The BFs are used to compare the null model to several alternative models of the PSR or the INTPSR in order to evaluate how much evidence the BF provides in favour of one model rather than the other. In addition, a nonparametric control chart with runs-rules has been used to investigate whether the process of the BFs is in an in-control or out-of-control state. The same simulated data in one run as in Section 4.4.1 was used to achieve this objective. It is concluded from the simulation experiments based on the PSR with a uniform distribution that the BF values are less than the threshold 3.2. We obtained this result regardless of whether the fitted model is correctly or incorrectly determined. However, this result does not provide evidence in favour of the null model against the alternative model. In contrast, the behaviour of the BF with the INTPSR and the normal distribution with larger mean shifts provides some outliers. Moreover, with respect to the simulation study, the suggestion of the mean and variance shifts should be quite small so that the sequential monitoring way using the Bayes factor will be effective in order to detect the outliers. In terms of applying the SPC, the proposed control scheme provided effective monitoring and was able to detect whether the process was in an as in-control or out-of-control state.

We applied the suggested procedure of the Bayesian monitoring model to both asthma and sudden infant death syndrome (SIDS) data. The obtained results of the behaviour of the BFs for all the proposed fitted models of both medical datasets showed that all the BFs values are less than the threshold 3.2. Moreover, the proposed control chart explains that the process of the BFs is in an in-control state. But when the BFs is calculated based on the INTPSR with a normal distribution, some outliers were detected for the LW particle filter algorithms with Poisson and mixture Poisson models with asthma data. However, the fitted models overall were accepted.

Chapter 6

Conclusions and further work

6.1 Conclusion

This thesis focuses on applications of Bayesian dynamic models in medical research. The principal aim is to provide a forecasting framework for count medical time series data. Sequential Monte Carlo techniques based on Liu and West and Storvik particle filter (PF) algorithms have been developed to address the problem of estimation of the hidden state and hyper-parameters for the dynamic generalised linear models (DGLM) when a new observation arrives. PF algorithms have a powerful theoretical foundation within the recursive Bayesian estimation framework, where they provide an effective method for approximating the marginal posterior distribution of interest in the DGLM. The aim of this chapter is to summarise what has been discussed in the thesis.

The literature on Bayesian time series models including the state space model (SSM) and the DGLM were introduced in Chapter 2. In addition, novel recursive Bayesian methods for tracking the state process of the DLM and the DGLM by

using the Kalman Filter (KF), the extended Kalman Filter (EKF) or the basic PF were presented.

In terms of handling the inference of the latent variable and the hyper-parameters of the DGLM, Chapter 3 presented two novel Bayesian algorithms based on PFs. The Liu and West and Storvik algorithms were proposed to learn about the joint posterior distribution of the state and hyper-parameters of the DGLM. We have applied these algorithms to the simulated and real medical data. With respect to the first simulation study, the tracking results of the model parameters showed that the LW.Pois algorithm outperformed the STK.Pois where the estimated unknown hyper-parameters are closer to the true values. In addition, the LW.Pois algorithm provided more accurate one-step-ahead forecasting compared to the STK.Pois algorithm, where the performance of the algorithms were evaluated based on the smaller value of the average Monte Carlo of the root mean squared error (RMSE). Beside, the obtained results by using the LW.Pois particle filter algorithm of the approximation posterior distribution of the hyper-parameters at the last time point are quite similar from the ones that were obtained by the MCMC. The second simulation experiment focused on investigating the performance of the Liu and West algorithm in terms of quality of parameter estimates and the accuracy of forecasting when the fitted model is correctly or incorrectly specified. Monte Carlo simulation results showed that the best results were obtained from the application of the LW. Pois algorithm, even if it was specified incorrectly.

The two proposed PF algorithms were applied to two medical datasets. With respect to the asthmatic dataset, the main target is to make a one-step-ahead forecast of weekly total medical visits for school aged children who suffer from asthma. The overall results calculated by applying all proposed fitted models

showed that the LW.Pois algorithm outperforms the other proposed models in terms of the performance accuracy of the prediction, where the predicted values calculated by the LW. Pois algorithm were closer to the actual values. This leads to the model having the smallest value of the RMSE compared to the others.

The objective of the second application was to examine if there is any effect of the environmental temperature on increasing the rate of sudden infant death syndrome (SIDS). In order to achieve this objective, we extended the DGLM so that it contains a set of covariates such as temperature and trend and seasonal components. The obtained result, based on applying the LW.Pois algorithm was a temperature coefficient estimate of -0.0354. This means that there is a negative impact of the environmental temperature on increasing the risk of SIDS. Additionally, this result was quite close to the result of Campbell (1994) which was computed using the Zeger (1988) method and the generalised Poisson model by Nelder and Wedderburn (1972).

In Chapter 4, the methodology of the model diagnostics of the non-Gaussian time series models by Smith (1985) was presented. This method is used to assess the performance of the fitted model based on the P-score residuals (PSR), which are calculated from the one-step-ahead predictive distribution of the fitted model. The inverse transformed P-scores residuals (INTPSR) for obtaining the i.i.d normally distributed random variable was discussed. The proposed diagnostic procedure was applied to the simulated data and two real medical datasets. In terms of the results obtained from the Monte Carlo study for the different simulation experiments, the behaviour of the P-score residuals (PSR) and the inverse transformed P-scores residuals (INTPSR) for both correctly and incorrectly specified models is quite similar. In addition, the PSR and INTPSR satisfy the assumption of a uniform distribution and standard normal distribution respec-

tively. With respect to the asthma and SIDS data, the obtained results of the model diagnostics for all the proposed models for both medical datasets showed a poor performance regarding the histogram shape of the PSR and INTPSR. This result has been extracted by looking at the histogram shape of the PSR and INTPSR over time. However, in the DGLM, the distribution shape of all observations over time does not reflect on the shape distribution of each single observation at each time point. In this case, it is therefore necessary to investigate the distributional assumption of the PSR and INTPSR at each time point.

In Chapter 5, we provided a Bayesian monitoring model for assessing sequentially the performance of the proposed fitted models. The Bayes factors (BFs) and nonparametric control chart were used as tools for achieving this target. The BF is used as a comparison measurement for providing evidence to support either the null model or the alternative model. The guideline of Jeffreys (1961) has been used to interpret the behaviour of the BF values. The proposed control procedure is designed based on the BFs. We adopted the binomial-type control chart with the proposed runs-rules with a window of length 4 and the guidelines of Jeffreys (1961) for detecting whether the process is in an in-control or out-of-control state. The BFs are calculated based on the PSR and INTPSR and the important feature of this chapter is where we combine the results obtained from Chapter 4 with monitoring with a nonparametric control chart. The suggested procedure of the Bayesian monitoring model was applied to simulated data and both real medical datasets. With respect to the obtained results of the BFs based on the PSR for both simulated and medical data, all values of the BFs through three different scenarios are shown under the threshold 3.2. This result does not provide an evidence in favour of the null model against the alternative model and this is because the threshold is not appropriate for us to make a decision. However, when the INTPSR with the normal distribution were used to calcu-

late the BFs, some outliers were detected at different time points, but the fitted model overall cannot be rejected. Beside, the mean shifts and variance shifts are suggested to be quite small in order to detect outliers. In terms of applying the nonparametric control chart, the proposed control scheme provided an efficient way of monitoring and detecting out-of-control signals of the process.

6.2 Further research directions

Based on our work in this thesis, we suggest some possible directions for future research that relate to this study as follows:

- A response distribution other than one belonging to the exponential family can be considered. This is a useful consideration, as other distributions may describe the count data more accurately, and hence may be more capable forecasts.

- In terms of model parameters inference, it is interesting to develop PF algorithms by involving MCMC step within the Liu and West algorithm. This incorporation probably leads to improve the performance of the Liu and West algorithm in order to provide accurate estimations of the unknown static parameters in the DGLM.

- Further to the work of the Bayesian monitoring presented in Chapter 5, it will be interesting to evaluate a performance of the proposed control procedure for investigating whether the process is considered as in-control state or not. This might be done by studying comparisons between our approach and more standard statistical process control methods for autocorrelated time series data.

- Although this research is focused on medical applications and our aim is to show how Bayesian time series analysis can be used in medical time series data, the methods that we have discussed here are general and they do not require any assumptions about the medical data; it is only collecting time series count data. Therefore, these methods can be applied in other kinds of count data in different disciplines.

Appendices

Appendix A

Derivation of the mean and variance of the importance function of Equations (3.5.1)–(3.5.1)

This appendix provides the derivation of the mean and variance of the importance function using the EKF. The DGLM is completely defined through the state model together with the observation model and the link function. The observation model can be expressed as

$$y_t = g(x_t) + \nu_t, \qquad \nu \sim N(0, \sigma_t^2),$$

where the mean and variance of the observation model are given by $E(y_t|x_t) = g(x_t)$ and $Var(y_t|x_t) = \sigma_t^2$. The link function for a Poisson model is defined as $\lambda_t = \exp(x_t)$. The State model is given by

$$x_t = \mu(1-\phi) + \phi x_{t-1} + \omega_t \qquad \omega_t \sim N(0, W_t),$$

where the mean and variance of the observation model are given by

$$E(x_t|x_{t-1}) = \mu(1-\phi) + \phi x_{t-1} = a_t \tag{A.0.1}$$

$$Var(x_t|x_{t-1}) = W_t = R_t \tag{A.0.2}$$

In the context of the DGLM, the observation model is expressed as follows: $y_t|x_t \sim Pois(\exp(x_t))$, where $E(y_t|x_t) = \exp(x_t)$ and $Var(y_t|x_t) = \exp(x_t)$. Matching both formulas of the mean and variance of the observation model, we can obtain the following:

$$\begin{aligned} E(y_t|x_t) &= f(x_t) = \exp(x_t) \\ Var(y_t|x_t) &= \sigma_t^2 = \exp(x_t) \end{aligned}$$

In order to linearize the observation equation, the first order Taylor expansion around the current state mean, $\hat{x}_t = E(x_t|x_{t-1})$ is used to achieve that.

$$y_t = g(\hat{x}_t) \mid_{x_t=\hat{x}_t} + \frac{\partial g(x_t)}{\partial x_t} \mid_{x_t=\hat{x}_t} (x_t - \hat{x}_t) + \nu_t, \tag{A.0.3}$$

where

$$g(\hat{x}_t) = g(x_t) \mid_{x_t=\hat{x}_t} = \exp(\hat{x}_t) = \exp(\mu(1-\phi) + \phi x_{t-1})$$

and

$$\frac{\partial g(x_t)}{\partial x_t} \mid_{x_t=\hat{x}_t} = \exp(\hat{x}_t) = \exp(\mu(1-\phi) + \phi x_{t-1})$$

Therefore, the linear observation equation in Equation (A.0.3) can be written as

follows:

$$\begin{aligned}
y_t &= \exp(\mu(1-\phi)+\phi x_{t-1}) + \exp(\mu(1-\phi)+\phi x_{t-1})[x_t - \mu(1-\phi) - \phi x_{t-1}] + \nu_t \\
&= \exp(\mu(1-\phi)+\phi x_{t-1}) + \exp(\mu(1-\phi)+\phi x_{t-1})x_t - \mu(1-\phi)\exp(\mu(1-\phi)+\phi x_{t-1}) \\
&\quad -\phi x_{t-1}\exp(\mu(1-\phi)+\phi x_{t-1}) + \nu_t \\
&= \exp(\mu(1-\phi)+\phi x_{t-1}) + \exp(\mu(1-\phi)+\phi x_{t-1})[\mu(1-\phi)+\phi x_{t-1}+\omega_t] - \\
&\quad \mu(1-\phi)\exp(\mu(1-\phi)+\phi x_{t-1}) - \phi x_{t-1}\exp(\mu(1-\phi)+\phi x_{t-1}) + \nu_t \\
&= \exp(\mu(1-\phi)+\phi x_{t-1}) + \exp(\mu(1-\phi)+\phi x_{t-1})\mu(1-\phi) + \\
&\quad \exp(\mu(1-\phi)+\phi x_{t-1})\phi x_{t-1} + \omega_t\exp(\mu(1-\phi)+\phi x_{t-1}) \\
&\quad -\mu(1-\phi)\exp(\mu(1-\phi)+\phi x_{t-1}) - \phi x_{t-1}\exp(\mu(1-\phi)+\phi x_{t-1}) + \nu_t \\
y_t &= \exp(\mu(1-\phi)+\phi x_{t-1}) + \exp(\mu(1-\phi)+\phi x_{t-1})\omega_t + \nu_t
\end{aligned}$$

Thus, the mean and variance of the linear observation process are given by:

$$\begin{aligned}
E(y_t|x_{t-1}) &= \exp(\mu(1-\phi)+\phi x_{t-1}) = f_t \\
Var(y_t|x_{t-1}) &= [\exp(\mu(1-\phi)+\phi x_{t-1})]^2 W_t + \sigma^2 \\
Var(y_t|x_{t-1}) &= [\exp(\mu(1-\phi)+\phi x_{t-1})]^2 W_t + \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t) = Q_t
\end{aligned}$$

Where $\sigma^2 = \exp(x_t) = \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t)$.

Given a previous state x_{t-1}, the calculation of a covariance between current state x_t and current observation y_t is given by:

$$\begin{aligned}
Cov(x_t, y_t|x_{t-1}) &= Cov(x_t, g(\hat{x}_t) + \frac{\partial g(x_t)}{\partial x_t}x_t - \frac{\partial g(x_t)}{\partial x_t}\hat{x}_t) \\
&= Cov(x_t, \frac{\partial g(x_t)}{\partial x_t}x_t) = \frac{\partial g(x_t)}{\partial x_t}Cov(x_t, x_t) \\
&= \frac{\partial g(x_t)}{\partial x_t}Var(x_t) = \exp(\mu(1-\phi)+\phi x_{t-1})W_t = \mathbb{A}_t
\end{aligned}$$

From above results, the importance function, in this case, follows a normal distribution $x_t|x_{t-1}, y_t \sim N(\mathfrak{m}_t, \mathbb{C}_t)$ where the mean and variance can be calculated as follows:

$$\mathfrak{m}_t = a_t + \frac{\mathbb{A}_t}{Q_t}(y_t - f_t)$$

$$\mathfrak{m}_t = \mu(1-\phi)+\phi x_{t-1}+\frac{\exp(\mu(1-\phi)+\phi x_{t-1})\, W_t\, [y_t - \exp(\mu(1-\phi)+\phi x_{t-1})]}{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, W_t + \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t)}$$

Where $\omega_t \sim N(0, W_t)$

$$\mathbb{C}_t = R_t - \frac{\mathbb{A}_t}{Q_t} Q_t \frac{(\mathbb{A}_t)^T}{Q_t} = R_t - \frac{\mathbb{A}_t(\mathbb{A}_t)^T}{Q_t}$$

$$\mathbb{C}_t = W_t - \frac{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, [W_t]^2}{[\exp(\mu(1-\phi)+\phi x_{t-1})]^2\, W_t + \exp(\mu(1-\phi)+\phi x_{t-1}+\omega_t)}.$$

Therefore $x_t | x_{t-1}, y_t \sim N(\mathfrak{m}_t, \mathbb{C}_t)$

Appendix B

Graphs

B.1 Resulting Graphs of The MCMC Algorithm and The Behaviour of The Bayes Factors for Simulation Experiments

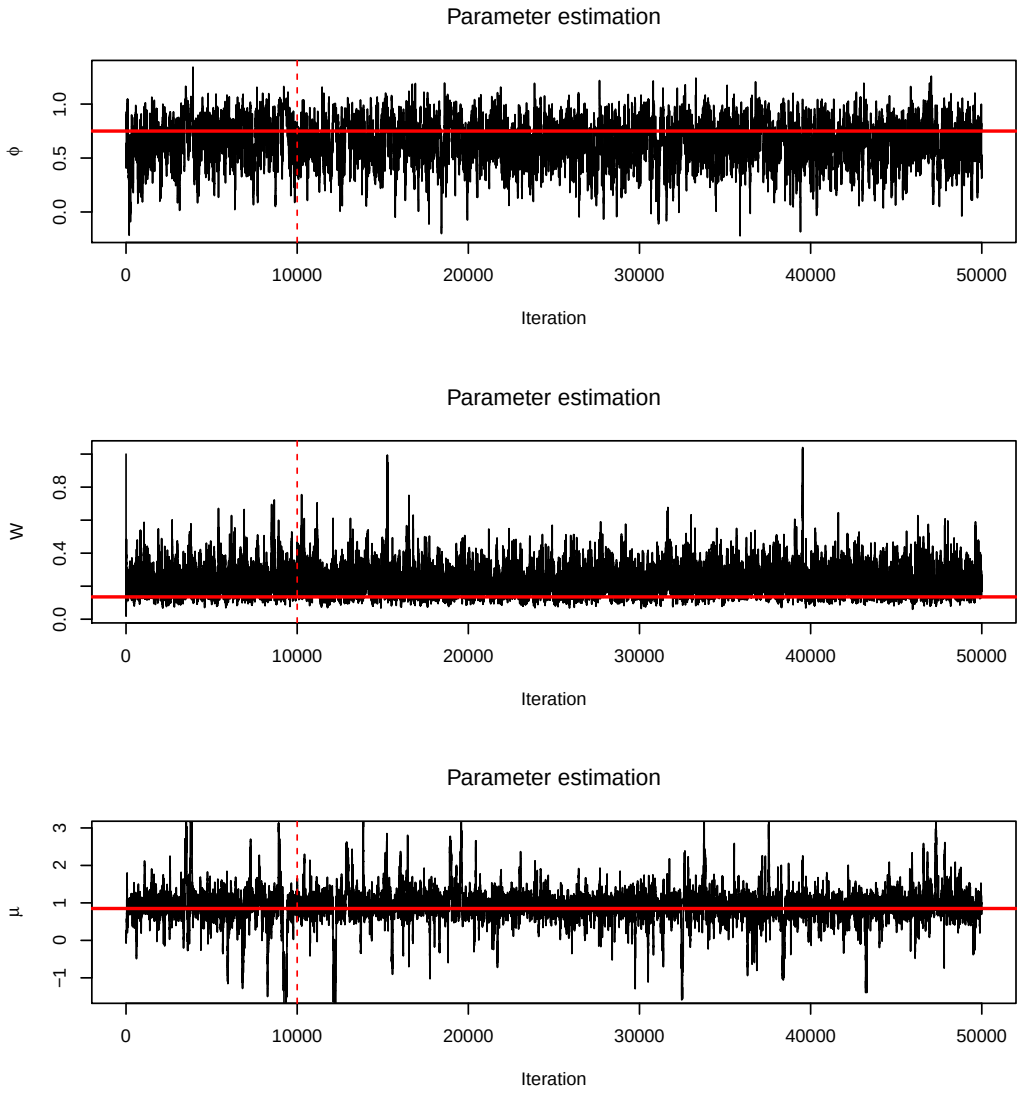

Figure B.1: *The trace plot for the MCMC method of the Poisson model parameters*

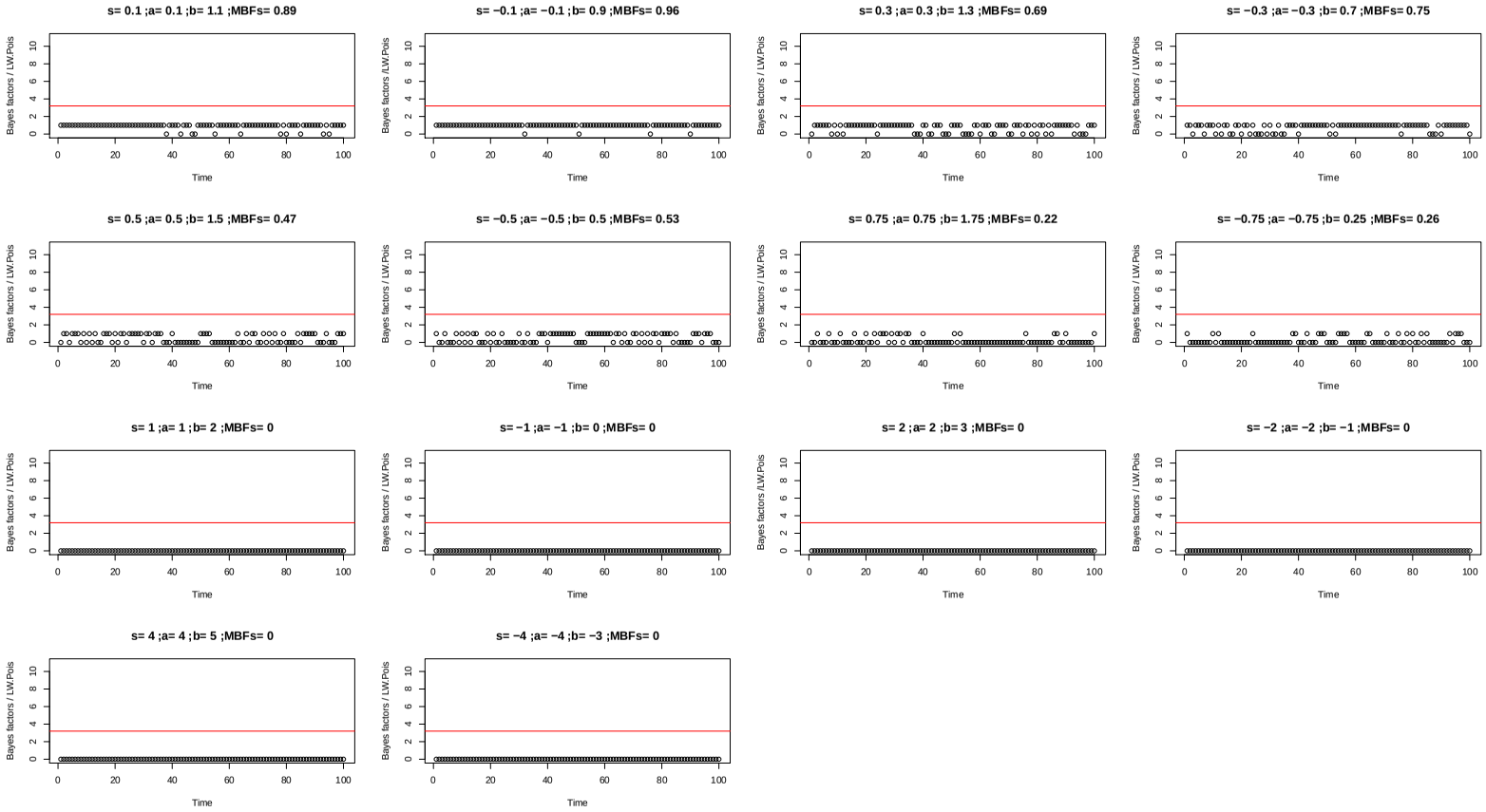

Figure B.2: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on proposed alternative models with different shifts from the mean of the PSR U(0,1).*

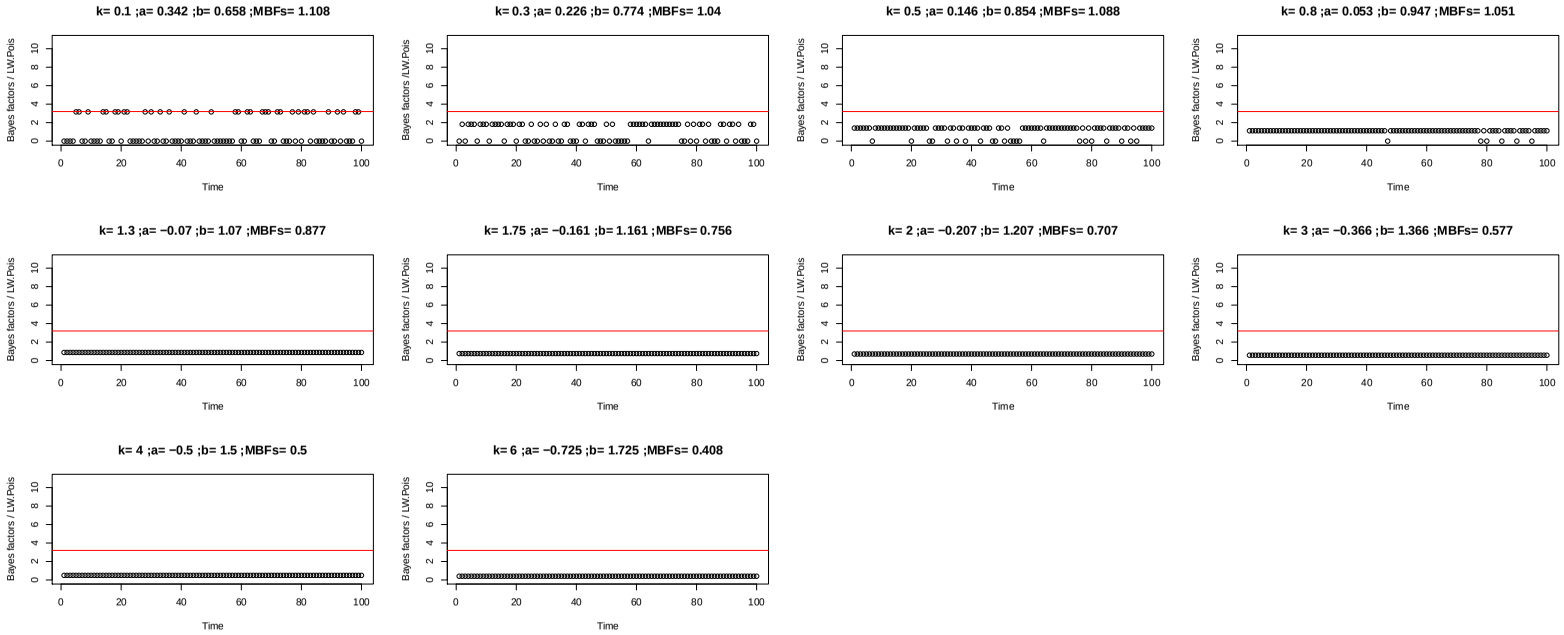

Figure B.3: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on proposed alternative models with different shifts from the variance of the PSR U(0,1).*

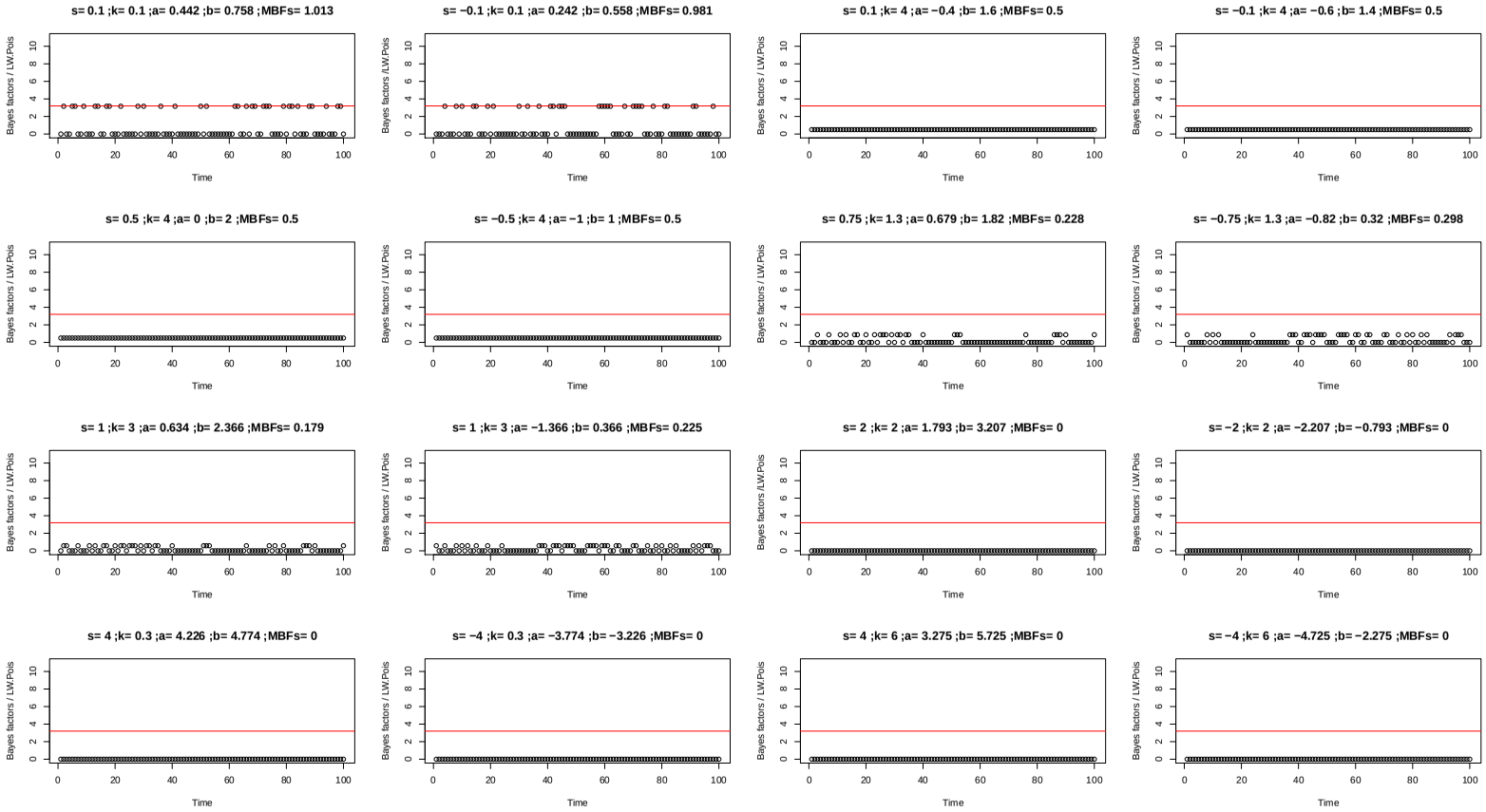

Figure B.4: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on proposed alternative models with different shifts from both the mean and the variance of the PSR U(0,1).*

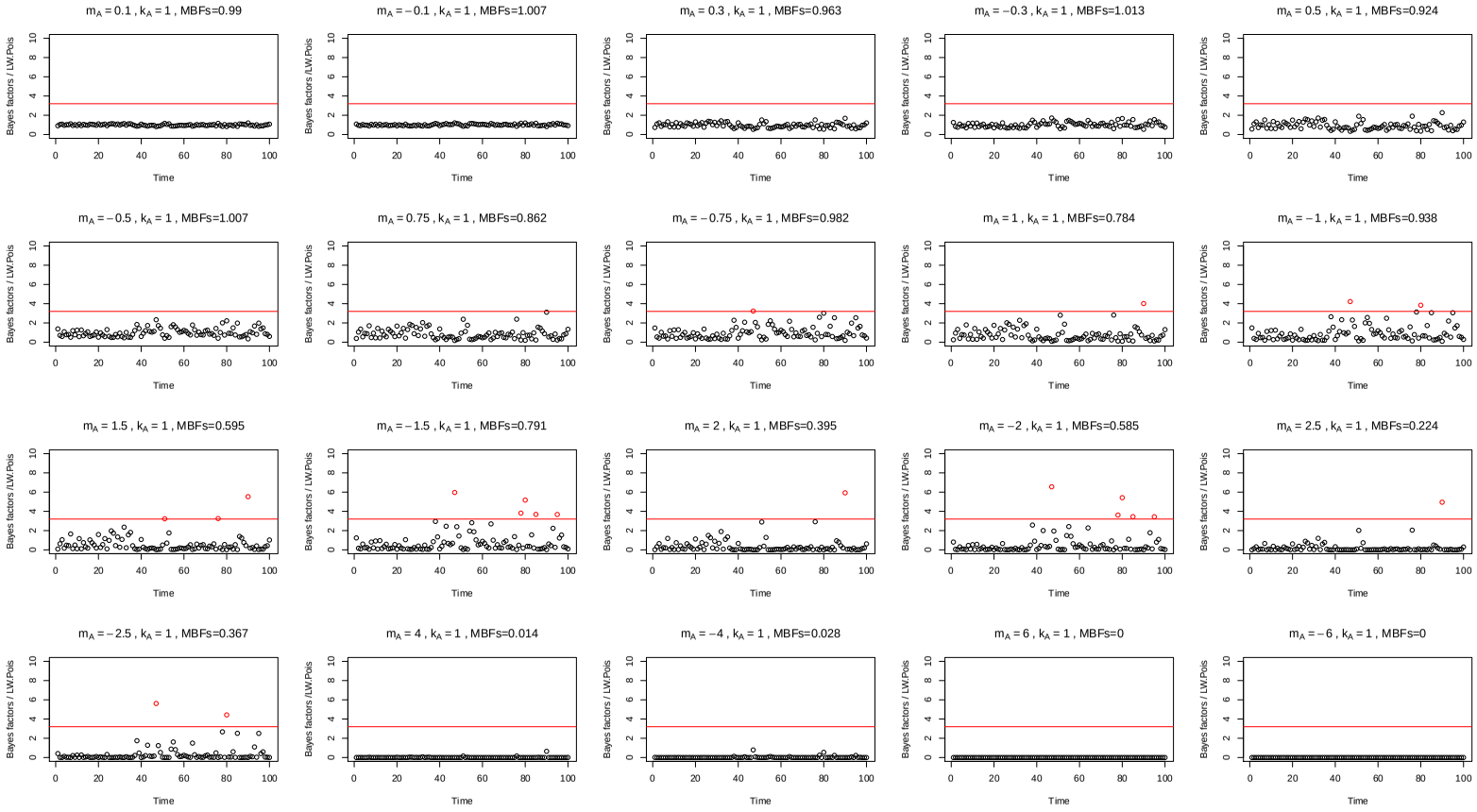

Figure B.5: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on some proposed alternative models with different shifts from the mean of the INTPSR N(0,1).*

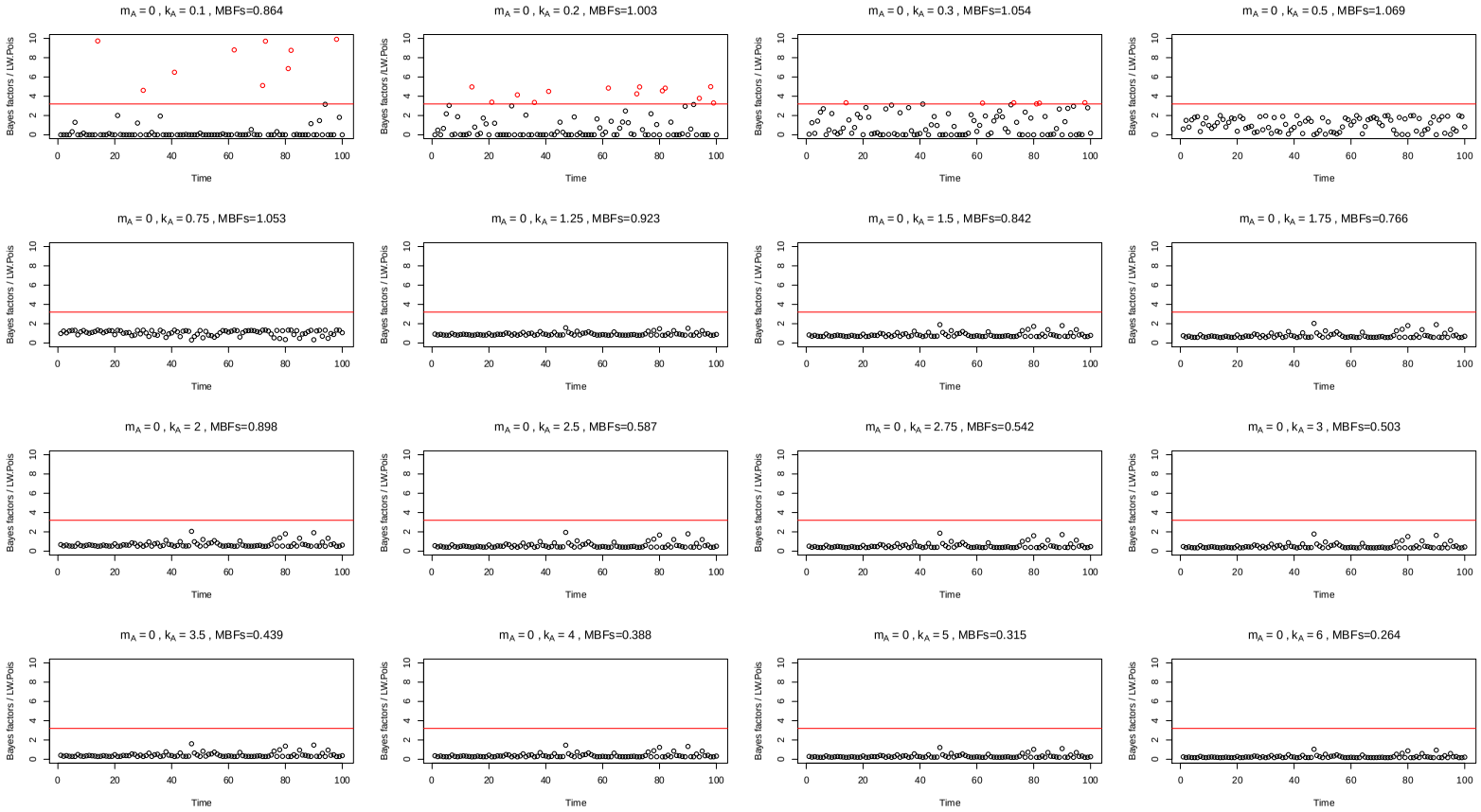

Figure B.6: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to a Poisson simulated data based on some proposed alternative models with different shifts from the variance of the INTPSR N(0,1).*

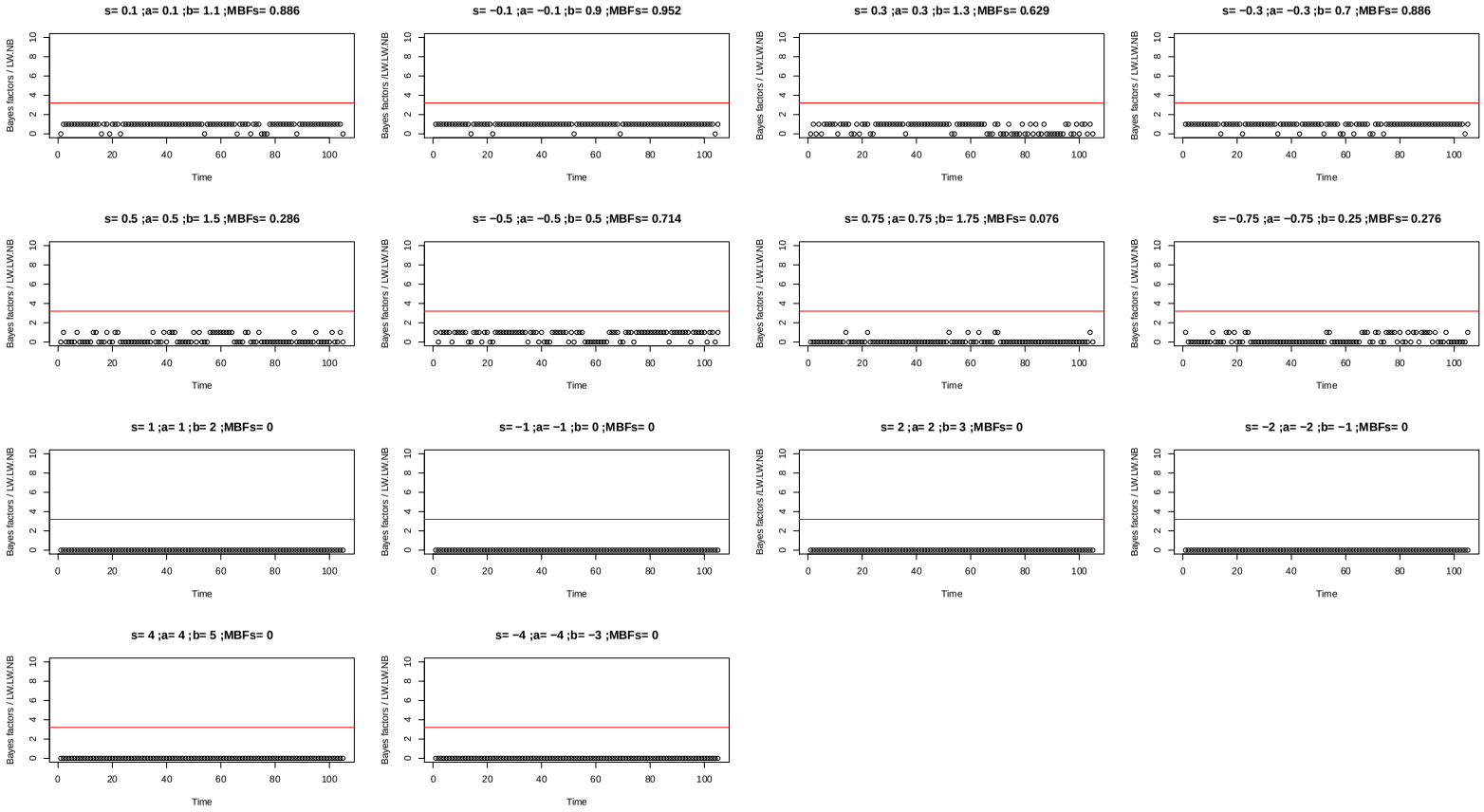

Figure B.7: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the PSR U(0,1).*

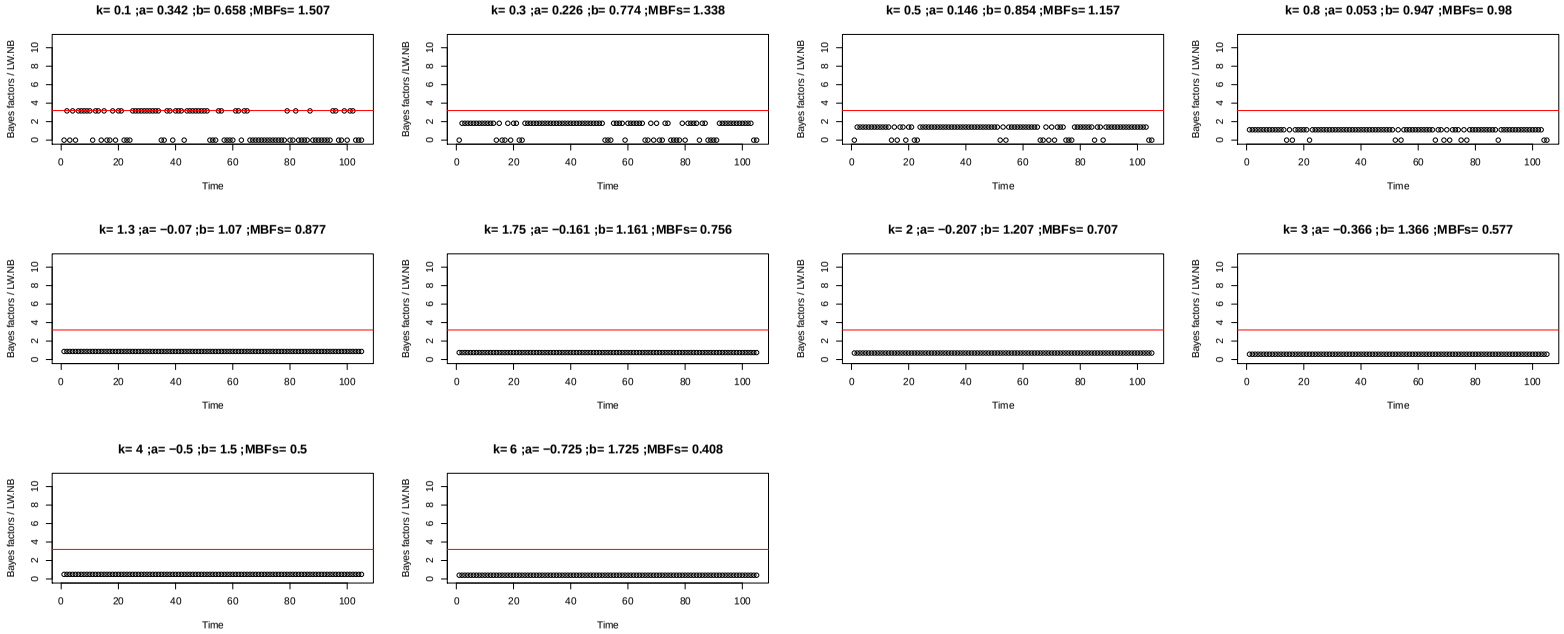

Figure B.8: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the PSR U(0,1).*

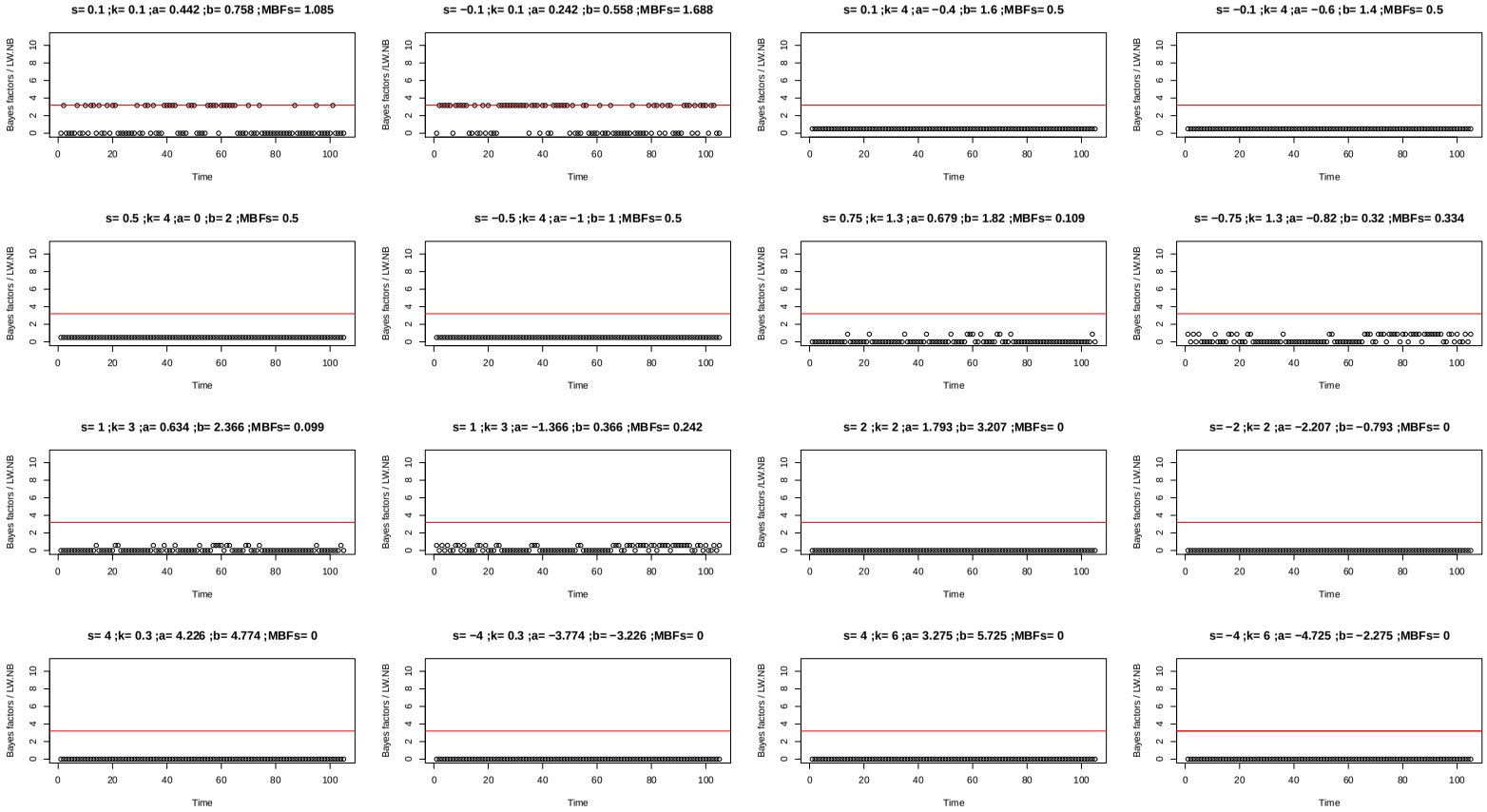

Figure B.9: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from both the mean and the variance of the PSR U(0,1).*

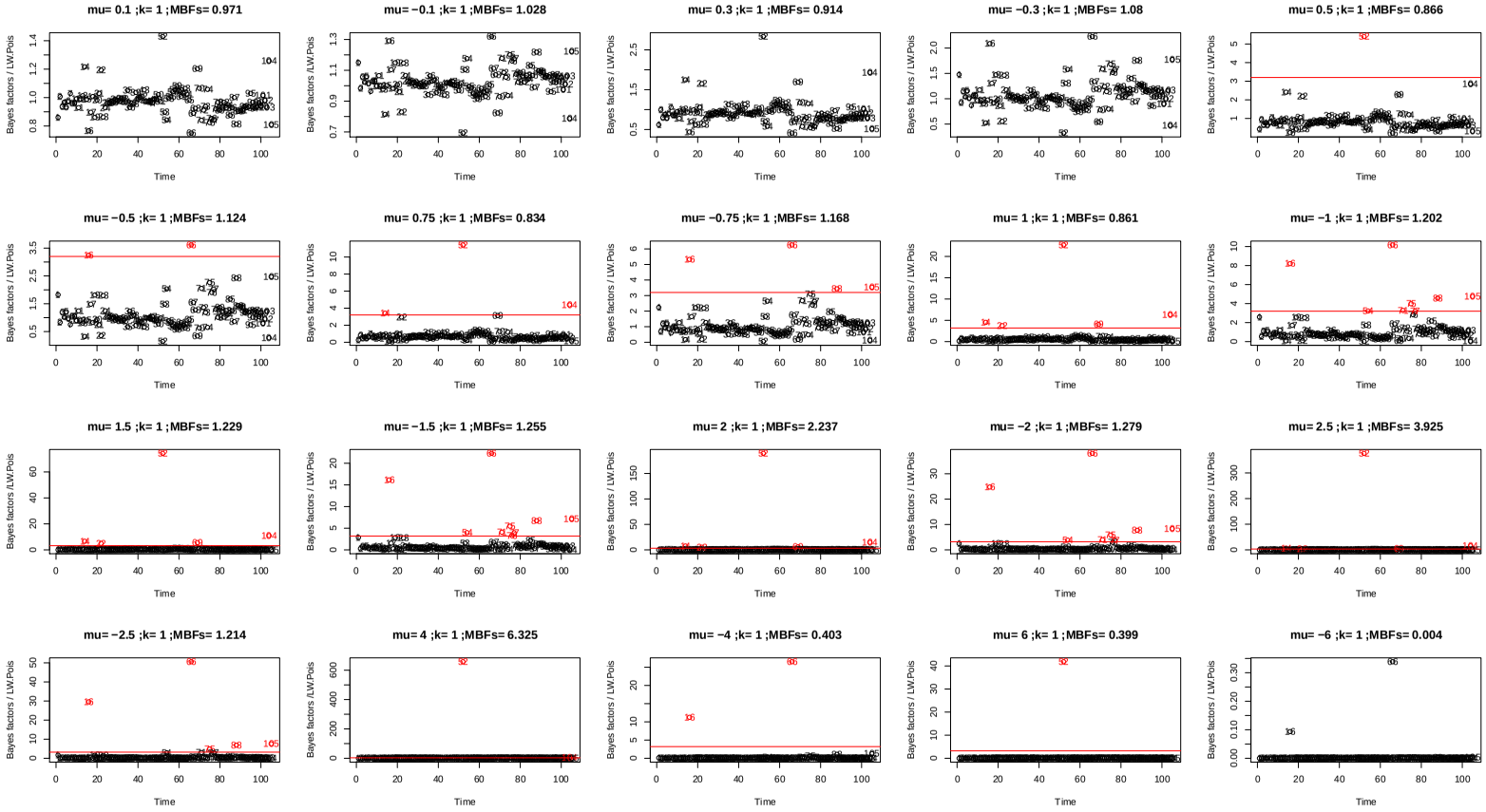

Figure B.10: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the INTPSR N(0,1).*

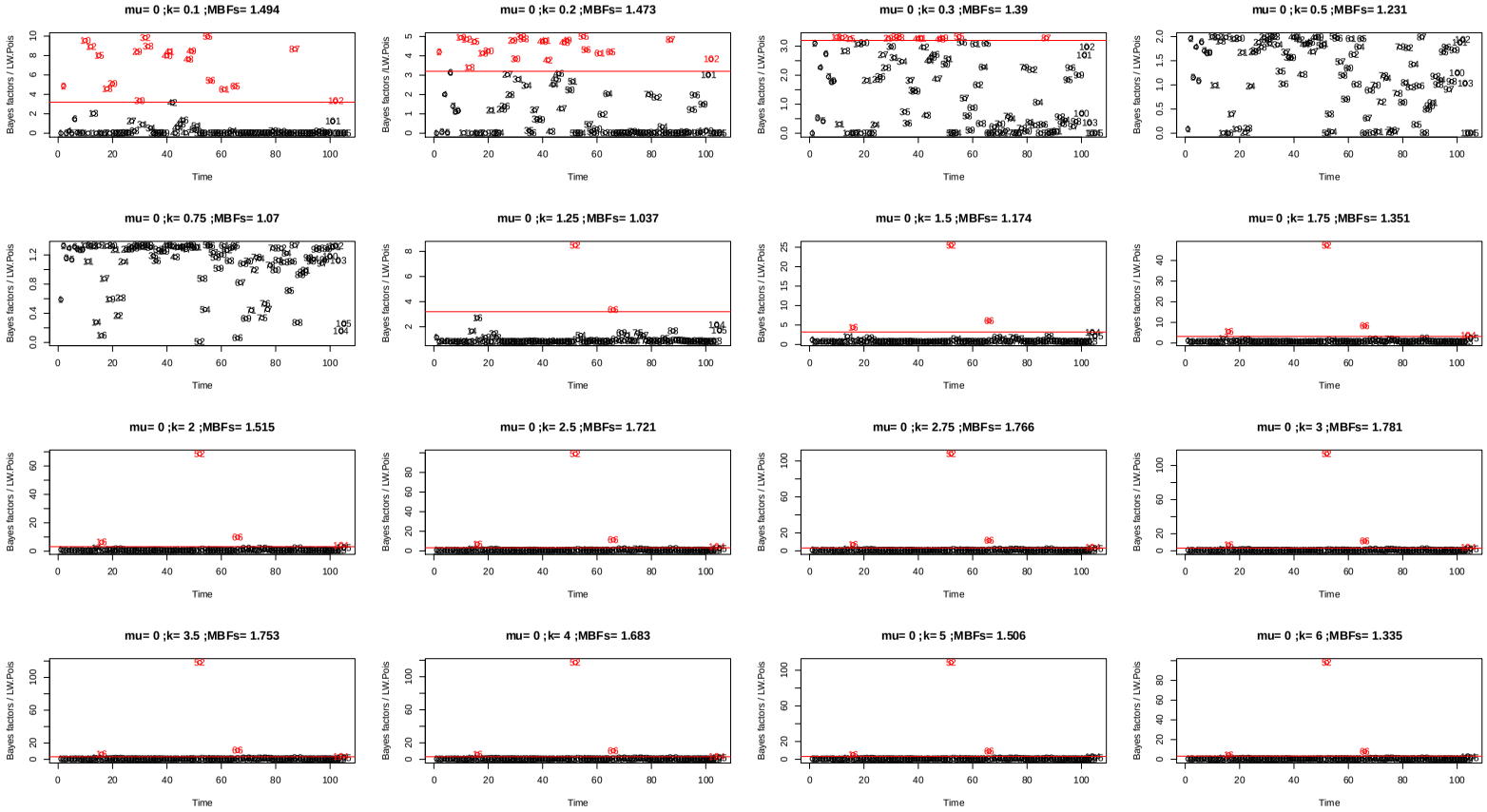

Figure B.11: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the INTPSR N(0,1).*

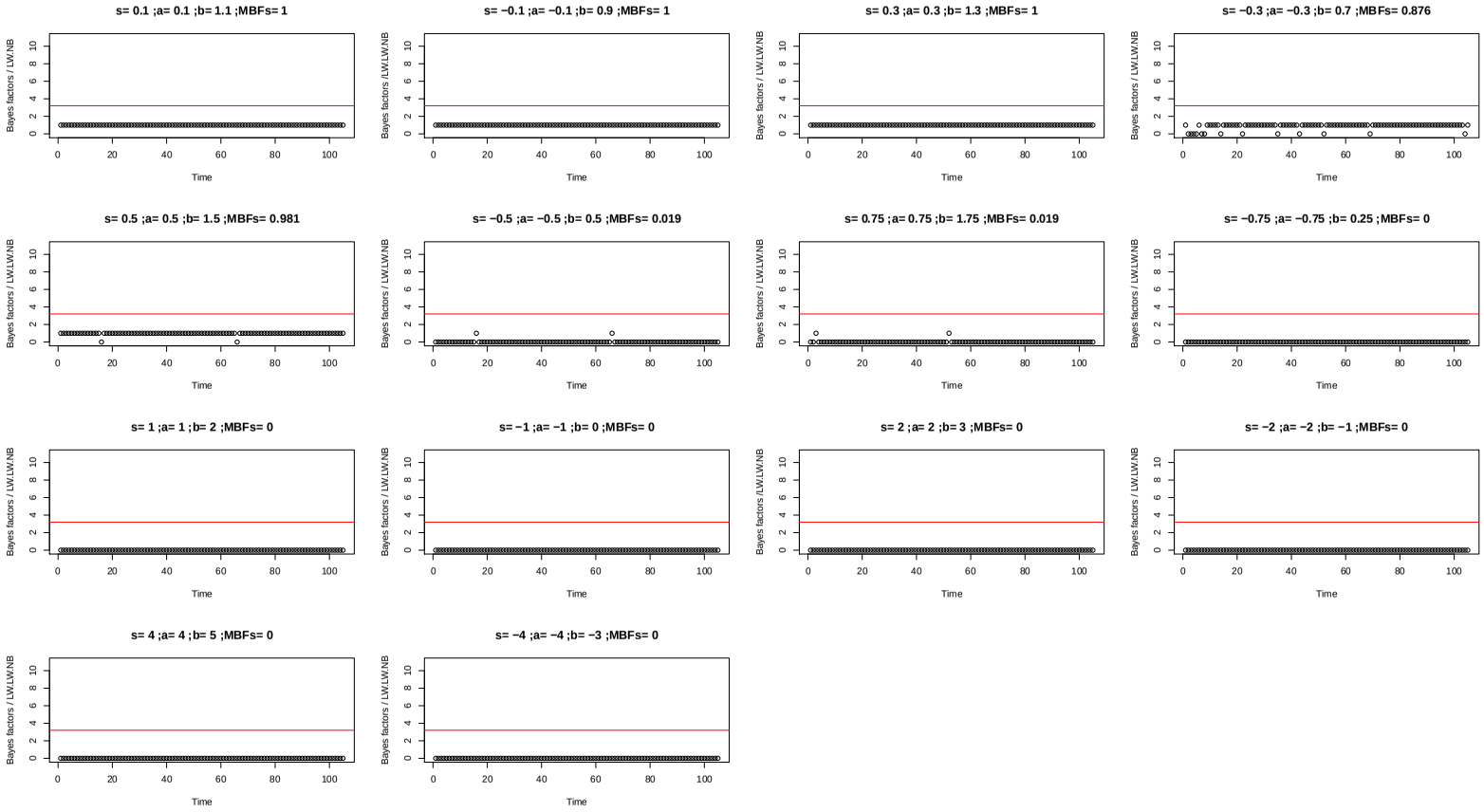

Figure B.12: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the PSR U(0,1).*

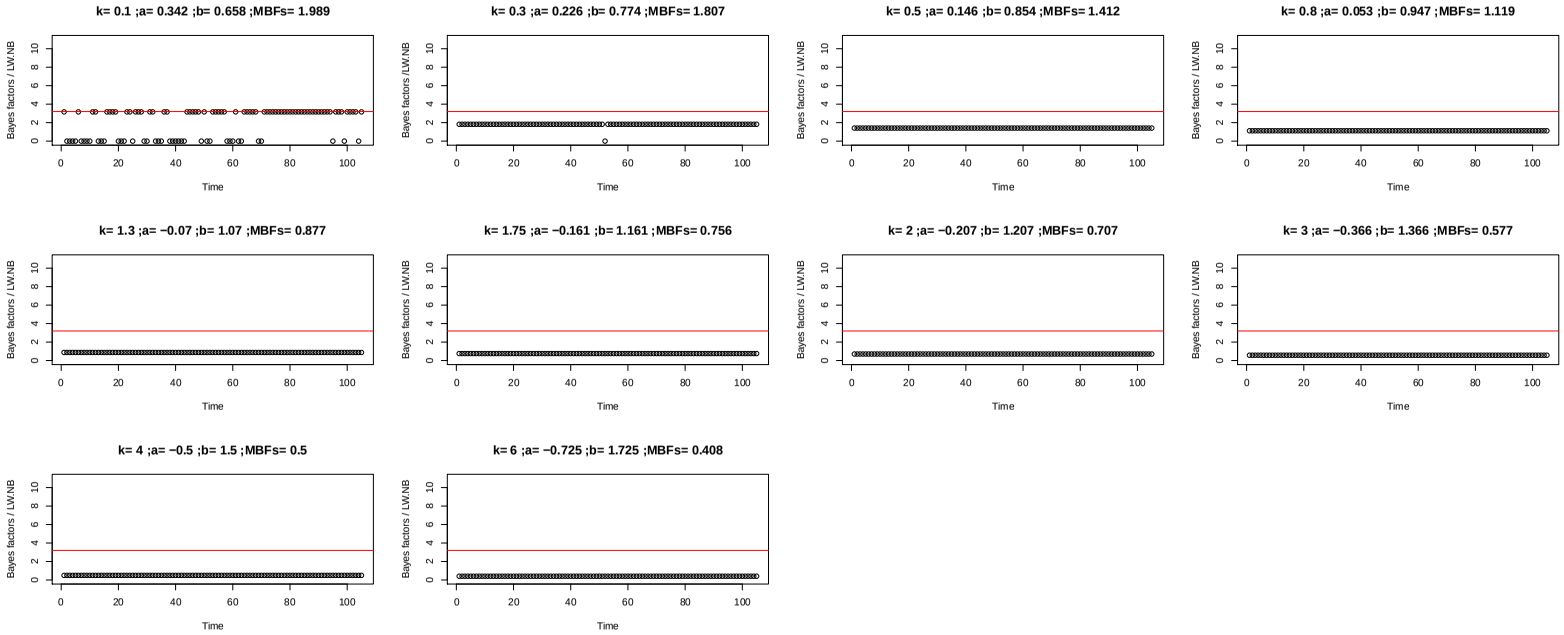

Figure B.13: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the PSR U(0,1).*

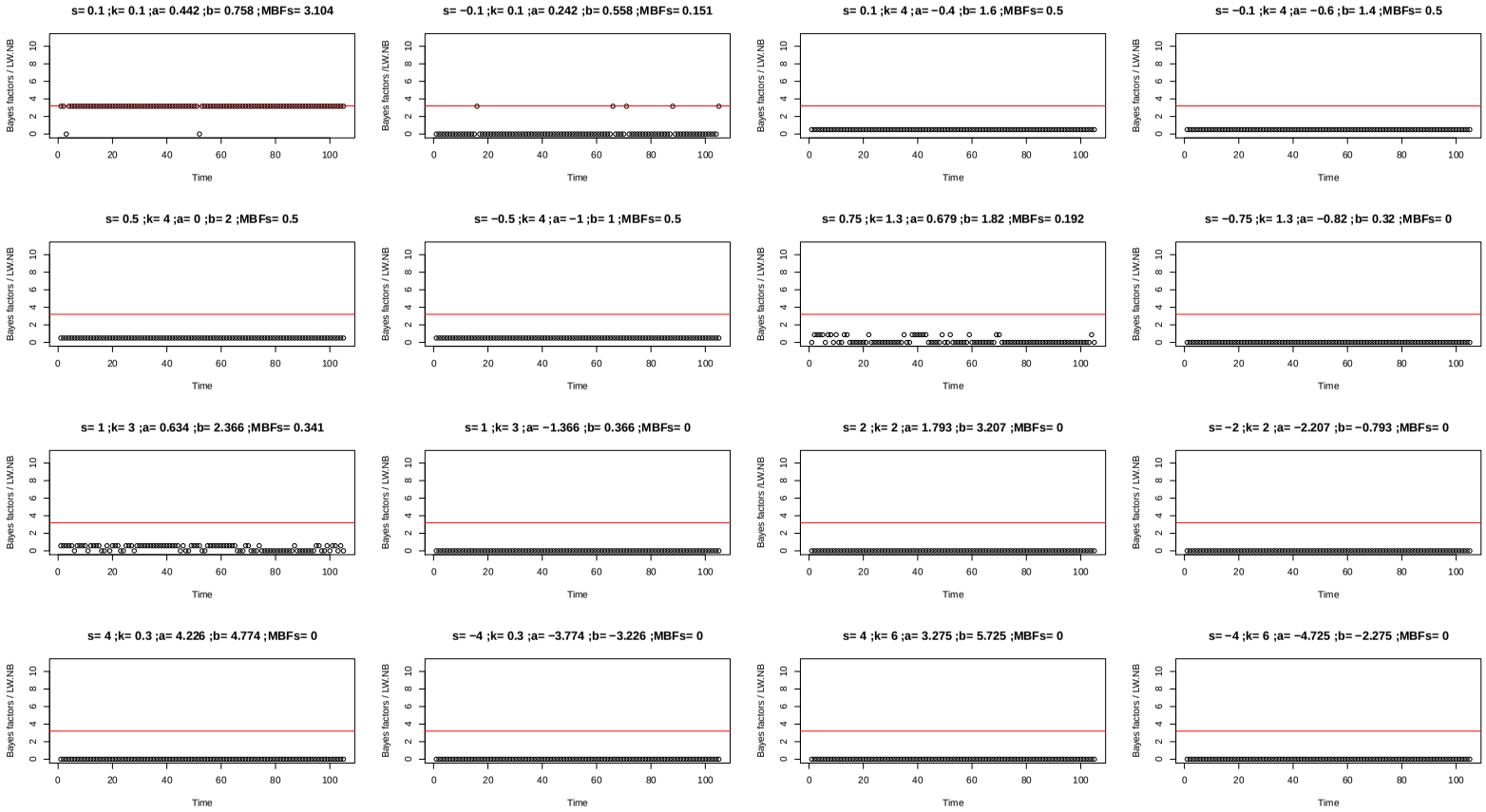

Figure B.14: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from both the mean and the variance of the PSR U(0,1).*

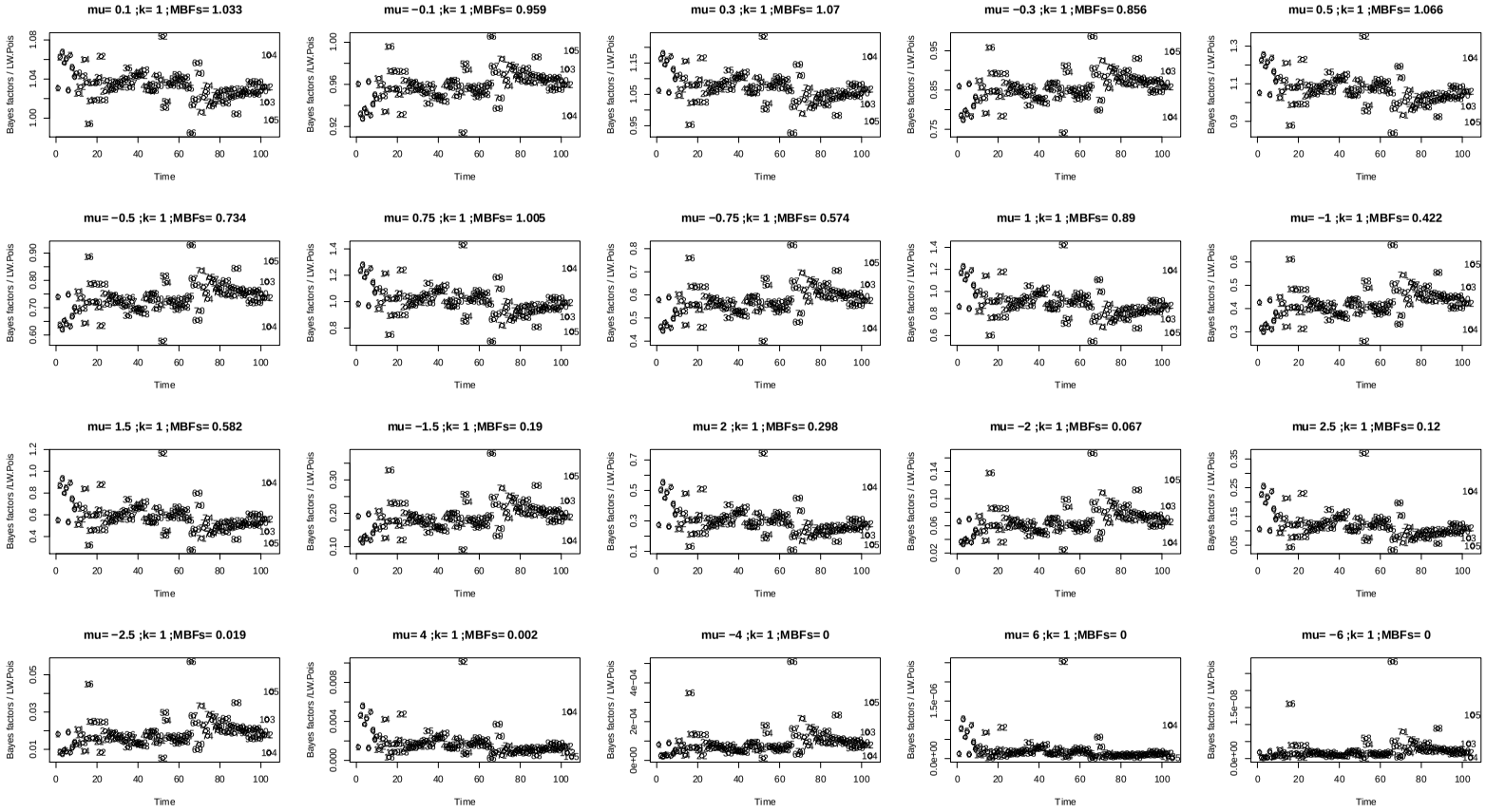

Figure B.15: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the INTPSR N(0,1).*

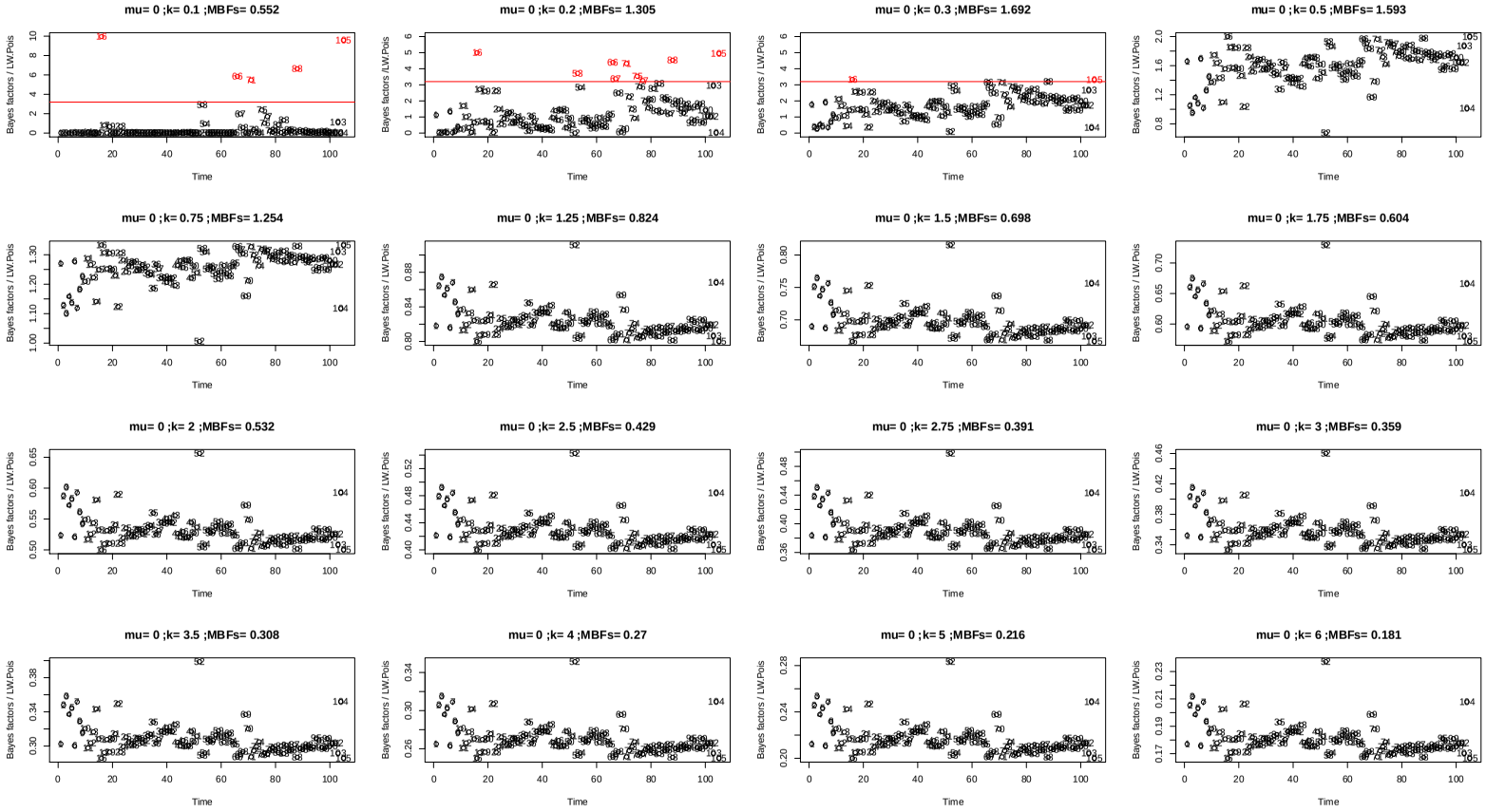

Figure B.16: *The behaviour of the Bayes factors for the LW.NB particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the INTPSR N(0,1).*

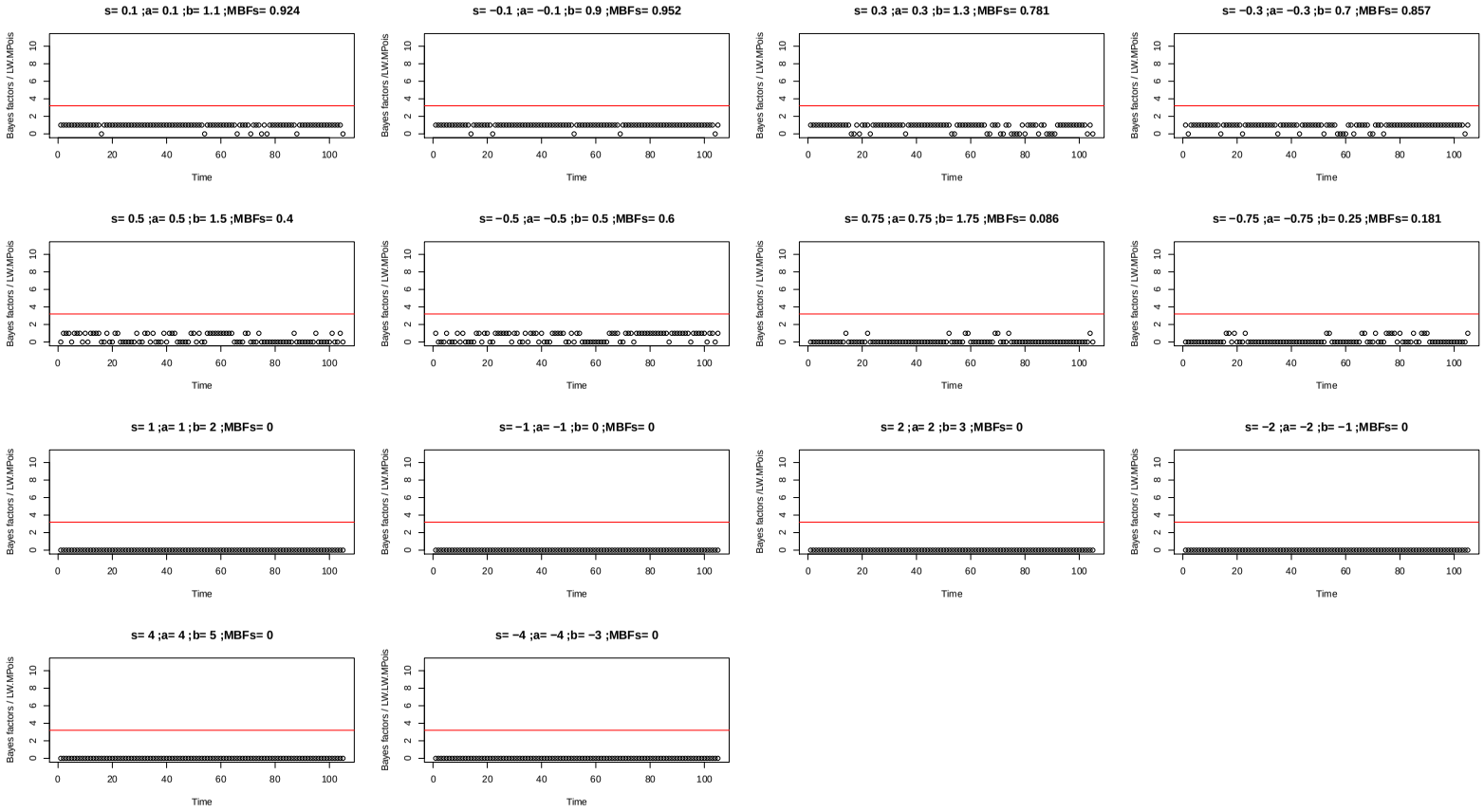

Figure B.17: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the PSR U(0,1).*

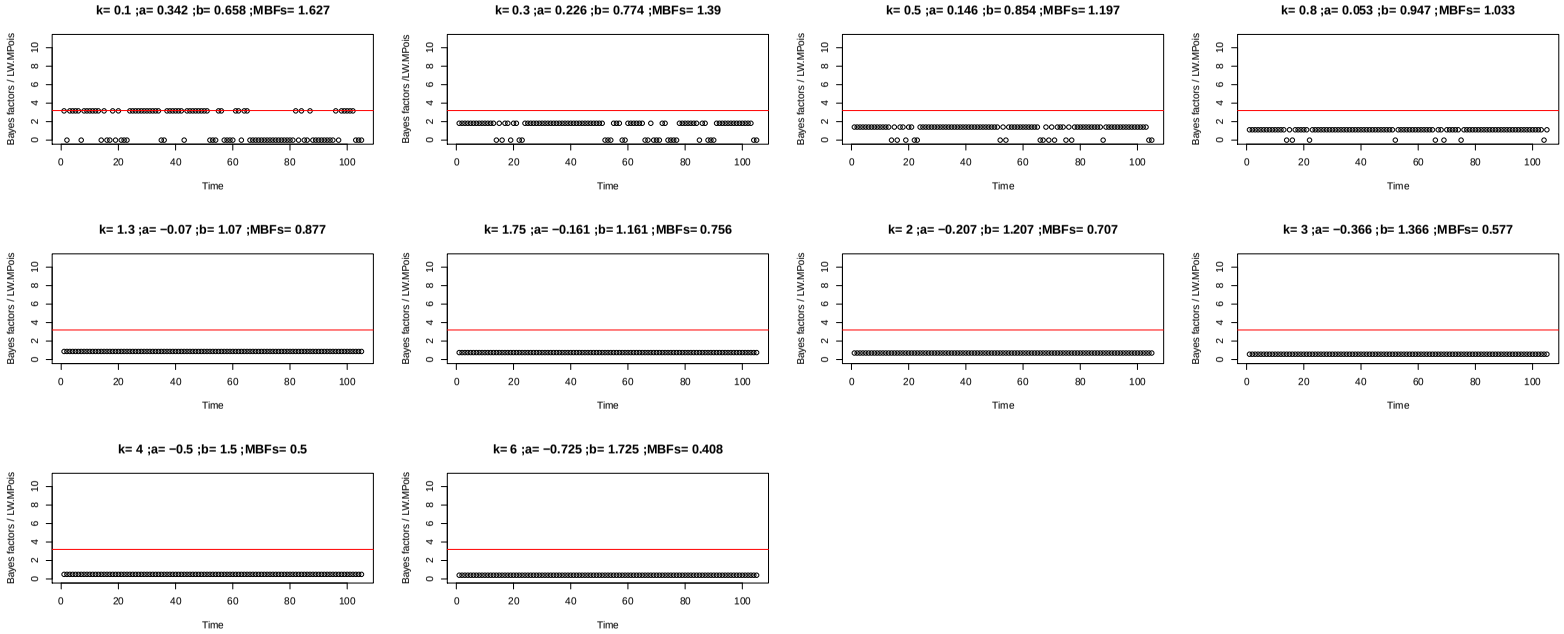

Figure B.18: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the PSR U(0,1).*

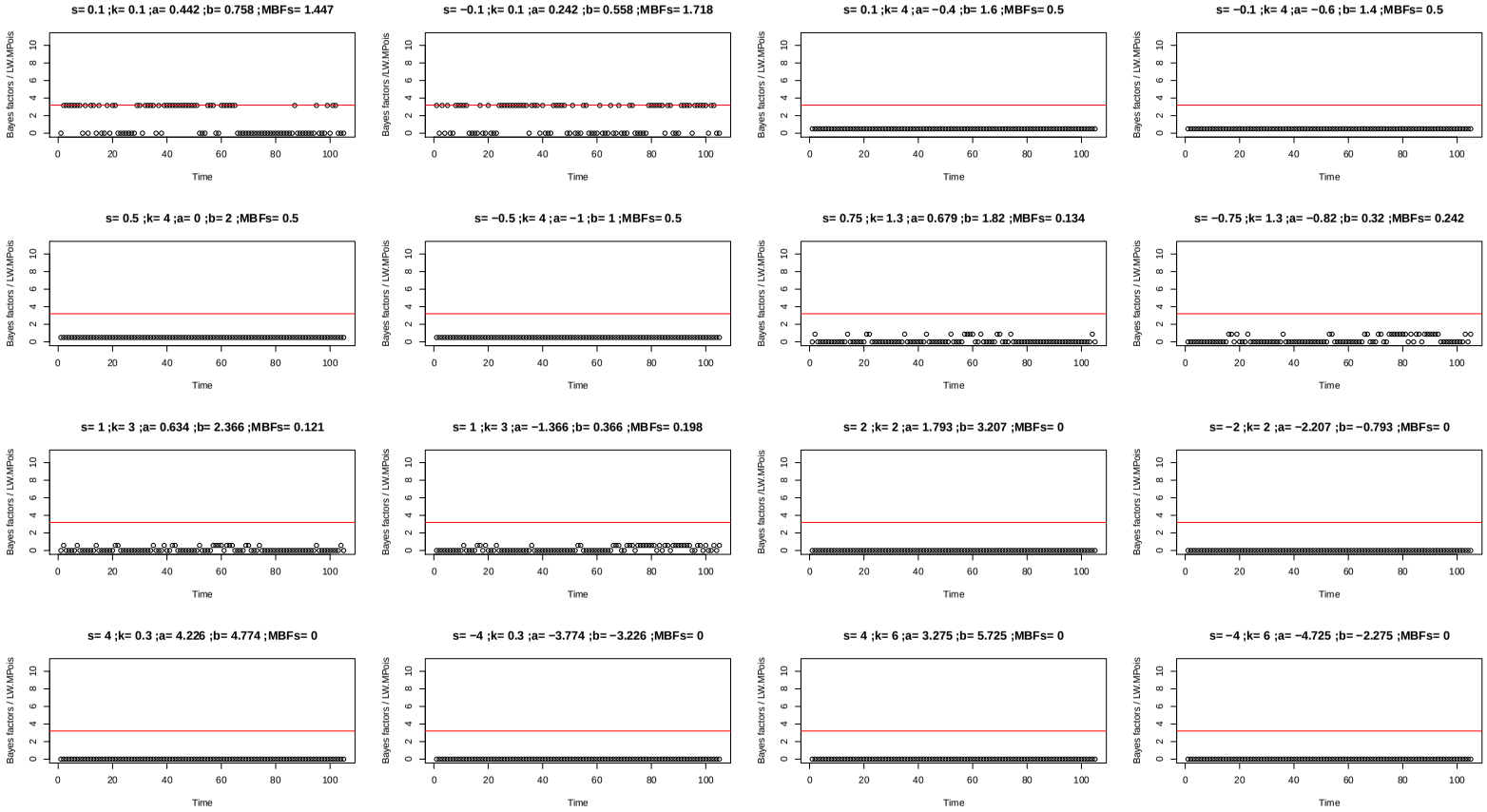

Figure B.19: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from both the mean and the variance of the PSR U(0,1).*

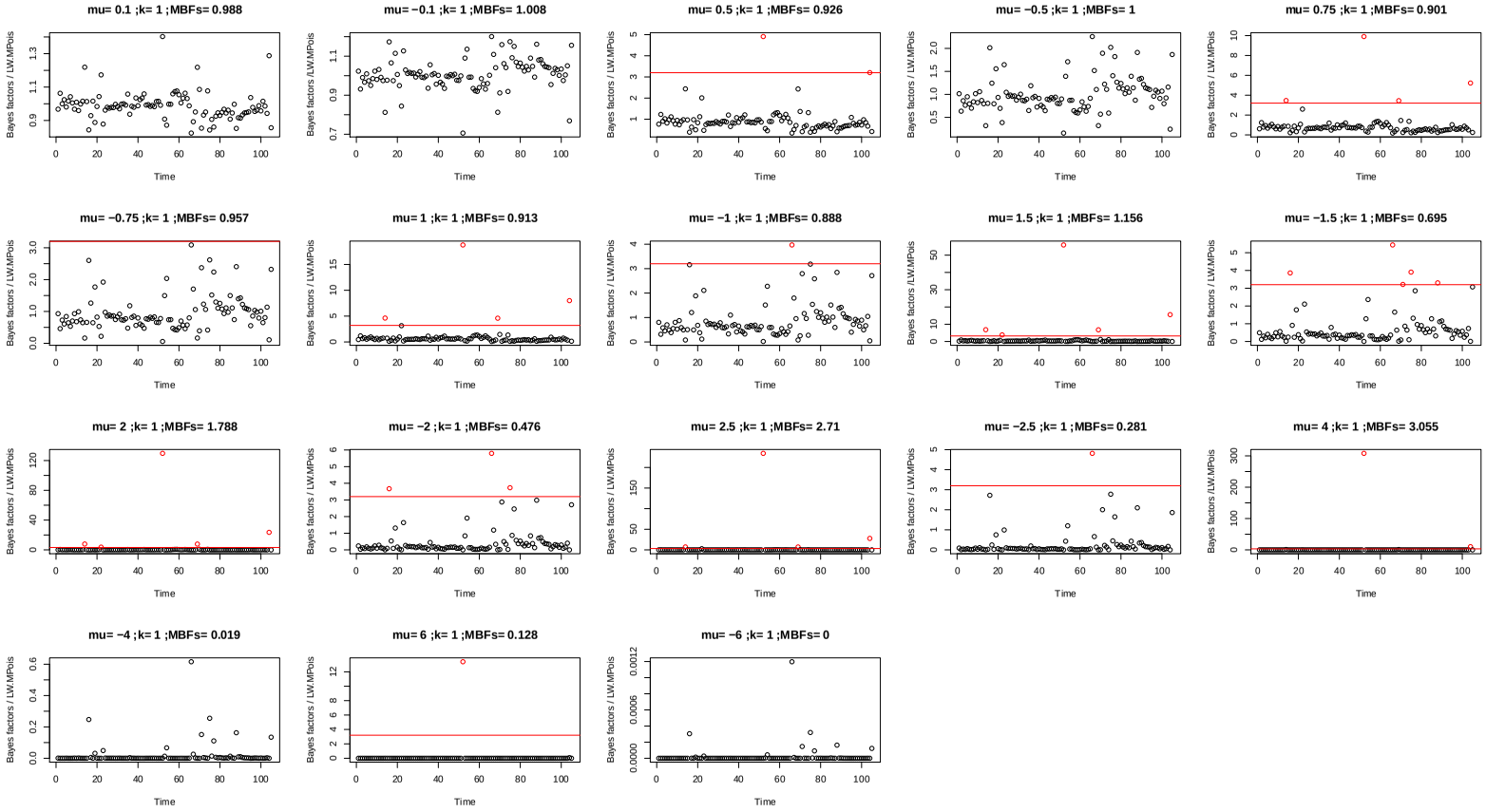

Figure B.20: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the mean of the INTPSR N(0,1).*

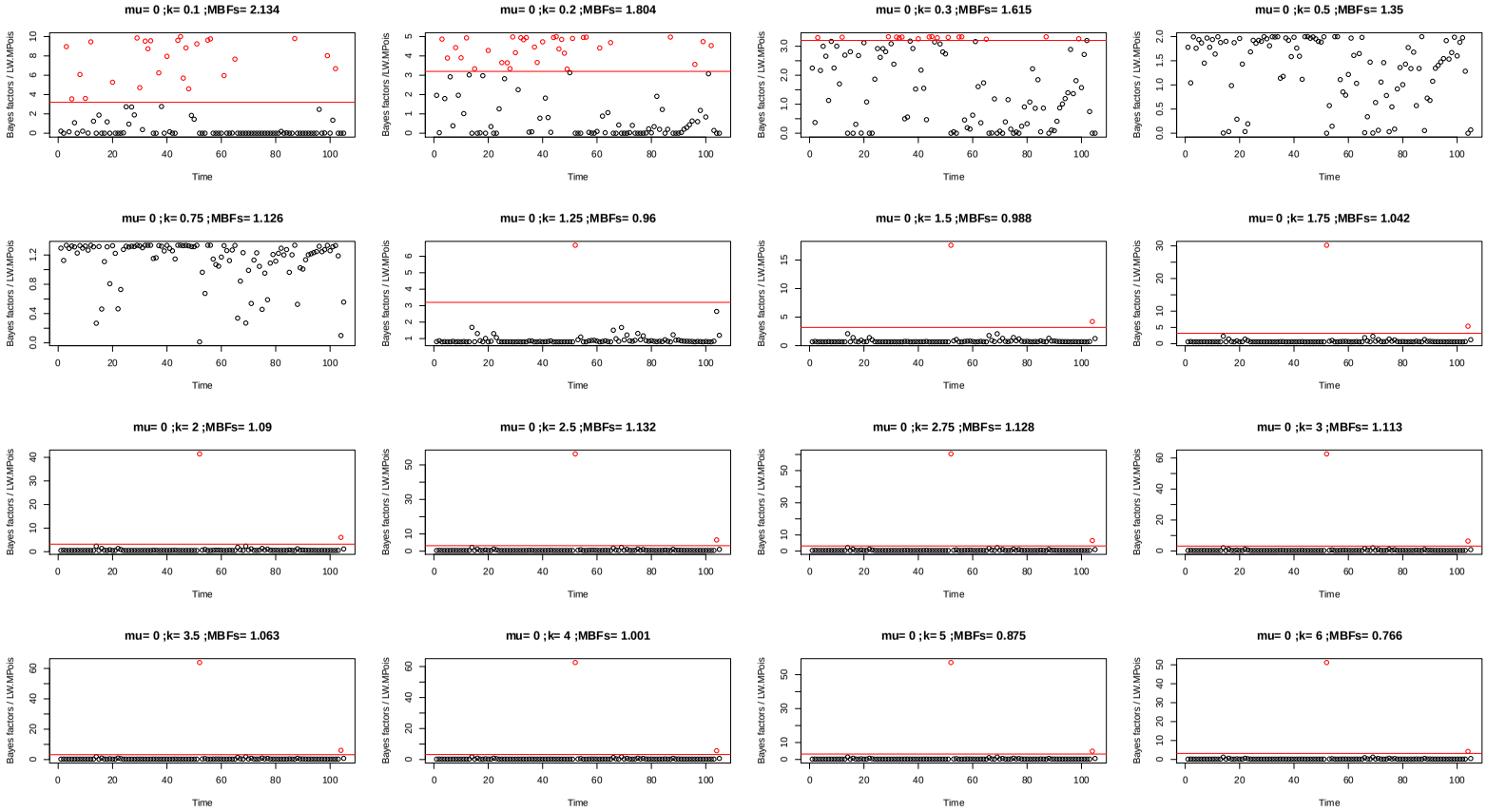

Figure B.21: *The behaviour of the Bayes factors for the LW.MPois particle filter algorithm to an asthma data based on some proposed alternative models with different shifts from the variance of the INTPSR N(0,1).*

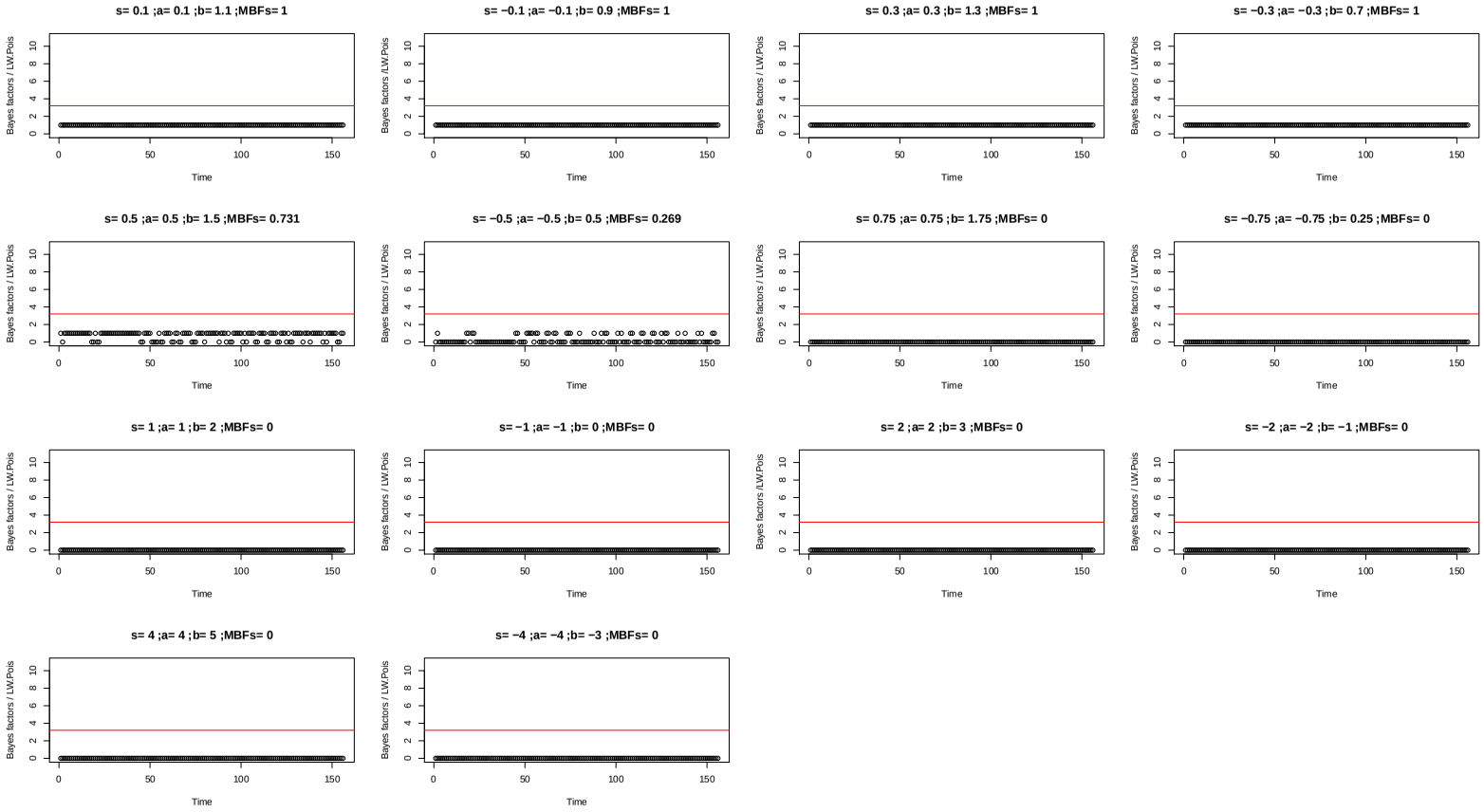

Figure B.22: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to SIDS data based on some proposed alternative models with different shifts from the mean of the PSR U(0,1).*

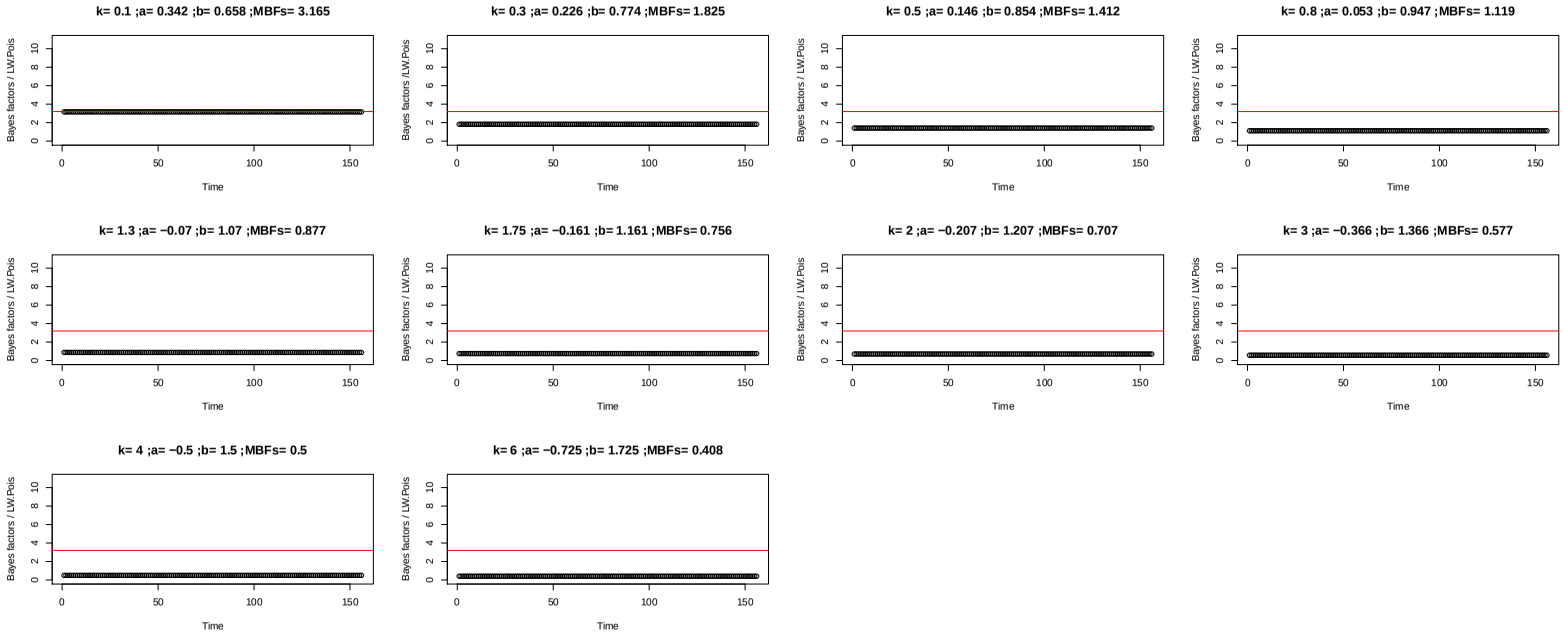

Figure B.23: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to SIDS data based on some proposed alternative models with different shifts from the variance of the PSR U(0,1).*

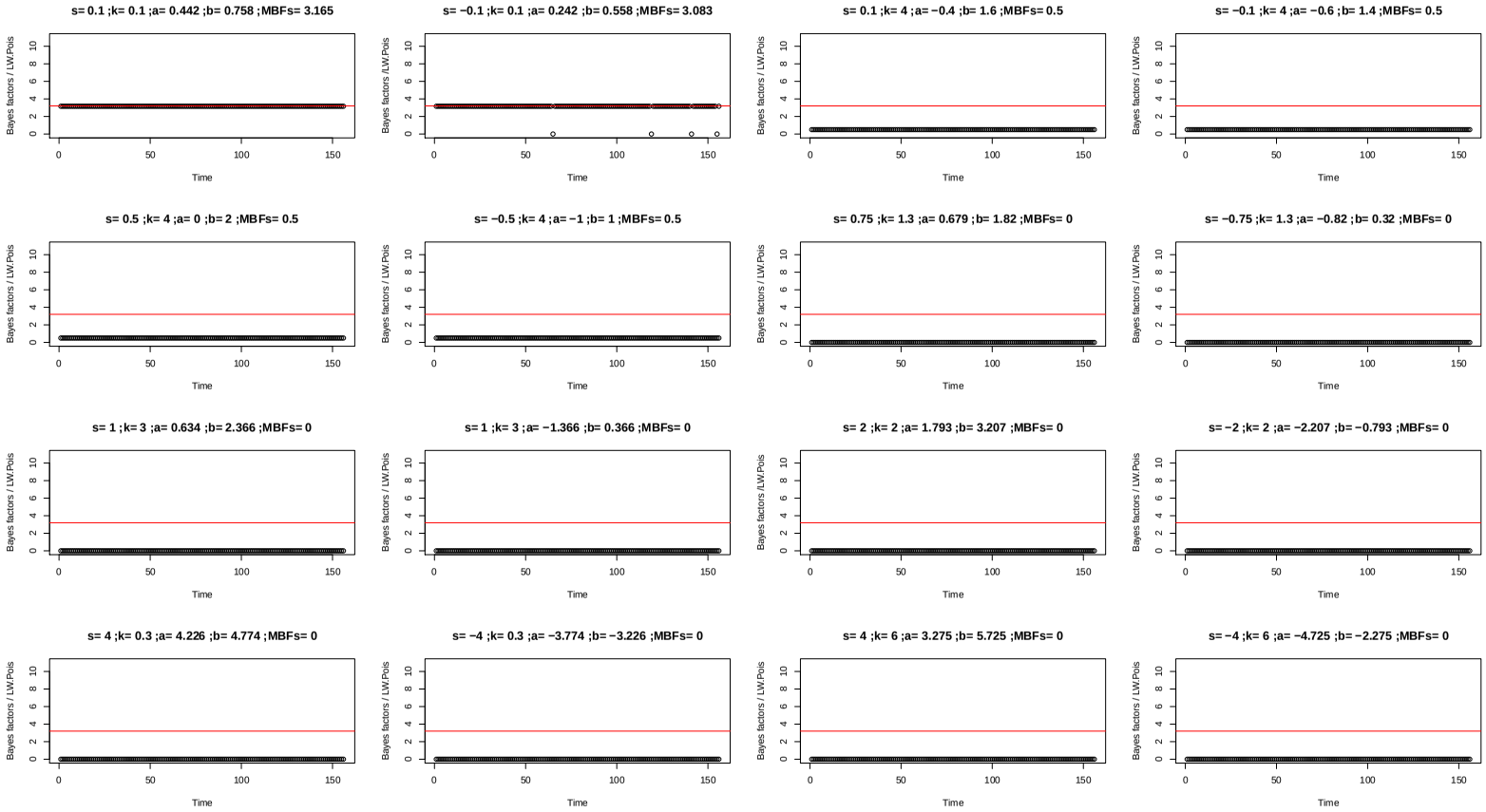

Figure B.24: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to SIDS data based on some proposed alternative models with different shifts from both the mean and the variance of the PSR U(0,1).*

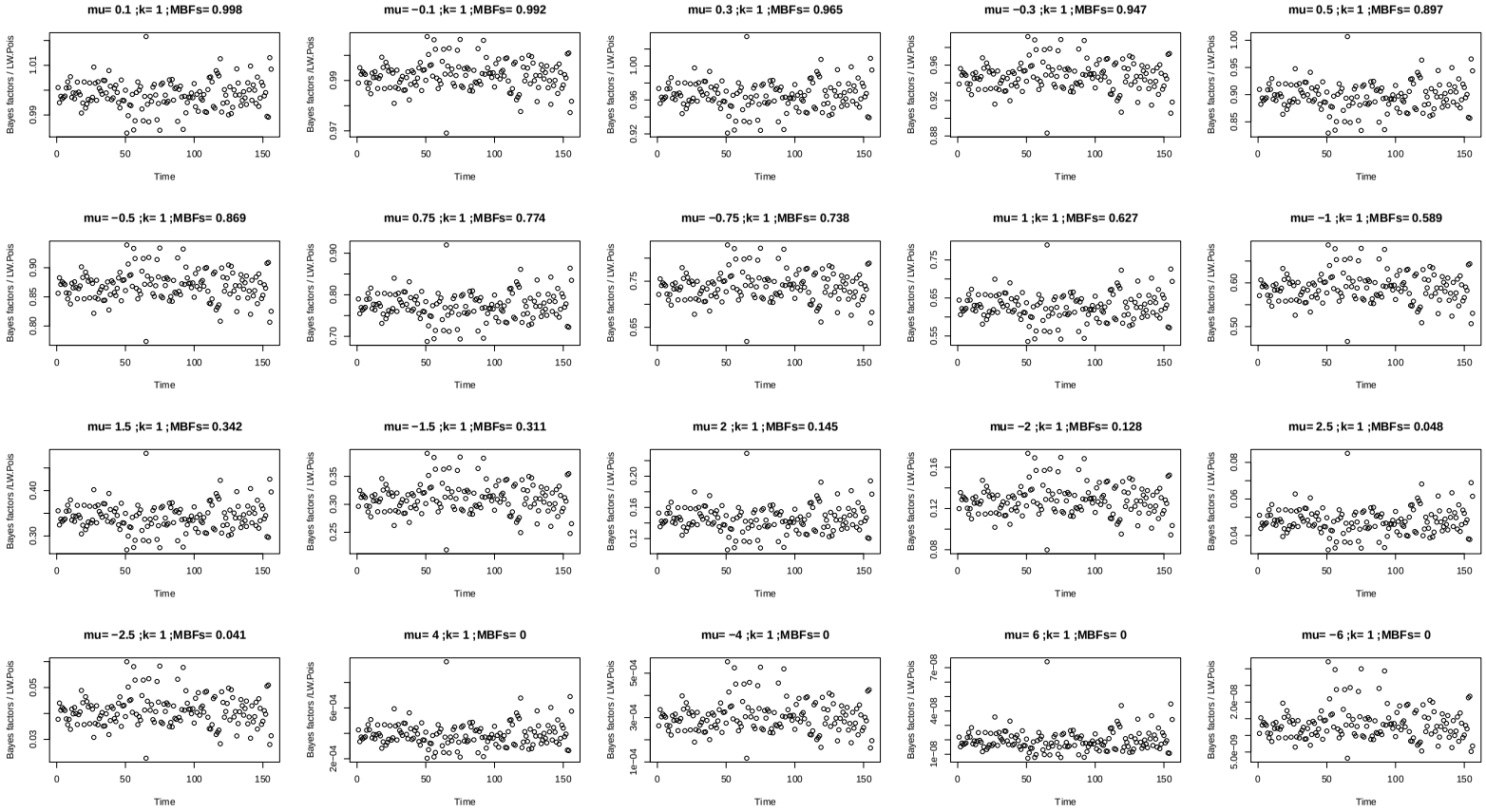

Figure B.25: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to SIDS data based on some proposed alternative models with different shifts from the mean of the INTPSR N(0,1).*

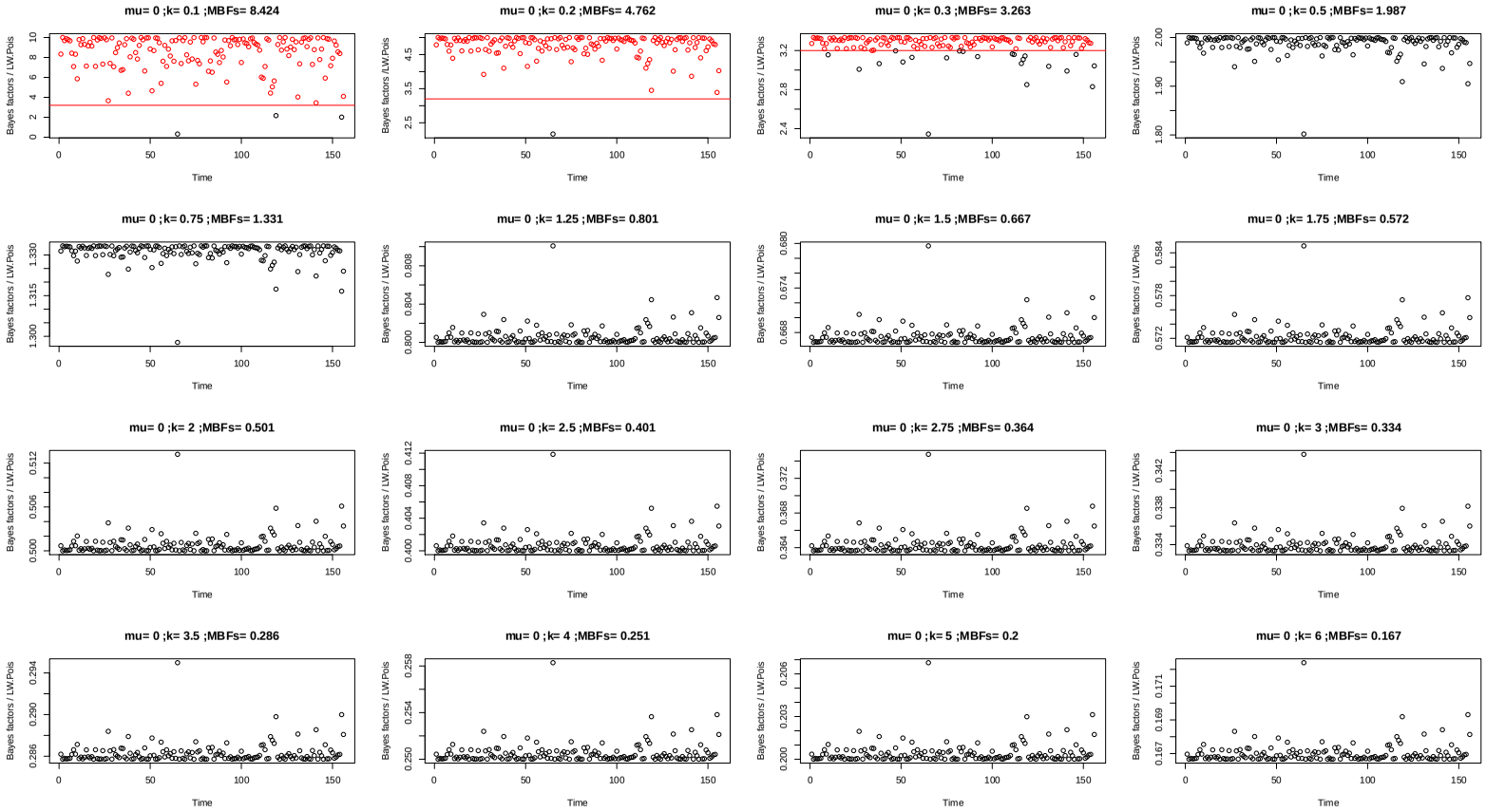

Figure B.26: *The behaviour of the Bayes factors for the LW.Pois particle filter algorithm to SIDS data based on some proposed alternative models with different shifts from the variance of the INTPSR N(0,1).*

Bibliography

Alwan, L. and Roberts, H. (1988). Time series modelling for statistical process control, *Journal of Business and Economic Statistics*, **6**: 86–95.

Amin, R., Reynolds, M. and Bakir, S. (1995). Nonparametric quality control charts based on the sign statistic, *Communications in Statistics: Theory and Methods*, **24 (6)**: 1597–1623.

Anderson, B. and Moore, J. (1979). *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall.

Anderson, O. (1976). *Time Series Analysis and Forecasting*, London: Butterworths.

Andrieu, C. and Doucet, A. (2002). Particle filtering for partially observed Gaussian state space models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64 (4)**: 827–836.

Andrieu, C., Doucet, A. and Tadic, V. (2005). On-line parameter estimation in general state-space models, *In Proc. IEEE Conf. Decision and Control/European Control Conf*, pp. 332–337.

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

Bakir, S. T. (2012). A nonparametric shewhart-type quality control chart for monitoring broad changes in a process distribution, *International Journal of Quality, Statistics, and Reliability*, **2012 (1)**: 1–10.

Bisgaard, S. and Kulahci, M. (2005). Quality quandaries: The effect of autocorrelation on statistical process control procedures, *Quality Engineering*, **17**: 481–489.

Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day; San Francisco.

Box, G. and Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association*, **65 (322)**: 1509–1526.

Campbell, M. J. (1994). Time series regression for counts: An investigation into the relationship between sudden infant death syndrome and environmental temperature, *Journal of the Royal Statistical Society, Series A: Statistics in Society*, **157**: 191–208.

Campbell, M. J., Holgate, S. T. and Johnston, S. L. (1997). Seasonality of asthma deaths, *Br. Med. J*, **315**: 1012.

Castagliola, P. and Tsung, F. (2005). Autocorrelated spc for non-normal situations, *Qual. Reliab. Engng. Int.*, **21**: 131–161.

Chakraborti, S. and Graham, M. (2007). Nonparametric control charts, *Encyclopedia of Statistics in Quality and Reliability*, **1**: 415–429.

Chakraborti, S., Human, S. and Graham, M. (2011). Nonparametric (distribution-free) quality control charts, *In Handbook of Methods and Appli-*

cations of Statistics: Engineering, Quality Control, and Physical Sciences., pp. 298–329.

Chakraborti, S., der Laan, P. V. and Bakir, S. (2001). Nonparametric control charts: An overview and some results, *Journal of Quality Technology*, **33 (3)**: 304–315.

Chakraborti, S., der Laan, P. V. and de Wiel, M. V. (2004). A class of distribution-free control charts, *Journal of the Royal Statistical Society. Series C: Applied Statistics*, **53 (3)**: 443–462.

Christensen, A., Christensen, S., Overvad, K., Rasmussen, L. and Dethlefsen, C. (2012). Modeling gradually changing seasonal variation in count data using state space models: a cohort study of hospitalization rates of stroke in atrial fibrillation patients in Denmark from 1977 to 2011, *BMC Medical Research Methodology*, **12:174**: doi:10.1186/1471–2288–12–174.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics*, **65**: 1254–1261.

Dawid, A. (1984). Statistical theory: The prequential approach, *Journal of the Royal Statistical Society. Series A*, **147**: 178–292.

DelMoral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers, *Journal of the Royal Statistical Society: Series B*, **68**: 411–436.

Dempster, A., Laird, N. and Rubin, N. (1977). Maximum likelihood estimation from incoplete data via the em algorithm(with discussion), *Journal of the Royal Statistical Society. Series B*, **39**: 1–38.

Diebold, F., Gunther, T. and Tay, A. (1998). Evaluating density forecasts with

applications to financial risk management, *International Economic Review*, **39**: 863–883.

Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practise*, Spring-Verlag; New York.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, **10**: 197–208.

Doucet, A. and Johansen, A. (2009). A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of Nonlinear Filtering*, **12**: 656–704.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods*, Oxford University Press.

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62 (1)**: 3–56.

Fahrmeir, L. (1992). Conditional mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models, *Journal of the American Statistical Association*, **87**: 501–509.

Fearnhead, P. and Clifford, P. (2003). On-line inference for hidden Markov models via particle filters, *Journal of the Royal Statistical Society: Series B*, **65**: 887–899.

Frühwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models, *Environmental and Ecological Statistics*, **3 (4)**: 291–309.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models, *Biometrika*, **85 (1)**: 215–227.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, CRC Press.

Gargallo, P. and Salvador, M. (2003). Monitoring residual autocorrelations in dynamic linear models, *Communications in Statistics: Simulation and Computation*, **32**: 1079–1104.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Transactions on pattern analysis and machine intelligence*, **(6)**: 721–741.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Journal of the Econometric Society*, **57**: 1317–1339.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1993). Markov chain monte carlo in practice.

Gordon, N., Salmond, D. and Smith, A. (1993). Novel approach to nonlinear-nonGaussian Bayesian state estimation, *IEE Proceedings F-Radar and Signal Processing*, **140 (2)**: 107–113.

Grech, V., Balzan, M. and Distefano, S. (2004). Paediatric wheezy admissions at and around school holiday periods, *Malta Med J*, **16**: 23–26.

Handschin, J. and Mayne, D. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering, *International Journal of Control*, **9**: 547–559.

Handschin, J. and Mayne, D. (1970). Monte Carlo techniques for prediction and filtering of non-linear stochastic process, *Automatic*, **6**: 555–563.

Harrison, P. and Stevens, C. (1971). A Bayesian approach to short-term forecasting, *Operations Research Quarterly*, **22**: 341–362.

Harrison, P. and Stevens, C. (1976). Bayesian forecasting(with discussion), *Journal of the Royal Statistical Society. Series B (Methodological)*, **38**: 205–247.

Harvey, A. (1989). *Forecasting Structrual Time Series Models and Kalman Filter*, Cambridge university press.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika*, **57 (1)**: 97–109.

Hol, J. D., Schon, T. B. and Gustafsson, F. (2006). On resampling algorithms for particle filters, in *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE*, pp. 79–82, IEEE.

Hulin, W. (2000). Modelling the hiv epidemic: A state-space approach, *Mathematical and Computer Modelling*, **32**: 197–215.

Jazwinski, A. H. (1973). *Stochastic processes and filtering theory*, Academic Press.

Jeffreys, H. (1935,1961). Some tests of significance, treated by the theory of probability, in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, pp. 203–222, Cambridge Univ Press.

Jeffreys, H. (1961). *Theory of Probability*, Oxford: Oxford University Press.

John, E. (2002). The neurophysics of consciousness, *Brain Res. Rev.*, **39**: 1–28.

Johnston, N. W., Johnston, S., Duncan, J. M., Greenea, J. M. and Kebadze, T. (2005). The September epidemic of asthma exacerbations in children: a search for etiology, *J Allergy Clin Immunol*, **115**: 132–138.

Julious, S., Campbell, M., Bianchi, S. and Murray-Thomas, T. (2011). Seasonality of medical contacts in school-aged children with asthma: Association with school holiday, *Public Health*, **125 (11)**: 769–776.

Julious, S., Osman, L. and Jiwa, M. (2007). Increases in asthma hospital admissions associated with the end of the summer vacation for school age children with asthma in two cities from England and Scotland, *Public Health*, **121**: 485–484.

Kalman, R. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, **82**: 35–45.

Kalman, R. and Bucy, R. (1961). New results in linear filtering and prediction theory, *Journal of Basic Engineering*, pp. 95–108.

Kass, R. and Raftery, A. (1995). Bayes factors, *Journal of American Statistical Association*, **90**: 773–795.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series analysis*, New York: John Wiley.

Kimes, D., Levine, E., Timmins, S. and Weiss, S. (2004). Temporal dynamics of emergency and hospital admissions of pediatric asthmatics, *Environ Res*, **94**: 7–17.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models., *Journal of Computational and Graphical Statistics*, **5**: 1–25.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Springer Verlag, New York,.

Kong, A., Liu, J. and Wong, W. (1994). Sequential imputations and Bayesian missing data problems, *Journal of the American Statistical Association*, **99**: 278–288.

Lincoln, D., Morgan, G., Sheppeard, V., Jalaludin, B., Corbeet, S. and Beard, J. (2006). Childhood asthma and return to school in Sydney, Australia, *Public Health*, **120**: 854–862.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*, Springer- Verlag; New York.

Liu, J. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association*, **93**: 1032–1044.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering, in *Sequential Monte Carlo methods in practice*, pp. 197–223, Springer.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models, *Biometrika*, **65 (2)**: 297–303.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *The journal of chemical physics*, **21 (6)**: 1087–1092.

Migon, H. S., Gamerman, D., Lopes, H. F. and Ferreira, M. A. (2005). Dynamic models, *Handbook of Statistics*, **25**: 553–588.

Montgomery, D., Johnson, L. and Gardiner, J. (1990). *Forecasting and Time Series Analysis*, New York: McGraw-Hill.

Montgomery, D. and Mastranglo, C. (1991). Some statistical process control methods for autocorrelated data, *Journal of Quality Technology*, **23**: 179–293.

Myatt, D., Nasuto, S. and Maybank, S. (2006). Towards the automatic reconstruction of dendritic trees using particle filters. in: Nsspw: Nonlinear statistical signal processing workshop classical, unscented and particle filtering methods, *IEEE*, pp. 193–196.

Myers, K., Brockwell, A. and Eddy, W. (2007). State-space models for optical imaging, *Statistics in Medicine*, **26**: 3862–3874.

Nelder, J. (1998). A large class of models derived from generalized linear models, *Statistics in Medicine*, **17**: 2747–2753.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models, *Journal of Royal Statistical Socity. Series A (General)*, **135**: 370–384.

Page, E. (1954). Continuous inspection scheme, *Biometrika*, **41**: 100–115.

Petris, G., Petrone, S. and Campagnoli, P. (2009). *Dynamic Linear Models with R*, Springer, New York.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association*, **94**: 590–599.

Pontabry, J. and Rousseau, F. (2011). Probabilistic tractography using q-ball modeling and particle filtering, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 209–216, Springer.

Priestley, M. (1981). *Spectral Analysis and Time Series*, New York:Academic Press.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, **77 (2)**: 257–286.

Ristic, B., Arulampalam, S. and Gorden, N. (2004). *Beyond the Kalman Filter: Particle Filter for Tracking Applications*, Artech House, London.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*, Springer Science & Business Media.

Robert, C. P. and Casella, G. (2010). *Introducing Monte Carlo methods with R*, Springer Science & Business Media.

Roberts, S. (1959). Control chart tests based on geometric moving averages., *Technometrics*, **1**: 239–250.

Rosenblatt, M. (1952). Remarks on a multinariate transformation, *The Annals of Mathematical Statistics*, **23**: 470–472.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*, ASQ Quality Press.

Shumway, R. and Stoffer, D. (2006). *Time Series Analysis and Its Application with R Examples*, Springer.

Singh, S. and Prajapati, D. (2011). Behavior of cusum chart for autocorrelated data, *International Journal of Engineering Sciences Research*, **2 (4)**: 1–8.

Smith, J. (1985). Diagnostic checks of non standard time series models, *Journal of Forecasting*, **4**: 283–291.

Storr, J. and Lenney, W. (1989). School holidays and admissions with asthma, *Arch Dis Child*, **64**: 103–107.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters, *IEEE Transactions on Signal Processing*, **50**: 281–290.

Stoumbos, Z. and Reynolds, J. (2000). Robustness to nonnormality and autocorrelation of individuals control charts, *Journal of Statistical Computation and Simulation*, **66 (2)**: 145–187.

Thode, H. C. (2002). *Testing for normality*, vol. 164, CRC press.

Triantafyllopoulos, K. (2008). Dynamic generalized linear models for non-Gaussian time series forecasting, *arXiv e-prints:0802.0219*.

Triantafyllopoulos, K. (2009). Inference of dynamic generalized linear models: On-line computataion and appraisal, *International Statistical Review*, **77 (3)**: 430–450.

Triantafyllopoulos, K. and Bersimis, S. (2016). Phase ii control charts for autocorrelated processes, *Quality Technology & Quantitative Management*, **13 (1)**: 88–108.

Triantafyllopoulos, K. and Montana, G. (2011). Dynamic modelling of meanreverting spreads for statistical arbitrage, *Computational Management Science*, **8**: 23–49.

Ullo, S., Murino, V., Maccione, A., Berdondini, L. and Sona, D. (2015). Bridging the gap in connectomic studies: A particle filtering framework for estimating structural connectivity at network scale, *Medical Image Analysis*, **21**: 1–14.

West, M. (1986). Bayesian model monitoring, *Journal of the Royal Statistical Society. Series B*, **48**: 70–78.

West, M. (1993). Approximating posterior distributions by mixtures, *Journal of the Royal Statistical Society*, **B 55 (2)**: 409–422.

West, M. and Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag; London.

West, M., Harrison, P. and Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting, *American Statistical Association*, **80**: 73–96.

Wonga, K., Galkab, A., Yamashited, O. and Ozaki, T. (2006). Modelling non-stationary variance in eeg time series by state space garch model, *Computers in Biology and Medicine*, **36**: 1327–1335.

Yang, H.-F., Descombes, X., Kervrann, C., Medioni, C. and Besse, F. (2013). Tracking growing axons by particle filtering in 3d+ t fluorescent two-photon microscopy images, in *Computer Vision–ACCV 2012*, pp. 272–283, Springer.

Yap, P.-T., Gilmore, J. H., Lin, W. and Shen, D. (2011). Longitudinal tractography with application to neuronal fiber trajectory reconstruction in neonates, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 66–73, Springer.

Yuan, L., Zheng, Y. F., Zhu, J., Wang, L. and Brown, A. (2012). Object tracking with particle filtering in fluorescence microscopy images: Application to the motion of neurofilaments in axons, *Medical Imaging, IEEE Transactions on*, **31 (1)**: 117–130.

Zeger, S. (1988). A regression model for time series of counts, *Biometrika*, **75**: 621–629.

Zhang, N. (1997). Detection capability of residual control chart for stationary process data, *Journal of Applied Statistics*, **24 (4)**: 475–492.