Bayesian Forecasting and Dynamic Linear Models

Jordan Oakley

A Dissertation presented for the degree of
Master of Research in Mathematical Sciences

Statistics and Probability
Department of Mathematical Sciences
Durham University
England

February 2019

Dedicated to

Zahra Qureshi

Bayesian Forecasting and Dynamic Linear Models

Jordan Oakley

Submitted for the degree of Master of Research
in Mathematical Sciences
February 2019

Abstract

Dynamic models offer a powerful framework for the modelling and analysis of time series, especially noisy time series, which are subject to abrupt changes in pattern. They are used in many time series applications from finance and econometrics, to biological series used in clinical monitoring. In this thesis we describe in detail how dynamic models can be used to model time series, following work from West and Harrison [1]. In particular, we will focus our attention to a specific problem of monitoring kidney failure in patients that have just had cardiac surgery. This work is in joint collaboration with the cardiac surgery unit at the University Hospital of South Manchester. The particular problem studied is that of developing an on-line statistical procedure to monitor the progress of kidney function in individual patients who have recently had heart surgery.

Declaration

The work in this thesis is based on research carried out at the Department of Mathematical Sciences, Durham University, UK. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

First, a big thank you to my supervisors, Dr Camila Caiado and Professor Michael Goldstein, for their time, patience, and expertise throughout my studies and for giving me the opportunity to work on such an amazing project during my first year in statistics, and further thanks to Dr Camila Caiado for the opportunities to attend conferences throughout the year, and for obtaining funding for my studies. In addition, I would like to thank Dr Samuel Howitt for his help and expertise throughout my studies.

I am also very grateful to Dr Jochen Einbeck for many hours of discussions about dynamic models and also for the opportunity to assist in teaching undergraduate statistics.

I would like to thank Professor Bärbel Finkenstädt Rand for sending me notes on Bayesian Forecasting and Intervention, which helped with my initial understanding of Bayesian forecasting and dynamic models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Acute kidney injury (AKI), as defined by the KDIGO (Kidney Disease Improving Global Outcomes) guidelines (Table 1.1) occurs in up to 75% of patients on the general intensive care unit (ICU) [3, 4] and in up to 30% of patients following cardiac surgery on the cardiac intensive care unit (CICU) [5]. It is essential that AKI is detected early and treated promptly, since without quick identification and treatment, abnormal levels of salts and chemicals can build up in the body, which affects the ability of other organs to work properly [6]. If a patient's kidneys fail completely, a patient may require temporary, or long term support from a dialysis machine (renal replacement therapy) thus we will aim to identify AKI as early as possible. In this chapter we will introduce the current world guidelines for monitoring kidney function and describe limitations with these criteria. This will lead us to consider monitoring kidney function by using statistical models in order to accurately detect and diagnose AKI.

1.1 Acute Renal Failure (ARF) and the KDIGO Guidelines

The KDIGO guidelines are clinical practice guidelines for the diagnosis, evaluation, prevention, and treatment of kidney disease. The concept of acute renal failure

(ARF) has undergone significant review over recent years. Evidence suggests that changes in the urine output and blood chemistries indicate injury to the kidney or impairment of kidney function [7]. These changes are warnings of serious clinical consequences [8, 9, 10, 11, 12], but traditionally, most studies only emphasise the most severe reduction in kidney function. It has only been recently that minor decreases of kidney function have been recognised as potentially important in the critically ill [7]. Identifying and intervening in patients with minor decreases in kidney function is clinically important as this can prevent kidney failure and mortality.

The glomerular filtration rate (GFR) is widely accepted as the most accurate measure of kidney function [7]. However, GFR is difficult to measure and is commonly estimated from the serum level of filtration markers, such as creatinine [7], which is often only measured once per day for patients in our dataset since it requires a blood test which takes time to analyse. Patients can be classified as having ARF within hours after having heart surgery, sometimes before the first measurement of serum creatinine is made, therefore identifying ARF in these patients is very difficult using serum creatinine alone. In this thesis we consider other variables (urine output), that are easy to measure on a regular basis, that can help identify which patients are likely to suffer kidney injury.

1.2 AKI: Acute Kidney Injury/Impairment

Due to recognition of the importance of minor decreases in kidney function, the RIFLE (Risk, Injury, Failure, Loss, End stage renal disease) criteria was defined. By redefining the acute changes in renal function more broadly, RIFLE criteria moves beyond ARF and focuses on a wide range of levels of kidney injury. The term acute kidney injury (AKI) was proposed to encompass minor changes in renal function as well as the requirement for renal replacement therapy (RRT) [13]. AKI covers the whole spectrum from minor to severe renal conditions. Moreover, AKI includes patients without actual kidney damage but with functional impairment relative to the physiologic response to surgery. Including such patients in the classification is clinically useful because these are precisely the patients that may benefit from early intervention. AKI by serum creatinine and urine output are classifications used worldwide for identifying and monitoring kidney deterioration, and the different levels of AKI are shown in Table 1.1.

Stage of AKI	Serum Creatinine	Urine Output
1	$1.5 - 1.9$ times baseline or $\geq 26.5 \mu mol/l$ increase	$< 0.5 ml/kg/hr$ for ≥ 6 hours
2	$2.0 - 2.9$ times baseline	$< 0.5 ml/kg/hr$ for ≥ 12 hours
3	3 times baseline or $\geq 353.6 \mu mol/l$ Initiation of RRT	$< 0.3 ml/kg/hr$ for ≥ 24 hours or Anuria for ≥ 12 hours

Table 1.1: KDIGO AKI criteria for adult patients. If one or the other criteria is met then a patient is considered suffering AKI

1.3 Monitoring Cardiac Patients

In the early stages following cardiac surgery, some patients will experience complications, which may result in deterioration in function of the patient's kidneys. Clinicians are therefore concerned with identifying methods for monitoring patients in order to detect sudden changes in performance of the patient's kidneys.

The urine output criteria classify patients as suffering stage one AKI if their urine output is less than $0.5 ml/kg/hr$ for six consecutive hours (see Table 1.1). This criterion is used worldwide but studies in general ICU patients have suggested that this urine output criterion may be too sensitive as patients diagnosed with stage one AKI by urine output alone have better outcomes than those who meet both (stage one) urine output and the serum creatinine criteria together [4, 2]. Ralib *et al* demonstrated that a threshold of $0.3 ml/kg/hr$ for six consecutive hours (which is what we will refer to when stating "severe oliguria") was more closely associated with adverse outcomes in general ICU patients [2]. However, if this threshold is used instead of the existing AKI stage one urine output definition (instead of using $0.5 ml/kg/hr$ for 6 consecutive hours), patients may suffer harm in the time required for severe oliguria to be observed (see Figure 1.1).

Figure 1.1 shows a urine output time series (a series of observations taken sequentially over time) for a patient that has just entered CICU after heart surgery. The blue dot, at hour 8, is the hour at which this person would have been classified as suffering stage one AKI by urine output (see Table 1.1) and the red dot, at hour 24, is the hour at which this person would have been classified as suffering severe oliguria, as defined in [2]. By the time the classification of severe oliguria is made,

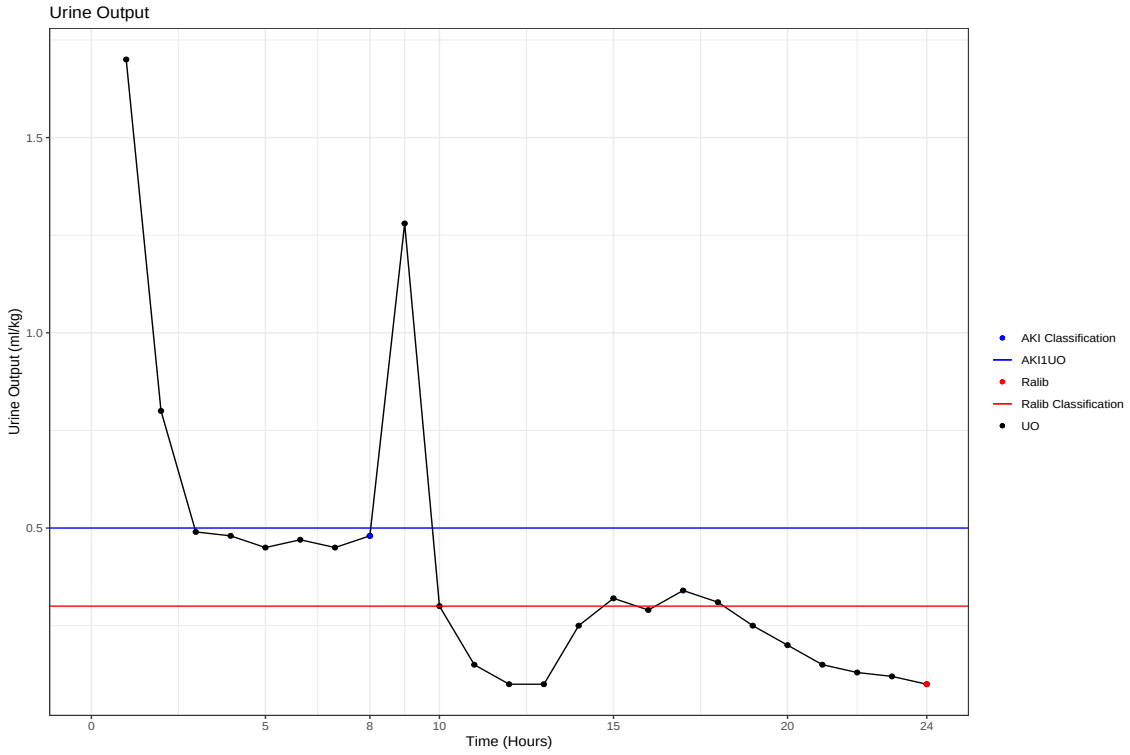

Figure 1.1: AKI Stage 1 Urine Output and Ralib Classifications. The blue line represents the stage one AKI urine output threshold (see Table 1.1) and the red line represents the threshold defined by Ralib [2]. The blue dot, at hour 8, is the time at which this patient would have been classified as suffering stage one AKI. The red dot, at hour 24, is the time at which this patient would have been classified as suffering severe oliguria

harm may have already occurred to the patient's kidneys in the time gap between classification of stage one AKI by urine output and severe oliguria. For example, in Figure 1.1, the time gap between classifications for AKI and severe oliguria is 16 hours. During this period a patient's kidneys can drastically deteriorate and harm may occur. By the time a patient is classified as suffering severe oliguria it may be too late for minor intervention methods to normalise kidney function and hence more complicated interventions are required.

We will therefore forecast severe oliguria (using statistical models), recalling that severe oliguria is more closely associated with adverse outcomes than the AKI stage one urine output criteria [2], in order to prevent harm from occurring whilst waiting for severe oliguria to be observed.

1.4 Data Description

In this study, we use prospectively collected data from the cardiac unit at the University Hospital of South Manchester. Data collected from adult patients admitted between January 2013 and November 2017 was analysed. The data collected contains many variables, such as weight, height, age, gender, urine output, and biochemistries for the patients.

Since a patient's urine output is dependant upon their weight we model the normalised urine output, namely urine output per kilogram. However, we have observed that urine output does not depend on age, or gender and hence these variables, amongst others, will not be included in our analysis.

1.5 Structure

This thesis is concerned with the prediction of severe oliguria and its consequences using dynamic linear models. Our methods of clinical prevention will be compared to methods currently used to detect and avert the consequences caused by renal dysfunction.

In this chapter we have introduced our motivation behind monitoring kidney function and we will then return to this area in Chapter 5 and analyse clinical problems for monitoring kidney function in detail. In Chapters 2 and 3 we discuss and analyse static linear models and dynamic linear models, respectively. Chapter 3 continues to discuss benefits of dynamic modelling when compared to static modelling. Chapter 3 concludes by discussing time series analysis and presents ideas and concepts of the Bayesian analysis of the dynamic linear model. In Chapter 4 we explore the mathematical structure of building complex dynamic models in detail.

In Chapter 5 we return to the problem of monitoring kidney function and we discuss how dynamic models can be used to prevent adverse outcomes associated with kidney failure, comparing our methods of clinical prevention to current worldwide guidelines. In Chapters 6 and 7 we discuss how to monitor model forecast performance and propose a framework for modelling noisy time series which are subject to abrupt changes in pattern.

Chapter 2

Linear Models

In this chapter we will discuss and analyse static linear models and we will also encounter scenarios where static linear models drastically fail to capture the dynamics of the system of interest.

2.1 Linear Regression

Linear regression is a method where a function of a response variable is modelled by a weighted linear combination of predictors [14]. Generally, the linear model, in matrix notation, is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\nu}, \tag{2.1.1}$$

where $\boldsymbol{Y}$ is an n-dimensional vector of responses; $\boldsymbol{X}$ is an $(n \times p)$ design matrix, where each column represents a predictor variable; $\boldsymbol{\theta}$ is a p-dimensional parameter vector; and $\boldsymbol{\nu}$ is an n-dimensional residual vector.

Fitting the model using least squares estimates, assuming linearity, i.e. $\mathrm{E}[\nu_i] = 0$ for $i = 1, \ldots, n$, where ν_i is the ith component of the residual vector $\boldsymbol{\nu}$; homoscedasticity, i.e. $\mathrm{Var}(\nu_i) = \sigma^2$ for $i = 1, \ldots, n$; and independence between observations, i.e.

$\text{Cov}(\nu_i, \nu_j) = 0, \forall i \neq j$, yields estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ given by (see Appendix **??**) [**?**]

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \\ \hat{\sigma}^2 &= \frac{\hat{\boldsymbol{\nu}}'\hat{\boldsymbol{\nu}}}{n-p} = \frac{\text{RSS}}{n-p}, \end{aligned} \tag{2.1.2}$$

where RSS is the residual sum of squares. These estimates are frequentist estimates, and are only used in this chapter.

Assumptions can be made about (2.1.1) for simplicity and tractability. A linear model, with normally distributed residuals, can be written as

$$\boldsymbol{Y} \sim \text{N}_n(\boldsymbol{X}\boldsymbol{\theta}, \sigma^2\boldsymbol{I}), \tag{2.1.3}$$

where $\boldsymbol{I}$ is the $(n \times n)$ identity matrix. The assumptions of (2.1.3) are that the residuals are assumed to have mean zero, $\text{E}[\nu_i] = 0$; the variance of the observations is assumed constant, $\text{Var}(\nu_i) = \sigma^2$; the residuals are assumed to be uncorrelated, $\text{Cov}(\nu_i, \nu_j) = 0, \forall i \neq j$; and that ν_i are normally distributed.

2.2 Urine Output Series: Static Linear Regression

Let us consider the linear regression model of a scalar response

$$Y = \mu + \beta t + \nu, \tag{2.2.1}$$

for some constants μ and β, the response Y is urine output per kilogram for a patient, and we assume $\nu \sim \text{N}(0, \sigma^2)$. In the notation of Equation (2.1.3), $\boldsymbol{\theta} = (\mu, \beta)'$, and

$$\boldsymbol{X} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix},$$

where $t_1, \ldots, t_n$ are the times at which the urine outputs were recorded. We want to be able to make predictions about urine output (per kilogram) for the next six hours, every hour, for a patient (i.e. predicting severe oliguria, see Section 1.3). These predictions will give the expected urine output for hours $t+1, \ldots, t+6$, given that we have observed urine outputs $1, \ldots, t$, and 95% prediction intervals for the

forecasts. This will give us point predictions and also a measure of the uncertainty of the predictions.

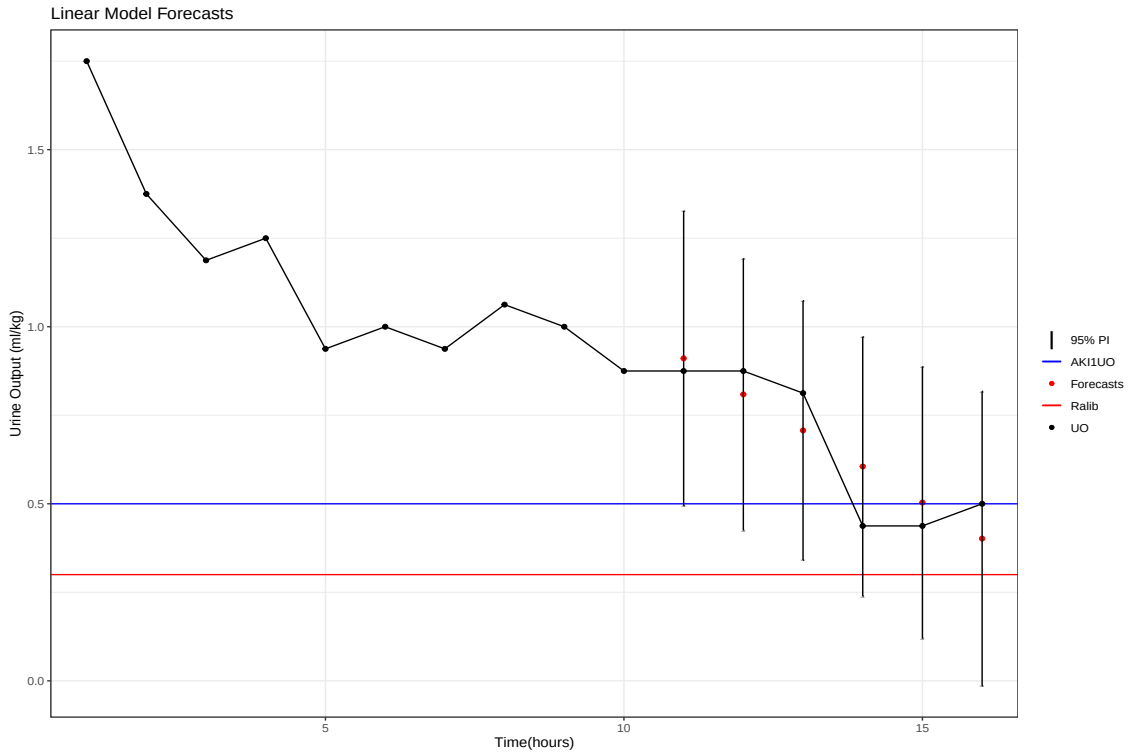

Figure 2.1: Linear model point forecasts with corresponding 95% prediction intervals. The blue and red lines represent the AKI stage one urine output threshold and Ralib's threshold, respectively

Figure 2.1 shows six-step forecasts at time $t = 11$. The prediction intervals obtained from the static linear model are too large to be useful to clinicians and also include negative values, e.g. the six-step ahead 95% prediction interval ranges from -0.014 to 0.82. It is not possible to have a negative urine output and so the intervals are physically implausible as well as being too wide. This is not desirable, but a way to overcome this is to transform the data.

2.3 Data Transformations

Many patients, after having spent hours in theatre for heart surgery, exhibit an exponential decay in urine output to begin with. This nonlinearity violates an assumption of the regression model defined in Equation (2.1.3). A logarithmic transformation resolves this problem by linearising this decay. Furthermore, the regression model

imposes no constraints that force the predictions made to be positive. As a result, by using the regression model defined in Equation (2.1.3), the urine output forecasts can be negative (see Figure 2.1). This is physically impossible and makes it difficult to interpret forecasts. However, if we take the logarithm of the response, fit the model, and then transform back to the raw scale, the resulting predictions are strictly positive.

When making a logarithmic transformation care must be taken for the zeros in the data. For the system that we are considering, a urine output recording of zero, corresponding to anuria (failure of the kidneys to produce urine) is possible and happens in a large proportion of patients. As a result we add a small constant to the data before taking logarithms. We consider the transformation $Y \longmapsto \log(Y+\varepsilon)$.

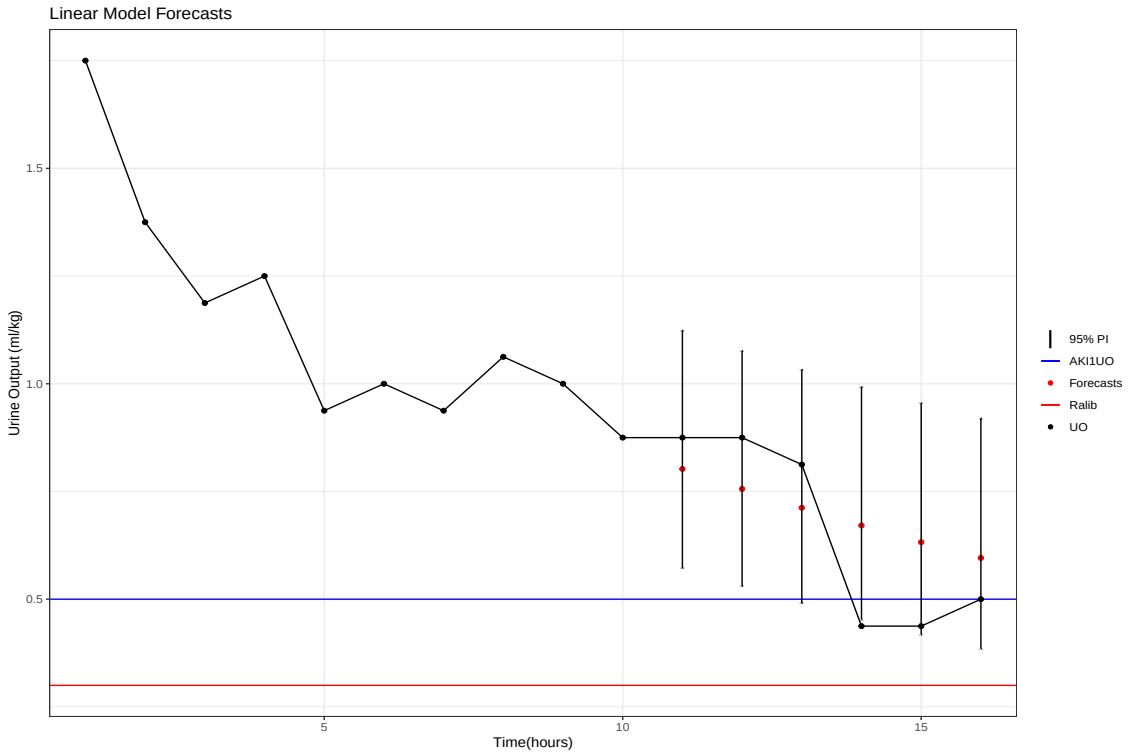

Figure 2.2: Linear model point forecasts with corresponding 95% prediction intervals. The prediction intervals are not symmetric anymore after applying a logarithmic transformation

2.4 Urine Output Series: Transformation

Following the discussion above we transform $Y \longmapsto \log(Y + \epsilon)$ and refit the (transformed) linear model

$$\log(Y + \varepsilon) = \mu + \beta t + \nu, \tag{2.4.1}$$

for some constants ε, μ, and β, the response $\log(Y + \varepsilon)$, where $\varepsilon = 0.1$ (this value was chosen to match the value of ε used in Sections 5, 6, and 7. Appendix D shows a sensitivity analysis for different choices of epsilon) is the transformed urine output per kilogram for a patient, and we assume $\nu \sim \mathrm{N}(0, \sigma^2)$.

Figure 2.2 shows six-step forecasts at time $t = 11$. The prediction intervals are shorter than in Figure 2.1 and are now more useful to clinicians measuring the uncertainty in the predictions, e.g. the six-step ahead 95% prediction interval ranges from 0.39 to 0.92 (note that the figure and forecasts in this discussion relate to the observation scale, not to the transformation scale on which the analysis is being performed: Communication is always in terms of the quantities directly meaningful to the investigator). We can also see that the prediction intervals are not symmetric anymore but do not extend into negative values and hence the prediction intervals are now physically useful to interpret.

We see that the (transformed) static linear model is adequate for modelling and performing forecasts in a steady case. However, the patient that we are considering starts with a simple downward trend and so a static model is (not surprisingly) able to capture the dynamics of the trend quite well. However, often the series of urine outputs for a patient is extremely noisy as a result of biological variation and errors arising in the collection, measurement and processing of the data. Furthermore, these series may be subject to several different types of abrupt change. If a patient has a downward trend it is not uncommon that the patient is given an intervention to increase their urine output. In this scenario the trend of the urine output can change drastically. The sign and rate of change of the slope can change frequently and the linear model soon becomes incapable of modelling such a complex system. In addition, many unknown, random, biological sudden changes occur in many urine output series which can create a lot of noise in the series and can cause sudden changes in the trend of the series.

2.5 The Downfall of Static Linear Models

In some circumstances a static linear model is a perfectly valid model to use, however, if the trend or shape of the data changes over time, these models soon become inadequate. Let us consider what happens when modelling the same patient as in the previous examples, but over a longer period of time. Once again, we consider the transformed data and remind the reader that the figure and forecasts in the discussion relate to the observation scale, not to the transformation scale on which the analysis is being performed.

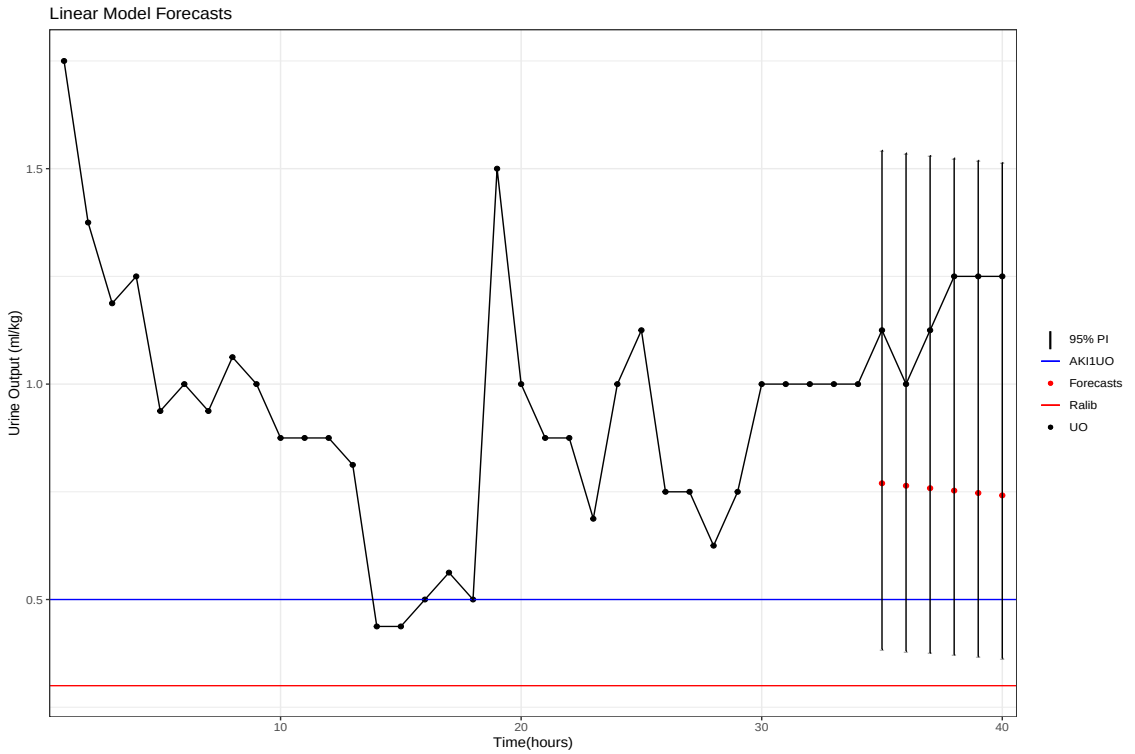

Figure 2.3: The Downfall of Static Linear Models. The point forecasts are unable to adapt to the changing trend of the urine output series and prediction intervals are too wide to be useful to clinicians

From Figure 2.3 it is clear that the patient's urine output drastically increases at hour 20 and at this point the static linear model starts to break and is unable to recognise the newer changing trends. Even 15 hours after the high urine output at hour 20, after observing the urine output at hour 35, the static linear model still predicts a downward trend. Static models use all observations (up to and including the current time) when calculating the forecast points and intervals. This is unwanted since a

urine output at time, $t - 15$ say, is unlikely to be useful when forecasting ahead at time t. The most useful observations will be those most recent to the current forecast time. Figure 2.3 shows that the predicted points are far from adequate and the prediction intervals are far too wide to be of any use to clinicians, e.g. the six-step ahead 95% prediction interval ranges from 0.36 to 1.51. Retrospectively, we can see that there is both a level change after hour 20 and then a slope change which is not recognised by the static model. For complex, noisy, time series data, it is essential that model parameters are able to evolve over time to capture changing trends and to make precise forecasts.

In summary, we see that the static linear model could not adapt to the noisy time series data quickly enough which lead to imprecise forecasts. The static model only suffices when a patient has a roughly constant trend but even simply changing from one steady trend to another is slow to be recognised by a static model. It is rarely the case that a patient has a roughly constant trend and if it is the case it is only for a short period of time. Therefore, it would be useful if we could incorporate our subjective knowledge into the model to make the predictions more useful to clinicians, and for the model parameters to sequentially update simultaneously with the system in order to capture the dynamics of the ever changing time series. In the example considered, when predicting the next six urine outputs at time t, the most recent urine outputs will be the most informative about what the forecasted urine outputs will be. We do not want past evidence of unhealthy decline influencing our forecasts when the trend has changed and the kidneys are performing adequately, since this can mislead clinicians. The idea of information loss over time as new information becomes available is the key to modelling complex time varying systems, and leads us to the next chapter which introduces dynamic linear models.

Chapter 3

Dynamic Linear Models

In this chapter we discuss time series and introduce dynamic linear models. We examine the nature of time series and we present ideas and concepts of Bayesian forecasting with dynamic models.

3.1 The Nature of Time Series

A time series is a series of observations taken sequentially over time. In a standard regression model the order in which observations are included in the data set is irrelevant. It is the ordering property of observations that distinguishes time series from non-time series data. Observations made at some time have effects on observations at later times.

Three basic model forms describe the majority of time series and forecasting situations [1]. They are models for time trends, systematic cyclic variation, and regressions. Combinations of these forms via block structuring (see Section 4.7) provides a large class of dynamic models suitable for modelling time series in many applications. Trend models are the simplest component models which represent a system with a straightforward linear progression.

3.2 Dynamic Models

The class of dynamic linear models is a class of models which contains the class of static models. Static models are models with one set of parameters whose values are fixed across all observational units. In some circumstances this assumption is valid, but in some circumstances it is a dangerous assumption to make, especially in the urine output time series (see Section 2.5). Time always brings changing circumstances and new considerations. For example, changes from healthy to deteriorating kidneys or vice versa. Dynamic models are a powerful way to handle time series problems since they are formulated to allow for changes in parameter values over time to reflect changing trends as new information becomes available.

We can think of additional information arriving sequentially as increasing our knowledge about the system and how the system is changing with time. If we make a new observation, say at time t, this gives us additional information about our model that we did not have at time $t-1$. However, not all information that we have about a system will be useful to us at time t when making decisions and forecasts about the future. If we were forecasting the response of urine output this coming hour, the urine output level from 10 hours ago will be less useful to us when making decisions about a patient's current kidney function than the level of urine output from the most recent observations. The idea of parameters evolving over time to adapt to new trends and to "forget" older trends is the key to dynamic models. This information loss is what distinguishes dynamic models from static models and allows dynamic models to capture changing trends by evolving to reflect new information. When building dynamic models we define a form for parameters of a series which is only appropriate locally [1], (see Figure 3.1). The aim being to "forget" older trends and to put more emphasis on local information. Static models have relationships that apply globally, due to the static parameters. Figure 3.1 is an example of data that changes markedly over time. It is clear that the trend in region one is completely different to the trend in region two. If we fit a static linear model (blue line) to the data we see that, by using all observations (up to and including time $t = 6$), we drastically under forecast at time $t = 7$ and we completely miss the changing trends of the data. We see that a model that is able to adapt to different regions is required.

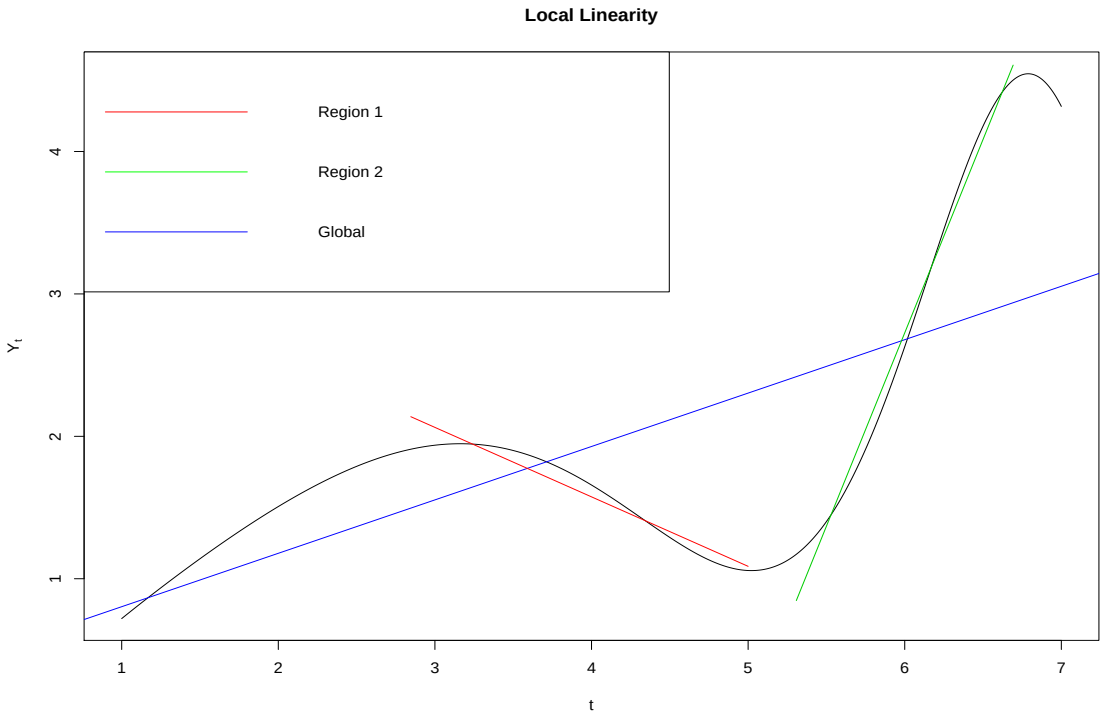

Figure 3.1: Local linearity of dynamic linear models compared to a static model with global parameters (adapted from [1])

3.3 The Dynamic Linear Model

For a response variable Y, and a vector of regressor variables $\boldsymbol{x}$, the equation for a single observation from a static linear model is usually written as

$$Y = \boldsymbol{x}'\boldsymbol{\theta} + \nu, \tag{3.3.1}$$

where $\boldsymbol{x}'$ represents the transpose of $\boldsymbol{x}$, $\boldsymbol{\theta}$ is a vector of unknown parameters and ν is an unobservable stochastic error term.

The equation for a single observation from a dynamic linear model is usually written as

$$Y_t = \boldsymbol{x}_t'\boldsymbol{\theta}_t + \nu_t, \tag{3.3.2}$$

where $\boldsymbol{x}_t$ is a column vector at time t and $\boldsymbol{x}_t'$ is its transpose. This equation is often referred to as the observation equation [1]. The dynamic model exhibits two

generalisations over the static model. Firstly, the time ordering of the observation sequence, $Y_1, Y_2, \ldots, Y_t$, as noted in Section 3.1, distinguishes time series from non-time series data. Secondly, each time has an individual parameterisation, $\boldsymbol{\theta}_t$, as noted in Section 3.2, evolving parameters are essential to capture changing trends. For example, it is often the case that the patient, after being in theatre for hours for surgery, will start with a high urine output, due to the patient's urine output not being measured for hours during surgery, and a decline in level immediately afterwards. If the patient's urine output gradually levels out, at a healthy, but roughly constant level, we do not want the early decline in urine output to influence our forecasts for this patient at the current time, and hence a model with time varying parameters is required. Figure 3.2 shows the difference between a static model and a dynamic model in the scenario described above. In this figure the patient starts with a steady downward trend but then, after a jump in urine output at hour 71, levels out to a healthy steady trend. By "forgetting" the initial unhealthy trend the dynamic model is able to rapidly adapt to the changing trend of the urine output data (Figure 3.2 (right)). Conversely, we can see that the static model (Figure 3.2 (left)) breaks down at hour 13 (when the slope in the data gets less steep), and is completely unable to adapt to the trend change at hour 71. As a result, the static model could misleadingly alarm experts that the patient's kidneys are failing by providing such poor, inaccurate, forecasts.

The parameter sets for each observation in the static linear model are the same set. Conversely, in the dynamic linear model, parameter sets are distinct for each observation, but are stochastically related through the system equation [1]. The system equation describes the evolution of parameters through time. The evolution of the parameters is a first order Markov process,

$$\text{System Evolution Equation:} \qquad \boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{3.3.3}$$

where $\boldsymbol{G}_t$ is a matrix of known coefficients and $\boldsymbol{\omega}_t$ is a vector of uncorrelated, unobservable stochastic error terms [1]. The system evolution matrix, $\boldsymbol{G}_t$, defines a known relationship of the state vector at time t with its value at time $t-1$. Through the system equation information on the state vector is propagated through time. The presence of the stochastic element, $\boldsymbol{\omega}_t$, adds random noise to the propagation and is crucial to dynamic modelling. The amount of movement in the parameters is described by the stochastic vector $\boldsymbol{\omega}_t$, the more uncertain one is about the parameter

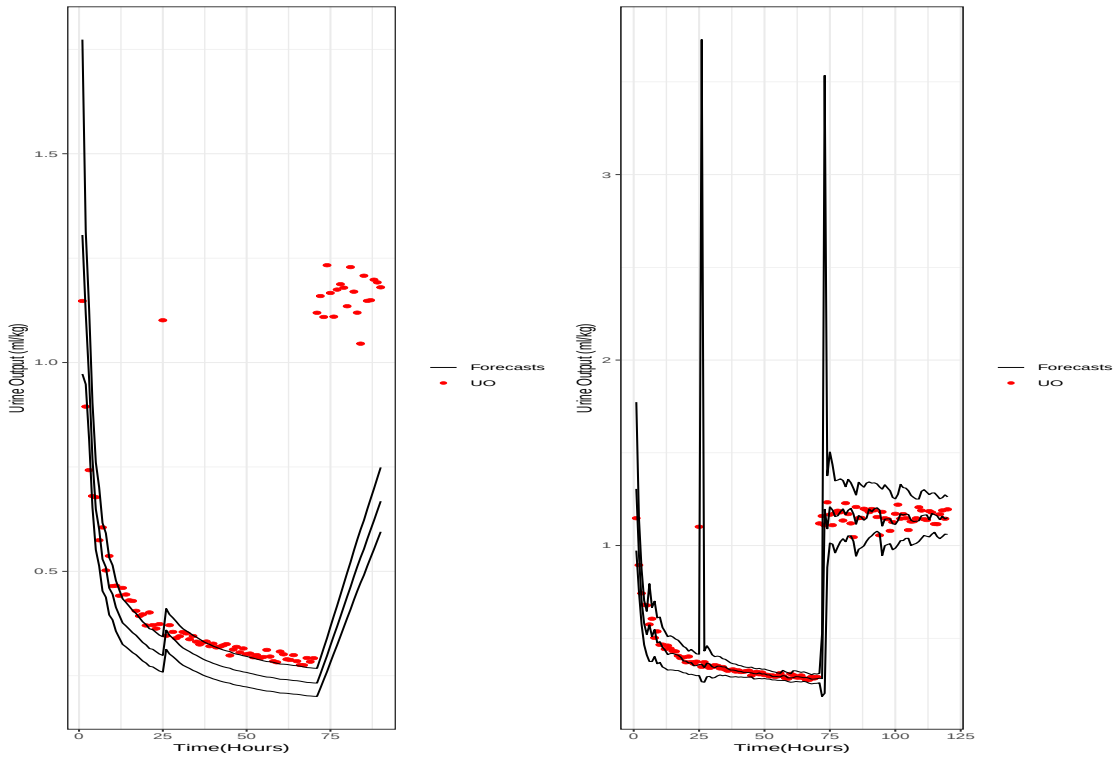

Figure 3.2: Static Model (left) versus Dynamic Model (right). Urine output observations are shown by red dots with corresponding point forecasts and 95% prediction intervals shown by black lines

values the larger the components of the stochastic error vector $\boldsymbol{\omega}_t$, and conversely the more certain one is about the parameter values the smaller the components of the stochastic error vector $\boldsymbol{\omega}_t$. In Section 6 we will see how increasing values in the stochastic vector $\boldsymbol{\omega}_t$, to reflect increased uncertainty about the future, allows dynamic models to adapt more rapidly to changing trends.

Generally the observation and the evolution equations can be written as [1]:

$$\begin{aligned} \text{Observation Equation:} \quad & Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathrm{N}(0, V_t), \\ \text{System Evolution Equation:} \quad & \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}(0, \boldsymbol{W}_t). \end{aligned} \tag{3.3.4}$$

3.4 Urine Output Series: Regression Model

Consider the following example of a dynamic regression model. The model we will use for illustration is a two component model in which the components are a level and a regression variable, where once again, the response represents the (transformed) urine output per kilogram

$$\log(Y_t + 0.1) = \mu_t + \beta_t \text{CVP}_t + \nu_t, \qquad \nu_t \sim \text{N}(0, V), \tag{3.4.1}$$

where the response is composed of an underlying level, μ_t, and an amount determined by CVP, the central venous pressure. Here the evolution equations are described by random walks (see section 4.6.3)

$$\begin{aligned} \mu_t &= \mu_{t-1} + \omega_{\mu t}, \qquad & \omega_{\mu t} &\sim \text{N}(0, W_{\mu t}), \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, \qquad & \omega_{\beta t} &\sim \text{N}(0, W_{\beta t}). \end{aligned} \tag{3.4.2}$$

The amount of movement in the level over time is determined by the stochastic term $\omega_{\mu t}$, which is governed by the variance, $W_{\mu t}$. As this variance increases (or decreases), the more (or less) volatile the series becomes. At the limit $W_{\mu t} = 0$ there is no volatility in the level and the dynamic parameter, μ_t, reduces to the static parameter, μ. At the limit $W_{\mu t} \longrightarrow \infty$ there is complete "information loss" from time $t-1$ to time t, allowing the level parameter to rapidly adapt to future data (see Section 6).

In this example (see Equations (3.3.4)), $\boldsymbol{F}_t' = (1, \text{CVP}_t)$, $\boldsymbol{\theta}_t = (\mu_t, \beta_t)$, $\boldsymbol{G}_t = \boldsymbol{G} = \boldsymbol{I}_2$, $V_t \equiv V$, and the system variance matrix is

$$\boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} & W_{\mu\beta t} \\ W_{\mu\beta t} & W_{\beta t} \end{pmatrix}, \tag{3.4.3}$$

where $W_{\mu\beta t} = \text{Cov}(\omega_{\mu t}, \omega_{\beta t}) = \text{Cov}(\omega_{\beta t}, \omega_{\mu t}) = W_{\beta\mu t}$.

3.4.1 Overview

The observation and system evolution equations (3.3.4), illustrate the concepts and important features of the class of dynamic linear models (DLMs). This class of DLMs is described and analysed in detail in Chapter 4, but here we provide a

basis for what follows in later chapters. The principles of Bayesian forecasting and dynamic modelling in this thesis involve

1. Parametric models with a probabilistic representation of information about parameters with the means to incorporate external information (see Section 6.2);
2. A sequential model definition utilising conditional independence (see Figure 3.3);
3. Probability distributions describing forecasts (see Section 4.2.1);
4. Model monitoring (see Chapter 6);
5. Dynamic mixture models (see Chapter 7).

We are interested in a scalar time series, Y_t, which represents urine output per kilogram. At time $t-1$ the current information is $D_{t-1} = \{D_{t-2}, Y_{t-1}\}$. At time $t-1$ the information relevant to predicting future observations is represented in the posterior probability distribution $(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$. Given information up to time $t-1$ the conditional distribution $(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$ is sufficient for predicting future forecasts [1]. There will also be occasions when external information is included by the forecaster. In these cases, changes in $\boldsymbol{\theta}_t$ may be required to allow the model to adapt to new circumstances. These occurrences can be very difficult to recognise and are discussed in Chapter 6.

The next modelling step is to relate the current information about the state into the future using the system evolution equation so that predictive distributions $(Y_{t+k} \mid D_{t-1})$, $k \geq 0$, can be derived. This is achieved by specifying a sequential relation to combine the forecasted system distributions $(\boldsymbol{\theta}_{t+k} \mid \boldsymbol{\theta}_{t-1}, D_{t-1})$ with the observation relations $(Y_{t+k} \mid \boldsymbol{\theta}_{t+k}, D_{t-1})$. In combination with the posterior distribution, $(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$, these distributions enable us to derive a full joint forecast distribution.

A crucial property enabling dynamic modelling is conditional independence (see Figure 3.3). The key structural feature is that given $\boldsymbol{\theta}_t$, the past present and future are mutually independent. Also, given information up to time t, D_t, all of the information concerning the future is contained in the posterior parametric distribution

$(\boldsymbol{\theta}_t \mid D_t)$. If we are only concerned with normal DLMs then we can go further and say that if $(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t)$ then the pair $\{\boldsymbol{m}_t, \boldsymbol{C}_t\}$ contains all of the relevant information about the future [1].

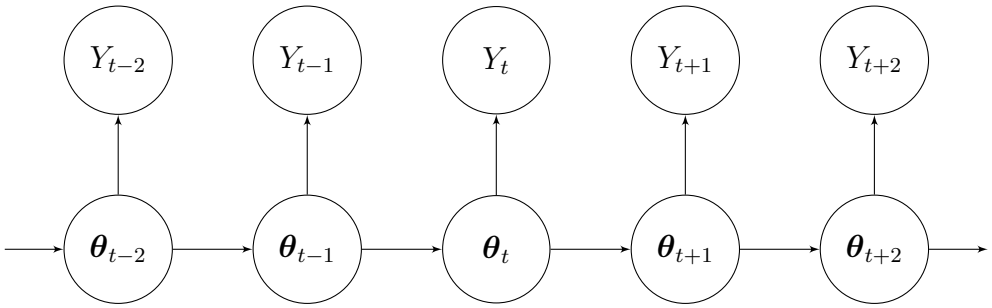

Figure 3.3: The DLM Conditional Independence Structure, adapted from [1]

Conditional independence features strongly in dynamic model building. As we shall see in Section 4.6 the principle of superposition states that any linear combination of independent normal DLMs is a DLM. In the case of the two-component DLM, with $\boldsymbol{\theta}_t' = (\theta_{t1}', \theta_{t2}')$, the two series representing different components of the DLM, $\{\theta_{t+i,1}, i > 0\}$ and $\{\theta_{t+i,2}, i > 0\}$ evolve independently (see Figure 3.4). This important consequence allows us to construct a complex DLM from a linear combination of simple DLMs [1].

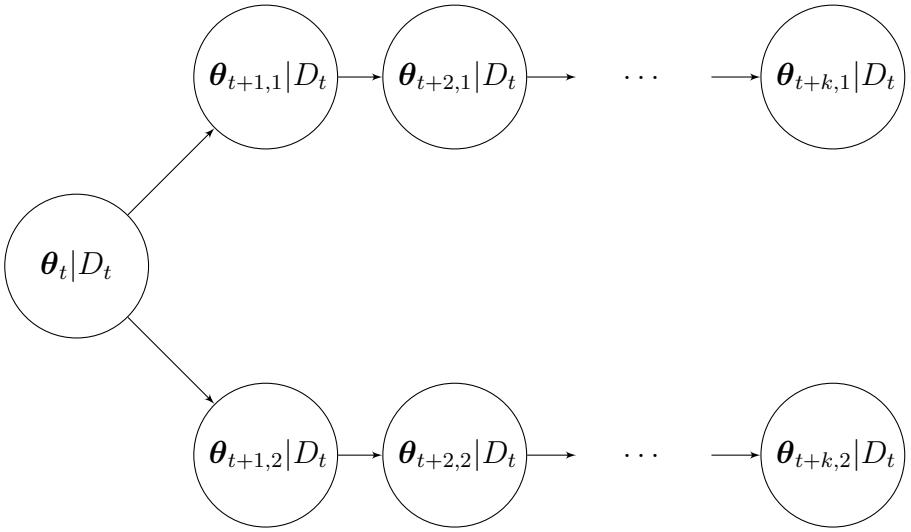

Figure 3.4: The Parametric Conditional Independence Structure, adapted from [1]

Using Bayes rule the joint distribution for the observations and parameters at time t may be derived using

$$p(Y_t, \boldsymbol{\theta}_t \mid D_{t-1}) = p(Y_t \mid \boldsymbol{\theta}_t, D_{t-1}) p(\boldsymbol{\theta}_t \mid D_{t-1}). \tag{3.4.4}$$

The one-step ahead forecast is given by the marginal distribution $(Y_t \mid D_{t-1})$, and the posterior distribution, $(\boldsymbol{\theta}_t \mid D_t)$, is the conditional distribution $(\boldsymbol{\theta}_t \mid D_{t-1}, y_t)$.

There are many external factors that can affect this and using Bayesian models this external information can be routinely incorporated into the system at any time. Bayesian dynamic models also allow for model monitoring. This involves using the proposed model routinely unless exceptional circumstances arise (see Chapter 6). These circumstances can occur in two ways. The first is when relevant external information is received which drastically changes the system. This can be routinely handled with Bayesian dynamic models and is sometimes referred to as a forward intervention (see Section 6.2). The second type of exception is model feedback. This occurs when a monitoring system is used to keep track of model performance and the monitoring system flags a significant inadequacy warning when the routine model performs poorly compared to some other alternative models. When the monitoring system flags a warning, diagnostics are performed to help correct the routine model. Model inadequacies can be very difficult to recognise and are discussed in Chapter 6. Automatic model monitoring schemes can be used to detect and diagnose model deficiencies and are used to monitor model performance.

Automatic model monitoring schemes will then lead us to using dynamic mixture models (see Chapter 7) to more concretely handle exceptional circumstances and to provide a more powerful framework for modelling noisy time series.

3.4.2 Definitions and Notation

The general DLM is defined for a vector observation $\boldsymbol{Y}_t$. However in this thesis we are only interested in the univariate case. Let Y_t be an observation on a time series. Following from the model defined by Equations (3.3.4) we have the following definitions and notation.

Definition 3.4.2.1. The general normal dynamic linear model is characterised by a set of quadruples [1]

$$\{\boldsymbol{F}, \boldsymbol{G}, V, \boldsymbol{W}\}_t = \{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\} \tag{3.4.5}$$

for each time t, where $\boldsymbol{F}_t$ is a known $(n \times 1)$ matrix, $\boldsymbol{G}_t$ is a known $(n \times n)$ matrix,

V_t is a known (1×1) variance, and $\boldsymbol{W}_t$ is a known $(n \times n)$ variance matrix (in Sections 4.8 and 4.9 we discuss how to incorporate learning procedures for unknown variances $\boldsymbol{W}_t$ and V_t respectively). This quadruple defines the model relating Y_t to the $(n \times 1)$ parameter vector $\boldsymbol{\theta}_t$ at time t, and the $\boldsymbol{\theta}_t$ sequence through time, via the sequentially specified distributions

$$\begin{aligned} (Y_t \mid \boldsymbol{\theta}_t) &\sim \mathrm{N}(\boldsymbol{F}_t'\boldsymbol{\theta}_t, V_t) \\ (\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, D_{t-1}) &\sim \mathrm{N}(\boldsymbol{G}_t\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_t). \end{aligned} \tag{3.4.6}$$

The conditional independence structure shown in Figure 3.3 applies. Given $\boldsymbol{\theta}_t$, Y_t is independent of all other observations and parameter values. In general, given the present, the future is independent of the past.

Chapter 4

Analysis of the DLM

In this chapter we present ideas and concepts of the Bayesian analysis of the dynamic linear model.

4.1 Model Form and Notation

The general univariate dynamic linear model is written as

$$\begin{aligned} &\text{Observation Equation:} && Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, && \nu_t \sim \mathrm{N}(0, V_t), \\ &\text{System Evolution Equation:} && \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, && \boldsymbol{\omega}_t \sim \mathrm{N}(0, \boldsymbol{W}_t), \\ &\text{Initial Information:} && (\boldsymbol{\theta}_0 \mid D_0) \sim \mathrm{N}(\boldsymbol{m}_0, \boldsymbol{C}_0), \end{aligned} \tag{4.1.1}$$

for some prior moments $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$. The time origin $t = 0$ is just an arbitrary label and applies particularly when the data $Y_1, Y_2, \ldots$ represents the continuation of a previously observed series. In such cases the initial prior is viewed as sufficiently summarising the information from the past. The initial state vector, $\boldsymbol{\theta}_0$, is concretely interpreted as the final state vector from the historical data. Otherwise, $\boldsymbol{\theta}_0$ has no such interpretation and the model may be equivalently initialised by specifying a normal prior, $(\boldsymbol{\theta}_1 \mid D_0)$, for the first state vector [1, 15]. Moreover, Y_t denotes the observation series at time t, $\boldsymbol{F}_t$ is a vector of known constants (the regression vector); $\boldsymbol{\theta}_t$ denotes the vector of model state parameters; ν_t is a stochastic error term having a normal distribution with zero mean and observational variance V_t;

$\boldsymbol{G}_t$ is a matrix of known coefficients that defines the systematic evolution of the state vector across time; and $\boldsymbol{\omega}_t$ is a stochastic error term having a normal distribution with zero mean and evolution covariance matrix $\boldsymbol{W}_t$. The two stochastic series $\{\nu_t\}$ and $\{\boldsymbol{\omega}_t\}$ are assumed to be temporally independent and mutually independent, i.e. $\text{Cov}[\nu_s, \nu_t]$, $\text{Cov}[\boldsymbol{\omega}_s, \boldsymbol{\omega}_t]$, $\forall s \neq t$, and $\text{Cov}[\nu_s, \boldsymbol{\omega}_t]$, $\forall s, t$ are zero. In this thesis we will only consider models with a constant evolution matrix, $\boldsymbol{G}_t \equiv \boldsymbol{G}$ and we only consider univariate time series Y_t.

4.2 Updating: Prior to Posterior Analysis

Prior information on the state vector for time $t+1$, given information up to time t, D_t, is summarised as a normal distribution with mean $\boldsymbol{a}_{t+1} = \text{E}[\boldsymbol{\theta}_{t+1} \mid D_t]$ and covariance matrix $\boldsymbol{R}_{t+1} = \text{Var}(\boldsymbol{\theta}_{t+1} \mid D_t)$,

$$\boldsymbol{\theta}_{t+1} \mid D_t \sim \text{N}(\boldsymbol{a}_{t+1}, \boldsymbol{R}_{t+1}), \tag{4.2.1}$$

where D_t denotes the state of knowledge at time t [15]. From the prior information, forecasts are generated using the observation equation.

4.2.1 Forecasting One Step Ahead

The forecast quantity Y_{t+1} is a linear combination of normally distributed variables, $\boldsymbol{\theta}_{t+1}$ and ν_{t+1}, and is therefore also normally distributed. The forecast mean and variance are:

$$\begin{aligned}
\text{E}[Y_{t+1} \mid D_t] &= \text{E}[\boldsymbol{F}'_{t+1}\boldsymbol{\theta}_{t+1} + \nu_{t+1} \mid D_t] \\
&= \text{E}[\boldsymbol{F}'_{t+1}\boldsymbol{\theta}_{t+1} \mid D_t] + \text{E}[\nu_{t+1} \mid D_t] \\
&= \boldsymbol{F}'_{t+1}\boldsymbol{a}_{t+1} \\
&= f_{t+1}, \\
\text{Var}[Y_{t+1} \mid D_t] &= \text{Var}[\boldsymbol{F}'_{t+1}\boldsymbol{\theta}_{t+1} + \nu_{t+1} \mid D_t] \\
&= \text{Var}[\boldsymbol{F}'_{t+1}\boldsymbol{\theta}_{t+1} \mid D_t] + \text{Var}[\nu_{t+1} \mid D_t] \\
&= \boldsymbol{F}'_{t+1}\boldsymbol{R}_{t+1}\boldsymbol{F}_{t+1} + V_{t+1} \\
&= Q_{t+1}.
\end{aligned} \tag{4.2.2}$$

Thus the one-step ahead forecast distribution is given by $(Y_{t+1} \mid D_t) \sim \mathrm{N}(f_{t+1}, Q_{t+1})$ [15]. Note that we used the assumption that the observation disturbance, ν_{t+1}, is uncorrelated with the state, $\boldsymbol{\theta}_{t+1}$.

4.2.2 Forecasting k-Steps Ahead

Forecasting k-steps ahead requires the prior information to be projected into the future through repeated application of the evolution equation. Given information up to time t, D_t, and the state prior for time $t+1$, the prior for time $t+2$, $p(\boldsymbol{\theta}_{t+2} \mid D_t)$ is obtained by applying the evolution equation. We find:

$$\boldsymbol{\theta}_{t+2} = \boldsymbol{G}\boldsymbol{\theta}_{t+1} + \boldsymbol{\omega}_{t+2}, \qquad \boldsymbol{\omega}_{t+2} \sim \mathrm{N}(0, \boldsymbol{W}_{t+2}). \tag{4.2.3}$$

Linearity ensures that this two step ahead prior will be normal. The moments are given by

$$\begin{aligned}
\mathrm{E}[\boldsymbol{\theta}_{t+2} \mid D_t] &= \boldsymbol{G}\mathrm{E}[\boldsymbol{\theta}_{t+1} \mid D_t] + \mathrm{E}[\boldsymbol{\omega}_{t+2} \mid D_t] \\
&= \boldsymbol{G}\boldsymbol{a}_{t+1} \\
&= \boldsymbol{a}_t(2) \\
\mathrm{Var}[\boldsymbol{\theta}_{t+2} \mid D_t] &= \boldsymbol{G}\mathrm{Var}[\boldsymbol{\theta}_{t+1} \mid D_t]\boldsymbol{G}' + \mathrm{Var}[\boldsymbol{\omega}_{t+2} \mid D_t] \\
&= \boldsymbol{G}\boldsymbol{R}_{t+1}\boldsymbol{G}' + \boldsymbol{W}_{t+2} \\
&= \boldsymbol{R}_t(2).
\end{aligned} \tag{4.2.4}$$

Note that we used the assumption that the evolutional disturbance, $\boldsymbol{\omega}_{t+2}$, is uncorrelated with the state, $\boldsymbol{\theta}_{t+1}$.

This procedure can be extended to find the k-step ahead state predictions. The state prior for time $t+k$ (given information up to and including time t, D_t) is given by

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathrm{N}(\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)), \tag{4.2.5}$$

where, for $k \geq 2$, the mean and variance are given by

$$\begin{aligned} \boldsymbol{a}_t(k) &= \boldsymbol{G}^{k-1}\boldsymbol{a}_{t+1} \\ \boldsymbol{R}_t(k) &= \boldsymbol{G}^{k-1}\boldsymbol{R}_{t+1}\boldsymbol{G}'^{k-1} + \sum_{j=2}^{k} \boldsymbol{G}^{k-j}\boldsymbol{W}_{t+j}\boldsymbol{G}'^{k-j}. \end{aligned} \tag{4.2.6}$$

Note that the sum runs from $j = 2$. This is seen by observing that $\boldsymbol{R}_{t+1} = \mathrm{Var}[\boldsymbol{\theta}_{t+1} \mid D_t] = \boldsymbol{G}\mathrm{Var}[\boldsymbol{\theta}_t \mid D_t]\boldsymbol{G}' + \mathrm{Var}[\boldsymbol{\omega}_{t+1} \mid D_t] = \boldsymbol{G}\boldsymbol{R}_t\boldsymbol{G}' + \boldsymbol{W}_{t+1}$ and so we see that the stochastic evolution variance, $\boldsymbol{W}_{t+1}$, is already included in the state prior variance $\boldsymbol{R}_{t+1}$. Given this forecast for the state, the corresponding forecast for the observation series is obtained from the observation equation as

$$(Y_{t+k} \mid D_t) \sim \mathrm{N}(f_t(k), Q_t(k)), \tag{4.2.7}$$

where (by using the observation equation) the moments are defined as,

$$\begin{aligned} f_t(k) &= \mathrm{E}[Y_{t+k} \mid D_t] \\ &= \mathrm{E}[\boldsymbol{F}'_{t+k}\boldsymbol{\theta}_{t+k} + \nu_{t+k} \mid D_t] \\ &= \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k), \\ Q_t(k) &= \mathrm{Var}[Y_{t+k} \mid D_t] \\ &= \mathrm{Var}[\boldsymbol{F}'_{t+k}\boldsymbol{\theta}_{t+k} + \nu_{t+k} \mid D_t] \\ &= \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + V_{t+k}. \end{aligned} \tag{4.2.8}$$

It is now a good point to discuss covariance between forecasts. We will now determine an expression for the covariance between two observations in the series [15]. The covariance $\mathrm{Cov}(Y_{t+i}, Y_{t+j} \mid D_t) = Q_t(i, j)$, where, for $i > j$, the covariances are defined by

$$\begin{aligned} Q_t(i, j) &= \boldsymbol{F}'_{t+i}\boldsymbol{C}_t(i, j)\boldsymbol{F}_{t+j}, \\ \boldsymbol{C}_t(i, j) &= \boldsymbol{G}^{i-j}\boldsymbol{R}_t(j). \end{aligned} \tag{4.2.9}$$

To see this result consider two future values of the observation series, given information up to time t, D_t

$$\begin{aligned} Y_{t+i} &= \boldsymbol{F}'_{t+i}\boldsymbol{\theta}_{t+i} + \nu_{t+i}, \\ Y_{t+j} &= \boldsymbol{F}'_{t+j}\boldsymbol{\theta}_{t+j} + \nu_{t+j}, \end{aligned} \tag{4.2.10}$$

where $i > j$.

The covariance between the state at time $t+i$ and time $t+j$ can be derived by application of the system equation. For a constant evolution matrix $\boldsymbol{G}$

$$\begin{aligned}
\boldsymbol{\theta}_{t+i} &= \boldsymbol{G}\boldsymbol{\theta}_{t+i-1} + \boldsymbol{\omega}_{t+i} \\
&= \boldsymbol{G}(\boldsymbol{G}\boldsymbol{\theta}_{t+i-2} + \boldsymbol{\omega}_{t+i-1}) + \boldsymbol{\omega}_{t+i} \\
&= \boldsymbol{G}^2\boldsymbol{\theta}_{t+i-2} + \boldsymbol{G}\boldsymbol{\omega}_{t+i-1} + \boldsymbol{\omega}_{t+i} \\
&= \boldsymbol{G}^3\boldsymbol{\theta}_{t+i-3} + \boldsymbol{G}^2\boldsymbol{\omega}_{t+i-2} + \boldsymbol{G}\boldsymbol{\omega}_{t+i-1} + \boldsymbol{\omega}_{t+i} \\
&= \cdots = \boldsymbol{G}^j\boldsymbol{\theta}_{t+i-j} + \boldsymbol{G}^{j-1}\boldsymbol{\omega}_{t+i-j+1} + \cdots + \boldsymbol{\omega}_{t+i} \\
&= \boldsymbol{G}^j\boldsymbol{\theta}_{t+i-j} + \sum_{l=0}^{j-1} \boldsymbol{G}^l\boldsymbol{\omega}_{t+i-l} \\
&= \boldsymbol{G}^{i-j}\boldsymbol{\theta}_{t+j} + \sum_{l=0}^{i-j-1} \boldsymbol{G}^l\boldsymbol{\omega}_{t+2j-l} \qquad \text{(relabelling} \quad j = i-j) \\
&= \boldsymbol{G}^{i-j}\boldsymbol{\theta}_{t+j} + \boldsymbol{W}_t(i,j),
\end{aligned} \tag{4.2.11}$$

where $\boldsymbol{W}_t(i,j)$ is a linear combination of the innovation terms for times $t+j+1, \ldots, t+i$.

The covariance between these two observations is (noting that the covariances $\mathrm{Cov}(\nu_s, \nu_t)$, $\mathrm{Cov}(\boldsymbol{\omega}_s, \boldsymbol{\omega}_t)$ for all $s \neq t$, and $\mathrm{Cov}(\nu_s, \boldsymbol{\omega}_t)$ for all s, t are zero):

$$\begin{aligned}
\mathrm{Cov}(Y_{t+i}, Y_{t+j} \mid D_t) &= \mathrm{Cov}(\boldsymbol{F}'_{t+i}\boldsymbol{\theta}_{t+i} + \nu_{t+i}, \boldsymbol{F}'_{t+j}\boldsymbol{\theta}_{t+j} + \nu_{t+j} \mid D_t) \\
&= \mathrm{Cov}(\boldsymbol{F}'_{t+i}\boldsymbol{\theta}_{t+i}, \boldsymbol{F}'_{t+j}\boldsymbol{\theta}_{t+j} \mid D_t) + \mathrm{Cov}(\boldsymbol{F}'_{t+i}\boldsymbol{\theta}_{t+i}, \nu_{t+j} \mid D_t) \\
&+ \mathrm{Cov}(\nu_{t+i}, \boldsymbol{F}'_{t+j}\boldsymbol{\theta}_{t+j} \mid D_t) + \mathrm{Cov}(\nu_{t+i}, \nu_{t+j} \mid D_t) \\
&= \boldsymbol{F}'_{t+i}\mathrm{Cov}(\boldsymbol{\theta}_{t+i}, \boldsymbol{\theta}_{t+j} \mid D_t)\boldsymbol{F}_{t+j} \\
&+ \mathrm{Cov}(\boldsymbol{F}'_{t+i}\boldsymbol{G}^{i-j}\boldsymbol{\theta}_{t+j} + \boldsymbol{W}_t(i,j), \nu_{t+j} \mid D_t) \\
&= \boldsymbol{F}'_{t+i}\boldsymbol{C}_t(i,j)\boldsymbol{F}_{t+j}.
\end{aligned} \tag{4.2.12}$$

Note that on the fifth line of (4.2.12) we use the substitution $\boldsymbol{\theta}_{t+i} = \boldsymbol{G}^{i-j}\boldsymbol{\theta}_{t+j} + \boldsymbol{W}_t(i,j)$ and we then use the assumption that the observation disturbance, ν_{t+j}, is uncorrelated with the state, $\boldsymbol{\theta}_{t+j}$.

Using Equations (4.2.11) and (4.2.12) the desired covariance between observations

can be written as:

$$\begin{aligned}\boldsymbol{F}'_{t+i}\text{Cov}(\boldsymbol{\theta}_{t+i}, \boldsymbol{\theta}_{t+j} \mid D_t)\boldsymbol{F}_{t+j} &= \boldsymbol{F}'_{t+i}\text{Cov}(\boldsymbol{G}^{i-j}\boldsymbol{\theta}_{t+j} + \boldsymbol{W}_t(i,j), \boldsymbol{\theta}_{t+j} \mid D_t)\boldsymbol{F}_{t+j} \\ &= \boldsymbol{F}'_{t+i}\boldsymbol{G}^{i-j}\text{Var}(\boldsymbol{\theta}_{t+j} \mid D_t)\boldsymbol{F}_{t+j} \\ &= \boldsymbol{F}'_{t+i}\boldsymbol{G}^{i-j}\boldsymbol{R}_t(j)\boldsymbol{F}_{t+j}.\end{aligned} \tag{4.2.13}$$

As a special case (for the model structure that we will use to model the urine output time series, see Chapter 5), take

$$\boldsymbol{F}_t = \boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{4.2.14}$$

and note that

$$\boldsymbol{G}^{i-j} = \begin{pmatrix} 1 & i-j \\ 0 & 1 \end{pmatrix}. \tag{4.2.15}$$

Then we can express, for $i > j$

$$\text{Cov}(Y_i, Y_j \mid D_t) = \boldsymbol{F}'\boldsymbol{G}^{i-j}\boldsymbol{R}_t(j)\boldsymbol{F} = R_{\mu,t}(j) + (i-j)R_{\mu\beta,t}(j), \tag{4.2.16}$$

where

$$\boldsymbol{R}_t(j) = \begin{pmatrix} R_{\mu,t}(j) & R_{\mu\beta,t}(j) \\ R_{\mu\beta,t}(j) & R_{\beta,t}(j) \end{pmatrix}. \tag{4.2.17}$$

4.3 Joint Probabilities

Now that we have expressions for k-step ahead forecasts and for covariances between forecasts we can find joint probability distributions and hence calculate joint probabilities.

In Section 2.4 we considered a (transformed) regression model which we wanted to use to make forecasts about the urine output (per kilogram) for the next six hours for a patient. It is of interest to us to predict the joint probability that the next six urine outputs will all be below $\log(0.3) ml/kg$ (see Section 1.3), since this is an indication of the likelihood that a patient's kidneys will be in a state of severe oliguria, indicating that a patient's kidneys are not working sufficiently. This

requires a joint normal distribution of the form:

$$\boldsymbol{Z}_t \mid D_t = \begin{pmatrix} Z_{t+1} \\ \vdots \\ Z_{t+6} \end{pmatrix} \mid D_t \sim \mathrm{N}\left[\begin{pmatrix} f_t(1) \\ \vdots \\ f_t(6) \end{pmatrix}, \quad \begin{pmatrix} Q_t(1) & \dots & Q_t(1,6) \\ \vdots & \ddots & \vdots \\ Q_t(1,6) & \dots & Q_t(6) \end{pmatrix} \right]. \tag{4.3.1}$$

In shorthand

$$\boldsymbol{Z}_t \mid D_t \sim \mathrm{N}[\boldsymbol{f}_t, \boldsymbol{Q}_t], \tag{4.3.2}$$

where $\boldsymbol{f}_t$ is a (6×1) vector of the six point forecasts and $\boldsymbol{Q}_t \equiv [Q_t(i,j)]_{1 \le i \le 6, 1 \le j \le 6}$ is the (6×6) covariance matrix, where, for $i > j$, $Q_t(i,j) = \mathrm{Cov}(Z_{t+i}, Z_{t+j} \mid D_t)$ (see Equations (4.2.9)); and for $i = j$, $Q_t(i,j) = Q_t(i) = \mathrm{Var}(Z_{t+i} \mid D_t)$, where $Z_{t+i} = \log(Y_{t+i} + 0.1)$.

4.4 Posterior Distribution

The model likelihood, is the conditional forecast distribution evaluated at an observed value. It has the normal form

$$\begin{aligned} L(\boldsymbol{\theta}_t \mid Y_t = y_t, V_t) &\propto p(Y_t = y_t \mid \boldsymbol{\theta}_t, V_t) \\ &\sim \mathrm{N}(\boldsymbol{F}'\boldsymbol{\theta}_t, V_t). \end{aligned} \tag{4.4.1}$$

The prior information is combined with the likelihood using Bayes' theorem to yield the posterior distribution on the state,

$$p(\boldsymbol{\theta}_t \mid D_t) = p(\boldsymbol{\theta}_t \mid D_{t-1}, y_t) = \frac{p(Y_t = y_t \mid \boldsymbol{\theta}_t, V_t) p(\boldsymbol{\theta}_t \mid D_{t-1})}{p(Y_t = y_t)}. \tag{4.4.2}$$

For the dynamic model the state posterior is the product of two normal density functions yielding another normal density [15]. Recalling that $(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{a}_t, \boldsymbol{R}_t)$ we obtain

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &\propto p(Y_t = y_t \mid \boldsymbol{\theta}_t, V_t) p(\boldsymbol{\theta}_t \mid D_{t-1}) \\ &\propto \exp\{-0.5 V_t^{-1} (y_t - \boldsymbol{F}'\boldsymbol{\theta}_t)^2\} \times \exp\{-0.5 (\boldsymbol{\theta}_t - \boldsymbol{a}_t)' \boldsymbol{R}_t^{-1} (\boldsymbol{\theta}_t - \boldsymbol{a}_t)\} \\ &\propto \exp\{-0.5 (\boldsymbol{\theta}_t - \boldsymbol{m}_t)' \boldsymbol{C}_t^{-1} (\boldsymbol{\theta}_t - \boldsymbol{m}_t)\}, \end{aligned} \tag{4.4.3}$$

where the moments are defined as

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t e_t, \\ \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t, \\ \boldsymbol{A}_t &= \boldsymbol{R}_t \boldsymbol{F} / Q_t, \\ e_t &= y_t - f_t. \end{aligned} \tag{4.4.4}$$

The above result states that the posterior distribution $(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t)$ with moments defined in (4.4.4) [15]. This result is derived in Appendix A.

These moments are key to why dynamic models are more powerful than static linear models. The posterior mean $\boldsymbol{m}_t$ is the prior mean $\boldsymbol{a}_t$ plus a multiple of the one-step ahead forecast error. The amount of adjustment from the prior mean is determined by the adaptive factor, $\boldsymbol{A}_t$ [15]. The adaptive factor is determined by the size of the state prior variance, $|\boldsymbol{R}_t|$, and the observation variance $Q_t = \boldsymbol{F}' \boldsymbol{R}_t \boldsymbol{F} + V_t$. The larger the observation variance compared to the state prior variance, the smaller the adaptive factor and as $\boldsymbol{A}_t \to \boldsymbol{0}$ the posterior mean is approximately the prior mean, i.e. $\boldsymbol{m}_t \approx \boldsymbol{a}_t$. The intuition here is that (recalling that $(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{a}_t, \boldsymbol{R}_t)$ and $(Y_t \mid D_{t-1}) \sim \mathrm{N}(f_t, Q_t)$) if $Q_t \gg |\boldsymbol{R}_t|$ then we are more certain about the parameter values than we are about the forecasts. This means that large forecast errors are possible. Since the data up to time $t-1$ has given us an adequate state estimate (since here $Q_t \gg |\boldsymbol{R}_t|$) we do not want the estimate to be largely changed by a large forecast error and so intuitively $\boldsymbol{A}_t$ should be small. Conversely, if $|\boldsymbol{R}_t| \gg Q_t$ then the state prior variance is large compared to the observation variance. This means we are uncertain about the state parameter but the observation variance is small (compared to the state prior variance) and so the observation has a lot of useful information for the state and so adjustment from the prior to the posterior should reflect that useful observation.

State posterior variances are smaller than the corresponding prior variances because we have gained more information from the observation at time t. The only exceptions to this are two special cases [15]. One case is when $\boldsymbol{F} = 0$ when the observation y_t is completely uninformative on the state and the posterior variance is identical to the prior variance. This applies only to the regression components since trends and seasonal components have constant, nonzero regression vectors. The second case is when an observation is missing. Posterior moments are then equal to prior

moments. If y_t is missing then $p(\boldsymbol{\theta}_t \mid D_t) = p(\boldsymbol{\theta}_t \mid D_{t-1}, y_t) = p(\boldsymbol{\theta}_t \mid D_{t-1})$.

Once an observation is made, and posterior distribution is calculated, we repeat the cycles of prior to forecast to posterior to next prior. These stages characterise the routine analysis of the DLM and we now have the means to update our recurrence relations and forecast distributions.

The recurrence relations, for a constant and known evolution matrix $\boldsymbol{W}_t = \boldsymbol{W}$, and a known constant observational variance $V_t = V$, are described in Algorithm 1 (adapted from West and Harrison [1]).

4.5 Bayesian Updating of the DLM

To illustrate the essence of dynamic models we will consider how to update our beliefs using Bayes' rule. For simplicity and illustration purposes we will consider updating an initial prior normal distribution with data that follows a normal distribution with known precision.

We are interested in making inferences about the level of urine output for a patient, μ. In our dataset we have measurements of a patient's urine output, $\boldsymbol{y} = \{y_1, \ldots, y_n\}$. Assume that our given measurement system has known precision, λ, with unknown mean, μ, and $y_i \sim \mathrm{N}(\mu, \lambda^{-1})$, where each y_i is independently distributed. Thus,

$$
\begin{aligned}
l(\mu \mid \boldsymbol{y}, \lambda) &= \prod_{i=1}^{n} p(y_i \mid \mu, \lambda^{-1}), \\
&= (2\pi\lambda^{-1})^{-n/2} \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right], \\
&= (2\pi\lambda^{-1})^{-n/2} \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{n} [(y_i - \bar{y}) - (\mu - \bar{y})]^2 \right], \\
&= (2\pi\lambda^{-1})^{-n/2} \exp\left[-\frac{\lambda}{2} \left\{ \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2(\mu - \bar{y}) \sum_{i=1}^{n} (y_i - \bar{y}) + n(\mu - \bar{y}^2) \right\} \right], \\
&= (2\pi\lambda^{-1})^{-n/2} \exp\left[-\frac{n\lambda}{2} [s^2 + (\mu - \bar{y})^2] \right],
\end{aligned}
\tag{4.5.1}
$$

Algorithm 1 Recurrence Relations for known $\boldsymbol{W}$ and known V

Input: $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, $\boldsymbol{W}$, and V

1: At time t the posterior for the state is given by

$$(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t). \tag{4.4.5}$$

2: Set $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$.

3: The prior for the state at time $t+k$, for $k > 0$, given information up to time t, D_t, is given by

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathrm{N}(\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)), \tag{4.4.6}$$

where

$$\begin{aligned} \boldsymbol{a}_t(k) &= \boldsymbol{G}\boldsymbol{a}_t(k-1), \\ \boldsymbol{R}_t(k) &= \boldsymbol{G}\boldsymbol{R}_t(k-1)\boldsymbol{G}' + \boldsymbol{W}. \end{aligned} \tag{4.4.7}$$

4: The prior for the forecasts at time $t+k$, for $k > 0$, given information up to time t, D_t, is given by

$$(Y_{t+k} \mid D_t) \sim \mathrm{N}(f_t(k), Q_t(k)), \tag{4.4.8}$$

where

$$\begin{aligned} f_t(k) &= \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k), \\ Q_t(k) &= \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + V. \end{aligned} \tag{4.4.9}$$

5: Then, after observation y_{t+1} is observed we update Equation (4.4.5) to time $t+1$

$$(\boldsymbol{\theta}_{t+1} \mid D_{t+1}) \sim \mathrm{N}(\boldsymbol{m}_{t+1}, \boldsymbol{C}_{t+1}) \tag{4.4.10}$$

where we set $\boldsymbol{a}_{t+1}(0) = \boldsymbol{m}_{t+1}$ and $\boldsymbol{R}_{t+1}(0) = \boldsymbol{C}_{t+1}$, where

$$\begin{aligned} e_t(1) &= y_{t+1} - f_t(1), \\ \boldsymbol{A}_{t+1} &= \boldsymbol{R}_t(1)\boldsymbol{F}_{t+1}/Q_t(1), \\ \boldsymbol{m}_{t+1} &= \boldsymbol{a}_t(1) + \boldsymbol{A}_{t+1}e_t(1), \\ \boldsymbol{C}_{t+1} &= \boldsymbol{R}_t(1) - \boldsymbol{A}_{t+1}\boldsymbol{A}'_{t+1}Q_t(1). \end{aligned} \tag{4.4.11}$$

where $\bar{y} = \sum_{i=1}^{n} y_i/n$ and $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/n$. Thus we have

$$l(\mu \mid \boldsymbol{y}, \lambda) \propto \exp\left[-\frac{n\lambda}{2}(\mu - \bar{y})^2 \right]. \tag{4.5.2}$$

Assume that our prior uncertainty about μ can be described by a normal distribution, $\mu \sim \mathrm{N}(a, R)$, so that its pdf is

$$p(\mu \mid a, R) = \sqrt{\frac{R^{-1}}{2\pi}} \exp\left[-\frac{R^{-1}}{2}(\mu - a)^2 \right]. \tag{4.5.3}$$

Updating our prior beliefs through Bayes' rule, we obtain

$$\begin{aligned}
p(\mu \mid \bar{y}, \lambda, a, R) &\propto l(\mu \mid \boldsymbol{y}, \lambda) \times p(\mu \mid a, R), \\
&\propto \exp\left[-\frac{n\lambda}{2}(\mu - \bar{y})^2 \right] \exp\left[-\frac{R^{-1}}{2}(\mu - a)^2 \right], \\
&\propto \exp\left\{ -\frac{1}{2}[n\lambda(\mu - \bar{y})^2 + R^{-1}(\mu - a)^2] \right\}, \\
&\propto \exp\left\{ -\frac{1}{2}[n\lambda(\mu^2 - 2\bar{y}\mu + \bar{y}^2) + R^{-1}(\mu^2 - 2a\mu + a^2)] \right\}, \\
&\propto \exp\left\{ -\frac{1}{2}\left[(n\lambda + R^{-1})\left(\mu^2 - 2\mu \frac{n\lambda\bar{y} + aR^{-1}}{n\lambda + R^{-1}} \right) + \bar{y}^2 n\lambda + a^2 R^{-1} \right] \right\}.
\end{aligned} \tag{4.5.4}$$

By completing the quadratic and dropping terms that do not depend on μ, we obtain

$$p(\mu \mid \bar{y}, \lambda, a, R) \propto \exp\left[-\frac{n\lambda + R^{-1}}{2}\left(\mu - \frac{n\lambda\bar{y} + aR^{-1}}{n\lambda + R^{-1}} \right)^2 \right]. \tag{4.5.5}$$

This is the kernel of a normal distribution for μ with mean, m, and precision, C^{-1}, given by

$$\begin{aligned}
m &= \frac{n\lambda}{n\lambda + R^{-1}}\bar{y} + \frac{R^{-1}}{n\lambda + R^{-1}}a, \\
C^{-1} &= n\lambda + R^{-1}.
\end{aligned} \tag{4.5.6}$$

Notice that the posterior mean is a weighted average of the prior and the sample means, with weights depending on the number of observations, and the precisions of the prior and the likelihood. A key point to note is that for a sample size, n, large enough, the posterior mean will be dominated by the sample mean and the posterior precision will hardly be influenced by the initial prior precision. So, when enough relevant information is accumulated, the initial prior influence in the posterior is overpowered by that contained in the sample. However, this is not necessarily the

case when n is relatively small.

To see the role of the initial prior distribution, as more data is collected, we will consider updating from prior to posterior for four different sample sizes, $n = 1, 10, 25, 50$. We will assume that the initial prior distribution for μ is given by $\mu \sim \mathrm{N}(-4, 1/2)$ and that $\lambda = 1$ and that $\bar{y} = 0$ for all n.

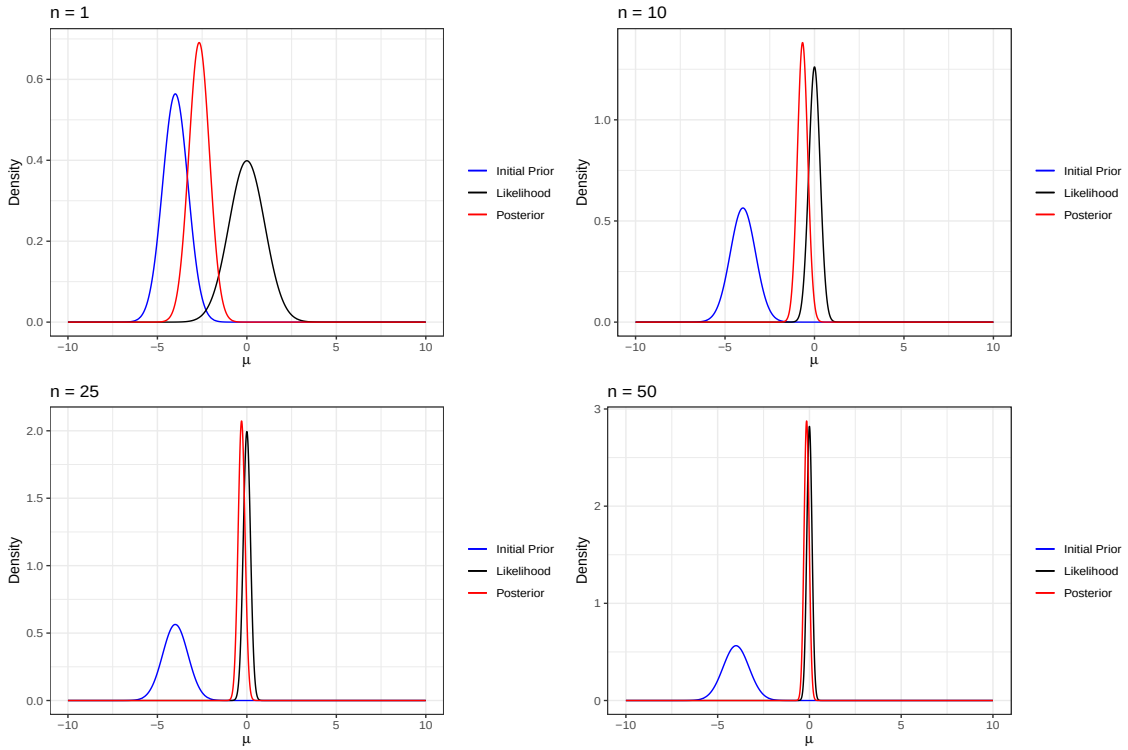

Figure 4.1: Role of initial prior distribution as n increases. Blue lines represent prior distributions, red lines represent posterior distributions, and black lines represent the likelihoods. We see that as the sample size increases, the posterior distribution resembles the likelihood more and more

From Figure 4.1 we see that for $n = 1$, representing the first update from prior to posterior, the prior and information from observation y_1 are quite different and hence the posterior lies in between both distributions (see Equations (4.5.6)). We also see that, as the sample size increases, the posterior distribution resembles the likelihood more and more, regardless of the initial prior. This illustrates a key point about DLMs and Bayesian analysis. The initial prior has little influence on the posterior distributions as n increases.

4.6 Component Forms

Building a suitable model for a time series is a difficult task in Bayesian modelling. In this section we will see several component forms and outline the mathematical structure of building more complex dynamic models.

4.6.1 Polynomial Trend Components

The simplest polynomial trend model is the first order polynomial model, or the level model. Observed series values are stochastically distributed about a time varying constant:

$$\begin{aligned} Y_t &= \mu_t + \nu_t, & \nu_t &\sim \mathrm{N}(0, V_t), \\ \mu_t &= \mu_{t-1} + \omega_t, & \omega_t &\sim \mathrm{N}(0, W_t). \end{aligned} \tag{4.6.1}$$

The system equation defines the level to be a simple random walk through time [15]. Here we see that the regression vector, $\boldsymbol{F}$, and the system evolution matrix, $\boldsymbol{G}$ are both 1 (see Equations (4.1.1)). A second order polynomial trend model allows for systematic growth or decline in level. That is, a model with both level and slope. The additional parameter quantifies the change in level over time,

$$\begin{aligned} Y_t &= \mu_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, \end{aligned} \tag{4.6.2}$$

where $\nu_t \sim \mathrm{N}(0, V_t)$ and $\boldsymbol{\omega}_t \sim \mathrm{N}(0, \boldsymbol{W}_t)$ where

$$\boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} & W_{\mu\beta t} \\ W_{\mu\beta t} & W_{\beta t} \end{pmatrix}.$$

The state vector $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$, where μ_t and β_t represent the level and the rate of change in level at time t, respectively. Here we see that the regression vector and system matrix are given by (see Equations (4.1.1))

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{4.6.3}$$

Higher order polynomial DLMs are defined by extension [16]. The class of polynomial models of order p have p-dimensional state vectors, $\boldsymbol{F} = (1, 0, \ldots, 0)'$ and a $(p \times p)$ evolution matrix $\boldsymbol{G}$ given by the $(p \times p)$ Jordan form (diagonal and superdiagonal entries are one, and all other entries are zero)

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 \\ 0 & 1 & 1 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}. \tag{4.6.4}$$

Then we have

$$\begin{aligned} Y_t &= \theta_{1t} + \nu_t \\ \theta_{1t} &= \theta_{1,t-1} + \theta_{2,t-1} + \omega_{1t} \\ \theta_{2t} &= \theta_{2,t-1} + \theta_{3,t-1} + \omega_{2t} \\ &\vdots \qquad\qquad \vdots \\ \theta_{jt} &= \theta_{j,t-1} + \theta_{j+1,t-1} + \omega_{jt} \\ &\vdots \qquad\qquad \vdots \\ \theta_{pt} &= \theta_{p,t-1} + \omega_{pt}. \end{aligned} \tag{4.6.5}$$

The element θ_{jt} of the state vector represents the jth difference of the trend in the series at time t.

4.6.2 Seasonal Component Models

Modelling seasonal patterns in time series requires a periodic component form. One way of representing such a form is to isolate an underlying trend from periodic movement about that trend [15]. Over a complete cycle the effects sum to zero since the trend, which is the average of the factors over the cycle, contains the overall series movement over that time span. The seasonal effects model defines parameters to measure seasonal departures from a trend. For example, a set of seasonal factors 1, 1.4, 0.8, and 1.2 is equivalent to a trend of 1.1 and seasonal effects -0.1, 0.3, -0.3, and 0.1.

Suppose that we have quarterly data Y_t, for $t = 1, 2, \ldots$ on the sales of a store

which show a cyclic behaviour [17]. Assume for simplicity that the series has zero mean (a non-zero mean is modelled separately by adding a first order polynomial component, or if there is an underlying level that may grow or decay over time, by adding a second order polynomial component, (see Section 4.7)), so we consider the series as purely seasonal. Here we describe the quarterly data by introducing seasonal deviations from the mean, expressed by different coefficients α_i for different quarters, $i = 1, 2, 3, 4$. So, if Y_{t-1} refers to the first quarter of the year and Y_t to the second quarter, we assume

$$\begin{aligned} Y_{t-1} &= \alpha_1 + \nu_{t-1} \\ Y_t &= \alpha_2 + \nu_t \\ Y_{t+1} &= \alpha_3 + \nu_{t+1} \end{aligned} \tag{4.6.6}$$

and so on. This model can be written as a DLM by writing $\boldsymbol{\theta}_{t-1} = (\alpha_1, \alpha_4, \alpha_3, \alpha_2)'$ and $\boldsymbol{F} = (1, 0, 0, 0)'$.

The state equation must rotate the components of $\boldsymbol{\theta}_{t-1}$ into the vector $\boldsymbol{\theta}_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)'$, so that $Y_t = \boldsymbol{F}\boldsymbol{\theta}_t + \nu_t = \alpha_2 + \nu_t$. This required permutation of the state vector can be obtained by a permutation matrix $\boldsymbol{G}$ defined by

$$\boldsymbol{G} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.6.7}$$

In general, for a seasonal model with n seasons, we have $\boldsymbol{F} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{G} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}, \tag{4.6.8}$$

an $(n \times n)$ matrix.

4.6.3 Regression Components

Regression components are easily added to a DLM through the regression vector. For example, regression on CVP (central venous pressure) with an underlying level, looks like

$$\begin{aligned} Y_t &= \mu_t + \beta_t \mathrm{CVP}_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, \end{aligned} \tag{4.6.9}$$

where the regression vector $\boldsymbol{F}_t = (1, \mathrm{CVP}_t)$ and the system matrix is, $\boldsymbol{G} = \boldsymbol{I}_2$, the (2×2) identity matrix. The regression coefficients have a simple random walk evolution. Regressing on several variables, $X_1, ..., X_q$, has the form

$$\begin{aligned} Y_t &= \beta_{1t} X_{1t} + \cdots + \beta_{qt} X_{qt} + \nu_t, \\ \beta_{it} &= \beta_{i,t-1} + \omega_{it}, \qquad i = 1, \ldots, q, \end{aligned} \tag{4.6.10}$$

where the regression vector $\boldsymbol{F}_t = \boldsymbol{X}_t = (X_{1t}, \ldots, X_{qt})'$, and the system matrix is the $(q \times q)$ identity matrix, $\boldsymbol{G} = \boldsymbol{I}_q$ [15].

4.7 Superposition: Block Structured Models

Derived from the principle that any linear combination of independent linear models is a linear model, the superposition principle gives us a means of constructing more complex dynamic models. Component forms for trend, seasonal, and regression are the building blocks for constructing models of complex time series behaviour. The linear additive structure of the (normal) DLM enables component models to be brought together in a straightforward way.

A linear growth plus regression model is a DLM that is often used when the dataset has many explanatory variables describing a response variable. We could represent this series as

$$Y_t = Y_{Tt} + Y_{Rt} + \nu_t, \tag{4.7.1}$$

where

$$\begin{aligned} Y_{Tt} &= \boldsymbol{F}'_{Tt}\boldsymbol{\theta}_{Tt}, \\ \boldsymbol{\theta}_{Tt} &= \boldsymbol{G}_{Tt}\boldsymbol{\theta}_{T,t-1} + \boldsymbol{\omega}_{Tt}, \end{aligned} \tag{4.7.2}$$

represents the trend component and

$$\begin{aligned} Y_{Rt} &= \boldsymbol{F}'_{Rt}\boldsymbol{\theta}_{Rt}, \\ \boldsymbol{\theta}_{Rt} &= \boldsymbol{G}_{Rt}\boldsymbol{\theta}_{R,t-1} + \boldsymbol{\omega}_{Rt}, \end{aligned} \tag{4.7.3}$$

represents the regression component. From Equations (4.7.2) and (4.7.3) we can see that the state variables for each component evolve independently (see Figure 3.4). The observation is a linear combination of these components

$$\begin{aligned} Y_t &= \boldsymbol{F}'_{Tt}\boldsymbol{\theta}_{Tt} + \boldsymbol{F}'_{Rt}\boldsymbol{\theta}_{Rt} + \nu_t, \\ &= \boldsymbol{F}'_t\boldsymbol{\theta}_t + \nu_t, \end{aligned} \tag{4.7.4}$$

where $\boldsymbol{F}_t = (\boldsymbol{F}_{Tt}, \boldsymbol{F}_{Rt})'$ and $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{Tt}, \boldsymbol{\theta}_{Rt})'$. For the system equation we can write

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{4.7.5}$$

where $\boldsymbol{\omega}_t = (\boldsymbol{\omega}_{Tt}, \boldsymbol{\omega}_{Rt})'$. The evolution and system variance matrices have the diagonal form:

$$\boldsymbol{G}_t = \begin{pmatrix} \boldsymbol{G}_{Tt} & 0 \\ 0 & \boldsymbol{G}_{Rt} \end{pmatrix}, \quad \boldsymbol{W}_t = \begin{pmatrix} \boldsymbol{W}_{Tt} & 0 \\ 0 & \boldsymbol{W}_{Rt} \end{pmatrix}. \tag{4.7.6}$$

Many complex dynamic models can be built in the same way. The following holds [1]

Theorem 4.7.0.1. Consider h time series Y_{it}, generated by DLMs

$$M_i = \{\boldsymbol{F}_{it}, \boldsymbol{G}_{it}, V_{it}, \boldsymbol{W}_{it}\}, \qquad i = 1, \ldots, h. \tag{4.7.7}$$

This quadruple defines the model relating Y_{it} to the $(n_i \times 1)$ parameter vector, $\boldsymbol{\theta}_{it}$, at time t and the observation and evolution error series are ν_{it} and $\boldsymbol{\omega}_{it}$, respectively. The state vectors are distinct, and for all $i \neq j$, the series ν_{it} and $\boldsymbol{\omega}_{it}$ are temporally independent and mutually independent.

Then the series

$$Y_t = \sum_{i=1}^{h} Y_{it}, \tag{4.7.8}$$

follows the n-dimensional DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\}$, where $n = n_1 + \cdots + n_h$ and the quadruple and the state vector, $\boldsymbol{\theta}_t$, are given by

$$\boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{1t} \\ \vdots \\ \boldsymbol{\theta}_{ht} \end{pmatrix}, \quad \boldsymbol{F}_t = \begin{pmatrix} \boldsymbol{F}_{1t} \\ \vdots \\ \boldsymbol{F}_{ht} \end{pmatrix}, \tag{4.7.9}$$

$$\boldsymbol{G}_t = \begin{pmatrix} \boldsymbol{G}_{1t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{G}_{ht} \end{pmatrix}, \quad \boldsymbol{W}_t = \begin{pmatrix} \boldsymbol{W}_{1t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{W}_{ht} \end{pmatrix}, \tag{4.7.10}$$

and

$$V_t = \sum_{i=1}^{h} V_{it}. \tag{4.7.11}$$

4.8 Block Discounting

In the preceding analysis we have used knowledge of the system evolution covariances, $\boldsymbol{W}_t$. In practice this evolution variance will be unknown and difficult to specify. The system evolution covariance matrix can be updated in the same manner that the parameters are updated. We could specify an initial prior distribution and then, using Bayes' theorem, update this distribution when observational information becomes available. However, this gets very difficult to implement on a routine basis, but fortunately a practical solution called information discounting exists to capture the evolution of the system covariances [15]. Information discounting is an important method to handle models with unknown evolution variances, $\boldsymbol{W}_t$. This practical solution captures the evolution of the system variance and a Bayesian learning approach.

Earlier we noted that as information ages its value diminishes. This ageing process

is modelled in the DLM through the system evolution. Information discounting increases uncertainty when forecasting into the future. Consider Equations (4.1.1), with $G = 1$, in particular the system evolution equation

$$\text{System Evolution Equation:} \qquad \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}(0, \boldsymbol{W}_t). \tag{4.8.1}$$

The mechanism of the dynamic linear model is to add extra variance to the state posterior distribution to yield the prior for the next time.

$$\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) + \boldsymbol{W}_t. \tag{4.8.2}$$

Another way to model information loss is in percentage or in discount terms [15]. In other words, we might want to quantify the loss of information as an $\alpha\%$ increase in uncertainty as information ages by one time period, where $\alpha \geq 0$. We could write

$$\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = (1 + \alpha)\mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}). \tag{4.8.3}$$

To see that Equations (4.8.2) and (4.8.3) are equivalent set $\boldsymbol{W}_t = \alpha\mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$. Here we define the discount factor $\delta = (1+\alpha)^{-1}$. The discount factor varies between 0 and 1 and for a discount factor $\delta \in (0, 1]$ the information loss through the evolution process is summarised as

$$\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \delta^{-1}\mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}). \tag{4.8.4}$$

For a 15% information loss, $\delta \approx 0.87$ and when $\delta = 1$ we have $\mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$, that is, the prior variance at time t is the same as the posterior variance at time $t-1$. This is practically unrealistic since forecasting to the future always brings additional uncertainty. The idea of information discounting is given information up to time t the prior variance for time $t+1$ should be larger than the posterior variance for time t. In the absence of further knowledge after time t this captures the essence of future uncertainty [15].

The state prior variance at any time is computed as a function of the most recent posterior variance determined by a discount factor, $\delta \in (0, 1]$. The discount factor represents the amount of information loss attributed to future advancement, more generally (recalling $(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$),

$$\begin{aligned} \mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) &= \delta^{-1}\boldsymbol{G}\mathrm{Var}(\boldsymbol{\theta}_{t-1} \mid D_{t-1})\boldsymbol{G}' \\ &= \delta^{-1}\boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}'. \end{aligned} \tag{4.8.5}$$

In addition we have (see Equation (4.4.7) with $k = 1$ and replacing t with $t-1$ and replacing $\boldsymbol{W}$ with $\boldsymbol{W}_t$),

$$\boldsymbol{R}_t = \text{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}' + \boldsymbol{W}_t = \boldsymbol{P}_t + \boldsymbol{W}_t, \tag{4.8.6}$$

where $\boldsymbol{C}_{t-1}$ is the posterior variance at time $t-1$. The matrix $\boldsymbol{P}_t = \boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}'$ can be thought of as the prior variance in a DLM with no evolution error, i.e. in a model with $\boldsymbol{W}_t = \boldsymbol{0}$, where the state vector is stable and has no stochastic variation. This situation is practically useless, however, it can be assumed that $\boldsymbol{R}_t = \frac{\boldsymbol{P}_t}{\delta}$ for $\delta \in (0, 1]$, and so the prior variance at time t is that of a model without stochastic error times a factor which inflates such variance. When $\delta = 1$ we have a static model. Using $\boldsymbol{R}_t = \text{Var}(\boldsymbol{\theta}_t \mid D_{t-1}) = \boldsymbol{P}_t + \boldsymbol{W}_t$ and $\boldsymbol{R}_t = \frac{\boldsymbol{P}_t}{\delta}$, we have

$$\boldsymbol{W}_t = \frac{1-\delta}{\delta}\boldsymbol{P}_t. \tag{4.8.7}$$

So we see that discounting variances is equivalent to setting the evolution variance as a proportion of the posterior variance,

$$\boldsymbol{W}_t = \left(\frac{1}{\delta} - 1\right)\boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}'. \tag{4.8.8}$$

Low values of the discount factor δ are consistent with high variability in the $\boldsymbol{\theta}_t$ sequence, while high values, with $\delta \geq 0.9$ are more practically useful [16]. This representation is general and for models with multiple components the recommended discount strategy is to proceed component by component. Separate discount factors are specified for each component and individual component evolution variance matrices computed. The overall state evolution variance matrix is then set to the block diagonal composition of these individual elements. For example, in the trend plus regression component model considered in Section 4.7, we could define two discount factors for these components, δ_T and δ_R. The evolution covariance matrix would be

$$\begin{aligned} \boldsymbol{W}_t &= \begin{pmatrix} \boldsymbol{W}_T & 0 \\ 0 & \boldsymbol{W}_R \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{1}{\delta_T} - 1\right)\boldsymbol{G}_T\boldsymbol{C}_{T,t-1}\boldsymbol{G}'_T & 0 \\ 0 & \left(\frac{1}{\delta_R} - 1\right)\boldsymbol{G}_R\boldsymbol{C}_{R,t-1}\boldsymbol{G}'_R \end{pmatrix}, \end{aligned} \tag{4.8.9}$$

where $\boldsymbol{C}_{T,t-1}$ is the posterior covariance matrix for the trend component at time $t-1$, and $\boldsymbol{C}_{R,t-1}$ is the posterior covariance matrix for the regression component at

time $t-1$.

Using block discounting to structure the evolution covariance allows for separate discount factors. This allows for separate components of the DLM to have different variability inflations and to evolve more or less quickly than other components. The trend variation could be seen to be more stable than regression variation, for example.

4.8.1 Practical Discount Strategy

Discounting is a practical solution that captures the evolution of the system variance and a Bayesian learning approach. When forecasting one-step ahead we do not need to compute $\boldsymbol{W}_t$ explicitly since $\boldsymbol{R}_t = \boldsymbol{P}_t/\delta$. However, when forecasting k-steps ahead at time t it is not the case that repeat application with the same discount factor will produce the relevant sequence of variance matrices [1]. For example, consider a DLM with $\boldsymbol{F}_t = \boldsymbol{F}$ and $\boldsymbol{G}_t = \boldsymbol{G}$ with one discount factor, δ, for each of its components, repeated application of the system equation, given information up to time t, D_t (see Equations (4.1.1) with $\boldsymbol{G}_t = \boldsymbol{G}$) and using Equation (4.8.8) leads to

$$\begin{aligned}
\text{Var}(\boldsymbol{\theta}_{t+1} \mid D_t) &= \text{Var}(\boldsymbol{G}\boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1} \mid D_t) = \boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}' + \boldsymbol{W}_{t+1} \\
&= \boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}' + \left(\frac{1}{\delta} - 1\right)\boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}' = \frac{\boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}'}{\delta} = \boldsymbol{R}_t(1), \\
\text{Var}(\boldsymbol{\theta}_{t+2} \mid D_t) &= \text{Var}(\boldsymbol{G}\boldsymbol{\theta}_{t+1} + \boldsymbol{\omega}_{t+2} \mid D_t) = \boldsymbol{G}\frac{\boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}'}{\delta}\boldsymbol{G}' + \boldsymbol{W}_{t+2} \\
&= \frac{\boldsymbol{G}^2\boldsymbol{C}_t\boldsymbol{G}^{2'}}{\delta} + \left(\frac{1}{\delta} - 1\right)\boldsymbol{G}\boldsymbol{R}_{t+1}\boldsymbol{G}' \\
&= \frac{\boldsymbol{G}^2\boldsymbol{C}_t\boldsymbol{G}^{2'}}{\delta} + \left(\frac{1}{\delta} - 1\right)\frac{\boldsymbol{G}^2\boldsymbol{C}_t\boldsymbol{G}^{2'}}{\delta} = \frac{\boldsymbol{G}^2\boldsymbol{C}_t\boldsymbol{G}'^2}{\delta^2} = \boldsymbol{R}_t(2).
\end{aligned} \tag{4.8.10}$$

This process can be repeated for up to k-steps into the future and we obtain

$$\text{Var}(\boldsymbol{\theta}_{t+k} \mid D_t) = \boldsymbol{R}_t(k) = \frac{\boldsymbol{G}^k\boldsymbol{C}_t\boldsymbol{G}'^k}{\delta^k}. \tag{4.8.11}$$

The use of δ^k as a discount factor k-steps ahead implies an exponential decay in information, and this is not strictly consistent with the DLM in which the information decays arithmetically through the addition of future evolution error variance matrices [1]. Hence, this discount approach must be adapted when forecasting more

than one-step ahead.

A practical approach is suggested in West and Harrison (1997) [1]. This approach assumes that the one-step ahead evolution variance matrix is appropriate for forecasting k-steps into the future, determining a constant step-ahead variance matrix. The resulting discount procedure is as follows.

1. Given $(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t)$, calculate $\boldsymbol{W}_{t+1} = \boldsymbol{P}_{t+1}(1-\delta)/\delta$, where $\boldsymbol{P}_{t+1} = \boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}'$.

2. In forecasting k-steps ahead, adopt the conditionally constant variance

$$\mathrm{Var}(\boldsymbol{\omega}_{t+k} \mid D_t) = \boldsymbol{W}_{t+k} = \boldsymbol{W}_{t+1}, \qquad (k = 1, ...). \tag{4.8.12}$$

   Thus, step-ahead forecast distributions will be based on the addition of evolution errors with the same variance matrix, $\boldsymbol{W}_{t+1}$ for all k.

3. When Y_{t+1} is observed the posterior $(\boldsymbol{\theta}_{t+1} \mid D_{t+1})$ can be derived and then $\boldsymbol{P}_{t+2} = \boldsymbol{G}\boldsymbol{C}_{t+1}\boldsymbol{G}'$ and thus $\boldsymbol{W}_{t+2}$ can be calculated. Thus forecasting ahead from time $t+1$, we have

$$\mathrm{Var}(\boldsymbol{\omega}_{t+k} \mid D_{t+1}) = \boldsymbol{W}_{t+k+1} = \boldsymbol{W}_{t+2}, \qquad (k = 1, ...). \tag{4.8.13}$$

4. Continue in this manner at time $t+2$, and so on.

Prior to Section 4.8 we assumed that $\boldsymbol{W}_t$ was known, and constant, for all times t. This implied that the evolution matrices were independent of the history of the series. Now, with the proposed discount strategy, the evolution matrices have been modified to depend on the current state of information, D_t, and this allows a learning discount strategy to be applied representing the evolution matrices as a proportion of the most recent state posterior variance. This allows DLMs to express more (less) uncertainty at times where we are less (more) certain about the values of the parameters.

4.9 Variance Learning

So far we have assumed a known, constant observational variance $V_t \equiv V$. In many applications the observation variance will not be constant over the full range of a time series. Similarly to the evolution variance a learning mechanism is necessary. According to Bayes' rule,

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{4.9.1}$$

and in order to fully know the posterior distribution we must be able to perform the integral appearing in the denominator. This is sometimes unfeasible but fortunately, in some circumstances, performing the integral in the denominator can be avoided. One way to avoid doing this integral is to ensure that $p(\boldsymbol{\theta} \mid \boldsymbol{x})$ and $p(\boldsymbol{\theta})$ have the same functional form, i.e. they belong to the same family.

Definition 4.9.0.1. A family $\mathscr{F}$ of probability distributions on Θ is said to be conjugate for a likelihood function $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ if, for every prior $p \in \mathscr{F}$, the posterior distribution, $p(\boldsymbol{\theta} \mid \boldsymbol{x})$, also belongs to $\mathscr{F}$.

Using a conjugate prior distribution an analytical solution to the learning mechanism for the observational variance exists [15]. This involves using normal-gamma conjugate analysis. Defining the constant unknown variance DLM, in terms of the precision $\phi_t = V_t^{-1}$, as

$$\begin{aligned} \text{Observation Equation:} \quad & Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, & \nu_t &\sim \mathrm{N}(0, \phi_t^{-1}), \\ \text{System Evolution Equation:} \quad & \boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \mathrm{N}(0, \boldsymbol{W}_t^*\phi_t^{-1}). \end{aligned} \tag{4.9.2}$$

The scaling of the system disturbance covariance by the unknown observation variance is necessary for a conjugate analysis [15]. Setting $\boldsymbol{W}_t = \boldsymbol{W}_t^*\phi_t^{-1}$ recovers the normal form of the system equation as in Equations (4.1.1).

4.9.1 Prior Information

At time t we now have two distributions describing the prior information given the information up to time $t-1$, D_{t-1}. One describing the state and the other describing

the scale parameter

$$\begin{aligned}(\boldsymbol{\theta}_t \mid D_{t-1}, \phi_t) &\sim \mathrm{N}(\boldsymbol{a}_t, \boldsymbol{R}_t^* \phi_t^{-1}), \\ (\phi_t \mid D_{t-1}) &\sim \mathrm{Ga}(n_{t-1}/2, d_{t-1}/2).\end{aligned} \tag{4.9.3}$$

Setting $\boldsymbol{R}_t = \boldsymbol{R}_t^* \phi_t^{-1}$ recovers the normal form of the prior as in Equation (4.2.1) (with t replaced with $t+1$). The parameters of the gamma prior on the scale represent the degrees of freedom, n_{t-1}, and sums of squared errors, d_{t-1}, with mean equal to the ratio of these quantities, n_{t-1}/d_{t-1}.

4.9.2 Forecasts

The conditional forecast distribution is, like before, normal but now we make the scale factor explicit,

$$(Y_t \mid D_{t-1}, \phi_t) \sim \mathrm{N}(f_t, Q_t^* \phi_t^{-1}), \tag{4.9.4}$$

where the observation scale-free forecast variance is

$$Q_t^* = 1 + \boldsymbol{F}_t' \boldsymbol{R}_t^* \boldsymbol{F}_t. \tag{4.9.5}$$

From standard normal-gamma theory, unconditionally the forecast has a T distribution on n_{t-1} degrees of freedom [15],

$$(Y_t \mid D_{t-1}) \sim \mathrm{T}_{n_{t-1}}(f_t, Q_t), \tag{4.9.6}$$

where the mean and scale parameter are given by

$$\begin{aligned}f_t &= \boldsymbol{F}_t' \boldsymbol{a}_t \\ Q_t &= S_{t-1} + \boldsymbol{F}_t' \boldsymbol{R}_t^* \boldsymbol{F}_t S_{t-1},\end{aligned} \tag{4.9.7}$$

where $S_{t-1} = d_{t-1}/n_{t-1}$.

We note that the expected forecast, f_t, is the same as the known constant variance scenario. The observational variance learning mechanism does not change the point forecasts. The aim of this mechanism being that the observational variance can learn and evolve simultaneously with the other parameters and hence, when necessary, the prediction intervals can be smaller (when more confident) or larger (when more uncertain) than in the constant observational variance case, depending on the

state of the series at different times. This learning mechanism illustrates a better understanding of the uncertainty in the system rather than assuming the observational variance to be constant throughout. Notice that the forecast scale has the same algebraic form as in the case of the known constant variance case in Equations (4.4.9) (with $k = 1$ and t replaced by $t + 1$), except the now unknown observational variance is estimated by its prior expected value, $S_{t-1} = d_{t-1}/n_{t-1}$ [15].

4.9.3 Posterior Information

The posterior distribution on the scale parameter is obtained by using Bayes' theorem

$$\begin{aligned} p(\phi_t \mid D_t) = p(\phi_t \mid D_{t-1}, y_t) &\propto p(y_t \mid D_{t-1}, \phi_t) p(\phi_t \mid D_{t-1}) \\ &\propto (\phi_t^{-1} Q_t^*)^{-1/2} \exp\left\{ -\frac{1}{2} (\phi_t^{-1} Q_t^*)^{-1/2} (y_t - f_t)^2 \right\} \\ &\times \phi_t^{n_{t-1}/2-1} \exp\left\{ -\frac{\phi_t}{2} d_{t-1} \right\} \\ &\propto \phi_t^{(n_{t-1}-1)/2} \exp\left\{ -\frac{\phi_t}{2} \left[\frac{e_t^2}{Q_t^*} + d_{t-1} \right] \right\} \\ &\sim \mathrm{Ga}\left(\frac{1}{2}(n_{t-1}+1), \frac{1}{2} \left[\frac{e_t^2}{Q_t^*} + d_{t-1} \right] \right) \\ &\sim \mathrm{Ga}(n_t/2, d_t/2), \end{aligned} \tag{4.9.8}$$

where

$$\begin{aligned} n_t &= n_{t-1} + 1 \\ d_t &= \frac{e_t^2}{Q_t^*} + d_{t-1}. \end{aligned} \tag{4.9.9}$$

The degrees of freedom parameter increases by one as an additional piece of information has been processed, and the rate parameter is incremented by the square of the (scaled) forecast error [15]. The expected value for the posterior precision at time t is

$$\mathrm{E}[\phi_t \mid D_t] = \frac{n_t/2}{d_t/2} = \frac{n_t}{d_t} = \frac{1}{S_t}. \tag{4.9.10}$$

Hence the expected value for the posterior variance at time t is S_t, and S_0 is the prior estimate for the observational variance at time $t = 0$. The intuition of Equations

(4.9.9) being that if the squared forecast error e_t^2 is sufficiently small compared to the scale-free forecast variance Q_t^* then this indicates that the observation at time t is in accord with the forecast distribution (4.9.4), i.e., y_t is not an outlier and is in the range of values where we expected it to be. This in accord forecast is reflected in the updating of the observational variance. In the limiting case if $e_t = 0$ then $d_t = d_{t-1}$ and the posterior variance at time t, $S_t = d_{t-1}/(n_{t-1}+1) < S_{t-1} = d_{t-1}/n_{t-1}$. This is as expected because we want our prediction intervals to be shorter when we are more confident because our forecasts errors are small and our forecasts are in accord to what we expect (e.g. in stable periods). Conversely, we want our prediction intervals to be larger, expressing more uncertainty, when a forecast is not in accord to what we expect. Moreover, the posterior variance at time t will be smaller than the posterior variance at time $t-1$ precisely when

$$\begin{aligned} \frac{d_t}{n_t} &< \frac{d_{t-1}}{n_{t-1}} \\ \implies n_{t-1}d_t &< d_{t-1}(n_{t-1}+1) \\ \implies n_{t-1}\left(\frac{e_t^2}{Q_t^*} + d_{t-1}\right) &< d_{t-1}(n_{t-1}+1) \\ \implies \frac{e_t^2}{Q_t^*} &< \frac{d_{t-1}}{n_{t-1}} = S_{t-1}. \end{aligned} \tag{4.9.11}$$

Note that in many texts [1,3,4,5], until a variance discounting method is adapted (see Section 4.9.4), ϕ_t is written without the time subscript, ϕ. The reason for this can be seen by considering the prior distribution at time $t+1$

$$(\phi_{t+1} \mid D_t) \sim \text{Ga}(n_t/2, d_t/2). \tag{4.9.12}$$

We can see that the expected value for the prior precision at time $t+1$ is the same as the expected value for the posterior precision at time t (see Equations 4.9.8), as expected. We can also observe that

$$\text{Var}(\phi_{t+1} \mid D_t) = 2\frac{n_t}{d_t^2} = \text{Var}(\phi_t \mid D_t). \tag{4.9.13}$$

Indicating that the prior information at time $t+1$ does not exhibit any additional uncertainty compared to the posterior at time t. This is unrealistic as the passage of time always brings additional uncertainty. For this reason many texts drop the subscript indicating that there is no need to distinguish between priors and posteriors until a variance discounting mechanism is introduced. However, we will use sub-

scripts to distinguish between the unknown precision posteriors at different times. Since the posterior at time t will have a different point estimate and variance to the posterior at time $t+1$, it is useful to distinguish between these with a subscript.

To obtain the posterior for the state we consider the joint distribution with an observation. Since both conditional distributions, $(Y_t \mid D_t, \phi_t)$ and $(\boldsymbol{\theta}_t \mid D_t, \phi_t)$ are normal their joint distribution is also normal. The covariance of the state and the observation at time t, given information up to time $t-1$, D_{t-1}, and ϕ_t, is derived using the observation equation

$$\begin{aligned} \mathrm{Cov}(Y_t, \boldsymbol{\theta}_t \mid D_{t-1}, \phi_t) &= \mathrm{Cov}(\boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \boldsymbol{\theta}_t \mid D_{t-1}, \phi_t) \\ &= \boldsymbol{F}_t' \mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}, \phi_t) = \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1}, \end{aligned} \tag{4.9.14}$$

and similarly

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\theta}_t, Y_t \mid D_{t-1}, \phi_t) &= \mathrm{Cov}(\boldsymbol{\theta}_t, \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t \mid D_{t-1}, \phi_t) = \mathrm{Var}(\boldsymbol{\theta}_t \mid D_{t-1}, \phi_t)\boldsymbol{F}_t \\ &= \boldsymbol{R}_t^* \phi_t^{-1} \boldsymbol{F}_t = \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} = (\boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1})'. \end{aligned} \tag{4.9.15}$$

Hence the joint distribution is given by

$$\begin{pmatrix} Y_t \\ \boldsymbol{\theta}_t \end{pmatrix} \mid D_{t-1}, \phi_t \sim \mathrm{N}\left[\begin{pmatrix} f_t \\ \boldsymbol{a}_t \end{pmatrix}, \begin{pmatrix} \phi_t^{-1} Q_t^* & \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} \\ \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} & \boldsymbol{R}_t^* \phi_t^{-1} \end{pmatrix} \right]. \tag{4.9.16}$$

Using properties of the multivariate normal distribution, conditioning the state on the observed value yields a normal distribution which is given by

$$(\boldsymbol{\theta}_t \mid D_{t-1}, Y_t = y_t, \phi_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t^* \phi_t^{-1}), \tag{4.9.17}$$

where the moments are updated from their prior values with the scale conditioning made explicit and are given by (see Appendix C)

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} (\phi_t^{-1} Q_t^*)^{-1} (y_t - f_t) \\ &= \boldsymbol{a}_t + \boldsymbol{R}_t^* \boldsymbol{F}_t e_t / Q_t^*, \end{aligned} \tag{4.9.18}$$

$$\begin{aligned} \boldsymbol{C}_t^* \phi_t^{-1} &= \boldsymbol{R}_t^* \phi_t^{-1} - \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} (\phi_t^{-1} Q_t^*)^{-1} \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} \\ &= \boldsymbol{R}_t^* \phi_t^{-1} - \boldsymbol{R}_t^* \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} / Q_t^*. \end{aligned} \tag{4.9.19}$$

The marginal posterior state distribution, for a p-dimensional state vector unconditional of the scale parameter, is obtained by integrating over the scale parameter,

ϕ_t, in the joint distribution,

$$p(\boldsymbol{\theta}_t \mid D_t) = \int p(\boldsymbol{\theta}_t, \phi_t \mid D_t) d\phi_t. \tag{4.9.20}$$

Using Bayes' theorem this integral can be expressed in terms of the state conditional distribution and the marginal distribution for the scale parameter.

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &= \int p(\boldsymbol{\theta}_t \mid D_t, \phi_t) p(\phi_t \mid D_t) d\phi_t \\ &\propto \int |\boldsymbol{C}_t^* \phi_t^{-1}|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'[\boldsymbol{C}_t^* \phi_t^{-1}]^{-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t) \right\} \\ &\times \phi_t^{n_t/2-1} \exp\{-\phi_t d_t/2\} d\phi_t. \end{aligned} \tag{4.9.21}$$

Since $\boldsymbol{C}_t^* \phi_t^{-1}$ is a covariance matrix the leading diagonal terms must be positive and hence for a p-dimensional state vector, $\boldsymbol{\theta}_t$, the determinant $|\boldsymbol{C}_t^* \phi_t^{-1}| \propto \phi_t^{-p}$ and hence $|\boldsymbol{C}_t^* \phi_t^{-1}|^{-1/2} \propto \phi_t^{p/2}$. Therefore (4.9.21) is proportional to

$$\int \phi_t^{(n_t+p)/2-1} \exp\left\{ -\frac{\phi_t}{2}\left[(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'\boldsymbol{C}_t^{*-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t) + d_t\right] \right\} d\phi_t. \tag{4.9.22}$$

Using Equation (G.0.3) (with $n = (n_t+p)/2$ and $d = [(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'\boldsymbol{C}_t^{*-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t) + d_t]/2$) it follows that (4.9.22) is proportional to

$$\left[(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'\boldsymbol{C}_t^{*-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t) + d_t\right]^{-\left(\frac{n_t+p}{2}\right)}. \tag{4.9.23}$$

Writing $S_t = d_t/n_t$ (the point estimate of the variance scale parameter), (4.9.23) can be expressed, by dividing by d_t, as

$$\left[1 + \frac{(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'(\boldsymbol{C}_t^* S_t)^{-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t)}{n_t}\right]^{-\left(\frac{n_t+p}{2}\right)}. \tag{4.9.24}$$

Using Equation (G.0.5) this can be recognised as a p-dimensional T distribution on n_t degrees of freedom with mean $\boldsymbol{m}_t$ and scale $\boldsymbol{C}_t = \boldsymbol{C}_t^* S_t$. That is,

$$(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{T}_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t). \tag{4.9.25}$$

4.9.4 Variance Discounting

In Section 4.8 we stressed the importance of dynamic change when modelling the stochastic element of the system equation, which describes parametric change. Similar arguments apply to the scale parameter as to the system state. The variance learning model discussed in Section 4.9 does not allow for such a dynamic, and it was noted in Section 4.9.3 that the prior (for the scale parameter) at time t has the same moments as the posterior at time $t-1$. This is unrealistic as forecasting to the future always brings additional uncertainty. The variance learning in Section 4.9 can be extended to allow for stochastic change in the scale using the information discounting strategy. Posterior information on the scale at time $t-1$ is described by a gamma distribution,

$$(\phi_{t-1} \mid D_{t-1}) \sim \text{Ga}(n_{t-1}/2, d_{t-1}/2), \tag{4.9.26}$$

where the mean and variance of the posterior information for the scale parameter are given by

$$\begin{aligned} \text{E}[\phi_{t-1} \mid D_{t-1}] &= \frac{n_{t-1}}{d_{t-1}} \\ \text{Var}(\phi_{t-1} \mid D_{t-1}) &= 2\frac{n_{t-1}}{d_{t-1}^2}. \end{aligned} \tag{4.9.27}$$

The system equation specifies a formal model for stochastic evolution of the state, but that is very difficult to do for the scale parameter. This is due to the gamma distribution not having the convenient mathematical properties that the normal distribution has [15]. However, we can still apply an information discounting strategy to the scale parameter. What matters is to have the time t prior information suitably adjusted from the time $t-1$ posterior information to reflect the additional uncertainty due to temporal advancement. Discounting the precision scale parameter by discounting both parameters of the gamma distribution achieves this aim [15]. For a variance component discount factor, δ_V, define the prior information on the scale at time t as

$$(\phi_t \mid D_{t-1}) \sim \text{Ga}(\delta_V n_{t-1}/2, \delta_V d_{t-1}/2). \tag{4.9.28}$$

The mean of the prior is exactly the same as before when we did not have a discount information approach for the observational variance, as expected. However, the

variance is inflated by a factor of $(\delta_V)^{-1}$ (see Equations (4.9.27)),

$$\begin{aligned} \mathrm{E}[\phi_t \mid D_{t-1}] &= \frac{\delta_V n_{t-1}/2}{\delta_V d_{t-1}/2} = \frac{n_{t-1}}{d_{t-1}}, \\ \mathrm{Var}(\phi_t \mid D_{t-1}) &= \frac{\delta_V n_{t-1}/2}{(\delta_V d_{t-1}/2)^2} = 2(\delta_V)^{-1}\frac{n_{t-1}}{d_{t-1}^2}. \end{aligned} \tag{4.9.29}$$

4.9.5 Observational Variance Practical Discount Strategy

Similarly to Section 4.8.1, discounting the observational variance in this way is an elegant way of coping with the observational variance series, V_t. Once again, when forecasting one-step ahead, at time t, there are no issues and using an observational discount factor, δ_V, the prior variance, at time $t+1$, is inflated by a factor of $(\delta_V)^{-1}$ (compared to the posterior variance at time t). However, inflating the k-step ahead prior variance by a factor of δ^{-k} implies an exponential growth in variance, which is not consistent with the DLM. Hence, this discount approach must be applied with thought when forecasting more than one-step ahead.

We use an approach that uses multiples of the one-step ahead observational variance for forecasting k-steps into the future. The resulting discount procedure is then as follows.

1. Given

$$(\phi_t \mid D_t) \sim \mathrm{Ga}(n_t/2, d_t/2). \tag{4.9.30}$$

   Discount the precision scale parameter by discounting both parameters of the gamma distribution. This results in

$$(\phi_{t+1} \mid D_t) \sim \mathrm{Ga}(\delta_V n_t/2, \delta_V d_t/2), \tag{4.9.31}$$

   and hence

$$\begin{aligned} \mathrm{E}[\phi_{t+1} \mid D_t] &= \mathrm{E}[\phi_t \mid D_t], \\ \mathrm{Var}(\phi_{t+1} \mid D_t) &= (\delta_V)^{-1}\mathrm{Var}(\phi_t \mid D_t). \end{aligned} \tag{4.9.32}$$

2. In forecasting k-steps ahead, adopt the variance

$$\mathrm{Var}(\phi_{t+k} \mid D_t) = k \times \mathrm{Var}(\phi_{t+1} \mid D_t) = k(\delta_V)^{-1}\mathrm{Var}(\phi_t \mid D_t). \tag{4.9.33}$$

3. Forecast in this manner at time $t+1$ and so on.

Prior to Section 4.9 we assumed that V_t was known, and constant, for all times t. This implied that the observational variances, V_t, were independent of the history of the series. Now, with the proposed discount strategy, the observational variances have been modified to depend on the current state of information, D_t, and this allows a learning discount strategy to be applied by inflating the prior variances by a discount factor δ_V^{-1}. This allows DLMs to express more (less) uncertainty at times with more (less) stochastic variation.

4.9.6 Recurrence Relations

Unconditional forecast, state prior and posterior distributions are modified to represent the information discounting strategies discussed in Sections 4.8.1 and 4.9.5. The forecasting algorithms, given an unknown evolution matrix, $\boldsymbol{W}_t$, and an unknown observational variance, V_t, where the evolution matrix and observational variance have discount learning mechanisms with discount factors δ and δ_V respectively, are described in Algorithm 2 (adapted from West and Harrison [1]).

Note that in Section 4.8 it was mentioned that block discounting allows for separate discount factors. This allows for separate components of the DLM to have different variability inflations and to evolve more or less quickly than other components. In this case (see Section 4.7),

$$\boldsymbol{W}_t = \begin{pmatrix} \boldsymbol{W}_{1t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{W}_{ht} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{C}_t = \begin{pmatrix} \boldsymbol{C}_{1t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{C}_{ht} \end{pmatrix}, \tag{4.9.42}$$

and Equation (4.9.35) in Algorithm 2 becomes

$$\boldsymbol{W}_{i,t}(k) \equiv \boldsymbol{W}_{i,t}(1) = (1/\delta_i - 1)\boldsymbol{G}_i\boldsymbol{C}_{i,t}\boldsymbol{G}_i', \tag{4.9.43}$$

for $i = 1, \dots, h$.

Algorithm 2 Recurrence Relations for unknown $\boldsymbol{W}_t$ and unknown V_t

Input: $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, δ, δ_V, n_0, and d_0

1: At time t the posterior for the state and the posterior for the scale parameter are given by

$$\begin{aligned}(\boldsymbol{\theta}_t \mid D_t) &\sim \mathrm{T}_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t),\\ (\phi_t \mid D_t) &\sim \mathrm{Ga}(n_t/2, d_t/2),\end{aligned} \tag{4.9.34}$$

where the point estimate for the variance at time t is given by $S_t = d_t/n_t$ and we set $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$. The evolution matrix for time $t+k$, for $k > 0$, given information up to time t, D_t, is calculated using Equation (4.8.8)

$$\boldsymbol{W}_t(k) \equiv \boldsymbol{W}_t(1) = (1/\delta - 1)\boldsymbol{G}\boldsymbol{C}_t\boldsymbol{G}'. \tag{4.9.35}$$

2: The prior for the state at time $t+k$, for $k > 0$, given information up to time t, D_t, is given by

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathrm{T}_{\delta_V n_t}(\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)), \tag{4.9.36}$$

where

$$\begin{aligned}\boldsymbol{a}_t(k) &= \boldsymbol{G}\boldsymbol{a}_t(k-1),\\ \boldsymbol{R}_t(k) &= \boldsymbol{G}\boldsymbol{R}_t(k-1)\boldsymbol{G}' + \boldsymbol{W}_t(k).\end{aligned} \tag{4.9.37}$$

3: The prior for the forecasts at time $t+k$ given information up to time t, D_t, is given by

$$(Y_{t+k} \mid D_t) \sim \mathrm{T}_{\delta_V n_t}(f_t(k), Q_t(k)), \tag{4.9.38}$$

where

$$\begin{aligned}f_t(k) &= \boldsymbol{F}'_{t+k}\boldsymbol{a}_t(k),\\ Q_t(k) &= \boldsymbol{F}'_{t+k}\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + kS_t.\end{aligned} \tag{4.9.39}$$

4: Then, after observation y_{t+1} is observed we update Equation (4.9.34) to time $t+1$

$$\begin{aligned}(\boldsymbol{\theta}_{t+1} \mid D_{t+1}) &\sim \mathrm{T}_{n_{t+1}}(\boldsymbol{m}_{t+1}, \boldsymbol{C}_{t+1}),\\ (\phi_{t+1} \mid D_{t+1}) &\sim \mathrm{Ga}(n_{t+1}/2, d_{t+1}/2),\end{aligned} \tag{4.9.40}$$

where the point estimate for the variance at time $t+1$ is given by $S_{t+1} = d_{t+1}/n_{t+1}$ and we set $\boldsymbol{a}_{t+1}(0) = \boldsymbol{m}_{t+1}$ and $\boldsymbol{R}_{t+1}(0) = \boldsymbol{C}_{t+1}$, where

$$\begin{aligned}e_t(1) &= y_{t+1} - f_t(1),\\ n_{t+1} &= \delta_V n_t + 1,\\ d_{t+1} &= \delta_V d_t + S_t e_t^2(1)/Q_t(1),\\ S_{t+1} &= d_{t+1}/n_{t+1}\\ \boldsymbol{A}_{t+1} &= \boldsymbol{R}_t(1)\boldsymbol{F}_{t+1}/Q_t(1),\\ \boldsymbol{m}_{t+1} &= \boldsymbol{a}_t(1) + \boldsymbol{A}_{t+1}e_t(1),\\ \boldsymbol{C}_{t+1} &= (S_{t+1}/S_t)[(\boldsymbol{R}_t(1) - \boldsymbol{A}_{t+1}\boldsymbol{A}'_{t+1}Q_t(1)].\end{aligned} \tag{4.9.41}$$

Chapter 5

Monitoring Renal Dysfunction: An Application of Dynamic Linear Models

In this chapter we show how dynamic models can be used to predict severe oliguria and its consequences, comparing our methods of clinical prevention to the KDIGO stage one urine output criterion (see Table 1.1) and Ralib's criterion [2]. The work in this chapter is in joint collaboration with Dr Samuel Howitt, and results produced by Dr Samuel Howitt are clearly referenced.

5.1 Introduction

It was mentioned in Section 1 that acute kidney injury (AKI), defined by the KDIGO guidelines (Table 1.1), occurs in up to 75% of patients on the general intensive care unit (ICU) [3, 4] and in up to 30% of patients following cardiac surgery on the cardiac intensive care unit (CICU) [5]. The KDIGO stage one urine output criterion for AKI ($< 0.5ml/kg/hr$ for 6 consecutive hours) lacks specificity (true positive rate) when identifying patients at risk of adverse renal outcomes [18, 19], meaning that many patients who are classified as suffering kidney injury, according to the stage one

urine output criterion, do not go on to suffer adverse renal outcomes, and hence a better classification method is needed. In this chapter we develop and validate a novel monitoring system to analyse urine output in order to identify those at risk of severe oliguria ($< 0.3ml/kg/hr$ for 6 consecutive hours) and its consequences. We discuss in detail how dynamic models can be used to identify a group of patients at increased risk of adverse outcomes due to renal dysfunctioning and the advantages of using dynamic models over current methods.

The most widely used tool which can identify patients at risk of adverse outcomes related to renal dysfunctioning are the KDIGO AKI guidelines (Table 1.1). These guidelines were designed to identify the severity of AKI according to urine output and serum creatinine concentration. The evidence for risk stratification by serum creatinine is stronger than that for stratification according to urine output. Serum creatinine within the guidelines have been shown to identify cardiac surgery patients with increased risk of a prolonged length of stay (> 10 days) [20], and increased risk of mortality [21, 22]. However, recall from Section 1.1 that serum levels, such as creatinine are only measured once per day for most patients in our dataset, since it requires a blood test which takes time to analyse. Patients can be classified as having ARF within hours after having heart surgery, and identifying ARF in these patients is very difficult using sparsely measured variables such as serum creatinine alone. Moreover, experts made the decision to start renal replacement therapy (RRT) within 66 hours for 90% of patients in our study, giving only two measurements of serum creatinine for most patients. Hence, in this thesis, we model and monitor kidney deterioration by using urine outputs, which are easy to measure on a regular basis.

Studies which have attempted to validate the urine output guidelines as a risk stratification tool for cardiac surgery patients have shown that patients who suffer AKI by urine output alone have outcomes that are only marginally worse than those who do not suffer AKI [18, 19]. This may be because a large proportion of patients who undergo cardiac surgery experience AKI stage one kidney injury (see Table 1.1) as part of the physiological response to surgery. This response is short-term and does not lead to adverse outcomes. The KDIGO guidelines assign higher levels of AKI to those who suffer more prolonged periods of low urine output, however, at least 12 hours of oliguria (low urine output) must pass before a higher classification of AKI can be made. In this setting the delays in identification of higher levels of AKI may

lead to harm to the kidneys which could be minimised if detected earlier.

Ralib *et al* demonstrated that a threshold of $0.3ml/kg/hr$ for six consecutive hours was more closely associated with adverse outcomes in general ICU patients [4, 2]. However, if this threshold is used instead of the existing AKI stage one urine output definition, patients may suffer harm in the six hours of oliguria required for identification before the increased risk is identified (see Figure 1.1 and discussion). Dynamic models have been used in similar studies [23, 24] and underpin the approach that we will use to model the urine output series. Using dynamic models it is possible to use the definition proposed by Ralib (which is able to identify a group of patients more likely to suffer adverse outcomes), but instead of waiting for six consecutive urine outputs below $0.3ml/kg/hr$ to be observed, we forecast the next six urine outputs and at each hour calculate the joint probability (see Section 4.3) that these six forecasted urine outputs are all below $0.3ml/kg/hr$. Using dynamic models we can forecast severe oliguria, aiming to intervene before damage to the kidneys occurs.

The clinician wishes to have a decision rule to intervene if a patient is at high risk of suffering severe oliguria, i.e.

$$\Pr(Y_{t+1} < T, \ldots, Y_{t+6} < T \mid D_t) \geq p, \tag{5.1.1}$$

for some threshold T and probability p, and where $Y_{t+1}, \ldots, Y_{t+6}$ are the urine outputs at times $t+1, \ldots, t+6$, given information up to time t, D_t. In this study we take $T = 0.3ml/kg$ (Ralib's threshold) and use $p = 0.8$. This value of p was chosen by the clinical team from the University Hospital of South Manchester, with the aim to produce a model with high specificity. If the probability of suffering severe oliguria is higher than 0.8, at time t, then this will be a high risk warning and the patient will be considered high risk by our model. This approach will be used to identify patients at increased risk of severe oliguria. The above decision rule is specified so that we can compare the first high risk warning raised by our model to the AKI guidelines and to Ralib's criterion. However, our model (with model monitoring, see Section 6) is developed into an R shiny application (see Section 8) and is more than a single decision rule. Clinicians can easily use the shiny application to monitor the progress of patients over time. The model automatically runs and requires no coding experience nor experience in statistics. Using the application clinicians can

potentially intervene before a patient is considered "high risk", if say, the probability of severe oliguria is rapidly increasing over time, for example. Clinicians can observe and monitor how a patient reacts to a drug (aimed at increasing urine output) or another intervention, and intervene accordingly using the application and expert experience. The R shiny application is discussed in Section 8.

Using dynamic models, which update in real time, we identify patients at risk of suffering severe oliguria on the cardiac intensive care unit (CICU). The model's predictions are updated on an hourly basis according to the measured hourly urine output. Using these methods a patient could be identified as being likely to suffer severe oliguria before it happens. The first aim of this study is to test the feasibility of realtime screening of patient data to identify those at risk early enough to allow intervention to prevent or lessen harm caused by prolonged oliguria. The secondary aim of the study is to compare the model's ability to identify patients at risk of adverse outcomes with the stage one urine output criterion within the KDIGO guidelines (see Table 1.1). The third aim of the study is to compare the model's ability to identify patients at risk of adverse outcomes with the criterion proposed by Ralib *et al* (observing six consecutive urine outputs $< 0.3ml/kg$).

We will test the model's (defined in Section 5.2) performance throughout a patient's CICU admissions up to 72 hours. This time frame was chosen by the clinical team from the University Hospital of South Manchester. Our analyses will classify a success if the model identifies a patient who is going to suffer severe oliguria within 12 hours of prediction and also if patients classified as low risk by our model do not go on to suffer severe oliguria. The threshold of 12 hours was chosen by the clinical team from the University Hospital of South Manchester as an acceptable monitoring period following a high risk classification.

Prospectively collected data, from the University Hospital of South Manchester, of 3602 adult patients admitted to the cardiac intensive care unit (CICU) following cardiac surgery between January 2013 and November 2017 were analysed. Patients receiving mechanical circulatory support (MCS) or cardiac transplantation were excluded (228). Patients who received RRT (renal replacement therapy) preoperatively were also excluded (4). This left 3370 eligible patients. Of the eligible patients, 981 were chosen at random to be assigned to the development group and the remaining 2389 were assigned to the validation group ($\approx 30\% : 70\%$). Stratified sampling was

used where the sets were chosen randomly and the sets have a 3 : 7 ratio of RRT (patients that had renal replacement therapy) and non-RRT patients (patients that did not have renal replacement therapy). Data from the development set were used to estimate model parameters (see Appendix D). Data from the validation set were used to test the three aims of our study described above. Model discrimination and calibration when identifying those likely to suffer severe oliguria within 12 hours were tested at 12, 24, 36, 48 and 72 hours using data from the validation dataset. Model calibration is used to test the models performance at different time stamps to see if the model performs adequately at all times.

5.1.1 Missing Data

In this analysis, at the request of doctors, where hourly urine output recordings were missing, the next recorded urine output was divided by the number of hours that had elapsed since the previous reading and this value was substituted for the hourly values for these hours. When patient weight was missing these values were imputed using the median weight for a patient of that gender. Patient weight was missing for 13 (1.3%) and 23 (1.0%) of the patients in the development and validation sets respectively.

5.2 Model Development

We will now apply the models and methods discussed in Section 4 to identify patients likely to suffer severe oliguria. A patient's urine level stochastically varies over time, and the functioning of a patient's kidneys can cause the urine output level to grow and decline. For these reasons we decided to start with a second-order polynomial dynamic model (see Section 4.6). The first component will capture the expected level for the urine output and the second component allows for a systematic change in level and captures the changing decline and growth dynamics of the urine output recordings.

In Section 2.2 it was discussed that the prediction intervals for the static model can include negative values and that many patients exhibit an exponential decay in urine output to begin with. Hence we, once again, transform the response to

$Y \longmapsto \log(Y + \varepsilon)$, where ε is constant (see Section 2.3 for discussion). A sensitivity analysis was performed for different values of epsilon, see Appendix D, and this analysis led us to choose $\varepsilon = 0.1$.

5.2.1 Second Order Polynomial Model

Using theory described in Section 4.6 we begin modelling the transformed urine output data using a second-order polynomial model with an unknown evolution variance series, $\boldsymbol{W}_t$, and an unknown observational variance series, V_t. Both of these unknown variances are modelled using information discounting (see Sections 4.8 and 4.9.4 respectively). The model is given below

$$\begin{aligned} \log(Y_t + 0.1) &= \mu_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}. \end{aligned} \tag{5.2.1}$$

The state vector $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$, where μ_t allows for systematic variation about a time varying level and β_t allows for systematic growth and decline of the level, where $\nu_t \sim \mathrm{N}(0, V_t)$ and $\boldsymbol{\omega}_t \sim \mathrm{T}_{n_{t-1}}(0, \boldsymbol{W}_t)$ where

$$V_t^{-1} \mid D_t \sim \mathrm{Ga}(\delta_V n_t/2, \delta_V d_t/2), \quad \boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} + W_{\beta t} & W_{\beta t} \\ W_{\beta t} & W_{\beta t} \end{pmatrix}, \tag{5.2.2}$$

where $W_{\mu t} = C_{\mu,t-1}(\delta_\mu^{-1} - 1)$ and $W_{\beta t} = C_{\beta,t-1}(\delta_\beta^{-1} - 1)$, where $\mathrm{Var}(\mu_{t-1}|D_{t-1}) = C_{\mu,t-1}$ and $\mathrm{Var}(\beta_{t-1}|D_{t-1}) = C_{\beta,t-1}$. Having separate discount factors, δ_μ and δ_β, for the level and growth parameters respectively, allows the parameters to evolve at different rates. At certain times the level could evolve more quickly than the slope and vice versa. The observation and evolution discounts chosen are $\delta_V = 0.95$, $\delta_\mu = 0.80$, and $\delta_\beta = 0.90$ respectively (see Appendix D for a sensitivity analysis performed on δ_V, δ_μ, and δ_β). Note that higher discount factors usually lead to better forecasts, but noisy data often pulls the discount factors downwards, and this is unwanted. Techniques to prevent this are discussed in Chapters 6 and 7.

5.2.2 Parameter Estimation and Diagnostics

Several measures of forecasting accuracy can be used to compare models formally. Commonly used criteria are the mean absolute deviation (MAD), and the mean square error (MSE), defined respectively by the following formula [15]

$$\begin{aligned} \text{MAD} &= \frac{1}{n}\sum_{t=1}^{n}|e_t|, \\ \text{MSE} &= \frac{1}{n}\sum_{t=1}^{n}e_t^2. \end{aligned} \tag{5.2.3}$$

Noting that these quantities should always be calculated on the untransformed scale, with the same units used for communication. This is required to understand the measures on the scale most useful to us, but also to prevent a model being under or over credited by using smaller or larger units.

The criterion for estimating the initial parameters in our model is to minimise the number of type two errors (see below). Diagnostics are shown in Appendix D (for the discount factors) alongside: the type one errors, the MAD, and the MSE. We now define type one and type two errors in this context.

Definition 5.2.2.1. A type one error is recorded when the model predicts that a patient will suffer severe oliguria in six hours when, in fact, they do not suffer severe oliguria in six hours. That is, when the next six urine outputs are not all below $0.3ml/kg/hr$ but we predict that

$$\Pr(Y_{t+1} < 0.3, \ldots, Y_{t+6} < 0.3 \mid D_t) \geq 0.8. \tag{5.2.4}$$

where $D_t = \{D_{t-1}, y_t\}$.

Definition 5.2.2.2. A type two error is recorded when the model predicts that a patient will not experience severe oliguria in six hours when the patient is observed suffering severe oliguria six hours later. In other words, when the next six observed urine outputs are all below $0.3ml/kg/hr$ and

$$\Pr(Y_{t+1} < 0.3, \ldots, Y_{t+6} < 0.3 \mid D_t) < 0.8, \tag{5.2.5}$$

where $D_t = \{D_{t-1}, y_t\}$.

The proportion of type one and type two errors will both ideally be as small as possible. However, it will be the case that some patients experience oliguria for five hours or the patient's kidneys are in a deteriorating state and the model classifies this patient at high risk of severe oliguria at said times (see Figure 5.1 (left)). If the patient (even slightly) improves for just one hour because of an intervention or for other reasons, even if they stay at an unhealthy level or go back to deteriorating afterwards, this is classified as a type one error (see Figure 5.1 (left)). Alerting these patients as at high risk of suffering severe oliguria is classified as a type one error but as we can see from the example described above, some type one errors will be unavoidable and are not a cause for concern. In fact, some of these type one errors are the reason why we are using dynamic models. For example, suppose that a patient has five consecutive urine outputs below $0.3ml/kg$, an outlying observation around $0.56ml/kg$, and then returns back to a deteriorating level and finally has six urine outputs below $0.3ml/kg$ (see Figure 5.1 (left)), and assume that our DLM alerts that this patient is at "high risk" at hours 6, 7, 8 and 9. This patient will not be considered as suffering severe oliguria until hour 16, however, it is clear that the one-off outlying observation should not remove the past evidence of this patient's deteriorating kidneys. Although this will cause four type one errors, the model has successfully identified a patient with deteriorating kidney function at hours 6, 7, 8 and 9. Consider now Figure 5.1 (right), if the model identifies this patient as high risk at hours 6, 7, 8 and 9, this would cause four type one errors (again). However, in real time, this urine output series would be a cause for concern to clinicians, at hours 6, 7, 8 and 9. Hence we use type one errors with caution, knowing that some are not a cause for concern.

On the other hand, a problem arises if, say, (see Figure 5.1 (right)), the model still signals that the patient is at high risk at, say hour 14. It may be the case that this patient is identified as likely to suffer severe oliguria in the deterioration stage, however, if the model fails to adapt to changing parameter values and continues to forecast that this patient is likely to suffer severe oliguria (from past evidence of deteriorating kidneys), then this is a more drastic type one error reflecting poor model forecasting performance, see Figure 3.2 (left). Poor model performance is discussed in Chapter 6.

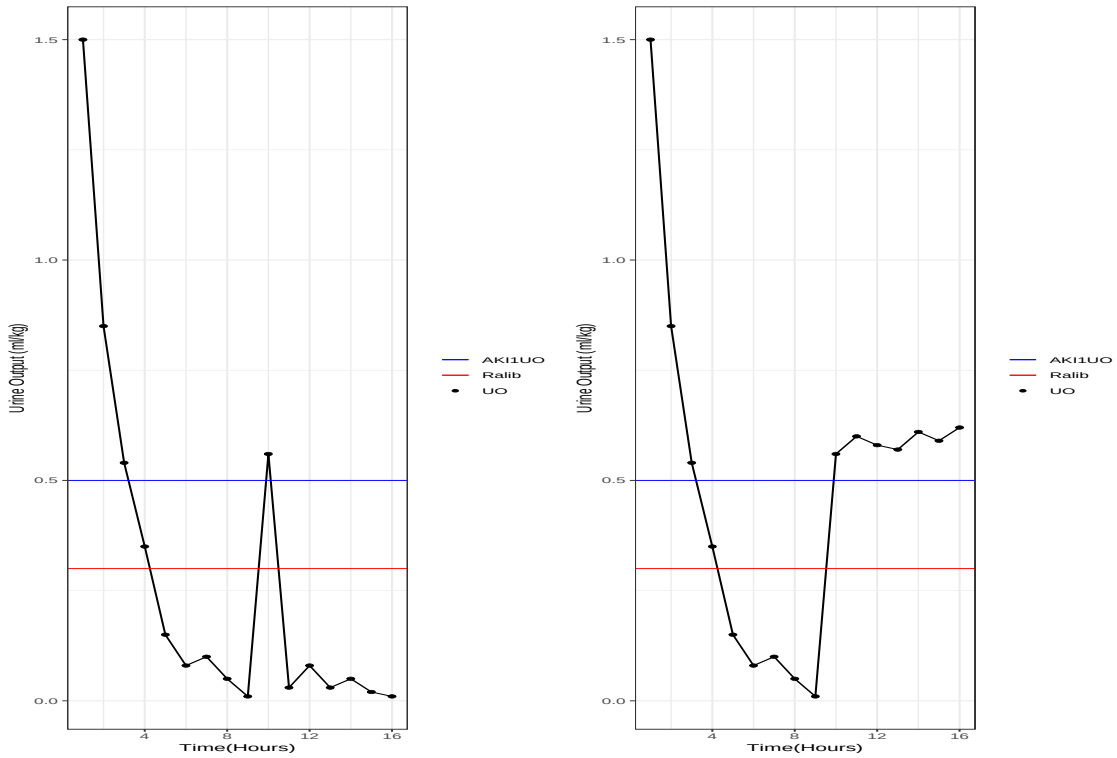

Figure 5.1: Type Two Errors. Figures illustrate times at which type two errors can occur. Figure (left) shows an example of when a type two error can occur due to an outlier. Figure (right) shows an example of when a type two error can occur due to poor model forecasting

In contrast, type two errors are classified as more severe, by the clinical team from the University Hospital of South Manchester, than type one errors, since these patients may be overlooked and this is a drastic model failure. This could be due to the model slowly adapting from a healthy trend to an unhealthy trend, reflecting poor model forecasting performance. In order to be clinically useful the model will be designed to minimise the number of type two errors, so that a "low risk" classification can be trusted. Type one errors, MAD, and MSE will also be considered in parameter estimation.

5.2.3 Prior Elicitation

Data processing through Bayes' Rule aids in updating beliefs. By using an initial prior we are able to update our beliefs, as new data becomes available, to obtain a posterior distribution. Then as more data becomes available this posterior distribution becomes our prior distribution and so on. With dynamic models as data

collection continues the initial prior information gets overwhelmed by the data. In other words, when enough relevant information is accumulated, the initial prior influence on the current posterior distribution is swamped by that contained in the sample (see Section 4.5). The prior distribution required to start the recurrence relations will now be constructed from data analysis, and expert opinions.

After a patient has been in an operating theatre, it is expected that their urine output will be higher than usual due to a long surgery. Analysing the development set shows that the first urine output for a patient is, on average, around $1.5ml/kg$. Also, the urine output for the following hours should decrease since the patient's urine outputs are measured regularly again. Data analysis leads us to choose (on the log-scale) $\boldsymbol{m}_0 = (m_{\mu,0}, m_{\beta,0}) = (0.55, -0.2)$ and

$$\boldsymbol{C}_0 = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.001 \end{pmatrix}. \tag{5.2.6}$$

This reflects that we expect the first urine output after surgery to be 1.42 and 95% of the time we expect the first urine output after surgery to lie in the interval $[0.71, 2.27]$ (using algorithms summarised in Section 4.9.6), on the raw scale. This agrees with the data and also with what experts expect after a patient has just had heart surgery. This reflects a large amount of uncertainty and this is due to not having any preoperative data to understand the functioning of patient's kidneys before operation; and also due to not knowing how patient's kidneys will react to having cardiac surgery and medication. The functioning of the patient's kidneys post surgery is very much uncertain and hence we reflect this in our prior beliefs. However, having a large variance to reflect great uncertainty about the future allows dynamic models to rapidly adapt to future observations (see Chapter 6). It is best to be cautious, and have large variances, so that the model can adapt quickly and perform well for future forecasts.

The same diagnostics shown in Appendix D were performed with values of expected initial level varying from $0 \leq m_{\mu,0} \leq 1$, expected initial slope varying from $-0.5 \leq m_{\beta,0} \leq 0$, and initial variances for the level and slope ranging from $0.001 < C_{\mu,0} < 0.1$, $0.0001 < C_{\beta,0} < 0.1$, respectively (recall that these values are on the log-scale). These tables are omitted to avoid showing repetitive analysis, however following the

criteria, mentioned in Section 5.2.2, leads to choosing the initial prior below

$$(\boldsymbol{\theta}_0 \mid D_0) \sim \mathrm{T}_{n_0}\left[\begin{pmatrix}0.55\\-0.2\end{pmatrix}, \begin{pmatrix}0.01 & 0\\0 & 0.001\end{pmatrix}\right]. \tag{5.2.7}$$

We take the prior value for the observational variance to be $S_0 = 0.1$, with $n_0 = 20$ and $d_0 = 2$ (see Equation (4.9.10)). Once again, this is chosen by following the criteria defined in Section 5.2.2.

5.3 Results

The results of the model's performance in the validation group (2389 patients) are now presented. To allow comparison of the model's output with the existing categorical KDIGO classification, patients were assigned to either a high risk or a low risk group based on their first high risk warning. The analyses tested the model's ability to identify which patients would suffer severe oliguria (UO $< 0.3ml/kg$ for 6 hours) within 12 hours of the high risk prediction. Risk classifications made during the last 12 hours of a patient's admission were disregarded since it was not possible to ascertain whether severe oliguria had occurred following discharge from CICU.

The model discriminated very well between those who went on to develop severe oliguria and those who did not at multiple time points during the first 72 hours in CICU. In the early stages of CICU admissions the model overestimated risk but calibration was very good following the first 24 hours. In the first 24 hours the risk predicted was higher than risk observed. The poor calibration of the model in the first 24 hours may be due to the recovery from the initial physiological response to surgery (it is problematic that we do not know when a patient will recover from an initial physiological response, and this can cause models to breakdown (see Chapters 6 and 7)). Interventions made in response to falling urine outputs will also have played a role in the model overestimating risk. However, if a clinician intervenes when our model identifies a patient as highly likely to suffer severe oliguria, and this patient recovers, then this is a success. Intervention effects, and physiological responses, will be less evident as time from surgery increases, since it is unlikely that a patient will have a physiological response lasting longer than 24 hours, also clinicians are likely to have exhausted minor interventions in the first 24 hours and hence after that time point interventions are unlikely to affect the model's

performance. Interventions and physiological responses can drastically affect the level and slope of a patients urine output and these sudden changes in the series can be difficult to model. This and other sudden changes are discussed in Chapters 6 and 7. Table 5.1 below shows model performance over time, by considering observed severe oliguria (SO) compared to predicted severe oliguria (within 12 hours).

Time Point (number of patients still on CICU)	Observed SO (% of patients)	Predicted SO (% of patients)	O:E ratio
12 hours (1947)	61 (3.1)	153 (7.9)	0.40
24 hours (1694)	57 (3.4)	91 (5.4)	0.63
36 hours (1137)	51 (4.5)	40 (3.5)	1.28
48 hours (909)	54 (5.9)	46 (5.1)	1.17
72 hours (545)	35 (6.4)	30 (5.5)	1.17

Table 5.1: Calibration of model's predictive performance, produced by Dr Samuel Howitt. Table shows DLM performance over time by comparing observed severe oliguria to predicted severe oliguria

5.4 Discussion

In the validation group, 2088 (87.4%) patients suffered at least one hour of urine output below $0.3ml/kg$. Severe oliguria was experienced in 197 (8.2%) patients. Renal replacement therapy was required in 89 (3.7%) patients, for 19 (0.8%) patients RRT was initiated within three hours of arrival on CICU and these patients were excluded from the RRT analyses. In all 19 excluded cases the decision to start RRT had been made prior to arrival on CICU by the anaesthetist responsible for intraoperative care; hence there is no clinical need for the model to predict RRT in these cases. Prolonged length of stay (PLOS) was observed in 589 (24.7%) patients and 36 (1.5%) patients died prior to hospital discharge. The results of the model are summarised in Table 5.2.

Group (2389)	RRT (70)	PLOS (589)	Hospital Mortality (36)
High Risk (202)	37 (18.3%)	113 (55.9%)	13 (6.4%)
Low Risk (2187)	33 (1.5%)	476 (21.8%)	23 (1.1%)

Table 5.2: Outcome of patients according to dynamic model, produced by Dr Samuel Howitt. Table shows proportion of patients who required RRT, had a PLOS, and died in hospital for patients classified as high risk by our model and for patients classified as low risk by our model

The model defined in Section 5.2 classified 202 (8.46%) patients in the validation group as being at high risk and assigned 2187 (91.54%) patients to the low risk group. Patients in the high risk group were at increased risk of RRT (OR 17.8, 95% CI 9.7-33), PLOS (OR 4.0, 95% CI 2.9-5.7) and hospital mortality (OR 5.3, 95% CI 3.0-13.0) compared with those in the low risk group (see Table 5.2). The odds ratio (OR) is the ratio of the odds of an event (here RRT, PLOS, or hospital mortality) for patients considered high risk compared to patients considered low risk. We see that, 95% of time, patients considered high risk are 9.7-33 times more likely to need RRT, 2.9-5.7 times more likely to have a prolonged length of stay (PLOS $>$ 10 days) in hospital, and 3-13 times more likely to die in hospital.

It is clear from Table 5.2 that the dynamic model is able to distinguish between patients likely to suffer adverse outcomes compared to those that are unlikely to suffer adverse outcomes. Hence, completing our first aim of testing the feasibility of realtime screening of patient data to identify those at risk early enough to allow intervention to prevent or lessen harm caused by prolonged oliguria.

The KDIGO guidelines classified 628 (26.3%) patients as suffering AKI stage one by urine output, that is, the KDIGO guidelines identified 628 (26.3%) patients as "high risk" and 1761 (73.7%) patients as "low risk". The results of classification by the model and the KDIGO criterion are summarised in Table 5.3. The subgroup of 451 (71.8% of the AKI stage one urine output high risk group) patients classified as "high risk" by the KDIGO guidelines, who met the AKI urine output stage one criterion but were classified as low risk by our model, experienced rates of RRT (4.0%), PLOS (33.3%) and mortality (2.7%) which were significantly lower than those classified as high risk by our dynamic model (see Table 5.2). Our second aim was to compare the model's ability to identify patients at risk of adverse outcomes with the stage one urine output criterion within the KDIGO guidelines. Table 5.3 confirms that our model outperforms the current AKI stage one urine output criterion.

In addition, when used to predict future RRT requirement, the dynamic model has a higher positive predictive value (PPV - the proportion of true positives) and model specificity than the KDIGO AKI stage one urine output criterion. The negative predictive value (NPV - the proportion of true negatives) for the DLM and the KDIGO AKI stage one urine output criterion was found to be the same. Furthermore, the model's classification was almost identical to that achieved by classification accord-

Group	N (%)	RRT	PLOS	Hospital Mortality
Low risk and no AKI by UO	1736 (72.7%)	15 (0.9%)	326 (18.8%)	11 (0.6%)
Low risk and AKI by UO	451 (18.9%)	18 (4.0%)	150 (33.3%)	12 (2.7%)
High risk and no AKI by UO	25 (1.0%)	3 (12.0%)	14 (56.0%)	2 (8.0%)
High risk and AKI by UO	177 (7.4%)	34 (19.2%)	97 (54.8%)	11 (6.2%)

Table 5.3: Comparison of model against KDIGO AKI urine output stage one guideline, produced by Dr Samuel Howitt. Table compares rates of adverse outcomes for patients classified as high or low risk by our model to patients classified as high or low risk by the KDIGO AKI stage one urine output guideline

ing to observed rather than predicted severe oliguria (see Table 5.4).

Outcome	Criterion	Sensitivity	Specificity	PPV	NPV
RRT	AKI-UO	0.74	0.75	0.08	0.99
	Dynamic Model	0.53	0.93	0.18	0.99
	Severe Oliguria	0.41	0.94	0.18	0.98

Table 5.4: Comparison of model, KDIGO AKI urine output stage one guideline and severe oliguria, produced by Dr Samuel Howitt. Table compares our DLM, the AKI stage one urine output criterion, and Ralib's criterion when identifying patients who went on to need RRT

From the tables above we can see that the model defined in Section 5.2 is able to identify a group of patients that are more likely to suffer adverse outcomes. Patients assigned to the high risk group were significantly more likely to suffer adverse outcomes than those assigned to the low risk group. The model outperformed the existing KDIGO AKI stage one urine output criterion when identifying those at risk of poor outcomes. As in previous studies we have found that the KDIGO AKI stage one urine output criterion is too sensitive, since many of those identified as high risk by the criterion went on to have good outcomes [19, 25].

The dynamic model's high risk group contained under a third as many patients as the "high risk" group identified by the KDIGO criterion (202 vs 628). The rate of adverse outcomes was higher in the dynamic model's high risk group than in the AKI group. A large subgroup ($n = 451$) met the AKI criterion but were classified as low risk by the model (see Table 5.3). Outcomes for these patients were significantly

better than for the group classified as high risk by our model (see Table 5.2). This suggests that most of the patients with the highest risk within the AKI group were also identified as high risk by our model, whereas lower risk patients in the AKI group were not identified by our model. This illustrates the models ability to outperform the AKI stage one urine output criterion. As expected, as the AKI group contained nearly three times as many patients as the dynamic model's high risk group, the AKI classification was more sensitive when identifying those who would go on to require RRT; however, specificity and positive predictive values were greater for the dynamic model (see Table 5.4).

The increased risk associated with predicted or observed oliguria confirms similar findings from Ralib *et al's* study in general ICU patients [2]. Our third aim was to compare the model's ability to identify patients at risk of adverse outcomes with the criterion proposed by Ralib *et al.* In our study, risk stratification was not significantly improved when classifications were made according to observed rather than predicted severe oliguria. This suggests that there is no advantage to using observed rather than predicted oliguria as the classifier. Conversely, there is a clear advantage to using our dynamic model's predictions; the model provides a warning of severe oliguria before it occurs, allowing time to deliver treatments to prevent severe oliguria and its consequences. The median (IQR) time from high risk classification to severe oliguria of 3.0 $(2.0 - 4.3)$ hours would be enough to allow interventions aimed at preserving a healthy, constant renal output. In reality patients for whom risk of severe oliguria is increasing are likely to be reviewed before a probability of 0.8 is reached, affording even more time for intervention.

While these results are a great start, analyses of urine output alone cannot identify all patients at risk of adverse outcomes related to renal dysfunctioning. There were 33 patients (see Table 5.2) who received RRT while they were classified as being at low risk of suffering severe oliguria by our model because their urine output was maintained around or above $0.3ml/kg/hr$. Indicating that the model is forecasting the urine output series accurately, however, it is also missing other variables that could be contributing to adverse outcomes related to renal dysfunctioning. Analysing these patients identified deranged biochemistry (elevated urea and/or creatinine concentrations) $(n = 27)$, fluid overload $(n = 11)$, elevated lactate $(n = 4)$ and sepsis $(n = 1)$ as the reasons for RRT initiation. Further work should focus on integrating the novel analysis of urine output described in this study with bio-

chemistries to identify patients at risk of adverse outcomes who have a healthy urine output level.

In Section 5.1 we mentioned that experts made the decision to start RRT within 66 hours for 90% of patients in our study. This meant that, for most patients, we did not have enough creatine measurements to make inferences about the performance of a patient's kidneys. However, including creatine measurements, alongside urine output for all patients, in particular for those who spend a long length of time in hospital (> 5 days), could help identify a large percentage of the 33 patients that were missed while maintaining a healthy urine output series.

5.5 Limitations

As this is a retrospective study we observed interventions that were made with the intention of normalising urine output (and not just the adverse outcomes that we were aiming to detect). However, if our model signals a high risk warning and this patient is given an intervention (such as a drug to normalise urine output) then these high risk warnings are also successes, since they were given because the patient's urine output was deteriorating. A total of 30 (14.9%) patients who were classified as high risk by our model received diuretics during the 6 hours before or the 6 hours post classification. Many patients also received a fluid challenge; although the retrospective nature of this study made it impossible to classify these interventions as it occurred in various forms including oral intake and intravenous fluids with differing strengths. Data on these interventions could not be reliably incorporated into the analysis since the data did not include the strengths nor the form of intake. These two crucial pieces of information would be needed to include feed-forward interventions (see Section 6.2).

5.6 Conclusions

In this chapter we have shown that using a dynamic linear model to identify those at risk of severe oliguria following cardiac surgery can identify a group of patients who are significantly more likely to suffer adverse outcomes. Similarly to previous studies [19], we found that the current AKI stage one urine output criterion is

not well suited to use in cardiac surgery patients during the postoperative period since it appears to be too sensitive. Therefore there is a clinical need for a better means of screening patients on the CICU in order to identify those at risk of adverse outcomes related to postoperative renal dysfunctioning. Classifying patients as high risk once they had suffered more severe oliguria, (Ralib *et al* [2]), would likely create a high risk group that would suffer higher rates of adverse outcomes. However, this classification method would be less clinically useful because by the time six consecutive urine outputs below $0.3ml/kg$ are observed and the patient is classified as high risk, harm may already have occurred (see Figure 1.1 and discussion). For this reason we used a dynamic model to predict severe oliguria rather than waiting for severe oliguria to occur. The disadvantage of predicting severe oliguria is that the predictions are not accurate for all patients (although in Section 5.2.2 we discussed that this is not always a cause for concern). In order to ensure that the model was clinically useful, it was designed to provide a low false negative rate at the expense of a higher false positive rate, so that a low risk classification could be trusted. Despite this strategy the model's high risk group is still much smaller than the AKI by urine output stage one group (see Table 5.3). In addition, the higher risk group identified by our model has significantly higher rates of all adverse outcomes studied. Furthermore, we have shown that there is no advantage to using observed severe oliguria as opposed to forecasted severe oliguria, however, there is a huge advantage in using dynamic models since the model provides a high risk warning of severe oliguria before it occurs (median (IQR) time 3.0 (2.0-4.3) hours before). This allows experts enough time to intervene. In addition, the model is currently built into a shiny application (with model monitoring, see Chapter 6) which allows clinicians to monitor kidney deterioration with ease and to monitor how quickly a patient's kidneys are deteriorating. This way clinicians can use their own subjectivity to intervene and also monitor the progress and success of interventions (see Chapter 8).

5.6.1 Forecast Analysis

In Section 5.2.1 we noted that higher discount factors usually produce better forecast performance than lower discount factors. However, the time series that we are considering is extremely noisy. The analysis shown in Appendix D suggests the discount factors $\delta_V = 0.95$, $\delta_\mu = 0.8$, and $\delta_\beta = 0.9$. This kind of analysis is a com-

promise because any structural changes other than smooth dynamic variation pull component discount factors in a downward direction. Although using the suggested discount factors allow for the structural changes to be adapted to, it also means that during routine intervals there is excessive uncertainty during evolution, resulting in imprecise forecasts. When the forecast system is explicitly on the lookout for structural changes, such a compromise over discounting is not necessary. In Chapter 6 we discuss an explicit monitoring system to monitor model performance, which can result in not having to compromise by using lower discount factors. Monitoring forecast performance provides invaluable information on model adequacy and indicates possible changes in the time series structure.

5.7 Model Diagnostics

One of the most important aspects of statistical modelling is the analysis of model residuals. When examining forecast residuals one is looking for evidence that the model is lacking in some way: outlying points, clusters of similarly signed residuals (indicating that your model is either underpredicting or overpredicting), clusters of greater or less variability (i.e. periods of larger than "normal" or smaller than "normal" residuals), and any other kind of structure.

Since we have thousands of patients we are unable to show diagnostics for each individual patient. However, below we show the forecast residuals for each of the six forecasts series, $(e_{t+1} \mid D_t), \ldots, (e_{t+6} \mid D_t)$, over time, for a patient.

We see that the scatterplots for the forecast errors for each forecast are centred around zero. This indicates that the expected value, for each forecast, is zero (as required). The scatterplots also have no signs of clusters. This indicates that the model residuals are not correlated, and also that the model is not underpredicting nor overpredicting.

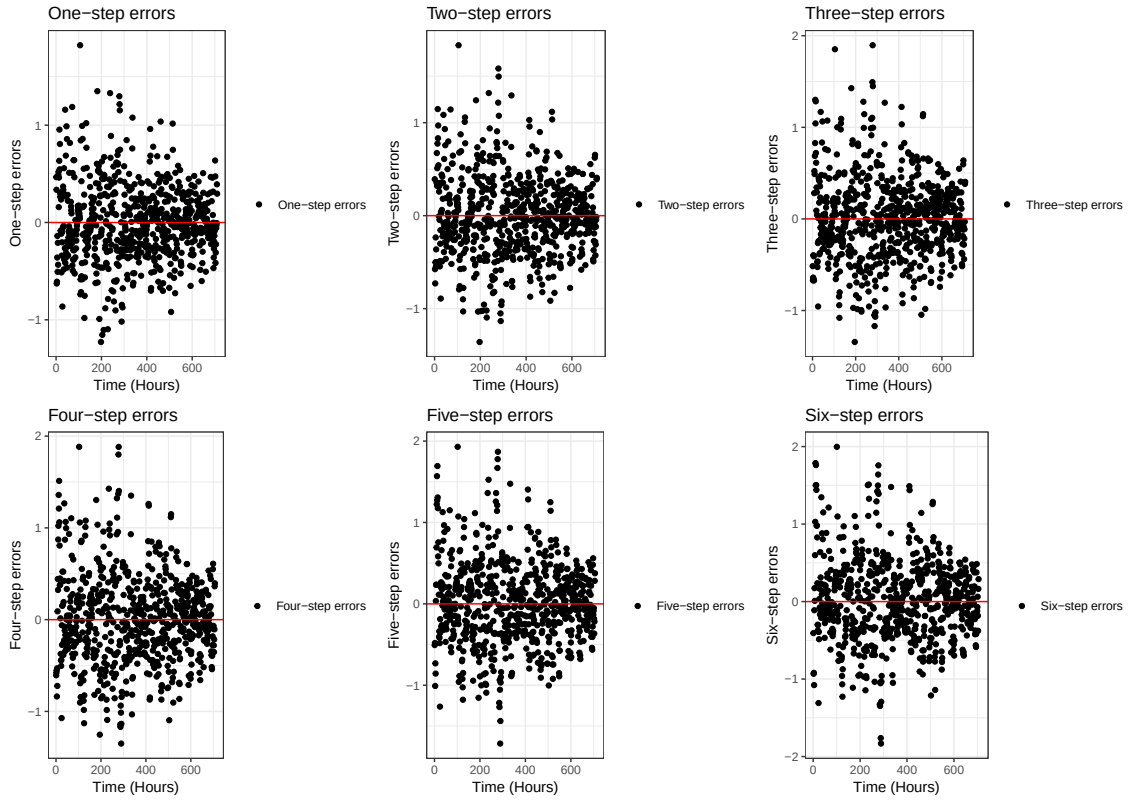

Figure 5.2: Forecast errors for the six future forecasts for a patient. Plot indicates that each forecast is centred around zero. There are also no signs of correlation, and no signs of underpredicting nor overpredicting

Chapter 6

Intervention and Monitoring

In this chapter we introduce the idea of interventions, and consider how routine interventions can be incorporated into existing DLMs and why such interventions may be necessary to sustain predictive performance. Statistical interventions, such as those described by Brown [26], Box et al [27], and Harrison and Veerapen [28] have been used to adapt forecasting systems to maintain efficient forecasting performance.

Interventions can be roughly classified as feedforward or feedback [1]. Feedforward interventions are anticipatory. Feedback interventions are corrective, responding to events that had not been foreseen or adequately allowed for (see Figure 6.1). Corrective actions often arise when it is seen that forecasting performance has deteriorated, warning the forecaster to diagnose any problems. In these situations, any information which was not available beforehand must be used retrospectively to attempt to adjust the model appropriately to the current, local conditions. Forecasting systems operate according to the principal of management by exception [1]. That is, a statistical model is routinely used to process data and information, providing forecasts that are used unless exceptional circumstances arise. Exceptional circumstances occur in two forms. The first relating to known external information, corresponding to feedforward interventions (such as fluid interventions when the clinician knows the amount and strength of the fluid). The second relating to model monitoring which is used to detect deterioration in forecasting performance. We focus on the latter when encountering unknown abrupt changes in the urine output series and take

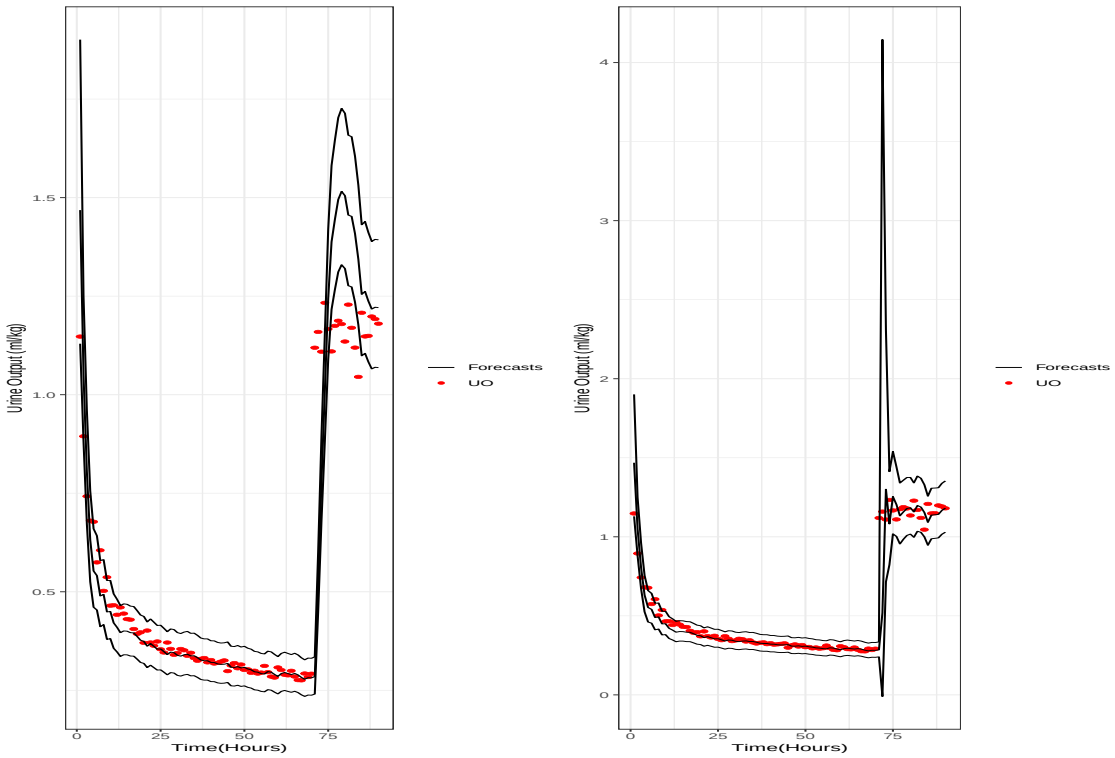

Figure 6.1: Corrective Feedback Intervention. Model (left) closed to intervention vs model (right) with corrective intervention. Model (right) increases prior variances at hour 71 after noticing model deterioration

the form of an automatic, statistical error analyses that continually monitors model performance and issues signals of breakdown when necessary. When forecast performance deteriorates the model forecasts can completely miss the dynamics of the system and struggle to ever recover (see Figure 6.1 (left)). In some circumstances the reason for model deterioration may be unknown. In this situation increasing the variance in the system allows the model to be more adaptive to new data, so that changes that may have taken place are rapidly identified and estimated. Thus, models can be self-correcting if uncertainties about parameters can be significantly increased at points of suspected change [1].

For reference we remind the reader of the general univariate dynamic linear model (with a constant evolution matrix),

$$\begin{aligned} Y_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, & \nu_t &\sim \mathrm{N}(0, V_t), \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{W}_t). \end{aligned} \tag{6.0.1}$$

The historical data, D_{t-1}, (including past observations and any previous interven-

tions) is summarised in terms of the posterior distribution for $\boldsymbol{\theta}_{t-1}$

$$(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}). \tag{6.0.2}$$

The (one-step ahead) prior for the state vector at time $t-1$ is,

$$(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{a}_t, \boldsymbol{R}_t), \tag{6.0.3}$$

with $\boldsymbol{a}_t = \boldsymbol{G}\boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}' + \boldsymbol{W}_t$.

6.1 Types of Intervention

6.1.1 Ignoring Observation y_t

The first type of intervention that we will consider involves treating y_t as an outlier [1]. This could be due to a blocked catheter creating a huge spike in urine output, or it could be a spike due to diuretics or a fluid challenge which are one-off high values, and any other interventions in the environment of the time series that may lead to a single observation being discrepant and unrelated to the rest of the series. In these cases it is suggested that the observation should not be used in updating the model for forecasting the future, since it provides no relevant information. If y_t is classified as an outlier, then y_t is uninformative about the future and so should be given no weight in updating the model distributions. Thus, $D_t = \{I_t, D_{t-1}\}$ is equivalent to D_{t-1}. The posterior for the state vector at time t is just the prior, with $(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t)$, where $\boldsymbol{m}_t = \boldsymbol{a}_t$ and $\boldsymbol{C}_t = \boldsymbol{R}_t$.

Formally, this can be modelled in the DLM format by viewing the observation as having a very large variance V_t. If we let V_t tend to infinity, or V_t^{-1} tend to zero, in the model equations, the observation will provide no information for $\boldsymbol{\theta}_t$ (nor for the scale parameter in the case of variance learning). In the updating equations, the one-step ahead forecast variance Q_t tends to infinity with V_t, and hence the posterior for $\boldsymbol{\theta}_t$ is just the prior, as required. Thus,

$$I_t = \{V_t^{-1} = 0\}. \tag{6.1.1}$$

Immediately following such an intervention, it may be that the series will develop in a different way than currently forecast. After an event, described above, such as a blocked catheter or an intervention such as diuretics or a fluid challenge, the patient's urine output may continue at the higher, healthier level; or the urine output may decrease below the initial spike, slowly or rapidly. The parameters in the DLM considered may drastically change or they may not change at all. If the patient's deteriorating kidneys do not respond well to the intervention it is expected that the patient's urine output will continue to be below a healthy level. Conversely, if the patient's kidneys do respond well then it is expected that the parameters will drastically change. In order to adapt to the changing pattern after the omitted observation, an additional intervention to increase uncertainty about components of $\boldsymbol{\theta}_t$ may be required. This second type of intervention is discussed in the following section.

Figure 6.2 shows the types of possible change after observing a wild observation in the second order polynomial model given by Equations (5.2.1). After observing the observation shown in red, at hour 17, there are three possibilities. This observation could be an outlier and the series could continue as usual and no parametric changes are required (shown by the purple urine outputs). The wild observation could be followed by a parametric change. This could result in a change in level parameter (shown by the green urine outputs) or a change in slope parameter (shown by the blue urine outputs). Until more observations are available, from the point of view of the forecaster at hour 17, we are uncertain which of the changes will take place. As a result, a further adjustment to the state priors is required (alongside ignoring y_t) in order to allow the model to adapt to any of the three above possible changes.

6.1.2 Additional Evolution Noise

Changes in conditions that affect the development of the series subsequently increases uncertainty about the future. This is reflected by increased uncertainties about some or all of the existing model parameters [1]. After a patient has a spike in their urine output it is unknown whether the patients urine output level will remain at a healthy level and continue around that level, or if the patients urine output will reduce back to an unhealthy, deteriorating level, or some other trend. We are more uncertain about the future after such an event and the uncertainty in

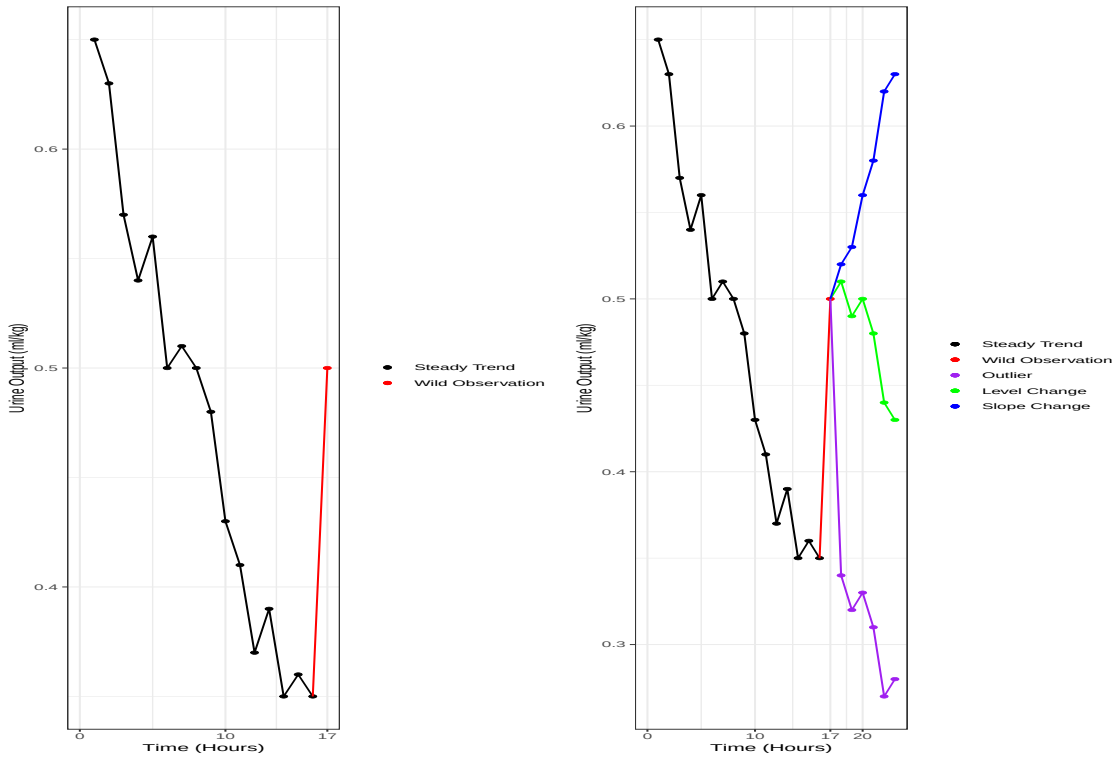

Figure 6.2: Types of change after a wild observation (shown in red). The purple, green and blue dots indicate that the red observation is an outlier, a level change, or a slope change, respectively

the parameters will be greater. One way to model this is to increase the prior variances of some (if it is evident that certain components are not going to be affected) or all of the components of the model [1]. This reflects the view that although something has changed, it is difficult to attribute the change to a particular component. In extremely noisy time series, like the urine output series, it is often the case that all elements of the covariance matrix, $\boldsymbol{R}_t$, are increased to reflect increased uncertainty about all parameters without changing the prior mean, $\boldsymbol{a}_t$, that would indicate the direction of change. This leads to great uncertainty about the entire state vector and is a technique to be used in cases with completely unknown sources of change. Therefore, unless completely uncertain about the source of change, it is best to increase uncertainties only on those components that are viewed as potentially subject to major change. In the series that we consider it is often the case that we do not know the source of uncertainty and that we have to increase the entire prior covariance matrix. The sudden changes that are frequently encountered in the urine output series are due to random biological variation and physiological responses. Many patients after having heart surgery will exhibit a decreasing urine

output as a physiological response to surgery and not due to harm to the kidneys. It is not known when this response will end but it causes the series to be extremely noisy and exhibit sudden, random changes. In order to allow our model to adapt to future trends, after such a random event, it will be the case that we must increase all parameter prior variances to incorporate the three possible changes (see Figure 6.2) that could occur after an unknown, wild observation.

All such interventions to the system evolution equations can be represented in DLM form by extending the model to include a second evolution of the state vector in addition to that in the general DLM defined by Equations (6.0.1) [1]. Suppose the intervention information is given by

$$I_t = \{\boldsymbol{h}_t, \boldsymbol{H}_t\}, \tag{6.1.2}$$

where $\boldsymbol{h}_t$ is the mean vector and $\boldsymbol{H}_t$ is the covariance matrix of a random quantity, $\boldsymbol{\xi}_t$, with

$$\boldsymbol{\xi}_t \sim \mathrm{N}(\boldsymbol{h}_t, \boldsymbol{H}_t). \tag{6.1.3}$$

The intervention is implemented by adding the additional noise term to the system evolution equation shown in Equations (6.0.1). Equivalently, the post-intervention prior distribution is defined via the extended evolution equation

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t + \boldsymbol{\xi}_t. \tag{6.1.4}$$

Thus,

$$(\boldsymbol{\theta}_t \mid I_t, D_{t-1}) \sim \mathrm{N}(\boldsymbol{a}_t^*, \boldsymbol{R}_t^*), \tag{6.1.5}$$

where $\boldsymbol{a}_t^* = \boldsymbol{a}_t + \boldsymbol{h}_t$ and $\boldsymbol{R}_t^* = \boldsymbol{R}_t + \boldsymbol{H}_t$ [1]. In the case that the source of uncertainty is unknown, $\boldsymbol{h}_t = \boldsymbol{0}$. Writing the evolution equations in this form allows for many practically useful variance inflations. This is very flexible and allows for cases of complete uncertainty through to subjective uncertainty where some of the elements of $\boldsymbol{H}_t$ can be set to 0, meaning that those parameters are unlikely to be subject to change (although there will be some correlation due to the off-diagonal terms in $\boldsymbol{R}_t$). Although it is difficult in general to be able to say that a parameter is not going to have any additional uncertainty after an intervention. In general it is better to be more cautious, and increasing parameter variances significantly (in addition to ignoring y_t by letting V_t tend to infinity) so that the model will be adaptive

to future data, rapidly identifying and estimating changes. The theorem below concretely confirms that the interventions can be written in the usual DLM form [1] (so that the interventions can be routinely updated in the recurrence relations given by Algorithms 1 and 2).

Theorem 6.1.2.1. From West and Harrison [1]. Conditional on information $\{I_t, D_{t-1}\}$, the DLM (described by Equations (6.0.1), (6.0.2), and (6.0.3)) holds with the system evolution equation in (6.0.1) amended according to Equation (6.1.4), written now as

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*, \tag{6.1.6}$$

where $\boldsymbol{\omega}_t^* = \boldsymbol{\omega}_t + \boldsymbol{\xi}_t$ is distributed as

$$\boldsymbol{\omega}_t^* \sim \mathrm{N}(\boldsymbol{h}_t, \boldsymbol{W}_t^*), \tag{6.1.7}$$

where $\boldsymbol{W}_t^* = \boldsymbol{W}_t + \boldsymbol{H}_t$.

This theorem confirms that the interventions can be written in the usual DLM form, with a generalisation to possibly non-zero mean for the evolution stochastic error series. A special case of this theorem is when $\boldsymbol{h}_t = \boldsymbol{0}$, when the addition of $\boldsymbol{\xi}_t$ simply increases the uncertainty about $\boldsymbol{\theta}_t$. An automatic intervention technique described in Section 6.5.1 uses this approach and we use it as an automatic monitoring technique for the urine output series.

6.2 Forward Intervention

Intervening in an on-line analysis to incorporate external information is achieved by adjusting the model based prior $p(\boldsymbol{\theta}_t \mid D_{t-1})$ to a new prior $p(\boldsymbol{\theta}_t \mid D_{t-1}, I_t)$, where I_t is the external information from the intervention [15]. The urine output data has many forms of sudden change which can cause the time series to be extremely noisy. When a patient's kidneys are deteriorating it is routine that the patient is given more or stronger fluids to hopefully restore the functioning of the kidneys. In this study we do not know the strengths of the fluids used and so this makes it extremely difficult to predict the effects of such interventions. However, when a

patient is given a fluid intervention we can say that uncertainty about the future is increased. Forward interventions, in the hands of an experienced forecaster, are an extremely useful way to incorporate subjective information into dynamic models.

To help convey forward interventions we will consider a time that a clinician gives a patient a "fluid challenge". The clinician will be expecting an increase in urine output at the next recording (fluid interventions that include a drug called furosemide are expected to work within 15 minutes) if the patient's kidneys are not severely failing. However, there is a lot of uncertainty about whether the mean parameters will change much at all or if they will drastically change (due to not knowing the strength of the interventions and also due to uncertainty about the performance of the patient's kidneys, although in application the clinician will know how strong the intervention is). This is not an atypical scenario for noisy time series and one way of modelling an intervention of this nature is to drastically increase the prior variances and to not adjust the prior means (setting $\boldsymbol{h}_t = \mathbf{0}$ in Equation (6.1.3)).

Suppose that (see Section 5.2.1) we model the (transformed) urine output time series by using a second order polynomial model. That is,

$$\begin{aligned} \log(Y_t + 0.1) &= \mu_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, \end{aligned} \tag{6.2.1}$$

where μ_t allows for systematic variation about a time varying level and β_t allows for systematic growth and decline of the level, where $\nu_t \sim \mathrm{N}(0, V_t)$ and $\boldsymbol{\omega}_t \sim \mathrm{T}_{n_{t-1}}(0, \boldsymbol{W}_t)$ where

$$V_t^{-1} \mid D_t \sim \mathrm{Ga}(\delta_V n_t/2, \delta_V d_t/2), \quad \boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} + W_{\beta t} & W_{\beta t} \\ W_{\beta t} & W_{\beta t} \end{pmatrix}, \tag{6.2.2}$$

where $W_{\mu t} = C_{\mu,t-1}(\delta_\mu^{-1} - 1)$ and $W_{\beta t} = C_{\beta,t-1}(\delta_\beta^{-1} - 1)$, where $\mathrm{Var}(\mu_{t-1}|D_{t-1}) = C_{\mu,t-1}$ and $\mathrm{Var}(\beta_{t-1}|D_{t-1}) = C_{\beta,t-1}$.

Suppose that our routine prior beliefs at time $t-1$ about the urine output level and growth parameters are $a_{\mu t} < 0.3$ and $a_{\beta t} < 0$, respectively. This is a scenario where a clinician may intervene with a "fluid challenge". Our prior beliefs are summarised

below

$$(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathrm{T}_{\delta_V n_{t-1}} \left[\begin{pmatrix} a_{\mu t} \\ a_{\beta t} \end{pmatrix}, \quad \begin{pmatrix} R_{\mu t} & R_{\mu\beta t} \\ R_{\mu\beta t} & R_{\beta t} \end{pmatrix} \right]. \tag{6.2.3}$$

Various strengths and compounds can have various levels of effects on different patients and it is difficult to estimate an increase in expected level. However, we know that there is some additional uncertainty in the next forecast and this can be expressed with a forward intervention by increasing prior variances. Our post-intervention prior is determined directly as

$$(\boldsymbol{\theta}_t \mid D_{t-1}, I_t) \sim \mathrm{T}_{\delta_V n_{t-1}} \left[\begin{pmatrix} a_{\mu t} \\ a_{\beta t} \end{pmatrix}, \quad \begin{pmatrix} R_{\mu t} & R_{\mu\beta t} \\ R_{\mu\beta t} & R_{\beta t} \end{pmatrix} + \boldsymbol{H}_t \right], \tag{6.2.4}$$

where $\boldsymbol{H}_t$ is commonly set to be a multiple of the most recent posterior covariance matrix, $\boldsymbol{C}_{t-1}$ (see Section 6.5.3 for more details on how to choose $\boldsymbol{H}_t$). Figure 6.3 shows the difference between including a subjective forward intervention and not including the additional subjective information. We see that the model (left) that does not incorporate this external information breaks down, when the intervention is given at hour 70, and is unable to recover. Conversely, we see that the model (right), that does incorporate this external information, prepares for additional future uncertainty by increasing prior variances at hour 70 and allows the model to rapidly adapt to the new future trend.

In Section 4.5 we considered the role of the initial prior distribution in a Bayesian updating analysis as n increases. We saw that as the sample size increases the posterior distribution resembles the likelihood more and more.

We now investigate what happens when updating from prior to posterior when we observe a level change. This is a common scenario when modelling noisy time series and is encountered frequently in the urine output series. We will consider the effect of a level change for a DLM that is closed to intervention, and the effect of a level change for a DLM that is open to intervention.

Suppose that we are modelling a stable time series and suppose that after t observations our posterior distribution, at time t, resembles the likelihood at time t (see plot for $n = 50$ in Figure 4.1).

If there is an unexpected level change in the series we see that (see Figure 6.4) the

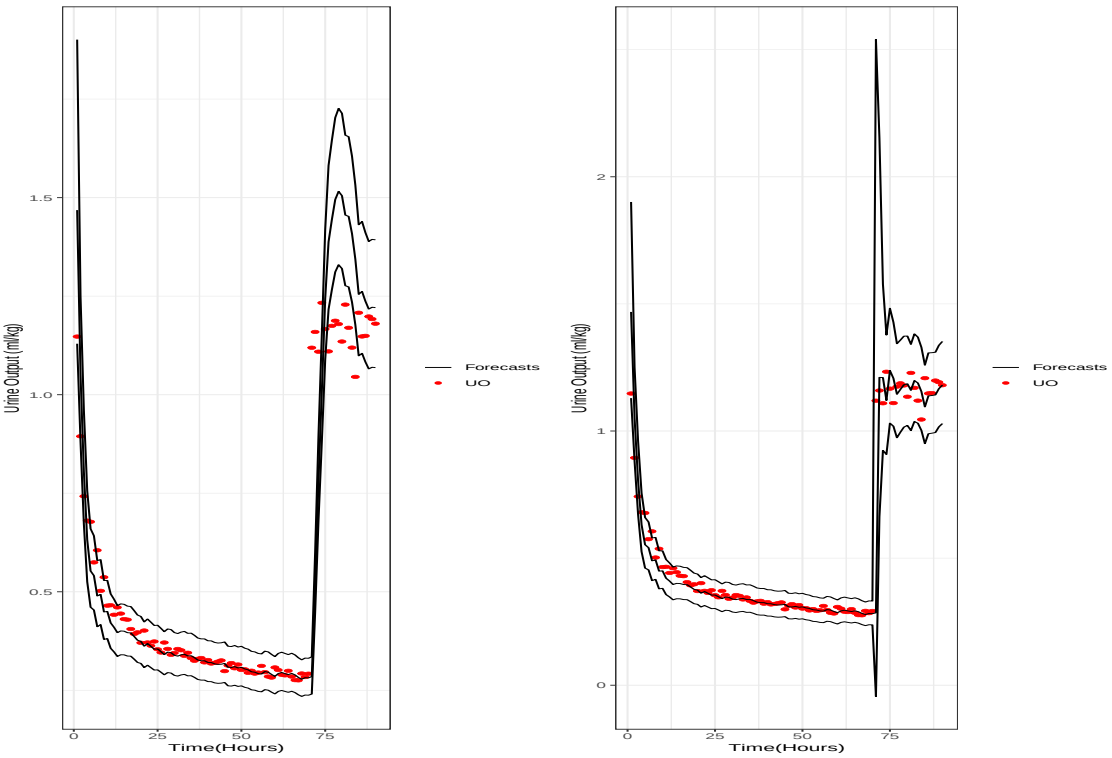

Figure 6.3: Intervention (right) Vs No Intervention (left). Forecasts for model closed to interventions are poor and prediction intervals are not useful to clinicians. Model (right) increases prior variances at hour 70 allowing model to rapidly adapt to the future trend

model is very slow to adapt to the new trend (see also Figure 6.3 (left)). This slow adaption due to a sudden level change can cause a DLM to break if not handled correctly. From Figure 6.4 we can see that after the level change, the posterior distribution (red) places density on incorrect regions of μ and has high confidence in these regions which could lead to misleading inferences.

However, if we are aware of this level change in advance (or if we account for it retrospectively like in Section 6.5) and increase prior variances we see from Figure 6.5 that the model is quickly able to adapt to the new trend and, once again, the posterior distribution resembles the likelihood (immediately). Figure 6.5 shows the updated prior (and prior to posterior updates) after including interventions $(\boldsymbol{\theta}_t \mid D_{t-1}, I_t)$ as opposed to Figure 6.4 representing a prior (and prior to posterior updates) with no intervention $(\boldsymbol{\theta}_t \mid D_{t-1})$.

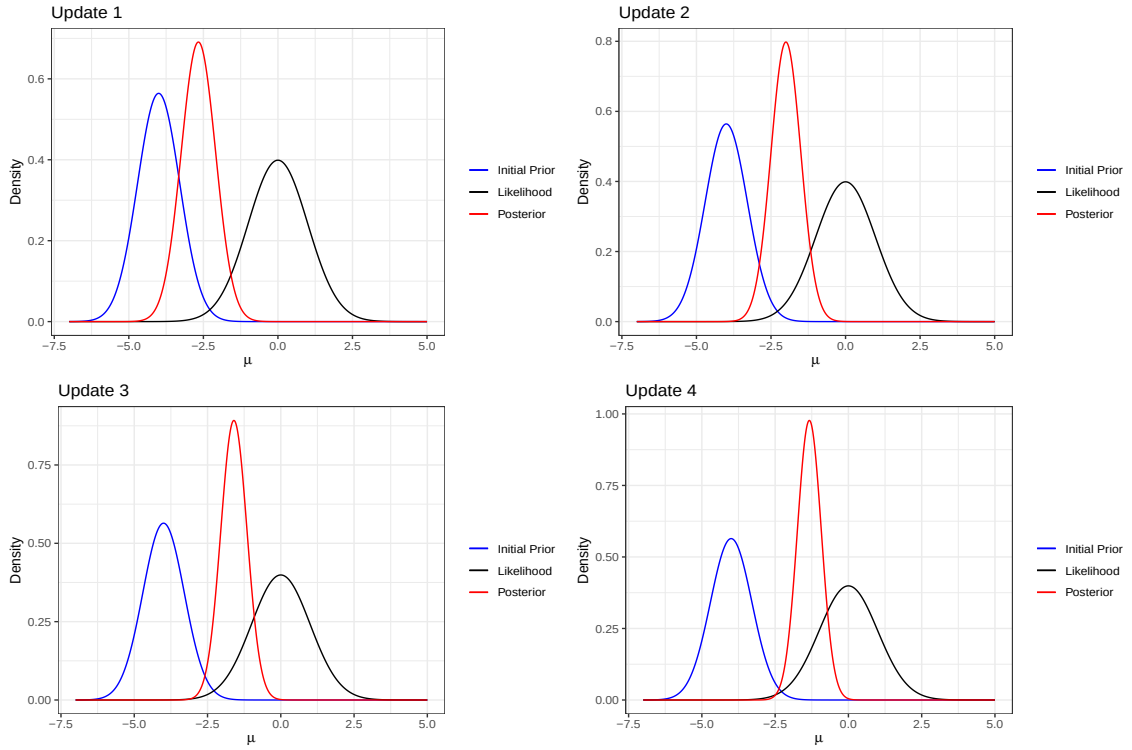

Figure 6.4: Model Breakdown. Blue lines represent prior distributions, red lines represent posterior distributions, and black lines represent the likelihoods. We see that the level is slow to adapt to the future trend. Note that this is not the updates for the series considered in Figure 6.3 (left). This Figure is for illustration purposes only

6.3 Model Performance

6.3.1 Bayes' factors for model assessment

We consider models operating subject to continual monitoring to detect deteriorations in predictive performance that indicate some form of model breakdown (in particular, changes in parameters). We now discuss automatic methods of sequentially monitoring the forecasting activity to detect breakdowns. This is essential when modelling patients since many of the clinicians around the hospital will not know how to statistically intervene, nor have the time to do so, and so automatic methods are required. Model assessment examines the extent to which the observed values of the time series are consistent with forecasts based on the model. In the DLM framework, the focus is on consistency of each observation with the corresponding one-step ahead forecast distribution [1]. Equivalently, the assessment can

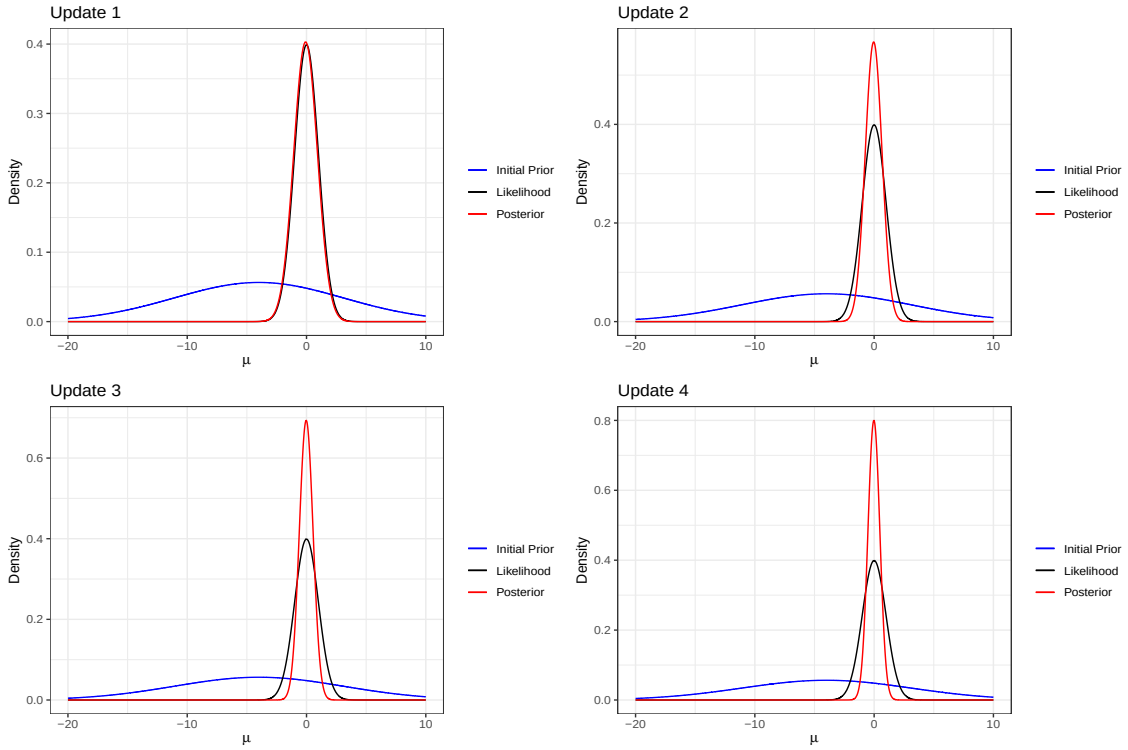

Figure 6.5: Forward intervention. Blue lines represent prior distributions, red lines represent posterior distributions, and black lines represent the likelihoods. We see that the level rapidly adapts to the future trend. Note that this is not the updates for the series considered in Figure 6.3 (right). This Figure is for illustration purposes only

be made on the basis of the standardised, one-step ahead forecast errors, measuring the extent to which they deviate from standard normal (in the case of known observational variance), or T distributed (in the case of unknown observational variance), quantiles. Assessing model performance relative to the performance of one or more alternative models is key to assessing model performance and model breakdown [1]. The alternative models under consideration are specifically designed, by the forecaster, to account for the expected changes in the time series under consideration. In the urine output series these alternative models will account for parameter changes (changes in level or growth) and outlying observations. Specific models are chosen to be based on the specific forms of departure anticipated from the routine model and model comparisons are made by using Bayes' factors, we define these below.

Consider any two models, denoted by M_0 and M_1 with the same mathematical structure, differing only through the values of defining parameters (e.g. different discount factors). That is, $M_0 = \{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_{0t}, \boldsymbol{W}_{0t}\}$, and $M_1 = \{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_{1t}, \boldsymbol{W}_{1t}\}$,

recalling that both the observational and evolution variances can be modelled with discount factors (see Sections 4.8 and 4.9.4). M_0 is the routine model that is used and monitored to detect for any deterioration in forecast performance and M_1 (there can be multiple alternative models) is an alternative model that is used to assess model performance of M_0. At time $t-1$ the models both provide a predictive distribution for Y_t given D_{t-1}. Since we now have different models we formally include this in the conditioning and write

$$p(Y_t \mid D_{t-1}, M_i) \equiv p_i(Y_t \mid D_{t-1}), \tag{6.3.1}$$

where $i = 0, 1$ in this illustration but any number of suitable models can be considered and D_{t-1} is the historical information that is common to the two models at time $t-1$. We now describe how Bayes' factors use these densities to compare model performance [1].

Definition 6.3.1.1.

- The Bayes' factor for M_0 versus M_1 based on the observed value of Y_t is defined as

$$H_t = p_0(Y_t \mid D_{t-1})/p_1(Y_t \mid D_{t-1}). \tag{6.3.2}$$

- For integers $k = 1, \ldots, t$, the Bayes' factor for M_0 versus M_1 based on the sequence of k consecutive observations $Y_t, Y_{t-1}, \ldots, Y_{t-k+1}$ is defined as

$$H_t(k) = \prod_{r=t-k+1}^{t} H_r = \frac{p_0(Y_t, Y_{t-1}, \ldots, Y_{t-k+1} \mid D_{t-k})}{p_1(Y_t, Y_{t-1}, \ldots, Y_{t-k+1} \mid D_{t-k})}. \tag{6.3.3}$$

These Bayes' factors provide measures of predictive performance of M_0 relative to M_1. For each k, $H_t(k)$ measures the evidence provided by the most recent (up to and including time t) k consecutive observations, and is used to detect slower changes in parameter values. Some features of Bayes' factors include:

- Setting $k = 1$ in $H_t(k)$ leads to $H_t(1) = H_t$, for all t and taking $k = t$ gives the Bayes' factor based on all of the data, $H_t(t)$.
- The Bayes' factor for M_1 versus M_0 are the reciprocals of those for M_0 versus M_1, $H_t(k)^{-1}$.

- Evidence for or against the model, M_0, accumulates multiplicatively as data are processed. For each $t > 1$,

$$H_t(k) = H_t H_{t-1}(k-1), \qquad (k = 2, \ldots, t). \tag{6.3.4}$$

- On the log scale, evidence is additive, with

$$\log(H_t(k)) = \log(H_t) + \log(H_{t-1}(k-1)), \qquad (k = 2, \ldots, t). \tag{6.3.5}$$

- Following Jeffreys (1961) [29], a log Bayes' factor of 1 (-1) indicates evidence in favour of model 0 (1), a value of 2 or more (-2 or less) indicates strong evidence in favour of model 0 (1). The value 0 indicates no evidence either way.

Note that we do not need to construct a fully specified alternative model for the data, a suitable sequence of alternative, one-step ahead forecast densities to provide the denominator for the Bayes' factors are all that is needed (see Section 6.5.2).

6.3.2 Cumulative Bayes' Factors

The overall Bayes' factor, $H_t(t)$, is used as a measure of overall model assessment. However, with dynamic models, the focus is on local model performance. We do not want high Bayes' factors from the past (reflecting strong evidence that the routine model is outperforming the alternatives) affecting Bayes' factors calculated now. Hence, the individual Bayes' factors H_t and the cumulative measures $H_t(k)$ for $k < t$ are key for assessing dynamic model performance [1]. Suppose that $t = 11$ and that $H_r = 2$ for $r \leq 10$ meaning that the first 10 observations are well in accord with the routine model M_0, their individual Bayes' factors each being 2. Thus the cumulative Bayes' factor, representing the cumulative evidence for M_0 relative to M_1 from the first 10 observations, $H_{10}(10) = 2^{10} = 1024$. Then if Y_{11} is an outlier under M_0, even if it is out in an extreme tail of the forecast distribution, the cumulative Bayes' factor $H_{11}(11)$ will likely still exceed 1, thus the cumulative Bayes' factor indicates no evidence against M_0, although by definition it is clearly an observation that is not in accord to the predictions of M_0. We would require $H_{11} \leq 1/1024$, indicating that Y_{11} is 1024 times more likely to come from the alternative model M_1 compared to M_0, to even begin to doubt M_0. The evidence for the routine model

from earlier observations dominates the evidence against the standard model at time $t = 11$ in the cumulative measure. Hence the need to consider the individual Bayes' factors as they arise for the case of outlying observations. However, in some cases, when there are gradual changes in the time series, the individual Bayes' factors will be smaller than one (indicating model deterioration) but not small enough to prompt a warning signal. These small changes in the time series can be identified by looking back over recent observations and calculating cumulative Bayes' factors, but discarding past information that favoured M_0 so that the cumulative factor is not dominated by evidence in favour of the routine model [1].

Recall earlier how we mentioned that a urine output observation from 10 hours prior to the current time is less useful than the most recent observations when forecasting the future. The same is true when monitoring model performance, so that past evidence does not affect current decisions. We should only be considering local Bayes' factors when monitoring forecast performance. A small individual Bayes' factor $H_t(1) = H_t$ provides a warning of a possible outlier or the onset of change in the series at time t. A small $H_t(k)$ for $k > 1$ is indicative of possible changes having taken place $k - 1$ steps back in the past. The following theorem illustrates the methods described above (adapted from [1]).

Theorem 6.3.2.1. With H_t and $H_t(k)$ as in Definition 6.3.1.1, let

$$L_t = \min_{1 \le k \le t} H_t(k), \tag{6.3.6}$$

with $L_0 = 1$ so that $L_1 = H_1(1)$. Then the quantiles L_t are updated sequentially over time by

$$L_t = H_t \min\{1, L_{t-1}\} \tag{6.3.7}$$

for $t > 1$. The minimum at time t is taken at $k = l_t$, with $L_t = H_t(l_t)$, where the integers l_t are updated sequentially via

$$l_t = \begin{cases} 1 + l_{t-1}, & \text{if } L_{t-1} < 1, \\ 1, & \text{if } L_{t-1} \ge 1. \end{cases} \tag{6.3.8}$$

Proof: By definition $H_t(1) = H_t$ and for $2 \le k \le t$, $H_t(k) = H_t H_{t-1}(k-1)$. Thus

$$\begin{aligned} L_t &= \min\{H_t, \min_{2\le k\le t} H_t H_{t-1}(k-1)\} \\ &= H_t \min\{1, \min_{2\le k\le t} H_{t-1}(k-1)\} \\ &= H_t \min\{1, \min_{1\le j\le t} H_{t-1}(j)\} \\ &= H_t \min\{1, L_{t-1}\}, \end{aligned} \tag{6.3.9}$$

as stated. By (6.3.6) we can write $L_t = H_t(l_t)$, where $1 \le l_t \le t$. By definition $L_1 = H_1 = H_1(1) = H_1(l_1)$, where $l_1 = 1$. Consider time $t = 2$, using (6.3.7) we can write

$$L_2 = \begin{cases} H_2(1)L_1, & \text{if } L_1 < 1, \\ H_2(1)H_1(l_1), & \\ = H_2(1 + l_1), & \\ H_2(1), & \text{if } L_1 \ge 1. \end{cases} \tag{6.3.10}$$

Noting that we used the multiplicative property (6.3.4) $H_t(k) = H_t H_{t-1}(k-1)$. In other words, $L_2 = H_2(l_2)$, where

$$l_2 = \begin{cases} 1 + l_1, & \text{if } L_1 < 1, \\ 1, & \text{if } L_1 \ge 1. \end{cases} \tag{6.3.11}$$

Consider now time $t = s$, we can write

$$L_s = \begin{cases} H_s(1)L_{s-1}, & \text{if } L_{s-1} < 1, \\ = H_s(1)H_s(l_{s-1}), & \\ = H_s(1 + l_{s-1}), & \\ H_s(1), & \text{if } L_{s-1} \ge 1. \end{cases} \tag{6.3.12}$$

In other words, $L_s = H_s(l_s)$, where

$$l_s = \begin{cases} 1 + l_{s-1}, & \text{if } L_{s-1} < 1, \\ 1, & \text{if } L_{s-1} \geq 1. \end{cases} \tag{6.3.13}$$

□

The sequence of local Bayes' factors, L_t, provides the basis of a local monitoring scheme, with the run-length, l_t, indicating how long ago the model deterioration may have begun by counting the number of recent observations contributing to the minimum cumulative Bayes' factor. If there is evidence in favour of the routine model at time $t-1$, the model is deemed adequate and future judgement will ignore the past. At the next time, t, the local Bayes' factor is set to the Bayes' factor for just that time, $L_t = H_t$, and decisions about possible inadequacies of the model are based on Y_t alone and the local Bayes' factor is not affected by past evidence favouring the routine model [15]. If this single Bayes' factor is "very small" (we define "very small" to be a Bayes' factor that breaks a pre-specified threshold, see Section 6.4. The Bayes' factor threshold is chosen by the forecaster and will depend on experience and on the system under consideration), then the observation, Y_t, is a possible outlier or may indicate a parametric change. If this single Bayes' factor exceeds one, then the model is adequate and the system rolls forward once more as before. Furthermore, if this Bayes' factor lies between these two extremes then there is some evidence against the model, but not enough to signal a warning. In this scenario, the evidence is held over for combined judgement with later evidence by using local Bayes' factors. If there is inconclusive evidence against the routine model before time t, so that the local Bayes' factor is less than one, $L_{t-1} < 1$, then evidence from the current observation is cumulated through the Bayes' factor product $L_t = H_t L_{t-1}$, and the run length increases by one, $l_t = l_{t-1} + 1$. If L_t now breaks the Bayes' factor threshold, then the l_t most recent observations together suggest evidence of reduced forecast performance. If L_t exceeds one, then, the model is deemed adequate and previous concerns are forgotten. If there is still no decision either way, the evidence is held over once again [15]. Finally, if the local Bayes' factor, L_t, is consistently less than one but does not break the pre-specified threshold, then a run-length threshold can be specified. That is, if $L_t < 1$ and $l_t > l$

(where l is chosen by the forecaster), then a warning can be raised, indicating a slow parametric change causing deterioration to model performance.

6.4 Urine Output Series: Corrective Feedback

Consider a series of forecasts that begin favouring the routine model, e.g., this could be a patient who has a steady downward trend and is forecasted adequately by the routine model. Following Jeffreys (1961) [29] we will consider a monitoring system that signals a warning when the cumulative Bayes' factor falls below a threshold $\tau = \exp(-2) \approx 0.135$ which corresponds to -2 on the log-scale. If the cumulative Bayes' factor falls below $\tau = 0.135$ this will warn that there is significant evidence for the alternative model and against the routine model.

At time $t = 1$ (recalling that $L_0 = l_0 = 1$)

$$\begin{aligned} H_1 = 1.5 \implies L_1 &= H_1 \times \min(1, L_0) = H_1 = 1.5, \\ l_1 &= 1. \end{aligned}$$

At time $t = 2$

$$\begin{aligned} H_2 = 2 \implies L_2 &= H_2 \times \min(1, 1.5) = H_2 = 2, \\ l_2 &= 1. \end{aligned}$$

As we can see, the favourable evidence in the first forecast is excluded from the assessment of the second. While forecasts favour the routine model, the local Bayes' factor equals the individual Bayes' factor for the current observation and the run length is one. No signals are generated. Suppose we now observe some poor forecasts. These poor forecasts could be due to an abrupt change in the urine output series. This could be due to the patient adjusting to the surgery they just had (overcoming the physiological response to surgery) or due to an increased amount of fluids given by clinicians.

At time $t = 3$

$$\begin{aligned} \tau = 0.135 < H_3 = 0.4 < 1 \implies L_3 &= H_3 \times \min(1, 2) = H_3 = 0.4, \\ l_3 &= 1. \end{aligned}$$

There is some evidence of model inadequacy, $L_3 < 1$, but there is not enough evidence to signal a warning, since $L_3 > \tau$. The evidence is held over until the next time. At time $t = 4$

$$\tau < H_4 = 0.3 < 1 \implies L_4 = H_4 \times \min(1, 0.4) = 0.12,$$
$$l_4 = 2.$$

The monitor signals a warning ($L_4 < \tau$) with evidence accumulated over two periods, $l_4 = 2$, a run-length of two. This indicates that two observations contributed to the evidence and that the model began to breakdown at time $t = (4-2+1) = 3$. At this point the forecaster (or the automatic diagnostics system) is warned of a potential problem and the model is adjusted accordingly before continuing. The monitor then resets to a neutral state before proceeding. The evidence that led to a warning at time $t = 4$ is removed because it has already been brought to the attention of the forecaster (or the automatic diagnostics system) and appropriate adjustments made. Just as a build-up of favourable evidence is prevented from overpowering current monitor performance, so is a build-up of unfavourable evidence once diagnostics have been performed [15].

At time $t = 5$

$$\tau < H_5 = 0.65 < 1 \implies L_5 = H_5 \times \min(1, 1) = H_5 = 0.65,$$
$$l_5 = 1.$$

At time $t = 6$

$$\tau < H_6 = 0.85 < 1 \implies L_6 = H_6 \times \min(1, 0.65) = 0.5525,$$
$$l_6 = 1 + 1 = 2.$$

Evidence is building up against the routine model, $L_6 < 1$ but it is not yet sufficient to cause a warning signal since $L_6 > \tau$, so evidence is held over. At time $t = 7$

$$H_7 = 3 \implies L_7 = H_7 \times \min(1, 0.5525) = 1.6575,$$
$$l_7 = 2 + 1 = 3.$$

Observation Y_7 strongly favours the routine model, $L_7 > 1$, and so removes the concern that had been building and the model performance is considered satisfactory, and so the evaluation next time will start from a neutral position and the run-length

starts over. At time $t = 8$

$$H_8 = 0.5 < 1 \implies L_8 = H_8 \times \min(1, 1.6575) = H_8 = 0.5,$$
$$l_8 = 1.$$

Note that if $H_7 = 1.5$ then $L_7 = 1.5 \times 0.5525 = 0.82875$ and although observation Y_7 would have favoured the routine model, the model performance would still have been inconclusive and the evidence would have, once again, been held over.

One final situation remains. It is possible for a series of observations to consistently provide evidence against the model, but in small enough quantities that the run-length grows considerably without the Bayes' factor threshold being breached. This can indicate slowly changing parameter values and should raise a performance warning. This slow change is made aware of by extending the monitor system to issue a signal whenever the run-length exceeds a preset limit in addition to signalling when the Bayes' factor threshold is passed.

6.5 FeedBack Intervention

6.5.1 Automatic Detection and Diagnosis

Identifying model breakdown is the first step in automatic monitoring, detection and diagnosis. The signals that are raised when a model is underperforming should prompt a user response in the form of a feedback intervention. The following logical scheme (adapted from [1]) provides a guide for the use of Bayes' factors in detecting and diagnosing model breakdowns.

At time t proceed as follows:

(A) Calculate the single Bayes' factor H_t. If $H_t \geq \tau$, then Y_t is viewed as consistent with M_0; proceed to (B) to assess the possibilities of model failure from changes in parameter values prior to time t. However, if $H_t < \tau$, then Y_t is a potential outlier and should be omitted from the analysis, being treated as a missing value. However, Y_t may be the onset of change in the model parameters and these changes must be allowed for after rejecting the observation. Thus the

need for intervention is identified and we can proceed to (C).

(B) Calculate the local cumulative Bayes' factor, L_t, and the corresponding run-length, l_t, to assess the possibility of changes prior to time t. If $L_t \geq \tau$ and $l_t < l$, then M_0 is satisfactory and proceed to (D) to perform standard updates. Otherwise, $L_t < \tau$ indicates change that should be addressed, requiring intervention; proceed to (C). Slower parametric changes can also be identified by signalling a possible breakdown of M_0 if either $L_t < \tau$ or $l_t > l$ (where l is chosen by the forecaster). The idea being that l recent observations may provide evidence marginally favouring M_1 over M_0, but this may be so small that $\tau < L_t < 1$, and so a run-length threshold is required. If $L_t > \tau$ but $l_t > l$ then this indicates change that should be addressed, requiring intervention; proceed to (C).

(C) Issue signal of model deterioration and perform model diagnostics, as in Section 6.1.2, to adapt the model for the future. Update the time index to $t+1$ for the next observation stage, and proceed to (A), reinitialising monitoring by setting $L_t = 1$. The prior $(\boldsymbol{\theta}_{t+1} \mid D_t, I_t)$ will be used at time $t+1$, where I_t is an external intervention as described in Section 6.1.

(D) Perform usual analysis and updating with M_0, proceeding to (A) at time $t+1$. The prior $(\boldsymbol{\theta}_{t+1} \mid D_t)$ will be used at time $t+1$.

We now specify the forms of intervention at points of possible changes. We have described interventions in Section 6.1 and we now focus on automatic alternatives for routine use.

6.5.2 Alternative Models for the DLM

Suppose that the routine model, M_0, to be assessed is a standard normal DLM producing one-step ahead forecast distributions $(Y_t \mid D_{t-1}) \sim \mathrm{N}(f_t, Q_t)$. Assessing consistency of the observed values Y_t with these distributions is equivalent to assessing the consistency of the standardised forecast errors $e_t/Q_t^{1/2} = (Y_t - f_t)/Q_t^{1/2}$ with the standard normal distribution [1]. Thus we will use Bayes' factors based on the predictive densities of the standardised forecast errors.

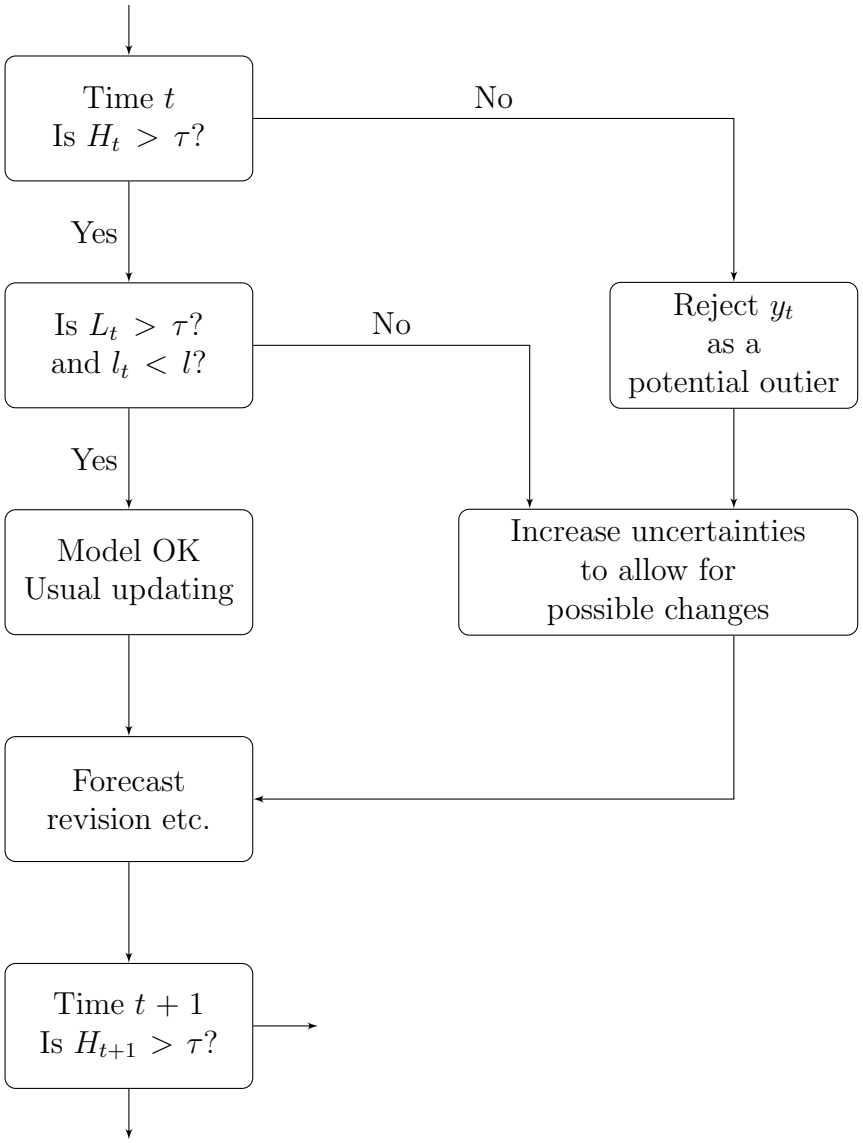

Figure 6.6: Monitoring system illustrating the steps performed at each time point to identify model deterioration

Recall that the focus is on the local performance of the model. It is the extent to which the current and most recent observations accord with the model that determines whether or not some form of intervention is required. Consider a single observation and hence a single forecast error, and without loss of generality take $f_t = 0$ and $Q_t = 1$, so that under M_0 the forecast distribution for $e_t = Y_t$ is

$$(e_t \mid D_{t-1}, M_0) \sim \mathrm{N}(0, 1), \tag{6.5.1}$$

and so

$$p_0(e_t \mid D_{t-1}) = (2\pi)^{-1/2} \exp\{-0.5e_t^2\}. \tag{6.5.2}$$

There are various alternative models, M_1, that provide for the types of departure from M_0 encountered in practice for the urine output series. The first alternative model can be used to capture a level change after a patient has received a successful intervention, or after a drastic increase marking the end of a physiological response to surgery. A suitable alternative model, M_1, to capture the level change, in which e_t has non-zero mean, h, is one with forecast distribution for e_t given by $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(h, 1)$ [1]. This reflects that the errors are shifted and $\mathrm{E}[(e_t \mid D_{t-1}, M_1)] = h$, representing a level shift by hml/kg. Thus we have

$$p_1(e_t \mid D_{t-1}) = (2\pi)^{-1/2} \exp\{-0.5(e_t - h)^2\}. \tag{6.5.3}$$

For any fixed shift h, the Bayes' factor at time t is

$$H_t = p_0(e_t \mid D_{t-1})/p_1(e_t \mid D_{t-1}) = \exp\{0.5(h^2 - 2he_t)\}. \tag{6.5.4}$$

Ranges of appropriate values of h will depend on the size of the error and the Bayes' factor threshold. For example, suppose that the error is positive and consider the point at which $H_t = 1$, indicating no evidence for or against M_0, from e_t alone (in Section 6.3.2 we discussed that although $H_t \geq 1$ the cumulative Bayes' factor could still be less than 1 and so there would be evidence against the routine model even though there is no evidence against M_0 from e_t. But here we are analysing the individual Bayes' factor). At this point, $\log(H_t) = 0$ and so we have $0.5(h^2 - 2he_t) = 0$ and hence, since h is non-zero, $h = 2e_t$. To be indifferent between M_0 and M_1 based from, say, $e_t \approx 2$, alone (at roughly the 97.5% point of the standard normal distribution) suggests that $h = 4$. This means that if we considered an alternative model, such that, $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(4, 1)$ and an error of $e_t = 2$ is observed, then $H_t = 1$, and there is no evidence, from $e_t = 2$ alone, against the routine model and assuming that the cumulative Bayes' factor is also above a chosen threshold and the run-length is below the chosen threshold (see Section 6.3.2) there would be no warning from the monitoring system and forecasts would be made for time $t + 1$ as usual.

Suppose that the forecaster specifies a threshold $H_t = \tau, (0 < \tau \ll 1)$ below which the evidence is accepted as a strong indication that e_t (alone) is inconsistent with M_0.

Using Equation (6.5.4), taking logs and rearranging, leads to a quadratic equation in h, namely

$$h^2 - 2he_t - 2\log(\tau) = 0. \tag{6.5.5}$$

Thus if we, once again, take $e_t = 2$ (at roughly the 97.5% point of the standard normal distribution) and, using a threshold of -2 for the log-Bayes' factor (Jeffreys (1961) [29]), i.e. taking $\tau = e^{-2} \approx 0.135$, we obtain

$$\begin{aligned} h^2 - 4h + 4 &= 0 \\ \implies h &= 2. \end{aligned} \tag{6.5.6}$$

Thus, $e_t \approx 2$ leads to strong evidence against M_0 in favour of M_1 when $h = 2$, and indifference between M_0 and M_1 when $h = 4$. This reflects that an observation at time t with a corresponding error $e_t \approx 2$ is equally likely to be from the standard normal distribution or from the $\mathrm{N}(4, 1)$ distribution but much more likely to be from the $\mathrm{N}(2, 1)$ distribution. Moreover, if we consider using $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(2, 1)$ as our alternative model then the Bayes' factor, at time t, is given by

$$H_t = p_0(e_t \mid D_{t-1})/p_1(e_t \mid D_{t-1}) = \exp\{0.5(4 - 4e_t)\}. \tag{6.5.7}$$

For $H_t < \tau \approx 0.135$ this would mean that

$$\begin{aligned} \log(H_t) = 0.5(4 - 4e_t) &< -2 \\ \implies e_t &> 2. \end{aligned} \tag{6.5.8}$$

Illustrating that if the error $e_t > 2$ is observed then this error will be enough evidence to breach the chosen threshold and indicate that the model performance is poor compared to M_1 and diagnostics can be performed. Finally, very importantly, if the forecaster decides that for a single observation to be considered an outlier, (following Jeffreys (1961) [29] and using a threshold of -2 for the log-Bayes' factor) and not in accord to the routine model, it must be outside of the 95% (highest density) prediction interval (which for positive e_t means above the 97.5% point of the standard normal distribution) at time t then the alternative model $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(2, 1)$ would suffice. Using $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(2, 1)$ for the automatic monitoring system allows the monitor to signal a warning and perform diagnostics at the levels defined above. Similarly, if the forecaster decides that an error outside of the 90% (highest density) prediction interval at time t is considered large enough for the model to be performing inadequately, then similar values of h can be defined via the above

analysis. Note if h is negative the sign of h must be reversed in the above findings.

Now consider a different form for the alternative model M_1 (although in practice a forecaster can define many alternative models). A useful alternative is the scale shift model, M_1, in which e_t has standard deviation k rather than unity [1], i.e. $(e_t \mid D_{t-1}, M_1) \sim \mathrm{N}(0, k)$, with

$$p_1(e_t \mid D_{t-1}) = (2\pi k^2)^{-1/2} \exp\{-0.5(e_t/k)^2\}. \tag{6.5.9}$$

Modelling a scale inflation with $k > 1$ provides a useful alternative model for modelling noisy time series like the urine output series. In Section 6.3.2 we discussed that when modelling on-line an outlier is indistinguishable from a parametric change. In this scenario we mentioned that a good response to such uncertainty is to drastically increase the prior variances and to leave the prior means unchanged (see Section 6.1.2). By increasing the prior variances substantially this allows the model to rapidly adapt to parametric changes (see Figure 6.5). In addition, the model will also be able to rapidly adapt if the wild observation corresponds to an outlier. The alternative model (6.5.9) captures such changes that the routine model does not. Using the alternative model (6.5.9), the Bayes' factor at time t is given by

$$H_t = k \exp\{-0.5e_t^2(1 - k^{-2})\}. \tag{6.5.10}$$

For the models M_0 and M_1 to be indifferent, $H_t = 1$, i.e. $\log(H_t) = 0$, implies that

$$\log(k) - 0.5e_t^2\left(\frac{k^2 - 1}{k^2}\right) = 0. \tag{6.5.11}$$

Once again, we will consider $e_t \approx 2$ meaning that e_t lies outside of the upper 97.5% point of the standard normal distribution. Substituting $e_t = 2$ gives

$$\begin{aligned} \log(k) - 2\left(\frac{k^2 - 1}{k^2}\right) &= 0 \\ \implies \left(\frac{k^2}{k^2 - 1}\right)\log(k) - 2 &= 0. \end{aligned} \tag{6.5.12}$$

An explicit expression for k does not exist, but we can use numerical techniques to find that the solution lies in the range $7 < k < 8$ (see Figure 6.7). For e_t to equal the threshold defined by Jeffreys (1961) (a log-Bayes' factor of -2 [29]) we would

require that

$$\begin{aligned} \log(k) - 0.5e_t^2\left(\frac{k^2-1}{k^2}\right) &= -2 \\ \implies \left(\frac{k^2}{k^2-1}\right)[\log(k)+2] - 0.5e_t^2 &= 0. \end{aligned} \tag{6.5.13}$$

It is not possible to obtain a value of k which satisfies the above when $|e_t| = 2$. We can see from Figure 6.7 that if the monitoring system was programmed to signal a warning when $|e_t| = 2$ the threshold would have to be around 0.45 and $k = 2$ (or higher thresholds for different values of k). If $k = 7$ and we observed an error of size $|e_t| = 2$ this would indicate (close to) indifference between M_0 and M_1. For the threshold $\tau = 0.135$ (the dashed horizontal line in Figure 6.7) we see that all values of k would produce an error if $|e_t| \gtrsim 2.9$.

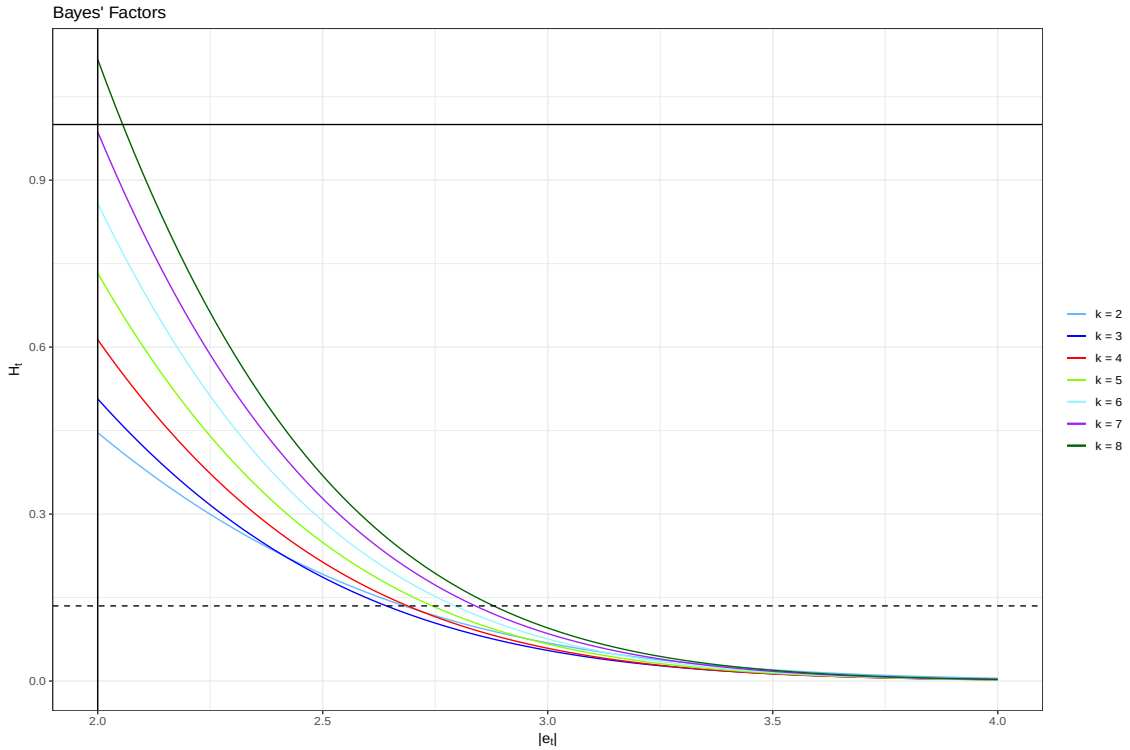

Figure 6.7: Bayes' factors H_t for normal scale inflation model plotted against $|e_t|$ for different scale inflations k. Black horizontal line at $H_t = 1$ represents indifference between M_0 and M_1. Black dashed line at ≈ 0.135 represents our chosen threshold to indicate poor model performance

A key point to make is that, above a certain level, the particular value of k is irrelevant. The point is that alternative models provide a larger variance than M_0, thus large errors will tend to be more consistent with M_1 no matter how large

the variance inflation is [1]. This is illustrated in Figure 6.7. The curves plotted provide Bayes' factors as a function of $|e_t|$ over the range of $2-4$, for $k = 2, 3, \ldots, 8$. This range of values of $|e_t|$ focuses attention on the region of observations that would make us doubt M_0. As expected, evidence in favour of M_0 decreases as $|e_t|$ increases. Notice from Figure 6.7 that as $|e_t|$ becomes extreme relative to M_0, the choices of k are indistinguishable. The Bayes' factors are all around 0.1 when $|e_t| = 2.9$.

6.5.3 Automatic Adaption in Cases of Parametric Change

Consider we are at (C) in the scheme described in Section 6.5.1, having identified that changes have occurred and M_0 is underperforming. Intervening by adding additional evolution noise (see Section 6.1.2) at the time of suspected change allows the routine model, M_0, to adapt to the changes automatically. In the case of the urine output series, when changes are abrupt and we cannot identify which parameters are subject to change, it is best to be cautious and increase all prior variances. This feedback intervention at (C) can be automatically achieved by following Theorem 6.1.2.1 with $\boldsymbol{h}_t = \mathbf{0}$ such that $\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1}+\boldsymbol{\omega}_t^*$, where, $\boldsymbol{\omega}_t^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{W}_t+\boldsymbol{H}_t)$ (if we are considering an unknown observational variance $\boldsymbol{\omega}_t^* \sim \mathrm{T}_{n_{t-1}}(\mathbf{0}, \boldsymbol{W}_t + \boldsymbol{H}_t)$). This is the form of intervention that we will use to adapt the model described in Section 5.2.

There are many ways to specify the additional variance matrix $\boldsymbol{H}_t$. One way to specify $\boldsymbol{H}_t$, when all parameters are subject to change, is in terms of the standard evolution variance matrix $\boldsymbol{W}_t$. Taking

$$\boldsymbol{H}_t = (\rho - 1)\boldsymbol{W}_t, \tag{6.5.14}$$

where $\rho > 1$, has the effect of inflating the standard variance of $\boldsymbol{\theta}_t$ by a factor of ρ [1].

An alternative specification can be seen by recalling from Section 4.8 that low values of discount factors are consistent with high variability in the $\boldsymbol{\theta}_t$ sequence. Thus reducing the discount factors, resulting in a more heavily discounted version of the matrix $\boldsymbol{W}_t$, when the monitoring system signals a warning is a way for the model to adapt to the onset of change. This is possible by specifying $\boldsymbol{H}_t$ so that $\boldsymbol{W}_t^* = \boldsymbol{W}_t + \boldsymbol{H}_t$ is of that form. If we consider $\boldsymbol{W}_t$ to take the form given by Equation (6.2.2), i.e. an evolution matrix with separate discount factors for the

level and growth parameters

$$\boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} + W_{\beta t} & W_{\beta t} \\ W_{\beta t} & W_{\beta t} \end{pmatrix}, \tag{6.5.15}$$

where $W_{\mu t} = C_{\mu,t-1}(\delta_\mu^{-1} - 1)$ and $W_{\beta t} = C_{\beta,t-1}(\delta_\beta^{-1} - 1)$, where $\text{Var}(\mu_{t-1} \mid D_{t-1}) = C_{\mu,t-1}$ and $\text{Var}(\beta_{t-1} \mid D_{t-1}) = C_{\beta,t-1}$. Then if we denote the reduced discount factors by δ_μ^* and δ_β^* then

$$\boldsymbol{W}_t^* = \begin{pmatrix} W_{\mu t}^* + W_{\beta t}^* & W_{\beta t}^* \\ W_{\beta t}^* & W_{\beta t}^* \end{pmatrix}, \tag{6.5.16}$$

where $W_{\mu t}^* = C_{\mu,t-1}(\delta_\mu^{*-1} - 1)$ and $W_{\beta t}^* = C_{\beta,t-1}(\delta_\beta^{*-1} - 1)$. This results in the following form for $\boldsymbol{H}_t = \boldsymbol{W}_t^* - \boldsymbol{W}_t$

$$\boldsymbol{H}_t = \begin{pmatrix} C_{\mu,t-1}(\delta_\mu^{*-1} - \delta_\mu^{-1}) + C_{\beta,t-1}(\delta_\beta^{*-1} - \delta_\beta^{-1}) & C_{\beta,t-1}(\delta_\beta^{*-1} - \delta_\beta^{-1}) \\ C_{\beta,t-1}(\delta_\beta^{*-1} - \delta_\beta^{-1}) & C_{\beta,t-1}(\delta_\beta^{*-1} - \delta_\beta^{-1}) \end{pmatrix}. \tag{6.5.17}$$

A similar discount lowering technique can be used for the observation equation. Recall from Section 4.9.4 that

$$\phi_t \mid D_{t-1} \sim \text{Ga}(\delta_V n_{t-1}/2, \delta_V d_{t-1}/2). \tag{6.5.18}$$

When a single Bayes' factor is lower than the Bayes' factor threshold at time t, it implies that the observation at time t is either an outlier or the onset of change. In this scenario, we ignore Y_t (as well as adding additional evolution variance). This can be achieved by letting S_t (the estimate for V_t) tend to infinity (see Section 6.1.1). Note that if the diagnostic threshold is broken because of more than one observation, from using local Bayes' factors with more than one observation or from breaking the run-length, then additional observation variance is not added, since these signals are not indicative of an outlier. One way of letting S_t tend to infinity is to reduce the observational discount factor, δ_V, to a lower value, δ_V^*.

One way to do this and also follow the DLM framework is by extending Theorem 6.1.2.1. The extended observation equation takes a similar form where

$$Y_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + \nu_t^*, \tag{6.5.19}$$

where $\nu_t^* = \nu_t + \psi_t$ (where ν_t and ψ_t are uncorrelated) is distributed as

$$\nu_t^* \sim \mathrm{N}(0, V_t^*), \tag{6.5.20}$$

where $V_t^* = V_t + \Psi_t$. Reducing the observational discount factor at times of possible outliers can be achieved by specifying $\Psi_t = V_t^* - V_t$, where

$$(\phi_t \mid D_{t-1}, I_t) \sim \mathrm{Ga}(\delta_V^* n_{t-1}/2, \delta_V^* d_{t-1}/2), \tag{6.5.21}$$

is the updated prior distribution.

Notice from Algorithm 2 that

$$\begin{aligned} n_t &= \delta_V n_{t-1} + 1, \\ d_t &= \delta_V d_{t-1} + S_{t-1} e_t^2 / Q_t, \\ S_t &= \frac{d_t}{n_t} = \frac{\delta_V d_{t-1} + S_{t-1} e_t^2 / Q_t}{\delta_V n_{t-1} + 1}. \end{aligned} \tag{6.5.22}$$

At times of possible outliers we would like

$$S_t^* = \frac{d_t^*}{n_t^*} = \frac{\delta_V^* d_{t-1} + S_{t-1} e_t^2 / Q_t}{\delta_V^* n_{t-1} + 1}. \tag{6.5.23}$$

This can be achieved by setting the additional point estimate for the unknown observational variance $S_t^\psi = S_t^* - S_t$. It follows that the additional point forecast for the observational variance is

$$\begin{aligned} S_t^\psi &= \frac{\delta_V^* d_{t-1} + S_{t-1} e_t^2 / Q_t}{\delta_V^* n_{t-1} + 1} - \frac{\delta_V d_{t-1} + S_{t-1} e_t^2 / Q_t}{\delta^* n_{t-1} + 1} \\ &= \frac{(\delta_V n_{t-1} + 1)(\delta_V^* d_{t-1} + S_{t-1} e_t^2 / Q_t) - (\delta_V^* n_{t-1} + 1)(\delta_V d_{t-1} + S_{t-1} e_t^2 / Q_t)}{(\delta_V^* n_{t-1} + 1)(\delta_V n_{t-1} + 1)} \\ &= \frac{(\delta_V^* - \delta_V)(d_{t-1} - n_{t-1} S_{t-1} e_t^2 / Q_t)}{(\delta_V^* n_{t-1} + 1)(\delta_V n_{t-1} + 1)}. \end{aligned} \tag{6.5.24}$$

The forms of $\boldsymbol{H}_t$ and S_t^ψ above allow us to automatically add additional evolution variance and observational variance when model deteriorations are detected.

6.6 Intervention Modelling

In Sections 5.2.1 and 5.6.1 it was mentioned that higher discount factors usually produce better forecast performance. However, due to the time series being extremely noisy this pulled component discount factors in a downward direction and so we had to make a compromise. We now add a monitoring system to detect model deterioration, and as a result we will not have to compromise by lowering component discount factors. The theory described throughout this chapter will be applied to the urine output series to allow the routine model described in Section 5.2 to handle outliers and to adapt to changing trends by reducing evolution and observation discount factors, using the methods defined in Section 6.5.3. The evolution equation is now adapted and takes the form

$$\begin{aligned} \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t + \boldsymbol{\xi}_t \\ &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*, \end{aligned} \tag{6.6.1}$$

where $\boldsymbol{\omega}_t^* = \boldsymbol{\omega}_t + \boldsymbol{\xi}_t$ is distributed as

$$\boldsymbol{\omega}_t^* \sim \mathrm{T}_{n_{t-1}}(\mathbf{0}, \boldsymbol{W}_t^*), \tag{6.6.2}$$

where $\boldsymbol{W}_t^* = \boldsymbol{W}_t + \boldsymbol{H}_t$, with $\boldsymbol{H}_t$ defined as in Equation (6.5.17). The observation equation takes a similar form where

$$Y_t = \boldsymbol{F}_t\boldsymbol{\theta}_t + \nu_t^*, \tag{6.6.3}$$

where $\nu_t^* = \nu_t + \psi_t$ (where ν_t and ψ_t are uncorrelated) is distributed as

$$\nu_t^* \sim \mathrm{N}(0, V_t^*), \tag{6.6.4}$$

where $V_t^* = V_t + \Psi_t$, with the estimate of Ψ_t, S_t^{ψ}, defined as in Equation (6.5.24). The second order polynomial model described in Section 5.2 is adapted to include an automatic monitoring system. The routine model, M_0 is described in detail in Section 5.2 and the alternative model M_1 (remembering that we could have many different alternative models) is designed to detect for abrupt changes in parameter values and for possible outliers. An alternative model that can account for possible outliers and parametric changes is the scale shift model defined in Section 6.5.2. A similar analysis to that in Section 6.5.2 can be performed for the case of unknown observational variance. The mathematics is not as convenient but defining

the standardised forecast distribution for $e_t = Y_t$ as

$$(e_t \mid D_{t-1}, M_0) \sim \mathrm{T}_{\delta_V n_{t-1}}(0, 1) \tag{6.6.5}$$

and so

$$p_0(e_t \mid D_{t-1}) = \frac{\Gamma(\frac{\delta_V n_{t-1}+1}{2})}{\sqrt{\delta_V n_{t-1}\pi}\Gamma(\frac{\delta_V n_{t-1}}{2})}\left(1 + \frac{e_t^2}{\delta_V n_{t-1}}\right)^{-\frac{n_t}{2}}. \tag{6.6.6}$$

Defining an alternative, M_1, with a larger scale than M_0, with forecast distribution for e_t

$$(e_t \mid D_{t-1}, M_1) \sim \mathrm{T}_{\delta_V n_{t-1}}(0, k) \tag{6.6.7}$$

and so

$$p_1(e_t \mid D_{t-1}) = \frac{\Gamma(\frac{\delta_V n_{t-1}+1}{2})}{\sqrt{\delta_V n_{t-1} k\pi}\Gamma(\frac{\delta_V n_{t-1}}{2})}\left(1 + \frac{e_t^2}{k\delta_V n_{t-1}}\right)^{-\frac{n_t}{2}}. \tag{6.6.8}$$

The Bayes' factor at time t is then

$$\sqrt{k}\left(1 + \frac{e_t^2}{\delta_V n_{t-1}}\right)^{-\frac{n_t}{2}}\left(1 + \frac{e_t^2}{k\delta_V n_{t-1}}\right)^{\frac{n_t}{2}}. \tag{6.6.9}$$

For illustration purposes, to plot Bayes' factors as a function of $|e_t|$ against k, we fix $n_t = n, \forall t$ (hence $\delta_V n_t = \delta_V n$ is fixed $\forall t$). Consider an observational discount factor $\delta_V = 0.95$ and $n_0 = 20$; we see that, $n_t = \delta_V n_{t-1} + 1 = 20, \forall t$. Figure 6.8 shows a plot of Bayes' factors as a function of $|e_t|$ against k. Once again, a key point to make is that above a certain level the particular value of k is irrelevant. The curves plotted provide Bayes' factors as a function of $|e_t|$ over the range of 2-4, for $k = 2, 3, \ldots, 8$. For the threshold $\tau = 0.135$ (the dashed horizontal line in Figure 6.8) we see that all values of k would produce an error if $|e_t| \gtrsim 3.25$. We will use a scale shift model with $k = 3$ to monitor the performance of the routine model used in Section 5.2.

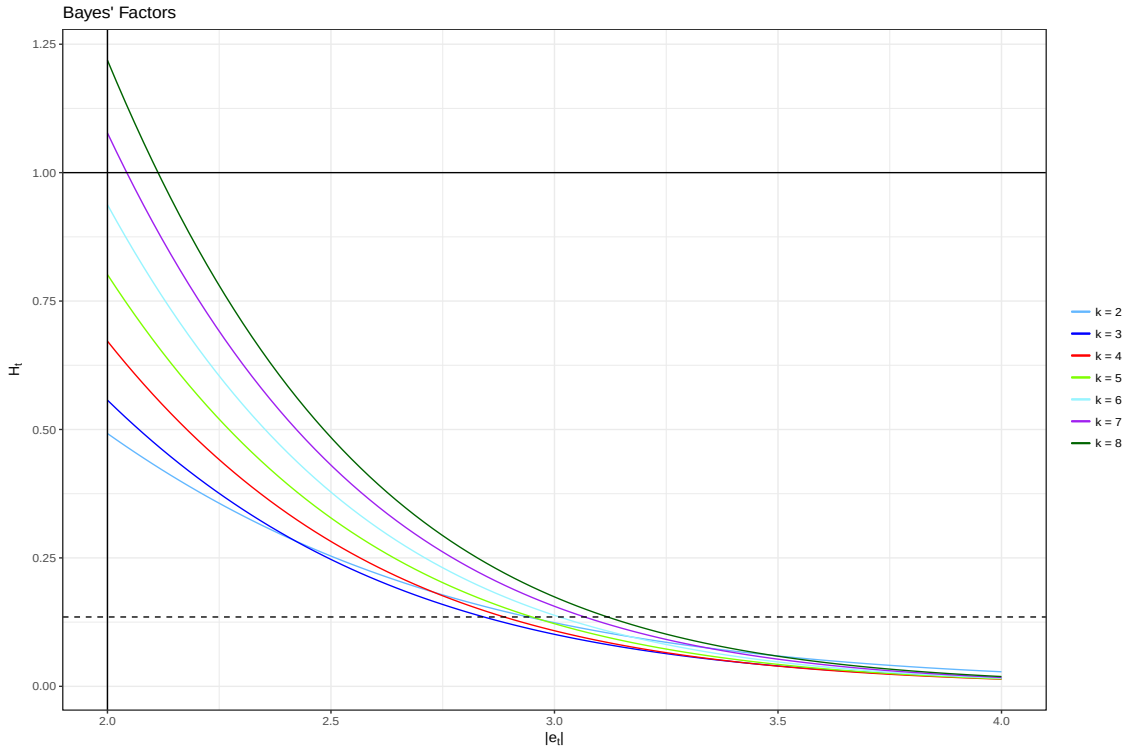

Figure 6.8: Bayes' factors H_t for T distribution scale inflation model plotted against $|e_t|$ for different scale inflations k. Black horizontal line at $H_t = 1$ represents indifference between M_0 and M_1. Black dashed line at ≈ 0.135 represents our chosen threshold to indicate poor model performance

6.7 Urine Output Series: Second-Order Polynomial Model with Monitoring

In Section 5 we used a second-order polynomial model to monitor the urine output time series in order to forecast severe oliguria. In Section 5.3 we discussed how in the early stages of CICU admissions the second-order polynomial model overestimated risk and discussed how this poor model performance was likely to be due to physiological responses to surgery or interventions attempting to normalise a patient's urine output. In this section we construct and use a model monitoring scheme to automatically correct the second-order polynomial model, used in Section 5, when poor model performance is detected. The results of forecasting severe oliguria and a discussion about its consequences are also given and results produced by Dr Samuel Howitt are clearly referenced.

6.7.1 Urine Output Series: Model Monitoring Scheme

The model used to monitor the urine output series was described in Section 5.2 and represented by Equations (5.2.1). It was also noted in Sections 5.2.1 and 5.6.1 that higher discount factors usually produce better forecast performance than lower discount factors, however, the time series that we are considering is extremely noisy which pulled component discount factors in a downward direction. The analysis shown in Appendix D suggested that the discount factors $\delta_V = 0.95$, $\delta_\mu = 0.80$, and $\delta_\beta = 0.90$, which minimise the number of type two errors, are suitable options. This kind of analysis was a compromise because any structural changes other than smooth dynamic variation pull component discount factors towards zero. Although using the suggested discount factors does allow for adaption to structural changes, it also means that during routine intervals there is excessive uncertainty during evolution resulting in imprecise forecasts. Now that the forecast system is explicitly on the lookout for structural changes, such a compromise over discounting is not necessary. In Section 6.6 we discussed an explicit monitoring system to monitor model performance which can result in not having to compromise by using lower discount factors. For convenience we remind the reader of the model described in Section 5.2.1

$$\begin{aligned} \log(Y_t + 0.1) &= \mu_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}. \end{aligned} \tag{6.7.1}$$

The state vector $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$, where μ_t allows for systematic variation about a time varying level and β_t allows for systematic growth and decline of the level, where $\nu_t \sim \mathrm{N}(0, V_t)$ and $\boldsymbol{\omega}_t \sim \mathrm{T}_{n_{t-1}}(0, \boldsymbol{W}_t)$ where

$$V_t^{-1} \mid D_t \sim \mathrm{Ga}(\delta_V n_t/2, \delta_V d_t/2), \quad \boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} + W_{\beta t} & W_{\beta t} \\ W_{\beta t} & W_{\beta t} \end{pmatrix}, \tag{6.7.2}$$

where $n_0 = 20$, $d_0 = 2$ (and hence $S_0 = 0.1$) and $\delta_V = 0.95$, also $W_{\mu t} = C_{\mu,t-1}(\delta_\mu^{-1} - 1)$ and $W_{\beta t} = C_{\beta,t-1}(\delta_\beta^{-1} - 1)$, where $\mathrm{Var}(\mu_{t-1} \mid D_{t-1}) = C_{\mu,t-1}$ and $\mathrm{Var}(\beta_{t-1} \mid D_{t-1}) = C_{\beta,t-1}$, where

$$(\boldsymbol{\theta}_0 \mid D_0) \sim \mathrm{T}_{n_0}\left[\begin{pmatrix} 0.55 \\ -0.2 \end{pmatrix}, \begin{pmatrix} 0.01 & 0 \\ 0 & 0.001 \end{pmatrix}\right], \tag{6.7.3}$$

and $\delta_\mu = 0.80$, and $\delta_\beta = 0.90$. We write

$$M_0 = \{\boldsymbol{F}, \boldsymbol{G}, V_t, \boldsymbol{W}_t\}, \tag{6.7.4}$$

where V_t and $\boldsymbol{W}_t$ are given by Equation (6.7.2) and (see Equation(4.6.3))

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{6.7.5}$$

We now use the model described above, but also include a model monitoring scheme. In addition we use a larger level discount factor $\delta_\mu = 0.9$ (chosen by using the criteria discussed in Section 5.2.2). That is, we will use the routine model, M_0, and assess forecast performance at each time point by using the detection scheme discussed in Section 6.5.1. The alternative model, which we use to assess the forecast performance of M_0 is a scale inflation model, M_1, which allows for the types of sudden changes that we expect in the urine output series (discussed in Sections 6.5.2 and 6.6). The Bayes' threshold is taken to be $\tau = 0.135$ [29], the run-length threshold is taken to be $l = 2$; this means that if the cumulative Bayes' factors are less than one, but do not break the threshold, τ, for three consecutive hours, then the model will signal a warning. The scale inflation of the alternative model is taken to be $k = 3$ (see Figure 6.8), this value is chosen as it is the first (scale inflation) alternative model to break the defined Bayes' threshold. Once the detection scheme flags a warning, indicating poor model performance, automatic diagnostics will be applied to correct the routine model. The automatic diagnostics take the form of reducing the discount factors of the unknown evolution matrix, $\boldsymbol{W}_t$, and of the unknown observational variance, V_t, as described in Section 6.5.3. The parameter estimates for the reduced discount factors (by using the criteria discussed in Section 5.2.2) are $\delta_\mu^* = \delta_\beta^* = 0.12$, and $\delta_V^* = 0.8$.

6.7.2 Model Monitoring Scheme: Results

The results of the model with monitoring are now presented, recalling that the validation group consisted of 2389 adult patients. We remind the reader that in order to allow comparison of the model's output with the existing stage one AKI classification (within the KDIGO guidelines), patients were assigned to either a high risk or a low risk group based on their first high risk warning. The analyses

tested the model's ability to identify which patients would suffer severe oliguria (UO $< 0.3ml/kg$ for 6 consecutive hours) within 12 hours of the high risk prediction.

Table 6.1 shows model performance over time, by considering observed severe oliguria (SO) compared to predicted severe oliguria.

Time Point (number of patients still on CICU)	Observed SO (% of patients)	Predicted SO (% of patients)	O:E ratio
12 hours (1947)	61 (3.1)	82 (4.2)	0.74
24 hours (1694)	57 (3.4)	61 (3.6)	0.93
36 hours (1137)	51 (4.5)	44 (3.9)	1.16
48 hours (909)	54 (5.9)	48 (5.3)	1.13
72 hours (545)	35 (6.4)	30 (5.5)	1.15

Table 6.1: Calibration of model's (with model monitoring) predictive performance, produced by Dr Samuel Howitt. Table shows DLM performance over time by calculating the proportion of the DLM's high and low risk predictions that were correct at different time points

In Section 5.3 we discussed how interventions and physiological responses are likely to have been two reasons why the model (without monitoring) overestimated risk for the first 24 hours. Comparing Table 5.1 to Table 6.1 we can see that the diagnostics performed after poor model performance is detected have improved this issue. The model is now able to retrospectively identify when physiological responses have ended, and the model can detect interventions used to normalise urine outputs, allowing the model to rapidly adapt to future trends. In addition, the model with monitoring performs better after 24 hours as well. From Tables 5.1 and 6.1 we can see that the DLM with model monitoring is calibrated better (at each time point) than the DLM with no model monitoring.

Chapter 7

Multi-Process Models

Multi-Process models offer a powerful framework for modelling and analysing time series which are subject to abrupt changes in pattern. In particular this approach can be used to provide on-line probabilities of whether changes have occurred in the series, as well as identifying the types of change that are involved. This methodology has been used to monitor the progress of kidney function in the past and is documented in West (1992) [30], Smith and West (1983) [23], Smith *et al* (1983) [31], and Trimble *et al* (1983) [32]. These documents were concerned with detecting and interpreting abrupt changes in the pattern of time series data. In this chapter we extend on the previous analysis in Chapter 6 to provide a more powerful model to capture the abrupt changes of the urine output series. Multi-process models are combinations of several DLMs, while individual DLMs are single process models. There are two classes of multi-process models, first introduced by Harrison and Stevens (1976) [33], and in this chapter we will focus on Class II multi-process models.

Recall from Definition 3.4.2.1 that a DLM is characterised by a quadruple

$$M_t : \qquad \{\boldsymbol{F}, \boldsymbol{G}, V, \boldsymbol{W}\}_t, \tag{7.0.1}$$

at each time t, conditional on initial information D_0. From now on, any defining parameters that are possibly subject to uncertainty are denoted by $\boldsymbol{\alpha}$. For example, in Section 6.5.3 we discussed a method where the discount factors were reduced to

allow the DLM in use to rapidly adapt to future trends. As a result, we could have denoted $\boldsymbol{W}_t$ by $\boldsymbol{W}_t(\boldsymbol{\alpha})$ and V_t by $V_t(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = \{\delta_\mu, \delta_\beta, \delta_V\}$, to represent uncertainty in the evolution matrix and observational variance (caused by uncertainty in the evolution and observational discount factors). We now represent the model on these uncertain quantities by writing

$$M_t = M_t(\boldsymbol{\alpha}). \tag{7.0.2}$$

For any given value of $\boldsymbol{\alpha}$, $M_t(\boldsymbol{\alpha})$ is a DLM for each time t. The uncertainty about the value of $\boldsymbol{\alpha}$ is what leads us to consider multi-process models [1]. Let $\mathcal{A}$ denote the set of possible values for $\boldsymbol{\alpha}$. If the size of $\boldsymbol{\alpha}$ is uncountably infinite, for example a discount factor is defined in the range $\delta \in (0, 1]$, then we can approximate $\mathcal{A}$ by a discrete set of values. For example, if $\boldsymbol{\alpha} = \delta$, a single discount factor, then we could approximate $\mathcal{A} \in (0, 1]$ by $\mathcal{A} = \{0.05, 0.10, 0.15, \ldots, 0.95, 1\}$.

The class of DLMs at time t is given by

$$\{M_t(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathcal{A}\}. \tag{7.0.3}$$

Then for some sequence of values $\boldsymbol{\alpha}_t \in \mathcal{A}$, $M_t(\boldsymbol{\alpha}_t)$ holds at time t. The description above reflects that there is no single DLM accepted as adequate for all times. This is usually the case when modelling noisy time series and this leads us to the next section discussing and analysing class II multi-process models.

7.1 Class II Multi-process Models

Extending on the theory described above that no single DLM is accepted as adequate for all times we define class II multi-process models.

Definition 7.1.0.1. Suppose that at each time t, $\boldsymbol{\alpha}$ takes a value in the discrete set $\mathcal{A} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k\}$. Then the Y_t series is said to follow a multi-process, class II model [1].

It remains to specify the probabilistic mechanisms by which a particular value of $\boldsymbol{\alpha}$ is chosen at each time.

Definition 7.1.0.2. In the multi-process class II framework, the value $\boldsymbol{\alpha} = \boldsymbol{\alpha}_j$ at

time t, defining the model $M_t(j)$, is selected with known probability

$$\pi_t(j) = \Pr[M_t(j) \mid D_{t-1}] = \Pr[\boldsymbol{\alpha} = \boldsymbol{\alpha}_j \text{ at time } t \mid D_{t-1}]. \tag{7.1.1}$$

The series Y_t is said to follow a multi-process, class II mixture model [1]. Here $\pi_t(j)$ could depend on the history of the process, but in this thesis we will assume fixed model selection probabilities

$$\pi_t(j) = \pi(j) = \Pr[M_t(j) \mid D_0] \tag{7.1.2}$$

for all t.

In other words, $M_t(j)$ has prior probability $\pi(j)$ at each time t no matter what the values of $Y_1, \ldots, Y_{t-1}$. Consider one of the discount factor components in the model monitoring technique discussed in Section 6.7.1. In this example $\boldsymbol{\alpha} = \delta_\mu$, and the parameter space $\mathcal{A} = \{0.9, 0.12\}$. If we define the prior probabilities of using models M_1 and M_2 to be $\pi(1) = 0.95$, and $\pi(2) = 0.05$, where M_1 and M_2 refer to the models with discount factors 0.9 and 0.12 respectively; this means that we expect sudden changes to happen around 5% of the time before observing any of the data. Using fixed model selection probabilities this means that at time t we still expect a sudden change with probability 0.05, irrespective of what has happened in the past. The probability of using model 1 at time t, $M_t(1)$ is 0.95, irrespective of D_{t-1}.

7.1.1 Fixed Selection Probability Models

For simplicity we will set $\alpha_j = j$ and refer to integer indices rather than parameters. As a result, the model index set is now $\mathcal{A} = \{1, \ldots, k\}$, and we refer to $M_t(j)$ as model j at time t [1]. We define, for each time t and integer h, $(0 \leq h < t)$, the probabilities

$$p_t(j_t, j_{t-1}, \ldots, j_{t-h}) = \Pr[M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}) \mid D_t]. \tag{7.1.3}$$

To gain a better understanding of dynamic mixture models, consider the position at time $t = 1$, assuming that $(\boldsymbol{\theta}_0 \mid D_0)$ has a normal or T distribution as usual. At time $t = 1$, there are k possible DLMs, $M_1(j_1)$, with prior probabilities $\pi(j_1)$,

$(j_1 = 1, \ldots, k)$. In DLM j_1 the state vector has posterior $p(\boldsymbol{\theta}_1 \mid M_1(j_1), D_1)$ and by Bayes' theorem, the DLM has posterior probability

$$p_1(j_1) = \Pr[M_1(j_1) \mid D_1] \propto p(Y_1 \mid M_1(j_1), D_0)\pi(j_1). \tag{7.1.4}$$

Unconditionally, inferences about the state vector are based on the discrete mixture

$$p(\boldsymbol{\theta}_1 \mid D_1) = \sum_{j_1=1}^{k} p(\boldsymbol{\theta}_1 \mid M_1(j_1), D_1)p_1(j_1), \tag{7.1.5}$$

a sum of k components. Proceeding to time $t = 2$, any of the k possible DLMs, $M_2(j_2)$, $(j_2 = 1, \ldots, k)$, may be selected, with probabilities $\pi(j_2)$. It is only possible to retain the components of standard DLM analyses if the models possible at time $t = 1$ are also considered. Hence, conditional on both $M_2(j_2)$ and $M_1(j_1)$ for some j_2 and j_1, the posterior for $\boldsymbol{\theta}_2$ given D_2, depending on j_2 and j_1 is denoted by $p(\boldsymbol{\theta}_2 \mid M_2(j_2), M_1(j_1), D_2)$. This mixture can be written as

$$p(\boldsymbol{\theta}_2 \mid M_2(j_2), D_2) = \sum_{j_1=1}^{k} p(\boldsymbol{\theta}_2, M_2(j_2), M_1(j_1), D_2) \Pr[M_1(j_1) \mid D_2]. \tag{7.1.6}$$

Thus, conditional on j_2 at time $t = 2$, the posterior is a mixture of k standard forms, depending on which model was obtained at time $t = 1$. Thus, unconditionally,

$$\begin{aligned} p(\boldsymbol{\theta}_2 \mid D_2) &= \sum_{j_2=1}^{k} p(\boldsymbol{\theta}_2 \mid M_2(j_2), D_2)p_2(j_2) \\ &= \sum_{j_2=1}^{k} \sum_{j_1=1}^{k} p(\boldsymbol{\theta}_2, M_2(j_2), M_1(j_1), D_2)p_2(j_2, j_1). \end{aligned} \tag{7.1.7}$$

Unconditionally the posterior for $\boldsymbol{\theta}_2$ depends on k^2 components. This continues as time goes by. Then at time t, the posterior density can be written as

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &= \sum_{j_t=1}^{k} p(\boldsymbol{\theta}_t \mid M_t(j_t), D_t)p_t(j_t) \\ &= \sum_{j_t=1}^{k} \sum_{j_{t-1}=1}^{k} p(\boldsymbol{\theta}_t, M_t(j_t), M_{t-1}(j_{t-1}), D_t)p_t(j_t, j_{t-1}). \end{aligned} \tag{7.1.8}$$

This can be written as

$$p(\boldsymbol{\theta}_t \mid D_t) = \sum_{j_t=1}^{k} \sum_{j_{t-1}=1}^{k} \cdots \sum_{j_1=1}^{k} p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_1(j_1), D_t) \times p_t(j_t, j_{t-1}, \ldots, j_1). \tag{7.1.9}$$

Thus to obtain the marginal posterior as a mixture, we need to consider all k^t possible combinations that may apply [1].

A problem arises with the above analysis. As t increases the number of possible combinations that need to be considered tends to infinity. After t observations there are k^t possible combinations of models that may apply. Schervish and Tsay (1988) performed an analysis with $k = 4$ and t up to approximately 200. Practically, however, the large amount of combinations required in some mixtures will be computationally unfeasible (especially when forecasting multiple steps ahead (see Section 7.3.1)). In the following sections we will discuss how to reduce the number of components by approximations, leading to more manageable mixtures.

7.1.2 Approximation of Mixtures

In this thesis we have emphasised many times that as time progresses, what has occurred in the past becomes less and less relevant for inferences made for the future. This key property of dynamic models is essential, and we will now apply this to mixtures. The possible models obtaining in the past losing relevance to inferences made at the current time t as t increases. In other words, the full conditional posterior $p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_1(j_1), D_t)$ will depend negligibly on early models $M_1(j_1)$, $M_2(j_2)$, etc., when t is large [1]. That is, for some $h \geq 1$, the full conditional posterior can be approximated by

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid & M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_1(j_1), D_t) \\ & \approx p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), D_t). \end{aligned} \tag{7.1.10}$$

This approximation means that the number of components in the mixture posterior at any time will not exceed k^{h+1}. Unconditionally, the full mixture (7.1.9) will be

approximated as

$$p(\boldsymbol{\theta}_t \mid D_t) = \sum_{j_t=1}^{k} \sum_{j_{t-1}=1}^{k} \cdots \sum_{j_{t-h}=1}^{k} p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), D_t) \times p_t(j_t, j_{t-1}, \ldots, j_{t-h}). \tag{7.1.11}$$

It now remains to calculate the posterior model probabilities weighting the posteriors in this mixture. Using Bayes' theorem,

$$p_t(j_t, j_{t-1}, \ldots, j_{t-h}) \propto \Pr[M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}) \mid D_{t-1}] \times p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), D_{t-1}). \tag{7.1.12}$$

The second term in Equation (7.1.12) is given by

$$p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), D_{t-1}) = \sum_{j_{t-h-1}=1}^{k} p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), M_{t-h-1}(j_{t-h-1}), D_{t-1}) \times \Pr[M_{t-h-1}(j_{t-h-1}) \mid D_{t-1}], \tag{7.1.13}$$

an average, with respect to models $h+1$ steps back, of the normal or T one-step predictive densities for Y_t under each of the models in the conditionings. The probabilities weighting these terms are obtained from

$$\Pr[M_{t-h-1}(j_{t-h-1}) \mid D_{t-1}] = \sum_{j_{t-1}=1}^{k} \cdots \sum_{j_{t-h}=1}^{k} p_{t-1}(j_{t-1}, \ldots, j_{t-h}). \tag{7.1.14}$$

The first term in (7.1.12) is similarly calculated by

$$\begin{aligned} &\Pr[M_t(j_t), M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}) \mid D_{t-1}] \\ &= \Pr[M_t(j_t) \mid M_{t-1}(j_{t-1}), \ldots, M_{t-h}(j_{t-h}), D_{t-1}] p_{t-1}(j_{t-1}, \ldots, j_{t-h}) \\ &= \pi(j_t) p_{t-1}(j_{t-1}, \ldots, j_{t-h}) \\ &= \pi(j_t) \sum_{j_{t-h-1}}^{k} p_{t-1}(j_{t-1}, \ldots, j_{t-h-1}). \end{aligned} \tag{7.1.15}$$

Further approximations to the mixture can also be made to reduce the number of components in (7.1.11). Three considerations are as follows [1]:

1. Ignore components that have very small posterior probabilities.

2. Combine components that are roughly equal into a single component, also combining probabilities.

3. Replace the contribution of a collection of components by a component that represents their contribution.

7.1.3 Kullback-Leibler Directed Divergence

In the previous section we discussed mixture approximating, or collapsing, in order to approximate the full posterior by only using models up to h-steps back. In this section we derive collapsing techniques and methods that are fundamental to the application of multi-process, class II models. To begin, assume that the density of the random vector $\boldsymbol{\theta}$ is the mixture

$$p(\boldsymbol{\theta}) = \sum_{j=1}^{k} p_j(\boldsymbol{\theta}) p(j). \tag{7.1.16}$$

Here the component densities may generally take any forms, although often they will have the same functional form, such as normal or T, differing only through defining parameters such as means, variances, etc. The probabilities $p(j)$ are known. Our aim is to approximate (7.1.16) by a density $p^*(\boldsymbol{\theta})$ and hence collapse the sum of k densities to one (generally the approximation can be used to reduce the summation to a number smaller than k but our aim is to completely collapse the summation to just one component). For example, a mixture of T densities may be approximated by a T density with location and scale to be chosen in some optimal way. Clearly $p^*(.)$ should be close to $p(.)$ and this brings in the need for a distance measure between densities to measure how close one density is to another.

Viewing $p(.)$ as the true density of $\boldsymbol{\theta}$ to be approximated by $p^*(.)$, consider the quantity

$$-\mathrm{E}\{\log[p^*(\boldsymbol{\theta})]\} = -\int \log[p^*(\boldsymbol{\theta})] p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{7.1.17}$$

For any approximating distribution with density $p^*(.)$, this entropy related quantity is a measure of the closeness of approximation to the true distribution $p(.)$ [1].

Choosing the approximating density to achieve a small value of (7.1.17) is equivalent to minimising the quantity $K(p^*)$ defined by

$$K(p^*) = \mathrm{E}\{\log\left[\frac{p(\boldsymbol{\theta})}{p^*(\boldsymbol{\theta})}\right]\} = \int \log\left[\frac{p(\boldsymbol{\theta})}{p^*(\boldsymbol{\theta})}\right] p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{7.1.18}$$

This is the Kullback-Leibler directed divergence between the approximating distribution whose density is $p^*(.)$ and the true distribution with density $p(.)$ [1]. Here we focus on continuous distributions and assume that $p(.)$ and $p^*(.)$ have the same support. As a result, $p(\boldsymbol{\theta}) > 0$ if and only if $p^*(\boldsymbol{\theta}) > 0$ and two key properties are that

- $K(p^*) \geq 0$ for all densities $p^*(.)$, and
- $K(p^*) = 0$ if and only if $p^* = p$.

We will use the Kullback-Leibler divergence to measure closeness of approximations to mixtures. Recall that if V_t is known the state and observation distributions are normal distributions and if V_t is unknown the state and observation distributions are T distributions. Since we are focusing on continuous distributions and assuming that $p(.)$ and $p^*(.)$ have the same support we will only need to consider measuring closeness of approximations to normal or T mixtures.

Example 7.1.3.1. From West and Harrison [1]. Suppose that $\boldsymbol{\theta}$ follows a multivariate normal distribution with a mean vector, $\mathrm{E}[\boldsymbol{\theta}]$, and variance matrix, $\mathrm{Var}(\boldsymbol{\theta})$, and that the approximating distribution is multivariate normal, $\mathrm{N}(\boldsymbol{m}, \boldsymbol{C})$. Then as a function of $\boldsymbol{m}$ and $\boldsymbol{C}$, the Kullback-Leibler divergence is given by (see Appendix E)

$$\begin{aligned} 2K(p^*) &= 2\{\mathrm{E}[\log\{p(\boldsymbol{\theta})\}] - \mathrm{E}[\log\{p^*(\boldsymbol{\theta})\}]\} \\ &= c + \log(|\boldsymbol{C}|) + \mathrm{trace}(\boldsymbol{C}^{-1}\mathrm{Var}(\boldsymbol{\theta})) + (\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m}), \end{aligned} \tag{7.1.19}$$

for some constant c that does not depend on $\boldsymbol{m}$ or $\boldsymbol{C}$. Noting $\mathrm{E}[\log\{p(\theta)\}]$ does not depend on $\boldsymbol{m}$ or $\boldsymbol{C}$.

Example 7.1.3.2. From West and Harrison [1]. Suppose that the q-vector $\boldsymbol{\theta}$ and

the scalar ϕ have a joint distribution that is a mixture of normal/gamma forms. The mixture has k components, the jth component being defined as follows:

1. Given ϕ, $(\boldsymbol{\theta} \mid \phi) \sim \mathrm{N}(\boldsymbol{m}(j), \boldsymbol{C}(j)/[S(j)\phi])$ for some mean vector $\boldsymbol{m}(j)$, variance matrix $\boldsymbol{C}(j)$, and estimate $S(j) > 0$ of ϕ^{-1}.

2. Marginally, $\phi \sim \mathrm{Ga}(n(j)/2, d(j)/2)$ for $n(j), d(j) > 0$, and $\mathrm{E}[\phi] = S(j)^{-1} = n(j)/d(j)$.

3. Integrating out ϕ gives the marginal, multivariate T distribution in model j, $\boldsymbol{\theta} \sim \mathrm{T}_{n(j)}(\boldsymbol{m}(j), \boldsymbol{C}(j))$.

Suppose that the mixture is to be approximated by $p^*(\boldsymbol{\theta}, \phi)$, a single, normal gamma distribution defined by parameters $\boldsymbol{m}$, $\boldsymbol{C}$, n, and d, giving $(\boldsymbol{\theta} \mid \phi) \sim \mathrm{N}(\boldsymbol{m}, \boldsymbol{C}/[S\phi])$, where $S = d/n$, $\phi \sim \mathrm{Ga}(n/2, d/2)$ and $\boldsymbol{\theta} \sim \mathrm{T}_n(\boldsymbol{m}, \boldsymbol{C})$. The Kullback-Leibler divergence with respect to the joint mixture distribution of $\boldsymbol{\theta}$ and ϕ can be written as (using Equation (7.1.18) and Bayes' theorem)

$$\begin{aligned} K(p^*) &= \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p(\boldsymbol{\theta}, \boldsymbol{\phi})]\} - \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta}, \boldsymbol{\phi})]\} \\ &= \textit{constant} - \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta} \mid \phi)] + \log[p^*(\phi)]\}, \end{aligned} \tag{7.1.20}$$

since $\mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p(\boldsymbol{\theta}, \boldsymbol{\phi})]\}$ does not depend on $\boldsymbol{m}$, $\boldsymbol{C}$, n, or d. Using the densities defined in 1. and 2. above results in (see Appendix F)

$$\begin{aligned} 2K(p^*) &= \textit{constant} - 2\mathrm{E}_\phi\{\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}\} - 2\mathrm{E}_\phi\{\log[p(\phi)]\} \\ &= \textit{constant} \\ &\quad - n\log(d/2) + 2\log(\Gamma(n/2)) - (n + q - 2)\mathrm{E}_\phi[\log(\phi)] + d\mathrm{E}[\phi] + \log(|S^{-1}\boldsymbol{C}|) \\ &\quad + S\sum_{j=1}^{k} S^{-1}(j)\{\mathrm{trace}(\boldsymbol{C}\boldsymbol{C}(j)) + (\boldsymbol{m}(j) - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{m}(j) - \boldsymbol{m})\}p(j). \end{aligned} \tag{7.1.21}$$

Minimising with respect to $\boldsymbol{m}$, $\boldsymbol{C}$, d and n leads to the following:

(a) Minimising with respect to $\boldsymbol{m}$ we find that

$$\boldsymbol{m} = \left\{ \sum_{j=1}^{k} S(j)^{-1} p(j) \right\}^{-1} \sum_{j=1}^{k} S(j)^{-1} \boldsymbol{m}(j) p(j). \tag{7.1.22}$$

(b) Minimising with respect to $\boldsymbol{C}$ leads to,

$$\boldsymbol{C} = S \sum_{j=1}^{k} S(j)^{-1} \{ \boldsymbol{C}(j) + (\boldsymbol{m} - \boldsymbol{m}(j))(\boldsymbol{m} - \boldsymbol{m}(j))' \} p(j). \tag{7.1.23}$$

(c) Minimising with respect to d leads to

$$S^{-1} = \mathrm{E}[\phi] = \sum_{j=1}^{k} S(j)^{-1} p(j). \tag{7.1.24}$$

We can then rewrite (a) and (b) as follows,

$$\begin{aligned} \boldsymbol{m} &= \sum_{j=1}^{k} \boldsymbol{m}(j) p^*(j), \\ \boldsymbol{C} &= \sum_{j=1}^{k} \{ \boldsymbol{C}(j) + (\boldsymbol{m} - \boldsymbol{m}(j))(\boldsymbol{m} - \boldsymbol{m}(j))' \} p^*(j), \end{aligned} \tag{7.1.25}$$

where the weights $p^*(j) = p(j)S/S(j)$ sum to unity.

(d) Minimising with respect to n leads to

$$\mathrm{E}_{\phi}[\log(\phi)] = \gamma(n/2) - \log(d/2). \tag{7.1.26}$$

7.2 Second-Order Polynomial Models with Exceptions

In this section we extend on the analysis discussed in Section 6.6 and build a more powerful framework for modelling with exceptions. The second order polynomial model discussed in Section 5.2.1 will be used unless exceptional events, namely outliers and changes in level and growth, occur.

To begin recall the second order polynomial model for the urine output series (see

Section 5.2.1)

$$\begin{aligned} \log(Y_t + 0.1) &= \mu_t + \nu_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, \end{aligned} \tag{7.2.1}$$

where the state vector $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$. Here μ_t allows for systematic variation about a time varying level and β_t allows for systematic growth and decline of the level, where $\nu_t \sim \mathrm{N}(0, V_t)$ and $\boldsymbol{\omega}_t \sim \mathrm{T}_{n_{t-1}}(0, \boldsymbol{W}_t)$ where

$$V_t^{-1} \mid D_t \sim \mathrm{Ga}(\delta_V n_t/2, \delta_V d_t/2), \quad \boldsymbol{W}_t = \begin{pmatrix} W_{\mu t} + W_{\beta t} & W_{\beta t} \\ W_{\beta t} & W_{\beta t} \end{pmatrix}, \tag{7.2.2}$$

where $W_{\mu t} = C_{\mu,t-1}(\delta_\mu^{-1} - 1)$ and $W_{\beta t} = C_{\beta,t-1}(\delta_\beta^{-1} - 1)$, where $\mathrm{Var}(\mu_{t-1} \mid D_{t-1}) = C_{\mu,t-1}$ and $\mathrm{Var}(\beta_{t-1} \mid D_{t-1}) = C_{\beta,t-1}$. In addition (see Section 4.6.1),

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{7.2.3}$$

This model can be described by the quadruple

$$\{\boldsymbol{F}, \boldsymbol{G}, V_t, \boldsymbol{W}_t\}. \tag{7.2.4}$$

Consider modelling outliers and changes in parameter values in a series thought to behave generally according to the model described above. An outlying observation, Y_t, may be modelled via a large observational error ν_t. A single, extreme value of ν_t is an exception in the model described above but perfectly consistent with a model with a large enough observational variance [1]. If outliers are expected to occur some percentage of the time, then the alternative DLM

$$\{\boldsymbol{F}, \boldsymbol{G}, V_t V(2), \boldsymbol{W}_t\}, \tag{7.2.5}$$

where $V(2) > 1$, will adequately model them, whilst (7.2.4) applies during routine periods.

In a similar fashion, changes in level μ_t much greater than predicted by (7.2.4) can be allowed for by replacing $\boldsymbol{W}_t$ with an alternative in which the variance term $W_{\mu t}$ is inflated. The alternative DLM

$$\{\boldsymbol{F}, \boldsymbol{G}, V_t, \boldsymbol{W}_t(3)\}, \tag{7.2.6}$$

where $\boldsymbol{W}_t(3)$ has an inflated level variance compared to $\boldsymbol{W}_t$ will adequately model them. This DLM will be used for the times of a level change. With this DLM $W_{\mu t}$ can be far greater in absolute value than in (7.2.4), allowing for significant changes in level.

Finally, to model jumps in growth of the series the alternative DLM

$$\{\boldsymbol{F}, \boldsymbol{G}, V_t, \boldsymbol{W}_t(4)\}, \tag{7.2.7}$$

where $\boldsymbol{W}_t(4)$ has an inflated growth variance compared to $\boldsymbol{W}_t$, will be used.

The above DLMs each apply at any given time with probability of applying given by fixed model probabilities (see Section 7.1.1). In this thesis we specify that

- the routine DLM (7.2.4) applies with probability 117/120;
- the outlier DLM (7.2.5) applies with probability 1/120;
- the level change DLM (7.2.6) applies with probability 1/120;
- the growth change DLM (7.2.7) applies with probability 1/120.

Meaning that the series is expected to use the routine DLM about 97.5% of the time. The chance of an outlier at any time is 0.83%. Level shifts are expected to occur 0.83% of the time and growth changes are expected to occur 0.833% of the time. Note that the values chosen are not estimated in any way, and are only used for illustrative purposes later in this section.

Here we have models $M_t(j)$ indexed by $\alpha_j = j$, $(j = 1, \ldots, 4)$.

7.2.1 Model Analysis

Recall that, in realtime, outliers are indistinguishable from the onset of a parametric change until more observations are available (see Figure 6.2). In this section we show how subsequent observations can be used to distinguish between the different types of change shown in Figure 6.2.

In this thesis we approximate mixtures by collapsing over models in the past by taking $h = 1$ (see Section 7.1.2). This approach has been used to approximate mixtures by Smith and West (1983). We have models

$$M_t(j_t) : \{\boldsymbol{F}, \boldsymbol{G}, V_t(j_t), W_t(j_t)\}, \quad (j_t = 1, \ldots, 4). \tag{7.2.8}$$

For each j_t, it is assumed that $M_t(j_t)$ applies at time t with fixed and pre-specified probability $\pi(j_t) = \Pr[M_t(j_t) \mid D_{t-1}] = \Pr[M_t(j_t) \mid D_0]$. Therefore at time t, the model is defined by observation and evolution equations

$$(Y_t \mid \boldsymbol{\theta}_t, M_t(j_t)) \sim \mathrm{T}_{\delta_V n_{t-1}}(\boldsymbol{F}'\boldsymbol{\theta}_t, V_t(j_t)) \tag{7.2.9}$$

and

$$(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, M_t(j_t)) \sim \mathrm{T}_{\delta_V n_{t-1}}(\boldsymbol{G}'\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_t(j_t)) \tag{7.2.10}$$

with probability $\pi(j_t)$. Also, at time $t = 0$, the initial prior for the state vector and scale factor are given by

$$\begin{aligned} (\boldsymbol{\theta}_0 \mid D_0) &\sim \mathrm{T}_{n_0}(\boldsymbol{m}_0, \boldsymbol{C}_0), \\ (\phi_0 \mid D_0) &\sim \mathrm{Ga}(n_0/2, d_0/2), \end{aligned} \tag{7.2.11}$$

where $\boldsymbol{m}_0$, $\boldsymbol{C}_0$, n_0, and d_0 are known and fixed at time $t = 0$ irrespective of possible models obtaining at any time. The initial point estimate S_0 of V is given by d_0/n_0. The algorithms and position at time $t-1$ are now summarised and follow from West and Harrison [1].

Historical information, D_{t-1}, is summarised in terms of a 4-component mixture posterior distribution for $\boldsymbol{\theta}_{t-1}$. Within each component, the usual conjugate normal/gamma analysis applies

1. For $j_{t-1} = 1, \ldots, 4$, model $M_{t-1}(j_{t-1})$ applied at time $t-1$ with posterior probability $p_{t-1}(j_{t-1})$. These probabilities are now known and fixed and $p_0(j_0) = \pi(j_0)$.

2. Given $M_{t-1}(j_{t-1})$ and D_{t-1}, $\boldsymbol{\theta}_{t-1}$ and ϕ_{t-1} are

$$\begin{aligned} (\boldsymbol{\theta}_{t-1} \mid M_{t-1}(j_{t-1}), D_{t-1}) &\sim \mathrm{T}_{n_{t-1}}(\boldsymbol{m}_{t-1}(j_{t-1}), \boldsymbol{C}_{t-1}(j_{t-1})), \\ (\phi_{t-1} \mid M_{t-1}(j_{t-1}), D_{t-1}) &\sim \mathrm{Ga}(n_{t-1}/2, d_{t-1}(j_{t-1})/2), \end{aligned} \tag{7.2.12}$$

where $S_{t-1}(j_{t-1}) = d_{t-1}(j_{t-1})/n_{t-1}$ is the estimate of $V_{t-1} = \phi_{t-1}^{-1}$ in model $M_{t-1}(j_{t-1})$. All quantities, except for the degrees on freedom parameter, n_{t-1}, depend on the model applying at time $t-1$. The degrees of freedom parameter, n_{t-1}, is common to each of the four possible models.

Evolving to time t, statements about $\boldsymbol{\theta}_t$, Y_t and ϕ_t depend on the combinations of possible models applying at both $t-1$ and t.

3. For each j_{t-1} and j_t we have

$$\begin{aligned}(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}) &\sim \mathrm{T}_{\delta_V n_{t-1}}(\boldsymbol{a}_t(j_{t-1}), \boldsymbol{R}_t(j_t, j_{t-1})),\\ (\phi_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}) &\sim \mathrm{Ga}(\delta_V n_{t-1}/2, \delta_V d_{t-1}(j_{t-1})/2),\end{aligned} \tag{7.2.13}$$

where $\boldsymbol{a}_t(j_{t-1}) = \boldsymbol{G}\boldsymbol{m}_{t-1}(j_{t-1})$, $\boldsymbol{R}_t(j_t, j_{t-1}) = \boldsymbol{G}\boldsymbol{C}_{t-1}(j_{t-1})\boldsymbol{G}' + \boldsymbol{W}_t(j_t)$, and δ_V allows for observational variance learning (see Section 4.9.4). Note that $\boldsymbol{a}_t(j_{t-1})$ does not differ across the $M_t(j_t)$ since $\boldsymbol{G}$ is common to these models.

4. The one-step ahead forecast distribution is given, for each possible combination of models, by

$$(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}) \sim \mathrm{T}_{\delta_V n_{t-1}}(f_t(j_{t-1}), Q_t(j_t, j_{t-1})), \tag{7.2.14}$$

where $f_t(j_{t-1}) = \boldsymbol{F}'\boldsymbol{a}_t(j_{t-1})$, and $Q_t(j_t, j_{t-1}) = \boldsymbol{F}'\boldsymbol{R}_t(j_t, j_{t-1})\boldsymbol{F} + S_{t-1}(j_{t-1})V(j_t)$. Noting that the means of the sixteen forecast distributions depend only on the models applying at time $t-1$ and hence take only four distinct values.

The forecast distribution unconditional on possible models at time $t-1$ and t involves the mixing of these standard T components with respect to their relevant probabilities. These probabilities are calculated as follows:

5. For each j_t and j_{t-1},

$$\begin{aligned}\Pr[M_t(j_t), M_{t-1}(j_{t-1}) \mid D_{t-1}] &= \Pr[M_t(j_t) \mid M_{t-1}(j_{t-1}), D_{t-1}]\\ &\quad \times \Pr[M_{t-1}(j_{t-1}) \mid D_{t-1}]\\ &= \pi(j_t)p_{t-1}(j_{t-1}),\end{aligned} \tag{7.2.15}$$

where we use the assumption that models apply at time t with constant probabilities irrespective of the past, so that $\Pr[M_t(j_t) \mid M_{t-1}(j_{t-1}), D_{t-1}] = \Pr[M_t(j_t) \mid D_0] = \pi(j_t)$.

6. The marginal predictive density for Y_t is a mixture of the 16 components in (7.2.14) with respect to the probabilities (7.2.15),

$$p(Y_t \mid D_{t-1}) = \sum_{j_t=1}^{4} \sum_{j_{t-1}=1}^{4} \pi(j_t) p_{t-1}(j_{t-1}) p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}). \tag{7.2.16}$$

Once the observation y_t is made, the prior distributions in 3. can be updated to posterior distributions.

7. The posterior distributions for the states and the scale are given by

$$\begin{aligned} (\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{T}_{n_t}(\boldsymbol{m}_t(j_t, j_{t-1}), \boldsymbol{C}_t(j_t, j_{t-1})), \\ (\phi_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{Ga}(n_t/2, d_t(j_t, j_{t-1})/2), \end{aligned} \tag{7.2.17}$$

where

$$\begin{aligned} \boldsymbol{m}_t(j_t, j_{t-1}) &= \boldsymbol{a}_t(j_{t-1}) + \boldsymbol{A}_t(j_t, j_{t-1}) e_t(j_{t-1}), \\ \boldsymbol{C}_t(j_t, j_{t-1}) &= [S_t(j_t, j_{t-1})/S_{t-1}(j_{t-1})] \\ &\quad \times [\boldsymbol{R}_t(j_t, j_{t-1}) - \boldsymbol{A}_t(j_t, j_{t-1}) \boldsymbol{A}_t'(j_t, j_{t-1}) Q_t(j_t, j_{t-1})], \\ e_t(j_{t-1}) &= y_t - f_t(j_{t-1}) \\ \boldsymbol{A}_t(j_t, j_{t-1}) &= \boldsymbol{R}_t(j_t, j_{t-1}) \boldsymbol{F} / Q_t(j_t, j_{t-1}) \\ n_t &= \delta_V n_{t-1} + 1 \\ d_t(j_t, j_{t-1}) &= \delta_V d_{t-1}(j_{t-1}) + \frac{S_{t-1}(j_{t-1}) e_t^2(j_{t-1})}{Q_t(j_t, j_{t-1})} \\ S_t(j_t, j_{t-1}) &= d_t(j_t, j_{t-1})/n_t. \end{aligned} \tag{7.2.18}$$

8. The joint posterior probabilities across the sixteen possible models are given

by

$$\begin{aligned} p_t(j_t, j_{t-1}) &= \Pr[M_t(j_t), M_{t-1}(j_{t-1}) \mid D_t] \\ &= c_t \Pr[M_t(j_t), M_{t-1}(j_{t-1}) \mid D_{t-1}] p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}) \\ &= c_t \pi(j_t) p_{t-1}(j_{t-1}) p(Y_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_{t-1}) \\ &= \frac{c_t \pi(j_t) p_{t-1}(j_{t-1})}{Q_t^{1/2}(j_t, j_{t-1})[\delta_V n_{t-1} + e_t^2(j_{t-1})/Q_t(j_t, j_{t-1})]^{n_t/2}}, \end{aligned} \tag{7.2.19}$$

where c_t is the constant of normalisation such that

$$\sum_{j_t=1}^{4} \sum_{j_{t-1}=1}^{4} p_t(j_t, j_{t-1}) = 1. \tag{7.2.20}$$

Inferences about $\boldsymbol{\theta}_t$ are based on the unconditional, sixteen component mixtures that average (7.2.17) with respect to the joint posterior model probabilities (7.2.19).

9. Thus

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &= \sum_{j_t=1}^{4} \sum_{j_{t-1}=1}^{4} p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) p_t(j_t, j_{t-1}), \\ p(\phi_t \mid D_t) &= \sum_{j_t=1}^{4} \sum_{j_{t-1}=1}^{4} p(\phi_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) p_t(j_t, j_{t-1}). \end{aligned} \tag{7.2.21}$$

This completes the evolution and updating up to time t. However, to proceed to time $t+1$, we need to remove the dependence of the joint posterior (7.2.21) on models at time $t-1$. If we do not remove this dependence and evolve (7.2.21) to time $t+1$ directly, the mixture will depend on $4^3 = 64$ components for $\boldsymbol{\theta}_{t+1}$. However, at the beginning of this section, we assumed that the effects of different models at time $t-1$ are negligible for time $t+1$. Hence, when evolving to time $t+1$, we need to collapse the sixteen component mixture (7.2.21) over possible models at time $t-1$.

For each $j_t = 1, \ldots, 4$, it follows that

$$\begin{aligned} p_t(j_t) = \Pr[M_t(j_t) \mid D_t] &= \sum_{j_{t-1}}^{4} p_t(j_t, j_{t-1}), \\ \Pr[M_{t-1}(j_{t-1}) \mid D_t] &= \sum_{j_t=1}^{4} p_t(j_t, j_{t-1}), \\ \Pr[M_{t-1}(j_{t-1}) \mid M_t(j_t), D_t] &= p_t(j_t, j_{t-1})/p_t(j_t). \end{aligned} \tag{7.2.22}$$

The first equation here gives the current model probabilities at time t. The second gives posterior probabilities over the possible models one-step back in time. These one-step back posterior probabilities are extremely useful for retrospective assessment to distinguish between outliers and parametric changes. The third equation, which we shall use to collapse the mixture (7.2.21) with respect to time $t-1$, gives the posterior probabilities of the models at time $t-1$ conditional on possible models at time t. Note that (7.2.21) can be written as

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &= \sum_{j_t=1}^{4} p(\boldsymbol{\theta}_t \mid M_t(j_t), D_t) p_t(j_t), \\ p(\phi_t \mid D_t) &= \sum_{j_t=1}^{4} p(\phi_t \mid M_t(j_t), D_t) p_t(j_t), \end{aligned} \tag{7.2.23}$$

where

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid M_t(j_t), D_t) &= \sum_{j_{t-1}=1}^{4} p(\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) p_t(j_t, j_{t-1})/p_t(j_t), \\ p(\phi_t \mid M_t(j_t), D_t) &= \sum_{j_{t-1}=1}^{4} p(\phi_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) p_t(j_t, j_{t-1})/p_t(j_t). \end{aligned} \tag{7.2.24}$$

In moving to time $t+1$ the posteriors (7.2.24) will have the required form (7.2.12) if each of the components in (7.2.24) is replaced by a single T distribution.

This is a very similar scenario to Example 7.1.3.2. We have

$$\begin{aligned} (\boldsymbol{\theta}_t \mid \phi_t, M_t(j_t), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{N}(\boldsymbol{m}_t(j_t, j_{t-1}), \boldsymbol{C}_t(j_t, j_{t-1})/[S_t(j_t, j_{t-1})\phi_t]) \\ (\boldsymbol{\theta}_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{T}_{n_t}(\boldsymbol{m}_t(j_t, j_{t-1}), \boldsymbol{C}_t(j_t, j_{t-1})), \\ (\phi_t \mid M_t(j_t), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{Ga}(n_t/2, d_t(j_t, j_{t-1})/2). \end{aligned} \tag{7.2.25}$$

In addition,

$$p(\boldsymbol{\theta}_t, \phi_t \mid M_t(j_t), D_t) = \sum_{j_{t-1}=1}^{4} p(\boldsymbol{\theta}_t, \phi_t \mid, M_t(j_t), M_{t-1}(j_{t-1}), D_t) p_t(j_t, j_{t-1}) p_t(j_t). \tag{7.2.26}$$

In other words, for each j_t, we have a 2-vector $\boldsymbol{\theta}_t$ and a scalar ϕ_t with a joint distribution that is a mixture of normal/gamma forms. Each mixture has $k = 4$ components and the j_{t-1}^{th} component is defined by Equations (7.2.25).

Consider each value of j_t in turn. If we consider $j_t = 1$, we have

$$\begin{aligned}
(\boldsymbol{\theta}_t \mid \phi_t, M_t(1), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{N}(\boldsymbol{m}_t(1, j_{t-1}), \boldsymbol{C}_t(1, j_{t-1})/[S_t(1, j_{t-1})\phi_t]) \\
(\boldsymbol{\theta}_t \mid M_t(1), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{T}_{n_t}(\boldsymbol{m}_t(1, j_{t-1}), \boldsymbol{C}_t(1, j_{t-1})), \\
(\phi_t \mid M_t(1), M_{t-1}(j_{t-1}), D_t) &\sim \mathrm{Ga}(n_t/2, d_t(1, j_{t-1})/2),
\end{aligned} \tag{7.2.27}$$

and

$$p(\boldsymbol{\theta}_t, \phi_t \mid M_t(1), D_t) = \sum_{j_{t-1}=1}^{4} p(\boldsymbol{\theta}_t, \phi_t \mid, M_t(1), M_{t-1}(j_{t-1}), D_t) p_t(1, j_{t-1}) p_t(1). \tag{7.2.28}$$

In other words, we have a 2-vector $\boldsymbol{\theta}_t$ and a scalar ϕ_t with a joint distribution that is a mixture of normal/gamma forms. The mixture has $k = 4$ components and the j_{t-1}^{th} component is defined by Equations (7.2.27). It is now evident that this is the same form as Example 7.1.3.2 (but with n_t fixed across models applying at time t). Note that in Example 7.1.3.2 the probability of model j applying was written as $p(j)$. In Equations (7.2.27), the probability of model j_{t-1} applying at time t given $M_t(1)$ is $p_t(1, j_{t-1})p_t(1)$ (see Equations (7.2.22)). The same arguments hold for $j_t = 2, 3, 4$. We are essentially applying Example 7.1.3.2 four times, one for each j_t. To this end, we show how to collapse over models applying at time $t - 1$.

Using part (c) of Example 7.1.3.2, for each j_t define the variance estimates $S_t(j_t)$ by

$$S_t(j_t)^{-1} = \sum_{j_{t-1}=1}^{4} S_t(j_t, j_{t-1})^{-1} p_t(j_t, j_{t-1})/p_t(j_t), \tag{7.2.29}$$

with

$$d_t(j_t) = n_t S_t(j_t). \tag{7.2.30}$$

Define the weights

$$p_t^*(j_t, j_{t-1}) = S_t(j_t) S_t(j_t, j_{t-1})^{-1} p_t(j_t, j_{t-1})/p_t(j_t), \tag{7.2.31}$$

noting that $\sum_{j_{t-1}=1}^{4} p_t^*(j_t, j_{t-1}) = 1$. Furthermore, using (a) and (b), in Example 7.1.3.2 the mean vectors $\boldsymbol{m}_t(j_t)$ and the variance matrices $\boldsymbol{C}_t(j_t)$ are given by

$$\begin{aligned}
\boldsymbol{m}_t(j_t) &= \sum_{j_{t-1}=1}^{4} \boldsymbol{m}_t(j_t, j_{t-1}) p_t^*(j_t, j_{t-1}), \\
\boldsymbol{C}_t(j_t) &= \sum_{j_{t-1}=1}^{4} \{\boldsymbol{C}_t(j_t, j_{t-1}) + (\boldsymbol{m}_t(j_t) - \boldsymbol{m}_t(j_t, j_{t-1}))(\boldsymbol{m}_t(j_t) - \boldsymbol{m}_t(j_t, j_{t-1}))'\} \\
&\quad \times p_t^*(j_t, j_{t-1}).
\end{aligned} \tag{7.2.32}$$

Since n_t is fixed across models applying at any time, we do not need to consider part (d) of Example 7.1.3.2.

For each j_t, the mixture posterior $p(\boldsymbol{\theta}_t, \phi \mid M_t(j_t), M_t(j_{t-1}), D_t)$ is then approximated by single normal/gamma distribution posteriors having marginals

$$\begin{aligned}
(\boldsymbol{\theta}_t \mid M_t(j_t), D_t) &\sim \mathrm{T}_{n_t}(\boldsymbol{m}_t(j_t), \boldsymbol{C}_t(j_t)), \\
(\phi_t \mid M_t(j_t), D_t) &\sim \mathrm{Ga}(n_t/2, d_t(j_t)/2).
\end{aligned} \tag{7.2.33}$$

These approximate the components in the mixture (7.2.24), thus collapsing from four to one standard normal/gamma component. This completes the cycle of evolution from time $t-1$ to time t.

7.3 Urine Output Series: Multi-Process Modelling

The model analysis is illustrated using the urine output series and we note that the model parameters and values given below have not been estimated in any way. This example is for illustration purposes only and is adapted from [1]. The second-order polynomial multiprocess model discussed in Section 7.2 is used. The four component multiprocess model at time t is as follows, with V_t, and $\boldsymbol{W}_t$ defined by Equations (7.2.2) (with $\delta_V = 0.9$):

1. Routine DLM: $\{\boldsymbol{F}, \boldsymbol{G}, V_t V_t(1), \boldsymbol{W}_t(1)\}$, having fixed model probability $\pi(1) = 117/120$, where $V_t(1) = 1$, and $\boldsymbol{W}_t(1) = \boldsymbol{W}_t$ with $\delta_\mu = \delta_\beta = 0.9$ (see Equations (7.2.2)).

2. Outlier DLM: $\{\boldsymbol{F}, \boldsymbol{G}, V_t V_t(2), \boldsymbol{W}_t(2)\}$, having fixed model probability $\pi(2) = 1/120$, where $V_t(2) = 1000$ is an inflated variance consistent with the occurrence of observations that would be extreme in the standard DLM, and $\boldsymbol{W}_t(2) = \boldsymbol{W}_t(1) = \boldsymbol{W}_t$.

3. Level Change DLM: $\{\boldsymbol{F}, \boldsymbol{G}, V_t V_t(3), \boldsymbol{W}_t(3)\}$, having fixed model probability $\pi(3) = 1/120$, where $V_t(3) = 1$, and $\boldsymbol{W}_t(3)$ is an evolution matrix consistent with level changes. In this example, $\boldsymbol{W}_t(3)$ takes the form of $\boldsymbol{W}_t$ but with $\delta_\mu = 0.01$.

4. Growth Change DLM: $\{\boldsymbol{F}, \boldsymbol{G}, V_t V_t(4), \boldsymbol{W}_t(4)\}$, having fixed model probability $\pi(3) = 1/120$, where $V_t(3) = 1$, and $\boldsymbol{W}_t(4)$ is an evolution matrix consistent with growth changes. In this example, $\boldsymbol{W}_t(3)$ takes the form of $\boldsymbol{W}_t$ but with $\delta_\beta = 0.01$.

Initial priors are defined by

$$\begin{aligned} (\boldsymbol{\theta}_0 \mid D_0) &\sim \mathrm{T}_{n_0}(\boldsymbol{m}_0, \boldsymbol{C}_0), \\ (\phi_0 \mid D_0) &\sim \mathrm{Ga}(n_0/2, d_0/2), \end{aligned} \tag{7.3.1}$$

where

$$\boldsymbol{m}_0 = \begin{pmatrix} 0.55 \\ -0.2 \end{pmatrix}, \quad \boldsymbol{C}_0 = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.001 \end{pmatrix}, \tag{7.3.2}$$

$n_0 = 20$ and $d_0 = 2$, so that the initial estimate for the observational variance is $S_0 = d_0/n_0 = 0.1$.

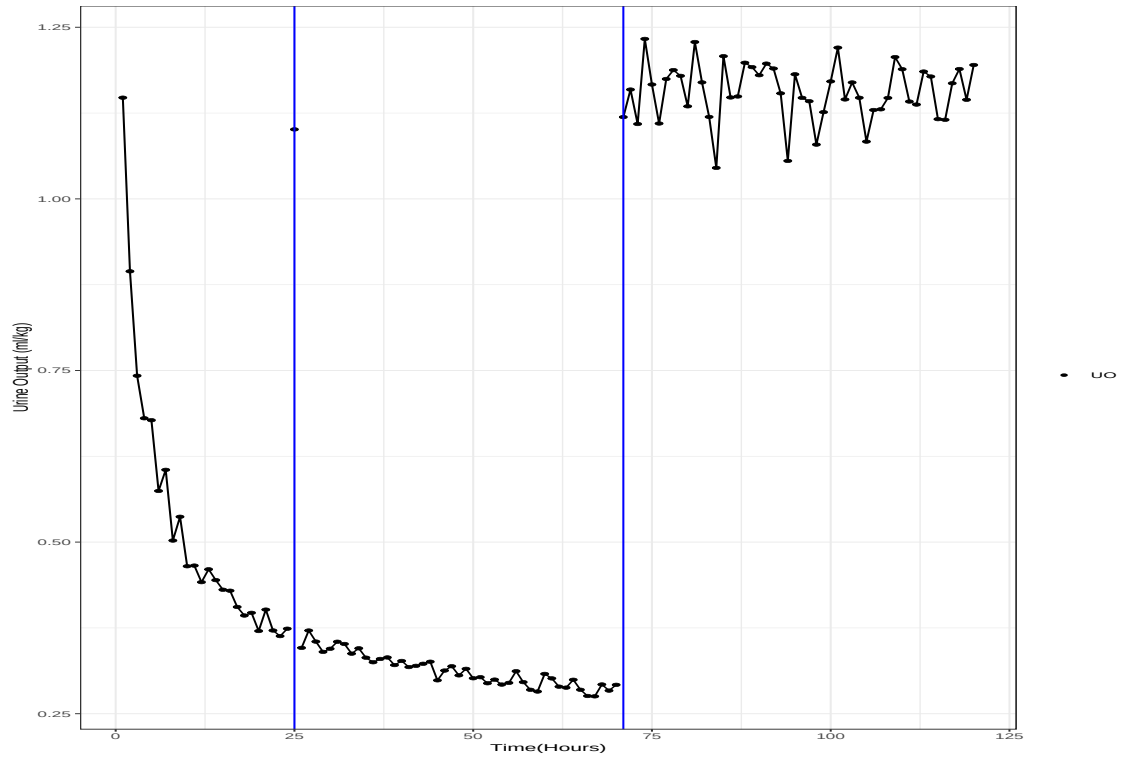

Figure 7.1: Urine output series with sudden changes at hours 25 and 71. The types of change correspond to an outlier, at hour 25, and a parametric change, at hour 71

Figure 7.1 shows a urine output series for a patient. The patient starts off with an exponential decay in urine output, and the urine output series has two sudden changes. The first sudden change, at hour 25, is an outlier, and the second sudden change, at hour 71 is a parametric change. Retrospectively these changes are clear and easy to identify. However, in real time, an outlier is indistinguishable from a parametric change (see Figure 6.2). Fortunately, using probabilistic mixtures, we have a means of distinguishing between types of change by using one-step back posterior probabilities.

Consider now the position at hour 24. Up until this time, the series is stable and is modelled well by the routine DLM, M_1. The posterior model probabilities shown in Figure 7.2 illustrate how the routine DLM dominates during this stable period and hence the routine DLM carries more weighting in the point forecasts and in the updating of parameters (see Equations (7.2.21) and (7.2.16)). Figure 7.2 shows that, during the initial stable period, the posterior model probabilities for models

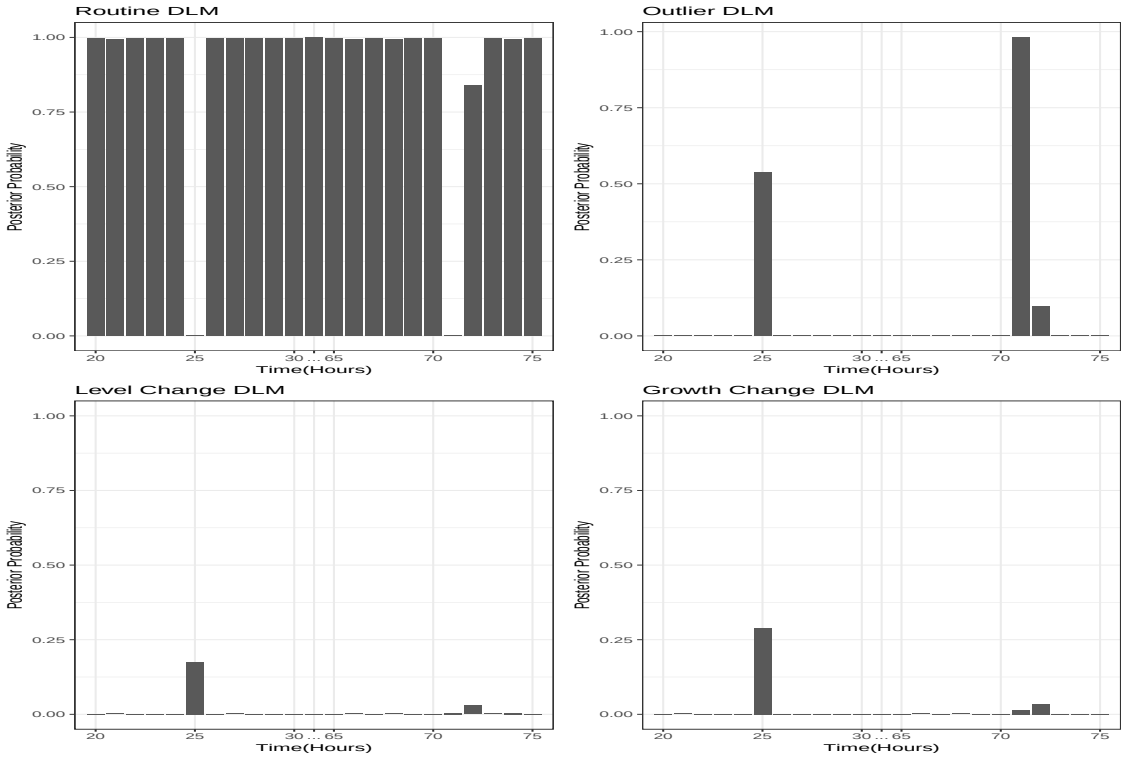

Figure 7.2: Posterior model probabilities showing (from left to right) the posterior probabilities of the routine, outlier, level change, and growth change DLMs respectively

M_2, M_3 and M_4 are very small. This reflects that the forecast distributions from the outlier model, and the level and growth change models, carry little weighting to the mixed forecasts for the first 24 hours.

Figure 7.3 displays the posterior density (7.2.23) of the level parameter, $p(\mu_{24} \mid D_{24})$, at hour 24. The four (state) components (7.2.12) are also plotted on the graph. Note that the routine (red) and level change (green) components are underneath the black (mixed) component. Even though the discount factor for the level change model $\delta_\mu(3) = 0.01$, and the discount factor for the routine model, $\delta_\mu(1) = 0.9$ we see that the level posterior for these models is (approximately) identical at time 24. This is because the discount factors only affect the prior distributions (through $\boldsymbol{R}_t(j_t, j_{t-1}) = \boldsymbol{G}\boldsymbol{C}_{t-1}(j_{t-1})\boldsymbol{G}' + \boldsymbol{W}_t(j_t)$). The level priors at time 25 would look very different for those conditioning on $M_{24}(1)$ and $M_{24}(3)$.

From Figure 7.3 we can see that the components are all located between approximately -0.791 and -0.747, and mostly spread between -0.9 and -0.6. The posterior based on $M_{24}(1)$ is the most peaked and has posterior probability $p_{24}(1) = 0.997$.

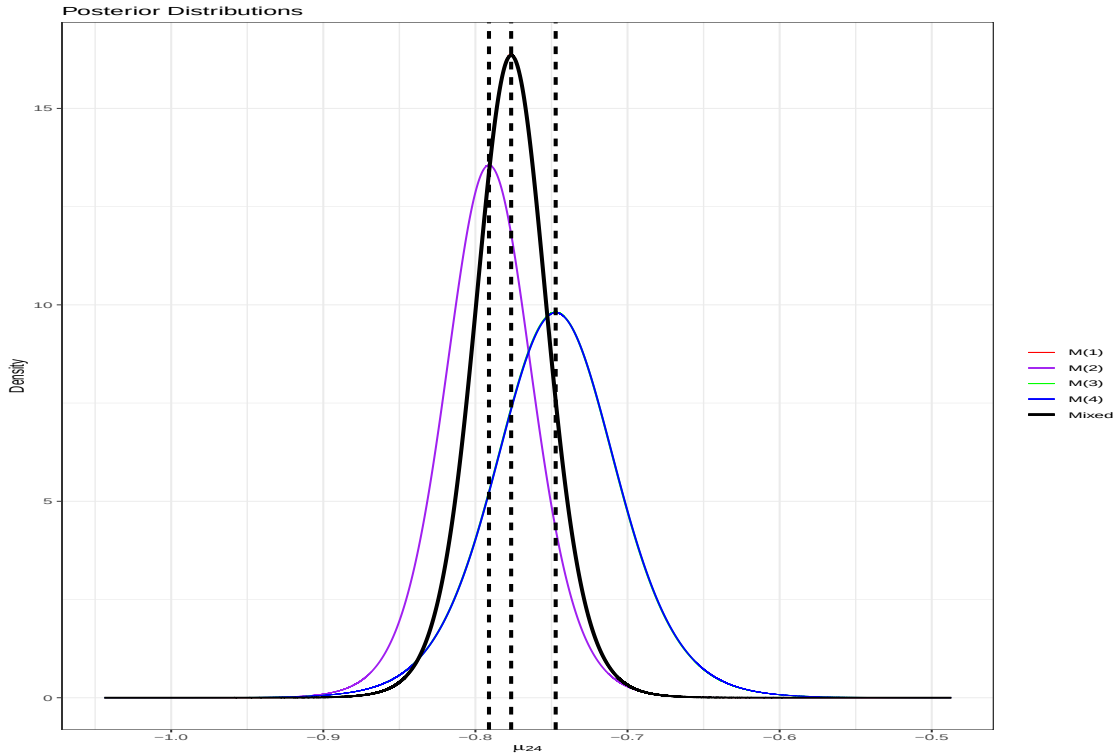

Figure 7.3: Posteriors for μ_{24} at $t = 24$. The black line represents the mixed posterior, the red, purple, green and blue lines represent the posteriors for the routine, outlier, level change, and growth change DLMs respectively. The three dashed lines, from left to right, represent $m_{24}(2) = -0.79, m_{24}(1) \approx m_{24}(3) \approx m_{24} = -0.78$ (where m_{24} is the median for the mixed level component at time $t = 24$), and $m_{24}(4) = -0.75$.

This reflects the previous stability of the series and consistency of the routine DLM. Thus, the mixture of these four components (Figure 7.3) closely resembles the routine component (so closely that the mixture overlaps with the posterior based on $M_{24}(1)$). This is standard in stable periods.

Figure 7.4 displays the one-step ahead forecast density (7.2.16), $p(Y_{25} \mid D_{24})$, at hour 24. The sixteen components (7.2.14) are also plotted. The mixture is shown in black and corresponds closely to the highly peaked, red components near the centre. The more diffuse, purple components correspond to the models that condition on the outlier model at time 25. The observational variance inflation factor of 1000 in the definition of $M_t(2)$ makes these components more spread. However, the outlier, level change, and growth change components have small probabilities and so contribute little to the mixture. This is to be expected in stable periods.

Moving now to hour 25. $Y_{25} = 1.10$, corresponding to 0.18 on the log-scale ($\log(Y_{25}+$

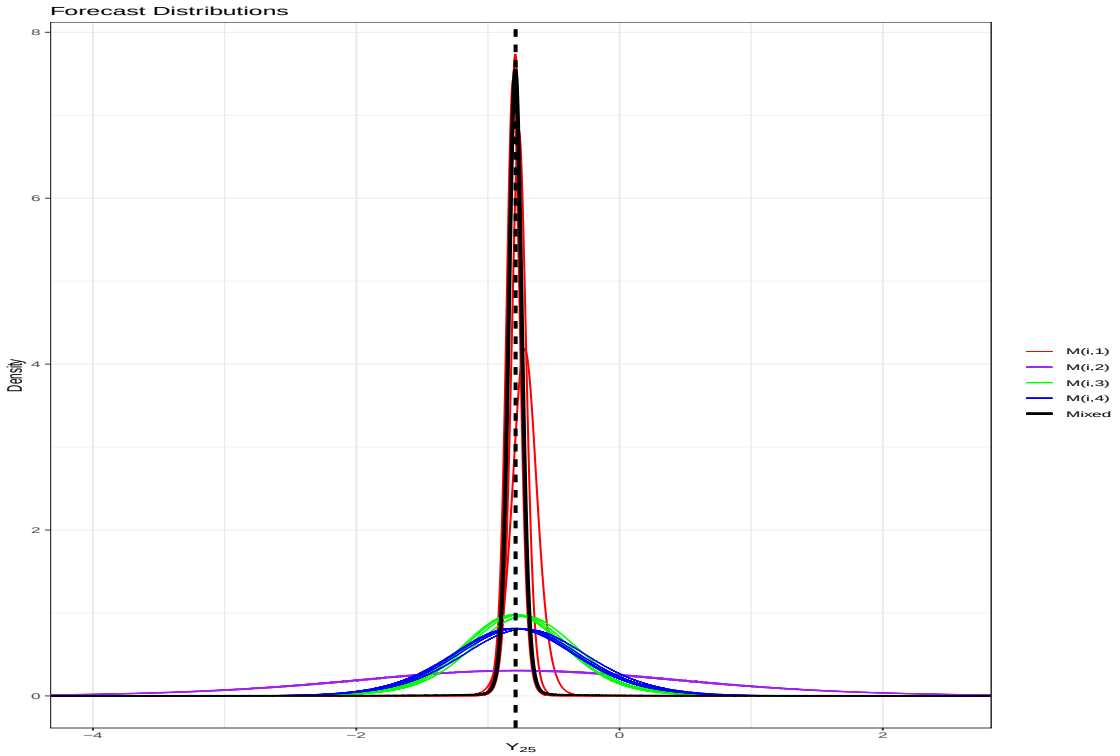

Figure 7.4: Forecasts for Y_{25} at $t = 24$. The black line represents the mixed forecast distribution, the red, purple, green and blue lines represent the forecast distributions that include $M_{25}(1)$, $M_{25}(2)$, $M_{25}(3)$, and $M_{25}(4)$, respectively. The dashed line represents $Y_{25} \approx -0.79$

0.1)), is a wild observation relative to the standard forecast distribution, $p(Y_{25} \mid D_{24})$, and most of the probability under the mixture density in Figure 7.4 is concentrated between -1 and -0.6, on the transformed scale, corresponding to around 0.27 and 0.45 on the untransformed scale. The alternative models, however, in particular the outlier components, give appreciable probability to values larger than 0.18. Hence, in updating to posterior model probabilities given D_{25}, the components that include $M_{25}(2)$ (and also to a lesser extent $M_{25}(3)$, and $M_{25}(4)$) will receive much increased weights. This is evident from Figure 7.2. The outlier, level change, and growth change models share most of the posterior probability at $t = 25$.

Figure 7.5 plots the posterior density, $p(\mu_{25} \mid D_{25})$, at hour 25 alongside the four individual components. This plot is very much different to Figure 7.3. The components are more disparate and the mixed posterior is bimodal, reflecting the ambiguity to whether the wild observation, Y_{25}, is an outlier or a parametric change. The (purple) peaked component located near -0.8 corresponds to the outlier model for Y_{25},

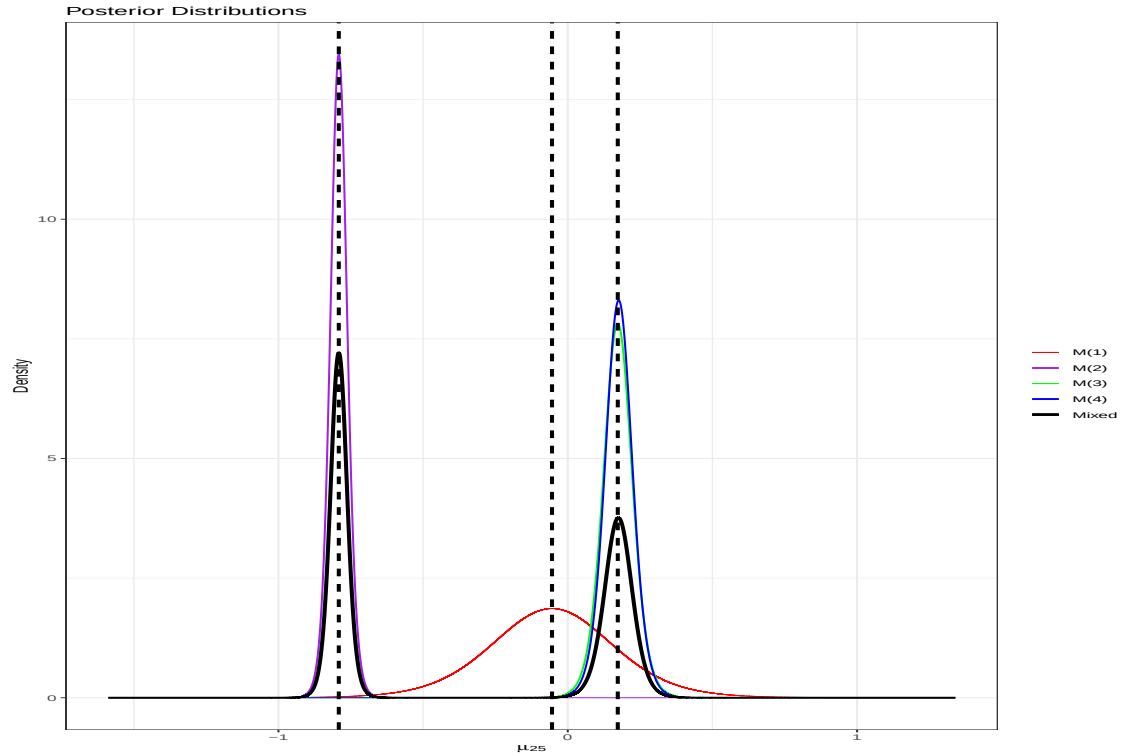

Figure 7.5: Posteriors for μ_{25} at $t = 25$. The black line represents the bimodal mixed posterior, the red, purple, green and blue lines represent the posteriors for the routine, outlier, level change, and growth change DLMs respectively. The three dashed lines, from left to right, represent $m_{25}(1) = -0.05, m_{25}(2) = -0.79$, and $m_{25}(3) \approx m_{25}(4) \approx 0.18$, respectively

this component represents the density $p(\mu_{25} \mid M_{25}(2), D_{25})$. In this scenario, the observation is considered an outlier and is ignored, the inference being that the level remains between (around) -0.9 and -0.7. That is, the state is unchanged and the wild observation is an outlier.

The two peaked (green and blue respectively) components (in Figure 7.5) located around the region 0.18 (corresponding to the observation Y_{25} on the transformed-scale) come from the level and growth change models. If the observation is not an outlier, and corresponds to a parametric change (in level and/or growth), then the inference is that the current level lies between 0 and 0.35. The fourth (red), more diffuse component located near -0.05 is $p(\mu_{25} \mid M_{25}(1), D_{25})$, the posterior from the routine model at time $t = 25$. The inference here being that if Y_{25} is a reliable observation and no level or growth change has occurred, then the posterior for the level is between the prior located near -0.8 and the likelihood from Y_{25} located around 0.18. Intuitively, conditional on $M_{25}(1)$, if observation Y_{25} is consistent with

the routine model, this means that observations that were considered extreme at time 24 are now consistent with the routine model. Hence, the updated routine model is more diffuse allowing for a larger range of values; indicating more (routine) uncertainty in the series. However, as can be seen from Figure 7.2, this density receives negligible weighting in the mixture at time 25.

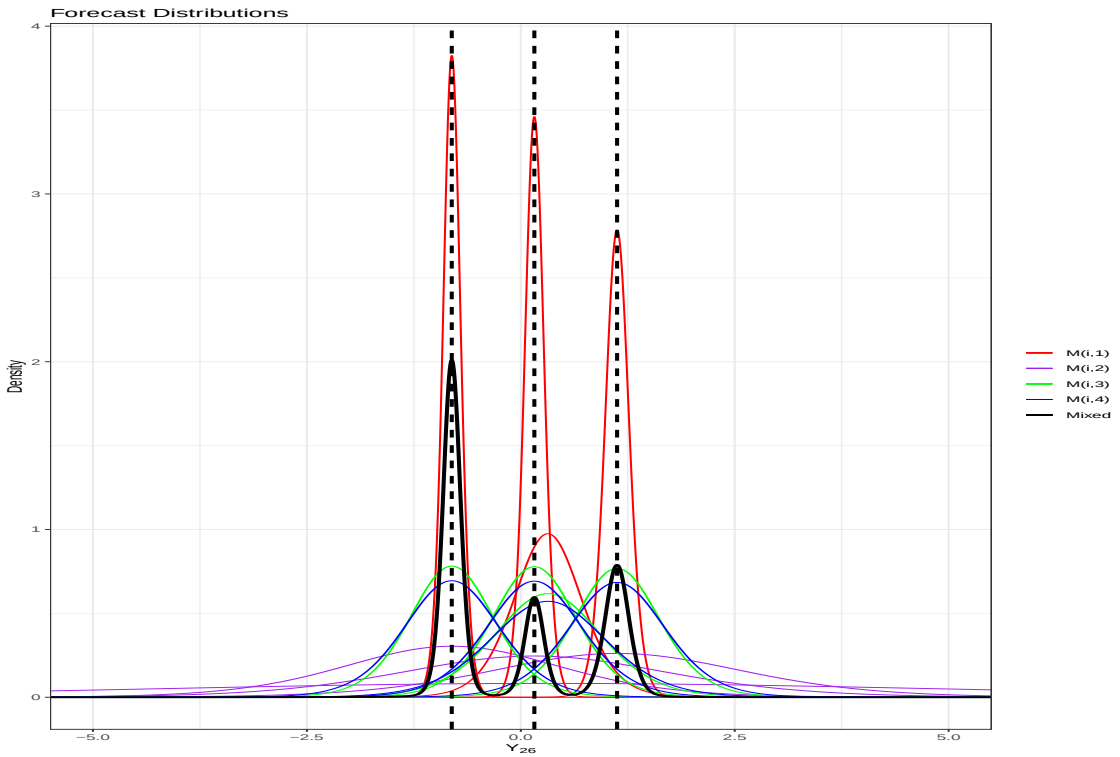

Figure 7.6: Forecasts for Y_{26} at $t = 25$. The black line represents the mixed forecast distribution, the red, purple, green and blue lines represent the forecast distributions that include $M_{26}(1)$, $M_{26}(2)$, $M_{26}(3)$, and $M_{26}(4)$ respectively. The three dashed lines, from left to right, represent $f_{26}(2) \approx Y_{26} \approx -0.81, f_{26}(3) = 0.16$ and $f_{26}(4) = 1.12$, respectively

The bimodal posterior at time 25 represents uncertainty to whether the extreme observation represents an outlier or a parametric change. Until another observation is made, there is a split between the two inferences (although in favour of the observation being an outlier, see Figure 7.2). Forecasting ahead to time $t = 26$, the ambiguity is once again clear from Figure 7.6. The forecast distribution, in Figure 7.6, is multimodal. The components located around -0.81, correspond to the forecasts that depend on the outlier model at time $t = 25$, that is $(Y_{26} \mid M_{25}(2), M_{26}(i))$ for $i = 1, 2, 3, 4$. Similarly, the components located around 0.16, correspond to the forecasts that depend on the level change model at time $t = 25$, that is $(Y_{26} \mid M_{25}(3), M_{26}(i))$ for $i = 1, 2, 3, 4$; and the components located

around 1.12, correspond to the forecasts that depend on the growth change model at time $t = 25$, that is $(Y_{26} \mid M_{25}(4), M_{26}(i))$ for $i = 1, 2, 3, 4$.

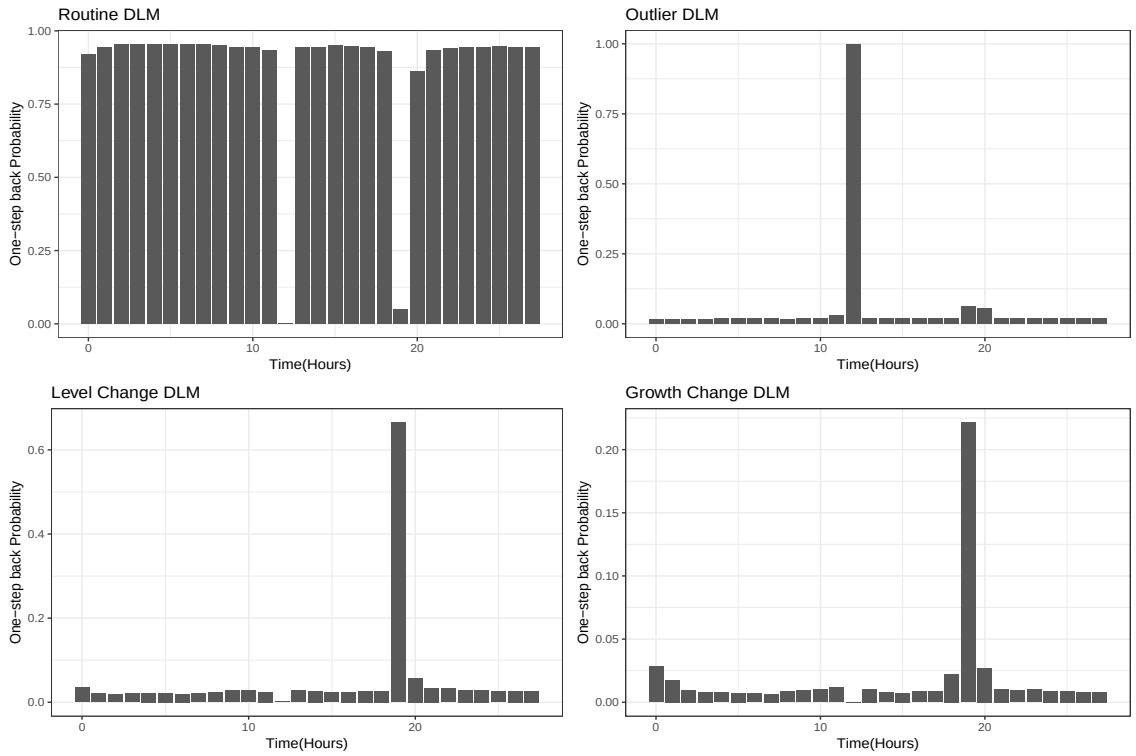

Figure 7.7: One-step back model probabilities showing (from left to right) the probabilities of the routine, outlier, level change, and growth change DLMs, respectively

An hour later we observe $Y_{26} = 0.35$, or -0.81 on the transformed-scale. It is now clear that observation $Y_{25} = 1.10$ was an outlier. Figure 7.7 represents one-step back posterior model probabilities. For each time, the bars in Figure 7.7 represent the retrospective probabilities $\Pr[M_{t-1}(j_{t-1}) \mid D_t]$ for $j_{t-1} = 1, \ldots, 4$. These probabilities are extremely useful for retrospective assessment of model occurrence at any time given one further observation. In this example, observing Y_{26} verifies the position at time 25, confirming that Y_{25} was in fact and outlier, with $\Pr[M_{25}(2) \mid D_{26}] \approx 1$. The outlier has been identified and is now ignored when updating to the posterior distributions at time 26.

Figure 7.8 (left) shows $p(\mu_{26} \mid D_{26})$, along with its four components, after having identified Y_{25} as an outlier. The corresponding one-step ahead forecast densities are shown in Figure 7.8 (right). It is clear from Figure 7.8 that the series is stable again, with unimodal distributions, once the extreme observation has been identified.

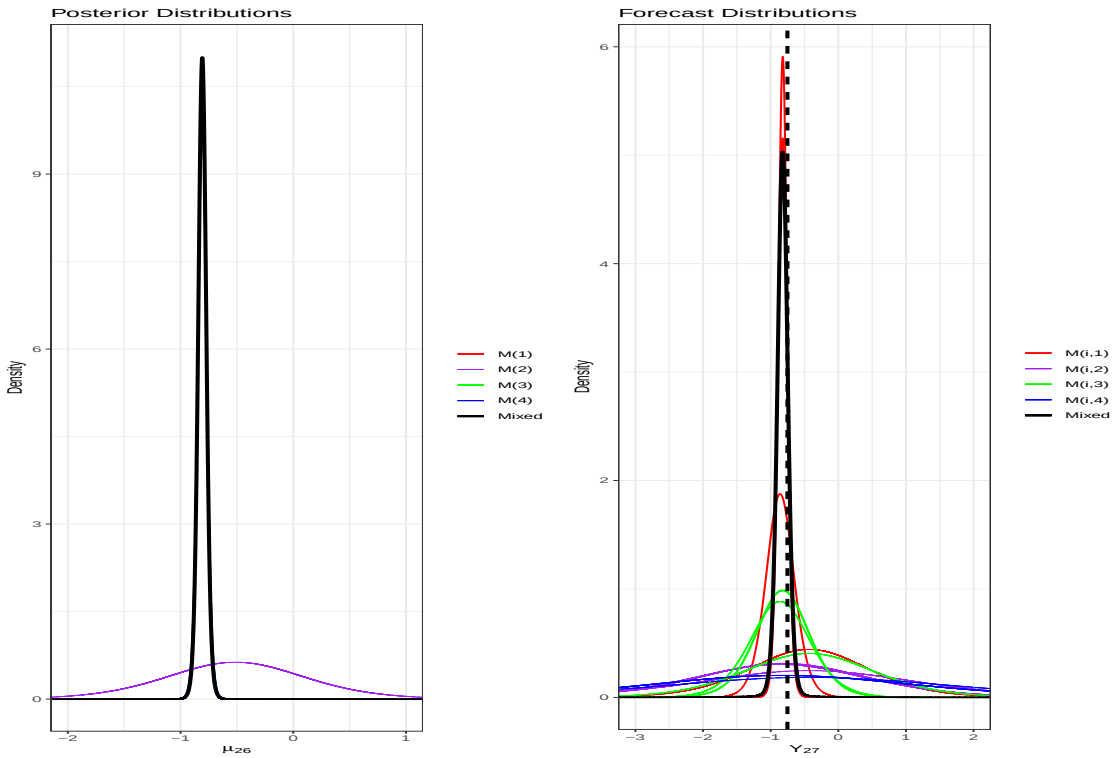

Figure 7.8: Perspective at time 26. Plots show posterior distribution (left) and (one-step) forecast distributions at time 26 after identifying Y_{25} as an outlier. Line colours for posteriors and forecasts are described in Figures 7.5 and 7.6, respectively

To summarise, the posteriors for model parameters tend to be very similar during stable periods, but most of the weighting in the posterior mixture is from the routine posterior distribution. These components separate out at the onset of an event due to ambiguity to whether the event corresponds to an outlier or a parametric change. A further observation usually identifies the event, and posteriors then update to reflect this, and the series reverts back to being consistent with the routine DLM.

To illustrate the analysis further, consider the parametric change at time $t = 71$. Figure 7.9 shows plots for $p(\mu_t \mid D_t)$, along with the four components $p(\mu_t \mid M_t(j_t), D_t)$, $j_t = 1, \ldots, 4$, for $t = 71$ and $t = 72$. Here the extreme observation Y_{71} initially leads to bimodality in the posterior for μ_{71}, see Figure 7.9 (left), and once again the posterior model probabilities favour the outlier model, see Figure 7.2. The observation is either an outlier or the onset of a parametric change, but the model cannot distinguish between the two until more information is available. Observing Y_{72} confirms a level change, from around -0.93 to 0.23 (from around 0.29 to 1.15 on the untransformed scale). Finally, current and one-step back probabilities indicate

the switching between models over time and the diagnosis of events.

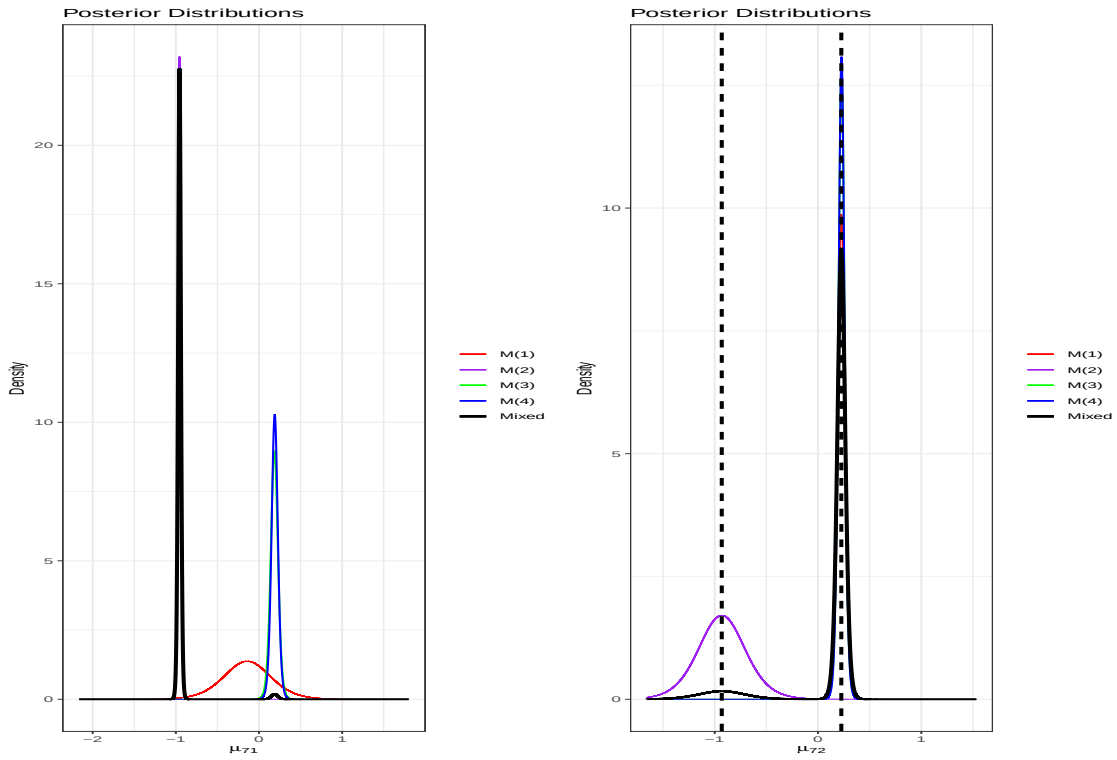

Figure 7.9: Posteriors at hours 71 and 72. Plot (left) shows ambiguity to whether observation Y_{71} corresponds to an outlier or parametric change. Plot (right) shows posteriors after one further observation identifies Y_{71} as a parametric change. The two dashed lines, from left to right, represent $m_{72}(2) = -0.93$, and $m_{72}(1) \approx m_{72}(3) \approx m_{72}(4) \approx 0.23$, respectively. Line colours are described in Figure 7.5

Figure 7.10 shows the urine output series with one-step ahead forecasts and corresponding 95% prediction intervals at each time point. From Figure 7.10 we can see that the forecasts after Y_{26} are unaffected by the outlier Y_{25}. This is because the multiprocess model was able to identify Y_{25} as an outlier after observing Y_{26}. Similarly, we can see that the forecasts are able to adapt to the parametric change at hour 71. After observing Y_{72} the multiprocess model is able to identify that the sudden change at hour 71 is a parametric change and not an outlier and the model adapts accordingly.

7.3.1 Forecasting k-Steps Ahead

Throughout this thesis we have emphasised that our aim is to forecast the next six urine outputs in order to predict severe oliguria to prevent adverse outcomes. Using

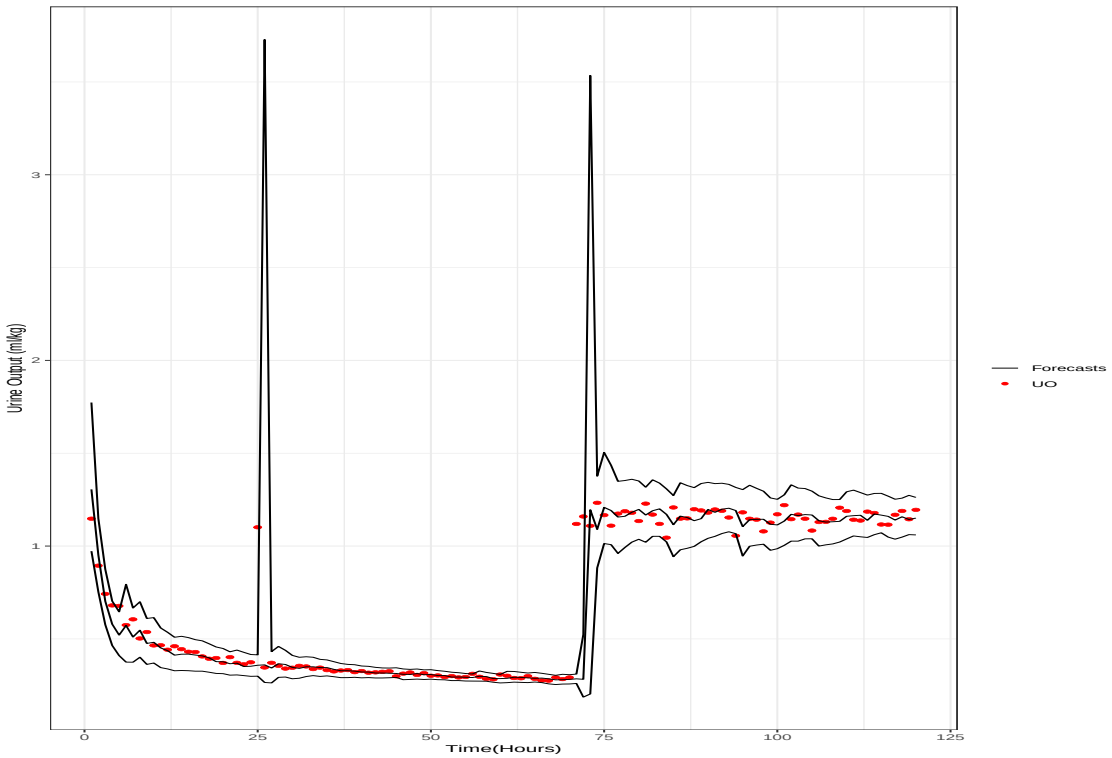

Figure 7.10: One-step ahead modal point forecasts (see Equation (7.2.16)) with corresponding 95% prediction intervals. Red dots represent urine output recordings and the black lines represent the modal forecasts and corresponding 95% prediction intervals

the four-component model discussed in Section 7.2, we will develop the methods required to forecast k-steps ahead.

Consider the position at time t with the posterior distributions summarised in (7.2.26). The one-step ahead forecast distribution for Y_{t+1} is given by (see Equation (7.2.16)) the 4^2 component mixture of T forecast densities

$$p(Y_{t+1} \mid D_t) = \sum_{j_{t+1}=1}^{4} \sum_{j_t=1}^{4} \pi(j_{t+1}) p_t(j_t) p(Y_{t+1} \mid M_{t+1}(j_{t+1}), M_t(j_t), D_t). \tag{7.3.3}$$

Consider now forecasting two steps ahead for Y_{t+2}. Given any combination of models at time $t, t+1, t+2$, it follows that the two-step ahead forecast distribution has density [1]

$$p(Y_{t+2} \mid M_{t+2}(j_{t+2}), M_{t+1}(j_{t+1}), M_t(j_t), D_t), \tag{7.3.4}$$

where j_t, j_{t+1} and j_{t+2} can each take values from 1 to 4. Once again, using Bayes'

rule, the mixing probabilities are given by

$$\begin{aligned}
&\Pr[M_{t+2}(j_{t+2}), M_{t+1}(j_{t+1}), M_t(j_t) \mid D_t], \\
&= \Pr[M_{t+2}(j_{t+2}) \mid M_{t+1}(j_{t+1}), M_t(j_t), D_t] \\
&\times \Pr[M_{t+1}(j_{t+1}), M_t(j_t) \mid D_t], \\
&= \Pr[M_{t+2}(j_{t+2}) \mid M_{t+1}(j_{t+1}), M_t(j_t), D_t] \\
&\times \Pr[M_{t+1}(j_{t+1}) \mid M_t(j_t), D_t] \Pr[M_t(j_t) \mid D_t], \\
&= \pi(j_{t+2})\pi(j_{j+1})p(j_t).
\end{aligned} \tag{7.3.5}$$

Using (7.3.5), the two-step ahead forecast distribution for Y_{t+2} is the 4^3 component mixture

$$\begin{aligned}
p(Y_{t+2} \mid D_t) = \sum_{j_{t+2}=1}^{4} \sum_{j_{t+1}=1}^{4} \sum_{j_t=1}^{4} &\pi(j_{t+2})\pi(j_{t+1})p_t(j_t) \\
&\times p(Y_{t+2} \mid M_{t+2}(j_{t+2}), M_{t+1}(j_{t+1}), M_t(j_t), D_t).
\end{aligned} \tag{7.3.6}$$

The above derivation can be extended for forecasting further ahead. The k-step ahead forecast distribution for Y_{t+k} is the 4^k component mixture

$$\begin{aligned}
p(Y_{t+k} \mid D_t) = \sum_{j_{t+k}=1}^{4} \cdots \sum_{j_{t+1}=1}^{4} \sum_{j_t=1}^{4} &\pi(j_{t+k}) \ldots \pi(j_{t+1})p_t(j_t) \times \\
&p(Y_{t+k} \mid M_{t+k}(j_{t+k}), \ldots, M_{t+1}(j_{t+1}), M_t(j_t), D_t).
\end{aligned} \tag{7.3.7}$$

Forecasting further ahead increases the number of components in the mixture to account for all possible models obtaining between time t and the forecast time point $t+k$ [1]. It is clear that, once again, the computational demand for mixture models can become computationally infeasible. Forecasting k-steps ahead requires a mixture of 4^k components. As k increases the number of components in the mixture tends to infinity.

In some models (such as the second-order polynomial model) features can be exploited to simplify the problem of summarising the forecast distributions given by (7.3.7) [1]. For example, the structure of the second-order polynomial multiprocess model is such that the models, $M_t(j)$, differ only through the evolution and observational variances for all t and all j. In particular, $\mathrm{E}[Y_{t+1} \mid M_{t+1}, M_t(j_t), D_t]$ does not depend on j_{t+1}, and $\mathrm{E}[Y_{t+k} \mid M_{t+k}, \ldots, M_{j_t}(j_t), D_t]$ does not depend on j_{t+l} for $l = 1, \ldots, k$. This means that there are only four distinct point forecasts for each

time point. Other simplifications mentioned in Section 7.1.2 can be used to reduce the number of calculations required in mixtures.

7.4 Multi-Process Models with Monitoring

Recall in Section 7.1.2 that we mentioned that a way of reducing computational demand of multiprocess models is to ignore components that have very small posterior probabilities. The reason being that computing mixtures at points of stability (when the routine model will dominate the mixture) is computationally redundant and we may ignore components that carry little weighting in the mixtures. This reduction method leads us to combine methods from this chapter with methods discussed in Chapter 6, as in Ameen and Harrison (1985b) [34].

Most of the time, a series will be stable over periods of observations. This will mean that most of the time, during mixtures, most of the weighting will come from the routine DLM and hence the multi-process will be redundant in these periods. Only at times of sudden changes and exceptional events, and in regions locally following these events, is the full multiprocess model required. The following approach recognises this and combines Chapters 6 and 7 to avoid the redundancy of mixture models but also utilises the multi-process framework simultaneously [1].

1. Model the series with the routine DLM subject to a monitoring system (such as the system described in Section 6.7). This way we can use the routine model for the stable periods and have a system in place to monitor the model's performance. If the monitoring system indicates that the data is consistent with the routine model, then continue to use the routine model as usual. At the same time, update alternative models that allow for the anticipated forms of change from the routine model (see Section 7.2 for an example regarding a specific type of model). However, do not perform any computationally costly mixtures at this point. Just update the alternative models for when they are required (see stage 2.).

2. When the monitoring system signals deterioration in forecast performance of the routine DLM, begin a period of using the multi-process (class II) approach. This will be when we use the alternative models from stage 1. As a reminder,

using only models from $h = 1$ steps back (or the value of h chosen by the forecaster) as approximations for mixtures; and the methods of reduction discussed in Section 7.1.2 are still used here.

3. Continue to use the multi-process model until the "unstable" period is over. This will be when the posterior model probabilities are essentially negligible for all alternative models, and hence are near unity for the routine model.

4. Switch back to stage 1. and continue.

The above scheme utilises the powerful framework of multi-process models and also provides a solution to the computational redundancy during stable periods.

Chapter 8

Shiny Application

8.1 Introduction

Shiny is an R package that builds interactive web applications from R. Shiny applications are an easy way to allow clinicians to use the models discussed throughout this thesis, without having to know how to code and without having to know any statistics. In this chapter we show how clinicians can use the shiny application to monitor the functioning of a patient's kidneys by using the DLM with model monitoring, discussed in Section 6.7. This application was designed to display and monitor information needed by clinicians at the University Hospital of South Manchester.

8.2 Using the Application

When a clinician runs the shiny application they are prompted by three dropdown boxes (see Figure 8.1). These boxes allow the clinician to choose how many steps ahead they would like to forecast, k (in our study this was 6); which patient they would like to monitor, "Select Patient"; and the hour at which they would like to monitor, "Enter Observation". The option to choose an hour to monitor allows the clinician to view what has happened in the past, and at the current time, t. Below these dropdown boxes is a summary of the patient. This is to allow clinicians to

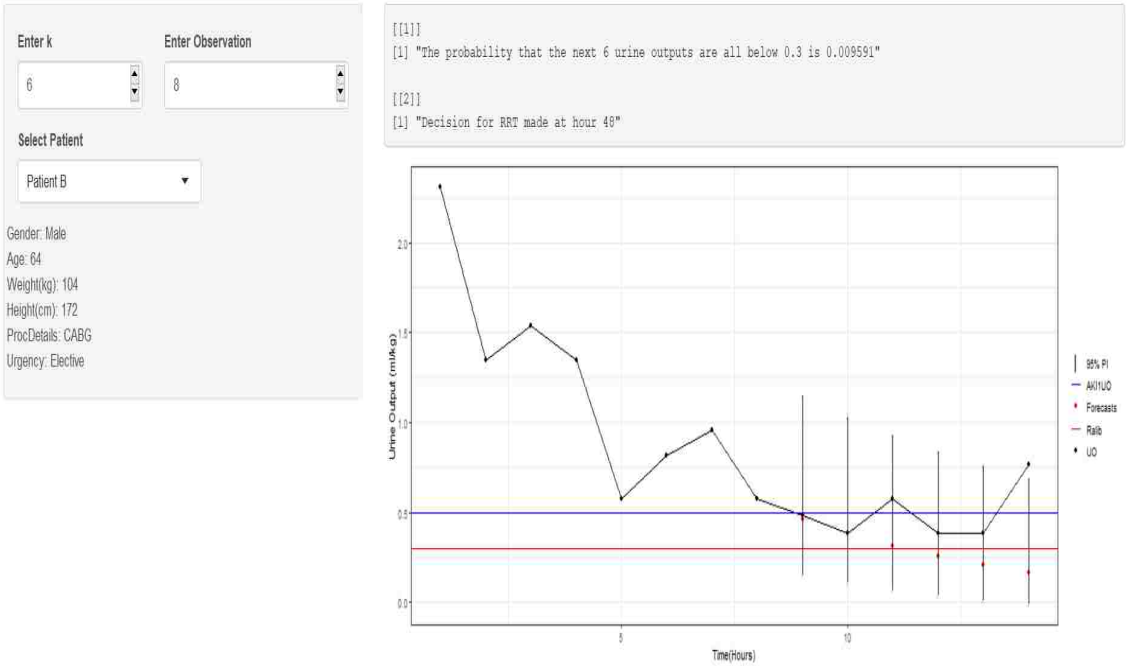

Figure 8.1: Shiny application. In the top left of the screen you can see dropdown boxes that clinicians can easily use to monitor patients. Below the dropdown boxes is a description of the patient's history. On the right side of the interface we see a box showing the joint probability of suffering severe oliguria at the chosen hour; and below this box we see a plot of the urine output time series with corresponding k-step ahead forecasts and corresponding 95% prediction intervals

easily view the patient's history. To the right we see a box that displays the joint probability that a patient will suffer severe oliguria in six hours time; this probability will be used by clinicians to make key decisions about a patient's kidneys. Below this warning box we see a plot of the patient's urine output; with k-step ahead point forecasts and corresponding prediction intervals at times $t+1, \ldots, t+k$.

Below the plot of the forecasts there is a table summarising key results of the model. Table 8.2 displays the forecast times; the point forecasts; the lower intervals; the upper intervals; the observations; the residuals; and the probabilities that a forecast is less than $0.3ml/kg$, i.e.

$$\Pr(f_{t+k} < 0.3 \mid D_t), \tag{8.2.1}$$

for $k = 1, \ldots, 6$. Below the summary table are (biochemistry) plots of interest to clinicians (see Figures 8.3 and 8.4). Having all of these available in one place is extremely convenient for clinicians and allows clinicians to monitor patients with

Time (Hours)	Forecasts	Lower Interval	Upper Interval	y	e	P
9	0.46	0.15	1.15	0.48	0.60	0.059871
10	0.38	0.11	1.03	0.38	-0.02	0.128784
11	0.32	0.07	0.93	0.58	0.35	0.230702
12	0.26	0.04	0.84	0.38	0.59	0.353603
13	0.21	0.01	0.76	0.38	0.59	0.479107
14	0.17	-0.01	0.69	0.77	0.58	0.592130

Figure 8.2: Below Figure 8.1 is a summary table. This table shows quantities of interest to the forecaster, namely the forecast times; the point forecasts; the lower intervals; the upper intervals; the observations; the residuals; and the probability that a forecast is less than $0.3ml/kg$

ease.

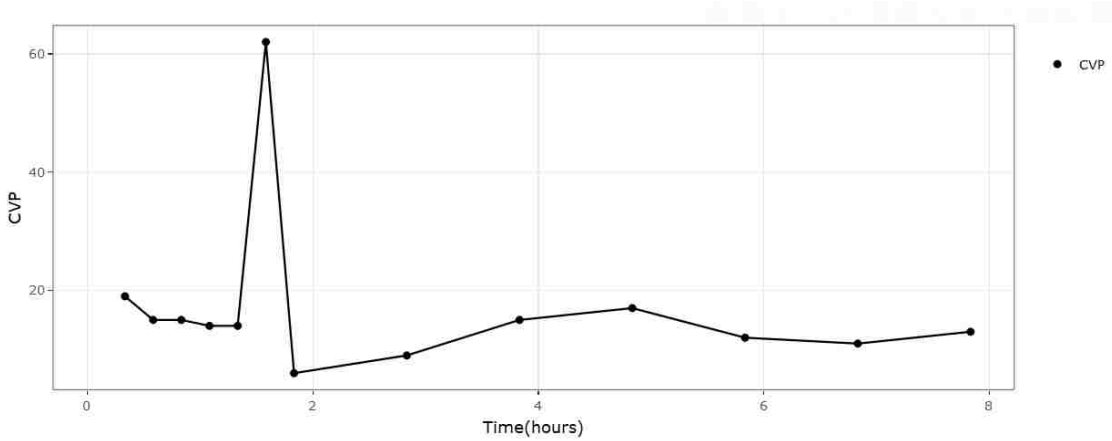

Figure 8.3: Below Figure 8.2 is a plot of the central venous pressure (CVP). This is a quantity that clinicians are often interested in when monitoring patients

Moreover, clinicians can use the application to monitor kidney function during periods when the patient is considered high risk by our DLM. In Figure 8.5 we see that the patient was considered high risk by our DLM at hour 37 (since the patient has been at high risk for 5 hours at hour 42). This high risk count will increase by one if the patient is consecutively considered high risk by our DLM. This allows clinicians to see if interventions used to normalise urine output are working or not. In addition, we see that our DLM indicates that this patient is at high risk 11 hours earlier than experts decided that this patient needed RRT. Furthermore, clinicians can use the application and their subjective expertise to intervene whenever they

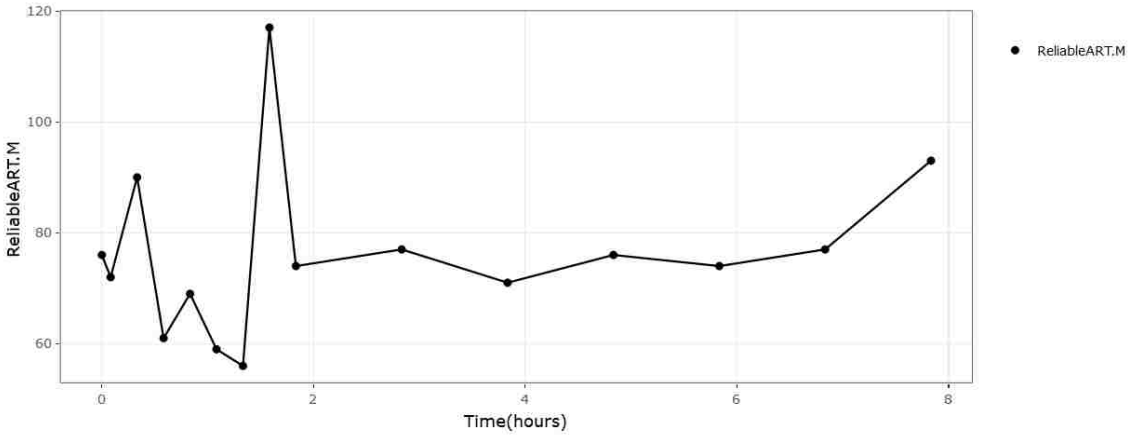

Figure 8.4: Below Figure 8.3 is a plot of the patient's arterial blood pressure. This is another quantity that is useful for clinicians

like. For example, by using the application clinicians can see how rapidly a patient's risk (of severe oliguria) is increasing and intervene accordingly, even before the risk of severe oliguria reaches 0.8. Or, perhaps a clinician gives more fluids to a patient considered high risk; further decisions can then be made (by monitoring this patient using the shiny application) if this patient is still considered high risk in the hours that follow.

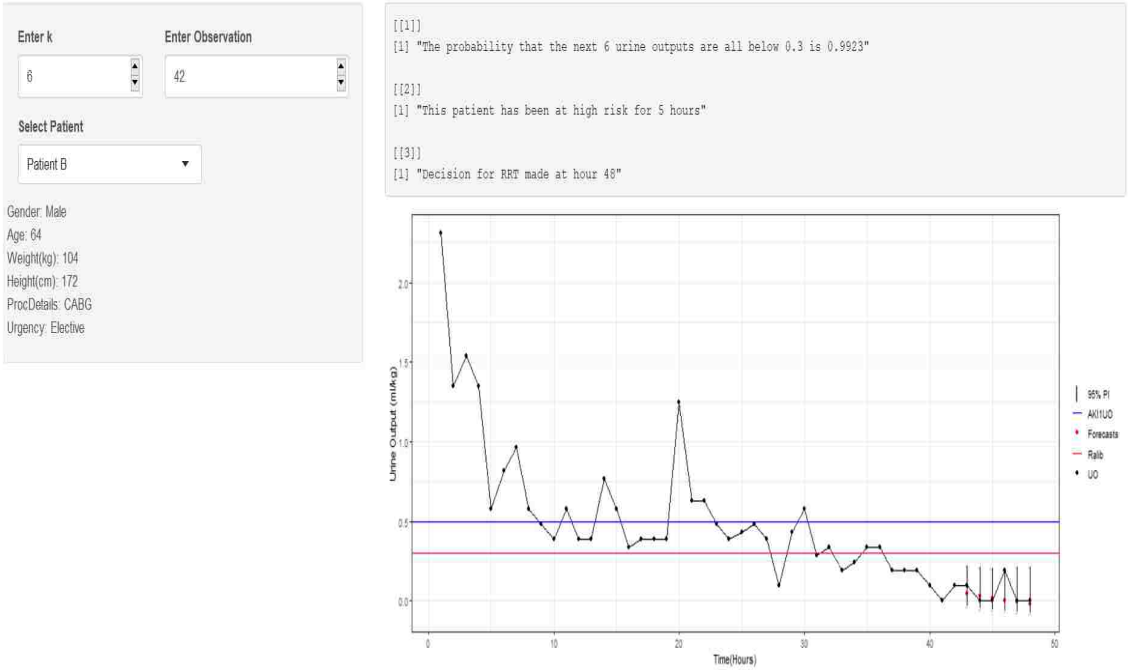

Figure 8.5: Shiny application showing how clinicians can monitor a high risk patient. The box above the urine output plot displays how long patients have been at high risk. This allows clinicians to monitor if interventions are working. In addition, for use in our study, the actual decision time to start RRT is also shown

Chapter 9

Conclusion

This thesis is concerned with the modelling of noisy time series using dynamic linear models. Throughout this thesis the particular application of monitoring renal failure was considered, with the aim of predicting severe oliguria in order to identify patients that are likely to suffer adverse outcomes related to kidney deterioration as early as possible.

Chapter 1 began by introducing the acute kidney injury (AKI) guidelines and how they are used to identify kidney injury and the possible adverse outcomes associated with kidney deterioration. The KDIGO AKI guidelines use two variables to identify kidney injury at different levels of severity. In our study it was observed that sparsely measured variables such as serum creatinine could not be used alone for identifying kidney injury, and this compelled us to monitor kidney injury using a patient's urine output. We discussed how recent studies have found the current AKI stage one urine output criterion to be too sensitive and this lead us to Ralib's work [2]. Ralib's study demonstrated that the stage one urine output criterion is too sensitive and Ralib proposed a lower threshold. We then suggested that by the time Ralib's criterion identifies a patient as likely to suffer adverse outcomes, harm may have already occurred, and hence the need for a better classification for patients likely to suffer adverse outcomes due to kidney injury. This lead us to using statistical models to forecast a patient's urine output time series with the aim of being able to identify a group of patients at higher risk of adverse outcomes before harm occurs.

Chapter 2 introduced (static) linear models and how they can be used to forecast the urine output time series. Data transformations were introduced to address the nonlinearities of the urine output data and to resolve negative, infeasible forecasts. We concluded this chapter by showing the downfall of static models when modelling noisy time series that are subject to sudden changes. This prompted the need for a model with time varying parameters that could adapt to changing trends.

Chapter 3 introduced dynamic linear models (DLMs) and compared DLMs to static models to show the power of a model with time varying parameters. We introduced the observation and system equations describing DLMs and discussed how the system evolution stochastic error term was key to modelling noisy time series. We discussed the conditional independence of DLMs and outlined an approach for building DLMs and modelling noisy time series.

Chapter 4 analysed the mathematical structure of DLMs in detail. We outlined how to update from prior to posterior distributions and showed how to forecast ahead using DLMs. We derived moments for forecasted state and observational distributions for a normal analysis with a known observation variance series, and a known evolution variance series. We derived crucial results, such as joint probabilities, which are used to determine whether a patient is likely to suffer severe oliguria or not. We then introduced component forms and how the principle of superposition can be used to construct more complex DLMs. We then analysed dynamic models with unknown evolution matrices and unknown observation error series. These methods allowed for dynamic models to adapt to regions with more or less stochastic variation and allowed model forecasts to represent more or less uncertainty, resulting in a more precise model.

In Chapter 5 we used DLMs to predict severe oliguria in order to model and monitor kidney deterioration to identify kidney injury and the possible adverse outcomes associated with kidney deterioration. Our aims were to use dynamic models to test the feasibility of realtime screening of patient data to identify those at risk early enough to allow intervention to prevent or lessen harm caused by prolonged oliguria; to compare the model's ability to identify patients at risk of adverse outcomes with the stage one urine output criterion within the KDIGO guidelines (see Table 1.1); and to compare the model's ability to identify patients at risk of adverse outcomes with the criterion proposed by Ralib *et al* (observing six consecutive urine outputs

$< 0.3ml/kg$). We used methodology developed in Chapter 4 to construct a second order polynomial model with learning mechanisms for the observation and evolution variance series. Prior elicitation and parameter estimation for the DLM were discussed. We then showed the results obtained by using the DLM and found that using dynamic models to model the urine output time series can identify a group of patients more likely to suffer adverse outcomes, and that the model outperforms the KDIGO stage one urine output criterion and Ralib's criterion.

In Chapter 6 we discussed model performance and causes of model breakdown. We showed how to use interventions to prevent model breakdown by using external, anticipated information. We then introduced methods for correcting an underperforming model retrospectively. We described an automatic model monitoring and diagnostics scheme that uses local Bayes' factors and alternative models to identify and correct model deterioration. We then incorporated an automatic model monitoring and diagnostics scheme to the DLM constructed in Chapter 5. This model was compared to the model used in Chapter 5.

In Chapter 7 we introduced more powerful models, namely multi-process dynamic models. These models illustrated how model mixing can be used to model noisy time series subject to outliers and parametric changes. We showed, in detail, how to use a multi-process dynamic model for a second-order polynomial structure. We then illustrated how to distinguish between types of change using retrospective probabilities. We concluded this chapter by showing how multi-process models can be used to model the urine output time series. We also discussed limitations of using multi-process dynamic models and proposed how one could use multi-process models with model monitoring to combat these limitations.

In Chapter 8 we described how clinicians can use a shiny application to easily monitor the functioning of a patient's kidneys. In this chapter we presented an application that was designed to meet the needs of clinicians at the University Hospital of South Manchester. The application displayed a patient's medical history, alongside the joint probability of suffering severe oliguria at the current, chosen time. The application displayed how long a high risk patient had been considered at risk by the DLM, making it easy for clinicians to monitor the success of minor interventions. In addition, the application showed a plot of the urine output time series with point forecasts and corresponding 95% prediction intervals. Furthermore, the

application showed a summary table and additional plots of interest to clinicians when monitoring a patient's kidneys.

In summary, this thesis described in detail how to construct a powerful dynamic model for modelling noisy time series. The methods provided can be used in a wide range of applications from finance and econometrics, to biological series used in clinical monitoring. Similarly to Smith and West (1983), we showed that a particular type of dynamic model can be used to model and monitor kidney injury and can be used as a useful classification for identifying patients at risk of adverse outcomes. Furthermore, the DLM with model monitoring, discussed in Section 6.7, is built into a shiny application. This shiny application can be used with ease by clinicians, and requires no statistical background nor coding experience. The shiny application can be used to monitor kidney function over time for patients and signals warnings at times of high risk of suffering severe oliguria.

Appendix A

Proof of Equation (4.4.3)

The state posterior for the general univariate dynamic model is given by

$$\begin{aligned} p(\boldsymbol{\theta}_t \mid D_t) &\propto P(Y_t = y_t \mid \boldsymbol{\theta}_t, V_t) p(\boldsymbol{\theta}_t \mid D_{t-1}) \\ &\propto \exp\{-0.5 V_t^{-1}(y_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t)^2\} \times \exp\{-0.5(\boldsymbol{\theta}_t - \boldsymbol{a}_t)'\boldsymbol{R}_t^{-1}(\boldsymbol{\theta}_t - \boldsymbol{a}_t)\} \\ &\propto \exp\{-0.5(\boldsymbol{\theta}_t - \boldsymbol{m}_t)'\boldsymbol{C}_t^{-1}(\boldsymbol{\theta}_t - \boldsymbol{m}_t)\}, \end{aligned} \tag{A.0.1}$$

where the moments are defined as

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t e_t, \\ \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t, \\ \boldsymbol{A}_t &= \boldsymbol{R}_t \boldsymbol{F}_t / Q_t, \\ e_t &= y_t - f_t. \end{aligned} \tag{A.0.2}$$

Proof: Taking natural logarithms, multiplying by -2, and expanding we find

$$\begin{aligned} -2\ln[p(\boldsymbol{\theta}_t \mid D_t)] &= (\boldsymbol{\theta}_t - \boldsymbol{a}_t)'\boldsymbol{R}_t^{-1}(\boldsymbol{\theta}_t - \boldsymbol{a}_t) \\ &\quad + (Y_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t)' V_t^{-1}(Y_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t) + \textit{constant} \\ &= (\boldsymbol{\theta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\theta}_t - \boldsymbol{\theta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{a}_t - \boldsymbol{a}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\theta}_t + \boldsymbol{a}_t'\boldsymbol{R}_t^{-1}\boldsymbol{a}_t) \\ &\quad + (Y_t' V_t^{-1} Y_t - Y_t' V_t^{-1}\boldsymbol{F}_t'\boldsymbol{\theta}_t - \boldsymbol{\theta}_t'\boldsymbol{F}_t V_t^{-1} Y_t + \boldsymbol{\theta}_t'\boldsymbol{F}_t V_t^{-1}\boldsymbol{F}_t'\boldsymbol{\theta}_t) + \textit{constant}. \end{aligned} \tag{A.0.3}$$

The constant not involving $\boldsymbol{\theta}_t$. This is a quadratic function of $\boldsymbol{\theta}_t$ and can be rearranged with a new constant (once again the constant not involving $\boldsymbol{\theta}_t$) and expressed

as

$$\boldsymbol{\theta}_t'(\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t')\boldsymbol{\theta}_t - 2\boldsymbol{\theta}_t'(\boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_t V_t^{-1} Y_t) + constant. \tag{A.0.4}$$

Above we were able to combine terms by noting that a (1×1) matrix is symmetric, that is, $c' = c$ where c is a constant. Namely,

$$\begin{aligned} \boldsymbol{\theta}_t' \boldsymbol{R}_t^{-1} \boldsymbol{a}_t &= \boldsymbol{a}_t' \boldsymbol{R}_t^{-1} \boldsymbol{\theta}_t \\ \boldsymbol{\theta}_t' \boldsymbol{F}_t V_t^{-1} Y_t &= Y_t' V_t^{-1} \boldsymbol{F}_t' \boldsymbol{\theta}_t. \end{aligned} \tag{A.0.5}$$

Now with $\boldsymbol{C}_t$ as defined above,

$$(\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t')\boldsymbol{C}_t = \boldsymbol{I}_n, \tag{A.0.6}$$

the $(n \times n)$ identity matrix, hence

$$\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t'. \tag{A.0.7}$$

Further, with $\boldsymbol{m}_t$ as defined above

$$\boldsymbol{C}_t^{-1} \boldsymbol{m}_t = \boldsymbol{R}_t^{-1} \boldsymbol{a}_t + \boldsymbol{F}_t V_t^{-1} Y_t. \tag{A.0.8}$$

See Appendix B for proof of results (A.0.7) and (A.0.8). Therefore,

$$-2\ln[p(\boldsymbol{\theta}_t \mid D_t)] = \boldsymbol{\theta}_t' \boldsymbol{C}_t^{-1} \boldsymbol{\theta}_t - 2\boldsymbol{\theta}_t' \boldsymbol{C}_t^{-1} \boldsymbol{m}_t + constant. \tag{A.0.9}$$

Once again we are able to combine terms by noting that a (1×1) matrix is symmetric, using $\boldsymbol{m}_t' \boldsymbol{C}_t^{-1} \boldsymbol{\theta}_t = \boldsymbol{\theta}_t' \boldsymbol{C}_t^{-1} \boldsymbol{m}_t$. Finally, noting that $\boldsymbol{m}_t' \boldsymbol{C}_t^{-1} \boldsymbol{m}_t$ is a constant not involving $\boldsymbol{\theta}_t$, we obtain

$$-2\ln[p(\boldsymbol{\theta}_t \mid D_t)] = (\boldsymbol{\theta}_t - \boldsymbol{m}_t)' \boldsymbol{C}_t^{-1} (\boldsymbol{\theta}_t - \boldsymbol{m}_t) + constant. \tag{A.0.10}$$

Dividing by -2 and exponentiating gives the final result:

$$p(\boldsymbol{\theta}_t \mid D_t) \propto \exp\{-0.5(\boldsymbol{\theta}_t - \boldsymbol{m}_t)' \boldsymbol{C}_t^{-1} (\boldsymbol{\theta}_t - \boldsymbol{m}_t)\}. \tag{A.0.11}$$

□

Appendix B

Proof of Equations (A.0.7) and (A.0.8)

Given $(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{a}_t, \boldsymbol{R}_t)$, $(Y_t \mid D_{t-1}) \sim \mathrm{N}(\boldsymbol{F}_t'\boldsymbol{\theta}_t, V_t)$, and $(\boldsymbol{\theta}_t \mid D_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t)$, we have

$$\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t', \tag{B.0.1}$$

and

$$\boldsymbol{C}_t^{-1}\boldsymbol{m}_t = \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_t V_t^{-1} Y_t, \tag{B.0.2}$$

Proof:

$$\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t', \tag{B.0.3}$$

and using $\boldsymbol{C}_t$ as stated in Section 4.4, $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t'Q_t$ we find:

$$\begin{aligned} \boldsymbol{I}_n = \boldsymbol{C}_t^{-1}\boldsymbol{C}_t &= (\boldsymbol{R}_t^{-1} + \boldsymbol{F}_t V_t^{-1} \boldsymbol{F}_t')\boldsymbol{C}_t \\ &= \boldsymbol{R}_t^{-1}(\boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t'Q_t) + \boldsymbol{F}_t V_t^{-1}\boldsymbol{F}_t'(\boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t'Q_t) \\ &= \boldsymbol{I}_n - \boldsymbol{R}_t^{-1}\boldsymbol{A}_t\boldsymbol{A}_t'Q_t + \boldsymbol{F}_t V_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t - \boldsymbol{F}_t V_t^{-1}\boldsymbol{F}_t'\boldsymbol{A}_t\boldsymbol{A}_t'Q_t. \end{aligned} \tag{B.0.4}$$

Cancelling common terms and using $\boldsymbol{A}_t$ as stated in Section 4.4, $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t/Q_t$ we find:

$$\begin{aligned} \boldsymbol{R}_t^{-1}\boldsymbol{A}_t\boldsymbol{A}_t'Q_t - \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{A}_t\boldsymbol{A}_t'Q_t = \boldsymbol{0} \\ \implies \boldsymbol{F}_t\boldsymbol{A}_t' - \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{A}_t' = \boldsymbol{0}. \end{aligned} \tag{B.0.5}$$

Using the definition of $Q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + V_t$, defined in Section 4.2.1, we can rearrange and multiply by V_t^{-1} to find $V_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t = Q_tV_t^{-1} - 1$. Using this result and cancelling common terms we find:

$$\begin{aligned} \boldsymbol{F}_t\boldsymbol{A}_t' - \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{A}_t' = 0 \\ \implies \boldsymbol{F}_t\boldsymbol{A}_t' - \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t + \boldsymbol{F}_t(Q_tV_t^{-1} - 1)\boldsymbol{A}_t' = 0 \\ \implies V_t^{-1}\boldsymbol{F}_t(\boldsymbol{F}_t'\boldsymbol{R}_t - Q_t\boldsymbol{A}_t') = 0. \end{aligned} \tag{B.0.6}$$

Finally noting that, from the definition in Section 4.4, $Q_t\boldsymbol{A}_t' = \boldsymbol{F}_t'\boldsymbol{R}_t$ (since $\boldsymbol{R}_t$ is symmetric $\boldsymbol{R}_t' = \boldsymbol{R}_t$) we see that the above result is, in fact, zero. Now using the definitions of $\boldsymbol{m}_t$, $\boldsymbol{C}_t^{-1}$ and Y_t in Section 4.4, $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_te_t$, $\boldsymbol{C}_t^{-1} = \boldsymbol{R}_t^{-1} + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'$, $e_t = Y_t - f_t$ we find:

$$\begin{aligned} \boldsymbol{C}_t^{-1}\boldsymbol{m}_t &= (\boldsymbol{R}_t^{-1} + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t')(\boldsymbol{a}_t + \boldsymbol{A}_te_t) \\ &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t - \boldsymbol{F}_tV_t^{-1}e_t + \boldsymbol{R}_t^{-1}\boldsymbol{A}_te_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{A}_te_t. \end{aligned} \tag{B.0.7}$$

Using, $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_tQ_t^{-1}$ we find:

$$\begin{aligned} \boldsymbol{C}_t^{-1}\boldsymbol{m}_t &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t - \boldsymbol{F}_tV_t^{-1}e_t + \boldsymbol{R}_t^{-1}\boldsymbol{A}_te_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{A}_te_t \\ &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t - \boldsymbol{F}_tV_t^{-1}e_t + \boldsymbol{F}_tQ_t^{-1}e_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_tQ_t^{-1}e_t. \end{aligned} \tag{B.0.8}$$

Finally using the definition of Q_t, $\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t = Q_t - V_t$ we find:

$$\begin{aligned} \boldsymbol{C}_t^{-1}\boldsymbol{m}_t &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t - \boldsymbol{F}_tV_t^{-1}e_t + \boldsymbol{F}_tQ_t^{-1}e_t + \boldsymbol{F}_tV_t^{-1}\boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_tQ_t^{-1}e_t \\ &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t - \boldsymbol{F}_tV_t^{-1}e_t + \boldsymbol{F}_tQ_t^{-1}e_t + \boldsymbol{F}_tV_t^{-1}(Q_t - V_t)Q_t^{-1}e_t \\ &= \boldsymbol{R}_t^{-1}\boldsymbol{a}_t + \boldsymbol{F}_tV_t^{-1}Y_t. \end{aligned} \tag{B.0.9}$$

□

Appendix C

Proof of Equation (4.9.17)

The joint distribution, conditional on ϕ_t, between an observation and the state at time t, given information up to time $t-1$, D_{t-1}, is

$$\begin{pmatrix} Y_t \\ \boldsymbol{\theta}_t \end{pmatrix} \mid D_{t-1}, \phi_t \sim \mathrm{N}\left[\begin{pmatrix} f_t \\ \boldsymbol{a}_t \end{pmatrix}, \begin{pmatrix} \phi_t^{-1} Q_t^* & \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} \\ \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} & \boldsymbol{R}_t^* \phi_t^{-1} \end{pmatrix} \right]. \tag{C.0.1}$$

Using properties of the multivariate normal distribution, conditioning the state on the observed value yields a normal distribution which is given by

$$(\boldsymbol{\theta}_t \mid D_{t-1}, Y_t = y_t, \phi_t) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{C}_t^* \phi_t^{-1}), \tag{C.0.2}$$

where the moments are updated from their prior values with the scale conditioning made explicit and are given by

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} (\phi_t^{-1} Q_t^*)^{-1} (y_t - f_t) \\ &= \boldsymbol{a}_t + \boldsymbol{R}_t^* \boldsymbol{F}_t e_t / Q_t^*, \end{aligned} \tag{C.0.3}$$

$$\begin{aligned} \boldsymbol{C}_t^* \phi_t^{-1} &= \boldsymbol{R}_t^* \phi_t^{-1} - \boldsymbol{R}_t^* \boldsymbol{F}_t \phi_t^{-1} (\phi_t^{-1} Q_t^*)^{-1} \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} \\ &= \boldsymbol{R}_t^* \phi_t^{-1} - \boldsymbol{R}_t^* \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{R}_t^* \phi_t^{-1} / Q_t^*. \end{aligned} \tag{C.0.4}$$

Proof: Suppose that a vector random variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)'$ has a multivariate normal distribution with pdf given by

$$\pi_{\boldsymbol{X}}(\boldsymbol{x}) = -\left(\frac{1}{2\pi}\right)^{k/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{C.0.5}$$

where $\boldsymbol{\Sigma}$ is the $(k \times k)$ covariance matrix, and $\boldsymbol{\mu}$ is the $(k \times 1)$ mean vector. First partition $\boldsymbol{X}$ into two components $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of dimensions d and $(k-d)$ respectively, that is, $\boldsymbol{X}' = (\boldsymbol{X_1'}, \boldsymbol{X}_2')$. Writing

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{C.0.6}$$

where $\boldsymbol{\Sigma}_{11}$ is $(d \times d)$, $\boldsymbol{\Sigma}_{22}$ is $(k-d) \times (k-d)$, and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$. Now, define a third variable $\boldsymbol{z} = \boldsymbol{x_2} + \boldsymbol{A}\boldsymbol{x}_1$ where $\boldsymbol{A} = -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$ is called the regression matrix of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$. Since all conditional distributions of a multivariate normal distribution are normal, all that is left to do is calculate the mean vector and covariance matrix. The covariance between $\boldsymbol{z}$ and $\boldsymbol{x}_1$ is given by

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{z}, \boldsymbol{x}_1) &= \mathrm{Cov}(\boldsymbol{x_2} + A\boldsymbol{x}_1, \boldsymbol{x}_1) \\ &= \mathrm{Cov}(\boldsymbol{x}_2, \boldsymbol{x}_1) + A\mathrm{Var}(\boldsymbol{x}_1, \boldsymbol{x}_1) \\ &= \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11} = \boldsymbol{0}. \end{aligned} \tag{C.0.7}$$

Therefore $\boldsymbol{z}$ and $\boldsymbol{x}_1$ are uncorrelated, and since they are jointly normal they are independent (See Appendix G, Correlations and Independence). Also $\mathrm{E}[\boldsymbol{z}] = \boldsymbol{\mu}_2 + \boldsymbol{A}\boldsymbol{\mu}_1$. Hence

$$\begin{aligned} \mathrm{E}[\boldsymbol{x}_2 \mid \boldsymbol{x}_1] &= \mathrm{E}[\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}_1 \mid \boldsymbol{x}_1] \\ &= \mathrm{E}[\boldsymbol{z} \mid \boldsymbol{x}_1] - \boldsymbol{A}\mathrm{E}[\boldsymbol{x}_1 \mid \boldsymbol{x}_1] \\ &= \mathrm{E}[\boldsymbol{z}] - \boldsymbol{A}\boldsymbol{x}_1 \\ &= \boldsymbol{\mu}_2 + \boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{x}_1) \\ &= \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1). \end{aligned} \tag{C.0.8}$$

Finally, we need to calculate the covariance matrix of the conditional distribution

$$\begin{aligned} \mathrm{Var}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) &= \mathrm{Var}(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}_1 \mid \boldsymbol{x}_1) \\ &= \mathrm{Var}(\boldsymbol{z} \mid \boldsymbol{x}_1) + \boldsymbol{A}\mathrm{Var}(\boldsymbol{x}_1 \mid \boldsymbol{x}_1)\boldsymbol{A}' - \boldsymbol{A}\mathrm{Cov}(\boldsymbol{x}_1, \boldsymbol{z}) - \mathrm{Cov}(\boldsymbol{z}, \boldsymbol{x}_1)\boldsymbol{A}' \\ &= \mathrm{Var}(\boldsymbol{z} \mid \boldsymbol{x}_1) = \mathrm{Var}(\boldsymbol{z}). \end{aligned} \tag{C.0.9}$$

Evaluating the variance of $\boldsymbol{z}$ gives the expression

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{z}) &= \mathrm{Var}(\boldsymbol{x_2} + \boldsymbol{A}\boldsymbol{x_1}) \\
&= \mathrm{Var}(\boldsymbol{x_2}) + \boldsymbol{A}\mathrm{Var}(\boldsymbol{x_1})\boldsymbol{A}' + \boldsymbol{A}\mathrm{Cov}(\boldsymbol{x_1}, \boldsymbol{x_2}) + \mathrm{Cov}(\boldsymbol{x_2}, \boldsymbol{x_1})\boldsymbol{A}' \\
&= \boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \qquad \text{(C.0.10)} \\
&= \boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - 2\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\
&= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.
\end{aligned}$$

Therefore the conditional distribution for $(\boldsymbol{X_1} \mid \boldsymbol{X_2} = \boldsymbol{x_2})$ is given by

$$(\boldsymbol{X}_2 \mid \boldsymbol{X}_1 = \boldsymbol{x}_1) \sim \mathrm{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}). \qquad \text{(C.0.11)}$$

Comparing terms to Equation (C.0.1) we see that $x_1 = y_t$, $\boldsymbol{x}_2 = \boldsymbol{\theta}_t$, $\mu_1 = f_t$, $\boldsymbol{\mu_2} = \boldsymbol{a}_t$, $\boldsymbol{\Sigma}_{11} = \phi_t^{-1}Q_t^*$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{F}_t'\boldsymbol{R}_t^*\phi_t^{-1}$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{R}_t^*\boldsymbol{F}_t\phi_t^{-1}$, and $\boldsymbol{\Sigma_{22}} = \boldsymbol{R}_t^*\phi_t^{-1}$ and hence we recover the conditional distribution with the moments defined as above.

□

Appendix D

Sensitivity Analysis

In this Appendix we show a sensitivity analysis for the discount factors, δ_μ, δ_β and δ_V, for the second order polynomial dynamic model discussed in Section 5.2.1. The criteria for estimating the initial parameters in this study is to minimise the number of type two errors. The number type one errors, the MAD, and the MSE are also considered (see Section 5.2.2 for definitions of the diagnostic measures).

Below are diagnostic tables for varying values of δ_μ fixing $\delta_\beta = 0.9$ and $\delta_V = 1$, meaning that the observational variance has no information discounting. The initial prior for the unknown observational variance is $V_0 = 0.1$ and so we start with $S_0 = 0.1$:

δ_μ	Type 1	Type 2	MAD	MSE
0.70	0.1065855	0.002914519	635.2311	1202.452
0.75	0.108088	0.002770723	634.5539	1194.602
0.80	0.1046256	0.002662096	634.1929	1187.282
0.85	0.0996045	0.003074558	634.1619	1180.404
0.90	0.1001783	0.003048288	634.4256	1173.759

Table D.1: Diagnostics Table for varying values of δ_μ. Type 1 represents the proportion of type one errors; type 2 represents the proportion of type two errors; MAD is the total mean absolute deviation for all patients in the development set; and MSE is the total mean square error for all patients in the development set, see Section 5.2.2 for definitions of diagnostic measures

The proportion of type one errors is given in the column named "Type 1", and the

proportion of type two errors is given in the column named "Type 2" (see Section 5.2.2 for definitions). Minimising the proportion of type two errors is our main diagnostic tool for the model to be clinically useful. The proportion of type one errors, the MAD, and the MSE will be considered as secondary measures. We can see from Table D.1 that the different values for δ_μ do not differ very much in any of the criteria considered. We choose $\delta_\mu = 0.8$ since this value has the smallest number of type two errors.

Now varying δ_β and fixing $\delta_\mu = 0.8$, and $\delta_V = 1$:

δ_β	Type 1	Type 2	MAD	MSE
0.70	0.1755006	0.003394694	892.7208	26560.71
0.75	0.1484301	0.00325695	788.7378	4610.318
0.80	0.1310137	0.003265415	718.8231	2229.61
0.85	0.1116297	0.003231378	669.3413	1512.267
0.90	0.1046256	0.002662096	634.1929	1187.282

Table D.2: Diagnostics Table for varying values of δ_β

From Table D.2 we can see that as δ_β decreases the MAD and MSE both increase. In addition, and most importantly, $\delta_\beta = 0.9$ has the smallest number of type two errors (and the smallest number of type one errors). Thus we choose $\delta_\beta = 0.9$.

Now varying the observational discount factor δ_V and fixing $\delta_\mu = 0.8$, and $\delta_\beta = 0.9$:

δ_V	Type 1	Type 2	MAD	MSE
0.80	0.0804836	0.002592593	634.1929	1187.282
0.85	0.0791978	0.002597999	634.1929	1187.282
0.90	0.0772865	0.00287948	634.1929	1187.282
0.95	0.0835598	0.002499021	634.1929	1187.282
1.00	0.1046256	0.002662096	634.1929	1187.282

Table D.3: Diagnostics Table for varying values of δ_V

From Table D.3, we see that, as expected, the point forecast predictions are the same (the MAD and MSE are unchanged) but since the prediction intervals are changing, due to the discount strategy applied to the observational variance, the proportion of type one and type two errors has changed. Minimising the proportion of type two errors, leads us to choose an observational variance discount factor of $\delta_V = 0.95$.

D.1 Epsilon

In Section 5.2.1 we transformed the urine output data. We added a small constant ε before taking logarithms. In this section we perform a sensitivity analysis on ε. We use the model discussed in Section 5.2.1 and use $\delta_\mu = 0.8$, $\delta_\beta = 0.9$, and $\delta_V = 0.95$.

ε	Type 1	Type 2	MAD	MSE
0.01	0.1972809	0.005213838	663.5979	1246.137
0.05	0.1636685	0.00584558	648.4412	1212.771
0.1	0.0835598	0.002499021	634.1929	1187.282
0.2	0.1045625	0.01141064	677.9085	1274.49
0.3	0.08755216	0.01558025	686.3824	1298.138

Table D.4: Diagnostics Table for varying values of ε

From Table D.4 we can see that $\varepsilon = 0.1$ has the lowest number of type one and type two errors. In addition, $\varepsilon = 0.1$ has the smallest MAD and the smallest MSE. Consequently, we choose $\varepsilon = 0.1$ and use the transformation $Y \longmapsto \log(Y + 0.1)$.

D.2 Other Transformations

In Section 2.3 we discussed how the urine output data is nonlinear. In order to linearise the data we used a logarithmic transformation. In this section we discuss other transformations that could also be used to address this issue. Power transformations have been used in many applications [15] (Chapter 4) , [35] and will be considered here.

Transformation	Type 1	Type 2	MAD	MSE
$\log(Y + 0.1)$	0.0835598	0.002499021	634.1929	1187.282
square root	0.1871771	0.003369517	720.6166	1304.78
3/4 power	0.1282401	0.007016891	678.4285	1206.623

Table D.5: Diagnostics Table for different transformations

Table D.5 shows the results of a sensitivity analysis performed using three different transformations of the urine output data. We considered $\log(Y + 0.1)$, a square root transformation, and a 3/4 power transformation. From Table D.5 we can see that the logarithmic transformation $\log(Y + 0.1)$ has a lower MAD and MSE compared to the power transformations considered. In addition, the logarithmic transformation

has the lowest number of type one and type two errors. As a result, we use the logarithmic transformation $\log(Y + 0.1)$ in this thesis.

Appendix E

Proof of Results in Example 7.1.3.1

Suppose that $\boldsymbol{\theta}$ has true distribution with mean vector, $\mathrm{E}[\boldsymbol{\theta}]$, and variance matrix, $\mathrm{Var}(\boldsymbol{\theta})$, and that the approximating distribution is multivariate normal, $\mathrm{N}(\boldsymbol{m}, \boldsymbol{C})$. The approximate and true densities are given by

$$\begin{aligned} p(\boldsymbol{\theta}) &= |2\pi \mathrm{Var}(\boldsymbol{\theta})|^{1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}])' \mathrm{Var}(\boldsymbol{\theta})^{-1} (\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}]) \right\} \\ p^*(\boldsymbol{\theta}) &= |2\pi \boldsymbol{C}|^{1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m})' \boldsymbol{C}^{-1} (\boldsymbol{\theta} - \boldsymbol{m}]) \right\}. \end{aligned} \tag{E.0.1}$$

Using Equation 7.1.18 we find

$$\begin{aligned} K(p^*) &= \mathrm{E}[\log(p(\boldsymbol{\theta}))] - \mathrm{E}[\log(p^*(\boldsymbol{\theta}))] \\ &= \mathrm{E}\left[-\frac{1}{2}\log(|2\pi \mathrm{Var}(\boldsymbol{\theta})|) - \frac{1}{2}(\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}])' \mathrm{Var}(\boldsymbol{\theta})^{-1} (\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}]) \right] \\ &\quad - \mathrm{E}\left[-\frac{1}{2}\log(|2\pi \boldsymbol{C}|) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m})' \boldsymbol{C}^{-1} (\boldsymbol{\theta} - \boldsymbol{m}) \right]. \end{aligned} \tag{E.0.2}$$

Multiplying by -2 and rearranging

$$\begin{aligned} 2K(p^*) &= \log\left(\frac{|\boldsymbol{C}|}{|\text{Var}(\boldsymbol{\theta})|}\right) - \text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\text{Var}(\boldsymbol{\theta})^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] \\ &\quad + \text{E}[(\boldsymbol{\theta} - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{m})] \\ &= c + \log(|\boldsymbol{C}|) + \text{E}[(\boldsymbol{\theta} - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{m})], \end{aligned} \tag{E.0.3}$$

where c does not depend on $\boldsymbol{m}$ or $\boldsymbol{C}$. Furthermore,

$$\begin{aligned} 2K(p^*) &= c + \log(|\boldsymbol{C}|) + \text{E}[(\boldsymbol{\theta} - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{m})], \\ &= c + \log(|\boldsymbol{C}|) + \text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}] + \text{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}] + \text{E}[\boldsymbol{\theta}] - \boldsymbol{m})] \\ &= c + \log(|\boldsymbol{C}|) + \text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] + \text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\boldsymbol{C}^{-1}(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})] \\ &\quad + \text{E}[(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] + \text{E}[(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})] \\ &= c + \log(|\boldsymbol{C}|) + \text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] + \text{E}[(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})]. \end{aligned} \tag{E.0.4}$$

In addition,

$$\begin{aligned} &\text{E}[(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] = \text{E}[\text{Tr}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] = \\ &\text{Tr}[\boldsymbol{C}^{-1}\text{E}(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])'(\boldsymbol{\theta} - \text{E}[\boldsymbol{\theta}])] = \text{Tr}[\boldsymbol{C}^{-1}\text{Var}(\boldsymbol{\theta})]. \end{aligned} \tag{E.0.5}$$

Putting all of this together we have

$$2K(p^*) = c + \log(|\boldsymbol{C}|) + \text{Tr}[\boldsymbol{C}^{-1}\text{Var}(\boldsymbol{\theta})] + \text{E}[(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\text{E}[\boldsymbol{\theta}] - \boldsymbol{m})]. \tag{E.0.6}$$

Appendix F

Proof of Results in Example 7.1.3.2

Using Equation (7.1.18) and Bayes' theorem we have that

$$\begin{aligned} K(p^*) &= \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p(\boldsymbol{\theta},\boldsymbol{\phi})]\} - \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta},\boldsymbol{\phi})]\} \\ &= constant - \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta} \mid \phi)] + \log[p^*(\phi)]\} \\ &= constant - \mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\} - \mathrm{E}_{\phi}\{\log[p^*(\phi)]\}. \end{aligned} \tag{F.0.1}$$

Firstly, consider the term $-\mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}$. This term can be written as

$$\mathrm{E}_{\boldsymbol{\theta},\phi}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\} = \mathrm{E}_{\phi}[\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}]. \tag{F.0.2}$$

From (E.0.6) we know $\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta})]\}$ and we can simply obtain $\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}$ by substituting $\boldsymbol{C}/S\phi$ for $\boldsymbol{C}$. Thus we have

$$\begin{aligned} -2\mathrm{E}_{\phi}\{\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}\} = \mathrm{E}_{\phi}\{c + \log(|\boldsymbol{C}/(S\phi)|) + \mathrm{Tr}[\{\boldsymbol{C}/(S\phi)\}^{-1}\mathrm{Var}(\boldsymbol{\theta})] \\ + \mathrm{E}_{\theta}[(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\{\boldsymbol{C}/(S\phi)\}^{-1}(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})]\}. \end{aligned} \tag{F.0.3}$$

By noting that $|\phi \boldsymbol{C}| = \phi^q |\boldsymbol{C}|$ where $\boldsymbol{C}$ is a $(q \times q)$ matrix and using logarithmic rules, we have that

$$\begin{aligned} -2\mathrm{E}_\phi\{\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}\} = c \\ + \log(|S^{-1}\boldsymbol{C}|) - q\mathrm{E}[\log(\phi)] + \mathrm{E}_\phi[S\phi \mathrm{Tr}(\boldsymbol{C}^{-1}\mathrm{Var}(\boldsymbol{\theta}))] \\ + \mathrm{E}_\phi[S\phi(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})]. \end{aligned} \tag{F.0.4}$$

Consider the final two terms in (F.0.4).

$$\begin{aligned} \mathrm{E}_\phi[S\phi \mathrm{Tr}(\boldsymbol{C}^{-1}\mathrm{Var}(\boldsymbol{\theta}))] &= \mathrm{E}_\phi[S\phi \mathrm{Tr}(\boldsymbol{C}^{-1}\mathrm{Var}(\boldsymbol{\theta}))] \\ &= S\sum_{j=1}^{k} S^{-1}(j)\mathrm{Tr}(\boldsymbol{C}^{-1}\boldsymbol{C}(j))p(j). \end{aligned} \tag{F.0.5}$$

The last step in (F.0.5) can be seen by considering a single component of the mixture $p^*(\boldsymbol{\theta}|\phi)$. The estimate for ϕ under model j is $S(j)^{-1}$. Also, the marginal multivariate T distribution in model j is given by $\boldsymbol{\theta} \sim \mathrm{T}_{n(j)}(\boldsymbol{m}(j), \boldsymbol{C}(j))$. That is, the true variance under model j is $\boldsymbol{C}(j)$. Similarly, the true mean under model j is $\boldsymbol{m}(j)$ and hence

$$\begin{aligned} \mathrm{E}_\phi[S\phi(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})'\boldsymbol{C}^{-1}(\mathrm{E}[\boldsymbol{\theta}] - \boldsymbol{m})] = \\ S\sum_{j=1}^{k} S^{-1}(j)(\boldsymbol{m}(j) - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{m}(j) - \boldsymbol{m})p(j). \end{aligned} \tag{F.0.6}$$

Hence

$$\begin{aligned} \mathrm{E}_\phi\{\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}\} = c + \log(|S^{-1}\boldsymbol{C}|) - q\mathrm{E}[\log(\phi)] + \\ S\sum_{j=1}^{k} S^{-1}(j)\{\mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}(j)) + (\boldsymbol{m}(j) - \boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{m}(j) - \boldsymbol{m})\}p(j). \end{aligned} \tag{F.0.7}$$

Marginally we have

$$p(\phi) = \frac{(d/2)^{n/2}}{\Gamma(n/2)}\phi^{n/2-1}\exp\{-\phi d/2\}. \tag{F.0.8}$$

Taking logarithms of both sides gives

$$\log[p(\phi)] = (n/2)\log(d/2) - \log(\Gamma(n/2)) + (n/2-1)\log(\phi) - \phi d/2. \tag{F.0.9}$$

Taking expectations with respect to ϕ and multiplying by -2 gives

$$-2\mathrm{E}_\phi\{\log[p(\phi)]\} = -n\log(d/2) + 2\log(\Gamma(n/2)) - (n-2)\mathrm{E}_\phi[\log(\phi)] + d\mathrm{E}[\phi]. \tag{F.0.10}$$

Thus we can write

$$\begin{aligned}2K(p^*) &= constant - 2\mathrm{E}_\phi\{\mathrm{E}_{\boldsymbol{\theta}}\{\log[p^*(\boldsymbol{\theta} \mid \phi)]\}\} - 2\mathrm{E}_\phi\{\log[p(\phi)]\} \\ &= constant \\ &- n\log(d/2) + 2\log(\Gamma(n/2)) - (n+q-2)\mathrm{E}_\phi[\log(\phi)] + d\mathrm{E}[\phi] + \log(|S^{-1}\boldsymbol{C}|) \\ &+ S\sum_{j=1}^{k} S^{-1}(j)\{\mathrm{Tr}(\boldsymbol{C}\boldsymbol{C}(j)) + (\boldsymbol{m}(j)-\boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{m}(j)-\boldsymbol{m})\}p(j).\end{aligned} \tag{F.0.11}$$

1. Differentiating with respect to $\boldsymbol{m}$ we obtain (using Equation (G.0.7))

$$2\frac{\partial K(P^*)}{\partial \boldsymbol{m}} = 2S\sum_{j=1}^{k} S(j)^{-1}\boldsymbol{C}^{-1}(\boldsymbol{m}(j)-\boldsymbol{m})p(j). \tag{F.0.12}$$

   Setting equal to zero we have

$$\sum_{j=1}^{k} S(j)^{-1}p(j)\boldsymbol{C}^{-1}\boldsymbol{m} = \sum_{j=1}^{k} S(j)^{-1}\boldsymbol{C}^{-1}\boldsymbol{m}(j)p(j). \tag{F.0.13}$$

   Multiplying both sides on the left by $\boldsymbol{C}$ and rearranging gives

$$\boldsymbol{m} = \left[\sum_{j=1}^{k} S(j)^{-1}p(j)\right]^{-1}\sum_{j=1}^{k} S(j)^{-1}\boldsymbol{m}(j)p(j). \tag{F.0.14}$$

2. Differentiating with respect to $\boldsymbol{C}$ we obtain (using Equations (G.0.8), (G.0.9), and (G.0.10))

$$\begin{aligned}2\frac{\partial K(P^*)}{\partial \boldsymbol{C}} &= \boldsymbol{C}^{-1} \\ &+ S\sum_{j=1}^{k} S(j)^{-1}\{-\boldsymbol{C}^{-1}\boldsymbol{C}(j)\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}(\boldsymbol{m}(j)-\boldsymbol{m}) \quad (\boldsymbol{m}(j)-\boldsymbol{m})'\boldsymbol{C}^{-1}\}p(j).\end{aligned} \tag{F.0.15}$$

   Setting equal to zero and multiplying on the left and right by $\boldsymbol{C}$ we have

$$\boldsymbol{C} - S\sum_{j=1}^{k} S(j)^{-1}\{\boldsymbol{C}(j) - (\boldsymbol{m}(j)-\boldsymbol{m})(\boldsymbol{m}(j)-\boldsymbol{m})'\}p(j) = \boldsymbol{0}. \tag{F.0.16}$$

Rearranging for $\boldsymbol{C}$ we have

$$\boldsymbol{C} = S\sum_{j=1}^{k} S(j)^{-1}\{\boldsymbol{C}(j) - (\boldsymbol{m}(j) - \boldsymbol{m})(\boldsymbol{m}(j) - \boldsymbol{m})'\}p(j). \tag{F.0.17}$$

3. Differentiating with respect to d we have

$$2\frac{\partial K(P^*)}{\partial d} = -\frac{n}{d} + \mathrm{E}_\phi[\phi]. \tag{F.0.18}$$

Setting equal to zero and rearranging we obtain

$$\mathrm{E}_\phi[\phi] = \frac{n}{d} = S^{-1} = \sum_{j=1}^{k} S(j)^{-1}p(j), \tag{F.0.19}$$

where $S(j)^{-1}$ is the estimate of ϕ^{-1} in model j. Using this result we can write $\boldsymbol{m}$ follows

$$\begin{aligned}\boldsymbol{m} &= \Big[\sum_{j=1}^{k} S(j)^{-1}p(j)\Big]^{-1}\sum_{j=1}^{k} S(j)^{-1}\boldsymbol{m}(j)p(j)\\ &= \sum_{j=1}^{k} SS(j)^{-1}\boldsymbol{m}(j)p(j)\\ &= \sum_{j=1}^{k} \boldsymbol{m}(j)p^*(j),\end{aligned} \tag{F.0.20}$$

where $p^*(j) = p(j)S/S(j)$. Similarly $\boldsymbol{C}$ can now be written as

$$\boldsymbol{C} = \sum_{j=1}^{k}\{\boldsymbol{C}(j) - (\boldsymbol{m}(j) - \boldsymbol{m})(\boldsymbol{m}(j) - \boldsymbol{m})'\}p^*(j). \tag{F.0.21}$$

4. Differentiating with respect to n we obtain

$$2\frac{\partial K(P^*)}{\partial n} = -\log(d/2) + \gamma(n/2) - \mathrm{E}_\phi[\log(\phi)], \tag{F.0.22}$$

noting that

$$2\frac{\partial \log(\Gamma(n/2))}{\partial n} = \frac{2\partial\Gamma(n/2)/\partial n}{\Gamma(n/2)}, \tag{F.0.23}$$

where, by using the chain rule, we have

$$2\frac{\partial\Gamma(n/2)}{\partial n} = 2\frac{\Gamma(n/2)}{\partial(n/2)}\frac{\partial(n/2)}{n} = \Gamma'(n/2). \tag{F.0.24}$$

Thus we have

$$2\frac{\partial \log(\Gamma(n/2))}{\partial n} = \frac{\Gamma' n/2}{\Gamma(n/2)} = \gamma(n/2). \tag{F.0.25}$$

Setting equal to zero and rearranging gives

$$\mathrm{E}_{\phi}[\log(\phi)] = \gamma(n/2) - \log(d/2). \tag{F.0.26}$$

Appendix G

Statistical Distributions and Useful Results

In this Appendix we provide essential distributions and theory that are used throughout this thesis.

Univariate Normal Distribution

A random quantity X is said to be normally distributed with mean μ and variance σ^2 if it has the probability density function

$$p(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad (\sigma > 0). \tag{G.0.1}$$

We use the notation $X \sim \mathrm{N}(\mu, \sigma^2)$.

Multivariate Normal Distribution

A random p-vector $\boldsymbol{X}$ is said to be jointly normal distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if it has the joint probability density function

$$p(\boldsymbol{X}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{p/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]. \tag{G.0.2}$$

We use the notation $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Correlations and Independence

In general, random variables may be uncorrelated but statistically dependent. But if a random vector has a multivariate normal distribution then any two or more of its components that are uncorrelated are independent.

Gamma Distribution

A positive random quantity ϕ is said to have a gamma distribution with parameters $n > 0$ and $d > 0$ if it has the probability density function

$$p(\phi) = \frac{d^n}{\Gamma(n)} \phi^{n-1} \exp(-\phi d), \tag{G.0.3}$$

where Γ is the gamma function. The mean and variance are given by

$$\begin{aligned} \mathrm{E}[\phi] &= \frac{n}{d}, \\ \mathrm{Var}(\phi) &= \frac{n}{d^2}. \end{aligned} \tag{G.0.4}$$

We use the notation $\phi \sim \mathrm{Ga}(n, d)$.

Multivariate Student-T Distribution

A random p-vector is said to have a joint Student-T distribution on n degrees of freedom with mode $\boldsymbol{m}$ and scale matrix $\boldsymbol{C}$ if it has the joint probability density function

$$p(\boldsymbol{X}) = \frac{\Gamma[(n+p)/2]}{\Gamma(n/2) n^{p/2} \pi^{p/2} |\boldsymbol{C}|^{1/2}} [n + (\boldsymbol{x} - \boldsymbol{m})' \boldsymbol{C}^{-1} (\boldsymbol{x} - \boldsymbol{m})]^{-(n+p)/2}. \tag{G.0.5}$$

We use the notation $\boldsymbol{X} \sim \mathrm{T}_n[\boldsymbol{m}, \boldsymbol{C}]$.

Properties of Covariance Matrices

If we have a p-dimensional random variable $\boldsymbol{X}$ and a q-dimensional random variable $\boldsymbol{Z}$, then the following hold:

1. $\mathrm{Cov}(A\boldsymbol{X} + \boldsymbol{a}, B\boldsymbol{Z} + \boldsymbol{b}) = A\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Z})B'$,
2. $\mathrm{Var}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{Var}(\boldsymbol{X}) + \mathrm{Var}(\boldsymbol{Y}) + \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) + \mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{X}) \quad \text{if } p = q.$

(G.0.6)

Vector Derivatives

Let $\boldsymbol{m}$ be an $n \times 1$ column vector and $\boldsymbol{C}$ be an $n \times n$ symmetric matrix, then [36]

$$\frac{\partial \boldsymbol{m}'\boldsymbol{C}\boldsymbol{m}}{\partial \boldsymbol{m}} = (\boldsymbol{C} + \boldsymbol{C}')\boldsymbol{m} = 2\boldsymbol{C}\boldsymbol{m} \tag{G.0.7}$$

since $\boldsymbol{C} = \boldsymbol{C}'$.

Matrix Derivatives Let $\boldsymbol{C}$ be an $n \times n$ symmetric matrix, then [36]

$$\frac{\partial \log |\boldsymbol{C}|}{\partial \boldsymbol{C}} = \boldsymbol{C}^{-1}. \tag{G.0.8}$$

For a symmetric $n \times n$ matrix $\boldsymbol{C}(j)$ [36]

$$\begin{aligned} \frac{\partial \mathrm{Tr}(\boldsymbol{C}^{-1}\boldsymbol{C}(j))}{\partial \boldsymbol{C}} &= -\boldsymbol{C}'^{-1}\boldsymbol{C}(j)\boldsymbol{C}'^{-1} \\ &= -\boldsymbol{C}^{-1}\boldsymbol{C}(j)\boldsymbol{C}^{-1}, \end{aligned} \tag{G.0.9}$$

since $\boldsymbol{C} = \boldsymbol{C}'$.

For an $n \times 1$ column vector $\boldsymbol{m}$ [37]

$$\begin{aligned} \frac{\partial \boldsymbol{m}'\boldsymbol{C}^{-1}\boldsymbol{m}}{\partial \boldsymbol{C}} &= \boldsymbol{C}'^{-1}\boldsymbol{m}\boldsymbol{m}'\boldsymbol{C}'^{-1} \\ &= \boldsymbol{C}^{-1}\boldsymbol{m}\boldsymbol{m}'\boldsymbol{C}^{-1}, \end{aligned} \tag{G.0.10}$$

since $\boldsymbol{C} = \boldsymbol{C}'$.

Appendix H

Notation

Notation	Description	Section
KDIGO	Kidney Disease Improving Global Outcomes	1
ICU	Intensive Care Unit	1
CICU	Cardiac Intensive Care Unit	1
ARF	Acute Renal Failure	1.1
GRF	Glomerular Filtration Rate	1.1
RIFLE	Risk, Injury, Failure, Loss, End stage renal disease	1.2
AKI	Acute kidney injury	1.2
RRT	Renal replacement therapy	1.2
Severe Oliguria	Six consecutive urine outputs below $0.3ml/kg$	1.3
Anuria	Failure of the kidneys to produce urine	2.1
CVP	Central venous pressure	3.3
D_t	Information up to time t	3.4.1
$\boldsymbol{F}_t$	Regression vector	4.1
$\boldsymbol{\theta}_t$	p-dimensional paramater vector	4.1
ν_t	Observational stochastic error series	4.1

Table H.1: Table of Notation and Terminology

Notation	Description	Section
V_t	Observational variance	4.1
$\boldsymbol{G}_t$	Evolution matrix	4.1
$\boldsymbol{\omega}_t$	Evolution stochastic error series	4.1
$\boldsymbol{W}_t$	Evolution variance	4.1
D_0	Historical information	4.1
$\boldsymbol{a}_t$	State prior mean vector	4.2
$\boldsymbol{R}_t$	State prior covariance matrix	4.2
f_t	Mean of forecasted observation distribution	4.2.1
Q_t	Variance of forecasted observation distribution	4.2.1
$\boldsymbol{m}_t$	State posterior mean vector	4.4
$\boldsymbol{C}_t$	State posterior covariance matrix	4.4
$\boldsymbol{A}_t$	Adaptive factor	4.4
I_t	External information	6.2
e_t	Difference between observation and forecast mean	4.4
δ	System discount factor $\in (0, 1]$	4.8
ϕ	Observational precision	4.9
n_t	$2\times$ Location of precision parameter, ϕ	4.9.1
d_t	$2\times$ Scale of precision parameter, ϕ	4.9.1
S_t	Observational variance estimate	4.9.2
δ_V	Observational discount factor $\in (0, 1]$	4.9.4
Oliguria	Low output of urine	5.1
$\boldsymbol{h}_t$	Mean vector of intervention, I_t	6.1.2
$\boldsymbol{H}_t$	Covariance matrix of intervention, I_t	6.1.2
$\boldsymbol{\xi}_t$	Additional evolution variance, $\boldsymbol{\xi}_t \sim \mathrm{N}(\boldsymbol{h}_t, \boldsymbol{H}_t)$	6.1.2

Table H.2: Table of Notation and Terminology Continued

Bibliography

[1] M. West and P. J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edition, 1997.

[2] Azrina Md Ralib, John W Pickering, Geoffrey M Shaw, and Zoltán H Endre. The urine output definition of acute kidney injury is too liberal. *Critical care*, 17(3):R112, 2013.

[3] Michael Joannidis, Barbara Metnitz, Peter Bauer, Nicola Schusterschitz, Rui Moreno, Wilfred Druml, and Philipp GH Metnitz. Acute kidney injury in critically ill patients classified by akin versus rifle using the saps 3 database. *Intensive care medicine*, 35(10):1692–1702, 2009.

[4] John A Kellum, Florentina E Sileanu, Raghavan Murugan, Nicole Lucko, Andrew D Shaw, and Gilles Clermont. Classifying aki by urine output versus serum creatinine level. *Journal of the American Society of Nephrology*, 26(9):2231–2238, 2015.

[5] Marc Vives, Duminda Wijeysundera, Nandor Marczin, Pablo Monedero, and Vivek Rao. Cardiac surgery-associated acute kidney injury. *Interactive cardiovascular and thoracic surgery*, 18(5):637–645, 2014.

[6] National Health Service. Acute kidney injury, September 2018.

[7] John A. Kellum, Norbert Lameire, Peter Aspelin, Rashad S. Barsoum, Emmanuel A. Burdmann, Stuart L. Goldstein, Charles A. Herzog, Michael Joannidis, Andreas Kribben, Andrew S. Levey, Alison M. MacLeod, Ravindra L. Mehta, Patrick T. Murray, Saraladevi Naicker, Steven M. Opal, Franz Schaefer, Miet Schetz, and Shigehiko Uchino. Kidney disease: Improving global

outcomes (kdigo) acute kidney injury work group. kdigo clinical practice guideline for acute kidney injury. *Kidney International Supplements*, 2(1):1–138, 3 2012.

[8] Glenn M Chertow, Elisabeth Burdick, Melissa Honour, Joseph V Bonventre, and David W Bates. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *Journal of the American Society of Nephrology*, 16(11):3365–3370, 2005.

[9] Eric AJ Hoste, Gilles Clermont, Alexander Kersten, Ramesh Venkataraman, Derek C Angus, Dirk De Bacquer, and John A Kellum. Rifle criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Critical care*, 10(3):R73, 2006.

[10] Andrea Lassnigg, Daniel Schmidlin, Mohamed Mouhieddine, Lucas M Bachmann, Wilfred Druml, Peter Bauer, and Michael Hiesmayr. Minimal changes of serum creatinine predict prognosis in patients after cardiothoracic surgery: a prospective cohort study. *Journal of the American Society of Nephrology*, 15(6):1597–1605, 2004.

[11] Elliott M Levy, Catherine M Viscoli, and Ralph I Horwitz. The effect of acute renal failure on mortality: a cohort analysis. *Jama*, 275(19):1489–1494, 1996.

[12] Shigehiko Uchino, Rinaldo Bellomo, Donna Goldsmith, Samantha Bates, and Claudio Ronco. An assessment of the rifle criteria for acute renal failure in hospitalized patients. *Critical care medicine*, 34(7):1913–1917, 2006.

[13] Ravindra L Mehta, John A Kellum, Sudhir V Shah, Bruce A Molitoris, Claudio Ronco, David G Warnock, and Adeera Levin. Acute kidney injury network: report of an initiative to improve outcomes in acute kidney injury. *Critical care*, 11(2):R31, 2007.

[14] Andrew Gelman. *Data analysis using regression and multilevel/hierarchical models*. Analytical methods for social research. Cambridge University Press, Cambridge ; New York, 2007.

[15] A. Pole, M. West, and P. J. Harrison. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman-Hall, 1994.

[16] R. Prado and M. West. *Time Series: Modelling, Computation and Inference*. Chapman and Hall/CRC Press, 2010.

[17] Giovanni Petris. *Dynamic linear models with R*. Use R! Springer, Dordrecht ; London, 2009.

[18] Liisa Petäjä, Suvi Vaara, Sasu Liuhanen, Raili Suojaranta-Ylinen, Leena Mildh, Sara Nisula, Anna-Maija Korhonen, Kirsi-Maija Kaukonen, Markku Salmenperä, and Ville Pettilä. Acute kidney injury after cardiac surgery by complete kdigo criteria predicts increased mortality. *Journal of cardiothoracic and vascular anesthesia*, 31(3):827–836, 2017.

[19] Marc-Gilbert Lagny, François Jouret, Jean-Noël Koch, Francine Blaffart, Anne-Françoise Donneau, Adelin Albert, Laurence Roediger, Jean-Marie Krzesinski, and Jean-Olivier Defraigne. Incidence and outcomes of acute kidney injury after cardiac surgery using either criteria of the rifle classification. *BMC nephrology*, 16(1):76, 2015.

[20] Joseph F Dasta, Sandra L Kane-Gill, Amy J Durtschi, Dev S Pathak, and John A Kellum. Costs and outcomes of acute kidney injury (aki) following cardiac surgery. *Nephrology Dialysis Transplantation*, 23(6):1970–1974, 2008.

[21] Maurício N Machado, Marcelo A Nakazone, and Lilia N Maia. Prognostic value of acute kidney injury after cardiac surgery according to kidney disease: improving global outcomes definition and staging (kdigo) criteria. *PloS one*, 9(5):e98028, 2014.

[22] Malene Kærslund Hansen, Henrik Gammelager, Carl-Johan Jacobsen, Vibeke Elisabeth Hjortdal, J Bradley Layton, Bodil Steen Rasmussen, Jan Jesper Andreasen, Søren Paaske Johnsen, and Christian Fynbo Christiansen. Acute kidney injury and long-term risk of cardiovascular events after cardiac surgery: a population-based cohort study. *Journal of cardiothoracic and vascular anesthesia*, 29(3):617–625, 2015.

[23] A. F. M. Smith and M. West. Monitoring renal transplants: An application of the multi-process kalman filter. *Biometrics*, 39:867–878, 1983.

[24] Graeme L Hickey, Stuart W Grant, Camila Caiado, Simon Kendall, Joel Dunning, Michael Poullis, Iain Buchan, and Ben Bridgewater. Dynamic prediction

modeling approaches for cardiac surgery. *Circulation: Cardiovascular Quality and Outcomes*, 6(6):649–658, 2013.

[25] Samuel H Howitt, Stuart W Grant, Camila Caiado, Eric Carlson, Dowan Kwon, Ioannis Dimarakis, Ignacio Malagon, and Charles McCollum. The kdigo acute kidney injury guidelines for cardiac surgery patients in critical care: a validation study. *BMC nephrology*, 19(1):149, 2018.

[26] Robert Goodell Brown. *Statistical forecasting for inventory control*. McGraw/Hill, 1959.

[27] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[28] P Jeff Harrison and Parma P Veerapen. A bayesian decision approach to model monitoring and cusums. *Journal of Forecasting*, 13(1):29–36, 1994.

[29] H Jeffreys. Theory of probability, 1961.

[30] M. West. Modelling with mixtures (with discussion). In *Bayesian Statistics 4*, pages 503–524. Oxford University Press, 1992.

[31] A. F. M. Smith, M. West, K. Gordon, M.S. Knapp, and I. Trimble. Monitoring kidney transplant patients. *The Statistician*, 32:46–54, 1983.

[32] I. Trimble, M. West, M. S. Knapp, R. Pownall, and A. F. M. Smith. Detection of renal allograft rejection by computer. *British Medical Journal*, 286:1695–1699, 1983.

[33] P. J. Harrison and C. F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):205–247, 1976.

[34] JRM Ameen and PJ Harrison. Discount bayesian multiprocess modelling with cusums. *Time Series Analysis: Theory and Practice*, 5:117–134, 1984.

[35] Víctor M Guerrero. Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1):37–48, 1993.

[36] Leow Wee Kheng. Matrix differentiaion, September 2018.

[37] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.