State-Space Models with GARCH Errors: Application to Health Data

Zayd Omar

Master of Science

Department of Mathematics and Statistics

McGill University

Montreal,Quebec

2019-08-10

This thesis is submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

ACKNOWLEDGEMENTS

First and foremost, I am tremendously grateful for all the guidance and contributions my supervisors, Professor David Stephens and Professor Alexandra Schmidt, have given me over the past two years and in writing this thesis. I will always be indebted to all the staff at McGill University who have helped me in my journey through the University. I would like to thank all my friends for all the comforts they have brought me and especially Shomoita Alam for all that she has taught me. Finally, I would especially like to thank my parents and my brother, without whom, I would not be where I am.

ABSTRACT

State-space models are used to study non-stationary data. However, in the presence on non-Gaussian error terms the standard state-space model does not apply. We investigate the properties of univariate and multivariate state-space models under conditional heteroskedasticity and multiple structural breaks. This allows us to extend the standard state-space models to heavy tailed data and allow for dynamic parameters. We develop a Gibbs sampling algorithm to carry out Bayesian inference on the parameters and the latent state vector. Finally, we carry out an empirical study on ICU data. We find that our models are better able to capture the variation in the data than the standard state-space models.

ABRÉGÉ

Les modèles d'espace d'ètats sont utilisés pour étudier des données non stationnaires. Cependant, en présence de termes d'erreur non gaussiens, le modèle d'espace d'états standard ne s'applique pas. Nous étudions les propriétés des modèles univariés et multivariés d'espace-état sous hétéroskédasticité conditionnelle et fractures structurelles multiples. Cela nous permet d'étendre les modèles d'espace d'états standard aux données à queue lourde et de prendre en compte les paramètres dynamiques. Nous développons un algorithme d'échantillonnage de Gibbs pour réaliser l'inférence bayésienne sur les paramètres et le vecteur d'état latent. Enfin, nous menons une étude empirique sur les données de l'unité de soins intensifs. Nous constatons que nos modèles sont mieux à même de rendre compte de la variation des données par rapport aux modèles despace à états standard.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

1.1 Introduction

Studying non-stationary time series is an important part of statistics. First introduced in Kalman et al. (1960) and Kalman and Bucy (1961), state-space models (SSM), form a rich class of models that may be used to study non-stationary data. These models have been studied for a while now and their flexibility in allowing for time varying parameters and structural changes make them extremely useful. These models have many applications in the field of economics, engineering and the health sciences. Kim et al. (1999), Hamilton (1994), Harvey (1990) and Durbin and Koopman (2012) have a list of applications and examples of problem where SSMs are used.

An SSM is a two component model. The first component is a latent or hidden state that evolves with time and the second component is the observation component that is driven by these latent states. The Gaussian-SSM (or standard SSMs) assumes that the distribution of the errors for these two components are Gaussian in nature. The assumption of having error terms that are Gaussian, gives SSMs many of their elegant properties. This assumption in fact makes the recursions described in Kalman et al. (1960) quite tractable and makes estimating these complex models fairly easy. However, this assumption also may be a disadvantage. While standard SSMs allow for time varying variance parameters, heteroskedasticity is still maintained under the Gaussianity assumption. We may very well want to allow for non-Gaussianity, however, in doing so we lose the tractability of the Kalman Filter and Kalman recursions.

We would like to extend SSMs to having heteroskedastic errors of a more general form instead of the strict Gaussianity assumption without completely abandoning the tractability that this assumption permits. This leads us to the idea of conditionally Gaussian error terms. Using the ideas introduced by Engle (1982) and Bollerslev (1986) we can extend the standard SSM to have errors that allow for a more general form of heteroskedasticity while maintaining conditional Gaussianity.

There is a great deal of literature for non-linear and non-Gaussian state space models. We refer the interested reader to Douc et al. (2014) for an exposition into these non-standard SSMs. More on special exponential family of dsitribution can be found in Gamerman et al. (2013). Further, Kitagawa (1987), Konishi and Kitagawa (2008) describes an extended version of the Kalman Filter in order to estimate these non-Gaussian models. However, as mentioned above these techniques lose their analytic tractability when in the non-Gaussian case.

The Auto-Regressive-Conditional-Heteroskedasticity (ARCH), introduced by Engle (1982) and Generalized-ARCH (GARCH) models introduced by Bollerslev (1986) have played crucial roles in modeling financial and economic time series data. The intuition behind these ideas is that, while the error terms might be uncorrelated, their higher order moments may have some kind of dependence. For example if we witness a huge market shock today, it is very likely that the volatility of this shock will persist tomorrow and maybe even for a longer period of time. Hence, we would like the volatility of tomorrow to depend on the volatility of today. In this way over time we are getting residuals that are uncorrelated but heteroskedastic in nature.

Our first goal is to extend the Gaussian-SSM to allow for GARCH errors. We call this model a GARCH-SSM. One can immediately see why this would be a useful model by studying

some of the properties of the Gaussian distribution. The Gaussain distribution has a kurtosis of 3, which leads to thin tails. As a result Gaussian distribution is not a good distribution to model data sets with outliers and large deviations. The GARCH model however, has a kurtosis greater than 3 and is better equipped to deal with large scale deviations from the mean. Gouriéroux (2012) gives a good exposition into the properties of the ARCH and GARCH models. It is this ability to deal with outliers and thicker tails that we would like our SSMs to have. Further we also want to be able to use the ability of SSMs to deal with structural changes to have structurally changing GARCH parameters.

Although extending the standard SSM to have GARCH errors is quite a natural extension, as will be shown later on, there has not been a lot of work looking into this. The best we have found about GARCH and SSMs is the work by Wong et al. (2006), which looks at model EEG data using state-variances that are GARCH. However, this does not look to extend the GARCH error assumption to the observation level of the SSM and also does not look at any structural breaks in the GARCH-SSM model. We show that this assumption of GARCH errors still allow us to use the Kalman recursions and estimate the SSMs quite elegantly without having to rely on the intractable estimation techniques that we would have to use if we chose other distributions.

Our second goal is to estimate these models using Bayesian techniques. However, one may also choose to use maximum likelihood to estimate these models. To estimate and do inference on the latent states, Frühwirth-Schnatter (1994), Carter and Kohn (1994) proposed a Forward Filtering Backward Sampling (FFBS) algorithm. Due to the nature of the GARCH-SSM the FFBS algorithm is still applicable for estimating the state vectors and is relied on heavily in estimating the states of the models shown later on. For general study of estimating SSMs we refer the reader to Prado and West (2010), West and Harrison (1997), Petris et al. (2009) which describe in details the Bayesian approach to estimating SSMs. For a

general introduction to Bayesian estimation and computation techniques a large number of resources are available. We refer the reader to Robert (2007), Gelman et al. (2013), Robert and Casella (2013), Bernardo and Smith (1994), Gamerman and Lopes (2006).

Our main inspiration for this project came from the *BrainIT* core data set provided by collaborators at McGill University Department of Epidemiology, Biostatistics and Occupational Health. The dataset contains various physiological readings from anonymous patients at the Intensive Care Unit (ICU). ICU's are complex environments, where patients are constantly monitored. This leads to the opportunity of extracting vast amounts of data from the ICU. The *BrainIT* data set contains a large amount of physiological data tracking patients in differing conditions.

The data obtained from the ICU is of interest to physicians and hospitals who may use this data for on-line detection of events that show changes in the status of a patient. The volume of data available in these data heavy situation may pose a challenge to physicians. Thus it would be of interest to design potentially multivariate models that summarize these heterogeneous data and allow for easier monitoring of patients. It is also important to be able to develop models that are able to handle missing data, which may be missing due to a variety of reasons.

While it is easy to think of a number of reasons why care givers would be interested in developing concise models to improve the quality of care, what is more difficult is to develop accurate models for these types of data. We show two sets of measurements obtained from a particular patient in the *BrainIT* data set.

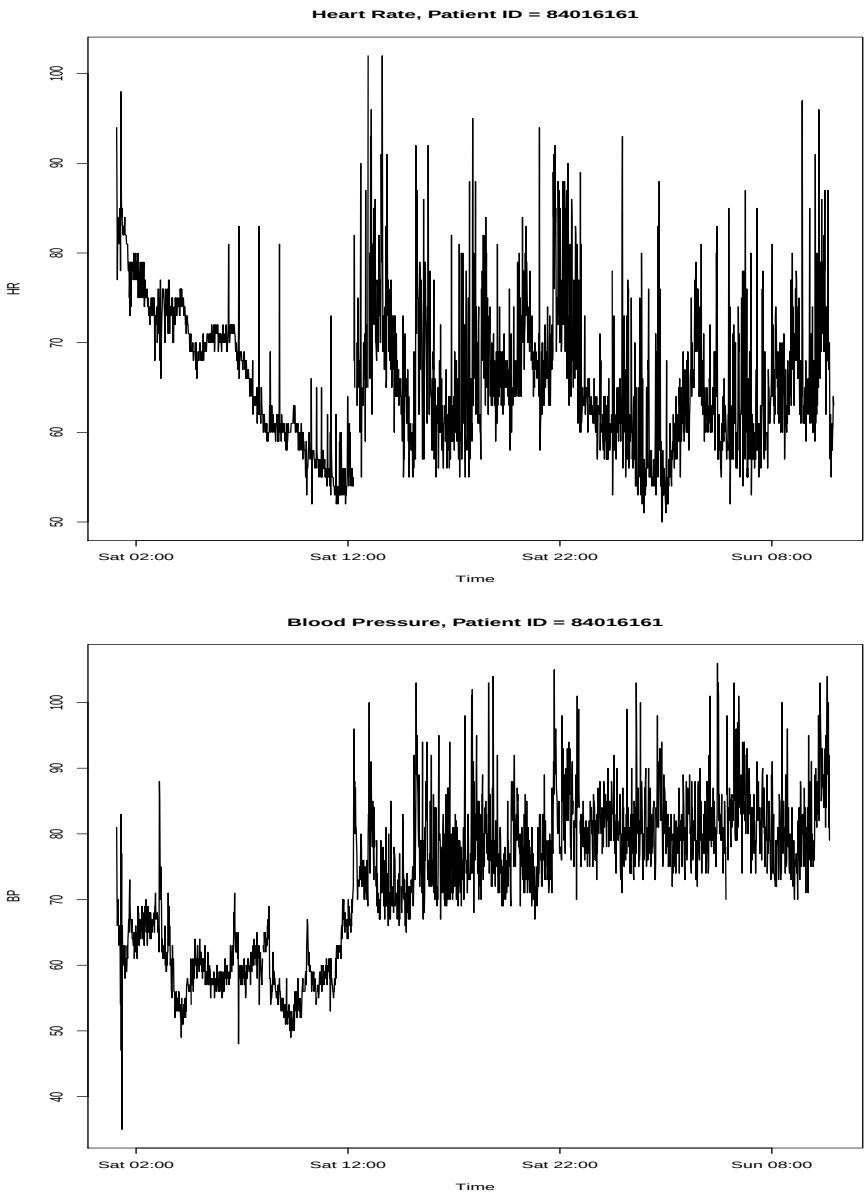

Figure 1–1: Left: Heart Rate of a particular patient; Right: Blood Pressure of the same patient. Both series show measurements taken at 4-minute intervals.

The data seems rather quite challenging to deal with. Clearly due to the non-stationarity nature of the data we cannot use simple classical time series modeling. Secondly, due to the presence of large deviations from the mean we also need to be able to use appropriate distributions to account for this behaviour. Further, do we actually believe that this is a

single regime model, i.e. do the same parameters operate on the data the entire time or do they change? If so, then can we say anything about the structural break? When did it happen? Why did it happen? So on and so forth.

Gordon and Smith (1990) applied SSMs to a variety of health data with missing values and with discontinuous change points. We wanted to build a similar type of model for our health series above that would be able to address the different properties shown in the series in Figure 1. What ever model we choose must be able to deal with the type of non-stationarity, structural breaks and the outliers seen in many of these data sets. Given the structure of the mean level of the model we chose to use SSMs, since they have the ability to deal with these types of non-stationary data. However, the standard Gaussian-SSM would be ill equipped to model the outliers present in many of our physiological time series. Our need to model the heavy-tailed behaviour of this data without losing the analytical tractability of the Kalman Filter estimates eventually led us to study a GARCH process and thereafter formulate the GARCH-SSM. This model retains the analytical tractability of the Kalman recursions while also retaining the ability to model the heavier tails.

It has been known for sometime that the physiological patterns in humans follow a complex interrelated patterns. As early as the 18^{th} century, research was being conducted into measuring and studying the correlation between the complex dynamical systems that operate within our bodies. Hales (1733) studied the correlation between the heart rate, blood pressure and respiration. While we are fortunate to have the *Brain IT* data that measures many more features than what Hales had the luxury of measuring, our choice of studying the heart rate and blood pressure is as much due to convenience as it is to the their connection with each other. We use convenience to refer to the completeness of the data. For many patients, other than the heart rate and blood pressure series, many of the other measurements, like respiration rate, are missing as much as of 50% or more of the data. On it own modeling

such incomplete data does not pose any difficulty. However, since we are interested in some multivariate modeling as well, trying to build such a model where one series is near complete while the other has 50% of its values missing seems like a rather impractical task. One can make the argument that we can just model both series from a time point where 'sufficient' amount of data is available. While that would be a sound argument we decided to try our hand instead with data that already is sufficiently complete. And we use the term sufficiently complete because we can see that in the heart rate series in Figure 1, slightly after 1200H there is gap in the series indicating the presence of missing data. Thus an appropriate model developed for this data must be able to handle missing data.

1.2 Outline of the Thesis:

The objective of this thesis is to build a model that addresses the aforementioned features of the health data available from the *Brain IT* data set. We organize this document in the following manner. Chapter 2 looks at the theory of SSMs; Chapter 3 looks at the estimation of SSMs through both the Bayesian and Frequentist viewpoints, where the emphasis is given on the Bayesian framework; Chapter 4 gives a brief theory of the conditional heteroskedasticity in the univariate and multivariate case; Chapter 5 looks at the extension of the standard SSM to the GARCH-SSM; Chapter 6 looks at the time varying parameters and SSMs with structural breaks; Chapter 7 shows a simulation study; Chapter 8 applies the GARCH-SSM to the *Brain IT* data.

CHAPTER 2
Basic Theory

2.1 State Space Models:

Consider a real valued $(n \times 1)$ vector time series $\{\boldsymbol{y}_t\}$ indexed by $t \in \mathbb{N}$, where $\mathbb{N}$ is countable. A state space representation of $\boldsymbol{y}_t$, is a representation that allows us to write the dynamics of $\boldsymbol{y}_t$ in terms of a latent or hidden state, $\boldsymbol{\theta}_t$, where, $\boldsymbol{\theta}_t$ is another real valued $(r \times 1)$ vector random variable indexed by $t \in \mathbb{N}$. In particular, the representation of $\{\boldsymbol{y}_t\}$ is given by,

$$\begin{aligned} \text{(Observation equation)} \qquad & \boldsymbol{y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim WN(\boldsymbol{0}, \boldsymbol{V}_t) \\ \text{(State equation)} \qquad & \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim WN(\boldsymbol{0}, \boldsymbol{W}_t), \end{aligned} \tag{2.1}$$

where, $\boldsymbol{F}_t$ is an $(n \times r)$ matrix known as the design matrix, $\boldsymbol{G}_t$ is an $(r \times r)$ matrix known as the state/evolution matrix, $\boldsymbol{v}_t$ is an $(n \times 1)$ vector white noise process, $\boldsymbol{V}_t$ is an $(n \times n)$ covariance matrix, $\boldsymbol{w}_t$ is an $(r \times 1)$ vector white noise process and $\boldsymbol{W}_t$ is an $(r \times r)$ covariance matrix.

We have the following assumptions of independence for the error terms.

$$\begin{aligned} E(\boldsymbol{v}_t\boldsymbol{v}_s') &= \begin{cases} \underset{(n\times n)}{\boldsymbol{V}_t} & t = s \\ \boldsymbol{0} & \text{otherwise} \end{cases} \\ E(\boldsymbol{w}_t\boldsymbol{w}_s') &= \begin{cases} \underset{(r\times r)}{\boldsymbol{W}_t} & t = s \\ \boldsymbol{0} & \text{otherwise} \end{cases} \end{aligned} \tag{2.2}$$

We further assume that $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are mutually uncorrelated at all lags, i.e.

$$E(\boldsymbol{v}_t \boldsymbol{w}_s') = \mathbf{0}, \qquad \forall t, s \in \mathbb{N} \tag{2.3}$$

The quadruple $(\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t)$ specifies the SSM at time t. We can also have that these parameters are time invariant, in which case we have that the quadruple may be written as $(\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t) = (\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W})$.

With this specification SSMs satisfy the following assumption,

A.1 $(\boldsymbol{\theta}_t)$ is Markovian, i.e. $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, ..., \boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$.

A.2 Conditional on $\boldsymbol{\theta}_t$, $\boldsymbol{y}_t$ is fully specified i.e $p(\boldsymbol{y}_t|\boldsymbol{\theta}_t, ..., \boldsymbol{\theta}_0, \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_1) = p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)$.

This gives us that the joint distribution of the state vector and the observation vector is then given by,

$$p(\boldsymbol{\theta}_{0:t}, \boldsymbol{y}_{1:t}) = p(\boldsymbol{\theta}_0) \prod_{i=1}^{t} p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i-1}) p(\boldsymbol{y}_i|\boldsymbol{\theta}_i) \tag{2.4}$$

When $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are Gaussian random variables, we have referred to them as Gaussian-SSM or standard SSM.

2.2 Examples of State Space Models:

2.2.1 Random Walk plus Drift Model:

Let y_t be a real valued time series whose state space representation is specified by the quadruple below,

$$(\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W}) = \Big\{1, 1, V, W\Big\}$$

The state-space representation of y_t is given by,

$$y_t = \theta_t + v_t \qquad v_t \sim N(0, V)$$
$$\theta_t = \theta_{t-1} + w_t, \quad w_t \sim N(0, W)$$

This type of Gaussian-SSM is alternatively known as a **first-order polynomial** SSM (univariate case).

2.2.2 Multivariate First Order Polynomial:

Let $\boldsymbol{y}_t$ be an (n $\times$ 1) vector time series whose state space representation is specified by the quadruple below,

$$(\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W}) = \left\{ \boldsymbol{I}_n,\ \boldsymbol{I}_n,\ diag(V_1, ..., V_n),\ diag(W_1, ..., W_n) \right\}$$

This gives us that,

$$\boldsymbol{y}_t = \boldsymbol{I}_n \boldsymbol{\theta}_t + \boldsymbol{v}_t \qquad \boldsymbol{v}_t \sim N(0, \boldsymbol{V})$$
$$\boldsymbol{\theta}_t = \boldsymbol{I}_n \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W})$$
$$\boldsymbol{\theta}_t = \begin{pmatrix} \theta_{t,1} \\ \vdots \\ \theta_{n,1} \end{pmatrix}$$

The multivariate first order polynomial models is then basically n univariate random walk plus drift models evolving together. More details about the general j^{th} order polynomial SSM may be found in West and Harrison (1997).

2.3 Estimating the Distributions of the State Vector:

Given a Gaussian SSM, our main task with these models is to make inference on the unobserved state vectors and the unknown variances. We could also use the data available to carry out forecasting exercises. Gaussianity allows us to use the normal equations to help

derive the conditional distributions for the unknown state vectors and the forecasts.

We give three definitions. The distribution of the state vector at time t conditional on the data up to time t is given by $\pi(\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t})$, is called the *filtering* distribution. The distribution of the forecast at time s conditional on the data up to time t, $s > t$ is given by $\pi(\boldsymbol{y}_s|\boldsymbol{y}_{1:t})$ and is called the *forecasting* distribution. The distribution of the state vector at time s conditional on the data up to time t, $s < t$ is given by $\pi(\boldsymbol{\theta}_s|\boldsymbol{y}_{1:t})$, and this distribution is called *smoothing* distribution.

2.4 Filtering Distributions:

Suppose that we have a DLM specified by the quadruple $(\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t)$ and that we start with the distribution, $\pi(\boldsymbol{\theta}_0) \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$. The following theorem describes the prior, posterior and one-step-ahead forecast distributions for the multivariate DLM.

The **filtering** distribution is given by,

1. The prior distribution at time t for the state vector is:

$$\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t-1} \sim N(\boldsymbol{a}_t, \boldsymbol{R}_t) \tag{2.5}$$

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

2. The one-step-ahead forecast for the observation vector is:

$$\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1} \sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t) \tag{2.6}$$

   Where $\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t$ and $\boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t$

3. The posterior distribution of the state at time t is then given by:

$$\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t} \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t) \tag{2.7}$$

Where,

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{e}_t \\ \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{K}_t \boldsymbol{Q}_t \boldsymbol{K}_t' \\ \boldsymbol{K}_t &= \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{Q}_t^{-1} \\ \boldsymbol{e}_t &= \boldsymbol{y}_t - \boldsymbol{f}_t \end{aligned} \tag{2.8}$$

We have $\boldsymbol{e}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t)$. The term $\boldsymbol{K}_t = \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{Q}_t^{-1}$ is known as the **Kalman gain matrix** or **gain** matrix for short. The solution above is known as the Kalman Filter recursion and was first proposed by Kalman et al. (1960) as a theory for optimal prediction. The proof of the above is simple and relies heavily on the normal equations and the fact that we have conditional independence of the state and observation vectors on account of A.1 and A.2 (see West and Harrison (1997) for more details) .

2.5 Smoothing:

The previous section was focused on using the current information to produce estimates for the future. However, often, it is just as important to use the current information to make estimations of the past in order to better understand the underlying process governing the system. Using the current information to make inference on the past values of the state vector is known as **smoothing**. The distribution of $(\boldsymbol{\theta}_{t-k}|\boldsymbol{y}_{1:t})$, $k \geq 1$, is called the **smoothing** distribution.

We described above using $\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)$ and $\boldsymbol{f}_t(k), \boldsymbol{Q}_t(k)$, for the filtered and smoothed distribution we describe them using $-k$.

The **smoothing** distribution is

$$\boldsymbol{\theta}_{t-k}|\boldsymbol{y}_{1:t} \sim N[\boldsymbol{a}_t(-k), \boldsymbol{R}_t(-k)], \tag{2.9}$$

where,

$$\begin{aligned} \boldsymbol{a}_t(-k) &= \boldsymbol{m}_{t-k} + \boldsymbol{B}_{t-k}[\boldsymbol{a}_t(-k+1) - \boldsymbol{a}_{t-k+1}] \\ \boldsymbol{R}_t(-k) &= \boldsymbol{C}_{t-k} - \boldsymbol{B}_{t-k}[\boldsymbol{R}_t(-k+1) - \boldsymbol{R}_{t-k+1}]\boldsymbol{B}'_{t-k} \\ \boldsymbol{B}_t &= \boldsymbol{C}_t\boldsymbol{G}'_{t+1}\boldsymbol{R}^{-1}_{t+1}, \end{aligned} \tag{2.10}$$

where,

$$\boldsymbol{a}_t(0) = \boldsymbol{m}_t, \qquad \boldsymbol{R}_t(0) = \boldsymbol{C}_t$$

2.6 Missing Data:

In the event of a missing data the filtering equations are easy to calculate. Let us first look at the univariate case. Suppose we have that y_t is missing. In terms of the distribution for the unknown state vector we have that $\pi(\boldsymbol{\theta}_t|y_{1:t}) = \pi(\boldsymbol{\theta}_t|y_{1:t-1})$. Formally, having a missing value at time t is the same as having $\boldsymbol{F}'_t = \mathbf{0}$ at time t. This is saying that there is no link between the state vector and the observation at time t. This affects the Kalman gain in (8) and $\boldsymbol{K}_t$ becomes 0. This results in the filtered values in (8) becoming,

$$E[\boldsymbol{\theta}_t|y_{1:t}] = E[\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t-1}] = \boldsymbol{m}_t = \boldsymbol{a}_t$$

$$Var[\boldsymbol{\theta}_t|y_{1:t}] = Var[\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t-1}] = \boldsymbol{C}_t = \boldsymbol{R}_t$$

This seems obvious, since there is no information at time t and hence all the information up to time $t-1$ is just carried forward.

We now describe missing values for the multivariate case. Suppose that $\boldsymbol{y}_t$ is a $p \times 1$ dimensional vector. There are two possibilities that we need to specify. The first one is when information is missing across the entire vector at time t, i.e. all the components of $\boldsymbol{y}_t$ are

missing. In this case the update is similar to the univariate case. The observation matrix $\boldsymbol{F}_t'$ is zero. This also results in the Kalman gain matrix to be a matrix of $0'$s. And the filtered updates are given as,

$$E[\boldsymbol{\theta}_t|y_{1:t}] = E[\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t-1}] = \boldsymbol{m}_t = \boldsymbol{a}_t$$
$$Var[\boldsymbol{\theta}_t|y_{1:t}] = Var[\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t-1}] = \boldsymbol{C}_t = \boldsymbol{R}_t$$

The second possibility is that only some of the observations at time t are missing. That is, for $\boldsymbol{y}_t$, only some components are missing while the others are present. In this case, we have that the components in the observation matrix $\boldsymbol{F}_t'$ that correspond to the missing values, are set to 0. This results in the a Kalman gain matrix, whose columns, corresponding to the missing values, contain zeros.

CHAPTER 3
Estimating the Parameters of a State Space Model:

3.1 Maximum Likelihood estimation:

Suppose that we have a multivariate time series $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$, whose distribution depends on some unknown parameter $\boldsymbol{\psi}$ and that in particular $\boldsymbol{y}_t$ has a Gaussian-SSM state space representation. To build the likelihood equation we use the conditional likelihood. Notice that by assumption A.2, which gives us that the $\boldsymbol{y_t}$ depends only on the state vector $\boldsymbol{\theta}_t$, if we had the state vectors then we could use (2.1) to build a fully Gaussian likelihood. However, since the state vectors are latent and in most case must be estimated we resort to the conditional likelihood method given by,

$$p(\boldsymbol{y}_1, ..., \boldsymbol{y}_n|\boldsymbol{\psi}) = \prod_{t=1}^{n} p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\psi}) \tag{3.1}$$

We know from previous results that $\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \psi \sim N(f_t, Q_t)$, where $\boldsymbol{f_t}$ and $\boldsymbol{Q_t}$ are obtained from the Kalman Recursions given in section 1.3.1, equation (6). This then gives us the joint distribution,

$$\begin{aligned} p(\boldsymbol{y}_1, ..., \boldsymbol{y}_n|\boldsymbol{\psi}) &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \det(\boldsymbol{Q_t})}} \exp\Big(\frac{(\boldsymbol{y}_t - \boldsymbol{f_t})'\boldsymbol{Q_t}^{-1}(\boldsymbol{y}_t - \boldsymbol{f_t})}{-2}\Big) \\ \implies l(\boldsymbol{\psi}) &= -\frac{1}{2}\sum_{t=1}^{n} \det(\boldsymbol{Q_t}) - \frac{1}{2}\sum_{t=1}^{n}\Big((\boldsymbol{y}_t - \boldsymbol{f_t})'\boldsymbol{Q_t}^{-1}(\boldsymbol{y}_t - \boldsymbol{f_t})\Big). \end{aligned} \tag{3.2}$$

Where the last line above is the log-likelihood. This can be numerically optimized to obtain the maxmum likelihood estimates (MLE) for the unknown parameters. Jensen et al. (1999) give regularity conditions needed for weak consistency of the MLE as well as their asymptotic normality. Further details on the asymptotic behaviour of the MLE can be found in Hannan

and Deistler (1988) and Caines (2018).

3.2 Bayesian Inference for SSMs:

While the maximum likelihood method is one approach to estimating SSM's, we can also use Bayesian techniques to estimate the parameters and make inference on them. In fact the latency of the state parameters makes the Bayesian method a natural approach for exploring the distributional properties of the unknown parameters and the latent state vectors.

For SSMs the assumptions A.1 and A.2 and the subsequent joint distribution of the observation and state vectors given by,

$$p(\boldsymbol{\theta}_{0:t}, \boldsymbol{y}_{1:t}|\boldsymbol{\psi}) = p(\boldsymbol{\theta}_0|\boldsymbol{\psi}) \prod_{i=1}^{t} p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i-1}, \boldsymbol{\psi}) p(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\psi}) \tag{3.3}$$

allows a straightforward way of carrying out a Bayesian estimation of the model. Based on a series of observations, $\boldsymbol{y}_1, ..., \boldsymbol{y}_t$ and some parameters $\boldsymbol{\psi}$, we would like to sample the full posterior $p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n|\boldsymbol{y}_{1:t})$.

This gives the posterior distribution for the state parameters as,

$$\begin{aligned} p(\boldsymbol{\theta}_{0:n}, \boldsymbol{\psi}|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{\theta}_{0:n}, \boldsymbol{y}_{1:t}, \boldsymbol{\psi})}{p(\boldsymbol{y}_{1:t})} \\ &\propto p(\boldsymbol{\psi}) p(\boldsymbol{\theta}_0|\boldsymbol{\psi}) \prod_{i=1}^{t} p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i-1}, \boldsymbol{\psi}) p(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\psi}). \end{aligned} \tag{3.4}$$

The goal now is to sample from the posterior distribution the parameters and possibly the latent states. It is almost always easier to sample from $p(\boldsymbol{\psi}|\boldsymbol{\theta}_{0:t}, \boldsymbol{y}_{1:t})$ than from just $p(\boldsymbol{\psi}|\boldsymbol{y}_{1:t})$, where $\boldsymbol{\psi}$ is a set of unknown parameters. In the same way to sample the unknown states we can sample from $p(\boldsymbol{\theta}_{0:t}|\boldsymbol{\psi}, \boldsymbol{y}_{1:t})$. However, notice that to sample from this latter distribution means that we would need to identify the cross correlations between ant $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_t$. However, we can get around this by noticing that,

$$p(\boldsymbol{\theta}_{0:n}|\psi, \boldsymbol{y}_{1:t}) = \prod_{i=0}^{t} p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1:t}, \psi, \boldsymbol{y}_{1:t}).$$

Where the last term is just $p(\boldsymbol{\theta}_t|\boldsymbol{y}_{1:t}, \psi)$. We have that,

$$p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1:t}, \boldsymbol{\psi}, \boldsymbol{y}_{1:t}) = p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1}, \boldsymbol{\psi}, \boldsymbol{y}_{1:i}) \sim N(\boldsymbol{h}_i, \boldsymbol{H}_i)$$

where,

$$\begin{aligned} \boldsymbol{h}_i &= \boldsymbol{m}_i + \boldsymbol{C}_i \boldsymbol{G}_{i+1} \boldsymbol{R}_{i+1}^{-1} (\boldsymbol{\theta}_{i+1} - \boldsymbol{a}_{i+1}) \\ \boldsymbol{H}_i &= \boldsymbol{C}_i - \boldsymbol{C}_i \boldsymbol{G}_{i+1}' \boldsymbol{R}_{i+1}^{-1} \boldsymbol{G}_{i+1} \boldsymbol{C}_i' \end{aligned} \tag{3.5}$$

This owes to the fact that by assumption A.1 we have that the $\boldsymbol{\theta}_t'$s are Markovian and by assumption we have that $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}$ is Gaussian, hence by inverting the distribution we can find the distribution of $\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t$.

Now that we have specified the distribution $p(\boldsymbol{\theta}_{0:t}|\boldsymbol{\psi}, \boldsymbol{y}_{1:t})$ conditional on the observation $\boldsymbol{y}$ and the unknown parameters $\boldsymbol{\psi}$ we can draw a sample of the state vectors. This type of sampling of the state vectors is known as *forward filtering and backward sampling* (FFBS) and was proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The FFBS algorithm is given below.

Algorithm 1: FFBS

1. We run the Kalman filter
2. Draw $\boldsymbol{\theta}_n \sim N(\boldsymbol{m}_n, \boldsymbol{C}_n)$, where $\boldsymbol{m}_n, \boldsymbol{C}_n$ can be found in (2.8)
3. For $t = n-1, ..., 0$ draw $\boldsymbol{\theta}_t \sim N(\boldsymbol{h}_t, \boldsymbol{H}_t)$, where $\boldsymbol{h}_t, \boldsymbol{H}_t$ can be found in (3.5)

However, drawing posterior samples of the state vector is not our only goal. We would also want to be able to sample the unknown parameters $\boldsymbol{\psi}$. To sample $\boldsymbol{\psi}$ from the posterior distribution given in (14), we can use a Gibbs Sampling or Metropolis Hastings algorithm depending on how complex the posterior distribution is. Suppose, like that the unknown

parameter is $\boldsymbol{\psi}$, where the unknown parameter can be decomposed into multiple components $\psi = (\psi_1, \psi_2, ...\psi_p)$. We set up the sampler as follows:

Algorithm 2: Posterior Sampling with FFBS step

1. Initialize the unknown parameters, $\psi_1^{(0)}, \psi_2^{(0)}, ..., \psi_p^{(0)}$
2. Draw the state parameters using FFBS using, $\boldsymbol{\theta}_{0:n}^{(0)} \sim p(\boldsymbol{\theta}_{0:n}^{(0)} | \boldsymbol{y}_{1:n}, \psi_1^{(0)}, ..., \psi_p^{(0)})$.
3. Then for iteration i,
   - Draw $\psi_1^{(i)} \sim p(\psi_1^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_2^{(i-1)}, ..., \psi_p^{(i-1)})$
   - Draw $\psi_2^{(i)} \sim p(\psi_2^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_1^{(i)}, \psi_3^{(i-1)}, ..., \psi_p^{(i-1)})$

     $\vdots$
   - Draw $\psi_j^{(i)} \sim p(\psi_j^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_1^{(i)}, ..., \psi_{j-1}^{(i)}, \psi_{j+1}^{(i-1)}, ..., \psi_p^{(i-1)})$
   - Draw the state parameters using FFBS using, $\boldsymbol{\theta}_{0:n}^{(i)} \sim p(\boldsymbol{\theta}_{0:n}^{(i)} | \boldsymbol{y}_{1:n}, \psi_1^{(i)}, ..., \psi_p^{(i)})$.
4. $i = i + 1$; Return to step 2

In algorithm 2 we are sampling each unknown parameter, ψ_j individually. We can modify and do a multivariate sampling as well. Suppose that, $\psi = (\psi_1, \psi_2, ...\psi_p) = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$, then do a blocked sampling. Depending on the form of the full conditional distributions given in Step 3 we can do a Gibb's sampling or use do a Metropolis-Hastings sampling regime. More details on the Gibbs-Sampling and Metropolis-Hastings sampling may be found in Marin and Robert (2007).

3.2.1 Example of a Univariate DLM:

We use a simple univariate first order DLM to give an illustration of how our posterior sampling can be done. As we have seen before the first order DLM is given by the quadruple $\{1, 1, V, W\}$. Here our unknown parameters are the observation and state variances, V and W, respectively (and obviously the latent state-vectors which can be obtained as a by-product of the estimation).

$$Y_i = \theta_i + v_i, \quad v_i \sim N(0, V)$$

$$\theta_i = \theta_{i-1} + w_i, \quad w_i \sim N(0, W)$$

The full posterior distribution is then given by,

$$\begin{aligned} p(\theta, V, W|Y_{1:n}) &\propto \mathcal{L}(\theta, V, W|\boldsymbol{y}) p(\theta|V, W)\pi_0(V)\pi_0(W) \\ &\propto \prod_{i=1}^{n} p(Y_i|\theta_i, V)p(\theta_i|\theta_{i-1}, W)p(\theta_0|W)\pi_0(V)\pi_0(W) \\ &\propto (V)^{-n/2} \exp\Big(\frac{\sum_{i=1}^{n}(Y_i - \theta_i)^2}{-2V}\Big) \times \\ &\qquad (W)^{-n/2} \exp\Big(\frac{\sum_{i=1}^{n}(\theta_i - \theta_{i-1})^2}{-2W}\Big) p(\theta_0|W)\pi_0(V)\pi_0(W) \end{aligned}$$

Given the form of the posterior this is how we would apply Algorithm 2 to simulate from the posterior distribution.

1. We specify an initial value for V and W, call these $V^{(0)}$ and $W^{(0)}$ (this is the Step 1 in Algorithm 2).
2. Then using these initial values we run the FFBS to obtain a sample of the state vectors $\boldsymbol{\theta}^{(0)}$ (this is the Step 2 in Algorithm 2).
3. Using these state vectors we then draw $V^{(1)}$ and $W^{(1)}$ from their full conditional posterior distributions (Step 3 in Algorithm 2).
4. Using these updated values of $V^{(1)}$ and $W^{(1)}$ we run the FFBS again to obtain $\boldsymbol{\theta}(1)$ and using the updated state vectors we update the sample V and W and so on and so forth.

First thing to note is that by choosing Inverse-Gamma priors for V and W we can do a conjugate analysis for V and W. Suppose that $\pi_0(V) \sim Inv - Gam(\alpha_y, \beta_y)$ and $\pi_0(W) \sim Inv - Gam(\alpha_S, \beta_S)$. Then,

$$p(V|\theta, W, Y) \propto (V)^{-\frac{n}{2}-\alpha_y-1} \exp\Big(\frac{\sum_{i=1}^{n}(Y_i-\theta_i)^2+2\beta_y}{-2V}\Big)$$
$$\propto Inv-Gamma\left(\frac{n}{2}+\alpha_y, \Big(\frac{\sum_{i=1}^{n}(Y_i-\theta_i)^2+2\beta_y}{-2}\Big)\right)$$

$$p(W|\theta, V, Y) \propto (W)^{-\frac{n}{2}-\alpha_S-1} \exp\Big(\frac{\sum_{i=1}^{n}(\theta_i-\theta_{i-1})^2+2\beta_S}{-2W}\Big)$$
$$\propto Inv-Gamma\left(\frac{n}{2}+\alpha_S, \Big(\frac{\sum_{i=1}^{n}(\theta_i-\theta_{i-1})^2+2\beta_S}{-2}\Big)\right)$$

3.2.2 Example of a Multivariate DLM:

Suppose now that we have a multivariate first order DLM. The observation at time t is the vector $\boldsymbol{y}_t$, where, $\boldsymbol{y}_t$ is a $p \times 1$ vector. The first order DLM in this multivariate case is specified by the quadruple $\{\boldsymbol{I}_p, \boldsymbol{I}_p, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_w\}$. $\boldsymbol{I}_p$ is the $p-$dimensional identity matrix. This model is akin to the univariate first oder model that we have seen above.

$$\boldsymbol{y}_t = \boldsymbol{\theta}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_v)$$
$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_w)$$

The posterior is then given by,

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_w|Y_{1:n}) \propto \left(\frac{1}{\sqrt{|\boldsymbol{\Sigma}_v|}}\right)^n \exp\Big(\frac{\sum_{i=1}^{T}(\boldsymbol{y}_i-\boldsymbol{\theta}_i)^\top \boldsymbol{\Sigma}_v^{-1}(\boldsymbol{y}_i-\boldsymbol{\theta}_i)}{-2}\Big)$$
$$\times \left(\frac{1}{\sqrt{|\boldsymbol{\Sigma}_w|}}\right)^n \exp\Big(\frac{\sum_{i=1}^{n}(\boldsymbol{\theta}_i-\boldsymbol{\theta}_{i-1})^\top \boldsymbol{\Sigma}_w^{-1}(\boldsymbol{\theta}_i-\boldsymbol{\theta}_{i-1})}{-2}\Big)$$
$$\times \pi_0(\boldsymbol{\Sigma}_v)\pi_0(\boldsymbol{\Sigma}_w)$$

Choosing that $\pi_0(\boldsymbol{\Sigma}_v)$ and $\pi_0(\boldsymbol{\Sigma}_w)$ are Inverse-Wishart priors with parameters, $(\nu_v, \boldsymbol{\Psi}_v)$ and $(\nu_w, \boldsymbol{\Psi}_w)$ respectively we can do a conjugate analysis for $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$. Where.

$$\pi_0(\boldsymbol{\Sigma}_v) \propto \exp\left(-\frac{1}{2} tr(\boldsymbol{\Psi}_v \boldsymbol{\Sigma}_v^{-1}) \right) \left(\frac{1}{|\boldsymbol{\Sigma}_v|} \right)^{(\nu_v + d_v + 1)/2}$$

$$\pi_0(\boldsymbol{\Sigma}_w) \propto \exp\left(-\frac{1}{2} tr(\boldsymbol{\Psi}_w \boldsymbol{\Sigma}_w^{-1}) \right) \left(\frac{1}{|\boldsymbol{\Sigma}_w|} \right)^{(\nu_w + d_w + 1)/2}$$

Given the form of the posterior this is how we would apply Algorithm 2 to simulate from the posterior distribution. As in the univariate case above, we start with some initial values for $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$ use the FFBS algorithm to draw an instance of the state, use these state vectors to update $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$ and then use these updated variance parameters to do the next FFBS update of the state and so on and so forth.

The full conditionals for the variance-covariance parameters are given as,

$$\begin{aligned}
p(\boldsymbol{\Sigma}_v | \boldsymbol{\Sigma}_w, Y_{1:n}, \boldsymbol{\theta}) &\propto \exp\left(\frac{\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\theta}_i)^\top \boldsymbol{\Sigma}_v^{-1} (\boldsymbol{y}_i - \boldsymbol{\theta}_i) + tr(\boldsymbol{\Psi}_v \boldsymbol{\Sigma}_v^{-1})}{-2} \right) \left(\frac{1}{|\boldsymbol{\Sigma}_v|} \right)^{(n + \nu_v + d_v + 1)/2} \\
&\propto \exp\left(\frac{\sum_{i=1}^n tr[(\boldsymbol{y}_i - \boldsymbol{\theta}_i)(\boldsymbol{y}_i - \boldsymbol{\theta}_i)^\top \boldsymbol{\Sigma}_v^{-1}] + tr(\boldsymbol{\Psi}_v \boldsymbol{\Sigma}_v^{-1})}{-2} \right) \left(\frac{1}{|\boldsymbol{\Sigma}_v|} \right)^{(n + \nu_v + d_v + 1)/2} \\
&\propto \exp\left(\frac{tr(\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\theta}_i)(\boldsymbol{y}_i - \boldsymbol{\theta}_i)^\top + \boldsymbol{\Psi}_v) \boldsymbol{\Sigma}_v^{-1})}{-2} \right) \left(\frac{1}{|\boldsymbol{\Sigma}_v|} \right)^{(n + \nu_v + d_v + 1)/2} \\
&\sim Inv - Wishart(T + \nu_v, \sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\theta}_i)(\boldsymbol{y}_i - \boldsymbol{\theta}_i)^\top + \boldsymbol{\Psi}_v)
\end{aligned}$$

A similar result holds for $\boldsymbol{\Sigma}_w$.

CHAPTER 4
Conditional Heteroskedasticity:

4.1 ARCH(p) and GARCH(p, q) Errors:

Autoregressive conditional heteroscedasticity (ARCH) models, first proposed by Engle (1982), and their extensions the generalized ARCH (GARCH), proposed by Bollerslev (1986), have proven to be very successful in modeling volatility of financial time series. A good survey of these types of models can be found in Bollerslev et al. (1994). GARCH models help explain the thick tailed distribution as seen in many financial data. The main idea is that conditional heteroskedacticity allows us to incorporate information from the past into the variance of the present.

A GARCH(p, q) process is defined by the following equations,

$$\begin{aligned} &X_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim N(0,1) \\ &\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + ... + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2, \qquad \alpha_i, \beta_j \in \mathbb{R}^+. \end{aligned} \tag{4.1}$$

A GARCH($p, 0$) process, i.e if $\beta_j = 0 \ \forall \ j$, is the equivalent of an ARCH(p) process. We expand on some of the properties of a GARCH(p, q) process. We would like to point out that in (4.1), ϵ_t need not be a standard normal random variable, it may have a Student's-T distribution, for example. However, for our purpose we assume that it has a standard normal distribution.

Let $m = max(p, q)$,

$$
\begin{aligned}
& X_t^2 - \sigma_t^2 = \sigma_t^2(\epsilon_t^2 - 1) := Z_t \\
\implies & X_t^2 - (\alpha_0 + \alpha_1 X_{t-1}^2 + ... + \alpha_m X_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_m \sigma_{t-m}^2) = Z_t \\
\implies & X_t^2 - \alpha_1 X_{t-1}^2 - ... - \alpha_m X_{t-m}^2 - \beta_1 \sigma_{t-1}^2 - ... - \beta_m \sigma_{t-m}^2 = \alpha_0 + Z_t \\
\implies & X_t^2 - \alpha_1 X_{t-1}^2 - ... - \alpha_m X_{t-m}^2 - \beta_1 (X_{t-1}^2 - Z_{t-1}) - ... - \beta_m (X_{t-1}^2 - Z_{t-m}) = \alpha_0 + Z_t \\
\implies & X_t^2 - (\alpha_1 + \beta_1) X_{t-1}^2 - ... - (\alpha_m + \beta_m) X_{t-m}^2 = \alpha_0 + Z_t + \beta_1 Z_{t-1} + ... + \beta_m Z_{t-m}.
\end{aligned}
$$

Here we have that $\alpha_i = 0$ for $i > p$ and $\beta_j = 0$ for $j > q$. Now it just remains to be verified that Z_t is white noise and then we have that X_t^2 is a standard ARMA(p,q) model. First define $\mathcal{F}_{t-1}^X$ to be an information set containing all the information up to time $t-1$. Notice that due to the assumption of normality of ϵ_t we have that,

$$
\begin{aligned}
X_t | \mathcal{F}_{t-1}^X &\sim N(0, \sigma_t^2) \\
\implies E(X_t^2 | \mathcal{F}_{t-1}^X) &= \sigma_t^2.
\end{aligned}
\tag{4.2}
$$

Thus X_t is conditionally Gaussian, where the conditioning is on all the previous information available. Next to show that Z_t is a white noise process observe the following,

$$
\begin{aligned}
E[Z_t] &= E[E[Z_t | \mathcal{F}_{t-1}^X]] \\
&= E[\sigma_t^2 E[\epsilon_t^2 - 1 | \mathcal{F}_{t-1}^X]] \\
&= 0 \\
E(Z_t Z_{t+h}) &= E[E[Z_t Z_{t+h} | \mathcal{F}_{t+h-1}^X]] \\
&= E\big[Z_t \sigma_{t+h}^2 E[\epsilon_{t+h}^2 - 1 | \mathcal{F}_{t+h-1}^X]\big] \\
&= 0.
\end{aligned}
$$

We also have that for a stationary GARCH(p, q), the unconditional variance of the process is then given by,

$$E[X_t^2 - (\alpha_1 + \beta_1)X_{t-1}^2 - ... - (\alpha_m + \beta_m)X_{t-m}^2] = E[\alpha_0 + Z_t - \beta_1 Z_{t-1} - ... - \beta_m Z_{t-m}]$$

$$\mu - \phi_1\mu - ... - \phi_p\mu = \alpha_0$$

$$\mu = \frac{\alpha_0}{1 - \phi_1 - ... - \phi_p},$$

where $\phi_i = \alpha_i + \beta_i$.

A sufficient and necessary condition for stationarity of the process is that $\sum_{i=1}^{m} \phi_i < 1$, see Bollerslev (1986) and Bollerslev et al. (1994) for more details on the properties of these models.

We mentioned earlier the *leptokurtic* properties of the GARCH model. For a GARCH$(1,1)$ process where the error terms are conditionally Gaussian, it is easliy seen that,

$$E(X_t^4|\mathcal{F}_{t-1}^X) = 3[E(X_t^2|\mathcal{F}_{t-1}^X)]^2 = 3(\sigma_t^2)^2.$$

For a GARCH$(1,1)$ process, where κ is the kurtosis, we have that,

$$E(X_t^4) = E[E[X_t^4|\mathcal{F}_{t-1}^X]]$$

$$\geq 3E[X_t^2]^2$$

$$\implies \kappa = \frac{E(X_t^4)}{[E(X_t)^2]^2} \geq 3.$$

With a bit more work we can actually see that,

$$\kappa = \frac{E[E[X_t^4|\mathcal{F}_{t-1}^X]]}{[E(X_t)^2]^2} = \frac{3E[(\sigma_t^2)^2]}{[E(X_t)^2]^2} = \frac{3Var[(\sigma_t^2)] + 3(E[(\sigma_t^2)])^2}{[E(X_t)^2]^2} = 3 + 3\frac{Var[E(X_t^2|\mathcal{F}_{t-1}^X)]}{[E(X_t)^2]^2}.$$

4.2 Asymmetric GARCH Model:

We can further generalize the GARCH(p,q) by allowing positive and negative innovations to have an asymmetric effect on the conditional variance. Take the following specification

for a GARCH$(1,1)$ process.

$$\begin{aligned} X_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_{11} X_{t-1}^2 \mathcal{I}_{X_t>0} + \alpha_{12} X_{t-1}^2 \mathcal{I}_{X_t<0} + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha_{11} X_{t-1}^2 \mathcal{I}_{\epsilon_t>0} + \alpha_{12} X_{t-1}^2 \mathcal{I}_{\epsilon_t<0} + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

In this case we are modeling the positive innovations to have a coefficient of $\alpha_{1,1}$ and the negative innovations to have a coefficient of $\alpha_{1,2}$. This specification was first introduced by Glosten et al. (1993) and is known as the GJR-GARCH$(1,1)$, named after the authors.

It is a simple exercise to show that under the weak stationarity assumption for σ_t^2 we have that,

$$\begin{aligned} & X_t^2 - \sigma_t^2 = \sigma_t^2(\epsilon_t^2 - 1) := W_t \\ & X_t^2 - (\alpha_0 + \alpha_{11} X_{t-1}^2 \mathcal{I}_{\epsilon_t>0} + \alpha_{12} X_{t-1}^2 \mathcal{I}_{\epsilon_t<0} + \beta_1 \sigma_{t-1}^2) = \sigma_t^2 \\ & X_t^2 - (\alpha_{11} X_{t-1}^2 \mathcal{I}_{\epsilon_t>0} + \alpha_{12} X_{t-1}^2 \mathcal{I}_{\epsilon_t<0}) + \beta_1 (X_{t-1}^2 - W_{t-1}) = \alpha_0 + W_t \\ & X_t^2 - (\alpha_{11} X_{t-1}^2 \mathcal{I}_{\epsilon_t>0} + \alpha_{12} X_{t-1}^2 \mathcal{I}_{\epsilon_t<0} + \beta_1 X_{t-1}^2) = \alpha_0 + W_t + \beta_1 W_{t-1} \\ \implies & E(X_t^2) - 0.5\alpha_{1,1} E(X_t^2) - 0.5\alpha_{1,2} E(X_t^2) + \beta E(X_{t-1}^2) = \alpha_0 \\ & E(X_t^2) = \frac{\alpha_0}{1 - 0.5\alpha_{1,1} - 0.5\alpha_{1,2} - \beta}. \end{aligned}$$

4.3 Estimating GARCH(p,q):

The GARCH(p,q) model is easy to estimate due to the form of the conditional likelihood.

$$\begin{aligned} L(\theta; X_1, ..., X_p, \sigma_{p-q+1}^2 = ... = \sigma_p^2 = 0) &= \prod_{t=p+1}^{n} p(X_t | X_{t-1:1}) \\ &= \prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\Big(\frac{X_t^2}{-2\sigma_t^2}\Big), \end{aligned} \tag{4.3}$$

where the unknown parameters to maximize the likelihood over are, $\theta = \alpha_0, ..., \alpha_p, \beta_1, ..., \beta_q$. However due to the non-linear nature of the likelihood in terms of the unknown parameter numerical methods are used. Notice from (4.3), σ_t^2 is a non-linear function in terms of α_0,

α_1 and β_1.

For a Bayesian estimation we would need to specify priors for the GARCH parameters and the posterior is then given by,

$$\pi(\alpha_0, \alpha_1, \beta_1 | X_{1:t}) \propto \prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi h_t}} \exp\Big(\frac{X_t^2}{-2h_t}\Big) \pi_0(\alpha_0)\pi_0(\alpha_1)\pi_0(\beta_1). \tag{4.4}$$

Given that the posterior distribution is a non-linear function in terms of α_0, α_1 and β_1, we cannot do a conjugate analysis and hence would have to use other simulation techniques to obtain posterior samples of the parameter.

4.4 Multivariate GARCH$(1,1)$:

Bollerslev et al. (1990) proposes a multivariate GARCH process known as the Constant Conditional Correlation GARCH process (CCC-GARCH). Suppose now that $\boldsymbol{X_t}$ is a $p \times 1$ dimensional process, where, the value of the i^{th} series at time t is $X_{i,t}$. The CCC-GARCH$(1,1)$ is formulated as the following,

$$\begin{aligned}
\boldsymbol{X}_t &= \boldsymbol{H}_t \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\boldsymbol{0}, \boldsymbol{I}_p) \\
\boldsymbol{H}_t^2 &= \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t', \quad \boldsymbol{H}_t = chol(\boldsymbol{H}_t^2) \\
\boldsymbol{D_t} &= diag(\sigma_{1,t}, ..., \sigma_{p,t}) \\
\boldsymbol{R_t} &= \begin{bmatrix} 1 & \rho_{1,2} & ... & \rho_{1,p} \\ \rho_{2,1} & 1 & ... & \rho_{2,p} \\ \vdots & & \ddots & \vdots \\ \rho_{p,1} & \rho_{p,2} & ... & 1 \end{bmatrix} \\
\sigma_{i,t}^2 &= \alpha_0^{(i)} + \alpha_1^{(i)} Y_{i,t-1}^2 + \beta_1^{(i)} \sigma_{i,t-1}^2.
\end{aligned} \tag{4.5}$$

Here $chol(\boldsymbol{H}_t^2)$ is the Cholesky decomposition of the matrix $\boldsymbol{H}_t^2$.

In this model specification what is essentially happening is that we have for $i = 1, ..., p$, the series $\boldsymbol{X}_i$ has a GARCH$(1, 1)$ process given by,

$$X_{i,t} = \sigma_{i,t}\epsilon_t$$
$$\sigma^2_{i,t} = \alpha_0^{(i)} + \alpha_1^{(i)} X^2_{i,t-1} + \beta_1^{(i)}\sigma^2_{i,t-1},$$

with the instantaneous correlation between $X_{i,t}$ and $X_{j,t}$ given by $\rho_{i,j}$, $i, j = 1, ..., p$ for all t.

The simplest case for this model is when $\rho_{i,j} = 0$, i.e. $\boldsymbol{R}_t$ is just the $p-$dimensional identity matrix. In that case we have that each of the series $\boldsymbol{X}_i$ is mutually independent from the other series. On the other hand when $\rho_{i,j}$ is not zero and is unknown, the parameter space increases quadratically. In a p dimensional model the unknown parameters are $\alpha_0^{(i)}, \alpha_1^{(i)}, \beta_1^{(i)}$ and $\{\rho_{i,j}\}^p_{j,i=1,j>i}$, which gives a total number of unknown parameters as $3p + p(p-1)/2$. In the bivariate case we have 7 unknown parameters which grows to 12 in the trivariate case.

However, despite the growth of the parameter space, this model is still quite flexible, because it allows us to apply restrictions or asymmetries on the individual GARCH series is desired. For example, suppose one hypothesizes that in a bivariate case, the first series has a GARCH$(1, 1)$ process and the second series has an asymmetric-GARCH process (as specified in the previous section), the CCC-GARCH model allows us to put in those restrictions.

4.5 Estimating the Multivariate GARCH$(1, 1)$ Model:

The estimation technique for the multivariate GARCH model is similar to that of the univariate GARCH model. The conditional likelihood is given by,

$$\mathcal{L}(\boldsymbol{X_{1:t}}; \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\rho}) = \prod_{i=1}^{t} \frac{1}{\sqrt{det(2\pi \boldsymbol{H}_i)}} \exp\Big(\frac{\boldsymbol{X}_i' \boldsymbol{H}_i^{-1} \boldsymbol{X}_i}{-2}\Big),$$

and is non-linear in the GARCH parameters and must be numerically optimized, which is a rather tall task given the dimensionality of the problem.

Similarly the Bayesian analysis can be done by specifying a prior distribution on the GARCH parameters and then sampling from the posterior using different sampling techniques. The posterior is given by,

$$\pi(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\rho}|X_{1:t}) \propto \prod_{i=1}^{t} \frac{1}{\sqrt{det(2\pi \boldsymbol{H}_i)}} \exp\Big(\frac{\boldsymbol{X}_i' \boldsymbol{H}_i^{-1} \boldsymbol{X}_i}{-2}\Big) \pi_0(\boldsymbol{\alpha}_0)\pi_0(\boldsymbol{\alpha}_1)\pi_0(\boldsymbol{\beta}_1).$$

CHAPTER 5
Extending the Gaussian SSM

5.1 State Space model with GARCH$(1,1)$ Errors:

In the previous chapters we discussed the Gaussian-SSMs and stadard GARCH models. In this chapter we extend the Gaussian-SSM to have GARCH errors at the observation level instead of a constant variance.

Suppose that y_t is a univariate time series with a state space representation specified by the design matrices $\{\boldsymbol{F}, \boldsymbol{G}\}$, a constant state level variance $\boldsymbol{W}$ and an observation variance which is a stationary GARCH$(1,1)$ process. The model may be written as,

$$\begin{aligned} y_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + z_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + w_t, \quad w_t \sim N(0, W) \\ z_t &= s_t \epsilon_t, \quad \epsilon_t \sim N(0,1) \\ s_t^2 &= \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 s_{t-1}^2. \end{aligned} \tag{5.1}$$

Given this specification we need to appropriately model the Kalman-Filter recursions to account for the changing variance at the observation level. This task turns out to be rather straight forward owing to the conditional nature of both the Kalman recursions and the GARCH errors, where, in both cases the conditioning at time t is done on all the previous information available.

Recall that the GARCH$(1,1)$ errors are conditionally Gaussian. The modified Kalman Recursion is given below. Suppose that we have the filtered distributions at time $t-1$, i.e.

$$\boldsymbol{\theta}_{t-1}|Y_{1:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$$

$$Y_{t-1}|\boldsymbol{\theta}_{t-1}, Y_{0:t-2} \sim N(\boldsymbol{F}'\boldsymbol{\theta}_{t-1}, S_{t-1}).$$

The one-step ahead distribution for the observation is given by, $Y_t|\boldsymbol{y}_{1:t-1} \sim N(f_t, Q_t)$. Where,

$$\begin{aligned}
f_t &= \boldsymbol{F}'\boldsymbol{a}_t \\
Q_t &= Var(Y_t|\boldsymbol{y}_{1:t-1}) \\
&= E(Var(Y_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) + Var(E(Y_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) \\
&= E(Var(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) + \boldsymbol{F}'\boldsymbol{R}_t\boldsymbol{F}.
\end{aligned}$$

The variance on the left hand term is given by,

$$\begin{aligned}
Var(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) &= E([z_t - E(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})]^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= E(z_t^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= E(s_t^2\epsilon_t^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= E(s_t^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= E(\alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 s_{t-1}^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= \alpha_0 + \alpha_1 E(z_{t-1}^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) + \beta_1 E(s_{t-1}^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) \\
&= \alpha_0 + \alpha_1 E((Y_{t-1} - \boldsymbol{F}'\boldsymbol{\theta}_{t-1})^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) + \beta_1 s_{t-1}^2 \\
&= \alpha_0 + \alpha_1 E(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{\theta}_{t-1} + \boldsymbol{F}'\boldsymbol{\theta}_{t-1}\boldsymbol{\theta}_{t-1}'\boldsymbol{F})|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) + \beta_1 s_{t-1}^2 \\
Var(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) &= \alpha_0 + \alpha_1(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{\theta}_{t-1} + \boldsymbol{F}'\boldsymbol{\theta}_{t-1}\boldsymbol{\theta}_{t-1}'\boldsymbol{F}) + \beta_1 s_{t-1}^2.
\end{aligned}$$

Next, taking the conditional expectation over this variance will give us,

$$\begin{aligned}
E(Var(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) &= E\Big[\alpha_0 + \alpha_1(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{\theta}_{t-1} + \boldsymbol{F}'\boldsymbol{\theta}_{t-1}\boldsymbol{\theta}_{t-1}'\boldsymbol{F}) + \beta_1 s_{t-1}^2\Big|\boldsymbol{y}_{1:t-1}\Big] \\
&= \alpha_0 + \alpha_1\Big(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{m}_{t-1} + \boldsymbol{F}'(\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1}\boldsymbol{m}_{t-1}')\boldsymbol{F}\Big) + \beta_1 s_{t-1}^2,
\end{aligned}$$

where we have that,

$$\begin{aligned} E(Var(z_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) &= E(Var(Y_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) \\ &= E(E(s_t^2|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})|\boldsymbol{y}_{1:t-1}) \\ &= E(s_t^2|\boldsymbol{y}_{1:t-1}) \\ &= s_t^2 \end{aligned} \tag{5.2}$$

$$\implies s_t^2 = \alpha_0 + \alpha_1\Big(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{m}_{t-1} + \boldsymbol{F}'(\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1}\boldsymbol{m}_{t-1}')\boldsymbol{F}\Big) + \beta_1 s_{t-1}^2.$$

Hence finally we have that,

$$Q_t = s_t^2 + \boldsymbol{F}'\boldsymbol{R}_t\boldsymbol{F}.$$

5.2 SSM with GARCH(p, q) Errors:

The details from the GARCH$(1, 1)$ case, allows us to quite easily extend the model to the GARCH(p, q) case. The GARCH(p, q) SSM is specified by,

$$\begin{aligned} y_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + z_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + w_t, \quad w_t \sim N(0, W) \\ z_t &= s_t\epsilon_t, \quad \epsilon_t \sim N(0, 1) \\ s_t^2 &= \alpha_0 + \alpha_1 z_{t-1}^2 + ... + \alpha_p z_{t-p}^2 + \beta_1 s_{t-1}^2 + ... + \beta_q s_{t-q}^2 \end{aligned} \tag{5.3}$$

Suppose that at time $t-1$ we have that,

$$\boldsymbol{\theta}_{t-1}|Y_{1:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$$

$$Y_{t-1}|\boldsymbol{\theta}_{t-1}, Y_{0:t-2} \sim N(\boldsymbol{F}'\boldsymbol{\theta}_{t-1}, S_{t-1})$$

Then following the same construction as in the GARCH$(1, 1)$ case, we have that, the Kalman Filter recursion needs to be modified only at the point of the one-step-ahead distribution. This is given by, $Y_t|\boldsymbol{y}_{1:t-1} \sim N(f_t, Q_t)$, where,

$$
\begin{aligned}
f_t &= \boldsymbol{F}'\boldsymbol{a}_t \\
Q_t &= Var(Y_t|\boldsymbol{y}_{1:t-1}) \\
&= s_t^2 + \boldsymbol{F}'\boldsymbol{R}_t\boldsymbol{F} \\
s_t^2 &= \alpha_0 + \alpha_1\Big(Y_{t-1}^2 - 2Y_{t-1}\boldsymbol{F}'\boldsymbol{m}_{t-1} + \boldsymbol{F}'(\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1}\boldsymbol{m}_{t-1}')\boldsymbol{F}\Big) + ... \\
&\quad + \alpha_p\Big(Y_{t-p}^2 - 2Y_{t-p}\boldsymbol{F}'\boldsymbol{m}_{t-p} + \boldsymbol{F}'(\boldsymbol{C}_{t-p} + \boldsymbol{m}_{t-p}\boldsymbol{m}_{t-p}')\boldsymbol{F}\Big) + \beta_1 s_{t-1}^2 + \\
&\quad ... + \beta_q s_{t-q}^2
\end{aligned}
$$

5.3 Multivariate SSM with GARCH Error:

Once we see the details in the case where the observation is univariate, the extension to the case where the observation is multivariate becomes straight forward. Here we have that the one-step ahead prediction and error term is given by,

$$
\begin{aligned}
\boldsymbol{f}_t &= \boldsymbol{F}'\boldsymbol{a}_t \\
\boldsymbol{Q}_{\boldsymbol{t}} &= \boldsymbol{S}_t^2 + \boldsymbol{F}'\boldsymbol{R}_t\boldsymbol{F}
\end{aligned}
$$

where,

$$
\begin{aligned}
&\boldsymbol{S}_t^2 = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t', \quad \boldsymbol{S}_t = chol(\boldsymbol{H}_t^2) \\
&\boldsymbol{D}_{\boldsymbol{t}} = diag(S_{1,t}, ..., S_{p,t}) \\
&\boldsymbol{R}_{\boldsymbol{t}} = \begin{bmatrix} 1 & \rho_{1,2} & ... & \rho_{1,p} \\ \rho_{2,1} & 1 & ... & \rho_{2,p} \\ \vdots & & \ddots & \vdots \\ \rho_{p,1} & \rho_{p,2} & ... & 1 \end{bmatrix} \\
&S_{i,t}^2 = \alpha_0^{(i)} + \alpha_1^{(i)}\Big(Y_{i,,t-1}^2 - 2Y_{i,t-1}(\boldsymbol{F}'\boldsymbol{m}_{t-1})_{(i)} + [\boldsymbol{F}'(\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1}\boldsymbol{m}_{t-1}')\boldsymbol{F}]_{(i)}\Big) + \beta_1 S_{i,t-1}^2
\end{aligned}
\tag{5.4}
$$

5.4 Bayesian Posterior Analysis:

We take the SSM with a univariate GARCH$(1,1)$ error at the observational level given in (5.1). The posterior distribution is then given by,

$$\begin{aligned} p(\boldsymbol{\theta}_{0:n}, \alpha_0, \alpha_1, \beta_1, \boldsymbol{W} | \boldsymbol{y}_{1:n}) &\propto \pi_0(\boldsymbol{W})\pi_0(\alpha_0, \alpha_1, \beta_1) \prod_{i=1}^{n} p(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{i-1}, \boldsymbol{W}) p(\boldsymbol{y}_i | \boldsymbol{\theta}_i, \alpha_0, \alpha_1, \beta_1) \\ &\propto \pi_0(\boldsymbol{W})\pi_0(\alpha_0, \alpha_1, \beta_1) \prod_{i=0}^{n} \left(\frac{1}{\sqrt{s_i^2}} \right) \exp\Big(\frac{(Y_i - \boldsymbol{F'\theta}_i)^2}{-2s_i^2} \Big) \times \\ &\prod_{i=0}^{n} \left(\frac{1}{|\boldsymbol{W}|} \right) \exp\Big(\frac{\sum_{i=1}^{n} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})^\top \boldsymbol{W}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})}{-2} \Big). \end{aligned}$$

By choosing the prior on $\boldsymbol{W}$ to be an Inverse-Wishart prior, allows us to do a conjugate analysis on the $\boldsymbol{W}$. The Inverse-Wishart prior with parameters degrees of freedom parameter ν and scale parameter $\boldsymbol{\Psi}$ is given by,

$$\pi_0(\boldsymbol{W}) \propto \exp\left(-\frac{1}{2} tr(\boldsymbol{\Psi W}^{-1}) \right) \left(\frac{1}{|\boldsymbol{W}|} \right)^{(\nu+d+1)/2}$$

where d is the dimension of $\boldsymbol{W}$. The full conditional distribution for $\boldsymbol{W}$ is then given by,

$$\begin{aligned} p(\boldsymbol{W} | \boldsymbol{\theta}_{0:n}, \alpha_0, \alpha_1, \beta_1, \boldsymbol{y}_{1:n}) &\propto \pi_0(\boldsymbol{W}) \prod_{i=0}^{n} \left(\frac{1}{|\boldsymbol{W}|} \right) \exp\Big(\frac{\sum_{i=1}^{n} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})^\top \boldsymbol{W}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})}{-2} \Big) \\ &\propto \left(-\frac{1}{2} tr\Big(\left[\sum_{i=1}^{n} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})^\top (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1}) + \boldsymbol{\Psi} \right] \boldsymbol{W}^{-1} \Big) \right) \left(\frac{1}{|\boldsymbol{W}|} \right)^{(n+\nu+d+1)/2} \\ &\propto Inv - Wishart(n + \nu, \sum_{i=1}^{n} (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1})^\top (\boldsymbol{\theta}_i - \boldsymbol{G\theta}_{i-1}) + \boldsymbol{\Psi}). \end{aligned}$$

On the other hand for the other parameters of the model, $\alpha_0, \alpha_1, \beta_1$, the full conditional distribution is not of some recognizable form and so we must use the Metropolis-Hastings algorithm, with a multivariate random walk proposal, to sample from it. We can choose independent Half-Cauchy priors for the $\alpha_0, \alpha_1, \beta_1$ parameters. The conditional posterior for the GARCH parameters is given by,

$$p(\alpha_0, \alpha_1, \beta_1 | \boldsymbol{W}, \boldsymbol{\theta}_{0:n}, \boldsymbol{y}_{1:n}) \propto \pi_0(\alpha_0, \alpha_1, \beta_1) \prod_{i=0}^{n} \left(\frac{1}{\sqrt{s_i^2}} \right) \exp\Big(\frac{(Y_i - \boldsymbol{F'\theta}_i)^2}{-2s_i^2} \Big)$$

We can see that this conditional posterior is not any easily recognized distribution and hence, to draw posterior samples of the GARCH parameters we must use the Metropolis-Hastings algorithm. It must also be noted that stationarity assumptions for the GARCH$(1,1)$ model imply that $\alpha_1 + \beta_1 < 1$ and this information can and should be incorporated into the prior information if so desired.

Finally to obtain posterior samples of the unknown state vectors we use FFBS, as we did in Algorithm 2. All of the above steps combined now allow us to estimate the model parameters.

CHAPTER 6
Simulation Study:

We run a few simulation studies of DLM's to show that we are able to estimate the parameters of the model well.

6.1 Symmetric GARCH-SSM:

We simulate a symmetric GARCH-SSM with different sample sizes. The estimates of the model parameters for different sample sizes, based on 4500 draws from the posterior distribution are given below. Below in brackets are the variances of the posterior samples. The true value of the parameters are: $\alpha_0 = 4, \alpha_1 = 0.1, \gamma_1 = 0.1, \beta_1 = 0.8, W = 0.2$.

Sample size	α_0	α_1	γ_1	β_1	W
n = 500	3.52	0.124	0.129	0.794	0.105
	(4.407)	(0.0022)	(0.0029)	(0.0048)	(0.001)
n = 2000	5.56	0.119	0.109	0.752	0.064
	(2.67)	(0.0008)	(0.0008)	(0.003)	(0.0002259)
n = 5000	3.94	0.113	0.0957	0.801	0.166
	(0.419)	(0.002)	(0.0002)	(0.0005)	(0.00066)

Table 6–1: Posterior median estimates of the parameter of a univariate **symmetric** GARCH-SSM.

The prior specification for the model is,

$$\pi_0(\alpha_0) \sim Cauchy(0,1)\mathcal{I}_{\alpha_0>0}$$

$$\pi_0(\alpha_1) \sim Cauchy(0,1)\mathcal{I}_{\alpha_1>0}$$

$$\pi_0(\beta_1) \sim Cauchy(0,1)\mathcal{I}_{\beta_1>0}$$

$$\pi_0(W) \sim IG(10,10)$$

The priors are kept quite vague. The choice of the Inverse-Gamma prior allows us to do a conjugate analysis for the parameter W. Due to the form of the posterior distribution for the GARCH parameters to sample from the posterior we use the Metropolis-Hastings algorithm. As the sample size grows we are better able to retrieve the true value of the parameters.

6.2 Asymmetric GARCH-SSM:

The next model that we simulate is a univariate SSM with asymmetric GARCH$(1,1)$ errors. Since by default a symmetric GARCH model is nested in the asymmetric model, if we are able to estimate the parameters of the asymmetric then we will be fine for the symmetric model. The parameter estimates for the asymmetric model based on 4500 draws from the posterior distribution are given below. Below in brackets are the variances of the posterior samples. The true value of the parameters are: $\alpha_0 = 4, \alpha_1 = 0.4, \gamma_1 = 0.1, \beta_1 = 0.2, W = 0.01$.

Sample size	α_0	α_1	γ_1	β_1	W
n = 500	2.82	0.276	0.092	0.433	0.0146
	(1.90)	(0.0122)	(0.00395)	(0.0503)	(0.0000237)
n = 2000	3.841	0.409	0.111	0.227	0.00812
	(0.247)	(0.00378)	(0.00163)	(0.00569)	0.000003
n = 5000	3.70	0.489	0.081	0.228	0.0128
	(0.069)	(0.002)	(0.0005)	(0.002)	(0.00004)

Table 6–2: Posterior median estimates of the parameter of a univariate **asymmetric** GARCH-SSM model.

We used the following priors for the model,

$$\pi_0(\alpha_0) \sim Cauchy(0,1)\mathcal{I}_{\alpha_0>0}$$

$$\pi_0(\alpha_1) \sim Cauchy(0,1)\mathcal{I}_{\alpha_1>0}$$

$$\pi_0(\gamma_1) \sim Cauchy(0,1)\mathcal{I}_{\gamma_1>0}$$

$$\pi_0(\beta_1) \sim Cauchy(0,1)\mathcal{I}_{\beta_1>0}$$

$$\pi_0(W) \sim IG(10,10)$$

We note that the priors for all the parameters are fairly vague. We have that the choice of the Inverse-Gamma prior for W leads to a conjugate analysis for W. From Table 2 we see that we are able to estimate the parameters of the model quite well. We use the Metropolis-Hastings algorithm to sample from the posterior distribution. As in the symmetric GARCH-SSM case the parameter estimates are better for larger sample sizes.

6.3 Bivariate Gaussian SSM:

We start with a simple bi-variate model.Since we have a Gaussian SSM by choosing Inverse-Wishart priors we can do a conjugate analysis. For both $\boldsymbol{V}$ and $\boldsymbol{W}$ we choose an $Inv-$

$Wishart(10, diag(10, 10))$ prior, which acts a weakly informative prior.

$$\boldsymbol{V} = \begin{bmatrix} 2 & -0.5 \\ -0.5 & 2 \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

In each case we have convergence, we ran 10000 simulations, with a burn in of 2000, and thinning of 10 (R seed = 201901).

Sample size	V_{11}	V_{22}	V_{12}	W_{11}	W_{22}	W_{12}
n = 500	2.05	1.50	-0.544	1.30	1.32	0.697
	(0.044)	(0.0347)	(0.022)	(0.0381)	(0.055)	(0.0299)
n = 2000	1.94	1.83	-0.523	1.26	0.986	0.494
	(0.0105)	(0.0088)	(0.0049)	(0.0093)	(0.0073)	(0.00394)
n = 5000	1.91	1.96	-0.447	1.182	0.935	0.440
	(0.0042)	(0.0038)	(0.00190)	(0.0032)	(0.00264)	(0.0013)

Table 6–3: Posterior median estimates of the parameter of a bi-variate Gaussian-SSM for different sample sizes. Below in brackets are the variances of the posterior samples.

We can see in Table 6-3, that as the sample size grows we are able to retrieve the true values of the parameters.

6.4 Bivariate GARCH-SSM:

Now we simulate a Bivariate GARCH-SSM. We use the CCC-GARCH specification given in the previous chapter. the model specification is,

$$\alpha_0^1 = 1, \alpha_1^1 = 0.1, \beta_1^2 = 0.8, W_{1,1} = 2$$

$$\alpha_0^2 = 5, \alpha_1^2 = 0.4, \beta_1^2 = 0.1, W_{2,2} = 0.5$$

$$\rho = -0.5, W_{1,2} = 0.$$

Sample size	α_0^1	α_1^1	β_1^1	W_{11}	ρ
	1.155	0.084	0.818	1.639	-0.56
	(0.3331)	(0.0011)	(0.0037)	(0.1301)	(0.0017)
n = 500	α_0^2	α_1^2	β_1^2	W_{22}	W_{12}
	4.210	0.274	0.231	0.831	0.174
	(1.0833)	(0.0049)	(0.0173)	(0.0437)	(0.0172)
	α_0^1	α_1^1	β_1^1	W_{11}	ρ
	1.667	0.092	0.742	2.143	-0.495
	(0.2287)	(0.0004)	(0.0031)	(0.0405)	(0.0004)
n = 2000	α_0^2	α_1^2	β_1^2	W_{22}	W_{12}
	4.557	0.269	0.235	0.546	0.084
	(0.4395)	(0.0016)	(0.0065)	(0.0056)	(0.0022)
	α_0^1	α_1^1	β_1^1	W_{11}	ρ
	0.844	0.052	0.861	2.092	-0.507
	(0.0649)	(0.0001)	(0.0009)	(0.0163)	(0.0002)
n = 5000	α_0^2	α_1^2	β_1^2	W_{22}	W_{12}
	4.973	0.270	0.164	0.879	0.153
	(0.2514)	(0.0006)	(0.0035)	(0.0073)	(0.0019)

Table 6–4: Posterior median estimates of the Bivariate GARCH-SSM for different sample sizes. Below in brackets are the variances of the posterior samples.

We used the following priors for the model,

$$\pi_0(\alpha_0^i) \sim Cauchy(0,1)\mathcal{I}_{\alpha_0^i>0}$$
$$\pi_0(\alpha_1^i) \sim Cauchy(0,1)\mathcal{I}_{\alpha_1^i>0}$$
$$\pi_0(\gamma_1^i) \sim Cauchy(0,1)\mathcal{I}_{\gamma_1^i>0}$$
$$\pi_0(\beta_1^i) \sim Cauchy(0,1)\mathcal{I}_{\beta_1^i>0}$$
$$\pi_0(\boldsymbol{W}) \sim Inverse - Wishart(10,10)$$
$$\pi_0(\rho) \sim Unif(-1,1).$$

6.5 Model Selection:

A key feature of estimating parameters is also deducing redundancies in the models. Model selection allows us to choose the best model given a choice of competing models. Particularly in our case we have three competing models, a Gaussian-SSM, a GARCH-SSM and an asymmetric-GARCH-SSM.

One of the main tools for model selections is the information criterion, in particular Akaike's information criteria (AIC), Bayes (or Schwarz') Information Criteria (BIC) use the maximum likelihood estimates and the log predictive distribution. Both AIC and BIC penalize for model complexity. We refer the reader Konishi and Kitagawa (2008) for more details.

Following Gelman et al. (2013) we use an information criteria proposed by Watanabe (2010), known as the Watanabe-Akaike Information criteria also known as the widely-applicable information Criteria (WAIC) to select our model. As mentioned in Gelman et al. (2013), the WAIC is a more Bayesian approach to the out of sample prediction performance and allows to assess the performance of our model using the posterior samples that we have drawn.

$$WAIC = lppd - p_{waic}$$

Where *lppd* is the *log pointwise predictive density* and p_{waic} is a bias correction. The *lppd* is given by,

$$\begin{aligned} lppd &= \sum_{i=1}^{n} \log \int p(y_i|\boldsymbol{\psi}) dp_{post}(\boldsymbol{\psi}) = \sum_{i=1}^{n} \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big) \\ p_{waic} &= 2 \sum_{i=1}^{n} \left\{ \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big) - \Big(E_{post} \log[p(y_i|\boldsymbol{\psi})] \Big) \right\} \end{aligned} \tag{6.1}$$

$$WAIC = 2 \sum_{i=1}^{n} \Big(E_{post} \log[p(y_i|\boldsymbol{\psi})] \Big) - \sum_{i=1}^{n} \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big)$$

From Gelman et al. (2013), we have that -2 times the $WAIC$ in (6.1) brings it on the deviance scale. This allows for a more straight forward comparison with the AIC and DIC. We do not however do this in our case. For us, a larger WAIC value implies a better model fit.

Here $\boldsymbol{\psi}$ is the unknown set of parameters. In our case we have that the unknown parameters is the set observation variances and the state variance. The integrations in (6.1) may be intractable since we have to integrate over the posterior distribution. However, the consistency of the sample means and the continuity of the *log*-function gives us that,

$$\begin{aligned} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i|\boldsymbol{\theta}_s) \right) &\xrightarrow{p} \log \Big(E_{post} p(y_i|\boldsymbol{\theta}) \Big) \\ \left(\frac{1}{S} \sum_{s=1}^{S} \log(p\big(y_i|\boldsymbol{\theta}_s)\big) \right) &\xrightarrow{p} E_{post} \log \Big(p(y_i|\boldsymbol{\theta}) \Big) \end{aligned}$$

This gives the computed $WAIC$ as,

$$cWAIC = 2 \sum_{i=1}^{n} \left(\frac{1}{S} \sum_{s=1}^{S} \log(p\big(y_i|\boldsymbol{\theta}_s)\big) \right) - \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i|\boldsymbol{\theta}_s) \right)$$

and $cWAIC$ converges in probability to $WAIC$ as S goes to infinity., where S is the size of the posterior sample.

In the following exercise we use the WAIC to select between a symmetric and an asymmetric GARCH model. We estimate both models and calculate the WAIC. The values are given below in Table 6-5. The true model was a symmetric GARCH model which has the following parameters $\alpha_0 = 4, \alpha_1 = 0.4, \beta_1 = 0.1, W = 1$. We see that in each case the symmetric model has a higher WAIC than the asymmetric one. Hence we will be choosing the symmetric model as the most appropriate model.

WAIC	$n = 500$	$n = 2000$	$n = 5000$
Symmetric	-1299.8	-5243.1	-13133.4
Asymmetric	-1301.5	-5244.5	-13135.0

Table 6–5: The WAIC was calculated for different sample sizes. The true model is a symmetric GARCH model.

In the following example we have that the true model is an asymmetric GARCH model given by the following parameters $\alpha_0 = 4, \alpha_1 = 0.4, \gamma_1 = 0.1, \beta_1 = 0.1, W = 1$. In this case we have that, for the larger sample sizes, the WAIC for the asymmetric model is larger than that of the symmetric model. Hence we are correctly choosing the asymmetric model as the most appropriate model for the larger sample sizes. However, for the smallest sample size, $n = 500$ we have that the symmetric model has the higher WAIC. Hence for the smallest sample size we would be choosing the incorrect model. In this case the performance of the WAIC has improved with sample size.

WAIC	$n = 500$	$n = 2000$	$n = 5000$
Symmetric	-1249.6	-5007.9	-12669.5
Asymmetric	-1252.5	-5002.6	-12652.5

Table 6–6: The WAIC was calculated for different sample sizes. The true model is an asymmetric GARCH model.

CHAPTER 7
Multiple Regime SSM:

So far in our study we have looked at SSMs whose parameters stay constant over time, i.e. without any structural breaks. One reason for this assumption is that it makes the analysis easier since we do not need to know how or when the parameters are evolving over time. In many cases we actually may know at what time point the parameters are changing. In Section 1, the equations of the Kalman Filter recursions are general enough to allow for time varying variance.

Suppose that we have a univariate time series from time $t = 1$ to $t = T$. If the model parameters do not vary over time, we say there is no change point (alternatively there is only one **regime** the model is operating under). We say there are two regimes if from some $k \in [1, T]$, between the time interval $t = [1, k]$ we have one set of parameters and between $t = [k + 1, T]$ we have another set of parameters. Here at $t = k$ we have a change point. More generally speaking if we have r regimes then each of the time intervals, $[1, k_1], [k_1+1, k_2], [k_2+1, k_3]..., [k_{r-1}+1, T]$ we have different parameters, where $k_1, k_2, ..., k_{r-1}$ are the change points. We take $k_0 = 0$ and $k_r = T$.

7.1 The Time-Varying SSM:

For simplicity let us assume that we have a univariate Gaussian SSM. In the single *regime* case we have that the parameters of the model are the unknown variances V and W. Now suppose that the model has 3 regimes. This means that under each of the regimes we have a different variances V_1, W_1 under the first regime, V_2, W_2 under the second regime and V_3, W_3 under the third regime. Hence the augmented parameter space of the three regime model is $\{V_1, W_1, V_2, W_2, V_3, W_3\}$. For the general r regime model we have the augmented parameter

space as, $\{V_1, W_1, ..., V_r, W_r\}$. In this case the model is specified as,

$$\begin{aligned} y_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + v_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \end{aligned} \tag{7.1}$$

where, the observation and state errors are distributed in the following manner,

$$v_t \sim \begin{cases} N(0, V_1), & k_0 \leq t < k_1 \\ \vdots & \vdots \\ N(0, V_r), & k_{r-1} \leq t < k_r \end{cases} \qquad \boldsymbol{w}_t \sim \begin{cases} N(0, \boldsymbol{W}_1), & k_0 \leq t < k_1 \\ \vdots & \vdots \\ N(0, \boldsymbol{W}_r), & k_{r-1} \leq t < k_r \end{cases}. \tag{7.2}$$

Here the change points $k_0, ..., k_r$ are assumed to be known. Thus, by correctly inputing the proper state and observation variances into the Kalman Filter we can carry on with the usual analysis that we have been doing so far.

An easy way to think of this model is to think of it as multiple pieces of data. Where $Y^1 = y_{1:k_1-1}$, $Y^2 = y_{k_1:k_2-1}$, ..., $Y^r = y_{k_{r-1}:T}$. Here $\{y_t\}$ is the original data. The individual series $Y^1, ..., Y_r$ glue together make up the original series $\{y_t\}$. Before mentioning the posterior distribution we describe the prior distribution for the parameters.

7.2 Estimating the Time-Varying SSM:

In the univariate single regime case we had chosen an Inverse-Gamma prior for V. If the state vector is univariate we also chose an Inverse-Gamma prior for W and if the state was multivariate we chose an Inverse-Wishart prior for $\boldsymbol{W}$. Using the same logic, for the multi-regime case we also choose Inverse-Gamma prior for the observation variances and an Inverse-Gamma prior or an Inverse-Wishart prior for the state variances for each of the regimes. We further assume that mutual independence between the priors, i.e. $V_i \perp V_j$, $\boldsymbol{W}_i \perp \boldsymbol{W}_j$ and $V_i \perp \boldsymbol{W}_j$ for all $i, j \in \{1, ..., r\}$ (To keep things consistent with the model

stated in (24) we keep the state errors as multivariate). This leads us to,

$$\pi_0(V_1, \boldsymbol{W}_1, ..., V_r, \boldsymbol{W}_r) = \pi_0(V_1)\pi_0(\boldsymbol{W}_1) \times ... \times \pi_0(V_r)\pi_0(\boldsymbol{W}_r).$$

The posterior distribution can then be written as,

$$\begin{aligned}
\pi_n(V_{1:r}, \boldsymbol{W}_{1:r}, \boldsymbol{\theta}|Y_{1:n}) &\propto \prod_{i=1}^{n} \pi(y_i|V_{1:r}, \boldsymbol{\theta}_i) \prod_{i=1}^{n} \pi(\boldsymbol{\theta}_i|\boldsymbol{W}_{1:r}, \boldsymbol{\theta}_{i-1})\pi_0(V_1, \boldsymbol{W}_1, ..., V_r, \boldsymbol{W}_r) \\
&\propto \prod_{i=1}^{k_1-1} \pi(y_i|V_1, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\boldsymbol{W}_1, \boldsymbol{\theta}_{i-1})\times \\
&\quad \prod_{i=k_1}^{k_2-1} \pi(y_i|V_2, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\boldsymbol{W}_2, \boldsymbol{\theta}_{i-1}) \times ...\times \\
&\quad \prod_{i=k_{r-1}}^{T} \pi(y_i|V_r, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\boldsymbol{W}_r, \boldsymbol{\theta}_{i-1}) \times \pi_0(V_1)\pi_0(\boldsymbol{W}_1) \times ... \times \pi_0(V_r)\pi_0(\boldsymbol{W}_r),
\end{aligned} \tag{7.3}$$

we can see from above that the posterior may be factorized based on the different regimes. This gives us the following conditional distributions,

$$\begin{aligned}
\pi_n(V_j|V_{(-j)}, \boldsymbol{W}_{1:r}, \boldsymbol{\theta}, Y_{1:n}) &\propto \pi_0(V_j) \prod_{i=k_{j-1}}^{k_j-1} \pi(y_i|V_j, \boldsymbol{\theta}_i) \\
\pi_n(\boldsymbol{W}_j|V_{1:r}, \boldsymbol{W}_{(-j)}, \boldsymbol{\theta}, Y_{1:n}) &\propto \pi_0(\boldsymbol{W}_j) \prod_{i=k_{j-1}}^{k_j-1} \pi(\boldsymbol{\theta}_i|\boldsymbol{W}_j, \boldsymbol{\theta}_{i-1}).
\end{aligned} \tag{7.4}$$

Thus the conditional distributions are quite simplified. On a final note, one can easily implement the multiple regime Gaussian SSM for multivariate observations.

Given the posterior distribution in (7.4) the analysis of the parameters $V_1, \boldsymbol{W}_1, ..., V_r\boldsymbol{W}_r$ maybe carried using Algorithm 2, i.e. we run the FFBS to get a sample of the state vectors and then conditional of this sample of the state vectors we draw the posterior samples of the variance parameters.

7.3 The Time-Varying GARCH-SSM:

Now we can formulate the GARCH-SSM to have time varying GARCH parameters. This is really quite easy to implement and the Kalman recursions for the time varying GARCH SSM are given next. For simplicity suppose that we have a univariate GARCH-SSM from time $t = 1, ..., T$. Again suppose that the change points are at time $k_1, k_2, ..., k_{r-1}$ and they are known. We assume a GARCH$(1, 1)$ model at the observation level.

$$\begin{aligned} y_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + z_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t \\ z_t &= s_t\epsilon_t, \quad \epsilon_t \sim N(0,1) \end{aligned} \tag{7.5}$$

$$s_t^2 = \begin{cases} \alpha_0^1 + \alpha_1^1 z_{t-1}^2 + \beta_1^1 s_{t-1}^2, & k_0 \le t < k_1 \\ \vdots & \vdots \\ \alpha_0^r + \alpha_1^r z_{t-1}^2 + \beta_1^r s_{t-1}^2, & k_{r-1} \le t < k_r \end{cases} \qquad \boldsymbol{w}_t \sim \begin{cases} N(0, \boldsymbol{W}_1), & k_0 \le t < k_1 \\ \vdots & \vdots \\ N(0, \boldsymbol{W}_r), & k_{r-1} \le t < k_r. \end{cases} \tag{7.6}$$

This gives us the following set of parameters $\{\alpha_0^1, \alpha_1^1, \beta_1^1, \boldsymbol{W}_1, ..., \alpha_0^r, \alpha_1^r, \beta_1^r, \boldsymbol{W}_r\}$.

Like before, an easy way to think of this model is to think of it as multiple separate pieces of data, $Y^1 = y_{1:k_1-1}$, $Y^2 = y_{k_1:k_2-1}$, ..., $Y^r = y_{k_{r-1}:T}$, glued together make up the original series y. And for each of these series we have a different set of parameters making up the model.

For clarity we give the full Kalman recursion equations. The **filtering** distribution is given by,

1. Assume that at time $t-1$ we have the distributions,

$$\boldsymbol{\theta}_{t-1}|y_{0:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$$

2. Then, the one-step-ahead forecast for the state vector is:

$$\boldsymbol{\theta}_t | y_{0:t-1} \sim N(\boldsymbol{a}_t, \boldsymbol{R}_t) \tag{7.7}$$

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

$$\boldsymbol{W}_t = \begin{cases} \boldsymbol{W}_1, & k_0 \leq t < k_1 \\ \vdots & \vdots \\ \boldsymbol{W}_r, & k_{r-1} \leq t < k_r \end{cases}$$

3. The one-step-ahead forecast for the observation vector is:

$$y_t | y_{0:t-1} \sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t), \tag{7.8}$$

where $f_t = \boldsymbol{F}_t' \boldsymbol{a}_t$ and $Q_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + \boldsymbol{s}_t^2$

$$s_t^2 = \begin{cases} \alpha_0 + \alpha_1 \Big(y_{t-1}^2 - 2 y_{t-1} \boldsymbol{F}' \boldsymbol{m}_{t-1} + \boldsymbol{F}' (\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1} \boldsymbol{m}_{t-1}') \boldsymbol{F} \Big) + \beta_1 s_{t-1}^2, & \leq t < k_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots & \vdots \\ \alpha_0^r + \alpha_1^r \Big(y_{t-1}^2 - 2 y_{t-1} \boldsymbol{F}' \boldsymbol{m}_{t-1} + \boldsymbol{F}' (\boldsymbol{C}_{t-1} + \boldsymbol{m}_{t-1} \boldsymbol{m}_{t-1}') \boldsymbol{F} \Big) + \beta_1^r s_{t-1}^2, & k_{r-1} \leq t < k_r \end{cases} \tag{7.9}$$

4. The filtered distribution at time t is then given by:

$$\boldsymbol{\theta}_t | y_{0:t-1} \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t) \tag{7.10}$$

Where,

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{e}_t \\ \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{K}_t \boldsymbol{Q}_t \boldsymbol{K}_t' \\ \boldsymbol{K}_t &= \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{Q}_t^{-1} \\ e_t &= y_t - f_t \end{aligned} \tag{7.11}$$

7.4 Estimating the Time-Varying GARCH-SSM:

In Section 4.2 we laid out the posterior distribution of the GARCH-SSM model and how to estimate the parameters. Just like in Section 4.2, to draw samples of the state vectors we use FFBS. The difference now is that the Kalman filter is run with the time varying parameters given in Section 6.2. For the single regime case we had chosen the an Half-Cauchy priors for the GARCH parameters and for the state variance we had chosen an Inverse-Gamma prior if it was univariate or an Inverse-Wishart prior if it was multivariate. Following this we choose mutually independent priors for the multi-regime GARCH-SSM. In particular we have that,

$$\begin{aligned} \pi_0(\alpha_0^1, \alpha_1^1, \beta_1^1, \boldsymbol{W}_1, ..., \alpha_0^r, \alpha_1^r, \beta_1^r, \boldsymbol{W}_r) = \pi_0(\alpha_0^1)\pi_0(\alpha_1^1)\pi_0(\beta_1^1)\pi_0(\boldsymbol{W}_1) \times ... \\ \times \pi_0(\alpha_0^r)\pi_0(\alpha_1^r)\pi_0(\beta_1^r)\pi_0(\boldsymbol{W}_r) \end{aligned}$$

Now we can describe the posterior distribution for the model as,

$$\begin{aligned} p(\boldsymbol{\theta}_{0:n}, \alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{W}_{1:r}) &\propto \pi_0(\alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{W}_{1:r}) \prod_{i=1}^{n} \pi(y_i | \alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_{1:r}, \boldsymbol{\theta}_{i-1}) \\ &\propto \prod_{i=1}^{k_1-1} \pi(y_i | \alpha_0^1, \alpha_1^1, \beta_1^1, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_1, \boldsymbol{\theta}_{i-1}) \times \\ &\quad \prod_{i=k_1}^{k_2-1} \pi(y_i | \alpha_0^2, \alpha_1^2, \beta_1^2, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_2, \boldsymbol{\theta}_{i-1}) \times ... \times \\ &\quad \prod_{i=k_{r-1}}^{T} \pi(y_i | \alpha_0^r, \alpha_1^r, \beta_1^r, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_r, \boldsymbol{\theta}_{i-1}) \times \\ &\quad \pi_0(\alpha_0^1)\pi_0(\alpha_1^1)\pi_0(\beta_1^1)\pi_0(\boldsymbol{W}_1) \times ... \times \pi_0(\alpha_0^r)\pi_0(\alpha_1^r)\pi_0(\beta_1^r)\pi_0(\boldsymbol{W}_r). \end{aligned} \tag{7.12}$$

Here $\alpha_0^{1:r} = \{\alpha_0^1, ..., \alpha_0^r\}$ and similarly for $\alpha_1^{1:r}$ and $\beta_1^{1:r}$.

Just like in the non-GARCH case, we see how the posterior distribution is factorizes based on the different regimes. This leads us to have the following conditional posterior distributions for the individual regime parameters,

$$\pi_n(\alpha_0^j, \alpha_1^j, \beta_1^j | \alpha_0^{(-j)}, \alpha_1^{(-j)}, \beta_1^{(-j)}, \boldsymbol{W}_{1:r}, \boldsymbol{\theta}, Y_{1:n}) \propto \pi_0(\alpha_0^j)\pi_0(\alpha_1^j)\pi_0(\beta_1^j) \prod_{i=k_{j-1}}^{k_j-1} \pi(y_i | \alpha_0^j, \alpha_1^j, \beta_1^j, \boldsymbol{\theta}_i)$$

$$\pi_n(\boldsymbol{W}_j | \alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{W}_{(-j)}, \boldsymbol{\theta}, Y_{1:n}) \propto \pi_0(\boldsymbol{W}_j) \prod_{i=k_{j-1}}^{k_j-1} \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_j, \boldsymbol{\theta}_{i-1}), \tag{7.13}$$

where $\alpha_0^{(-j)} = \{\alpha_0^1, ..., \alpha_0^{j-1}, \alpha_0^{j+1}, ..., \alpha_0^r\}$ and similarly for $\alpha_1^{(-j)}$ and $\beta_1^{(-j)}$.

Thus we see from above that the conditional distributions breaks down quite easily for each of the regimes. Since these conditional distributions are not tractable standard distribution with respect to the parameters $\alpha_0^j, \alpha_1^j, \beta_1^j$, we use the Metropolis-Hastings algorithm to sample from the posterior. For the $\boldsymbol{W}_j$ on the other hand an Inverse-Wishart prior is a conjugate prior.

7.5 Unknown Change Point:

So far in this section we have assumed that the change points $k_1, ..., k_{r-1}$ are known. However, in reality the position of the change points may be unknown. Further the number of change points present may also be unknown. We defer the discussion of the latter to section 6.5. For know let us assume that the number of change points is known but their positions unknown.

Like before, let $k_1, ..., k_{r-1}$ be the change points, whose values are unknown. We would like to place a prior distribution on the change points. Following Green (1995), we may choose the

prior on $k_1, ..., k_{r-1}$ to be the even-numbered order statistics from $2r-1$ points drawn from the discrete uniform distribution on $[0, T]$. Meaning, if $j_{(1)}, j_{(2)}, ..., j_{(2r-1)}$ are the order statistics obtained from the uniform distribution on $[0, T]$ then $k_1 = j_{(2)}, k_2 = j_{(4)}, ..., k_{r-1} = j_{(2r-2)}$. This formulation of taking the even ordered statistics gives a nice formulation for the joint distribution of $k_1, ..., k_{r-1}$.

$$\begin{aligned}
\pi_0(k_1, ..., k_{r-1}) &= \pi_0(j_2, j_4, ..., j_{2(r-1)}) \\
&= (2r-1)! \frac{(j_2)(j_4 - j_2)...(j_{2r-2} - j_{2(r-2)})}{T^{r-1}} \frac{(T - j_{2r-2})}{T} \times \left(\frac{1}{T}\right)^{r-1} \mathbb{I}_{t_{\min} \geq 0} \mathbb{I}_{t_{\max} \leq L} \\
&= (2r-1)! \frac{(k_1)(k_2 - k_1)...(k_{r-1} - k_{r-2})}{T^{r-1}} \frac{(T - k_r)}{T} \times \left(\frac{1}{T}\right)^{r-1} \mathbb{I}_{s_{(0)} \geq 0} \mathbb{I}_{s_{\max} \leq L} \\
&= (2r-1)! \frac{(k_1)(k_2 - k_1)...(k_{r-1} - k_{r-2})(T - k_{r-1})}{T^{2r-1}} \\
&= \frac{(2r-1)!}{T^{(2r-1)}} \prod_{i=1}^{r} (k_i - k_{i-1}).
\end{aligned} \tag{7.14}$$

Here we take $k_0 = 0$ and $k_r = T$. This prior has been adapted from Green (1995), which looked at an example of Bayesian change point analysis.

The resulting posterior distribution is very similar to what we have already seen so far.

$$
\begin{aligned}
\pi_n(\boldsymbol{\theta}_{0:n}, \alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{W}_{1:r}, k_{1:r-1} | y_{1:n}) \propto \\
\pi_0(\alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{W}_{1:r}, k_{1:r-1}) \prod_{i=1}^{T} \pi(y_i | \alpha_0^{1:r}, \alpha_1^{1:r}, \beta_1^{1:r}, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_{1:r}, \boldsymbol{\theta}_{i-1}) \\
\propto \prod_{i=1}^{k_1-1} \pi(y_i | \alpha_0^1, \alpha_1^1, \beta_1^1, k_1, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_1, \boldsymbol{\theta}_{i-1}, k_1) \times \\
\prod_{i=k_1}^{k_2-1} \pi(y_i | \alpha_0^2, \alpha_1^2, \beta_1^2, k_2, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_2, \boldsymbol{\theta}_{i-1}, k_2) \times ... \times \\
\prod_{i=k_{r-1}}^{T} \pi(y_i | \alpha_0^r, \alpha_1^r, \beta_1^r, k_r, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | \boldsymbol{W}_r, \boldsymbol{\theta}_{i-1}, k_r) \times \\
\pi_0(\alpha_0^1) \pi_0(\alpha_1^1) \pi_0(\beta_1^1) \pi_0(\boldsymbol{W}_1) \times ... \times \pi_0(\alpha_0^r) \pi_0(\alpha_1^r) \pi_0(\beta_1^r) \pi_0(\boldsymbol{W}_r) \times \\
\frac{(2r-1)!}{T^{(2r-1)}} \prod_{i=1}^{r} (k_i - k_{i-1}).
\end{aligned}
\tag{7.15}
$$

Like before the conditional distribution for the GARCH parameters are factorize quite easily based on the different regimes. To sample from the posterior distribution we follow our usual steps of running the FFBS to get the state vectors and then conditional on the state vectors we can use the Metropolis-Hastings algorithm to draw posterior samples of the other parameters.

Simulation Study:

We simulate a GARCH-SSM with two regimes. We have 2000 observations. The change point occurs at time $t = 1000$. The parameters of the model are,

$$
\begin{aligned}
\alpha_0^1 = 1, \alpha_1^1 = 0.3, \beta_1^1 = 0.1, W^1 = 0.04 \\
\alpha_0^2 = 2, \alpha_1^1 = 0.1, \beta_1^1 = 0.5, W^1 = 0.01
\end{aligned}
\tag{7.16}
$$

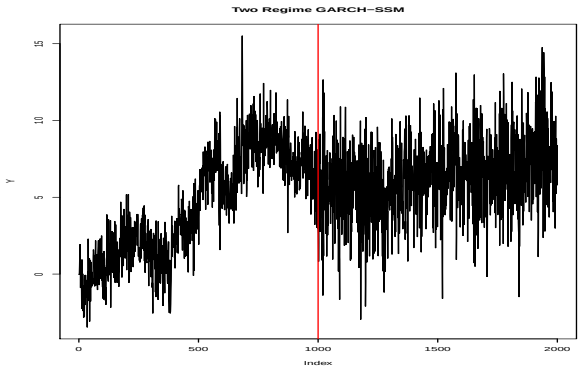

Figure 7–1: A two-regime GARCH SSM. The red line in the middle shows where the change point occurs.

Our goal here is to estimate the parameters operating under the two regimes and also estimate the change point. The estimated parameters are given in the table below. We can see that the parameters are pretty close to the true values. In terms of the estimate for the change point, we are estimating that the change point occurred at time $t = 995.63$ which is very close to the true value of the change point ($t = 1000$).

Regime 1	α_0^1	α_1^1	β_0^1	W^1
	0.9194	0.2528	0.2467	0.0280
	0.0487	0.0029	0.0199	0.00004
Regime 2	α_0^2	α_1^2	β_0^2	W^2
	1.9748	0.2399	0.4317	0.0091
	0.3495	0.0027	0.0140	0.00001
	Change Point		WAIC	
	995.63		-4102.823	

Table 7–1: Posterior median of the GARCH parameters based on 5000 draws from the posterior distribution.

We fitted two more models to this data. The first assumed that there was just one regime, i.e. no change point. The second assumed that there was three regimes operating. We would want to be able to reject both these models using WAIC. For the single regime model the WAIC was -4219.746. And for the three regime model the WAIC was -4087.151. The three-regime model has the highest WAIC. Ideally we would like the two-regime model to have the highest WAIC. In this case it is interesting to see the traceplots for the estimate of the change points in the two-regime and three-regime model. These are given in Figure 7-2 and Figure 7-3 respectively.

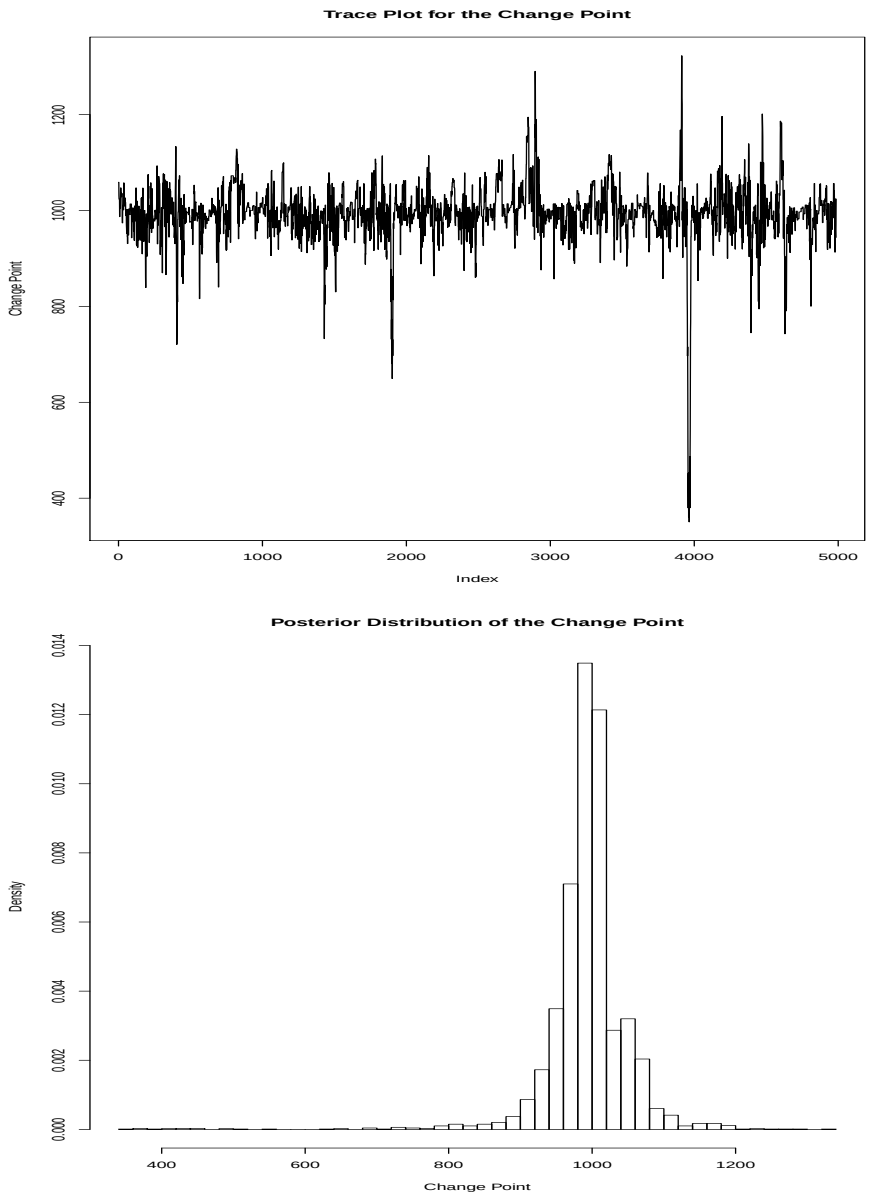

Figure 7–2: The traceplot (left) and the histogram (right) of the posterior samples of the change point. We can see that the chain has converged quite well. The histogram shows us that the posterior value for the change point is meaned around 1000.

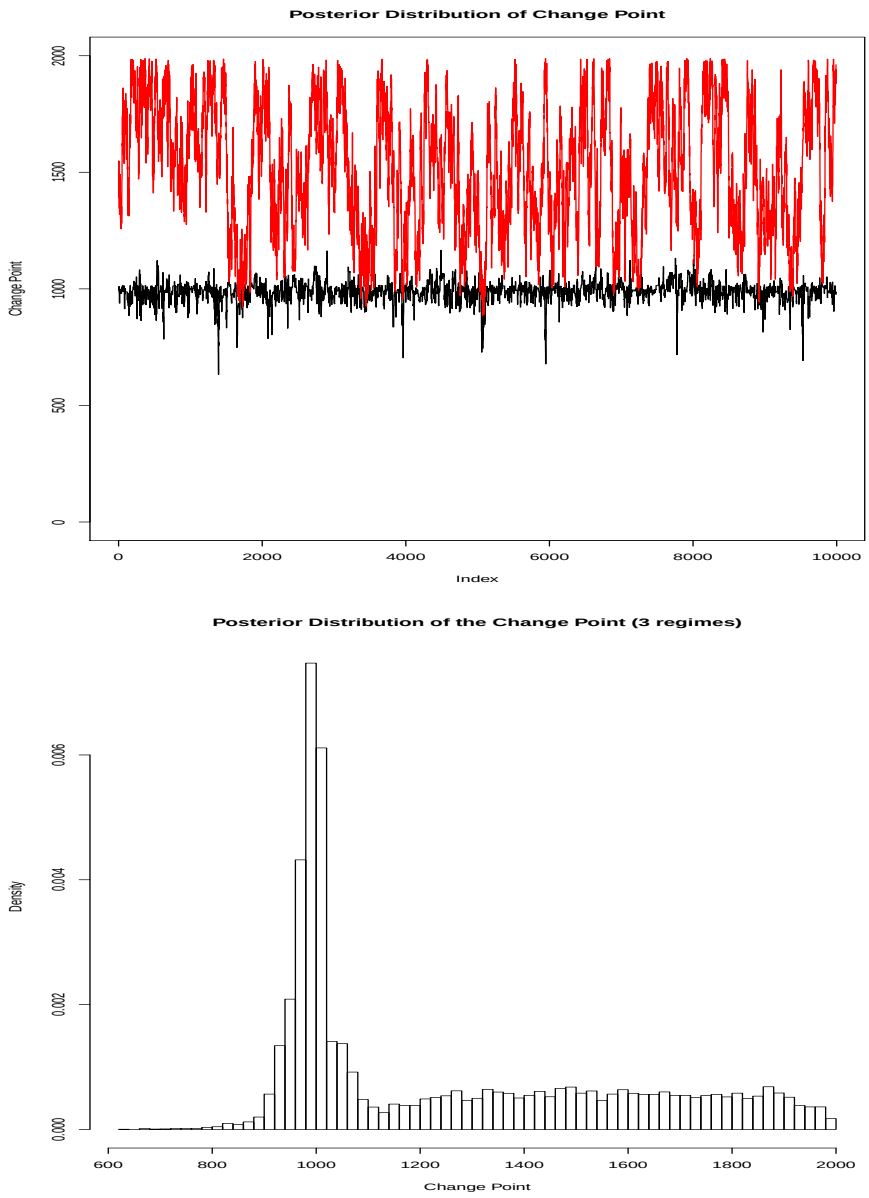

Figure 7–3: The traceplot (above) and the histogram (below) of the posterior samples of the change point. In black is the traceplot of the first change point and red is the traceplot of the second change point. We can see that the chain has converged quite well. The histogram shows us that the posterior value for the change point is meaned around 1000.

For the two-regime model we see that the traceplot is fluctuating very evenly around the value of 1000. We know that this is the true change point and is the only change point and

hence the two-regime model is correctly estimating the true value of the change point. For the three regime model, the first change point is being estimated around the point 1000. If we look at the black traceplot in Figure 7-3, which is the traceplot of the first change point we see that it is fluctuating within a narrow band around the point 1000, our true change point value. Where as, the red line, which, is the traceplot of the second change point, is uniformly fluctuating between the points 1000 to 2000. This means that the second change point is similarly high posterior probability along the interval 1000 to 2000.

A look at the estimates of the three-regime model shows us that the estimates of the parameters for the second and third regime in the two-regime model are extremely similar. Which again is no surprise, since in truth the model after time $t = 1000$ has follows the same regime. The first regime in the three-regime model is also being correctly estimated, with the parameter estimates close to the true parameters of the first regime given in (7.16).

Regime 1	α_0^1	α_1^1	β_0^1	W^1
	0.9363	0.2471	0.2212	0.0281
	0.04196	0.00259	0.01750	0.00004
Regime 2	α_0^2	α_1^2	β_0^2	W^2
	2.3558	0.2569	0.3362	0.0096
	0.89348	0.00886	0.02670	0.00001
Regime 3	α_0^3	α_1^3	β_0^3	W^3
	2.3318	0.2659	0.3718	0.0099
	0.91242	0.00872	0.01974	0.00001
Change Point 1		Change Point 2		WAIC
990.88		1518.405		-4087.151

Table 7–2: Posterior median of the GARCH parameters based on 5000 draws from the posterior distribution.

7.6 Unknown Number of Change Points:

So far we have assumed the number of change points to be known. However, it may be of interest to estimate the number of change points supported by the posterior distribution. Green (1995) uses the technology of reversible jump MCMC (RJMCMC) to tackle the problem of finding the number of change points in a particular data set. In this section we discuss how we may also use this technology in the context of estimating the number of regimes for our SSM. In one sense RJMCMC can be used as a method of assessing Bayesian model selection by helping us choose the number of regimes and location of the change points most supported by the posterior distribution. To keep things simple we assume a standard Gaussian first order SSM's (random walk plus noise SSM) with multiple regimes. We show later how this method can later be extended to GARCH-SSMs with multiple change points.

For clarity we specify the model here again.

$$\begin{aligned} y_t &= \theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t, \end{aligned} \tag{7.17}$$

where the observation and state errors are distributed in the following manner,

$$v_t \sim \begin{cases} N(0, V_1), & k_0 \leq t < k_1 \\ \vdots & \vdots \\ N(0, V_r), & k_{r-1} \leq t < k_r \end{cases} \qquad w_t \sim \begin{cases} N(0, W_1), & k_0 \leq t < k_1 \\ \vdots & \vdots \\ N(0, W_r), & k_{r-1} \leq t < k_r \end{cases}, \tag{7.18}$$

where r is now a parameter that must be estimated.

7.6.1 Prior Specification:

First we specify the prior for the number of regimes (i.e. nuber of change points in the model), r. The prior for the regimes is set to be a Truncated Poisson distribution. In particular we have that,

$$\pi_0(r) = \frac{\exp(-\lambda)\lambda^{r-1}}{(r-1)!}\mathbb{I}_{r\leq r_{\max}}\mathbb{I}_{r\geq 1} \quad r = 1, 2, ..., r_{\max}$$

where we set $r_{\max}$ is set deterministically.

Finally, we specify the priors of the variances conditional on the number of regimes. Like before we take independent Inverse-Gamma priors for the variances. So given that there are r regimes in the model we have that the prior variances are given by,

$$\pi_0(V_1, W_1, ..., V_r, W_r|r) = \pi_0(V_1)\pi_0(W_1) \times ... \times \pi_0(V_r)\pi_0(W_r)$$

where, $V_i \sim IG(\alpha_1, \beta_1)$ and $W_i \sim IG(\alpha_2, \beta_2)$, for all $i = 1, ..., r$.

The posterior distribution is thus given by,

$$\begin{aligned}
\pi_n(V_{1:r}W_{1:r}, r, k_{1:r-1}, \boldsymbol{\theta}|Y_{1:n}) &\propto \prod_{i=1}^{n} \pi(Y_i|V_{1:r}, \theta_i, k_{1:r-1}) \prod_{i=1}^{n} \pi(\theta_i|W_{1:r}, \theta_{i-1}, k_{1:r-1})\pi_0(V_{1:r}, W_{1:r}, k_{1:r-1}, r) \\
&\propto \mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{1:r-1})\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1})\pi_0(V_{1:r}, W_{1:r}|r)\pi_0(k_{1:r-1}|r)\pi_0(r) \\
&\propto \mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{1:r-1})\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1}) \prod_{j=1}^{r} \pi_0(V_j|r)\pi_0(W_j|r)\pi_0(k_{1:r-1}|r)\pi_0(r) \\
&\propto \mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{1:r-1})\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1}) \prod_{j=1}^{r} \pi_0(V_j|r)\pi_0(W_j|r)\pi_0(k_{1:r-1}|r)\pi_0(r)
\end{aligned}$$

7.6.2 Jumps:

Now that we have the priors for the model we describe the "moves" that our model can make. Our objective here is two folds. First we want to obtain information regarding the change points as supported by the data. By this we mean that we want to know the optimal number of change points suggested by the data and estimate their positions. Secondly given these change points change points we want to estimate the variances in each regime.

We define here the possible jumps that our model can make. Suppose that we have univariate DLM. Now suppose that in the current state we have k regimes. There are 4 possible moves that we can make. The first move is a **position** move, the second move is a **birth** move, the third move is a **death** move and finally the fourth move is call a **local** move.

The **position** move, keeps us in the same number of regimes but proposes to change the length one of the regimes. For this we randomly choose one of the change points $k_1, ..., k_{r-1}$ with uniform probability, say we choose k_j. We propose to then move is uniformly between $[k_{j-1}, k_{j+1}]$. The second possible move, the **birth** move, proposes to increase the number of regimes, from r to $r+1$. And the third move, the **death** move, proposes to decrease the number of regimes from r to $r-1$. The fourth move, the **local** move, keeps the regimes and the change points the same and updates the state and observational variances for that regime, using Gibbs sampling.

We quickly note some exceptions. If currently we have only one regime, i.e. $r = 1$, then we cannot make a death move. Similarly if we have that $r = r_{\max}$ then we cannot make a birth move.

In the next part we discuss each of these moves and their proposal and acceptance probabilities in more details.

7.6.3 Acceptance Probabilities:

We discuss the acceptance probabilities of each move. In general the acceptance probability of a reversible jump step is given by,

$$\alpha(\mathcal{M}_k, \mathcal{M}_{k'}) = \min\left\{1, \frac{\pi_n(\boldsymbol{\theta}_{k'})q(k', k)\overleftarrow{r}_{k'}}{\pi_n(\boldsymbol{\theta}_k)q(k, k')\overrightarrow{r}_k}|\boldsymbol{J}|\right\}$$

Here $\pi_n()$ is the posterior probability, $q()$ is the proposal probability, $\overleftarrow{r}$ is the backward probability, $\overrightarrow{r}$ is the forward probability and $|\boldsymbol{J}|$ is the Jacobian.

We start with the **local** move. Suppose that in our current state we have r regimes. Our parameters are the $V_1, ..., V_r$ observation variances, $W_1, ..., W_r$ state variances and $k_1, ..., k_{r-1}$ change points. For this move we keep the number of regimes and the current change points fixed and update the variances using a Gibbs sampling. This would be done in the usual way, we run the FFBS using our time varying variances and obtain a sample of the state vectors. Then for each regime we have that,

$$\pi_n(V_i|\boldsymbol{\theta}, r, k_1, ..., k_{r-1}) \propto Inv - Gam\left(\frac{k_i - k_{i-1}}{2} + \alpha_1, \frac{\sum_{j=k_{i-1}}^{k_i}(Y_j - \theta_j)^2}{2} + \beta_1\right)$$
$$\pi_n(W_i|\boldsymbol{\theta}, r, k_1, ..., k_{r-1}) \propto Inv - Gam\left(\frac{k_i - k_{i-1}}{2} + \alpha_2, \frac{\sum_{j=k_{i-1}}^{k_i}(\theta_j - \theta_{j-1})^2}{2} + \beta_2\right)$$

It is easy to arrive at these above distributions. To illustrate this we write the full posterior distribution below. We run the FFBS and we get a sample of the state vectors. As shown previously the full posterior distribution is given by,

$$\pi_n(V_{1:r}W_{1:r}, r, k_{1:r-1}, \boldsymbol{\theta}|Y_{1:n}) \propto \mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{1:r-1})\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1}) \prod_{j=1}^{r} \pi_0(V_j|r)\pi_0(W_j|r)\pi_0(k_{1:r-1}|r)\pi_0(r)$$

We since V_i, W_j are mutually independent priors for all i, j we have that the priors factorize out completely conditional on the number of regimes, r. When finding the conditional distribution of V_i, we see that in the likelihood, the only Y's that depend on the V_i are terms that are in the regime i, i.e. the $Y_i's$ that are in the interval $(k_{i-1}, k_i]$. Similarly the only terms that depend on the parameter W_i are the θ's that are in the interval $(k_{i-1}, k_i]$. Thus we get the conditional distributions that we have above.

The last thing that we have to specify is the probability of choosing this move. For now let us just suppose that we choose to make a **local** move when we have r regimes with probability c_r.

The next move that we describe is the **position** move. Suppose that currently, we are at r regimes. We run the FFBS, using the current value of parameters, and obtain a sample of the state vectors. Next, for the **position** move we choose to alter one of the change points, $k_1, ..., k_{r-1}$, which we pick randomly. Suppose we choose k_j. We propose a new value of k'_j, picked uniformly over the interval $[k_{j-1}, k_{j+1}]$. Then using these state vectors we calculate the posterior under the proposed value and current value. We use a Metropolis-Hastings step to accept or reject the proposed value of k'_j, with the acceptance probability being given by,

$$\min\left\{1, \frac{\mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{-j}, k'_j)\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1})(k'_j - k_{j-1})(k_{j+1} - k'_j)}{\mathcal{L}(\boldsymbol{y}|V_{1:r}, \boldsymbol{\theta}, k_{1:r-1})\mathcal{L}(\boldsymbol{\theta}|W_{1:r}, k_{1:r-1})(k_j - k_{j-1})(k_{j+1} - k_j)}\right\}$$

The probability of choosing to make a **position** move is given by p_r.

We now describe the **birth** move. As stated before, this type of move proposes to increase the number of regimes from the current value r to $r + 1$. First, we run the FFBS, using the current value of parameters (from the r regime model) and obtain a sample of the state vectors. Next, we uniformly choose a point in the interval $[1, T]$, call it k^*. With probability 1 we have that this point lies inside one of the r regimes, given by the intervals $[k_0, k_1], ..., [k_{r-1}, T]$. Suppose that it lies in the j^{th} regime, i.e., $k^* \in [k_{j-1}, k_j]$. Our new proposed change points are then given by, $k_1, ..., k_{j-1}, k^*, k_j, ..., k_{r-1}$. Now given these new intervals, $[k_{j-1}, k^*]$ and $[k^*, k_j]$ we need to also propose a new observation and state variances operating in these new proposed intervals. We choose to make a deterministic proposal for these new variances, however, one can also make a stochastic proposal.

For the interval $[k_{j-1}, k^*]$, we propose the variances to be based on the sample variances of the observation and the states operating in these intervals. This would thus be given by,

$$\begin{aligned} V_j^* &= \frac{1}{\#\{Y \in [k_{j-1}, k^*]\} - 1} \sum_{Y_i \in [k_{j-1}, k^*]} (u_i - \bar{u})^2 \\ W_j^* &= \frac{1}{\#\{\theta \in [k_{j-1}, k^*]\} - 1} \sum_{\theta_i \in [k_{j-1}, k^*]} (z_i - \bar{z})^2, \end{aligned} \tag{7.19}$$

where,

$$u_i = Y_i - \theta_i, \qquad z_i = \theta_i - \theta_{i-1}.$$

One recognizes that 7.19 is exactly the sample variance formula. A look at the model description given in (7.17) gives us the motivation of our proposal. Similarly, for the interval $[k^*, k_j]$, we have the new variances proposed using the sample variances within these intervals, given by,

$$\begin{aligned} V_{j+1}^* &= \frac{1}{\#\{Y \in [k^*, k_j]\} - 1} \sum_{Y_i \in [k^*, k_j]} (u_i - \bar{u})^2 \\ W_{j+1}^* &= \frac{1}{\#\{\theta \in [k^*, k_j]\} - 1} \sum_{\theta_i \in [k^*, k_j]} (z_i - \bar{z})^2, \end{aligned}$$

where,

$$u_i = Y_i - \theta_i, \qquad z_i = \theta_i - \theta_{i-1}.$$

This describes the full set of parameter augmentation to go from r regimes to $r+1$ regimes. We use a Metropolis Hastings step to accept or reject our proposed parameters. Before we go into the details about acceptance probabilities, we expose the process of a death step, since we will need this to describe the reverse move used to calculate the acceptance probability.

The **death** move allows us to move from r regimes to $r-1$ regimes. As before, first, we run the FFBS, using the current value of parameters (from the r regime model) and obtain a sample of the state vectors. Next, we randomly choose one of the r regimes, suppose we

choose regime j, which operates over the interval $[k_{j-1}, k_j]$ and then we collapse it. Which means that we remove the point k_j. This then gives us the single interval $[k_{j-1}, k_{j+1}]$. Currently we have the variances V_j, W_j and V_{j+1}, W_{j+1} operating on the intervals $[k_{j-1}, k_j]$ and $[k_j, k_{j+1}]$ respectively. Since we have merged these two intervals into a single interval we need to somehow propose a single variance operating in this new regime. We propose the new variance deterministically, by choosing the sample variance operating in the newly collapsed regime. This then gives us the variances as,

$$\begin{aligned} V_j^* &= \frac{1}{\#\{Y \in [k_{j-1}, k_{j+1}]\} - 1} \sum_{Y_i \in [k_{j-1}, k_{j+1}]} (u_i - \bar{u})^2 \\ W_j^* &= \frac{1}{\#\{\theta \in [k_{j-1}, k_{j+1}]\} - 1} \sum_{\theta_i \in [k_{j-1}, k_{j+1}]} (z_i - \bar{z})^2, \end{aligned} \tag{7.20}$$

where,

$$u_i = Y_i - \theta_i, \qquad z_i = \theta_i - \theta_{i-1}.$$

We choose the birth move with probability b_r and the death move with probability d_r.

The last thing we expose before discussing the acceptance probability at length is the calculation of the Jacobian, $\boldsymbol{J}$. Due to our choice of proposals in the dimension matching it is easy to see that the Jacobian, $|\boldsymbol{J}|$ is unity. In fact, this is not a coincidence but rather premeditated. However, depending on the choice of the proposals used, the Jacobian would change.

Now we have everything that we need to find the acceptance probability. Suppose we choose to make a **birth** move. We use FFBS to obtain a sample of the states. Then, given our current parameter values we propose the augmented parameters and accept these parameter

using with probability,

$$\min\left\{1, \frac{\pi_n(\boldsymbol{V}', \boldsymbol{W}', \boldsymbol{k}', \boldsymbol{\theta}) 1/(r+1) d_{r+1}}{\pi_n(\boldsymbol{V}, \boldsymbol{W}, \boldsymbol{k}, \boldsymbol{\theta})(1/T) b_r} |\boldsymbol{J}|\right\}$$

Here $\boldsymbol{V} = (V_1, ..., V_r)$, $\boldsymbol{W} = (W_1, ..., W_r)$, $\boldsymbol{k} = (k_1, ..., k_{r-1})$, $\boldsymbol{V'} = (V'_1, ..., V'_{r+1})$, $\boldsymbol{W'} = (W'_1, ..., W'_{r+1})$ and $\boldsymbol{k'} = (k'_1, ..., k'_r)$.

Finally the acceptance probability of a **death** move is given by,

$$\min\left\{1, \frac{\pi_n(\boldsymbol{V}', \boldsymbol{W}', \boldsymbol{k}', \boldsymbol{\theta})(1/T) b_r}{\pi_n(\boldsymbol{V}, \boldsymbol{W}, \boldsymbol{k}, \boldsymbol{\theta})(1/r) d_{r-1}} |\boldsymbol{J}|^{-1}\right\}$$

Here $\boldsymbol{V} = (V_1, ..., V_r)$, $\boldsymbol{W} = (W_1, ..., W_r)$, $\boldsymbol{k} = (k_1, ..., k_{r-1})$, $\boldsymbol{V'} = (V'_1, ..., V'_{r-1})$, $\boldsymbol{W'} = (W'_1, ..., W'_{r-1})$ and $\boldsymbol{k'} = (k'_1, ..., k'_{r-2})$.

Simulation:

We simulate a Gaussian SSM with four regimes. In this model $T = 2000$ and the change points are place at $t = 500, 1000, 1500$. For the first and the third regime we have that the observation variance is 100 and the state variance is 0.04. For the second and the third regime we have that the observation variance is 10 and the state variance is 0.01. We want to use the RJMCMC technology to estimate the value of the parameters and the number of change points.

We use the priors given previously. The priors for the variances are re-stated for convenience. $V_i \sim IG(\alpha_1, \beta_1)$ and $W_i \sim IG(\alpha_2, \beta_2)$, for all $i = 1, ..., r$. Where $\alpha_1 = 2.5$, $\beta_1 = 0.01$ and $\alpha_1 = 2.5$, $\beta_1 = 1$. For this model we set $r_{\max} = 10$, i.e. there can be a max 10 regimes (9 change points).

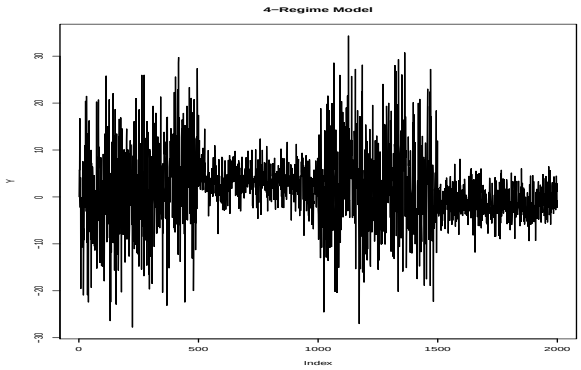

Figure 7–4: 4-regime model (3 change points)

The first plot we show is the histogram of the number of change points estimated by the model. We see from the histogram below that 3 change point model has the highest posterior probability. The 4 change point model also has a high posterior probability. Having any other number of change points has a very low posterior probability. Looking at the trace plots of estimates of the change points of these models paints a much clearer picture on which model is the clear winner.

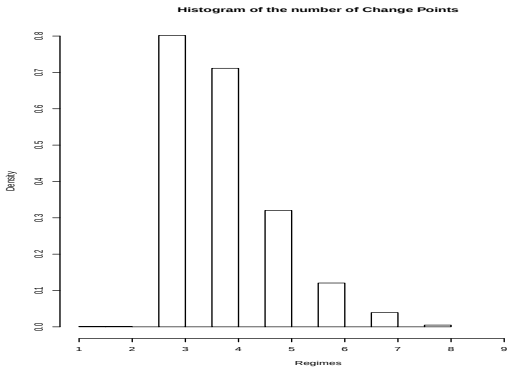

Figure 7–5: Histogram of the number of change points

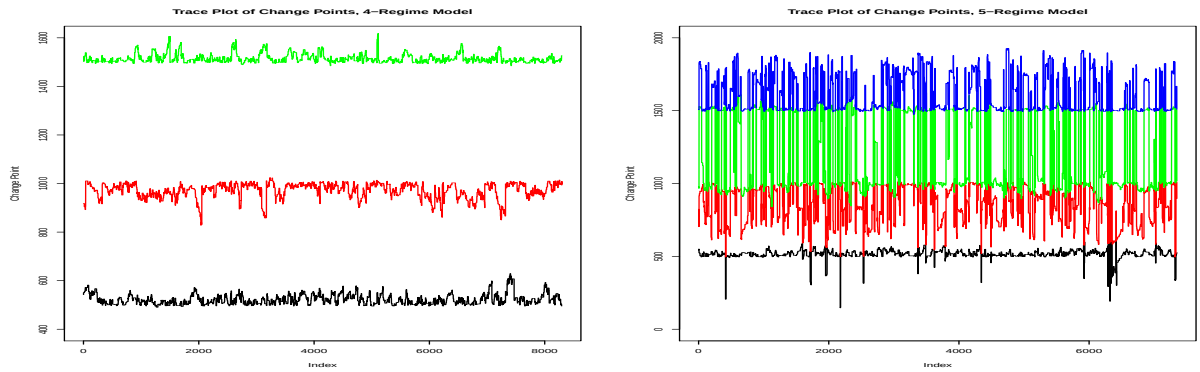

Figure 7–6: Left: Trace Plots of the estimated change points from the 3 change points model; Right: Trace Plots of the estimated change points from the 4 change points.

The plot on the left shows the trace plots of the change points from the three change point model. We see that each of the trace plots are quite stationary and fluctuating evenly around the true change point values of $500, 1000, 1500$. Clearly the change points are also quite well identified.

For the plot on the right, the 4 change point model, the first change point at $t = 500$ is well identified and estimated. However, if we look ad the next three change points they seem be a bit all over the place. Recall that in Figure 4, where the true model had only one change point, the second change point was uniformly fluctuating about. In this 4 change point model, we see a similar pattern, with 3 of the change points uniformly fluctuating between the intervals $(500, 1000)$, $(1000, 1500)$ and $(1500, 2000)$, unlike the 3 change point model where we see that the estimates have 'settled' in quite well. Thus we would be correct in choosing the three change point model as the best model in for this data, which is good since we know that this is the true model specification.

The estimates of the models are given in the table below.

Sample size	V^1	V^2	V^3	V^4
$T = 2000$	95.012	9.972	91.103	9.529
	(46.884)	(0.717)	(56.521)	(0.406)
	W^1	W^2	W^3	W^4
	0.188	0.157	0.317	0.114
	(0.010)	(0.003)	(0.029)	(0.002)
	Change Point 1	Change Point 2	Change Point 3	
	516.549	978.980	1510.193	

Table 7–3: Posterior median estimates from the 3-change point model.

The estimates of the observation variance and the change points are very close to the true values. However, the estimates of the state variances are biased. The estimates of the state variances can be improved by using a larger sample size and/or different priors.

One note of caution to mention is that the ability to properly estimate the number of regimes in the model and where the change points lie is dependent on the length of each regime. In our case here each regime had a length of 500 data points, which seems to be good enough to help us in correctly estimating the model. The model is also quite sensitive to the choice of prior as well. Below is the histogram of the posterior estimates of the number of change points when changing the rate parameter for the state prior from $\beta_2 = 1$ to $\beta_2 = 0.1$. We see that the 5 change point model is now the posterior mode.

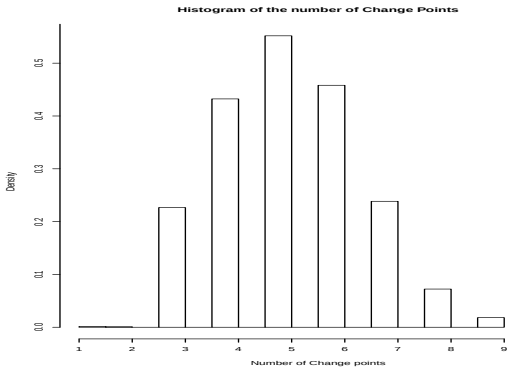

Figure 7–7: Histogram of the number of change points

CHAPTER 8
Analyzing the ICU Data:

In this section we will apply our GARCH-SSM to the heart rate (HRT) and blood pressure (BP) series. As mentioned earlier both these series show non-stationarity properties in mean and variance. We thus analyze the data using the methods developed in the previous sections. In terms of the overall structure of the model, for both the HRT and BP series we use the univariate first order polynomial SSM (random walk plus drift) with GARCH errors. Thus the model specification is,

$$\begin{aligned} y_t &= \theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t, \quad w_t \sim N(0, W) \\ v_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

We used the following priors for the model,

$$\begin{aligned} \pi_0(\alpha_0) &\sim Cauchy(0,1)\mathcal{I}_{\alpha_0 > 0} \\ \pi_0(\alpha_1) &\sim Cauchy(0,1)\mathcal{I}_{\alpha_1 0} \\ \pi_0(\beta_1) &\sim Cauchy(0,1)\mathcal{I}_{\beta_1 > 0} \\ \pi_0(W) &\sim IG(10,10) \end{aligned} \tag{8.1}$$

We will also want to analyze the data for evidence of any structural breaks. For the multi structure model we used the prior described in (37) for the unknown change points. And for the GARCH parameters for each of the regimes we used the prior given in (45).

A quick look at the data would suggest that there is at least one structural break in both series, which, have been re-plotted below for convenience. Further, the large spike in the data that are present throughout both series are characteristics of having volatility clusters.

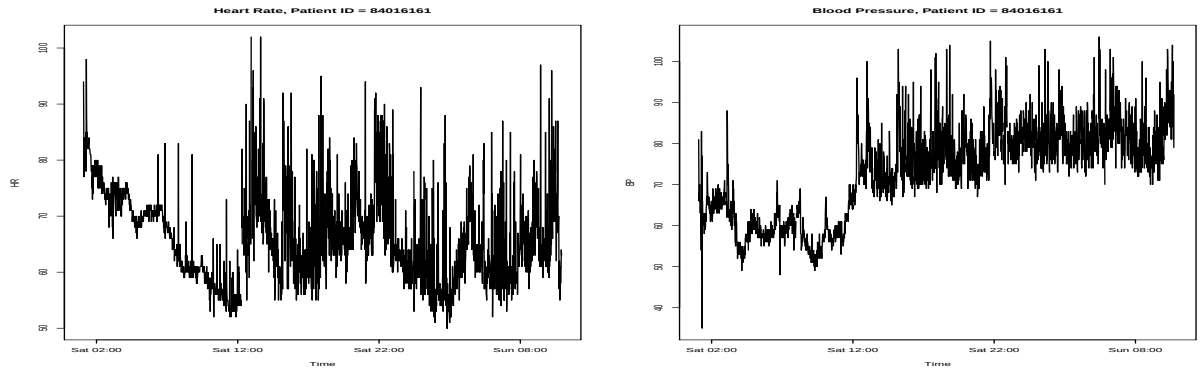

Figure 8–1: Left: Heart Rate; Right: Blood Pressure.

Slightly after Saturday 1200 hr there seems to be a structural break in both series.

Given these observations, we believe that a SSM with GARCH$(1,1)$ errors would be an appropriate model. The SSM can help us estimate the latent process driving the level of the model, while the GARCH$(1,1)$ errors would help us to model the heteroskedasticity present in the data. We compare this GARCH-SSM to the standard Gaussian SSM, which we can use as a benchmark. We also consider fitting a model with two regimes to account for the structural break seen in the data and compare the single regime model to the model with a structural break. Finally we fit a bivariate model.

8.1 Estimating the Heart Rate Series:

We start with the HRT series. We use the Metropolis-Hastings algorithm to sample from the posterior distribution to estimate the single regime GARCH-SSM. We simulated 50,000 samples from the posterior distribution and the estimates below are the posterior-median

estimates of the parameters. For the Gaussian-SSM we used Gibbs sampling to simulate from the posterior distribution since a conjugate analysis is available. This is computationally much faster than the Metropolis Hastings algorithm. Below are the estimates of the single regime HRT series.

GARCH-SSM					
Sample size	α_0	α_1	β_1	W	WAIC
$T = 2029$	1.235	0.305	0.689	0.653	-6220.22
	(0.0856)	(0.0017)	(0.0017)	(0.0080)	
Gaussian-SSM					
	V	W			WAIC
	26.395	0.7274			-6450.00
	(0.860)	(0.0168)			

Table 8–1: Posterior estimates of the parameters for the HRT series when fitting a GARCH-SSM and when fitting a Gaussian-SSM based on 5000 draws from the posterior distribution. In parenthesis are the standard errors of the estimates.

Comparing the WAIC of the GARCH model with that of the Gaussian model, we see that for the heart rate series, the GARCH model has a higher WAIC (-6220.22) than the standard Gaussian model (-6450.00). Recall that WAIC is a measure for predictive accuracy, thus in terms of prediction accuracy the GARCH SSM is more appropriate than the standard Gaussian SSM. From Table-9 above we see that for the heart rate series, the GARCH parameters are very close to the non-stationary region. The estimate for the unconditional variance for the heart rate at the observational level is,

$$Var(Y) = \frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1 - \hat{\beta}_1} = \frac{1.235}{1 - 0.305 - 0.689} = 205.83$$

In terms of interpretation for the parameter estimates, the values of $\hat{\alpha}_1$ and $\hat{\beta}_1$ prescribe how the variance/volatility will change over time. α_0 is the baseline variance. A high value for α_1 implies that large deviations will have a big impact on the variance, while a large value of β_1 allows for change in variance to persist over time. If α_1 was very small, then large deviations would not have much effect on the variance. Similarly if β_1 was very small, then we have that the variance would very quickly revert back to its mean. In our case we have that the estimate for $\alpha_1 = 0.305$ and the estimate for $\beta_1 = 0.689$. The high value of $\hat{\beta}_1$ means that variances will revert to its mean much slowly, i.e. high persistence in the variance.

We now fit a GARCH-SSM with two regimes to the HRT series.

Regime 1	α_0^1	α_1^1	β_0^1	W^1
	1.2959	0.5989	0.3054	0.1818
	(0.22893)	(0.01229)	(0.01038)	(0.00270)
Regime 2	α_0^2	α_1^2	β_0^2	W^2
	4.9377	0.2389	0.6680	0.5248
	(2.56028)	(0.00354)	(0.00576)	(0.01201)
Change Point 1		WAIC		
600.905		-6166.319		

Table 8–2: Posterior median of the GARCH parameters and the estimate of the change point based on 5000 draws from the posterior distribution. In parenthesis are the standard errors of the estimates.

The first thing we notice is that the WAIC for the two-regime model is -6166.319, which is higher than the WAIC for the other two models. Secondly, unlike the single regime estimates the GARCH parameter estimates for both regimes are not close to the non-stationary region. And we see that for each regime, the estimates are considerably different. The unconditional variance for each regime is,

$$Var(Y^1) = \frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1 - \hat{\beta}_1} = \frac{1.2959}{1 - 0.5989 - 0.3054} = 13.541$$
$$Var(Y^2) = \frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1 - \hat{\beta}_1} = \frac{4.9377}{1 - 0.2389 - 0.6680} = 53.037$$

For both regimes, we see that the unconditional variance is considerably less than the that of the single regime GARCH-SSM (which was 205.83). We also have that the second regime is considerably longer than the first regime, since the change point is estimate to be around time $t \approx 600$. Recall from our simulation study, how the sample size effects the parameter estimates. Thus if the estimated length of a regime is long then its parameter estimates will be better.

The figure below shows us the estimated values of the states and the residuals from fitted the GARCH model. The line in red in the plot below is the posterior mean of the state vectors over time. Both the single regime and two-regime model seem to have captured the mean level of the model well.

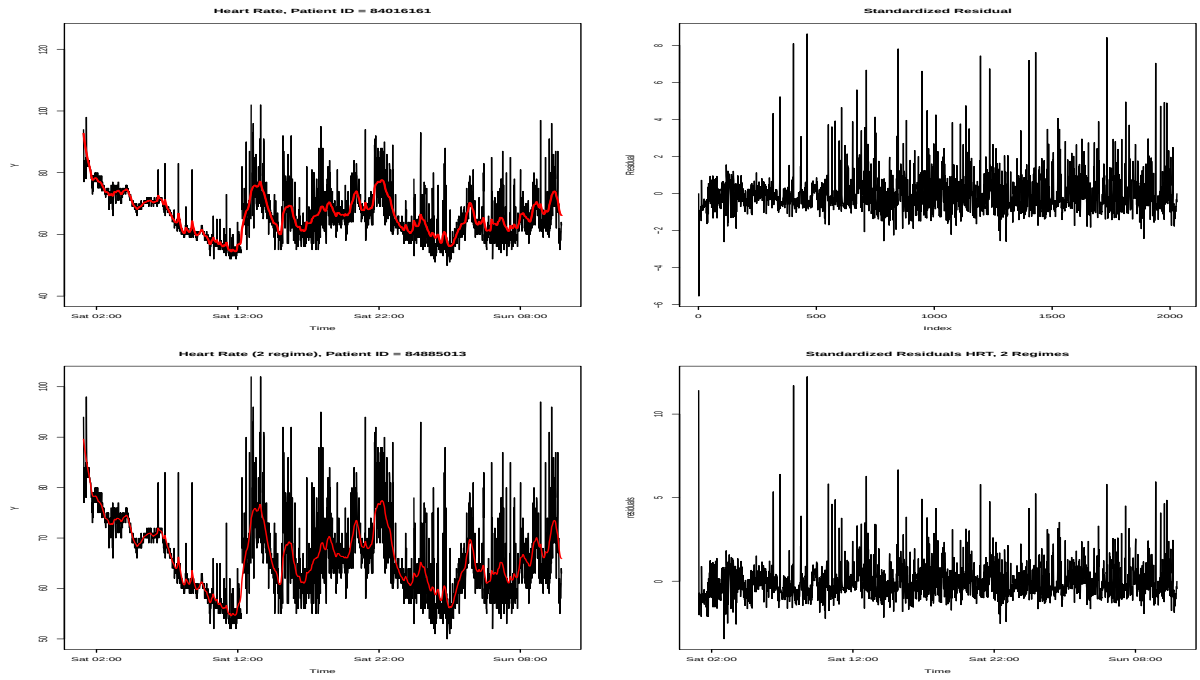

Figure 8–2: Left: Posterior estimates of the state vector; Right: standardized forecasting residuals. Top: Single Regime Model; Bottom: Two regime model.

The plot on the right hand side is the estimate of the standardized residuals from the one step ahead forecast. Given the raw residuals and the GARCH parameter estimates, we can calculate the variance at time t. Thus by scaling the raw residual at time t by the variance at that time gives us the standardized residual. We see that in this series our standardized residuals are evenly spread around zero. There still are a few number of large deviations from zero in both the single regime and the two-regime model. It is actually quite difficult to notice the difference between the two residual plots. So instead we look at the RMSE for both models. The single regime model has an RMSE of 1.2735 while two-regime model has an RMSE of 1.2554.

The Normal-QQ plot also shows us that our residuals seem to have very heavy tails. Just like the residual plots the QQ plots are hard to distinguish from each other.

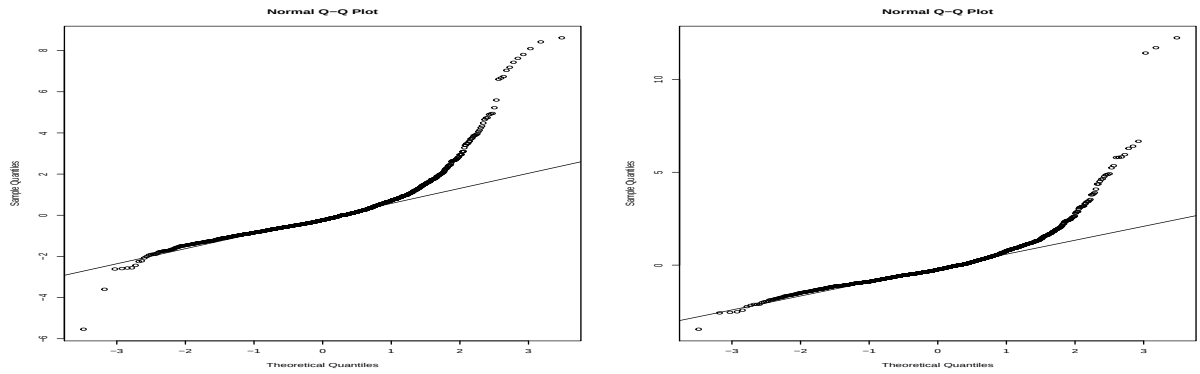

Figure 8–3: QQ-plot for the standardized residuals.

Overall the model seems to be fitting well. The residual plot shows why fitting a constant variance model like the standard Gaussian-SSM would be problematic. The large deviations that we see in the residual plot would not be properly captured by the constant variance Gaussian-SSM.

Change points analysis for the Heart Rate series:

In the last part we fixed the number of regimes to two and analyzed the data. In this part we can use our reversible jump technology from Chapter 7 to estimate the number of regimes in the model. Below is the histogram for the posterior estimate of the number of change points in the model. (We remind the reader that the reversible jump technology is assuming Gaussian-SSMs.)

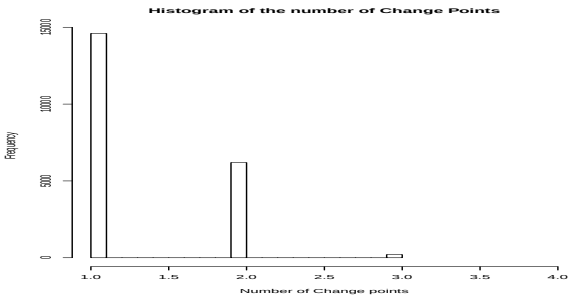

Figure 8–4: Plot for the posterior estimate of the regimes.

We see above, a large mass on having one change point. The two change point model has about half as much mass. The three change point model has negligible mass and higher change points have zero mass. In this model the $r_{\max}$, which is the maximum number of change points, was set at 10. We see that model is quite reasonable. Because if we look at the heart rate series in Figure 8-1, we see that there seems to be a structural break that occurs at around $t = 600$.

The posterior median estimate of the change point for the one change point model (two-regime model) is around 554.8. Which is very similar to the estimate we had of the change point in Table 8-2.

In the previous part we had used the WAIC to suggest that the two-regime model is better than the one regime model and the reversible jump is giving us a similar result. Thus from this, it would be a reasonable assumption to take the two-regime point model is probably the best model to fit this data.

8.2 Estimating the Blood Pressure Series:

Next we analyze the blood pressure series (BP). As before we start with the comparison between the single regime GARCH-SSM and the Gaussian-SSM before doing a two-regime model.

GARCH-SSM					
Sample size	α_0	α_1	β_1	W	WAIC
$T = 2029$	1.175	0.275	0.712	0.617	-6067.63
	(0.0972)	(0.0019)	(0.0019)	(0.0068)	
Gaussian-SSM					
	V	W			WAIC
	19.95	1.074			-6204.86
	(0.6062)	(0.0359)			

Table 8–3: Posterior estimates of the parameters for the GARCH SSM. In parenthesis are the standard errors of the estimates.

For this series as well, we see that using WAIC for model selection, we have that the GARCH model has a much higher WAIC (-6067.63) than the Standard Gaussian SSM (-6204.86), thus giving us that GARCH model is more appropriate than the standard Guassian-SSM.

For the blood pressure series, we have that the parameter estimates for $\hat{\alpha}_1 = 0.275$ and $\hat{\beta}_1 = 0.712$, which are also really close to the non-stationarity region. For the blood pressure series the estimate of the unconditional observational variance is,

$$Var(BP) = \frac{1.175}{1 - 0.275 - 0.712} = 90.38462.$$

Next we estimate the two-regime model and the parameter estimates are given in Table-12. We see in this case as well that the WAIC for the two-regime GARCH-SSM is -5954.372, which is higher than the single regime GARCH-SSM. In terms of the GARCH parameter

estimates, for both regimes the GARCH parameters are well away from the non-stationary region. We also have that the unconditional variance of each regime is less than the unconditional variance from the single regime model. The posterior median estimate for the change point is 692.84. Both regimes are long enough for the estimates to be reasonable.

$$Var(Y^1) = \frac{\hat{\alpha}_0^1}{1 - \hat{\alpha}_1^1 - \hat{\beta}_1^1} = \frac{2.0253}{1 - 0.5652 - 0.2130} = 9.1311$$
$$Var(Y^2) = \frac{\hat{\alpha}_0^2}{1 - \hat{\alpha}_1^2 - \hat{\beta}_1^2} = \frac{12.6645}{1 - 0.1826 - 0.4247} = 32.2499$$

Regime 1	α_0^1	α_1^1	β_0^1	W^1
	2.0253	0.5652	0.2130	0.4369
	(0.20836)	(0.01007)	(0.01144)	(0.00759)
Regime 2	α_0^2	α_1^2	β_0^2	W^2
	12.6645	0.1826	0.4247	0.2593
	(9.11390)	(0.00293)	(0.01261)	(0.00771)
Change Point 1		WAIC		
692.84		-5954.372		

Table 8–4: Posterior median of the GARCH parameters based on 5000 draws from the posterior distribution. In parenthesis are the standard errors of the estimates.

The residual plots below gives us that the standardized residuals are quite evenly spread across a mean of zero. Again it is difficult to differentiate between the residual plots of the single regime model and the residuals from the two-regime model. Thus we use the RMSE again. The RMSE of the single regime model is 1.1528 and the RMSE of the two-regime model is 1.1251. Both models seem to be capturing the mean level of the model quite well.

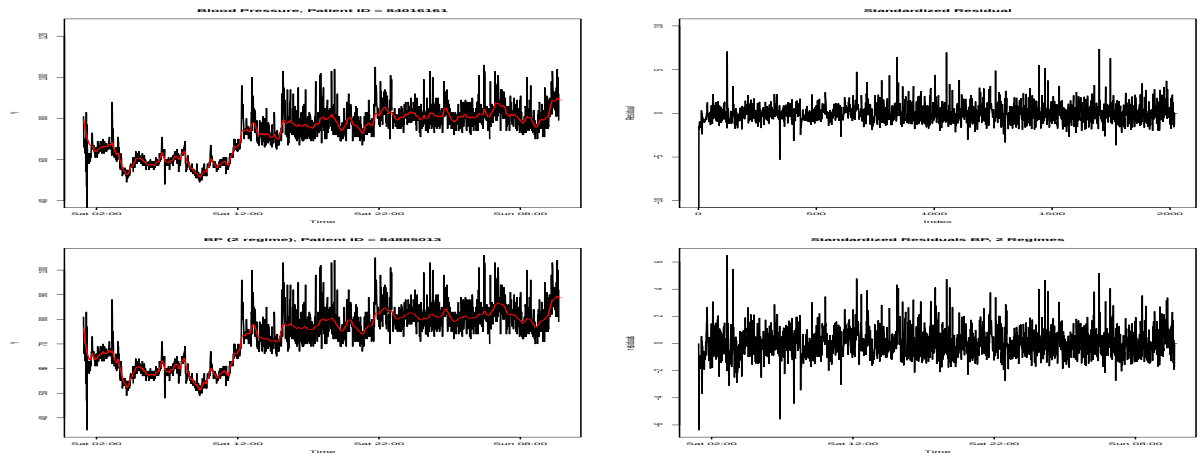

Figure 8–5: Left: Posterior estimates of the state vector; Right: standardized forecasting residuals. Top: Single Regime Model; Bottom: Two regime model.

The comments pertaining to the QQ plots are similar to that of the heart rate series. It is difficult to distinguish one plot from another. The one difference between the heart rate series and the blood pressure series is that the blood pressure series seems to contain less outliers compared to the heart rate series.

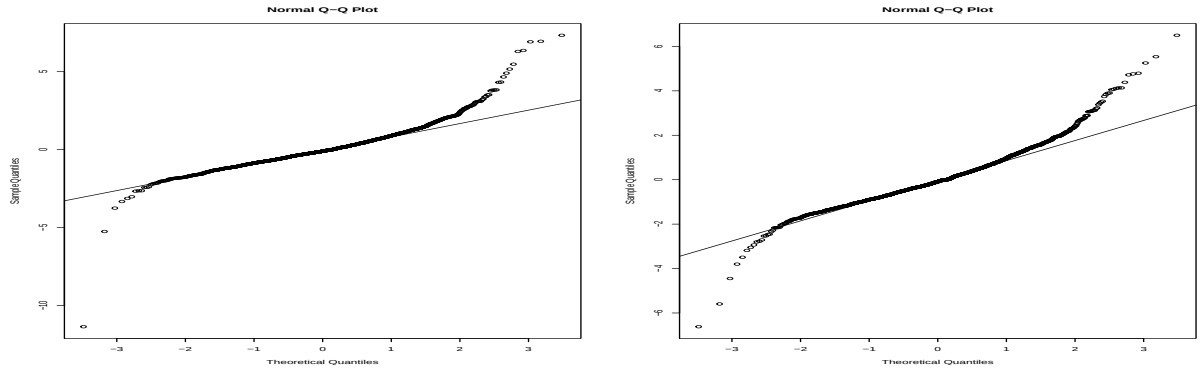

Figure 8–6: QQ-plot standardized forecasting residuals.

Change points analysis for the Blood Pressure series:

Like for the heart rate series, we do a estimate the number of regimes in the blood pressure series using the Reversible Jump technology. The histogram below gives the posterior estimate of the number of regimes.

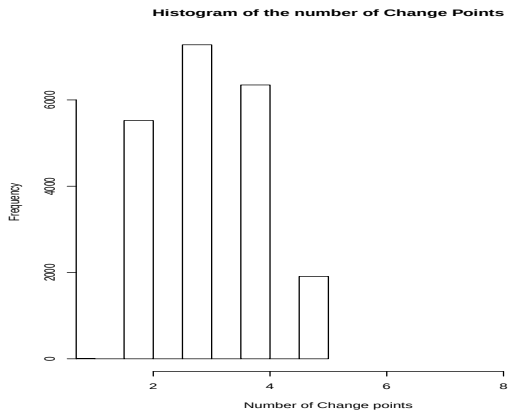

Figure 8–7: Plot for the posterior estimate of the regimes.

The results are interesting because unlike the heart rate series where we saw that the two-regime model was dominating, in this case we see the four-regime model (3 change-points model) being the dominating one closely followed by the three-regime and then the five regime model. In fact the two-regime model has almost zero mass in this analysis. Thus for the blood pressure series the two-regime model is perhaps not the most adequate model.

We give the estimate of the parameters from the four-regime model below. We see from the table that the middle change point around $t = 646$ is very close to the change point from Table 8-4. Other than that the model is picking up change points around $t = 452$ and a very late change point at $t = 1886$.

	Regime 1	Regime 2	Regime 3	Regime 4
V	9.288	1.887	26.196	29.213
	(3.678)	(0.171)	(2.397)	(20.272)
W	0.924	0.764	1.317	0.485
	(0.072)	(0.085)	(0.146)	(0.162)
	Change Point 1	Change Point 2	Change Point 3	
	452.194	646.115	1886.376	

Table 8–5: Posterior median estimates of the four-regime model for the blood pressure series. In parenthesis are the standard errors of the estimates.

8.3 Another Example:

While estimating the models we might also want to check if there is a general preference of the WAIC to always choose the GARCH-SSM. The example below shows that this need not be the case. The heart rate series from another patient is plotted below. The estimate for the single regime GARCH-SSM and the Gaussian-SSM are given in the table below. We see that the WAIC for the GARCH model (-12487.51) is lower than the WAIC for the standard Gaussian SSM (-12477.5). Thus for this patient we would feel comfortable using the Gaussian-SSM instead of the GARCH-SSM.

GARCH-SSM					
Sample size	α_0	α_1	β_1	W	WAIC
$T = 4194$	2.016	0.428	0.539	1.382	-12487.51
	(0.1027)	(0.0012)	(0.0013)	(0.0143)	
Gaussian-SSM					
	V	W			
$T = 4194$	13.000	3.776			-12477.5
	(0.1947)	(0.0830)			

Table 8–6: Posterior estimates of the parameters for the GARCH SSM. In parenthesis are the standard errors of the estimates.

8.4 Bivariate Model:

In this section we do a joint modeling of the HRT and BP series. Recall that from the previous sections we had that for both models the GARCH model is an appropriate model. We would thus expect that for the bivariate model a bivariate GARCH model should be more appropriate. Thus we estimate this model using the CCC-GARCH introduced in Chapter 4. We also estimate a bivariate Gaussian-SSM to see how it compares with the GARCH-SSM compares.

Sample size	V_{HR}	V_{BP}	$V_{HR,BP}$
$T = 2029$	25.168	19.813	1.213
	(0.873)	(0.534)	(0.358)
	W_{HR}	W_{BP}	$W_{HR,BP}$
	1.287	0.93	0.15
	(0.038)	(0.02)	(0.005)
	WAIC		
	-24676.8		

Table 8–7: Bivariate Gaussian-SSM estimates of the parameters. Here $V_{HR,BP}$ is the estimate of the correlation between the observation errors of the HRT series and the BP series and $W_{HR,BP}$ is the estimate of the correlation between the state errors of both series.

In the estimates above we have that

Sample size	α_0^{HR}	α_1^{HR}	β_1^{HR}	W^{HR}	ρ
$T = 2029$	0.876	0.237	0.749	3.871	0.039
	(0.0592)	(0.0019)	(0.0019)	(0.2650)	(0.0009)
	α_0^{BP}	α_1^{BP}	β_1^{BP}	W^{BP}	$W_{HR,BP}$
	0.350	0.173	0.818	2.660	0.824
	(0.0116)	(0.0009)	(0.0008)	(0.0978)	(0.0460)
	WAIC				
	-24522.08				

Table 8–8: Standard Gaussian SSM estimates of the parameters of other patients.

Recall that in this model ρ is the correlation between the GARCH errors of both series and $W_{HR,BP}$ is the correlation between the state observations of both series.

Like in the univariate case, we have that for the bivariate case the GARCH-SSM is a better model given the higher value of the WAIC for the GARCH model than the Gaussian model. In terms of the estimate of the correlation, ρ, the estimate of ρ is quite small thus indicating that there may not be any significant correlation between the observation errors.

CHAPTER 9
Conclusion:

In this thesis our goal was to develop a SSM able to deal with heteroskedastic data. The GARCH-SSM developed allows us to deal with time varying variance. This model is also able to deal with multiple structural breaks. We have been able to extend the model to the multivariate case as well. The Gaussian-SSM has been used as a benchmark to compare with the GARCH-SSM. Our analysis of the *Brain IT* data has shown that the GARCH-SSM with structural breaks is a more appropriate model for the HRT and BP series than the Gaussian-SSM.

We carried simulations studies to show that our model is able to estimate the parameters and the state vectors. We also showed how we would be able to estimate the a multiple regime model. Using the WAIC we showed how we might choose between the Gaussian-SSM, GARCH-SSM and multiple regime GARCH-SSM. Finally we saw how the reversible jump technology might be used to select the number of regimes that the model is operating under.

9.1 Future Work and Limitations:

While we implemented the reversible jump technology in the univariate case, applying it to the multivariate case would be an interesting problem to look at. In this thesis we also did not look at the reversible jump technology for the GARCH-SSM and this is also something that can be developed. In fact the reversible jump technology is quite versatile in terms of validating model adequacy and its use is not only limited to assess the number of regimes. We should also be able to use it to chose between a local level model and a linear growth model, for example. We saw in the univariate case during our simulation

that the estimation of the number of regimes is quite sensitive to the prior specification of the model. We suspect this would be a similar issue when testing different model hypotheses.

Currently there is not any software that can handle models with unknown types of structural breaks and multivariate models from a Bayesian point of view. This something that our model can handle and given their utility it might be worthwhile to develop such a software. However, that being said, these models are computationally quite challenging. The FFBS algorithm requires running the Kalman Filter at every iteration. While this is simple enough in the univariate case and when the series are not too long, in the higher dimensional case and/or with very large series the estimation becomes quite slow. Even with using `RCPP`, which integrates `C++` with `R` and works at lightening speed compared to standard `R`, estimating some these larger models is quite exhausting.

Further, with data that display a large amount of variability and large changes in trend it is often difficult to discern between where the variability in the series is coming from, the states or the observation.

In general our algorithm was not sensitive to starting values and in almost all cases was able to retrieve the true values of the parameter, however, for the observation error in many cases we actually see a significant correlation between the α_0 and β_1 parameter of the GARCH model. This may effect the convergence of the MCMC thus requiring larger numbers of posterior simulations.

REFERENCES

Bernardo, J. M. and Smith, A. F. (1994). *Bayesian Theory. 1994.*

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). Arch models. *Handbook of Econometrics*, 4:2959–3038.

Bollerslev, T. et al. (1990). Modelling the coherence in short-run nominal exchange eates: A multivariate generalized arch model. *Review of Economics and Statistics*, 72(3):498–505.

Caines, P. E. (2018). *Linear Stochastic Systems*, volume 77. SIAM.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear time series: Theory, methods and applications with R examples*. Chapman and Hall/CRC.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202.

Gamerman, D., dos Santos, T. R., and Franco, G. C. (2013). A non-gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34(6):625–645.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on rtocks. *The Journal of Finance*, 48(5):1779–1801.

Gordon, K. and Smith, A. (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, 85(410):328–337.

Gouriéroux, C. (2012). *ARCH models and financial applications*. Springer Science & Business Media.

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

Hales, S. (1733). Statical essays: Vol. 2 containing haemastaticks; or, an account of some hydraulick and hydrostatical experiments made on the blood and blood-vessels of animals. *London, W. Innys and R. Manby*, 1733.

Hamilton, J. D. (1994). State-space models. *Handbook of Econometrics*, 4:3039–3080.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, volume 70. SIAM.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Jensen, J. L., Petersen, N. V., et al. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *The Annals of Statistics*, 27(2):514–535.

Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108.

Kalman, R. E. et al. (1960). Contributions to the theory of optimal control. *Bol. soc. mat. mexicana*, 5(2):102–119.

Kim, C.-J., Nelson, C. R., et al. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.

Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041.

Konishi, S. and Kitagawa, G. (2008). *Information criteria and statistical modeling*. Springer Science & Business Media.

Marin, J.-M. and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian statistics*. Springer Science & Business Media.

Petris, G., Petrone, S., and Campagnoli, P. (2009). Dynamic linear models. In *Dynamic Linear Models with R*, pages 31–84. Springer.

Prado, R. and West, M. (2010). *Time series: Modeling, Computation and Inference*. Chapman and Hall/CRC.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media.

Wong, K. F. K., Galka, A., Yamashita, O., and Ozaki, T. (2006). Modelling non-stationary variance in eeg time series by state space garch model. *Computers in Biology and Medicine*, 36(12):1327–1335.