12-2018

Sequential Inference for Hidden Markov Models

Michael Ellis
University of Arkansas, Fayetteville

Sequential Inference for Hidden Markov Models

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Statistics and Analytics

by

Michael Ellis
University of Arkansas
Bachelor of Science in Mathematics, 2016

December 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Giovanni Petris, Ph.D.
Thesis Director

John Tipton, Ph.D.
Committee Member

Jyotishka Datta, Ph.D.
Committee Member

Abstract

In many applications data are collected sequentially in time with very short time intervals between observations. If one is interested in using new observations as they arrive in time then non-sequential Bayesian inference methods, such as Markov Chain Monte Carlo (MCMC) sampling, can be too slow. Increasingly, state space models are being used to model nonlinear and non-Gaussian systems. The structure of state space models allows for sequential Bayesian inference so that an approximation to the posterior distribution of interest can be updated as new observations arrive. In special cases, the exact posterior distribution can be updated through conjugate Bayesian inference. However, for the general state space model this is not possible. In quantitative finance hidden Markov models have been used to analyze and forecast percent log returns of an asset or a group of assets. In this thesis the Liu and West [2001] auxiliary particle filter is applied to sequentially update the posterior distribution of a hidden Markov model with unknown state and observation distribution parameters.

Acknowledgements

I would like to thank my advisor Dr. Giovanni Petris for his guidance and support. I am very grateful for the opportunity to have learned from him. I would also like to thank the University of Arkansas Department of Mathematics. I feel incredibly lucky to have had the opportunity to learn and grow during my time here. Thank you to the many professors who guided, encouraged and taught me during this journey. Thank you to my friends and family for their encouragement and belief in me and thank you for patiently listening while I discussed my work with you.

Table of Contents

1 Introduction

A time series is a set of observations collected sequentially in time. Commonly, it is assumed the observations were collected at equally spaced time intervals. In this case, the time series can be denoted as $y_{1:T} := \{y_t : t = 1, 2, ..., T\}$ where y_t was observed at time t. Examples of time series data include stock market prices, daily average temperature, and daily sales counts. Given a time series we are typically interested in analyzing the historical observations and forecasting future observations. Outside of time series literature, many statistical models assume the observations are independent of each other. However, when collecting observations sequentially through time it is very common to see temporal dependence between the observations. For example, Figure 1.1 shows the daily closing price of the S&P 500 Index from June 4th, 2017 to June 4th, 2018. Figure 1.1 shows the trend of the daily closing price changing through time with a large structural break between January and March, a pattern that would have been missed if time had been ignored. If the assumption of independence is relaxed and it is assumed that previous observations in time can provide information about the current and future observations then inference and forecasting can be greatly improved.

To handle temporal dependence classical time series models such as, autogressive moving average (ARMA) models, assume the observed time series is stationary. Loosely, a time series is stationary if the mean and variance are constant over time [Chatfield, 2003, West and Prado, 2010]. If the time series is clearly non-stationary then the data must be transformed so they are approximately stationary. This assumption is very limiting and does not allow for modeling of time series with evolving dynamics, as seen in Figure 1.1. Recently, extensive research has been devoted to developing state space models, that allow for modeling of non-stationary time series [Petris et al., 2009, West et al., 1994, West and Prado, 2010, Koopman and J., 2001].

State space models provide a flexible and intuitive approach for handling many problems in time series analysis and include ARMA models as a special case. In general, state space models assume that the observed time series $y_{1:T}$ is the noisy measurement of some underlying

Figure 1.1: Daily Closing Price of S&P 500 in US Dollars

and possibly unobserved structural or state process, denoted by $\theta_{0:T} := \{\theta_t : t = 0, 1, ..., T\}$. The goal then is to filter out the noise from the observations in order to estimate the latent state process. Furthermore, in many applications observations are received at very short time intervals and the structure of state space models allows for computationally efficient sequential Bayesian inference and forecasting [Petris et al., 2009].

Under certain conditions the posterior distribution of interest can be updated sequentially in closed form. For example, if the structure of the state space model is defined so that the observations and states evolve according to a linear function and the corresponding distributions are Gaussian then the posterior distribution can be updated using the Kalman filter [Kalman, 1960]. The current Gaussian posterior distribution becomes a prior distribution when a new observation is received and given the new observation the Kalman filter updates the mean and variance of the prior distribution to form the new Gaussian posterior distribution. However, for a general nonlinear and non-Gaussian state space model computing the closed form posterior distribution is not possible. The use of nonlinear and non-Gaussian state space models has greatly increased. Among many other applications these models have been used for automatic speech recognition, modeling volatility of stock price returns, and tracking animal movement [Zucchini et al., 2016, Petris et al., 2009]. And so, extensive research has been devoted to developing sequential Monte Carlo methods to compute and update approximations to the posterior distribution as new observations are received [Petris et al., 2009, Arulampalam et al., 2002, Cappe et al., 2007, Doucet et al., 2001].

In this thesis we will provide a brief overview the general state space model in Chapter 2. In addition, we will discuss optimal state estimation. In Chapter 3 we will outline several of the common and well known sequential Monte Carlo methods. Finally, in Chapter 4 we apply the Liu and West [2001] auxiliary particle filter to a hidden Markov model to model the volatility of daily log returns of the S & P 500.

2 General State Space Models

A general state space model is defined by

$$y_t \sim p(y_t|\theta_t, \phi), \tag{2.1}$$

$$\theta_t \sim p(\theta_t|\theta_{t-1}, \phi) \quad t \geq 1 \tag{2.2}$$

and an initial density $\theta_0 \sim p(\theta_0)$. y_t is the observation, θ_t is the latent state at time t, and ϕ is a vector of unknown parameters. The densities (2.1) and (2.2) are called the observation density and state density respectively. Using the above specification it is assumed that each y_t is conditionally independent of all other observations given the latent state θ_t and the underlying state process is a Markov chain. That is, θ_t depends only on the previous state θ_{t-1}.

Assuming we have observations $y_{1:T}$, the goal is to compute the posterior distribution $p(\theta_t, \phi|y_{1:T})$. If $t < T$ then we are using all available observations to estimate a state at a previous time period. This is referred to as smoothing. If $t = T$ then we are using all available observations to estimate the current state. This is referred to as filtering. And if $t > T$ then we are using all available observations to estimate a future state. This is referred to as forecasting [Petris et al., 2009].

2.1 State Estimation

In this section we derive the closed form equations to sequentially update the filtering distribution, $p(\theta_t|y_{1:t})$, once a new observation is received.

Using Bayes' theorem and the conditional independence of the observations the filtering distribution $p(\theta_t|y_{1:t})$ can be factored as

$$p(\theta_t|y_{1:t}) = \frac{p(\theta_t, y_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} = \frac{p(y_t|\theta_t)p(\theta_t|y_{1:t-1})}{\int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})d\theta_t}.$$

In the above equation $p(y_t|\theta_t)$ is the observation density defined by the model. However, the one step ahead forecast density, $p(\theta_t|y_{1:t-1})$, needs to be computed recursively. Using the Markov property of the states, the density can be factored as

$$p(\theta_t|y_{1:t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})d\theta_{t-1}.$$

The state density $p(\theta_t|\theta_{t-1})$ is defined by the model and $p(\theta_{t-1}|y_{1:t-1})$ is known from the previous step so it does not have to be recomputed. Therefore, updating the filtering distribution is a two step process. First, forecast the state one step ahead using equation 2.3 then once a new observation is received update the forecast once a new observation is received using equation 2.1 [Doucet et al., 2001].

$$p(\theta_t|y_{1:t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})d\theta_{t-1}. \tag{2.3}$$

$$p(\theta_t|y_{1:t}) = \frac{p(y_t|\theta_t)p(\theta_t|y_{1:t-1})}{\int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})d\theta_t} \quad t > 1. \tag{2.4}$$

Often an estimate for the entire state path up to time t is needed. The goal then is to compute the joint posterior distribution $p(\theta_{0:t}|y_{1:t})$ and sequentially update the estimate as new observations are received. As before, using an application of Bayes' theorem and the properties of the state space model the joint distribution can be factored as

$$p(\theta_{0:t}|y_{1:t}) = \frac{p(y_t|\theta_t)p(\theta_t|\theta_{t-1})}{p(y_t|y_{1:t-1})}p(\theta_{0:t-1}|y_{1:t-1}) \tag{2.5}$$

$$= \frac{p(y_t|\theta_t)p(\theta_t|\theta_{t-1})}{\int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})d\theta_t}p(\theta_{0:t-1}|y_{1:t-1}). \tag{2.6}$$

The density $p(\theta_{0:t-1}|y_{1:t-1})$ is known from the previous step, $p(y_t|\theta_t)$ and $p(\theta_t|\theta_{t-1})$ are defined by the model and $p(\theta_t|y_{1:t-1})$ can be computed using equation 2.4. Therefore, similar to how the filtering distribution is updated the posterior distribution $p(x_{0:t-1}|y_{1:t-1})$ can be updated in two steps, prediction and updating. First, use equation 2.7 to forecast the state one step ahead then use

equation 2.8 [Doucet et al., 2001].

$$p(\theta_t|y_{1:t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})d\theta_{t-1}. \tag{2.7}$$

$$p(\theta_{0:t}|y_{1:t}) = \frac{p(y_t|\theta_t)p(\theta_t|\theta_{t-1})}{\int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})d\theta_t} p(\theta_{0:t-1}|y_{1:t-1}) \quad t > 1. \tag{2.8}$$

Kalman [1960] introduced a solution, known as the Kalman Filter, for linear Gaussian state space models. However, in most cases computing the distributions defined in 2.7 and 2.8 is not possible given that they can involve high dimensional integrals. The use of nonlinear and non-Gaussian state space models has greatly increased. Therefore, extensive research has been devoted to developing Monte Carlo methods to sequentially update approximations of the filtering distribution for general state space models. This class of algorithms are commonly known as Sequential Monte Carlo methods or Particle Filters [Cappe et al., 2007, Doucet et al., 2001, Arulampalam et al., 2002, Petris et al., 2009].

3 Sequential Monte Carlo Methods

In this chapter we give an overview of some common sequential Monte Carlo methods. Similar summaries are given by [Cappe et al., 2007, Arulampalam et al., 2002, Doucet et al., 2001, Petris et al., 2009]. Recall that given the filtering density $p(\theta_t|y_{1:t})$ or the joint density $p(\theta_{0:t}|y_{1:t})$ at time t, the goal is to update the density when a new observation is received. Sequential Monte Carlo (SMC) methods achieve this by extending importance sampling.

3.1 Importance Sampling

Suppose we wish to calculate the expectation of a function $f(x)$ with respect to the density, $p(x)$. Computing the necessary integral is either not possible or computationally expensive. Instead we can introduce an importance density $q(x)$ and write the expectation of a function $f(x)$ with respect to $p(x)$ as

$$E_p[f(x)] = \int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx = E_q\left[f(x)\frac{p(x)}{q(x)}\right].$$

To calculate this expectation draw the samples from the importance density $q(x)$ and compute the Monte Carlo approximation

$$E_q\left[f(x)\frac{p(x)}{q(x)}\right] = \int f(x)\frac{p(x)}{q(x)}q(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})\frac{\tilde{w}^{(i)}}{\sum_{i=1}^{N}\tilde{w}^{(i)}}.$$

Where, $\tilde{w}^{(i)} = \frac{p(x^{(i)})}{q(x^{(i)})}$ are the unnormalize importance weights. The samples $\{x^{(i)} : i = 1, 2, ..., N\}$ are commonly referred to as particles. These particles and corresponding weights, denoted $\{(x^{(i)}, w^{(i)}) : i = 1, 2, ..., N\}$, are a Monte Carlo approximation of the density $p(x)$. That is, we can define the approximation to $p(x)$ as

$$\hat{p}(x) = \sum_{i=1}^{N} w^{(i)}\delta_{x^{(i)}},$$

where $\delta_{x^{(i)}}$ is the Dirac delta function centered at $x^{(i)}$ and $w^{(i)}$ are the normalized importance weights defined above.

For the special case in which we can draw samples directly from the target density $p(x)$ the importance density $q(x)$ will be equal to $p(x)$ and all of the importance weights will be equal to one. If the importance density is approximately equal to the target density then all of the normalized importance weights will be close to one. However, if the importance density is far from the target density then a small number of the particles will have large corresponding importance weights while most of the particles will have importance weights approximately equal to zero. This leads to a small effective sample size and a poor Monte Carlo approximation. To counteract this the particles can be resampled. By resampling, particles with large importance weights are replicated and particles with small importance weights are removed. There are several ways to resample the particles. An obvious approach is known as multinomial resampling. Given a Monte Carlo approximation, $\{(x^{(i)}, w^{(i)}) : i = 1, 2, , N\}$, the particles can be resampled by drawing a sample of size N from the set $\{x^{(i)} : i = 1, 2, ..., N\}$ where the probability of selecting $x^{(i)}$ is equal to the importance weight, $w^{(i)}$. This results in a new set of equally weighted particles that is now a new Monte Carlo approximation to the density $p(x)$. However, this resampling step increases the Monte Carlo variance so resampling methods designed to limit the increase in variance have been developed [Cappe et al., 2007, Petris et al., 2009].

3.2 Particle Filters

Particle filters extend importance sampling to sequentially update a Monte Carlo approximation to the joint density $p(\theta_{0:t}|y_{1:t})$. We begin by reviewing methods that assume the parameters of the state and observation densities are known. For a general state space model we would like to estimate the distribution by sampling directly from $p(\theta_{0:t}|y_{1:t})$; however, in most cases this is not possible and it is also increasingly inefficient as time goes on. As in importance sampling an importance density, $q(\theta_{0:t}|y_{1:t})$, can be introduced. To compute a Monte Carlo approximation to $p(\theta_{0:t}|y_{1:t})$ start by drawing a large number of samples from the importance density and then cal-

culate the corresponding unnormalized importance weights as,

$$\tilde{w}_t^{(i)} = \frac{p(\theta_{0:t}^{(i)}|y_{1:t})}{q(\theta_{0:t}^{(i)}|y_{1:t})}.$$

Finally, normalize the importance weights and the set $\{(\theta_{0:t}^{(i)}, w_t^{(i)}) : i = 1, 2, , N\}$ defines a Monte Carlo approximation to $p(\theta_{0:t}|y_{1:t})$. However, when a new observation is received the target density will change. Furthermore, it would be computationally efficient to update the current state path particles so that they provide an approximation to the new density rather than drawing an entirely new set of state paths each time we receive a new observation. Particle filters update the current Monte Carlo approximation to estimate the new density by assuming or constructing an importance density that evolves in a similar manner to the target density.

3.2.1 The Bootstrap Filter

Assume that at time t the set $\{(\theta_{0:t-1}^{(i)}, w_{t-1}^{(i)}) : i = 1, 2, , N\}$ is a Monte Carlo approximation to $p(\theta_{0:t-1}|y_{1:t-1})$ and a new observation, y_t, is received. In order to sequentially update the approximation it is assumed that the importance density can be factored as

$$q_t(\theta_{0:t}|y_{1:t}) = q_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})q_{t-1}(\theta_{0:t-1}|y_{1:t-1}).$$

The general idea is that the current samples $\theta_{0:t-1}^{(i)}$ were generated from $q_{t-1}(\theta_{0:t-1}|y_{1:t-1})$, the importance density at time $t-1$, and each of the state paths can by extend from $\theta_{0:t-1}^{(i)}$ to $\theta_{0:t}^{(i)}$ by drawing a new particle $\theta_t^{(i)}$ from $q_{t|t-1}(\theta_t|\theta_{t-1}^{(i)}, y_t)$. This assumption about the importance density allows us to update the unnormalized importance weights sequentially. The formula for updating

the unnormalized importance weights can be factored as

$$\begin{aligned}\tilde{w}_t^{(i)} &= \frac{p(\theta_{0:t}^{(i)}|y_{1:t})}{q_t(\theta_{0:t}^{(i)}|y_{1:t})} \\ &= \frac{p(y_t|\theta_t^{(i)})p(\theta_t^{(i)}|\theta_{t-1}^{(i)})}{q_{t|t-1}(\theta_t^{(i)}|\theta_{0:t-1}^{(i)}, y_{1:t})} \frac{p(\theta_{0:t-1}^{(i)}|y_{1:t-1})}{q_{t-1}(\theta_{0:t-1}^{(i)}|y_{1:t-1})} \\ &= \frac{p(y_t|\theta_t^{(i)})p(\theta_t^{(i)}|\theta_{t-1}^{(i)})}{q_{t|t-1}(\theta_t^{(i)}|\theta_{0:t-1}^{(i)}, y_{1:t})} w_{t-1}^{(i)}\end{aligned}$$

where, the importance density $q_{t|t-1}(\theta_t^{(i)}|\theta_{t-1}^{(i)}, y_t)$ is carefully chosen, $p(y_t|\theta_t^{(i)})$ and $p(\theta_t^{(i)}|\theta_{t-1}^{(i)})$ are defined by the model and $w_{t-1}^{(i)}$ are the importance weights computed at the previous step. This process defines an algorithm that updates the approximation of the filtering density at time $t-1$ to an approximation of the density at time t each time a new observation is received.

In summary, suppose at time t the set $\{(\theta_{0:t-1}^{(i)}, w_{t-1}^{(i)}) : i = 1, 2, ..., N\}$ defines a Monte Carlo approximation to $p(\theta_{0:t-1}|y_{1:t-1})$ and a new observation, y_t, is received. The approximation can be updated if, for $i = 1, 2, ..., N$:

- extend $\theta_{0:t-1}^{(i)}$ to $\theta_{0:t}^{(i)}$ by sampling $\theta_t^{(i)}$ from $q_{t|t-1}(\theta_t|\theta_{0:t-1}^{(i)}, y_{1:t})$,
- update the corresponding unnormalized importance weight, $\tilde{w}_t^{(i)}$, according to

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

After, normalizing the importance weights and the new set of particles and weights define a Monte Carlo approximation to $p(\theta_{0:t}|y_{1:t})$. The approximation can be written as $p(\theta_{0:t}|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{\theta_{0:t}^{(i)}}$. If only the filtering distribution is need then the weights remain the same and the state path up to the current time can be dropped. That is, the approximation to the filtering distribution is given by $p(\theta_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{\theta_t^{(i)}}$ [Petris et al., 2009].

When using the basic particle filter algorithm a proper importance density needs to be defined. In practice, even with a well defined importance density after several iterations almost all

of the importance weights are approximately zero so only a few particles contain almost all of the weight. To counteract this problem a resampling step as described in Section 3.1 can be included when the effective sample size falls below a certain threshold [Petris et al., 2009]. As in Petris et al. [2009], the effective sample size is defined as $N_{eff} = (\sum_{i=1}^{N}(w_t^{(i)})^2)^{-1}$ and resampling occurs when N_{eff} is less than half the number of samples drawn, N.

When defining the importance density it is common to use the state density. That is, $q_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t}) = p(\theta_t|\theta_{t-1})$. This choice is convenient because it is easy to draw samples from and the equations to update the importance weights simplify nicely. However, this choice often leads to poor performance [Petris et al., 2009]. The state density will likely be far from the target density and this can worsen the weight degeneracy problem or this can result in resampling particles that have relatively high importance weights but still provide a poor Monte Carlo approximation. A better choice for importance density is to use the new observation y_t to help guide the generation of new particles. If possible we could compute the density $q_{t|t-1}(\theta_t|\theta_{t-1}, y_t)$ directly. This is commonly defined as the optimal importance density in the sense that the conditional variance of the importance weights is zero [Cappe et al., 2007, Petris et al., 2009]. Unfortunately, directly computing the optimal importance density is not possible for a general state space model.

3.2.2 Auxiliary Particle Filter

Pitt and Shephard [1999] introduced the Auxiliary Particle Filter (APF) which updates the approximation to the posterior density while avoiding the potential problem of computing the optimal importance density for the bootstrap filter. Suppose at time $t-1$, $\hat{p}(\theta_{0:t-1}|y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{\theta_{0:t}^{(i)}}$

defines a Monte Carlo approximation to $p(\theta_{0:t-1}|y_{1:t-1})$ then it follows that

$$\begin{aligned} p(\theta_{0:t}|y_{1:t}) &= \frac{p(y_t|\theta_t)p(\theta_t|\theta_{t-1})}{p(y_t|y_{1:t-1})}p(\theta_{0:t-1}|y_{1:t-1}) \\ &\propto p(y_t|\theta_t)p(\theta_t|\theta_{t-1})p(\theta_{0:t-1}|y_{1:t-1}) \\ &\approx p(y_t|\theta_t)p(\theta_t|\theta_{t-1})\hat{p}(\theta_{0:t-1}|y_{1:t-1}) \\ &= \sum_{i=1}^{N} p(y_t|\theta_t)p(\theta_t|\theta_{t-1}^{(i)})w_{t-1}^{(i)}\delta_{\theta_{0:t-1}^{(i)}}. \end{aligned}$$

An auxiliary variable can be introduced and then the auxiliary target distribution is given by

$$\hat{p}(\theta_{0:t}, i|y_{1:t}) \propto p(y_t|\theta_t)p(\theta_t|\theta_{t-1}^{(i)})w_{t-1}^{(i)}\delta_{\theta_{0:t-1}^{(i)}}$$

To update the approximation Pitt and Shephard [1999] proposed using the importance density,

$$q_{t|t-1}(\theta_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)}p(y_t|\hat{\theta}_t^{(i)})p(\theta_t|\theta_{t-1}^{(i)})\delta_{\theta_{0:t-1}^{(i)}}.$$

Where, $\hat{\theta}_t^{(i)}$ is some likely value, such as the mean or mode, of the state density $p(\theta_t|\theta_{t-1}^{(i)})$. Using this importance density the Monte Carlo approximation can be updated if, for $k = 1, 2, ..., N$,:

- draw an auxiliary variable I_k from the set $\{i = 1, 2, ..., N\}$ with $p(I_k = i) \propto w_{t-1}^{(i)}p(y_t|\hat{\theta}_t^{(i)})$
- extend the state path from $\theta_{0:t-1}^{(I_k)}$ to $\theta_{0:t}^{(k)}$ by sampling $\theta_t^{(k)}$ from $p(\theta_t|\theta_{t-1}^{(I_k)})$
- update the unnormalized weights according to

$$\tilde{w}_t^{(k)} = \frac{w_{t-1}^{(I_k)}p(y_t|\theta_t^{(k)})p(\theta_t^{(k)}|\theta_{t-1}^{(k)})}{w_{t-1}^{(I_k)}p(y_t|\hat{\theta}_t^{(I_k)})p(\theta_t^{(k)}|\theta_{t-1}^{(k)})} = \frac{p(y_t|\theta_t^{(k)})}{p(y_t|\hat{\theta}_t^{(I_k)})}.$$

After discarding the auxiliary variables, normalizing the weights and resampling if needed and the new set of particles will define a Monte Carlo approximation to $p(\theta_{0:t}|y_{1:t})$ [Petris et al., 2009].

It is interesting to note that when an auxiliary variable is sampled certain state paths will have a relatively higher probability of being selected. A large value of $w_{t-1}^{(i)}$ suggest that the state $\theta_{t-1}^{(i)}$ is consistent with the previous observation. In theory $\hat{\theta}_t^{(i)}$ is a state that is likely to show up at time t given $\theta_{t-1}^{(i)}$. So, for each state path we find a likely next state for time t. If the previous state $\theta_{t-1}^{(i)}$ is consistent with the previous observation and the observation density given a likely next state in that path, $p(y_t|\hat{\theta}_t^{(i)})$, is large then that state path will have a relatively large probability of being selected. In theory we should select state paths that will likely evolve to a state that is "consistent" with the new observation [Liu and West, 2001].

3.2.3 Auxiliary Particle Filter with Unknown Parameters

Until now we have discussed methods for general nonlinear and non-Gaussian state space models with known observation and state density parameters. Now assume that the parameter vector ϕ is unknown. The goal is to update the estimate of the joint posterior density $p(\theta_{0:t-1}, \phi|y_{1:t-1})$ sequentially as new observations are received. A method proposed by Liu and West [2001] extends the auxiliary particle filter by using a kernel smoothing technique to approximate the marginal density of the unknown parameter vector, ϕ, given the observations $y_{1:t-1}$.

In a similar process to the APF, suppose at time $t-1$ the set $\{(\theta_{0:t-1}^{(i)}, \phi^{(i)}, w_{t-1}^{(i)}) : i = 1, 2, ..., N\}$ defines a Monte Carlo approximation to $p(\theta_{0:t-1}, \phi|y_{1:t-1})$ then it follows that,

$$p(\theta_{0:t}, \phi|y_{1:t}) \propto p(y_t|\theta_t, \phi)p(\theta_t|\theta_{t-1}, \phi)p(\theta_{0:t-1}, |y_{1:t-1}, \phi)p(\phi|y_{1:t-1}).$$

Kernel density smoothing methods take an importance sample and create a weighted mixture of a continuous distribution to define a continuous approximation to the density of interest. The particle filtering algorithm proposed by Liu and West [2001] uses a kernel density smoothing technique to approximate $p(\phi|y_{1:t-1})$. Given the importance sample $\{(\phi^{(i)}, w_{t-1}^{(i)}) : i = 1, 2, ..., N\}$ let $\bar{\phi}$ and V be the mean and variance matrix of the importance sample then

$p(\phi|y_{1:t-1})$ can be approximated as

$$p(\phi|y_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} N(m^{(i)}, h^2 V).$$

Where, the mean $m^{(i)} = a\phi^{(i)} + (1-a)\bar{\phi}$ and the variance h^2V of the normal kernel density are defined so that the variance of the approximation is not increased by using a mixture of normal distributions. And, $a = \sqrt{1-h^2}$ is defined by a smoothing parameter, h [Liu and West, 2001]. Liu and West [2001] suggest using a value of a between 0.974 and 0.995. It follows that the density of interest can be approximated as,

$$\begin{aligned} p(\theta_{0:t}, \phi|y_{1:t}) &\propto p(y_t|\theta_t, \phi)p(\theta_t|\theta_{t-1}, \phi)p(\theta_{0:t-1}, |y_{1:t-1}, \phi)p(\phi|y_{1:t-1}) \\ &\approx \sum_{i=1}^{N} w_{t-1}^{(i)} p(y_t|\theta_t, \phi)p(\theta_t|\theta_{t-1}, \phi)N(m^{(i)}, h^2V)\delta_{\theta_{0:t-1}^{(i)}}. \end{aligned}$$

Again, introduce auxiliary variable i so that the distribution of interest becomes

$$\hat{p}(\theta_{0:t}, \phi, i|y_{1:t}) \propto w_{t-1}^{(i)} p(y_t|\theta_t, \phi)p(\theta_t|\theta_{t-1}^{(i)}, \phi)N(m^{(i)}, h^2V)\delta_{\theta_{0:t-1}^{(i)}}.$$

To sample from the distribution introduce an importance density,

$$q_t(\theta_{0:t}, \phi, i|y_{1:t}) \propto w_{t-1}^{(i)} p(y_t|\theta_t = \hat{\theta}_t^{(i)}, \phi = m^{(i)})p(\theta_t|\theta_{t-1}^{(i)}, \phi)N(m^{(i)}, h^2V)\delta_{\theta_{0:t-1}^{(i)}}.$$

Where $\hat{\theta}_t^{(i)}$ is some likely value such as the mean, median or mode of the state density $p(\theta_t|\theta_{t-1} = \theta_{t-1}^{(i)}, \phi = m^{(i)})$. Then the approximation can be updated if, for $k = 1, 2, ..., N$:

- draw an auxiliary variable I_k from the set $\{i = 1, 2, ..., N\}$ with $p(I_k = i) \propto w_{t-1}^{(i)} p(y_t|\theta_t = \hat{\theta}_{t-1}^{(i)}, \phi = m^{(i)})$
- draw a new fixed parameter vector $\phi^{(k)}$ from $N(m^{(i)}, h^2V)$
- extend the state path from $\theta_{0:t-1}^{(I_k)}$ to $\theta_{0:t}^{(k)}$ by drawing $\theta_t^{(k)}$ from $p(\theta_t|\theta_{t-1} = \hat{\theta}_{t-1}^{(I_k)}, \phi = \phi^{(k)})$

- update the unnormalized importance weights according to

$$\begin{aligned}\tilde{w}_t^{(k)} &= \frac{w_{t-1}^{(I_k)} p(y_t|\theta_t = \theta_t^{(k)}, \phi = \phi^{(k)}) p(\theta_t|\theta_{t-1}^{(k)}, \phi) N(m^{(I_k)}, h^2 V)}{w_{t-1}^{(I_k)} p(y_t|\theta_t = \hat{\theta}_t^{(I_k)}, \phi = m^{(I_k)}) p(\theta_t|\theta_{t-1}^{(k)}, \phi) N(m^{(I_k)}, h^2 V)} \\ &= \frac{p(y_t|\theta_t = \hat{\theta}_t^{(k)}, \phi = \phi^{(k)})}{p(y_t|\theta_t = \hat{\theta}_t^{(I_k)}, \phi = m^{(I_k)})}\end{aligned}$$

After, discarding the auxiliary variables, normalizing the weights and resampling if needed the new set of importance weights and particles define a Monte Carlo approximation to $p(\theta_{0:t}, \phi|y_{1:t})$ [Liu and West, 2001].

The Liu and West [2001] particle filter used normal kernels to smooth the importance sample $\{(\phi^{(i)}, w_t^{(i)}) : i = 1, 2, ..., N\}$ however this does not work if the support of the parameter of interest is not the real line. If this is the case then we can either transform the parameter so that its support is the real line or we can use a different continuous density for the kernel. If a different continuous density is used then, as Petris et al. [2009] suggests, we can match the first two moments to $m^{(i)}$ and V. An example given in Petris et al. [2009] used a gamma kernel for unknown variance parameters. Given an importance sample $\{(\phi^{(i)}, w_{t-1}^{(i)}) : i = 1, 2, ..., N\}$ that approximates $p(\phi|y_{1:t-1})$. Then for $i = 1, 2, ..., N$ the we can match the mean $\mu^{(i)} = \frac{\alpha^{(i)}}{\beta^{(i)}}$ and variance $\sigma^{2^{(i)}} = \frac{\alpha^{(i)}}{\beta^{2(i)}}$ of the gamma distributions to the mean $m^{(i)}$ and variance $h^2 V$ of the normal distribution suggest by Liu and West [2001]. Solving for $\alpha^{(i)}$ and $\beta^{(i)}$ we get that

$$\alpha^{(i)} = \frac{m^{(i)^2}}{h^2 V} \tag{3.1}$$

$$\beta^{(i)} = \frac{m^{(i)}}{h^2 V}. \tag{3.2}$$

Then $p(\phi|y_{1:t-1})$ can be approximated as a weighted mixture of gamma distributions, $\sum_{i=1}^{N} w_{t-1}^{(i)} gam(\alpha^{(i)}, \beta^{(i)})$. And like the normal mixture used in Liu and West [2001] the gamma mixture has the same mean and variance as the discrete Monte Carlo approximation given by $\{(\phi^{(i)}, w_t^{(i)}) : i = 1, 2, ..., N\}$.

4 Application

In quantitative finance modeling and forecasting daily returns of an asset or a group of assets is a large area of research. Let p_t be an asset price at time t then the percent log return is defined as $y_t = 100 * (log(p_t) - log(p_{t-1}))$. In a linear Gaussian state space model the variance of the observations is assumed to be constant through time. However, the variance of log returns notoriously vary over time and so the assumption of constant variance is not appropriate [Koopman and J., 2001]. Historically, the autoregressive conditional heteroskedasticity (ARCH) model and the generalized autoregressive conditional heteroskedasticity (GARCH) model have been used to model the changing variance by assuming the current variance is a function of the previous return residuals [Engle, 1982], [Bollerslev, 1996]. More recently stochastic volatility models assume the variance evolves according to a stochastic process. Another approach to model the changing variance is to use a hidden Markov model (HMM) [Zucchini et al., 2016]. HHMs are another type of state space model that assume the latent states are discrete random variables.

4.1 A Hidden Markov Model for a Financial Application

In general Hidden Markov models are mixture models that allow for temporal dependence between observations. Commonly a m-state HMM assumes that the latent state process is a discrete Markov chain with a transition probability matrix Γ defined as,

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix}$$

where $\gamma_{ij} = p(\theta_{t+1} = j|\theta_t = i)$. That is, the probability of moving to state j given the process is currently in state i is entry γ_{ij} in the transition probability matrix. The observation density depends on the current state, $y_t \sim p(y_t|\theta_t = i)$ for $i = 1, 2, \ldots, m$.

When using HHMs to model the changing variance of log returns the discrete states correspond to observation densities that have different variances. So as the returns transition between states the expected variance of the returns changes. We can interrupt the latent states as different levels of variance. For example, the returns may have the smallest relative variance when in state one, the second smallest variance when in state two and so on.

We will use a two state HMM to model percent log returns. Log returns are known to have several“stylized facts" such as a kurtosis greater than three and significant autocorrelation in absolute or squared returns [Zucchini et al., 2016, Bulla, 2011]. Bulla [2011] showed that a HMM with at least one state corresponding to Student's t-distribution reproduces“most of the stylized facts better than or comparably well" as a HMM with only Gaussian components. Furthermore, under the assumption of efficient markets log returns have a conditional expected value of zero [Petris et al., 2009]. Using this assumption and the results from Bulla [2011], the hidden Markov model we will use for the percent log returns is defined by equation 4.1.

$$y_t|\theta_t = i \sim \begin{cases} N(0, \sigma_1^2), & \text{if } i = 1 \\ t_{\nu_2}(0, \lambda_2^{-1}), & \text{if } i = 2 \end{cases} \tag{4.1}$$

If the current state is one, $\theta_t = 1$, then the observations follow a normal distribution with mean zero and variance σ_1^2 and if the current state is two, $\theta_t = 2$, then observations will follow a t-distribution with mean zero, scale parameter λ_2^{-1}, and degrees of freedom ν_2. The latent states evolve as a Markov Chain with a transition probability matrix $\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$. That is, the probability of moving to state i given state j is γ_{ij}. For identifiability it is assumed that $\sigma_1^2 < \lambda_2^{-1}$ and $2 < \nu_2 \leq 20$. The general idea behind this model is that state one will account for observations in a state of low variance while state two will account for outliers and observations with high variance.

The t-distribution for observations in state two can be represented as a scale mixture of normal distributions by introducing a latent variable ω_t for each observation in that state. That

is, if

$$y_t|\omega_t, \lambda_2 \sim N(0, (\omega_t\lambda_2)^{-1})$$
$$\omega_t|\nu_2 \sim \text{gamma}(\nu_2/2, \nu_2/2)$$

then the marginal distribution of y_t is $t_{\nu_2}(0, \lambda_2^{-1})$. The ω_t's are commonly used to detect outliers. If ω_t is far from one then this signals that the corresponding observation, y_t, is likely a outlier [Petris et al., 2009].

4.1.1 Sequential Inference for Hidden Markov Models

In the sequential Monte Carlo methods described in Chapter three we assume that we have an initial approximation to posterior distribution that can be updated as new observations are received. With little to no prior information this initial approximation may be very poor. However, if we have previous observations we can use these to construct an approximation to the current posterior distribution. One suggestion is to use a Markov Chain Monte Carlo (MCMC) sampler to initialize the particle filter [Petris et al., 2009]. Luckily, financial data is extremely abundant so for the examples below we run an MCMC sampler on all available observations up to some time t then use the Liu and West [2001] auxiliary particle filter to sequentially update the approximation.

For the HMM defined in 4.1 we are interested in the posterior distribution $p(\theta_{0:t}, \omega_{1:t}, \phi|y_{1:t})$ where $\phi = (\sigma_1^2, \lambda_2, \nu_2, \Gamma)$. The Liu and West [2001] auxiliary particle filter allows us to update the marginal distributions of the unknown parameters in the parameter vector, ϕ by drawing samples from a smoothed approximation to the parameter's marginal distribution. The marginal distributions of σ_1^2, λ_2 and ν_2 will be approximated using a gamma kernel with shape, α, and rate, β, parameters defined as in equations 3.1 and 3.2 so that the variance of the approximation is not increased. The state transition probability matrix Γ has two free parameters. To approximate marginal distribution of γ_{11} and γ_{21} a logistic transformation will be applied with a normal

density kernel. At time $t-1$ a MCMC sampler is used to approximate the posterior of interest $p(\theta_{0:t-1}, \omega_{1:t-1}, \phi|y_{1:t-1})$ then as a new observation is received at time t the posterior is updated to $p(\theta_{0:t}, \omega_{1:t}, \phi|y_{1:t})$ by applying the Liu and West [2001] auxiliary particle filter. That is, for $k = 1, ..., N$,

- draw an auxiliary variable I_k from the set $\{i = 1, 2, ..., N\}$ with $p(I_k = i) \propto w_{t-1}^{(i)} p(y_t|\theta_t = \hat{\theta}_t^{(i)}, \omega_t = \hat{\omega}_t^{(i)}, \phi = m^{(i)})$
- update each of the parameters by drawing
  - $\sigma_1^{2(k)}$ from gamma$(\alpha_{\sigma_1^2}^{(I_k)}, \beta_{\sigma_1^2}^{(I_k)})$
  - $\lambda_2^{(k)}$ from gamma$(\alpha_{\lambda_2}^{(I_k)}, \beta_{\lambda_2}^{(I_k)})$
  - $\nu_2^{(k)}$ from gamma$(\alpha_{\nu_2}^{(I_k)}, \beta_{\nu_2}^{(I_k)})$
  - $logit(\gamma_{11}^{(k)})$ from $N(m_{\gamma_{11}}^{(I_k)}, h^2 V_{\gamma_{11}})$
  - $logit(\gamma_{21}^{(k)})$ from $N(m_{\gamma_{21}}^{(I_k)}, h^2 V_{\gamma_{21}})$
- extend the state parameters from
  - $\omega_{1:t-1}^{(I_k)}$ to $\omega_{1:t}^{(k)}$ by drawing $\omega_t^{(k)}$ from $gam(\nu^{(I_k)}/2, \nu^{(I_k)}/2)$
  - $\theta_{0:t-1}^{(I_k)}$ to $\theta_{0:t}^{(k)}$ by drawing $\theta_t^{(k)}$ from $p(\theta_t|\theta_{t-1} = \hat{\theta}_{t-1}^{(I_k)}, \phi = \phi^{(k)})$
- update the unnormalized importance weights according to

$$\tilde{w}_t^{(k)} = \frac{p(y_t|\theta_t = \theta_t^{(k)}, \omega_t = \omega_t^{(k)}, \phi = \phi^{(k)})}{p(y_t|\theta_t = \theta_t^{(I_k)}, \omega_t = \omega_t^{(I_k)}, \phi = m^{(I_k)})}.$$

After discarding the auxiliary variables, normalizing the weights and resampling if needed and the new set of particles will define an approximation to the posterior $p(\theta_{0:t}, \omega_{1:t}, \phi|y_{1:t})$. When drawing the auxiliary variable I_k with probability $p(I_k = i) \propto w_{t-1}^{(i)} p(y_t|\theta_t = \hat{\theta}_t^{(i)}, \omega_t = \hat{\omega}_t^{(i)}, \phi = m^{(i)})$, $\hat{\theta}_t^{(i)}$ is defined to be the previous state, $\theta_{t-1}^{(i)}$, in that state path. This implies the observation

is likely to remain in the same state. And $\hat{\omega}_t^{(i)}$ is defined to be the median of $gam(\nu_2^{(i)}/2, \nu_2^{(i)}/2)$. A value of 0.975 was used for the smoothing parameter a, $h = \sqrt{1-a^2}$.

4.2 A Simulated Example

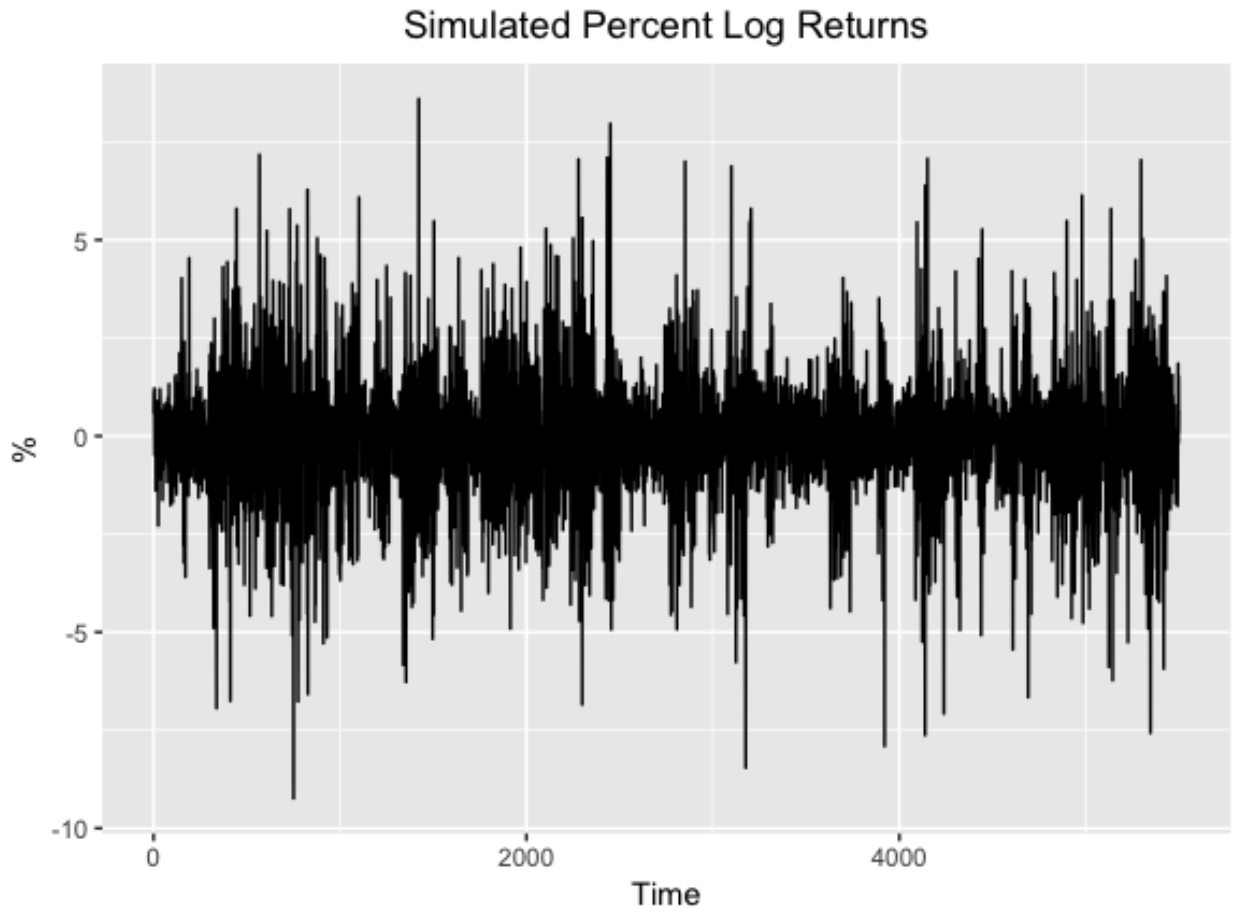

Figure 4.1: Simulated Daily Returns from HMM

To provided an example of estimating the known state and observation density parameters the Liu and West [2001] auxiliary particle filter is first applied to data simulated from the HMM defined in equation 4.1. Figure 4.1 shows 5500 data points simulated for the HMM defined by 4.1. A MCMC sampler using 70000 samples and a burnin period of 50000 samples was used to construct an approximation to the posterior at time t. The MCMC samples were then used to initialize the particle filter. Then for each new observation we update the approximation using the

Liu and West [2001] algorithm laid out in section 4.1.1. At each iteration 20,000 samples were used to approximate the distribution and the particles were resampled after every iteration.

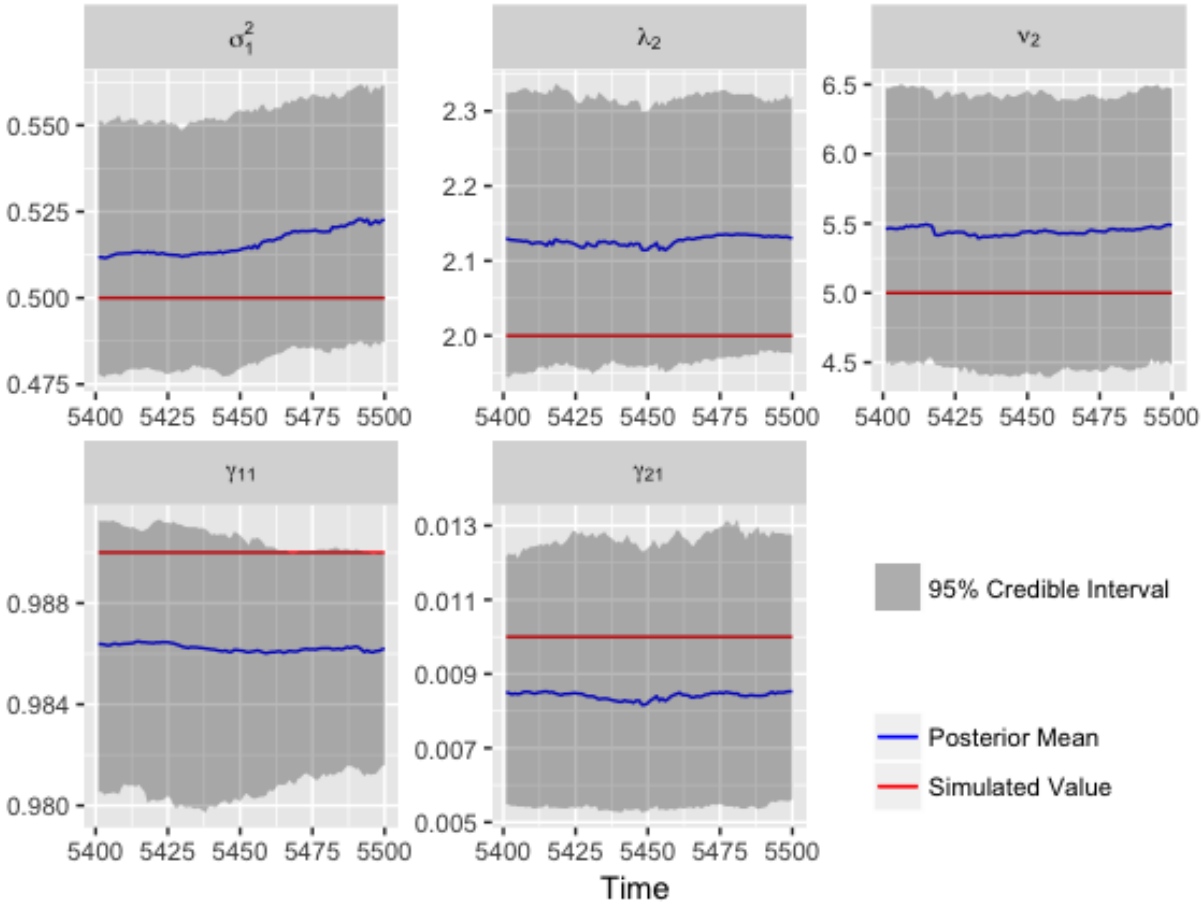

Figure 4.2: State and Observation Parameter Estimates for Simulated Data

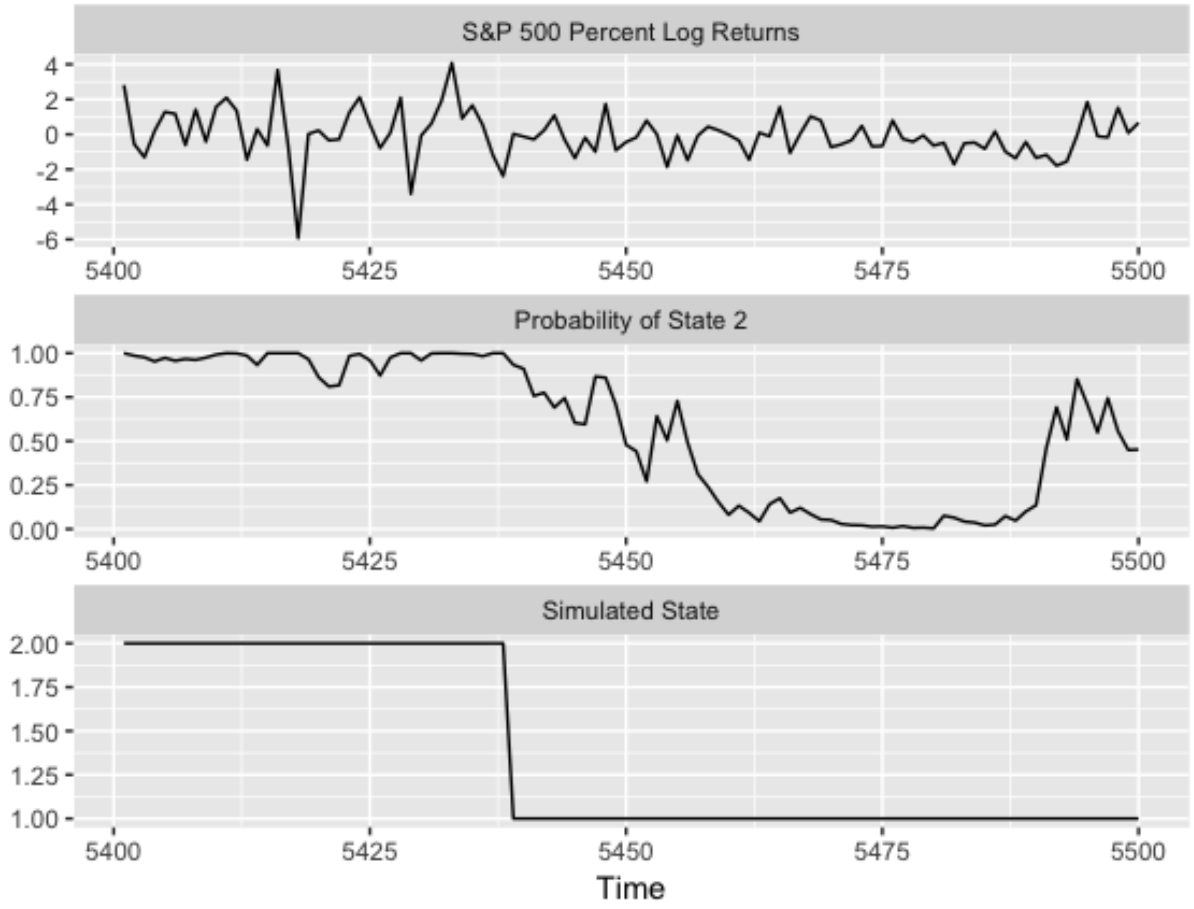

Figure 4.3: State Estimates for Simulated Data

4.2.1 Parameter and State Estimation for Simulated Data

Figures 4.2 show the 95% credible intervals in grey, the posterior mean in blue and the simulated value in red for each of the parameters in the observation and state densities over the 100 observations. Figure 4.3 shows the 100 simulated daily log returns with the probability of state two and the simulated state. State two corresponds to the state with highest variance so periods of high volatility in the log returns should correspond to periods of high probability of state two. There is a period of high volatility from observation 5400 to 5437 followed by a period of volatility from observation 5438 to 5500. Figure 4.3 shows the particle filter is identifying periods of high and low volatility though it takes a few observations for the particle filter to fully identify the change in volatility.

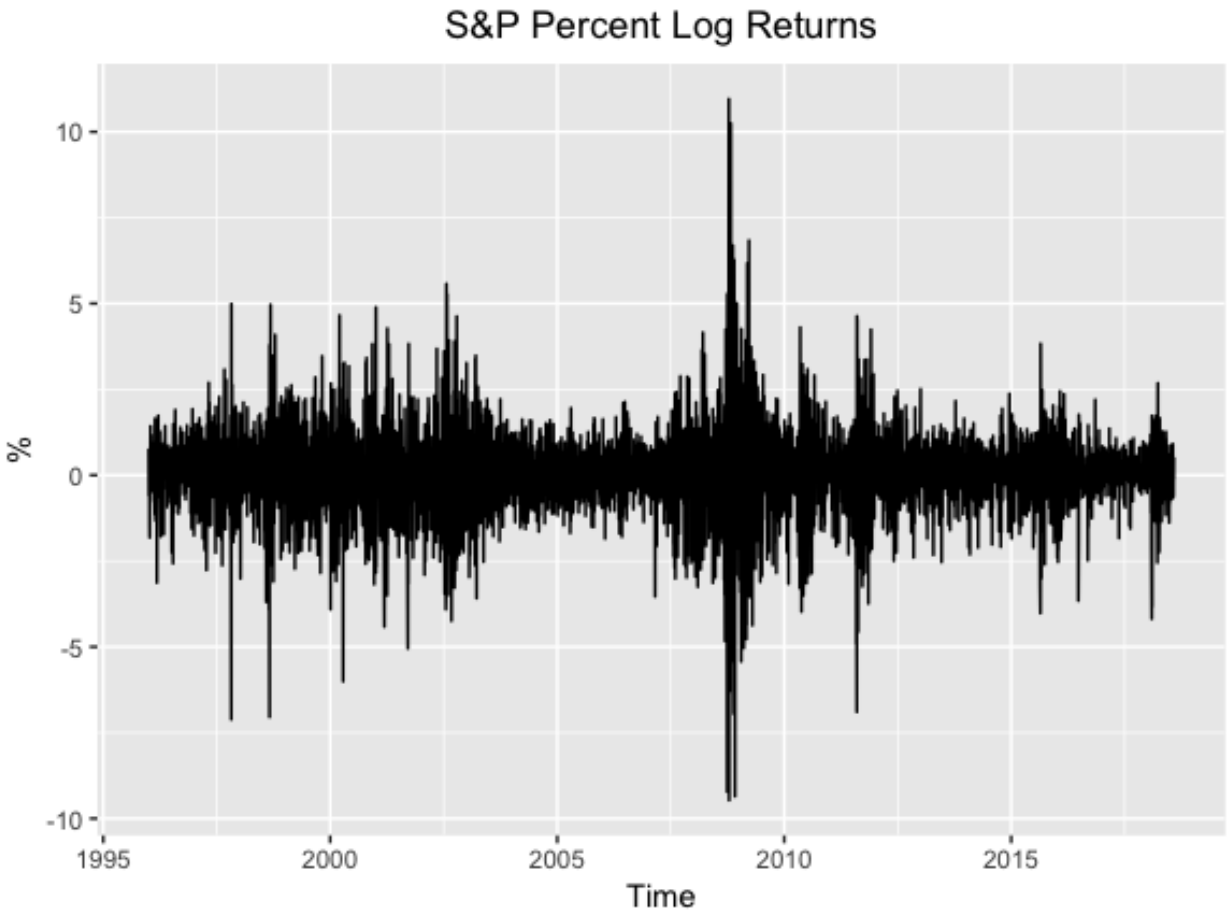

Figure 4.4: Daily Log Returns for S & P 500 Index

4.3 Daily Returns for S&P 500 index

Next the Liu and West [2001] auxiliary particle filter is applied to the daily log returns for the S&P 500 index. Figure 4.4 shows 5688 daily returns from January 2nd, 1996 to August 3rd, 2018. A MCMC sampler using 50,000 samples and a burnin of 25,000 samples was applied to the first 5588 observations to initialize the particle filter. Then the Liu and West [2001] auxiliary particle filter was used to update the posterior distribution over the final 100 observations (March 14th, 2018 to August 3rd, 2018). For the auxiliary particle filter a sample of size 20,000 was used and the particles were resampled at each iteration.

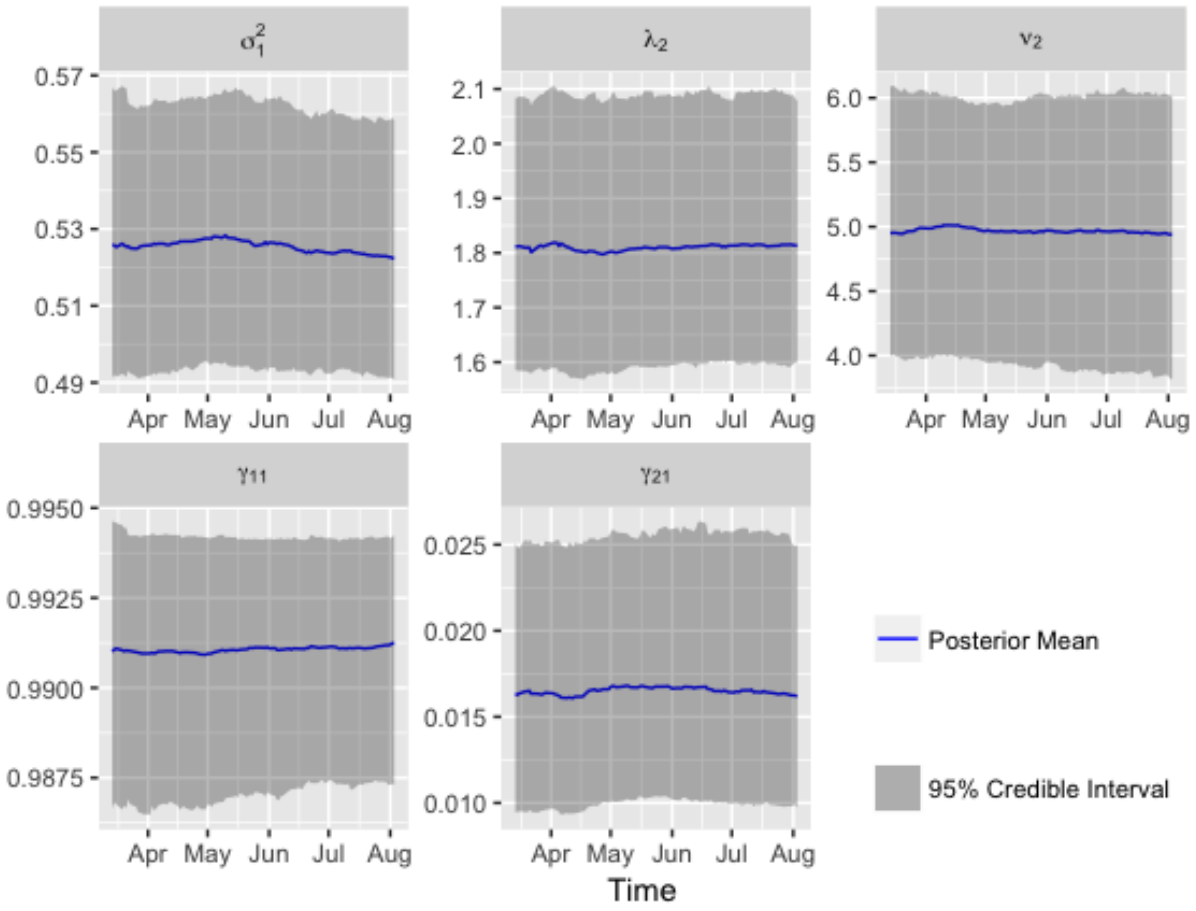

Figure 4.5: Parameter Estimates for S & P 500 Data

4.3.1 Parameter and State Estimation for S & P 500 Data

Figures 4.5 show the 95% credible intervals in grey and the posterior mean in blue for each of the parameters in the observation and state densities over the 100 days. Figure 4.6 shows the daily log returns of the S&P 500 over the 100 days with the probability of the state two at time t below. State two corresponds to the state with highest variance so periods of high volatility in the log returns should correspond to periods of high probability of state two. Around April of 2018 the S&P went though what appears to be a period of relatively high volatility and also during that period our HMM estimates with a high probability that the log returns were in state two, the state of high volatility.

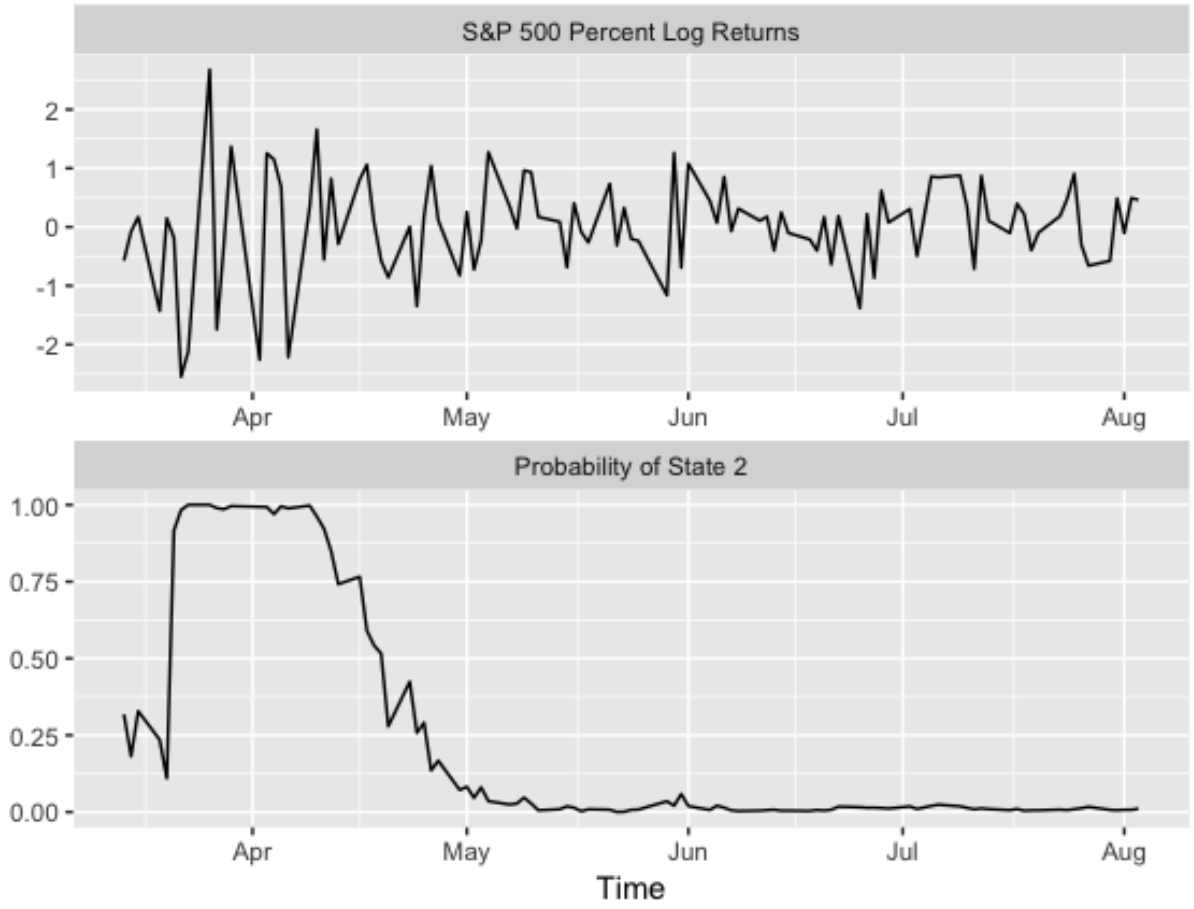

Figure 4.6: State Estimates for S & P 500 Data

4.4 Forecasting

In quantitative finance the interest is often in forecasting future volatility in order to price various types of options. The full forecast distribution can be produced by simulating the process ahead "k" steps. At time t, given an up-to-date approximation to $p(\theta_{0:t}, \omega_{1:t}, \phi|y_{1:t})$, assuming all importance weights are equal, we can produce a one step ahead forecast if for $n = 1, ..., N$

- draw $\theta_{t+1}^{(n)}$ from $p(\theta_{t+1}^{(n)}|\theta_t = \theta_t^{(n)})$
- Given $\theta_{t+1}^{(n)}$
  - Forecast the variance of the log returns, $var(y_{t+1})$
    - if $\theta_{t+1}^{(n)} = 1$ then $var(y_{t+1}) = \sigma_1^{2^{(n)}}$
    - if $\theta_{t+1}^{(n)} = 2$ then $var(y_{t+1}) = \frac{\nu_2^{(n)}}{\nu_2^{(n)}-2}\lambda_2^{-1^{(n)}}$
  - Forecast the log returns, y_{t+1}
    - if $\theta_{t+1}^{(n)} = 1$ then draw y_{t+1} from $N(0, \sigma_1^{2^{(n)}})$
    - if $\theta_{t+1}^{(n)} = 2$ then
      - draw $\omega_{t+1}^{(n)}$ from $gam(\nu_2^{(n)}/2, \nu_2^{(n)}/2)$
      - given $\omega_{t+1}^{(n)}$ draw y_{t+1} from $N(0, (\omega_{t+1}^{(n)}\lambda_2^{(n)})^{-1})$

Figure 4.7 shows the one step ahead forecast distribution for the states and percent log returns.

4.4.1 Evaluation of Forecasts

It is common to evaluate a forecast based on how well the point forecast, such as mean or median, minimize a given loss function. In the case of log returns, given the assumption of efficient markets, any point estimate is always expected to be zero so this type of evaluation is not helpful. Furthermore, the forecast of the log returns is the full forecast distribution. To fully evaluate forecast distribution we can check the coverage of the credible intervals over the 100 day period. In

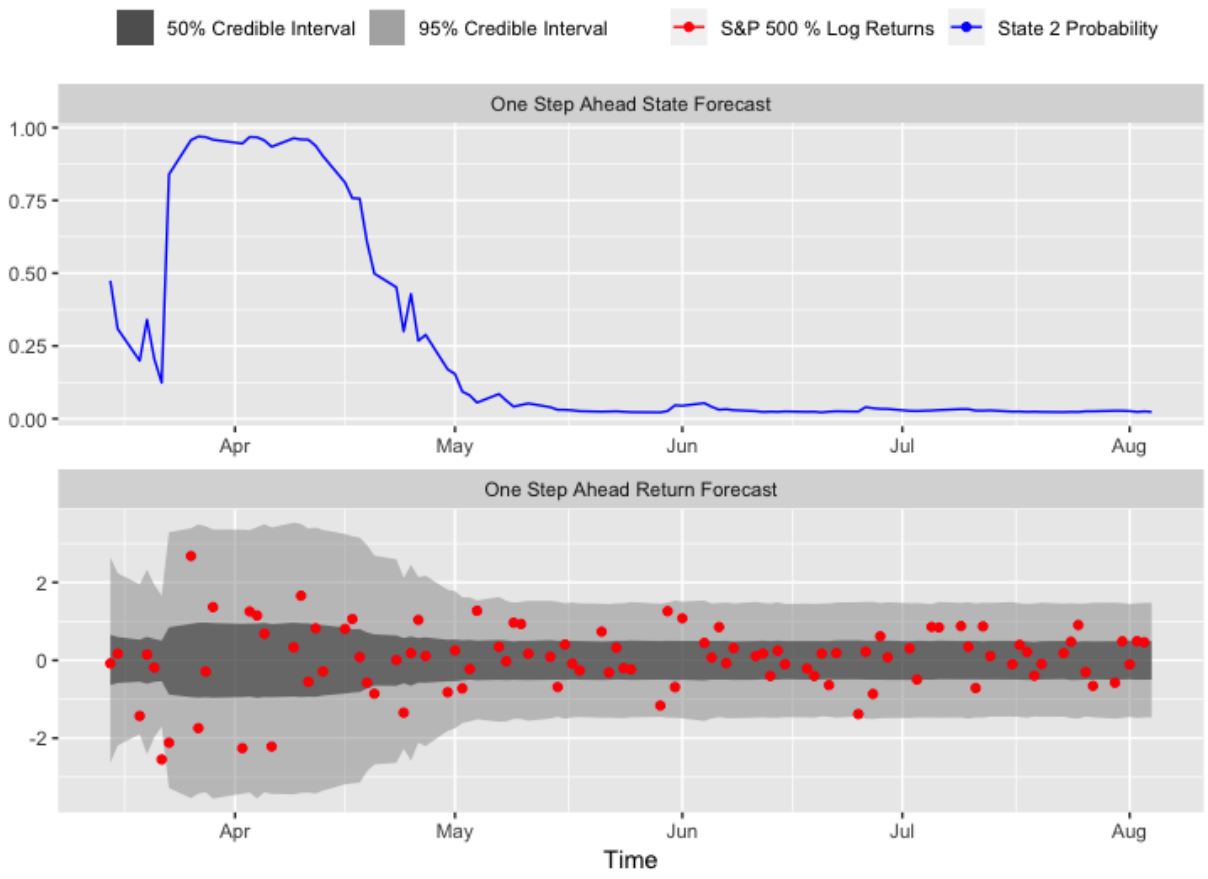

Figure 4.7: One Step Ahead Forecast Distributions

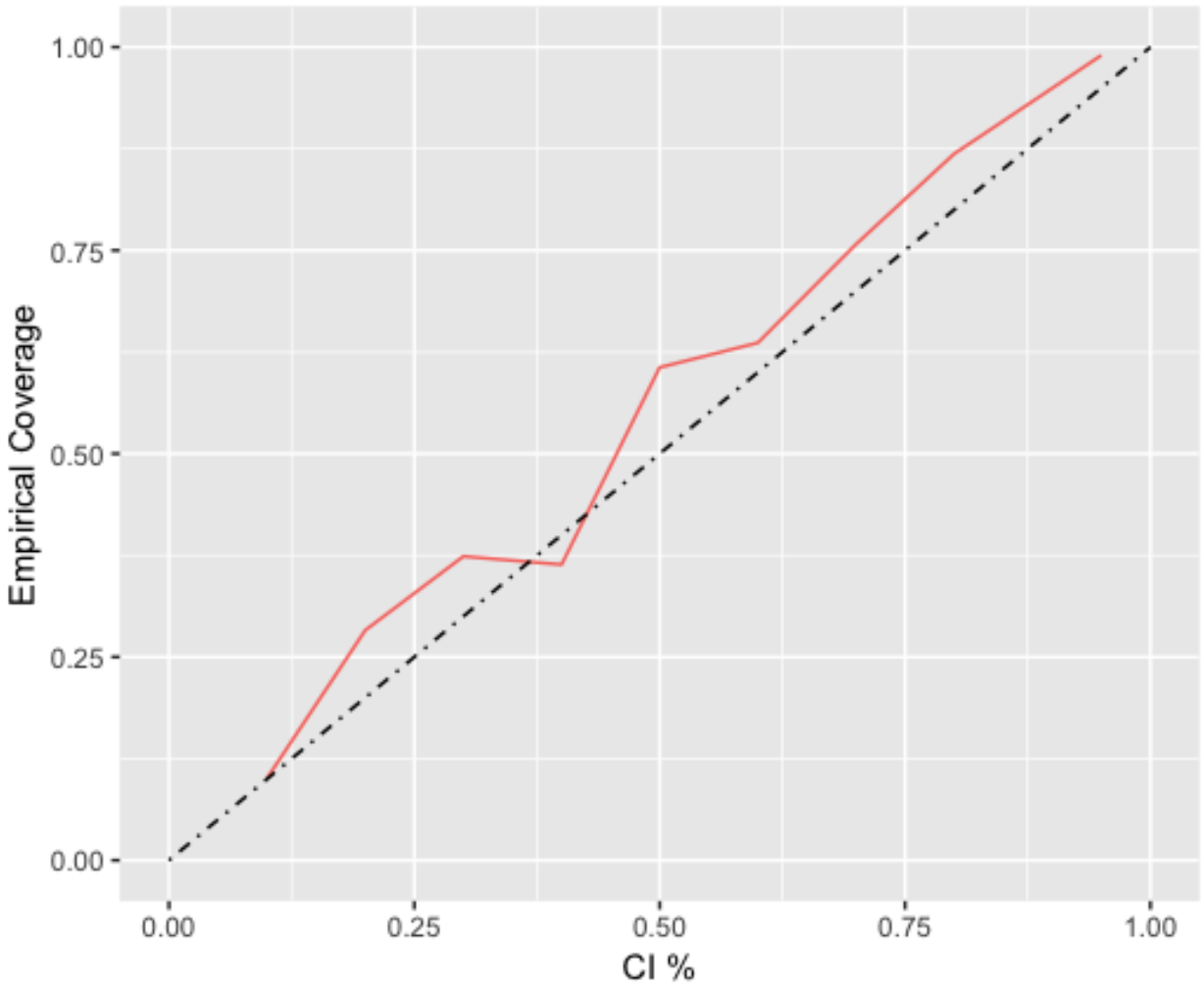

Figure 4.8: Evaluation of S & P 500 Percent Log Return Forecast Distributions

figure 4.8 the y axis is the proportion of observations within each of the estimated credible intervals and the x axis is the theoretical proportion. So the diagonal dotted line is the ideal scenario where $x\%$ of observations fall within the $x\%$ credible interval. The plot shows that the estimated credible intervals are consistently too wide. For example, the 60% credible interval contains approximately 63.6% of the observations in that 100 day period.

5 Conclusion

With today's technology data are being collected sequentially with increasingly short time intervals between observations. Non-sequential Bayesian inference methods, such as MCMC, are notoriously slow and if we want to account for new observations in analysis and forecasting as they arrive then these methods are not practical. State space models provide a frame work that allows for sequential inference for the posterior distributions of interest. In special cases the full closed form posterior distribution can be computed directly. However, for a general nonlinear and/or non-Gaussian state space model this is not possible. In this case, particle filters allow for sequential updating of the approximate posterior distribution as new observations arrive.

In this thesis we have applied the Liu and West [2001] auxiliary particle filter with unknown parameters to a two state hidden Markov model to analyze and forecast the volatility of the log returns of the S & P 500. The variance of the log returns is known to change through time and the hidden Markov model allows for changing variance as the returns transition through the latent states. Furthermore, Bulla [2011] showed that by including at least one t distribution in the HMM can reproduce some of the"stylized facts" of log return at least as well as a HMM with only Gaussian distributions. The model presented in this thesis uses a Gaussian distribution for periods when the log returns are in a state of low volatility and a t distribution for periods when the log returns are in a state of high volatility. The simulated example in 4.2 showed how the auxiliary particle filter tracks the parameter and state estimates. In section 4.3 the HMM was applied to the S & P 500 index. When evaluating forecasts of the log returns standard point estimate metrics are not helpful. Furthermore, we would like to evaluate the full forecast distributions. To this end the coverage of the estimated credible intervals were compared to theoretical credible interval coverage. Using this metric it appears that one step ahead forecast from the HMM produce credible intervals that are slightly too wide. In this thesis the number of discrete latent states was assumed to be two. The main goal being to provide an example of using the Liu and West [2001] auxiliary particle filter on a HMM with unknown state and observation parameters. Further re-

search to determine the proper number of latent states and mixture of normal distribution and t distributions could potentially reduce the coverage of the credible intervals. Another option is to assume the number of latent states is unknown and estimate them.

Bibliography

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

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1996.

J. Bulla. Hidden Markov models with t components. increased persistence and other aspects. *Quantitative Finance*, 11(3):459–475, March 2011.

O. Cappe, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, May 2007.

C. Chatfield. *The Analysis of Time Series*. Chapman and Hall/CRC, 6 edition, 2003.

A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 1960.

J. Durbin and S.J. Koopman *Time Series by State Space Methods*. Oxford University Press, 2 edition, 2001.

J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 10, page 197. Springer, 2001.

G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. Springer, 2009.

M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590, June 1999.

M. West and R. Prado. *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC, 1 edition, 2010.

M. West, A. Pole, and J. Harrison. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman and Hall/CRC, 1 edition, 1994.

W. Zucchini, I. L. MacDonald, and R. Langrock. *Hidden Markov Models for Time Series*. CRC Press, 2 edition, 2016.