Parameter Learning with Particle Filters

Vegan Pather

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

October 19, 2020

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of the Cape Town. It has not been submitted before for any degree or examination in any other University.

October 19, 2020

Abstract

Common applications of asset-pricing models in practice rely on recalibrating model parameters periodically for effective risk management. Yet, these model parameters are often assumed to be constant over time, thereby countering the notion of readjusting these values. A possible solution to this problem is to recalibrate at times where observed market prices cannot realistically match model prices based on parameter values at those times. This dissertation aims to test the effectiveness of a possible algorithm which can be used in optimally identifying such times. An overview is provided of the recently proposed particle filter with accelerated adaptation which has demonstrated rapid time detection for changes in parameter values and has been applied to regime-shifting and stochastic volatility models. Numerical and graphical evidence of parameter and volatility estimation will be provided under regime-shifting parameters for the Heston (1993) stochastic volatility model. The filter demonstrates rapid adaptation in estimating parameter values and accurate estimation of the volatility process. Furthermore, we provide a discussion for possible extensions towards a metric for optimal recalibration times.

Acknowledgements

I would like to thank my supervisor Dr. Ralph Rudd and my co-supervisor Andrew Soane for their invaluable insight and assistance throughout this whole process. I would also like to extend my sincere gratitude to the AIFMRM staff for providing me with the knowledge, facilities and confidence to get to this stage.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When pricing and hedging financial derivatives, a common approach relies on implementing models with time-invariant model parameters. For example, the option pricing formula by Black and Scholes (1973) assumes constant volatility over time. On the other hand, when hedging the risk of such derivatives, model parameters are re-estimated periodically, so as to match liquidly traded market instruments. Such calibration essentially implies time-varying parameters which contradict the initial model assumptions. This is driven by the need to reduce parameter error, as these parameters are generally used for hedging purposes. A possible solution would be to use a technique which can identify model parameter misspecifications as market data is collected sequentially through time. Such a technique which has seen extensive use in latent state estimation is known as filtering.

Filtering refers to the process of using data up to some time t to obtain information of a certain desired entity at that time. Under particle filtering, the entity of interest is the posterior distribution of an unobserved state (Chen, 2003). Applications of filters to sequential parameter learning have lead to various extensions to the traditional particle filter. Such approaches include the use of Markov Chain Monte Carlo techniques as per Berzuini, Best, Gilks and Larizza (1997), sufficient statistic methods by Storvik (2002) and the addition of random perturbation methods by Liu and West (2001). More recently, Gellert and Schlogl (2018) have proposed an extension to previous particle filtering techniques which allows for rapid detection of parameter changes based on financial data.

This project aims to determine the effectiveness of the accelerated particle filter in adapting to changes in parameter values of the Heston (1993) stochastic volatility model. This can serve as a useful tool for recalibration purposes as the accelerated particle filter need only use empirical data to detect parameter changes.

This dissertation begins by outlining the fundamental principles of particle filtering in relation to state estimation in Chapter 2. In particular, an overview will be given of the basic filtering algorithms in the literature, which will then be extended to the particle filter with accelerated adaptation as per Gellert and Schlogl (2018). Chapter 3 provides results for parameter learning for the Liu & West filter along with change detection for the accelerated filter applied to regime shifts of these parameters. Further analysis is given on changes to the rate of adaptation and possible extensions towards optimal time detection. Chapter 4 concludes.

Chapter 2

Particle Filtering

Particle filtering was introduced by Gordon, Salmond and Smith (1993) as a Bayesian filtering method which is a technique that uses prior knowledge and sequential observations in order to make probabilistic inferences. This Bayesian filtering method, was then labelled as "particle filtering" when it was implemented by Del Moral (1996) in numerically solving non-linear equations for interacting particle systems.

2.1 Preliminaries

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the discrete-time set $\mathbb{T} = \{0, 1, ..., n\}$. For every $t \in \mathbb{T}$ we define two sequences of $\mathcal{F}$-measurable random variables, otherwise known as stochastic processes, given by the observation set $y_{0:t} = \{y_0, ..., y_t\}$ and the unobserved states $x_{0:t} = \{x_0, ..., x_t\}$. These processes are also dependent on a d-dimensional parameter set Θ. The state process is first-order Markov which implies that future inferences are conditionally dependent only on the current state value, i.e., $p\left(x_t|x_{0:t-1}\right) = p\left(x_t|x_{t-1}\right)$ and the observation process y_t is conditionally independent given $x_{0:t}$ (Doucet, Godsill and Andrieu, 2000). The objective is to estimate the joint posterior density $p\left(x_t, \Theta_t|y_{0:t}\right)$[1] and this is used to find point estimates of the underlying latent state and parameters at some time t. This density can be found analytically when the state space is linear-Gaussian or finite (Nemeth, Fearnhead, Mihaylova and Vorley, 2012), however for more complex models a closed-form expression is generally unknown. Hence, this posterior density has commonly been approximated using a a set of N discrete random samples, $\left\{x_t^{(1)}, x_t^{(2)}, ..., x_t^{(N)}\right\}$, which are known as particles. The posterior can then be characterized by associating each particle with a weight $\pi_t^{(i)}$. The posterior is then approximated as

$$p\left(x_t, \Theta_t|y_{0:t}\right) \approx \sum_{i=1}^{N} \delta\left(x_t^{(i)} = x_t\right)\pi_t^{(i)},$$

where $\delta\left(\cdot\right)$ is the Dirac delta function.

[1] Note that the subscript t for Θ_t refers to the parameters of the posterior density at time t and does not imply time-varying parameters.

2.2 SIS Filter

The first step in building the filter requires an application of the sequential importance sampling algorithm (SIS) which has been the basis for most filtering algorithms found in literature (Arulampalam, Maskell, Gordon and Clapp, 2002). The SIS filter is generally used for latent state estimation, therefore the following assumes Θ is known, and is henceforth dropped for ease of notation. An initial discrete approximation of the true posterior is represented as follows:

$$\begin{aligned} p\left(x_t|y_{0:t}\right) &= \int \delta\left(x - x_t\right) p\left(x|y_{0:t}\right) dx \\ &= \mathbb{E}[\delta\left(x - x_t\right)] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \delta\left(x^{(i)} - x_t\right), \end{aligned}$$

where $x_t^{(i)}$ is drawn from the true posterior $p\left(x_t|y_{0:t}\right)$ and the first line follows from the definition of the Dirac delta function. Since we cannot always sample from the true posterior, it is common to sample from another distribution known as a proposal distribution which is given by the importance density $q\left(x_t|y_{0:t}\right)$ [2]. In order to correct this sample, we assign a normalized importance weight

$$\pi_t^{(i)} = \frac{p\left(x_t|y_{0:t}\right)}{q\left(x_t|y_{0:t}\right)},$$

to each particle. We now show how a recursive relationship can be established in order to estimate these weights. Consider estimating the posterior $p\left(x_{0:t}|y_{0:t}\right)$ using the importance density which has the factorized form

$$q\left(x_{0:t}|y_{0:t}\right) = q\left(x_t|x_{0:t-1}, y_{0:t}\right) q\left(x_{0:t-1}|y_{0:t-1}\right).$$

The posterior can also be factorized as [3]

$$p\left(x_{0:t}|y_{0:t}\right) = p\left(x_{0:t-1}|y_{0:t-1}\right) \frac{p\left(y_t|x_t\right) p\left(x_t|x_{t-1}\right)}{p\left(y_t|y_{0:t-1}\right)}. \tag{2.1}$$

Using the above equations, we can then find a recursive relationship between importance weights as follows

$$\begin{aligned} \pi_t^{(i)} &= \frac{p\left(x_{0:t}^{(i)}|y_{0:t}\right)}{q\left(x_{0:t}^{(i)}|y_{0:t}\right)} \\ &\propto \frac{p\left(y_t|x_t^{(i)}\right) p\left(x_t^{(i)}|x_{t-1}^{(i)}\right) p\left(x_{0:t-1}^{(i)}|y_{0:t-1}\right)}{q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{0:t}\right) q\left(x_{0:t-1}^{(i)}|y_{0:t-1}\right)} \\ &= \pi_{t-1}^{(i)} \frac{p\left(y_t|x_t^{(i)}\right) p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)}{q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{0:t}\right)}. \end{aligned}$$

[2] This is also known as the importance function.

[3] See Appendix A for derivation.

The proportionality in the second lines follows from the fact that $p(y_t|y_{0:t-1})$ has no dependency on the i^{th} index hence it is a constant. In this case, the entity of interest is $p(x_t|y_{0:t})$. Assuming that the importance density is first-order Markov, then $q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{0:t}\right) = q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_t\right)$. Therefore for non-normalised weights $\hat{\pi}_t$ the recursive relationship reduces to

$$\hat{\pi}_t^{(i)} = \pi_{t-1}^{(i)} \frac{p\left(y_t|x_t^{(i)}\right) p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)}{q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_t\right)}. \tag{2.2}$$

By applying (2.2), we arrive at the SIS algorithm adapted from Chen (2003).

Algorithm 2.1: SIS Filter

Initialise particles $x_0^{(i)}$ for $i = 1, .., N$ by sampling from a chosen prior distribution $p(x_0|\Theta)$. Set $\pi_0 = \frac{1}{N}$. For $t = 1, .., n$ apply the following:

1. *Propagation:* For $i = 1, .., N$ sample $x_t^{(i)} \sim q\left(x_t|x_{t-1}^{(i)}, y_t\right)$.
2. *Measurement:* For $i = 1, .., N$ evaluate

$$\hat{\pi}_t^{(i)} = \pi_{t-1}^{(i)} \frac{p\left(y_t|x_t^{(i)}\right) p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)}{q\left(x_t^{(i)}|x_{t-1}^{(i)}, y_t\right)}.$$

3. *Normalisation:* For $i = 1, .., N$ evaluate

$$\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum_{j=1}^N \hat{\pi}_t^{(j)}}.$$

4. *Prediction:* The posterior is approximated as $\sum_{i=1}^N \delta_{\{x_t^{(i)} = x_t\}} \pi_t^{(i)}$ and the expected value for time step t is given by $\sum_{i=1}^N \pi_t^{(i)} x_t^{(i)}$.

2.3 SIR Filter

The SIS algorithm suffers from a well-known problem known as particle impoverishment or weight degeneracy. Particle impoverishment refers to the decline in the number of non-zero weighted particles as the number of observations increase. Chen (2003) states that impoverishment occurs due to the weight distribution becoming more skewed over time, while Gellert and Schlogl (2018) assert that if a posterior estimation point is below the smallest available floating point number in the computing environment, then this results in a zero weight for that point. This impoverishment implies that only a few particles will contribute to the posterior estimation after many iterations through time. Figure 2.1 below demonstrates this weight degeneracy effect for a simple Gaussian model as per Gellert and Schlogl (2018) with the stochastic differential equation (SDE) given as $dy_t = \sigma dW_t$ where W_t is a standard Brownian Motion. The SIS algorithm is applied using 1000 particles with initial particles $\sigma^{(i)} = \frac{(0.3-0.01)i}{N}$ and an input value of $\sigma = 0.09$. The time interval is partitioned into 10^4 time steps with equal time intervals of $\Delta t = 0.001$. and the importance density is chosen to be the likelihood function given by

$$p\left(y_t|y_{t-1},\sigma\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_t-y_{t-1})^2}{2\sigma_t^2}}.$$

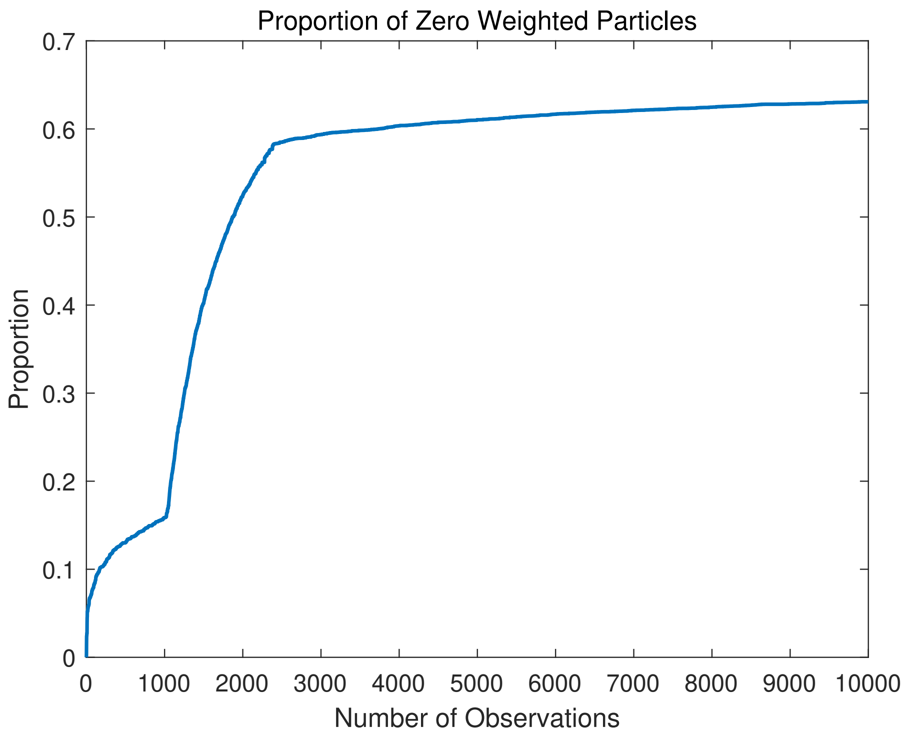

Fig. 2.1: Particle impoverishment effect for increasing number of observations.

As the number of observations increases from 1000 to 3000, there is a high increase in the proportion of zero weight particles. Figure 2.1 shows that approximately 60% of the particles do not contribute to the posterior estimate after 3000 time steps and the proportion also increases at a slower rate as it approaches 10^4 observations.

A frequent approach in mitigating particle impoverishment is to employ sample importance resampling (SIR) which is a method of discarding particles with low weights and duplicating particles with high weights, however this does introduce correlation between particles (Chen, 2003). This is implemented by redistributing new particles at each iteration after finding all weights and then using a resampling technique.

Hol, Schon and Gustafsson (2006) analysed four frequently used resampling techniques (namely stratified sampling, multinomial sampling, systematic sampling and residual sampling) and concluded that systematic resampling has the lowest computation complexity and superior resampling quality.

The first step in systematic resampling is to generate N ordered uniform numbers, $\left\{u^{(1)}, ..., u^{(N)}\right\}$. Then N cumulative sums of the weights, given by $\sum_{s=1}^{i} \pi_t^{(s)}$, can be computed and if the k^{th} ordered number falls between the i^{th} and $(i-1)^{th}$ sum, then that i^{th} weight is taken as the new k^{th} weight. Since multiple ordered numbers are likely to fall in the intervals where the i^{th} weight is relatively high, this leads to replication of high weighted particles.

Zaritskii, Svetnik and Šimelevič (1975) introduced the importance density as the function which minimizes the variance of the importance weights $\pi_t^{(i)}$ as this can limit the weight degeneracy effect. The optimal function which achieves this has been shown to be the transition density $p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)$ which reduces the measurement step in the SIS algorithm to $\hat{\pi}_t^{(i)} = \hat{\pi}_{t-1}^{(i)} p\left(y_t|x_t^{(i)}\right)$. Henceforth, the transition density will be used as the importance density. Another modification to the SIS filter is the introduction of an annealing factor, α, which controls the influence of previous weights as per

$$\hat{\pi}_t^{(i)} = \left(\hat{\pi}_{t-1}^{(i)}\right)^{\alpha} \frac{p\left(y_t|x_t^{(i)}\right) p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)}{q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_t\right)}.$$

Generally, $\alpha = 0$, in which case $\hat{\pi}_t^{(i)} = p\left(y_t|x_t^{(i)}\right)$. Furthermore, in order to determine the correct weights $\hat{\pi}_t^{(i)}$, it is required that

$$q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_t\right) \propto \hat{\pi}_{t-1}^{(i)} p\left(x_t^{(i)}|x_{t-1}^{(i)}\right).$$

Systematic resampling adjusts the posterior approximation such that the distribution is estimated using N equal weights. This ensures that $\hat{\pi}_{t-1}^{(i)} = \frac{1}{N}$ at each propagation step. Therefore the importance density remains proportional to $\hat{\pi}_{t-1}^{(i)} p\left(x_t^{(i)}|x_{t-1}^{(i)}\right)$. Algorithm 2.1 is updated for this resampling step and is given by Algorithm 2.2. In the following algorithms, modifications have a red text color.

Algorithm 2.2: SIR Filter

Initialise particles $x_0^{(i)}$ for $i = 1, .., N$ by sampling from a chosen prior distribution $p(x_0|\Theta)$. Set $\pi_0 = \frac{1}{N}$. For $t = 1, .., n$ apply the following:

1. *Propagation:* For $i = 1, .., N$ sample $x_t^{(i)} \sim q\left(x_t|x_{t-1}^{(i)}, y_t\right)$.
2. *Measurement:* For $i = 1, .., N$ evaluate $\hat{\pi}_t^{(i)} = p\left(y_t|x_t^{(i)}\right)$.
3. *Normalisation:* For $i = 1, .., N$ evaluate
$$\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum_{j=1}^{N} \hat{\pi}_t^{(j)}}.$$
4. *Resampling:* Sample new values of particles from the set $\left\{x_t^{(i)}\right\}_{i=1}^{N}$ according to the importance weights $\left\{\pi_t^{(i)}\right\}_{i=1}^{N}$. The new set of particles have the same weight equal to $\frac{1}{N}$.
5. *Prediction:* The posterior is approximated as $\sum_{i=1}^{N} \delta_{\{x_t^{(i)}=x_t\}} \pi_t^{(i)}$ and the expected value for time step t is given by $\sum_{i=1}^{N} \pi_t^{(i)} x_t^{(i)}$.

The above algorithm can be extended by incorporating the estimation of static parameters. This requires the resampling of these parameters and no additional propagation. Yet, for time-invariant parameters, particle impoverishment remains and this results in estimation of the posterior density around a single point, namely the maximum likelihood estimate (MLE). For a finite number of observations this likelihood estimate may not be the true value of the parameter (Gellert and Schlogl, 2018). In order to address this particle impoverishment, Gordon *et al.* (1993) proposed a method of adding random perturbations to each particle which implies that $\Theta_{t+1} = \Theta_t + \epsilon$ where ϵ is sampled from a d-dimensional multivariate normal distribution with mean 0 and covariance matrix A_{t+1}. This allows the fixed parameters to vary within their own state space which avoids the issue of the parameter posterior being approximated by a single point. Despite this, the choice of A_{t+1} has the potential for the posterior variance to deviate from the theoretical posterior variance which can reduce the accuracy of the estimate.

2.4 Liu & West Filter

In order to address this deviation Liu and West (2001) proposed a method which links the variance of the posterior to the variance of the random perturbation or kernel. They proposed a kernel smoothing method whereby the posterior is approximated as

$$p(\Theta|y_{0:t}) \approx \sum_{i=1}^{N} \pi_t^{(i)} \mathcal{N}\left(\Theta|m_t^{(i)}, h^2 V_t\right),$$

where $\mathcal{N}(\cdot|\mu,\Sigma)$ is a multivariate normal density with mean μ and covariance matrix Σ. Furthermore

$$m_t^{(i)} = a\Theta_t^{(i)} + (1-a)\bar{\Theta}_t,$$

with $\bar{\Theta}_t = \frac{1}{N}\sum_{i=1}^{N}\Theta_t^{(i)}$ and $a = \sqrt{1-h^2}$. The kernel is dependent on a smoothing parameter $h > 0$ chosen to be a decreasing value of N so that the sampled kernel value $\Theta^{(i)}$ converges towards the kernel location $m_t^{(i)}$. Liu and West (2001) found that this choice of kernel location addresses the deviation of variances between the estimated and theoretical posterior. Algorithm 2.2 is then updated as follows:

Algorithm 2.3: Liu & West Filter

Initialise particles $x_0^{(i)}$ and $\Theta_0^{(i)}$ by sampling from a set of prior distributions. Set $\pi_0 = \frac{1}{N}$. For $t = 1, .., n$:

1. *Propagation:* For $i = 1, .., N$ sample $x_t^{(i)} \sim p\left(x_t|x_{t-1}^{(i)}, \Theta_t\right)$.
2. *Measurement:* For $i = 1, .., N$ evaluate $\hat{\pi}_t^{(i)} = p\left(y_t|x_t^{(i)}\right)$.
3. *Normalisation:* For $i = 1, .., N$ evaluate
$$\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum_{j=1}^{N}\hat{\pi}_t^{(j)}}.$$
4. *Resampling:* Sample new values of particles from the set $\left\{x_t^{(i)}, \Theta_t^{(i)}\right\}_{i=1}^{N}$ according to the importance weights $\left\{\pi_t^{(i)}\right\}_{i=1}^{N}$.
5. *Kernel Smoothing:* Apply $\Theta_t^{(i)} \sim \mathcal{N}\left(m_t^{(i)}, h^2 V_t\right)$ where
$$m_t^{(i)} = a\Theta_t^{(i)} + (1-a)\bar{\Theta}_t,$$
$a = \sqrt{1-h^2}$ and $V_t = \frac{1}{N}\sum\left(\Theta_t^{(i)} - \bar{\Theta}_t\right)^2$.
6. *Prediction:* The expected values for (x_t, Θ_t) are given by $\frac{1}{N}\sum_{i=1}^{N} x_t^{(i)}$ and $\frac{1}{N}\sum_{i=1}^{N} m_t^{(i)}$ respectively.

2.5 Particle Filter with Accelerated Adaptation

Gellert and Schlogl (2018) found that the introduction of the kernel variance V_t allows for an expansion in the posterior distribution interval which in turn is capable of detecting changes in model parameters. Therefore, they deduced that the variance of the kernel V_t is related to the speed at which the filter can adapt to model parameter changes. The Liu and West filter, however, demonstrated very slow adaptation. Consequently, a noise term ϕ was added to this variance which resulted in increased adaptation speed at the cost of greater noise around the posterior expected value.They found that there was a selection bias for particles which deviated far from their initial values after kernel smoothing. By allowing ϕ to vary for each particle, the variance term would also be affected by this bias, resulting in high values of $\phi^{(i)}$ during adaptation and low values of $\phi^{(i)}$ for times when adaptation was unnecessary. This maintained the speed of adaptation while reducing noise. This implies initialising the noise term by drawing N samples from a $U(0;c)$ distribution, however this initial distribution range placed a bound on the speed of adaptation. To accelerate adaptation Gellert and Schlogl (2018) introduced a random growth perturbation such that the value of $\phi^{(i)}$ would increase during adaptation. This growth perturbation was added to the algorithm in the form of $\phi_{t+1}^{(i)} = \phi_t^{(i)} e^{\Delta\phi_t^{(i)}}$ where $\Delta\phi_t^{(i)} \sim \mathcal{N}(-\beta, \gamma)$, and β is a dampening parameter which is used to reduce noise after adaptation and control the adaptation speed. The algorithm for the accelerated filter is as follows:

Algorithm 2.4: Accelerated Particle Filter

Initialise particles $\left(x_0^{(i)}, \Theta_0^{(i)}\right)$ by sampling from a set of prior distributions and sample $\phi_0^{(i)} \sim U(0, c)$ for $i = 1, .., N$. Set $\pi_0 = \frac{1}{N}$. For $t = 1, .., n$:

1. *Propagation:* For $i = 1, .., N$ sample $x_t^{(i)} \sim p\left(x_t | x_{t-1}^{(i)}, \Theta_t\right)$.
2. *Measurement:* For $i = 1, .., N$ evaluate $\hat{\pi}_t^{(i)} = p\left(y_t | x_t^{(i)}\right)$.
3. *Normalisation:* For $i = 1, .., N$ evaluate
$$\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum_{i=1}^N \hat{\pi}_t^{(i)}}.$$
4. *Resampling:* Sample new values of particles from the set $\left\{x_t^{(i)}, \Theta_t^{(i)}\right\}_{i=1}^N$ according to the importance weights $\left\{\pi_t^{(i)}\right\}_{i=1}^N$.
5. *Noise Perturbation:* For $i = 1, .., N$, evaluate $\phi_t^{(i)} = \phi_{t-1}^{(i)} e^{\Delta\phi_t^{(i)}}$ where $\Delta\phi_t^{(i)} \sim \mathcal{N}(-\beta, \gamma)$.
6. *Kernel Smoothing:* Apply $\Theta_t^{(i)} \sim \mathcal{N}\left(m_t^{(i)}, h^2 V_t\right)$ where
$$m_t^{(i)} = a\Theta_t^{(i)} + (1-a)\bar{\Theta}_t,$$
$a = \sqrt{1-h^2}$ and $V_t = \frac{1}{N}\sum\left(\Theta_t^{(i)} - \bar{\Theta}_t\right)^2$.
7. *Prediction:* The expected values for (x_t, Θ_t) are given by $\frac{1}{N}\sum_{i=1}^N x_t^{(i)}$ and $\frac{1}{N}\sum_{i=1}^N m_t^{(i)}$ respectively.

As an example, Algorithm 2.4 is applied to the Gaussian model while incorporating a a change in σ at some time t^*. The dynamics of this model are given by
$$dy_t = \sigma_1 \mathbb{I}_{\{t<t^*\}} dW_t + \sigma_2 \mathbb{I}_{\{t\geq t^*\}} dW_t.$$

Figure 2.2 below shows how the accelerated filter compares to the baseline Liu & West Filter, when $\sigma_1 = 0.01, \sigma_2 = 0.02$ and $t^* = 5000$. Furthermore $h = 0.1$ and $c = \frac{0.0002}{N}$. The Liu & West filter shows accurate estimation of the initial parameter value and demonstrates a change detection at t^*. Despite this detection, the rate of adaptation is comparatively slower than the accelerated filter, which rapidly adapts to the change. Conversely, the accelerated filter estimate is much more volatile due to the introduction of the random perturbations.

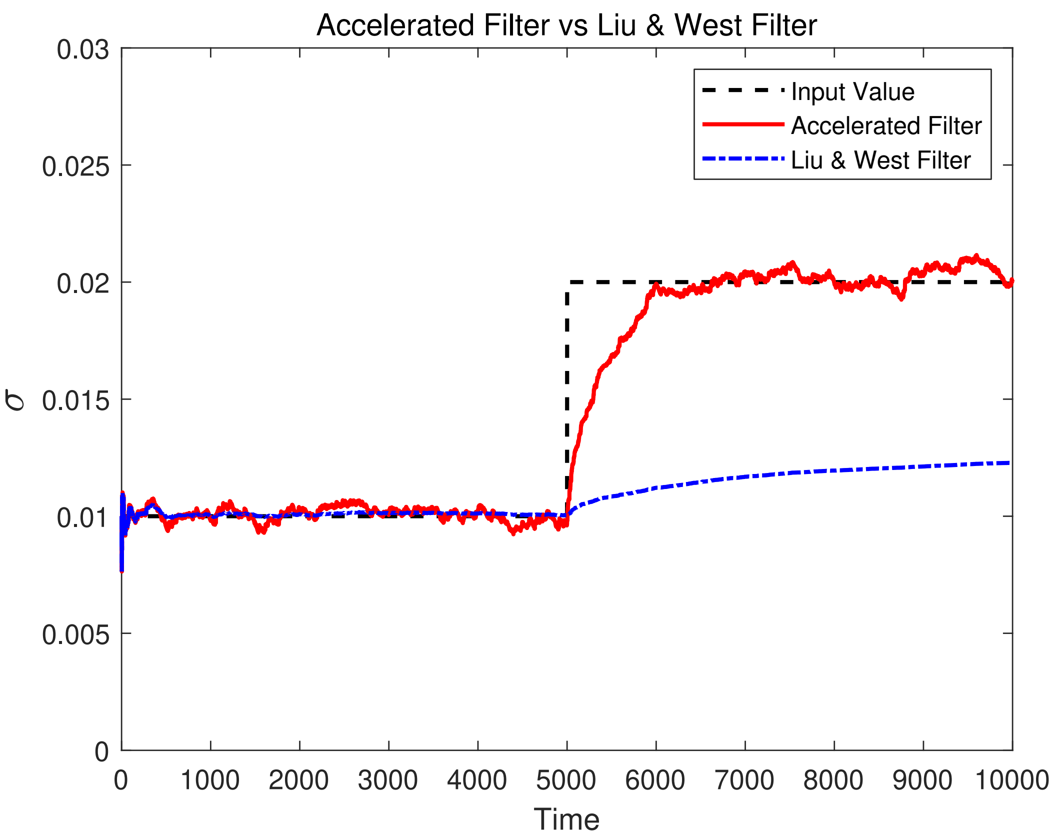

Fig. 2.2: Comparision of particle filter with accelerated adaptation and the Liu & West Filter.

2.6 Parameter Estimation of the Heston Model

The Heston stochastic volatility model is implemented to model the dynamics of a stock price process. Here, we are working under the risk-neutral measure $\mathbb{Q}$ instead of $\mathbb{P}$, since this is the measure used generally for portfolio hedging and pricing. In this case, the dynamics of this model are given by the following stochastic differential equations (SDEs):

$$dS_t = rS_t dt + \sqrt{\nu_t} S_t dW_t^S,$$

$$d\nu_t = \kappa\left(\theta - \nu_t\right) dt + \xi\sqrt{\nu_t} dW_t^\nu,$$

where S_t is the stock price process, ν_t is the volatility process, r is the risk-free rate of return, κ is the mean reversion rate, θ is the mean reversion level and ξ is the volatility of the volatility process. $W_t^S$and W_t^ν are correlated Brownian Motions with correlation ρ. The observation process is the log returns process $Y_t = \log\left(\frac{S_t}{S_{t-1}}\right)$ and the state process is ν_t. Due to the dependency on the number of observations, a partition $\{t_0, ..., t_n\}$ is implemented with a time step of $\Delta t_k = t_k - t_{k-1}$, for $k = 1, ..., n$. This system of SDEs is discretised using the full reflection Euler scheme given by

Lord, Koekkoek and Dijk (2010) as

$$Y_{t+\Delta t} = Y_t + \left(\mu - \frac{1}{2}\nu_t\right)\Delta t + \sqrt{\nu_t} Z_t^S \sqrt{\Delta t}, \tag{2.3}$$

$$\nu_{t+\Delta t} = \left|\nu_t + \kappa\left(\theta - \nu_t\right)\Delta t + \xi\sqrt{\nu_t} Z_t^{\nu} \sqrt{\Delta t}\right|, \tag{2.4}$$

where $Z_t^{\nu} = \rho Z_t^S + \sqrt{1-\rho^2} Z_t^1$ and Z_t^S, Z_t^1 are standard normal independent random variates. Using this scheme ensures that ν_t remains positive. The methodology serves as an extension to the filter applied in Gellert and Schlogl (2018) by implementing both dynamic and static states in the filter where ν_t is the latent state process and $\Theta = \{\kappa, \theta, \xi\}$ are the static states.

Chapter 3

Parameter Estimation and Change Detection

3.1 Parameter Estimation

Implementation of the particle filter was carried out in MATLAB R2019b. By using (2.3) and (2.4), the stock price and volatility processes were simulated using the input parameters in Table 3.1.

Tab. 3.1: Simulation input values

ν_0	S_0	r	κ	θ	ξ	ρ	n	Δt
0.3	100	0.1	3	0.1	0.4	-0.2	100 000	0.001

The Liu & West filter was then applied using a smoothing parameter of $h = 0.1$ and $N = 10^5$ particles. Initialisation of the algorithm requires drawing samples from a prior distribution which is taken as a uniform distribution with parameters a and b. For ν_0 and each parameter, N samples are drawn. Table 3.2 indicates the parameter values for the initial particle draws.

Tab. 3.2: Prior distribution parameters (Liu & West filter)

Value	ν_0	κ	θ	ξ
a	0.2	1	0.05	0.01
b	0.4	4	0.10	0.7

Figure 3.1 below demonstrates the accuracy of the filter in estimating volatility and the three parameters over time. The absolute value for the difference in the estimated and input volatility processes has been shown in the top left figure. For each static parameter, the initial estimates are set at different values to the benchmark input value. Each of the static parameters deviate far from the input value in the early stages due to a lack of information, yet θ and ξ demonstrate steady convergence to the correct values while κ underestimates the correct value.

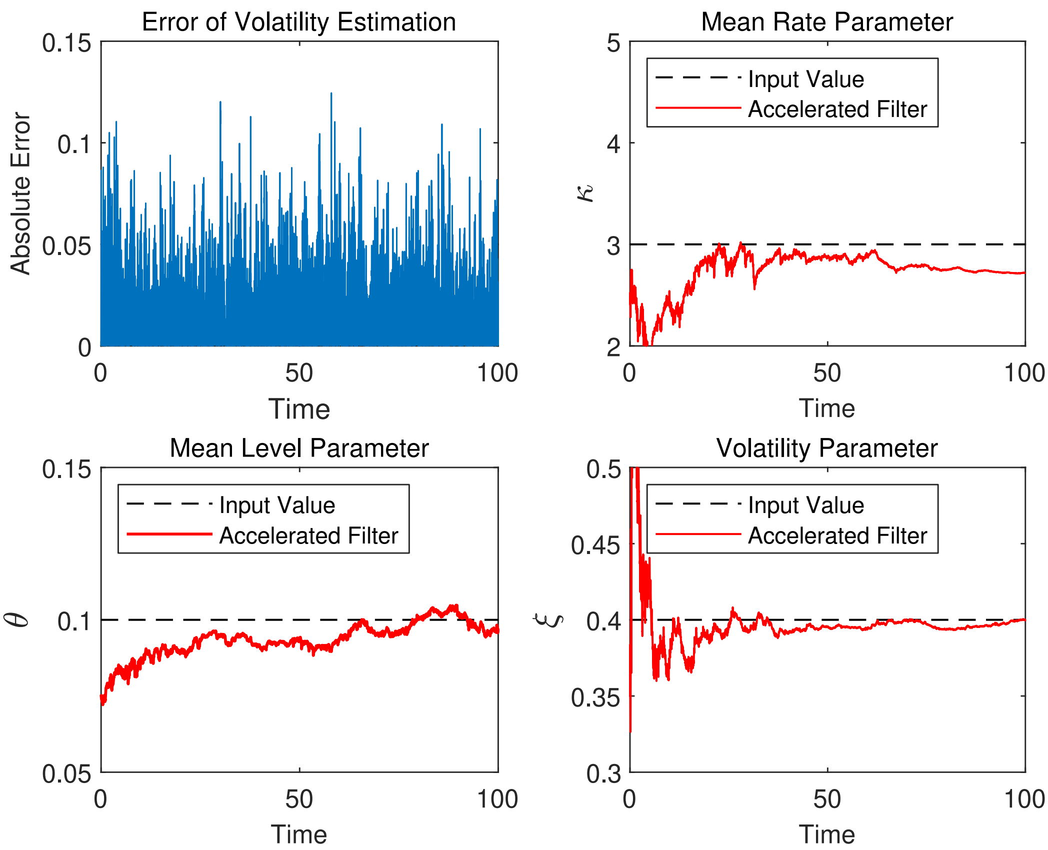

Fig. 3.1: Latent state and parameter estimation using the Liu & West filter.

Mean square error (MSE) was was used as a measure of accuracy for volatility estimation and is given by

$$\text{MSE} = \frac{1}{n+1} \sum_{t=0}^{n} (\nu_t - \hat{\nu}_t)^2,$$

where $\nu_t, \hat{\nu}_t$ is the observed and estimated volatility respectively. The MSE of the above volatility estimation was 5.6793×10^{-4}.

3.2 Change Detection under Regime Shifts

A regime shifting model is a model whereby one of the parameter values shifts at some time t^*, which is set to 5×10^4 in this dissertation. The accelerated filter is applied under regime shifts for each of the parameter values in order to test the filter's capability of detecting parameter changes. Table 3.3 provides the parameter values of the filter. It should be noted that the scale of c is adjusted to match the scale of the shfted parameter i.e., $c = \frac{2}{N}$ is used for the estimation of κ instead of $\frac{0.02}{N}$.

Tab. 3.3: Filter parameter values

N	h	c	γ	β
100 000	0.1	$\frac{0.02}{N}$	$\frac{0.01}{N}$	$\frac{0.0001}{N}$

Furthermore, the parameters of the prior distributions were adjusted such that the $t = 0$ estimates were centred around the true values of each parameter. The new values are given in Table 3.4.

Tab. 3.4: Prior distribution parameters (accelerated filter)

Value	ν_0	κ	θ	ξ
a	0.2	2	0.05	0.3
b	0.4	4	0.15	0.5

Figures 3.2 – 3.4 represent the results from the accelerated filter.

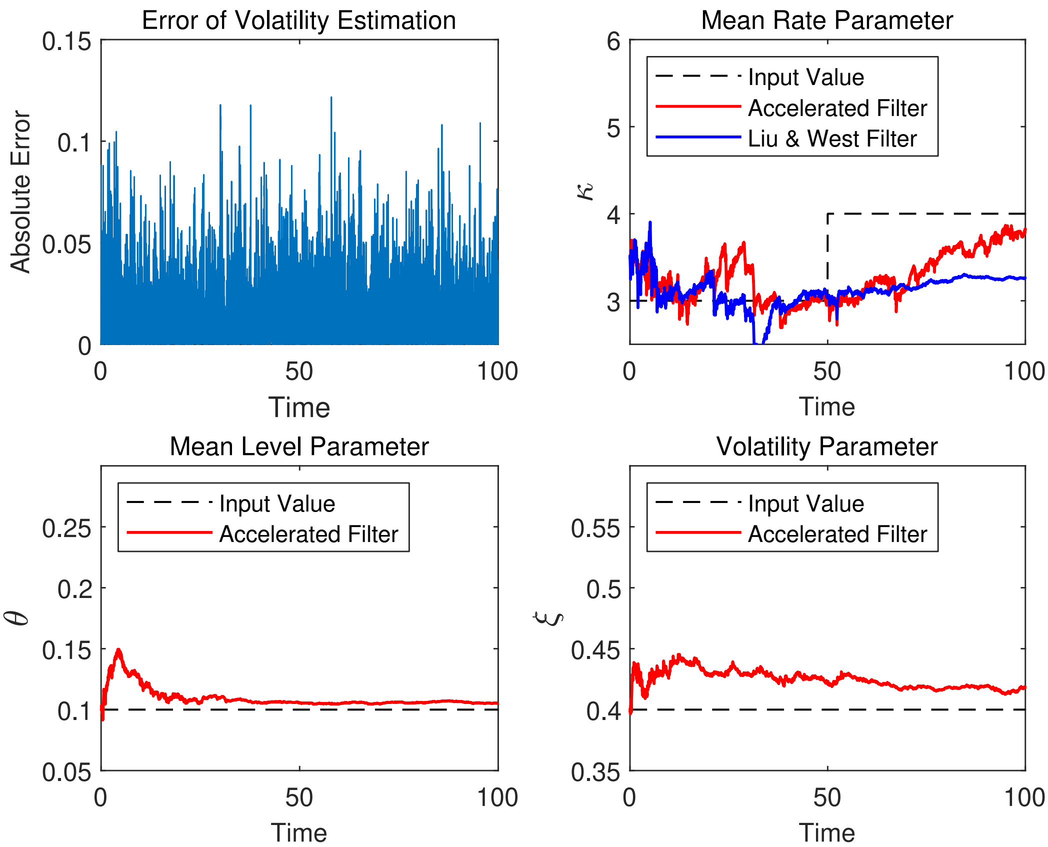

Fig. 3.2: Latent state and parameter estimation under regime shifted κ.

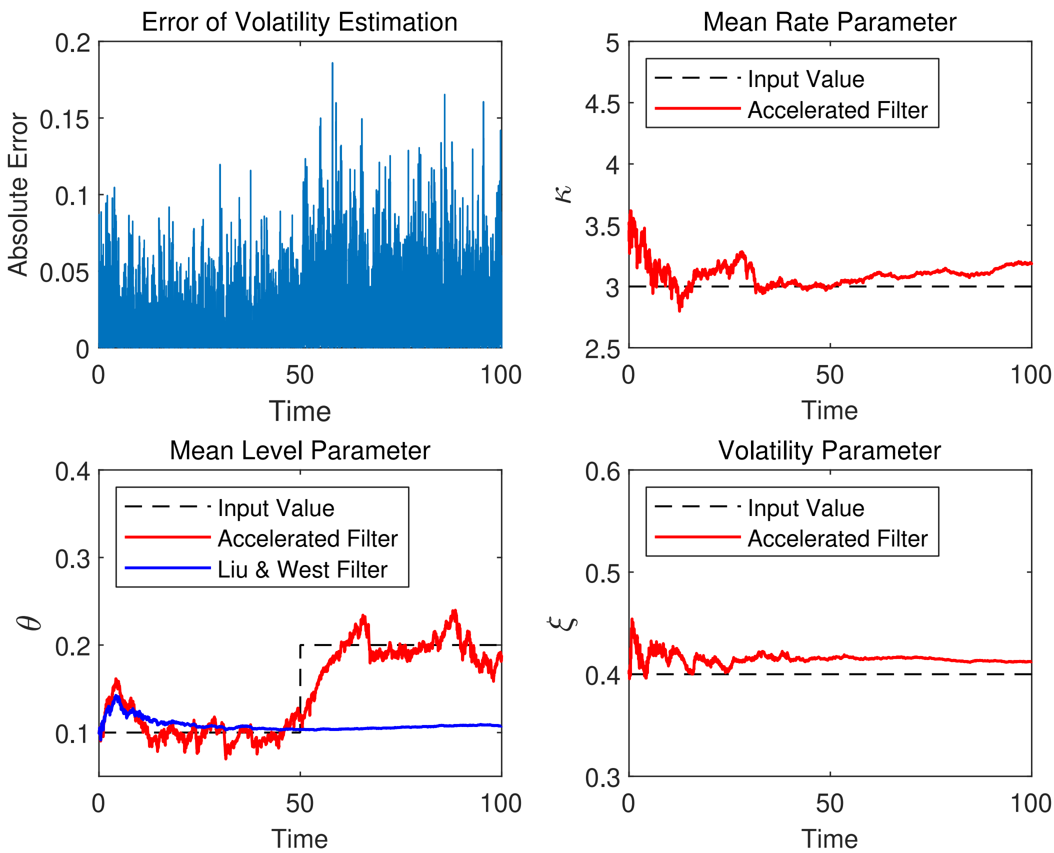

Fig. 3.3: Latent state and parameter estimation under regime shifted θ.

In each of the scenarios, the undisturbed parameters converge closely to their input values. In terms of speed of adaptation ξ has rapid adaptation to the shift while θ demonstrated slower speed. Nonetheless, κ has unstable estimates and lower accuracy, however Figure 3.2 shows that filter detects a paramter changes and shows convergence towards the new value. The accelerated filter also outperformed the Liu & West filter in terms of speed when adapting to the change in each scenario. Furthermore the error plots in each scenario indicate that the errors frequently fall below 0.1 in magnitude. Table 3.5 provides values for the MSE under each shift along with the comparative MSE without the accelerated filter. In each scenario the MSE for the accelerated filter is lower, since the volatility estimation incorporates estimates for the regime shifted parameter.

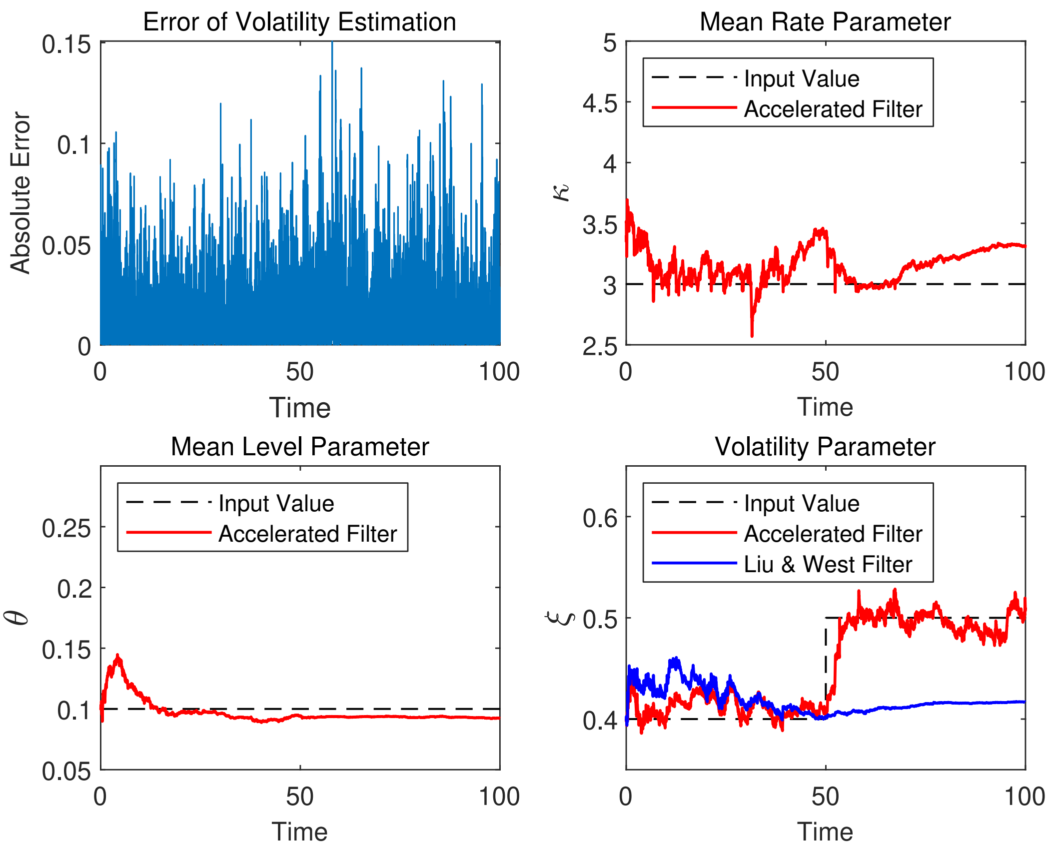

Fig. 3.4: Latent state and parameter estimation under regime shifted ξ.

Tab. 3.5: Mean square error $\left(\times 10^{-4}\right)$

Shifted Parameter	κ	θ	ξ
Liu & West	5.6037	10.8092	6.5944
Accelerated Filter	5.5675	10.3115	6.5607

The computation times for the full algorithm and the resampling step were recorded in Table 3.6. In each scenario, the resampling step had the highest proportion of computation time as compared to the other steps in the algorithm.

Tab. 3.6: Computational time (seconds)

Shifted parameter	κ	θ	ξ
Accelerated filter	2057	2071	2062
Resampling time	857	851	863

3.3 Changing the Rate of Adaptation

One of the influential parameters in the filter is c which controls the magnitude of the initial random perturbations. Gellert and Schlogl (2018) showed that the speed of adaptation is bound by the size of ϕ_t, hence an increase in c should increase the rate of adaptation. Figure 3.5 shows this effect for the estimation of θ, however there is also a significant increase in noise. Sensitivity plots of the other parameters are provided in Appendix B.

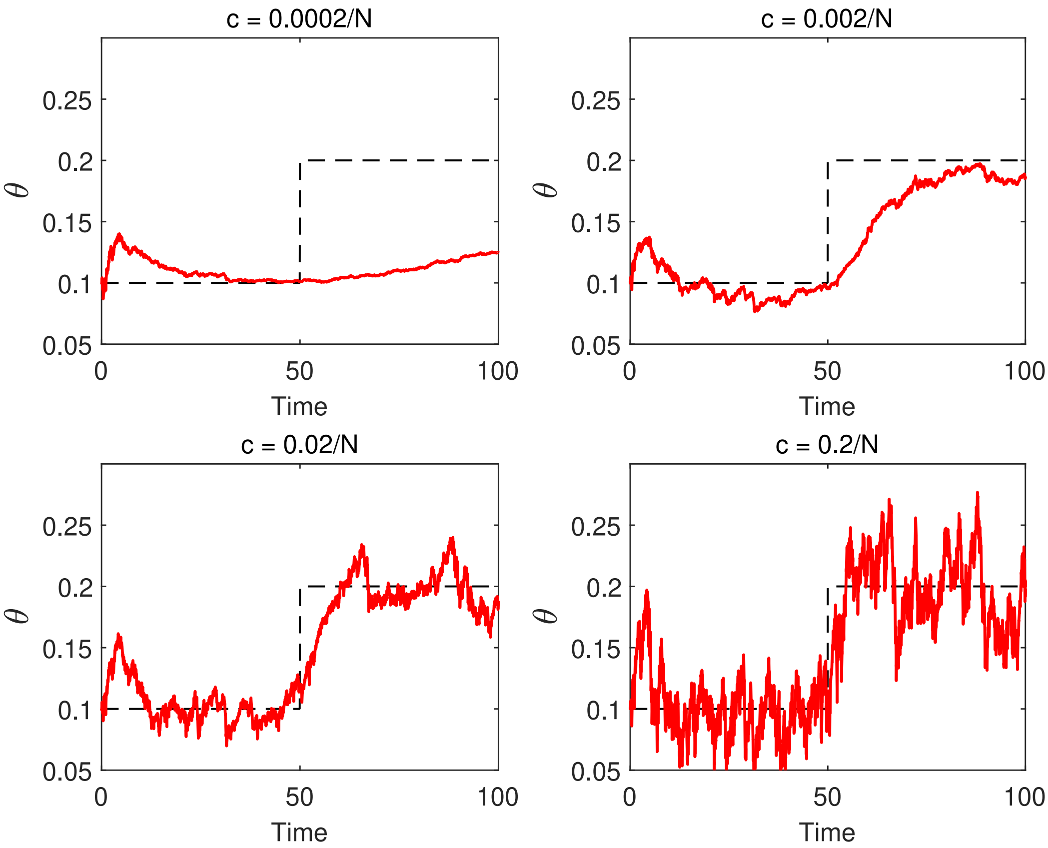

Fig. 3.5: Comparison of changes in the scale of random perturbations for θ.

3.4 Optimal Time Detection

The focus of this dissertation has been the sequential Monte Carlo estimation of a posterior density using simulated data. Considering the apparent discrepancy between maintaining model assumptions and hedging effectively, one may ask how this filter can be used as an effective risk management tool. Glasserman and Xu (2014) proposed an approach of robust risk measurement using relative entropy. Given the probability densities $p(x)$ and $\tilde{p}(x)$, relative entropy is defined as

$$\mathcal{R}\left(p, \tilde{p}\right) = \mathbb{E}[m \log(m)] = \int \frac{\tilde{p}(x)}{p(x)} \log\left(\frac{\tilde{p}(x)}{p(x)}\right) p(x)dx,$$

where $m = \frac{\tilde{p}(x)}{p(x)}$ is the likelihood ratio. Relative entropy measures the deviation between p and $\tilde{p}$. It has also been defined as the additional information required for which $\tilde{p}$ is preferred to p. By using a specified entropy limit based on other market distributions, optimal recalibration times may be determined if $\mathcal{R}\left(p, \tilde{p}\right)$ exceeds this limit. Therefore the posterior's fast adaptation could serve as an informative density to be used with relative entropy in further research.

Chapter 4

Conclusion

The aim of this dissertation was to test the effectiveness of a filtering technique which can adapt to changes in parameter values when the observed price data does not match the model's output.

A detailed overview of particle filtering was provided where we found that the SIS filter suffers from weight degeneracy. To address this problem, a systematic resampling procedure was introduced which worked well for state estimation, however weight degeneracy persisted for static parameter estimation. Liu and West (2001) addressed this issue by sampling parameter particle values from a normal kernel with a mean and variance linked to the estimated posterior. Gellert and Schlogl (2018) found a link between this variance and the speed of adaptation to changes in parameter values, which lead to the accelerated filter.

This accelerated filter was then applied to the Heston model with the aim of estimating both stochastic volatility and static parameters through time. The accelerated particle filter was able to detect changes under regime shifts for each parameter, while also estimating accurate values for the other parameters. In each case, the filter also provided accurate estimation of the volatility process. By changing the rate of adaptation, it was shown there is also a significant trade-off between the speed of adaptation and the volatility of the estimates. While the filter may not provide stable estimates through time, it acts as a clear heuristic for change detection. This allows for further extensions, such as relative entropy, which can be used as a measure for recalibrating model parameters at optimal times.

Bibliography

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

Berzuini, C., Best, N. G., Gilks, W. R. and Larizza, C. (1997). Dynamic conditional independence models and markov chain monte carlo methods, *Journal of the American Statistical Association* **92**(440): 1403–1412.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of political economy* **81**(3): 637–654.

Chen, Z. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond, *Statistics* **182**(1): 1–69.

Del Moral, P. (1996). Non-linear filtering: Interacting particle resolution, *Markov Processes and Related Fields* **2**(4): 555–581.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering, *Statistics and Computing* **10**(3): 197–208.

Gellert, K. and Schlogl, E. (2018). Parameter learning and change detection using a particle filter with accelerated adaptation.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk, *Quantitative Finance* **14**(1): 29–58.

Gordon, N. J., Salmond, D. J. and Smith, A. F. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation, *IEE Proceedings F*, Vol. 140, IET, pp. 107–113.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* **6**(2): 327–343.

Hol, J. D., Schon, T. B. and Gustafsson, F. (2006). On resampling algorithms for particle filters, *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, IEEE, pp. 79–82.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering, *Sequential Monte Carlo Methods in Practice*, Springer, pp. 197–223.

Lord, R., Koekkoek, R. and Dijk, D. V. (2010). A comparison of biased simulation schemes for stochastic volatility models, *Quantitative Finance* **10**(2): 177–194.

Nemeth, C., Fearnhead, P., Mihaylova, L. and Vorley, D. (2012). Particle learning methods for state and parameter estimation.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters, *IEEE Transactions on Signal Processing* **50**(2): 281–289.

Zaritskii, V., Svetnik, V. and Šimelevič, L. (1975). Monte carlo technique in problems of optimal information processing, *Automation and Remote Control* (12): 95–103.

Appendix A

Posterior Estimation

In this appendix, we derive the simplified expression for (2.1), using Bayes' rule. As per Chen (2003), the posterior can be expressed as follows:

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

Now since x_t is first-order Markov and y_t is conditionally independent given $x_{0:t}$, we have that

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

Appendix B

Sensitivity Plots

In this appendix, we provide the remaining sensitivity plots for κ, ξ due to changes in the rate of adaptation, given by c.

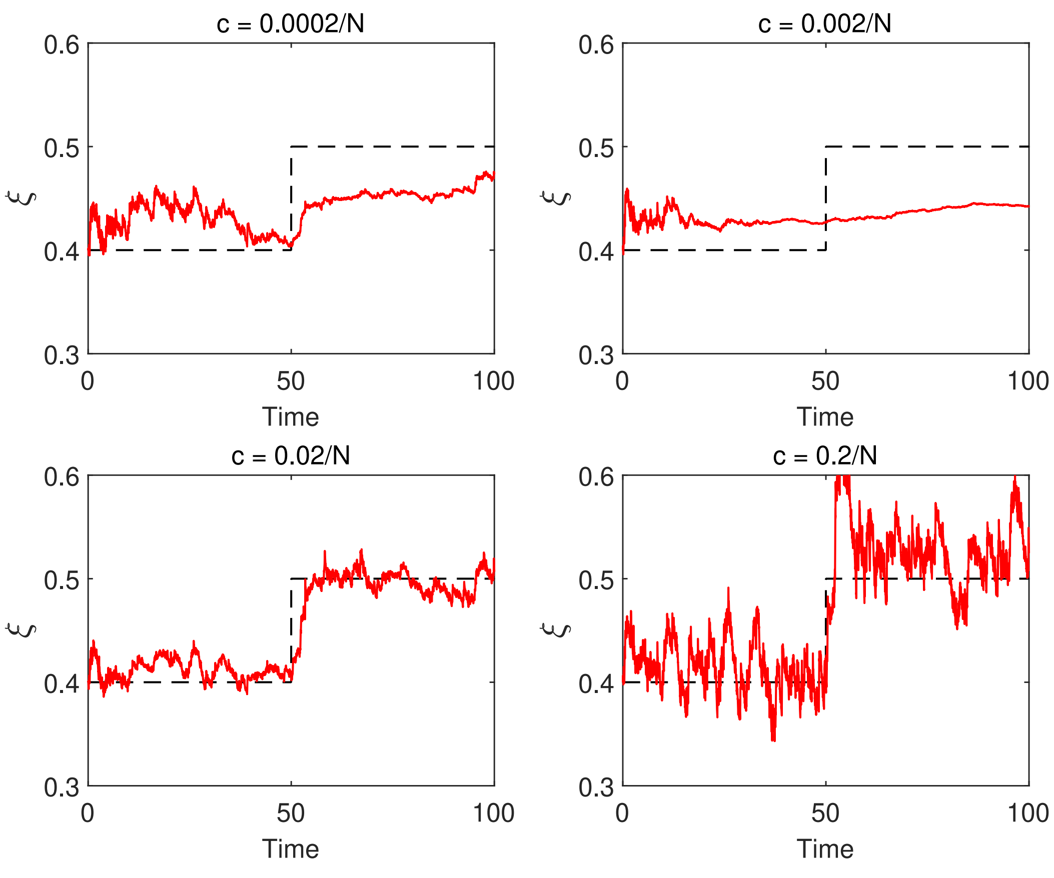

Fig. B.1: Comparison of changes in the scale of random perturbations for ξ.

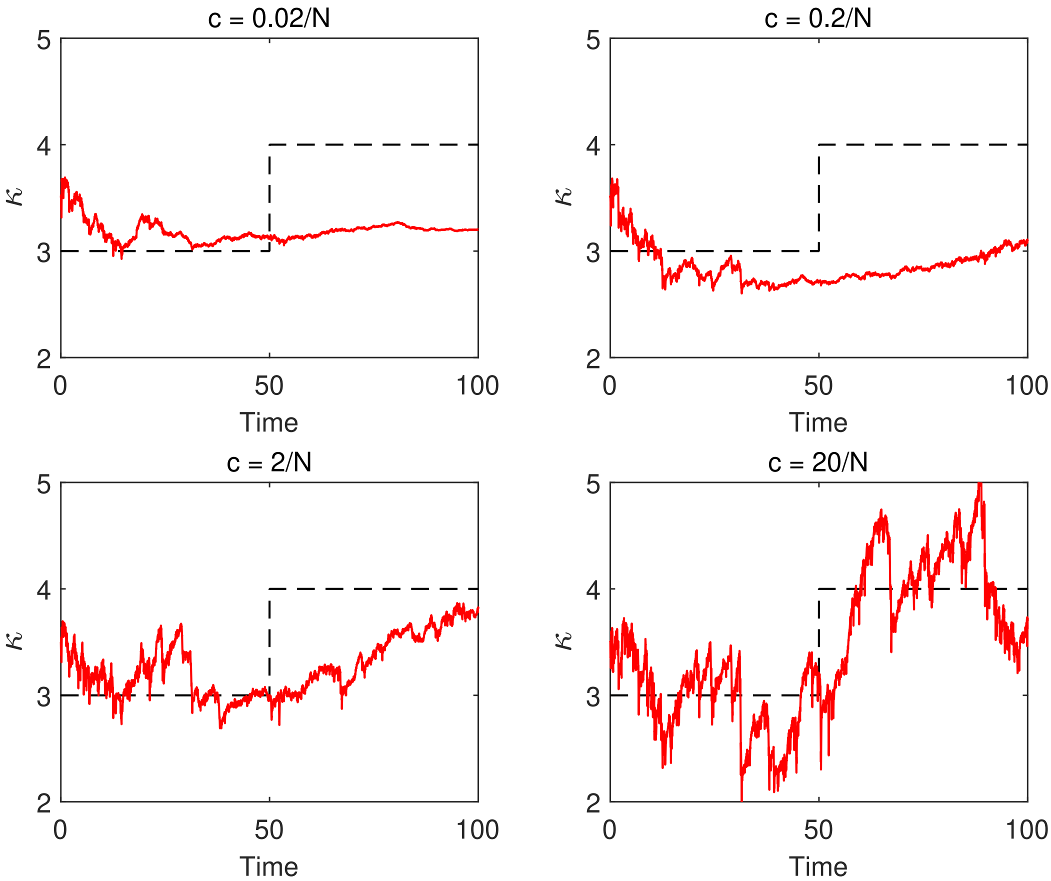

Fig. B.2: Comparison of changes in the scale of random perturbations for κ.

Figure B.1 shows a similar trade-off between volatility and speed as with θ. The estimates for κ are more volatile and deviates from the input values, however the filter does demonstrate change detection for higher values of c.