Parameter Estimation for Infinite Lévy Processes via Sequential Monte Carlo

Liew Kuang Chen Joel A0004624U

Department of Mathematics

Supervisor: Professor Zhou Chao

An honours thesis submitted in partial fulfillment for the degree of

Bachelor of Science (Honours) in Quantitative Finance

Acknowledgements

First, I would like to express my deepest appreciation to Professor Zhou Chao for his patient teaching and the time and effort that he took from his busy schedule to meet me for bi-weekly meetings. Thank you for the positive feedback and the encouragement throughout this period.

Second, I would also like to thank Associate Professor Ajay Jasra for taking time off to meet me even though I am not his FYP student. The great passion and knowledge that he possesses in the field of Sequential Monte Carlo had inspired me to strive for excellence in this FYP. Thank you for your patience and guidance.

Third, I want to give a special mention to my coursemates: Samuel, Nige, Shih Chen and Sylvanus. I enjoy the times spent discussing everything under the Sun. You guys have been my pillar of strength for the past 4 years.

Fourth, I would like to show appreciation to my family and my girlfriend, Li Ling. They have been very supportive and understanding throughout my University Life, without which I will not be who I am today.

Last, I would like to thank God for the many inspirations that He had given me during this ardous project journey. He has been my emotional fortitude and source of strength when things were not working out well in life. Thank you for bringing people into my life who had shaped me to become who I am today.

Abstract

Unforseen events can jolt Financial Markets and cause prices to 'jump'. As a result, our pricing models should be able to capture such characteristics or we risk mispricing financial instruments. For this paper, we will first introduce the concept of Lévy process. Next, the paper will familiarise the reader with the concept of Sequential Monte Carlo. Lastly, we will use this new simulation method to retrieve the posterior estimates and attempt to create a profitable trading strategy with the estimates.

Contents

5 Conclusion   31

6 Appendix   31

References   32

1 Introduction

1.1 Background

We live in a very uncertain World - Financial Crises, acts of Terrorism, uncertain Central Bank Policies, Geopolitical risks etc, all of which could result in spikes of volatility, and therefore jumps in Markets. As a result, the traditional idea of modelling log-prices using Brownian Motion does not fully capture the characteristics of prices in the Financial Markets. Hence, one can perhaps model these prices using the Lévy Process. In fact, it has been established by that the Lévy Process is an attractive alternative to Affine Jump Diffusion Models (Brownian motion and Compound Poisson process) [1]. Within the family of the Lévy Process, there are 2 main branches - Finite Activity Models and Infinite Activity Models. The work of Li, Wells and Yu (2006) shows that the Infinite Activity Models can approximate the AJDs and is more robust in capturing the dynamism of the daily returns of the S&P500 [2].

1.2 Project Scope

This paper firstly gives an introduction of the Lévy Process and Infinite Activity Models. Next, it acquaints one with the Sequential Monte Carlo Sampler by methodically building upon the concepts of Perfect Monte Carlo and Importance Sampling to derive the Sequential Importance Sampling and then the Sequential Monte Carlo Sampler. Next, this paper seeks to replicate the work of Jasra, Stephens, Doucet and Tsagaris (2011) - using the SMC Algorithm to retrieve the posterior parameters of interest. Lastly, this paper proposes a possible trading strategy to generate returns. In short my contributions for this project are:

- Combined and simplified the work of Li (2006) and Jasra (2011)

- simplified the posterior distribution from Li (2006) by introducing a result from Jasra (2011)

- the resulting posterior distribution has 1 lesser dimension to simulate from

- introduced the SMC Algorithm from Jasra (2011), instead of the MCMC from Li (2006), which has greatly cut down the time to get the estimate of the parameters.

- tested the SMC Algorithm to check if the posterior estimates lie within the 95% credible interval

- Proposed Trading Strategy

  - introduced a novel approach of using posterior parameters to generate returns from the instrument

  - provided a clear and logical framework for refining trading strategies

  - explored the use of simple moving average and exponential moving average with the posterior parameters to create trading systems

  - experimented on the use of multiple trading strategies to create a portfolio of trading strategies that seeks to outperform the traditional buy and hold

2   Preliminaries

2.1   The Lévy Process

In this paper, we will be using the Variance Gamma model from the family of Infinite Activity Lévy Processes. For the uninitiated, we will first go through the basics.

Definition 2.1. Lévy Process

A Lévy Process, X_t, is a cadlag stochastic process on $(\Omega, \mathcal{F}_t, P)$,with $X_0 = 0$ and has the following properties:

1. Independent increments: $X_{t+j+1} - X_{t+j} \perp X_{t+k+1} - X_{t+k}, \forall k, j \geq 0, k \neq j$

2. Stationary increments: The Law of $X_{t+h} - X_t$ does not change for any t.

3. Stochastic continuity: For any $\epsilon > 0, \lim_{h \to 0} P(|X_{t+h} - X_t| > \epsilon) = 0$

The Lévy Process also satisfies the Infinite Divisibility property.

Definition 2.2. Infinite Divisibility

For $n \geq 2$, a probability distribution F is infinitely divisibile if there exists n i.i.d. random variables $Y_1, Y_2, ...$ such that $\sum_{i=1}^{n} Y_i$ has distribution F.

One common distribution satisfying the infinite divisibility condition is the Normal Distribution. If $X \sim N(\mu, \sigma^2)$, then $X = \sum_{i=1}^{n} Y_i$,where $Y_i \sim N(\frac{\mu}{n}, \frac{\sigma^2}{n})$.

Due to fact that Lévy Processes are semimartingales and that their increments are independent, the process is Markovian and therefore in line with the Efficient Market Hypothesis. Hence, such processes are suitable to model asset prices [2].

2.1.1   Infinite Activity Processes - Variance Gamma

The probability densities of most Lévy processes are not known in closed forms. However, their characteristics function $\phi_{X_t}(u)$ is as follows:

$$\phi_{X_t}(u) = E[\exp^{iuX_t}] = \exp^{-t\psi_x(u)}, t \geq 0$$

where $\psi_x(u)$ is the characteristic exponent of X and satisfies the following Lévy-Khintchine formula [3]

$$\psi_x \equiv -i\mu u + \frac{\sigma^2 u^2}{2} + \int_{R_0}(1 - \exp^{iux} + iux1_{|x|<1})\pi(dx)$$

where $R_0 = R \setminus \{0\}$ and π is a R_0 Radon measure and

$$\int_{R_0} |x|^2\pi(dx) < \infty \qquad \int_{R_0} \pi(dx) < \infty$$

Unlike the Finite Activity models (such as the Merton Jump Model) under the family of Lévy Processes, the infinite-activity models allow infinite jumps within a finite time interval. Hence,

$$\int_{R_0} \pi(dx) \not< \infty$$

Within the category of Infinite Activity Lévy Processes, the jump process can have either finite or infinite variation. In the case of the finite/infinite variation, the sum of the absolute distance is finite/infinite over any finite time interval. For this paper, we will only be focusing on the Variance Gamma Process - a member from the Infinite Activity with Finite Variation. This process, Z_t, is created using a Brownian Motion, B_t, with drift γ and variance σ_j^2, subordinated by an independent gamma process, G_t, with parameter α.

$$dZ_t = \gamma dG_t + \sigma_j\sqrt{dG_t}dB_t$$

$$G_{t+t_0} - G_t \sim Ga(\tfrac{t}{\alpha}, \tfrac{1}{\alpha})$$

with the Lévy measure given by

$$\pi(dx) = \frac{A_\pm^2 \exp(-\frac{A_\pm}{B_\pm}|x|)}{B_\pm|x|}(dx)$$

where $A_\pm = \frac{1}{v}\sqrt{\frac{\gamma^2 v^2}{4} + \frac{\sigma^2}{2} \pm \frac{\gamma v}{2}}, B_\pm = A_\pm^2 v$. In the case where $\gamma = 0$, the jumps are symmetric about 0 [2]. As one can observe in the figure below, the Variance Gamma Process consists of many small jumps with occasional large jumps.
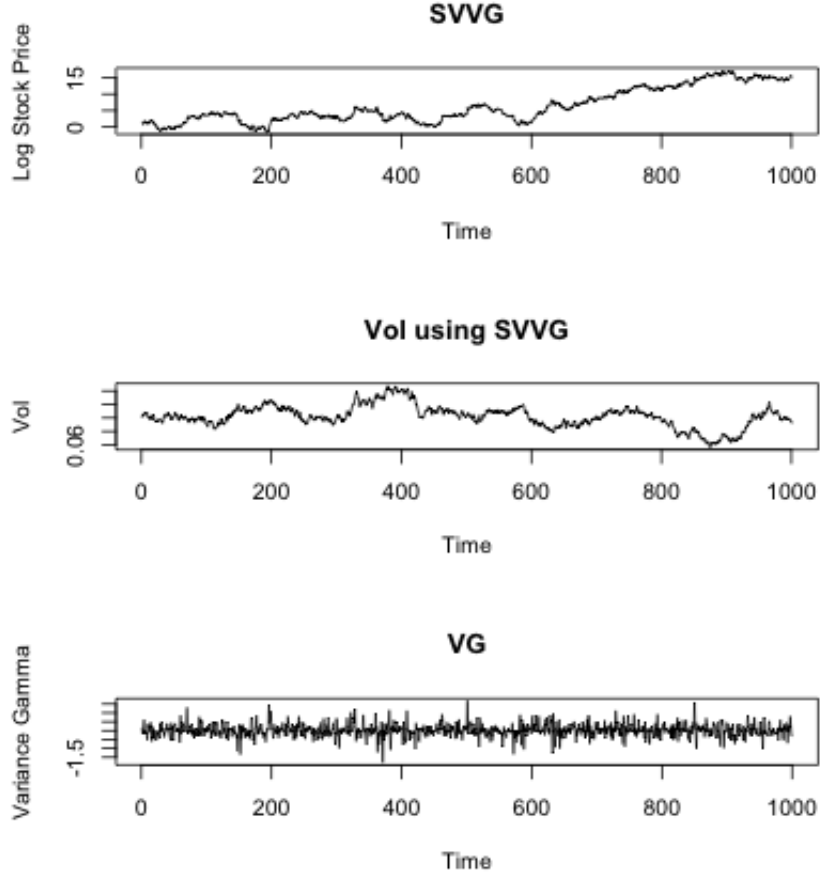
Figure 1: A Simulation of the Heston Model with Variance Gamma Process

2.2 Sequential Monte Carlo

2.2.1 Perfect Monte Carlo

Suppose we are able to simulate N i.i.d. random samples/particles $\{\theta_{0:t}^i, i = 1, 2, ..., N\}$ from a particular distribution (which in this paper, we are refering to the posterior distribution) - $p(\theta_{0:t}|y_{0:t})$. Then, the empirical estimate of this distribution is

$$P_N(d\theta_{0:t}|y_{0:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\theta_{0:t}^i}(d\theta_{0:t}) \tag{1}$$

9

where $\delta_{\theta_{0:t}^i}(d\theta_{0:t})$ refers to the delta-dirac mass located at $\theta_{0:t}^i$. One can then obtain the estimate of $E[f(\theta_{0:t})]$ via

$$E[f(\theta_{0:t})]_{est} = \int f(\theta_{0:t}^i) \, P_N(d\theta_{0:t}|y_{0:t}) = \frac{1}{N} \sum_{i=1}^{N} f(\theta_{0:t}^i) \tag{2}$$

If the posterior variance of $f(\theta_{0:t})$ satisfies $\sigma^2 := E[f(\theta_{0:t})^2] - E[f(\theta_{0:t})]^2 < +\infty$,

$$var(E[f(\theta_{0:t})]_{est}) = \frac{\sigma^2}{N} \tag{3}$$

and by the Strong Law of Large Numbers,

$$E[f(\theta_{0:t})]_{est} \xrightarrow[N\to\infty]{a.s.} E[f(\theta_{0:t})] \tag{4}$$

Also, if $\sigma^2 < +\infty$, the Central Limit Theorem holds

$$\sqrt{N}(E[f(\theta_{0:t})]_{est} - E[f(\theta_{0:t})]) \underset{N\to\infty}{\Longrightarrow} N(0, \sigma^2)] \tag{5}$$

Unfortunately, it often is difficult to sample from the posterior distribution, especially in high dimensional cases. Markov Chain Monte Carlo (MCMC) methods is a possible alternative, but are unsuited for recursive problems [4]. Hence, we will introduce a new method in this paper.

2.2.2 Importance Sampling

One classical method is the use of Importance Sampling method [5]. Suppose we are interested in evaluating $E[f(\theta_{0:t})]$. Using a proposal density $\pi(\theta_{0:t}|y_{0:t})$, from which we can easily sample from, we get the following:

$$
\begin{aligned}
E[f(\theta_{0:t})]_{est} &= \int f(\theta_{0:t}^i) \, P_N(d\theta_{0:t}|y_{0:t}) \\
&= \int f(\theta_{0:t}^i) p(\theta_{0:t}|y_{0:t}) \, d\theta_{0:t} \\
&= \int f(\theta_{0:t}^i) \frac{p(\theta_{0:t}|y_{0:t})}{\pi(\theta_{0:t}|y_{0:t})} \pi(\theta_{0:t}|y_{0:t}) \, d\theta_{0:t} \\
&= \int f(\theta_{0:t}^i) w(\theta_{0:t}^i) \pi(\theta_{0:t}|y_{0:t}) \, d\theta_{0:t}
\end{aligned}
\tag{6}
$$

Since we can easily simulate from N i.i.d. from $\pi(\theta_{0:t}|y_{0:t})$, we can get an estimate of $E[f(\theta_{0:t})]$

$$
\begin{aligned}
\hat{E}[f(\theta_{0:t})]_{est} &= \int f(\theta_{0:t}^i) \, \hat{P}(d\theta_{0:t}) \\
&= \frac{\frac{1}{N}\sum_{i=1}^N f(\theta_{0:t}^i) w(\theta_{0:t}^i)}{\frac{1}{N}\sum_{i=1}^N w(\theta_{0:t}^i)} \\
&= \sum_{i=1}^N f(\theta_{0:t}^i) W(\theta_{0:t}^i)
\end{aligned}
\tag{7}
$$

where $w(\theta_{0:t}^i)$ and $W(\theta_{0:t}^i)$ is known as the importance weights and normalised weights respectively. Note that we are now operating under the new measure

$$
\hat{P}(\theta_{0:t}|y_{0:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\theta_{0:t}^i}(d\theta_{0:t}) W(\theta_{0:t}^i)
\tag{8}
$$

The Importance Sampling method does not work well in this form (especially in high dimensional cases) [6]. In addition, under $\mathcal{F}_{t+1}$, one has to recompute the weights/normalised weights despite having the weights/normalised weights at $\mathcal{F}_t$ - the recursive problem

11

has yet to be solved! However, this has spawned new innovations, one which we will discuss and use in this paper.

2.2.3 Sequential Importance Sampling (SIS)

The Importance Sampling Method can be improved to include new information, $\mathcal{F}_{t+1}$, without recomputing $\{\theta_{0:t}^i, i = 1, 2, ..., N\}$. Using the following fact,

$$P(A_n, .., A_1|B) = P(A_n|A_{n-1}, .., B)P(A_{n-1}|A_{n-2}, .., B)..P(A_1|B) \tag{9}$$

we get,

$$\pi(\theta_{0:t+1}|y_{1:t+1}) = \pi(\theta_{t+1}|\theta_{0:t}, y_{1:t+1})\pi(\theta_{0:t}|y_{1:t}) \tag{10}$$

After iterating,

$$\pi(\theta_{0:t+1}|y_{1:t+1}) = \pi(\theta_0) \prod_{i=1}^{t+1} \pi(\theta_i|\theta_{0:i-1}, y_{1:i}) \tag{11}$$

Our weights can then be recursively updated using the following

$$W(\theta_{0:t+1}^i) \propto W(\theta_{0:t}^i) \frac{p(y_{t+1}|\theta_{t+1}^i)p(\theta_{t+1}^i|\theta_t^i)}{\pi(\theta_{t+1}^i|\theta_{0:t}^i, y_{1:t+1})} \tag{12}$$

One important case (which we will use) is when we adopt the prior distribution as the importance distribution at $t = 0$. However, one drawback of using the SIS method is that one is unable to get the Importance distribution $\pi_t(\theta_{1:t})$

$$\pi_t(\theta_{1:t}) = \int \pi_1(\theta_1) \prod_{k=2}^{t} K_k(\theta_{k-1}, \theta_k) \, d\theta_{1:t-1} \tag{13}$$

12

where $K_k(\theta_{k-1}, \theta_k)$ is known as the Markov Kernel. Although there are some approximations used for local random-walk moves, the complexity of the algorithm is $O(N^2)$ and one cannot compute $K_t(\theta_{t-1}, \theta_t)$ pointwise [7].

2.2.4 Sequential Monte Carlo Samplers (SMC)

The main idea of the Sequential Monte Carlo Samplers is to propose an auxiliary variable technique and introduce artificial backward Markov Kernels, L_{t-1} [7]. Given

$$\widetilde{p}_t(\theta_{1:t}) = \frac{\widetilde{\gamma}_t(\theta_{1:t})}{C_t}$$
$$\widetilde{\gamma}_t(\theta_{1:t}) = \gamma_t(\theta_t) \prod_{k=1}^{t-1} L_k(\theta_{k+1}, \theta_k)$$

where C_t is the normalisation constant for the unnormalised probability density $\widetilde{\gamma}_t(\theta_{1:t})$ and $\gamma_t(\theta_t)$ is the actual unnormalised probability density function of the distribution in question. Suppose, at $\mathcal{F}_t$, we have $\{W_t^i, \theta_{0:t}^i, i = 1, 2, ..., N\}$ approximating $\widetilde{p}_t(\theta_{1:t})$

$$\widetilde{p_t^N}(d\theta_{1:t}) = \sum_{i=1}^{N} W_t^i \delta_{\theta_{1:t}}(d\theta_{1:t})$$
$$W_t^i(\theta_{1:t}^i) = \frac{w_t^i(\theta_{1:t}^i)}{\sum_{i=1}^{N} w_t^i(\theta_{1:t}^i)}$$

where $w_t^i(\theta_{1:t}^i)$ is the standard Importance weights. Now, at $\mathcal{F}_{t+1}$, we bring the particles forwards via the Markov Kernel $K_{t+1}(\theta_t, \theta_{t+1})$. The unnormalised weights is computed using

$$w_{t+1}^i(\theta_{1:t+1}) = \frac{\gamma_{t+1}(\theta_{1:t+1})}{\pi_{t+1}(\theta_{1:t+1})} = w_t^i(\theta_{1:t})\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1}) \tag{14}$$

where the incremental weight, $\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1})$ equals to

$$\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1}) = \frac{\gamma_{t+1}(\theta_{t+1})L_t(\theta_{t+1}, \theta_t)}{\gamma_t(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})} \tag{15}$$

13

In the case of Markov Chain Monte Carlo Kernels, we can define the backward Markov Kernel suboptimally, by

$$L_t(\theta_{t+1}, \theta_t) = \frac{p_{t+1}(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})}{p_{t+1}(\theta_{t+1})} \tag{16}$$

If $p_t \approx p_{t+1}$, then

$$L_t(\theta_{t+1}, \theta_t) \approx L_t^{sub}(\theta_{t+1}, \theta_t) = \frac{p_t(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})}{\int p_t(\theta_t)K(\theta_t, \theta_{t+1})\, d\theta_t}$$

Thus, we will have the following incremental weights

$$\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1}) = \frac{\gamma_{t+1}(\theta_t)}{\gamma_t(\theta_t)} \tag{17}$$

3 Literature Review

3.1 The Model

For this paper, we will be focusing on the Stochastic Volatility Model (Heston) with Variance Gamma Jumps in Returns (SVVG). The model is as follows:

$$dY_t = \mu dt + \sqrt{v_t}[\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}] + dZ_t$$

$$dv_t = \kappa(v - v_t)dt + \sigma_v\sqrt{v_t}dW_{1t} \tag{18}$$

$$dZ_t = \gamma dG_t + \sigma_j\sqrt{dG_t}dB_t$$

Using Euler Discretization, we have the following:

$$Y_{t+1} = Y_t + \mu\Delta + \sqrt{v_t\Delta}\epsilon_{t+1}^y + Z_{t+1}$$

$$v_{t+1} = v_t + \kappa(v - v_t)\Delta + \sigma_v\sqrt{v_t\Delta}\epsilon_{t+1}^v \tag{19}$$

$$Z_{t+1} = \gamma G_{t+1} + \sigma_j\sqrt{G_{t+1}}\epsilon_{t+1}^Z$$

where both ϵ_{t+1}^y and ϵ_{t+1}^v follow $N(0,1)$ with $corr(\epsilon_{t+1}^y, \epsilon_{t+1}^v) = \rho$ while $\epsilon_{t+1}^y, \epsilon_{t+1}^v \perp G_{t+1}$. The Jump Process Z_{t+1} follows a variance gamma process where ϵ_{t+1}^Z follows $N(0,1)$ and $\epsilon_{t+1}^Z \perp \epsilon_{t+1}^y, \epsilon_{t+1}^v, G_{t+1}$. G_{t+1} follows a Gamma Distribution, $\Gamma(\alpha, \beta)$.

Hence, we have observations $(Y_t)_{t=0}^T$; latent variables $(v_t)_{t=0}^T, (Z_t)_{t=1}^T, (G_t)_{t=1}^T$; and parameters $\Theta = \{\mu, \rho, \kappa, v, \sigma_v, \gamma, \sigma_j\}$. For convenience, we let $\theta = \{\Theta, (v_t)_{t=0}^T, (Z_t)_{t=1}^T, (G_t)_{t=1}^T\}$.

3.2 Derivation of Posterior Density

Since $corr(\epsilon_{t+1}^y, \epsilon_{t+1}^v) = \rho$ and $\epsilon_{t+1}^y, \epsilon_{t+1}^v \sim N(0,1)$, conditioning on the values of Z_{t+1}, v_t and Θ, we get

$$\begin{pmatrix} Y_{t+1} - Y_t \\ v_{t+1} - v_t \end{pmatrix} | v_t, Z_{t+1}, \Theta \sim N\left(\begin{pmatrix} \mu\Delta + Z_{t+1} \\ \kappa(\theta - v)\Delta \end{pmatrix}, v_t\Delta \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix} \right) \tag{20}$$

As for the variance gamma process, Li (2006) shows that conditioning on G_{t+1} and Θ, we get

$$Z_{t+1} | G_{t+1}, \Theta \sim N(\gamma G_{t+1}, \sigma_j^2 G_{t+1})$$

$$G_{t+1} | \Theta \sim \Gamma(\frac{\Delta}{v}, v)$$

However, in Jasra (2011), it is shown that G_{t+1} can be integrated out [8]:

$$p(Z_{t+1} | \Theta) = \int p(Z_{t+1} | G_{t+1}, \Theta) p(G_{t+1} | \Theta) \, dG_{t+1}$$

$$= \frac{2 exp(\frac{\gamma Z_{t+1}^2}{\sigma})}{\alpha^{\frac{t-u}{\alpha}} \sqrt{2\pi} \Gamma(\frac{t-u}{\alpha}, \frac{1}{\alpha})} \left(\frac{Z_{t+1}^2}{\gamma^2 + 2\frac{\sigma^2}{\alpha}} \right)^{\frac{t-u}{2\alpha} - \frac{1}{4}} K_{\frac{t-u}{\alpha} - \frac{1}{2}} \left(\frac{\sqrt{Z_{t+1}^2(\gamma^2 + \frac{2\sigma^2}{\alpha})}}{\sigma^2} \right) \tag{21}$$

where $K_a(.)$ is the modified Bessel Function of the second kind, and $\sigma = \sigma_j\sqrt{t-u}$. In this paper, we will let $\alpha = t - u$. Thus, the simulation procedure is reduced by one dimension, and this reduces the simulation complexity, since there is one lesser latent variable to update.

As for the priors, we will follow the ones subscribed by Li (2006) [2]:

- $p(\Theta) = p(\mu)p(\gamma)p(\sigma_j)p(\kappa)p(\sigma_v, \rho)p(v)$

- $\mu, \gamma, \sigma_j \sim N(0, 1)$

- $\kappa, v \sim N(0, 1)$ truncated at 0

- Reparameterize (ρ, σ_v) to (ϕ, w), where $\phi = \sigma_v \rho$ & $w = \sigma_v^2(1 - \rho^2)$

  - $w \sim IG(1.0, 0.5)$

  - $\phi | w \sim N(0, 0.5w)$

Now, the posterior distribution is as follows:

$$p(\Theta, v_{1:n}, Z_{1:n}|Y_{0:n}) \propto p(\Theta) \prod_{i=0}^{n} p(Y_{i+1}, v_{i+1}|Y_i, v_i, Z_{i+1}, \Theta)p(Z_{i+1}|\Theta) \tag{22}$$

3.3 Simulation using a Sequence of Densities

Using the following densities,

$$\pi_k(\theta_{1:n}|y_{0:n}) \propto p(\Theta) \prod_{i=1}^{n} [p(y_i, v_i|y_{i-1}, v_{i-1}, z_i, \Theta)^{\zeta_k} p(z_i|\Theta)] \tag{23}$$

where $0 \leqslant \zeta_1 < \ldots < \zeta_p = 1$. The idea is to simulate from an 'easy' density, before shifting the particles to a much more difficult density. At ζ_1, the tempered posterior density focusses more on the priors than when compared to ζ_k where $k > 1$. Overtime, as the algorithm iterates, the tempered posterior focuses less on the prior and more on the liklihood. Hence, when $\zeta_p = 1$, the particles now exist in the target posterior density [9]. The reason for doing so is because if the densities of the two stages are very close to each other, $L_t^{sub}(\theta_{t+1}, \theta_t) \approx L_t(\theta_{t+1}, \theta_t)$ and we can use result 17.

17

3.4 Simulation Procedure

To start off the simulation, we first sample from the prior distribution π_0 and Importance sampling is then conducted as below

$$
\begin{aligned}
w_1(\theta^i) &= \frac{\pi_1(\theta_1^i|y_{0:n})}{\pi_0(\theta_1^i|y_{0:n})} \\
W_1(\theta^i) &= \frac{w_1^i}{\sum_{i=1}^M w_1^i}
\end{aligned}
\tag{24}
$$

At the second iteration, the particles are shifted from π_1 to π_2 via a kernel of invariant distribution $K_2(\theta_1^i, \theta_2^i)$

$$
w_2(\theta^i) = \frac{\pi_2(\theta_2^i|y_{0:n})}{\int \pi_1(\theta_1^i|y_{0:n})K(\theta_1^i, \theta_2^i)\, d\theta_1}
\tag{25}
$$

There are many choices of kernels, but in this paper we will adopt the Random Walk Metropolis Kernel. Hence, for $t \geqslant 2$,

$$
w_k(\theta^i) = W_{k-1}^i(\theta^i) \prod_{i=1}^n p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}}
\tag{26}
$$

For the Random Walk Metropolis Kernel, one has to specify the proposal variance ψ_{k+1} in order for one to update the parameter values. Using adaptive MCMC techniques, we approximate the mean and therefore the variance of the parameters at iteration k, for the next iteration $k+1$ [10]. Also, if the acceptance rate of the parameter is above 0.85, we tune up the variability by a factor of 5. Likewise, if the acceptance rate of the parameter is below 0.15, we tune down the variability by a factor of 1/5 [9].

3.5 Resampling

Like the problem encountered in the Sequential Importance Sampling, the variability of the weights increases as the iteration increases. This is termed as weight degeneracy [4]. Hence, to counter this problem, we resample the particles according to the normalised weights within the cloud of particles. After resampling, the weights of the particles are reset to 1. In this paper, we will be using the Multinomial Resampling method, even though more sophisticated methods exist. One point to note is that the resampling should not occur too often. When resampling occurs, the number of unique particles fall, hence reducing the particles' approximation of the target density. One criterion to measure the variability of the weights is the Effective Sample Size (ESS) [6].

$$
ESS_k = \frac{(\sum_{i=1}^{M} w_k^i)^2}{\sum_{i=1}^{M} (w_k^i)^2} = \frac{(\sum_{i=1}^{M} W_k^i \prod_{i=1}^{n} p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}})^2}{\sum_{i=1}^{M} (W_k^i \prod_{i=1}^{n} p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}})^2} \tag{27}
$$

The idea of this criterion is to give insight on the approximate number of samples relative to an independent simulation approach [9]. When ESS_k drops below some threshold l, we resample.

In this paper, at iteration k of the algorithm, we let $ESS_k = 0.95 ESS_{k-1}$. Using the bisection method, we are able to calculate ζ_k using 27.

3.6 The SMC Algorithm

Therefore, the algorithm is as follows: At iteration k=1, sample $\theta_1^i \sim \pi_1$ for $i = 1, .., M$ and compute

$$
w_1(\theta^i) = \frac{\pi_1(\theta_1^i|y_{0:n})}{\pi_0(\theta_1^i|y_{0:n})}
$$
$$
W_1(\theta^i) = \frac{w_1^i}{\sum_{i=1}^{M} w_1^i}
$$

If $ESS_k \leqslant l$, resample and set $W_1(\theta^i) = 1/M$. Set ψ_{k+1} for each kernel.

At iteration $k = 2, .., p$, set ζ_k and compute

$$w_k(\theta^i) = \frac{\pi_1(\theta_1^i | y_{0:n})}{\pi_0(\theta_1^i | y_{0:n})}$$

$$W_k(\theta^i) = \frac{w_k^i}{\sum_{i=1}^{M} w_k^i}$$

If $ESS_k \leqslant l$, resample and set $W_k(\theta^i) = 1/M$. If $k < l$, set ψ_{k+1} for each kernel [9].

4 Empirical Results

4.1 Simulation Tests

We now provide evidence that this SMC algorithm can accurately estimate the parameters of the SVVG Model. First, using the SVVG Model, we generate 100 paths of 150 days of daily data. The initial value of the volatility is chosen to be 1 and the rest of the latent variables are randomly chosen within their ranges. Then, we run the SMC algorithm with 2000 particles (and resample if $ESS_k < 1000$) and $\zeta_1 = 0.00001$ over these 100 paths to get us the posterior means of the parameters. The 100 different posterior parameter means are then used to generate the posterior densities of the parameter. The 95% highest posterior density interval is also calculated.

On average, the algorithm took 10-20 minutes for each sample path, resamples every 6-8 steps and runs for about 200 steps. From Table 1, 2 out of the 7 parameters do not lie within the 95% credible interval. This could be due to the choice of Δ in the Euler Discretization which would affect the accuracy of the posterior parameter estimates. For this simulation, Δ was chosen to be 1/250 to represent the daily close. One can choose a smaller Δ at the cost of higher complexity for the latent variables as one has to simulate data points between $(t, t + \Delta t)$. Note that from Table 1 and Figure 2, this SMC algorithm has been able to estimate μ quite accurately - a feature which we will use for our proposed Trading Strategy. From Figure 3, the acceptances are considered acceptable, since a Random-Walk Metropolis Kernel is used. However, there are some bouts of sensitivity of the rates during the middle of Iterations. But in general, the adaptive SMC Sampler has been able to prevent acceptance rates from going too low or too high.
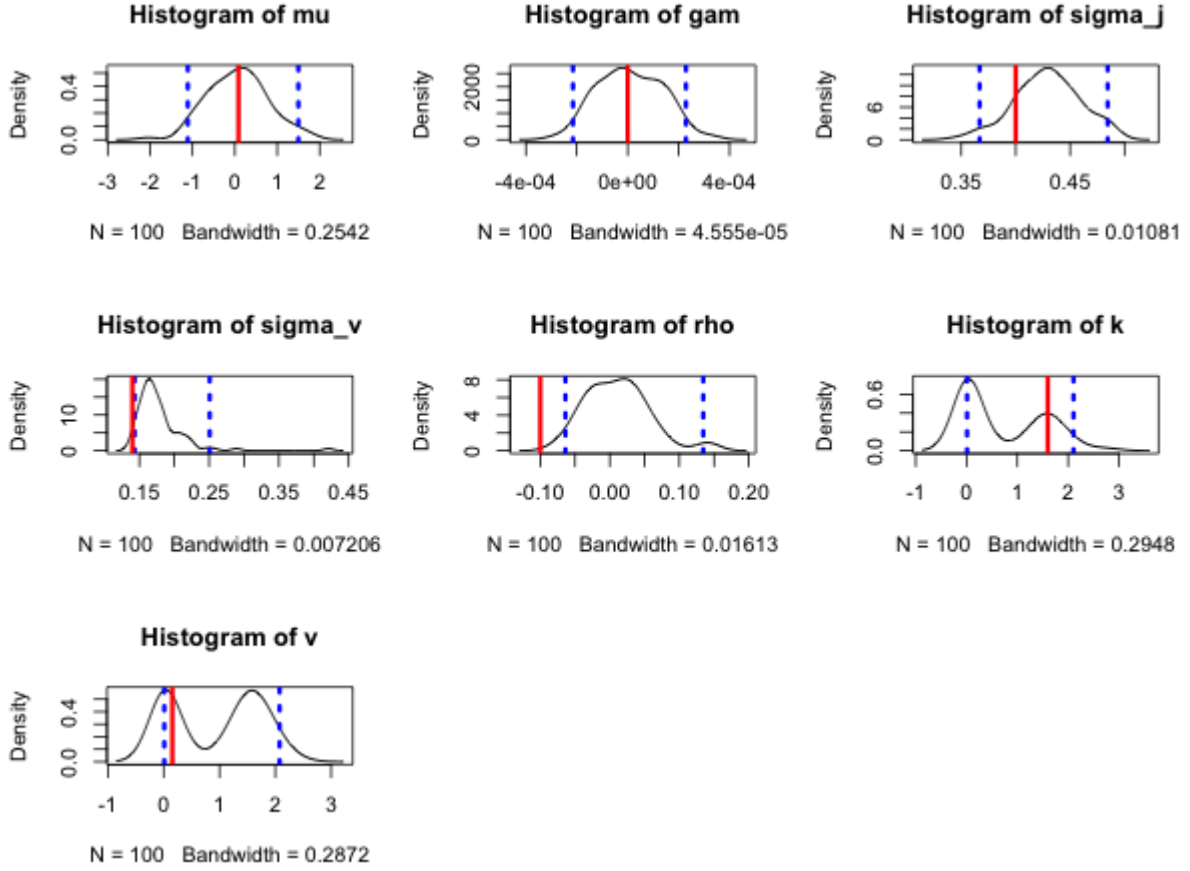
Figure 2: 95% Credible Interval is denoted by the blue dotted line. True value of the parameter of the SVVG Model denoted by the red solid line.

Parameter	True	0.025	Median	0.975
μ	0.085	-1.1108	0.0595	1.4978
γ	0.0	-0.0002	0.0000	0.0002
σ_j	0.4	0.3668	0.4293	0.4845
σ_v	0.14	0.1432	0.1693	0.2508
ρ	-0.1	-0.0638	0.0100	0.1344
κ	1.6	0.0107	0.0479	2.1061
v	0.155	0.0109	1.2968	2.0762

Table 1: 95% Credible Interval and Median
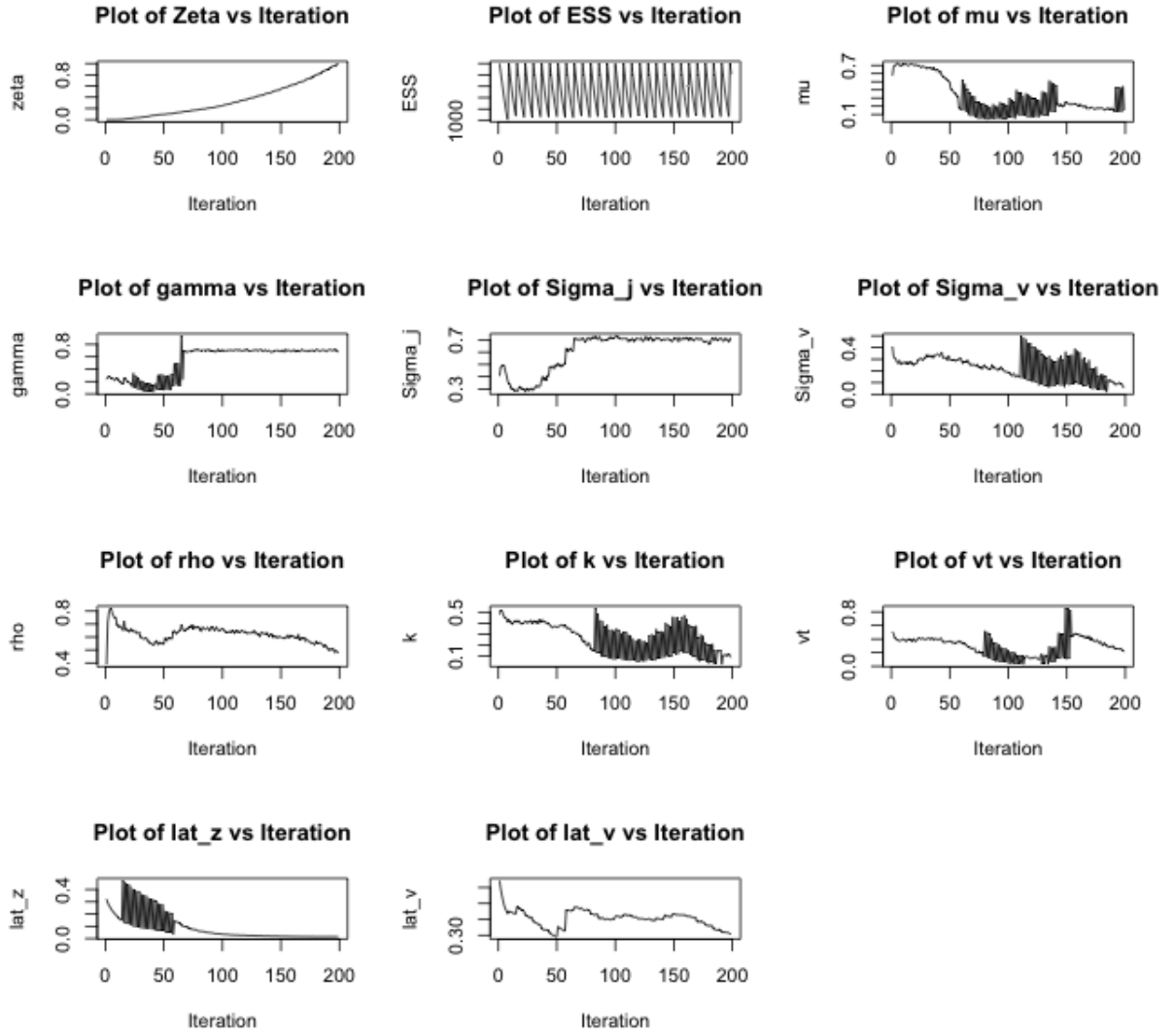
Figure 3: Plot of Acceptance Rates of Parameters vs Iteration

4.2 The Trading Strategy

For this section we will introduce a trading strategy using the result of the SMC Algorithm and see if it can outperform the traditional Buy and Hold (BnH) strategy. We will be using the daily data of S&P500 from 3/1/2007 to 4/3/2016. Then given a user defined lookback period, n, we take a rolling window of n Closes until 4/3/2016. Next, for each window, we perform the SMC algorithm to get us the posterior mean of μ, which is the posterior expectation of the log returns of the instrument. After that, we generate buy/sell signals if $\mu > \mu_{thres}/\mu < -\mu_{thres}$. For the trading strategy, we take $n = 30$, $\mu = 0$. Also, we assume that we can buy/sell S&P500 and that there are no transaction costs.

For comparison between strategies, the metric we will be using is called the CAR/MDD. It is simply the compounded annual return (CAR) divided by the max drawdown (MDD). Intuitively, a higher score means that our equity curve is smoother and therefore a better system.

4.2.1 Naive Trading Strategy

We model the trading strategy with zero transaction costs. The signal generated from the window of Closes is to be applied on the next day's Open. A positive/negative posterior mean would give a buy/sell signal for the next day. We will split the trading strategy into 2 separate portfolios that either buys or sells only. It is interesting to note that this naive strategy suffered a deep drawdown for the first 2.5 years, before it started to perform. Comparing between the buy and hold strategy and the Buy strategy, one can postulate that this is due to the initial weak performance of the Buy strategy.
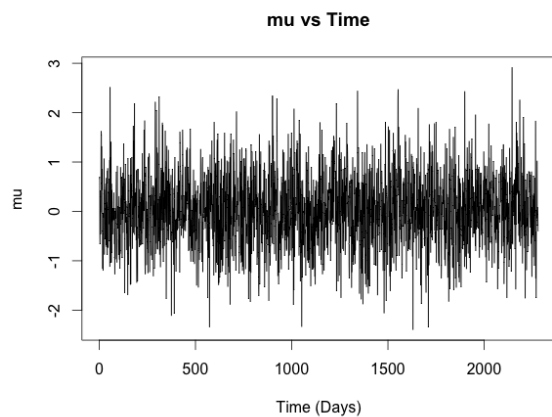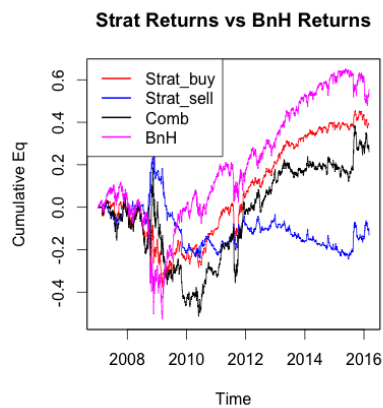
Figure 4: Plot of μ vs Time

Figure 5: Plot of Equity vs TIme

Strategy	CAR/MDD
Buy Only	0.08463839
Sell Only	-0.05490352
Combined	0.01378244
BnH	0.07773994

4.2.2 The Improved Strategy

We will now propose filters to filter out potential bad trades to try to increase profitability and improve the CAR/MDD. Using the first 500 observations as our training set (we select observations 1-500 for convenience), we will be using the following filters:

- Simple Moving Average (SMA) : To capture long term trends

- Exponential Moving Average (EMA): To capture short term movements

Our objective at the end of this is to create a portfolio of multiple trading strategies that hopes to produce superior returns, smoother equity curve and higher CAR/MDD ratio. In each of the trading strategies, we will optimise the input parameters (ie lookback period and μ_{thres}) over the grid to get the highest CAR/MDD for the training set.

Simple Moving Average Strategy

For the SMA trading strategy, we will use a simple moving average (with lookback period n) on the Open of SP500. The reason we calculate it on the Open is because the trading strategy can act upon the newest information, instead of using today's Close as a signal for tomorrow. Hence the trading rules are:

- Buy when Open is above the current SMA price & $\mu > \mu_{thres}$

- Sell when Open is below the current SMA price & $\mu < -\mu_{thres}$

Next we optimise the Buy Strategy and Sell Strategy of the training set over the grids : $\mu_{thres} \in [0, 1.0], n \in [3, 60]$. The optimal parameters for the Buy and Sell strategies are $(0.8, 60), (0.9, 51)$ respectively. Next, we plot the whole equity curve (training + test sets) below. One may observe that the Sell strategy does not trade as often as one may expect - perhaps this is favourable because the SP500 has been increasing from 2009-2016 (and that Sell Strategies will tend to underperform under such conditions).
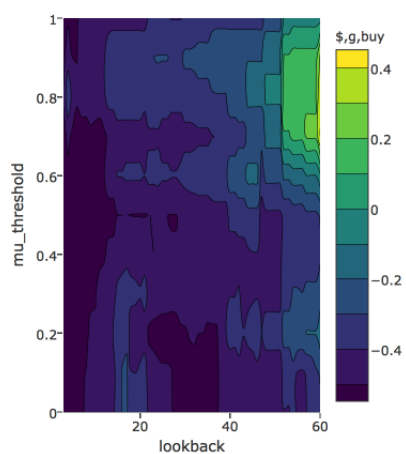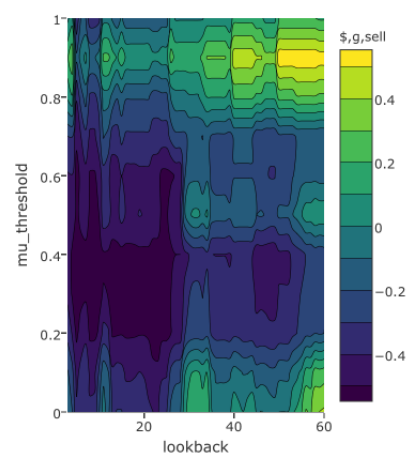
Figure 6: Contour for Buy Strategy (SMA)

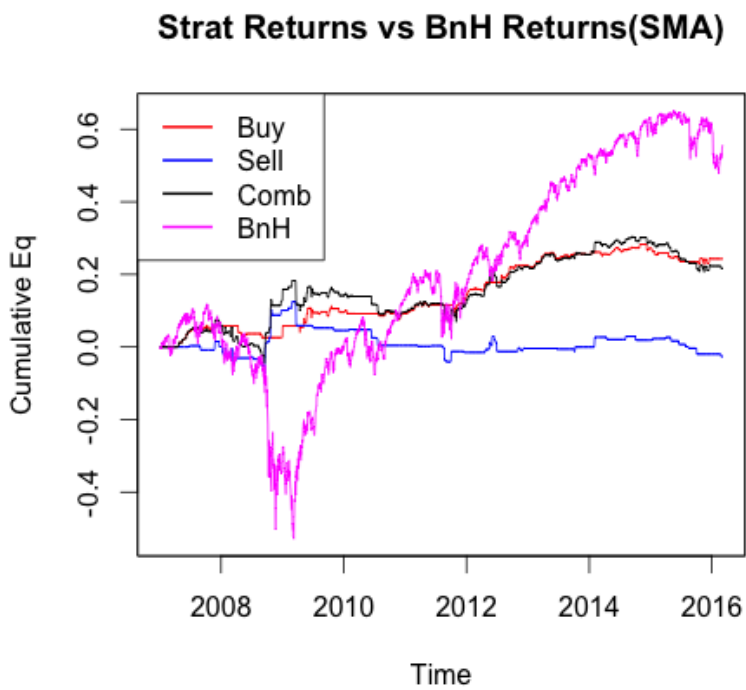Figure 7: Contour for Sell Strategy (SMA)

Figure 8: Equity Curve for Optimised Buy/Sell Strategy (SMA)

27

Strategy - SMA	CAR/MDD(Trg)	CAR/MDD(Test)
Buy Opt	0.4022432	0.34835462
Sell Opt	0.5756164	-0.124869920
Combined Opt	0.5475945	0.07000729
BnH	-0.3938852	0.502743

Table 2: Statistics for Optimised Strategy (SMA)

Exponential Moving Average Strategy

Like the SMA trading strategy, we will use an exponential moving average (with lookback period n) on the Open of SP500. The trading rules are:

- Buy when Open is above the current EMA price & $\mu > \mu_{thres}$

- Sell when Open is below the current EMA price & $\mu < -\mu_{thres}$

We optimise the Buy Strategy and Sell Strategy of the training set over the grids : $\mu_{thres} \in [0, 1.0], n \in [3, 60]$. The optimal parameters for the Buy and Sell strategies are $(0.4, 3), (0.7, 5)$ respectively. Next, we plot the whole equity curve (training + test sets) below.
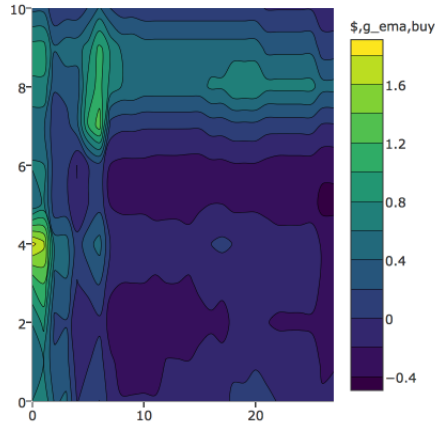

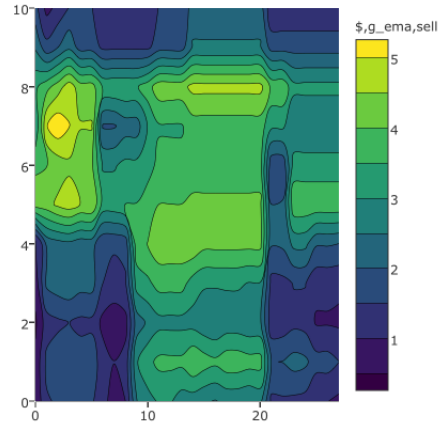Figure 9: Contour for Buy Strategy(EMA)

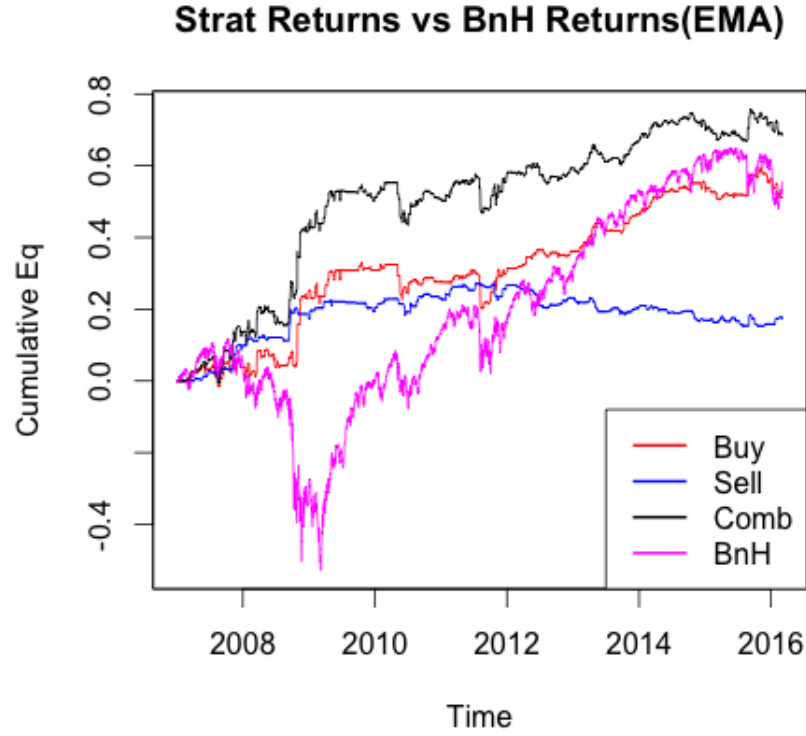Figure 10: Contour for Sell Strategy(EMA)

Figure 11: Equity Curve for Optimised Buy/Sell Strategy (EMA)

| Strategy - EMA	CAR/MDD(Trg)	CAR/MDD(Test)
Buy Opt	1.852921	0.3027411
Sell Opt	5.3704	-0.02609313
Combined Opt	2.951226	0.3053557
BnH	-0.3938852	0.5027435

Table 3: Statistics for Optimised Strategy (EMA)

4.2.3 Combined Strategies

Now that we have optmised the training sets of the 2 different strategies, we will now try to combine them into a single portfolio in hope of getting a smoother equity curve. We will optimise the strategies as follows:

- Step 1: Optimise weights w_1, w_2, w_3, w_4 (SMA Buy, SMA Sell, EMA Buy, EMA Sell respectively), based on the first 100 readings of the strategies.

- Step 2: Walk the portfolio for the next k days

- Step 3: Re-optimise the weights using the past 100 days

- Step 4: Iterate until end

We will set $k = 20$ for this example. The plots and relevant statistics are shown below:

Optimised Portfolio vs BnH
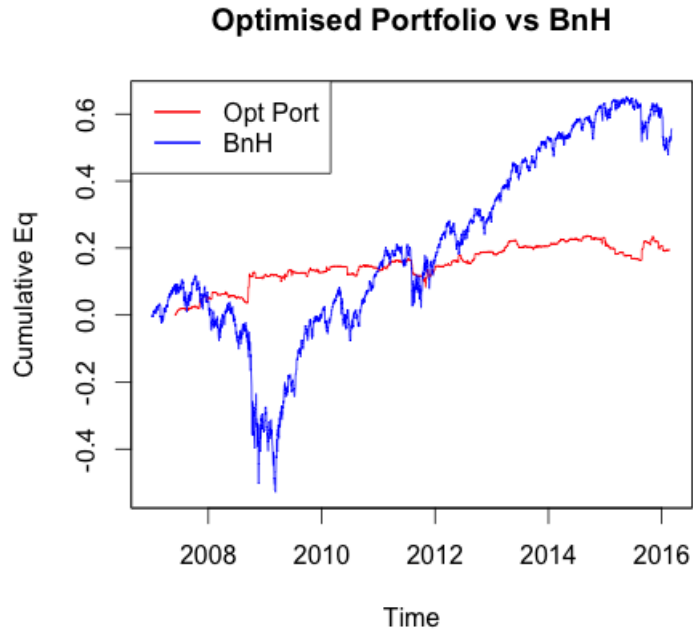
Strategy	CAR/MDD
Opt. Portfolio	0.2590368
BnH	0.07773994

5   Conclusion

In conclusion, we have introduced the Lévy process and Sequential Monte Carlo and applied the simulation method on the SP500 to obtain the posterior estimates. The posterior estimates have been shown to be relatively accurate and one may tune the accuracy up by decreasing the delta in the Euler Discretisation. As for the implementation of the trading strategies, it is unfortunate that the strategies are unable to exceed the traditional buy and hold in terms of absolute returns even though it scores higher for the CAR/MDD. One reason could be that the strategies themselves (without incorporating information of the posterior estimates) are not known to be very profitable. In all, this paper hopes to not only be a foundation for readers to get acquainted with the SMC and the Models that incorporates the Lévy Process, but also to inspire further experimentation with the posterior estimates and trading strategies.

6   Appendix

$P(A_n, .., A_1|B) = P(A_n|A_{n-1}, .., B)P(A_{n-1}|A_{n-2}, .., B)..P(A_1|B)$

Proof:

$$
\begin{aligned}
&P(A_n|A_{n-1}, .., B)P(A_{n-1}|A_{n-2}, .., B)..P(A_1|B) \\
&= \frac{P(A_n, A_{n-1}, .., B)}{P(A_{n-1}, A_{n-2}, .., B)} \frac{P(A_{n-1}, A_{n-2}, .., B)}{P(A_{n-2}, .., B)} ... \frac{P(A_1, B)}{P(B)} \\
&= \frac{P(A_n, .., A_1, B)}{P(B)} \\
&= P(A_n, .., A_1|B)
\end{aligned}
$$

If $p_t \approx p_{t+1}$, then $\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1}) = \frac{\gamma_{t+1}(\theta_t)}{\gamma_t(\theta_t)}$

Proof:

Using equations 15 and 16,

$$
\begin{aligned}
\widetilde{w}_{t+1}^i(\theta_t, \theta_{t+1}) &= \frac{\gamma_{t+1}(\theta_{t+1})L_t(\theta_{t+1}, \theta_t)}{\gamma_t(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})} \\
&= \frac{\gamma_{t+1}(\theta_{t+1})}{\gamma_t(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})} \cdot \frac{p_{t+1}(\theta_t)K_{t+1}(\theta_t, \theta_{t+1})}{p_{t+1}(\theta_{t+1})} \\
&= \frac{\gamma_{t+1}(\theta_{t+1})}{\gamma_t(\theta_t)} \cdot \frac{p_{t+1}(\theta_t)}{p_{t+1}(\theta_{t+1})} \\
&= \frac{\gamma_{t+1}(\theta_{t+1})}{\gamma_t(\theta_t)} \cdot \frac{\frac{\gamma_{t+1}(\theta_t)}{Z_{t+1}}}{\frac{\gamma_{t+1}(\theta_{t+1})}{Z_{t+1}}} \\
&= \frac{\gamma_{t+1}(\theta_t)}{\gamma_t(\theta_t)}
\end{aligned}
$$

where we have used the fact that $p_t = \frac{\gamma_{t+1}(\theta_t)}{Z_{t+1}}, Z_{t+1} :=$ the normalisation constant.

References

[1] P. Carr and L. Wu, "Time-changed lévy processes and option pricing," *Journal of Financial economics*, vol. 71, no. 1, pp. 113–141, 2004.

[2] H. Li, M. T. Wells, and L. Y. Cindy, "A bayesian analysis of return dynamics with lévy jumps," *Review of Financial Studies*, vol. 21, no. 5, pp. 2345–2378, 2008.

[3] R. Cont and P. Tankov, *Financial modelling with jump processes.* Chapman & Hall, 2004.

[4] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential monte carlo methods," in *Sequential Monte Carlo methods in practice.* Springer, 2001, pp. 3–14.

[5] J. Geweke, "Bayesian inference in econometric models using monte carlo integration," *Econometrica: Journal of the Econometric Society*, pp. 1317–1339, 1989.

[6] A. Smith, A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer Science & Business Media, 2013.

[7] P. Del Moral, A. Doucet, and A. Jasra, "Sequential monte carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[8] D. B. Madan, P. P. Carr, and E. C. Chang, "The variance gamma process and option pricing," *European finance review*, vol. 2, no. 1, pp. 79–105, 1998.

[9] A. Jasra, D. A. Stephens, A. Doucet, and T. Tsagaris, "Inference for lévy-driven stochastic volatility models via adaptive sequential monte carlo," *Scandinavian Journal of Statistics*, vol. 38, no. 1, pp. 1–22, 2011.

[10] C. Andrieu, É. Moulines *et al.*, "On the ergodicity properties of some adaptive mcmc algorithms," *The Annals of Applied Probability*, vol. 16, no. 3, pp. 1462–1505, 2006.