Parameter Inference with MCMC in Heston's Model

Group Number: 18

February 27, 2024

Abstract

The Wiener process is an important model in financial mathematics, where it is used for pricing assets. However, the model assumes these asset prices stay constantly volatile over time, which fails in many practical scenarios. To overcome this constraint, Heston's stochastic volatility model extends upon the Wiener process framework by modelling volatility with another stochastic process. Due to this hierarchical structure of many parameters, parameter estimation of Heston's model requires a full Bayesian treatment to avoid overfitting, and this must be done with MCMC. In this project, we will explore the motivation and theory of Heston's model from first principles. Next, we will explore the theory of MCMC and run simulations for sampling the posterior distribution of the parameters. This will be finally applied on a Google stock dataset, where Heston's models find the most applications.

1 Introduction

For financial corporations, asset pricing is a crucial practice that determines the current value of assets, such as stocks and their derivatives, in order to assess their potential future performance. This process provides investors with the information they need to make educated decisions about where to invest their money. By understanding a company's valuation and its potential for growth or decline, investors can strategically place their capital in investments that are likely to offer the best returns. The Wiener process and its extensions, such as Heston's model, are widely used to model these prices because of their strong history of good performance and high reliability.

This project has three goals that benefit the group and the wider STAT0009 cohort: **(1)** Firstly, wanted to understand the literature on Wiener process and Heston's model, which expands on ideas of stochastic processes in the course, STAT0013 and MATH0031. **(2)** Secondly, we aim to explore the applications of Bayesian inference and MCMC methods in finance, building upon the foundations established in STAT0008. **(3)**. Finally, our goal is to break down the concepts and re-implement the source code in a way that's easily understandable for Statistical Science students. We provide clear explanations for key derivations, primarily referencing the work of Cape et al. unless otherwise noted.

2 Background Theory

First, we need to review the fundamentals of stochastic processes and stochastic differential equations, emphasizing their connection and their relevance to financial asset pricing. A stochastic process is a collection of random variables, indexed by time, that represents the evolution of some random value or system over time. A stochastic differential equation (SDE) is a differential equation that integrates a stochastic process as a component. This equation typically has terms that reflect deterministic trends, similar to those outlined by traditional differential equations, and stochastic terms that account for the randomness. Both stochastic processes and SDEs model the random evolution of a system over time. The key similarity is their ability to incorporate randomness. However, SDEs provide a more structured mathematical framework by explicitly distinguishing between deterministic and random influences.

Stock prices are known to fluctuate unpredictably on a daily basis. A SDE, with its integration of both deterministic and random trends, is similar to the way stock prices are influenced by sophisticated market and economic dynamics. Therefore, using SDEs to model stock prices allows for a mathematical representation that accurately captures the unpredictable and dynamic nature of financial markets. One of the key stochastic process that is used for modelling stock prices is the Wiener process.

2.1 Wiener Process

Definition 1 (Weiner Process)**.** *A **Wiener Process** (also known as Brownian Motion) is a continuous-state stochastic process* $\{W_t : t \geq 0\}$ *with the following properties:*

1. $W_0 = 0$
2. *All non-overlapping increments of the process are independent. That is,* $W_{s_2} - W_{s_1}$ *is independent of* $W_{t_2} - W_{t_1}$ *for all* $0 \leq s_1 < s_2 \leq t_1 < t_2$*.*
3. *For* $0 \leq s < t$, $(W_{t+s} - W_s) \sim N(0, t)$*. The variance of the step is proportional with the time interval. That is, a process* $W = \{W_t\}_{t \geq 0}$ *following Brownian motion has increments that can be expressed as*
$$\Delta W = W_{t+\Delta t} - W_t = \sqrt{\Delta t} \cdot \varepsilon_t$$
where $\varepsilon_t \sim N(0, 1)$*.*

A Wiener process with drift μ and volatility σ incorporates linear growth and random fluctuations into its behavior. The value X_t of this adjusted process at time can be modelled by:

$$X_t = \mu t + \sigma W_t$$

The meaning of each term, with their impact, is visualised in Figure 1:

1. μt: This is the constant drift component. It represents a constant trend over time t. If μ is positive, there is a tendency for the process to increase over time; if μ is negative, there is a tendency to decrease.
2. σW_t: This term scales the standard Wiener process by the constant **volatility** σ. Volatility represents the degree of variation or fluctuation in the process's path over time. If σ is large, the process will exhibit larger fluctuations; if σ is small, the fluctuations will be more restrained.

The SDE governing the process can be presented as follows, with the detailed derivation provided in the Appendix A:

$$dX_t = \mu dt + \sigma dW_t \tag{1}$$

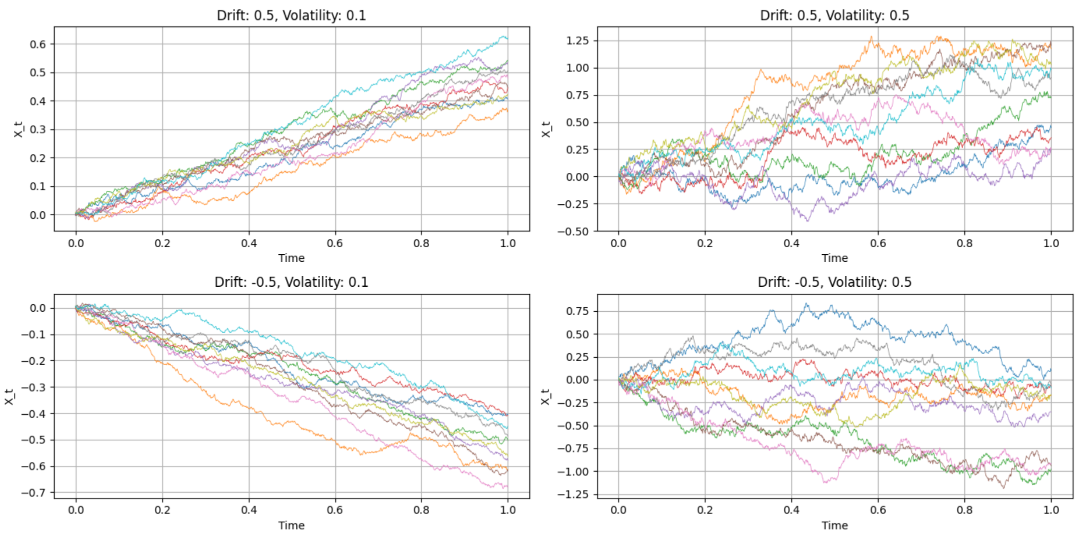

Figure 1: Simulated paths of a Wiener process with varying drift and volatility parameters. Top-left: Positive drift (0.5) with low volatility (0.1). Top-right: Positive drift (0.5) with high volatility (0.5). Bottom-left: Negative drift (-0.5) with low volatility (0.1). Bottom-right: Negative drift (-0.5) with high volatility (0.5). Each subplot shows multiple realisations to illustrate the variability of paths.

2.2 A brief introduction to Itô's lemma

Ito's lemma [5] is a key result in the field of stochastic calculus for calculating the derivative of functions of stochastic processes:

Definition 2 (Itô's Process)**.** *An Itô process is a stochastic process X_t which generalises the Wiener process to having drift and diffusion coefficients as a function of time and current state. It can be described by the SDE:*

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$$

- $\mu(X_t, t)$ *- the drift coefficient,*
- $\sigma(X_t, t)$ *- the diffusion coefficient,*
- W_t *- a Wiener process.*

Lemma 3 (Itô's Lemma)**.** *Let $f(X_t, t)$ be a function that is twice continuously differentiable in both X_t and t. Then, the differential of f is:*

$$df(X_t, t) = \left(\frac{\partial f}{\partial t} + \mu(X_t, t)\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2(X_t, t)\frac{\partial^2 f}{\partial X_t^2} \right) dt + \sigma(X_t, t)\frac{\partial f}{\partial X_t} dW_t$$

Intuitively, Itô's lemma acts like the "chain-rule" in stochastic calculus when a transformation is applied on the Itô process. This is important because it allows us to effectively manipulate stochastic processes, just as how we manipulate deterministic functions in classical calculus to analyse and solve differential equations. We will apply Itô's lemma to derive important results in the next sections.

2.3 Calibrating model parameters with historical data

From the definitions above, the parameters μ and σ are arbitrary. To calibrate, or optimise, the model parameters on historical data, we can use statistical point estimation methods such as Maximum Likelihood Estimation (MLE) or Least Squares (LS) [3]. Let $\{X_t\}_{t=1}^{T}$ represent the observed process values at discrete times $t = 1, 2, ..., T$. In LS, for example, the calibration aims to solve the following optimisation problem:

$$\min_{\mu,\sigma} \sum_{t=1}^{T-1} \left(X_{t+1} - X_t - \mu\Delta t - \sigma(W_{t+1} - W_t)\right)^2$$

where Δt is the time increment, $\mu\Delta t + \sigma(W_{t+1} - W_t)$ is the model increment, and $X_{t+1} - X_t$ is the observed data increment. By solving this, we obtain estimates for μ and σ that make the model's predictions as close as possible to the historical data. These estimates are then used to forecast future behavior of the process or to simulate new paths under the same stochastic dynamics.

2.4 Motivation for model extensions

The Wiener process model, however, assumes a constant volatility parameter σ, represented mathematically by the term σW_t. This is a significant limitation for modeling financial assets for several reasons:

- **Constant Volatility:** Volatility tends to behave differently, or in clusters, over time. Financial markets show periods where asset prices are highly volatile, followed by periods of low volatility. For instance, during the 2008 financial crisis, stock markets experienced extremely high volatility as prices fell sharply, followed by periods of relative calm. A constant σ cannot account for these changing conditions.

- **No Leverage Effect:** Often, when the market value of an asset decreases, its volatility increases, which is known as the leverage effect. For example, if a company's stock price drops significantly, it might indicate underlying problems within the company, leading to higher uncertainty (volatility). A constant volatility parameter cannot model this relationship between price levels and volatility.

These limitations motivate the use of more sophisticated models such as the Heston's model, which incorporates a stochastic volatility term.

3 Heston's Model for Stochastic Volatility

The Heston model [4] is an asset-pricing model that defines a joint process between the price of the asset and the variance. The model allows volatility to change over time and the volatility follows a mean-reverting stochastic process. This means that over time, the volatility tends towards its mean function.

Definition 4 (Heston Model)**.** *The Heston model is a dynamical system of SDEs of the form*

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^S \tag{2}$$

$$dV_t = \kappa(\theta - V_t)dt + \sqrt{V_t}\sigma dW_t^V \tag{3}$$

$$dW_t^S dW_t^V = \rho dt \tag{4}$$

- S_t *- price of the asset at time t*
- μ *- drift rate of the asset price*
- V_t *- volatility of the asset returns at time t*
- κ *- mean reversion speed coefficient*
- θ *- long term mean of volatility*
- σ *- variance of the asset volatility*
- W_t^S *and* W_t^V *- two Wiener processes for asset price and volatility*
- ρ *- the correlation between* W_t^S *and* W_t^V

Equation (2) describes the evolution of the asset's price S_t over time. It shows that the asset price changes due to a drift component $\mu S_t dt$ (representing expected returns) and a stochastic component $\sqrt{V_t} S_t dW_t^S$, which allows the volatility term of the path to vary based on a further SDE in equation (3).

The equation (3) models the volatility of the asset's returns V_t so that it has its own drift component $\kappa(\theta - V_t)dt$. Here, $\kappa(\theta - V_t)dt$ represents the difference between the long-term average volatility and the current volatility at time t. This forces mean reversion because if V_t is above θ, the term $\kappa(\theta - V_t)$ becomes negative, pulling V_t back down towards θ, and if V_t is below θ, it becomes positive, pushing V_t up towards θ.

Finally, the correlated Weiner processes, W^S and W^V in (4) captures the interconnected price-volatility dynamics observed in financial markets. This is particularly important for modeling the leverage effect.

As a result of all these distinct features, the Heston model demonstrates robustness in predicting stock prices under various market conditions. In tumultuous market conditions such as a financial crisis, it prices in loss of investor confidence, ensuring assets are valued more accurately than with models assuming constant volatility. As markets begin to stabilise or recover, the model anticipates reductions in prices by predicting volatility reversion. Moreover, in declining markets, it captures between falling stock prices and rising volatility, aiding investors in strengthening their portfolios against potential losses. This adaptability makes the Heston model a valuable tool for managing risk and making informed investment decisions.

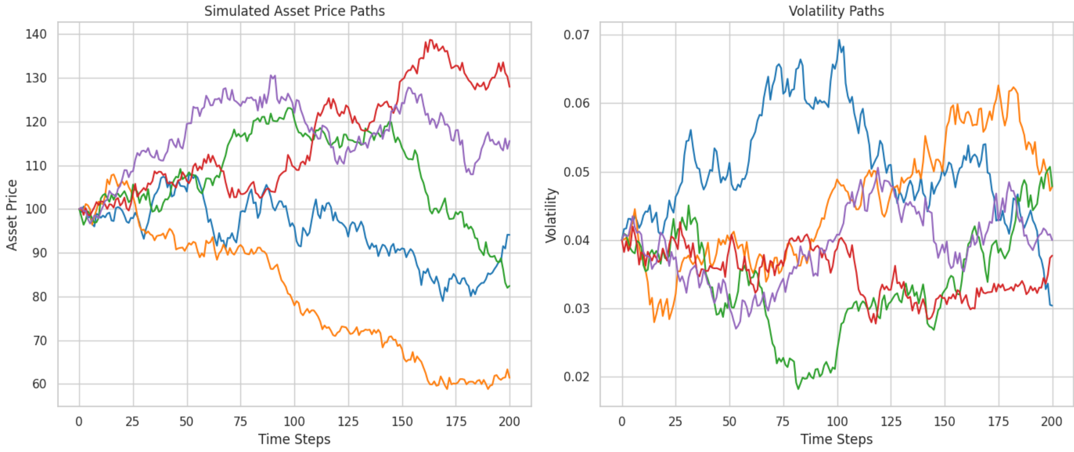

Figure 2: The Heston's model simulated asset prices (left) and the corresponding volatilities (right) over time, highlighting the model's characteristic stochastic price movements and volatility fluctuations.

4 Methodology

The Heston model, with an multi-parametric hierarchical structure, poses a risk of overfitting when calibrated using point estimates. Overfitting means the performance of the model does not transfer well from historical data to unseen data. Point estimation approaches such as MLE and LS have low bias, but high variance in the bias-variance tradeoff, resulting in overfitting.

Unlike point estimates, Bayesian inference is robust against overfitting. However, Bayesian inference is computationally costly with a large high-dimensional dataset. This is resolved by using Markov Chain Monte Carlo (MCMC). MCMC enables us to approximate the posterior distribution over the parameters by generating samples that represent the distribution. This gives us parameter estimates and their associated uncertainty, allowing us to average over multiple plausible models and reduce the variance of our parameter estimates.

This methodology section include three steps

1. We introduce the basic algorithms of MCMC - Gibbs sampling and Metropolis-Hasting.
2. We discretise the Heston's model in order to apply the MCMC algorithms.
3. We derive the posterior distribution of the parameters up to a constant that we can simulate on.

4.1 Bayesian Inference and Markov Chain Monte-Carlo

Bayesian inference is a method to update the probability of a parameter - the **posterior** based on prior belief - the **prior** and new evidence - the **likelihood**. It provides a principled way to update our beliefs in light of new data using Bayes' theorem:

Theorem 5 (Bayes Theorem)**.** *Let X be data generated/sampled from a population distribution $\mathcal{D}$, of which a is inferred from. Then*

$$P(a|X) = \frac{P(X|a)P(a)}{P(X)}$$

where:

- $P(a|X)$ *is the posterior probability of the parameter* a *given the data* X*,*
- $P(X|a)$ *is the likelihood of the data* X *given the parameter* a*,*
- $P(a)$ *is the prior probability of the parameter* a*,*
- $P(X)$ *is the marginal likelihood (evidence) of the data* X*.*

$$P(X) = \int P(X|a)P(a)da$$

The marginal likelihood $P(X)$ integrates over all possible values of a. For complex models or large parameter spaces, this integral becomes intractable due to the high-dimensional space over which the integration must be performed. MCMC algorithms, such as Metropolis-Hastings and Gibbs Sampling, offer a solution by generating samples from the posterior distribution $P(a|X)$ and constructing a Markov chain with the desired distribution as the equilibrium distribution. The only thing we need is the posterior distribution up to a proportional constant.

Metropolis-Hastings Algorithm

Intuition: The Metropolis-Hastings algorithm proposes a new state (parameter value) based on a proposal distribution and accepts this new state with a probability.

Algorithm 1 Metropolis-Hastings Algorithm

```
Initialize a^(0) arbitrarily
for t = 0 to T do
    Propose a' ~ q(a'|a^(t))
    Compute acceptance ratio accept = P(a'|X)q(a^(t)|a') / P(a^(t)|X)q(a'|a^(t))
    Set a^(t+1) = a' with probability min(1, accept)
    if a' is not accepted then
        Set a^(t+1) = a^(t)
```

Gibbs Sampling

Intuition: Gibbs sampling is a special case of the Metropolis-Hastings algorithm where the proposal distribution is always accepted. It works by updating one parameter at a time, conditional on all other parameters, which removes the computation of an acceptance ratio.

Algorithm 2 Gibbs Sampling Algorithm

```
Initialize a^(0) arbitrarily in the parameter space
for each iteration t do
    for each parameter a_i in a do
        Sample a_i^(t+1) from P(a_i|a_{-i}^(t), X), where a_{-i} denotes all parameters except a_i
    Update a^(t) to a^(t+1) for the next iteration
```

By iteratively applying these steps, both Metropolis-Hastings and Gibbs Sampling algorithms generate a sequence of samples that approximate the posterior distribution without directly computing the marginal likelihood. Constructing a fast-converging Markov chain with efficient sampling requires several considerations. Firstly, the chosen proposal distribution in the Metropolis-Hastings algorithm should closely matches the shape of the target distribution to facilitate higher acceptance rates and more representative sampling. For Gibbs Sampling, ensuring that the conditional distributions are easy to sample from greatly increases efficiency. Additionally, the initial state of the chain should be chosen to be close to the high-density regions of the target distribution to reduce burn-in time (the initial period before the chain converges). Tuning parameters, like the scale of the proposal distribution, and utilising adaptive techniques that adjust the proposal based on the history of the chain, can further enhance efficiency. Monitoring convergence diagnostics and autocorrelation within the chain helps in assessing efficiency and ensuring that the samples generated are a good representation of the target distribution.

4.2 Model discretisation

We discretise the model to make it compatible with discrete historical data, as stock tickers operate in discrete time. In practice, it is common to model prices using logarithms as they capture the percentage change in value, offering a clearer view of investment performance over time. Specifically, log-returns are preferred in finance for understanding how returns compound. To model the log-returns of asset prices, Itô's lemma is applied to transform the stochastic differential equation of S_t into a log-return process. The derivation can be found in the appendix A and the differential equation for the log-return, $d(\ln\, S_t)$, is expressed as:

$$d(\ln\, S_t) = \left(\mu - \frac{1}{2}V_t\right) dt + \sqrt{V_t} dW_t^S \tag{5}$$

To discretise the equation, we use the concept:

$$X_t = X_{t-1} + \Delta X_{t,t-1} = X_{t-1} + \mu(X_{t-1}, t-1)\Delta t + \sigma(X_{t-1}, t-1)\Delta W_t^S.$$

Let $Y_t = \ln S_t - \ln S_{t-1}$. By property 3 of the Wiener Process, we have $\Delta W_t = \sqrt{\Delta t} \cdot \epsilon_t^S$, where $\epsilon_t^S \sim \mathcal{N}(0, 1)$, also

$$\begin{aligned}
\ln S_t - \ln S_{t-1} &= \left(\mu - \frac{1}{2}V_{t-1}\right) \Delta t + \sqrt{V_{t-1}} dW_t^S \\
Y_t &= \left(\mu - \frac{1}{2}V_{t-1}\right) \Delta t + \sqrt{V_{t-1}}\sqrt{\Delta t}\epsilon_t^S
\end{aligned}$$

By the same reason:

$$\begin{aligned}
V_t - V_{t-1} &= \kappa(\theta - V_{t-1})\Delta t + \sqrt{V_{t-1}}\sqrt{\Delta t}\epsilon_t^V \\
\epsilon_t^V &\sim \mathcal{N}(0, \sigma_V^2) \\
Corr(\epsilon_t^S, \epsilon_t^V) &= \rho
\end{aligned}$$

Without loss of generality, we assume $\psi = \rho\sigma_V$ and $\Omega = \sigma_V^2(1 - \rho^2)$. These substitutions allow us to simplify the equations further.

$$\begin{aligned}
Y_t &= \mu\Delta t - \frac{1}{2}V_{t-1}\Delta t + \sqrt{V_{t-1}}\sqrt{\Delta t}\varepsilon_t^S \\
V_t &= \kappa\theta\Delta t + (1 - \kappa\Delta t)V_{t-1} + \sigma\sqrt{V_{t-1}}\sqrt{\Delta t}\varepsilon_t^V
\end{aligned}$$

Rearranging the equations above:

$$\begin{aligned}
\epsilon_t^S &= \frac{Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t}{\sqrt{V_{t-1}}\sqrt{\Delta t}} \\
\epsilon_t^V &= \frac{V_t - \kappa\theta\Delta t - (1 - \kappa\Delta t)V_{t-1}}{\sqrt{V_{t-1}}\sqrt{\Delta t}} \\
(\epsilon_t^S, \epsilon_t^V) &\sim \mathcal{N}\left((0, 0), \begin{bmatrix} 1 & \rho\sigma_V \\ \rho\sigma_V & \sigma_V^2 \end{bmatrix}\right) = \left((0, 0), \begin{bmatrix} 1 & \psi \\ \psi & \psi^2 + \Omega \end{bmatrix}\right)
\end{aligned} \tag{6}$$

By this discretisation process, we have developed a method to find the log-returns $\{Y_1, \ldots, Y_n\}$ from $\{S_1, \ldots, S_n\}$. This method effectively transforms the series of stock prices into a series of log-returns, capturing the percentage changes in price in a way that is more suitable for statistical analysis and modelling.

4.3 Posterior distribution of model parameters

We need to apply Bayes' Theorem to derive the posterior distribution over model parameters. As we use MCMC to sample this posterior, we only need to derive it up to a constant of proportionality. The formula is therefore:

$$P(\mathbf{a}|X) \propto P(X|\mathbf{a})P(\mathbf{a}) \tag{7}$$

As described in section (4.2), we run inference on parameters $\mathbf{a} = \{\mu, \psi, \Omega, \theta, \kappa\}$. To achieve this, we apply Bayes' rule (7) by deriving the likelihood for each parameter and multiplying them with the proposed priors for each parameter. With the process involves lengthy algebraic simplification, we opt to use the derived results from Cape et al.. Our primary focus here will be on the resulting posterior form, which will serve as the target distribution for our MCMC algorithm. We assume normal, Inverse Gamma and truncated normal for the priors as they are conjugate priors with light exponential tails, aiding in low bias estimation and fast convergence.

Likelihood distribution of Y_t and V_t

From section 4.2, we see that Y_t and V_t can be written as functions of $\epsilon_t^S, \epsilon_t^V$, but we know from equation (6) that the two terms are jointly Gaussian. Therefore, Y_t and V_t are also jointly Gaussian, conditioning on all the paraters in $\mathbf{a}$. The simplied likelihood distribution is:

$$P(Y_t, V_t|\mathbf{a}) = \Omega^{-T/2}\left(\prod_{t=1}^{T}\frac{1}{V_{t-1}\Delta t}\right)\exp\left(-\frac{1}{2\Omega}\sum_{t=1}^{T}\left[(\Omega+\psi^2)(\epsilon_t^S)^2 - 2\psi\epsilon_t^S\epsilon_t^V + (\epsilon_t^V)^2\right]\right).$$

Posterior distribution of μ

Let the prior of μ be $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Then the posterior of μ is derived using (7). We compute

$$P(\mu|Y_t, V_t, \psi, \Omega, \theta, \kappa) \propto P(Y_t, V_t|\mu, \kappa, \theta, \psi, \Omega) \cdot P(\mu)$$

giving a posterior distribution $\mu \sim \mathcal{N}(\mu^*, \sigma^{*2})$, where

$$\mu^* = \frac{\sum_{t=1}^{T}((\Omega+\psi^2)(Y_t + \frac{1}{2}V_{t-1}\Delta t)/\Omega V_{t-1}) - \sum_{t=1}^{T}(\psi(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1})/\Omega V_{t-1}) + \mu_0/\sigma_0^2}{\Delta t\sum_{t=1}^{T}((\Omega+\psi^2)/\Omega V_{t-1}) + 1/\sigma_0^2}$$

and:

$$\sigma^{*2} = \frac{1}{\Delta t\sum_{t=1}^{T}((\Omega+\psi^2)/\Omega V_{t-1}) + 1/\sigma_0^2}$$

Posterior distribution of ψ and Ω

Let the priors of ψ and Ω be

- $\Omega \sim \mathcal{IG}(\tilde{\alpha}, \tilde{\beta})$, an inverse gamma distribution.
- $\psi|_\Omega \sim \mathcal{N}(\psi_0, \Omega/p_0)$, a normal distribution conditional on Ω.

The posterior distribution of Ω is given by:

$$\Omega \sim \mathcal{IG}(\alpha_*, \beta_*)$$

where:

$$\alpha_* = \frac{T}{2} + \tilde{\alpha},$$
$$\beta_* = \tilde{\beta} + \frac{1}{2}\sum_{t=1}^{T}(\epsilon_t^V)^2 + \frac{1}{2}p_0\psi_0^2 - \frac{1}{2}\frac{(p_0\psi_0 + \sum_{t=1}^{T}\epsilon_t^S\epsilon_t^V)^2}{p_0 + \sum_{t=1}^{T}(\epsilon_t^S)^2}.$$

The posterior distribution of ψ given Ω is:

$$\psi|_\Omega \sim \mathcal{N}(\psi^*, \sigma_\psi^{*2})$$

where

$$\psi^* = \frac{p_0\psi_0 + \sum_{t=1}^{T}\epsilon_t^S\epsilon_t^V}{p_0 + \sum_{t=1}^{T}(\epsilon_t^S)^2},$$
$$\sigma_\psi^{*2} = \frac{\Omega}{p_0 + \sum_{t=1}^{T}(\epsilon_t^S)^2}.$$

Posterior distribution of κ and θ

Let the prior of θ be $\theta \sim \mathcal{N}(\theta_0, \sigma_\theta^2)$. Then the posterior of θ is $\theta \sim \mathcal{N}(\theta^*, \sigma_\theta^{*2})$ where:

$$\theta^* = \frac{\sum_{t=1}^T((\kappa)(V_t - (1-\kappa\Delta t)V_{t-1})/\Omega V_{t-1}) - \sum_{t=1}^T(\psi(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)\kappa/\Omega V_{t-1}) + \theta_0/\sigma_\theta^2}{\Delta t \sum_{t=1}^T(\kappa^2/\Omega V_{t-1}) + 1/\sigma_\theta^2}$$

and:

$$\sigma_\theta^{*2} = \frac{1}{\Delta t \sum_{t=1}^T(\kappa^2/\Omega V_{t-1}) + 1/\sigma_\theta^2}$$

Let the prior of κ be $\kappa \sim \mathcal{N}(\kappa_0, \sigma_\kappa^2)$. Then the posterior distribution of κ is $\kappa \sim \mathcal{N}(\kappa^*, \sigma_\kappa^{*2})$, where:

$$\kappa^* = \frac{\sum_{t=1}^T((\theta - V_{t-1})(V_t - V_{t-1})/\Omega V_{t-1}) - \sum_{t=1}^T(\psi(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)(\theta - V_{t-1})/\Omega V_{t-1}) + \kappa_0/\sigma_\kappa^2}{\Delta t \sum_{t=1}^T((V_{t-1}-\theta)^2/\Omega V_{t-1}/\Omega V_{t-1}) + 1/\sigma_\kappa^2}$$

and:

$$\sigma_\kappa^{*2} = \frac{1}{\Delta t \sum_{t=1}^T((V_{t-1}-\theta)^2/\Omega V_{t-1}) + 1/\sigma_\kappa^2}$$

Posterior distribution of V_t, the state variable of variance

We have that

$$\begin{aligned}
&P(V_t|Y, V_{t+1}, V_{t-1}, \kappa, \theta, \psi, \Omega, \mu) = \\
&\frac{1}{V_t\Delta t}\exp\Bigg(-\frac{1}{2\Omega}\frac{(\Omega+\psi^2)(\frac{1}{2}V_t\Delta t + Y_{t+1} - \mu\Delta t)^2}{V_t\Delta t} \\
&\quad -\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_t\Delta t + Y_{t+1} - \mu\Delta t)(-(1-\kappa\Delta t)V_t - \kappa\theta\Delta t + V_{t+1})}{V_t\Delta t} \\
&\quad -\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_t - \kappa\theta\Delta t + V_{t+1})^2}{V_t\Delta t}\Bigg) \\
&\cdot\exp\Bigg(-\frac{1}{2\Omega}\frac{-2\psi(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1})}{V_{t-1}\Delta t} \\
&\quad -\frac{1}{2\Omega}\frac{(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1})^2}{V_{t-1}\Delta t}\Bigg)
\end{aligned}$$

and by denoting the two exponentials as $\mathcal{A}$ and $\mathcal{B}$, we have

$$P(V_t|Y, V_{t+1}, V_{t-1}, \kappa, \theta, \psi, \Omega, \mu) = \frac{1}{V_t\Delta t} \times \mathcal{A} \times \mathcal{B}$$

By the Markov property, V_{T+1} only depends on V_T. Thus for the derivation of V_{T+1}, we only keep the second exponential part in the posterior distribution, i.e. $P(V_{t+1}|Y, V_t, \kappa, \theta, \psi, \Omega, \mu) = \frac{\mathcal{B}}{V_{t+1}\Delta t}$. Similarly for V_0, the posterior distribution only depends on the parameter and the first exponential part, i.e. $P(V_0|Y, V_1, \kappa, \theta, \psi, \Omega, \mu) = \frac{\mathcal{A}}{V_0\Delta t}$.

Next, we will derive the posterior distributions. We choose Gaussian, Inverse Gamma and truncated Gaussian for inducing geometric ergodicity in the simulation.

MCMC sampling for the posterior distribution

The posterior distributions of wanted parameters are derived from the former part of section (4.2). We then use MCMC to obtain draws from the posterior distribution $P(\mu, \kappa, \theta, \psi, \Omega, V|Y)$. Sampling from the normal distribution is a straightforward process even when conditioned on the other parameters, so we used Gibbs sampler for these parameters. Firstly, we set the initial values of each parameter $\{\mu^{(0)}, \psi^{(0)}, \Omega^{(0)}, \kappa^{(0)}, \theta^{(0)}, V_0^{(0)}, V_1^{(0)}, ..., V_t^{(0)}\}$. Here, the values of $V_t^{(0)}$ are generated using a truncated normal distribution. Then, we use MCMC to update the parameters. In our analysis, we derive the conditional distribution of the parameter μ based on the given values and subsequently draw a sample $\mu^{(1)}$ from this distribution. The progression involves updating our current state to include the obtained values $\{\mu^{(1)}, \psi^{(0)}, \Omega^{(0)}, \kappa^{(0)}, \theta^{(0)}, V_0^{(0)}, V_1^{(0)}, ..., V_{T+1}^{(0)}\}$, where the $V_{T+1}^{(0)}$ is assigned the value 0 for future shifting. Following this, we adopt a similar approach to draw samples for $\psi^{(1)}$ and $\Omega^{(1)}$, iteratively advancing the state of the Markov chain.

Variables κ and θ follow truncated normal distribution, so we use the techniques cited in Cape et al. to obtain our draws.

For other parameters, the distribution of the state space $\{V_0, ..., V_T\}$, we decided to apply a random walk Metropolis-Hastings approach.

At 0th step, we initiate the state space with proposal value:

$$V_0^{(0)}, V_1^{(0)}, ..., V_{T+1}^{(0)}$$

Next, we run the algorithm for n steps. For step g$\in \{1, ..., n\}$, we obtain the proposal density for

$$V_t^{*(g)} = V_t^{(g-1)} + \mathcal{N}_t, \text{ where } \mathcal{N}_t \sim \mathcal{N}(0, \sigma_N^2) \text{ or } t(v)$$

where t $\in \{1, ..., T\}$ and $V_t^{(g-1)}$ is our proposal for $V_t^{(g)}$. We should note that the σ_N^2 is picked parameter for reject rate. Next, we can then derive the likelihood of the proposed volatility $V_t^{*(g)}$ from Cape et al.:

$$\begin{aligned}
\pi(V_t^{*(g)}) = & \frac{1}{V_t^{*(g)}\Delta t} \exp\Bigg(-\frac{1}{2\Omega V_t^{*(g)}\Delta t}(\Omega + \psi^2)\left(\frac{1}{2}V_t^{*(g)}\Delta t + Y_{t+1} - \mu\Delta t\right)^2 \\
& - \frac{1}{2\Omega V_t^{*(g)}\Delta t}\left(-2\psi\left(\frac{1}{2}V_t^{*(g)}\Delta t + Y_{t+1} - \mu\Delta t\right)\left(-\left(1-\kappa\Delta t\right)V_t^{*(g)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)}\right)\right) \\
& - \frac{1}{2\Omega V_t^{*(g)}\Delta t}\left(-\left(1-\kappa\theta\Delta t\right)V_t^{*(g)} - \kappa\Delta t + V_{t+1}^{(g-1)}\right)^2 \\
& - \frac{1}{2\Omega V_{(t-1)}^{*(g)}\Delta t}\left(-2\psi\left(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}^{(g)}\Delta t\right)\left(V_t^{*(g)} - \kappa\theta\Delta t - \left(1-\kappa\Delta t\right)V_{t-1}^{(g)}\right)\right) \\
& - \frac{1}{2\Omega V_{(t-1)}^{*(g)}\Delta t}\left(V_t^{*(g)} - \kappa\theta\Delta t - \left(1-\kappa\Delta t\right)V_{t-1}^{(g)}\right)^2\Bigg)
\end{aligned}$$

and current volatility estimate $V_t^{(g-1)}$:

$$\begin{aligned}
\pi(V_t^{(g-1)}) = & \frac{1}{V_t^{(g-1)}\Delta t} \exp\Bigg(-\frac{1}{2\Omega}\frac{(\Omega + \psi^2)\left(\frac{1}{2}V_t^{(g-1)}\Delta t + Y_{t+1} - \mu\Delta t\right)^2}{V_t^{(g-1)}\Delta t} \\
& - \frac{1}{2\Omega}\frac{-2\psi\left(\frac{1}{2}V_t^{(g-1)}\Delta t + Y_{t+1} - \mu\Delta t\right)\left(-(1-\kappa\Delta t)V_t^{(g-1)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)}\right)}{V_t^{(g-1)}\Delta t} \\
& - \frac{1}{2\Omega}\frac{\left(-(1-\kappa\Delta t)V_t^{(g-1)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)}\right)^2}{V_t^{(g-1)}\Delta t} \\
& - \frac{1}{2\Omega}\frac{-2\psi\left(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}^{(g)}\Delta t\right)\left(V_t^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}^{(g)}\right)}{V_{t-1}^{(g)}\Delta t} \\
& - \frac{1}{2\Omega}\frac{\left(V_t^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}^{(g)}\right)^2}{V_{t-1}^{(g)}\Delta t}\Bigg)
\end{aligned}$$

the acceptance rate is defined as follows:

$$\alpha(V_t^{*(g)}, V_T^{*(g-1)}) = min\left(\frac{\pi(V_t^{*(g)})}{\pi(V_t^{*(g-1)})}, 1\right)$$

Then, a uniform random variable U is stochastically selected from $\mathcal{N}[0,1]$

$$\text{If } U < \alpha(V_t^{*(g)}, V_t^{(g-1)}), V_t^{(g)} = V_t^{(*(g))},$$
$$\text{Otherwise, } V_t^{(g)} = V_t^{(g-1)}$$

With the probability of acceptance rate, $\alpha(V_t^{*(g)}, V_T^{*(g-1)})$, we accept the proposed volatility $V_t^{*(g)}$. By iterating the steps above for $t \in \{1, 2, 3, ..., T\}$, we can find $\{V_1^{(g)}, ..., V_T^{(g)}\}$ for each step $g \in \{1, ..., n\}$. V_0 and V_{T+1} are obtained as in the previous section.

5 Experiment and Result

Our aim is to illustrate the methodology described in the previous section and apply it to the share price of the American technology company Google.

Data Analysis

Data of Google´s (stock ticker: GOOG) stock price was collected daily from 2021-02-23 to 2024-02-22 over the span of three years. In total, there are 755 observations in total, with no empty entries. The dataset consists of 7 covariates corresponding to each column:

- **Date**: The date of the trading session.
- **Open**: The opening price of the stock for the day.
- **High**: The highest price of the stock during the day.
- **Low**: The lowest price of the stock during the day.
- **Close**: The closing price of the stock for the day.
- **Adj Close**: The adjusted closing price of the stock for the day, accounting for any corporate actions, i.e. decisions taken by a company's board of directors that bring about changes to its stock.
- **Volume**: The total volume of stock units traded during the day.

Here, we only use the adjusted closing price in Heston's model, as it ensures a more accurate reflection of a stock's value by accounting for changes such as dividends and stock splits [2]. We observe the behaviour of this quantity in Figure 3.

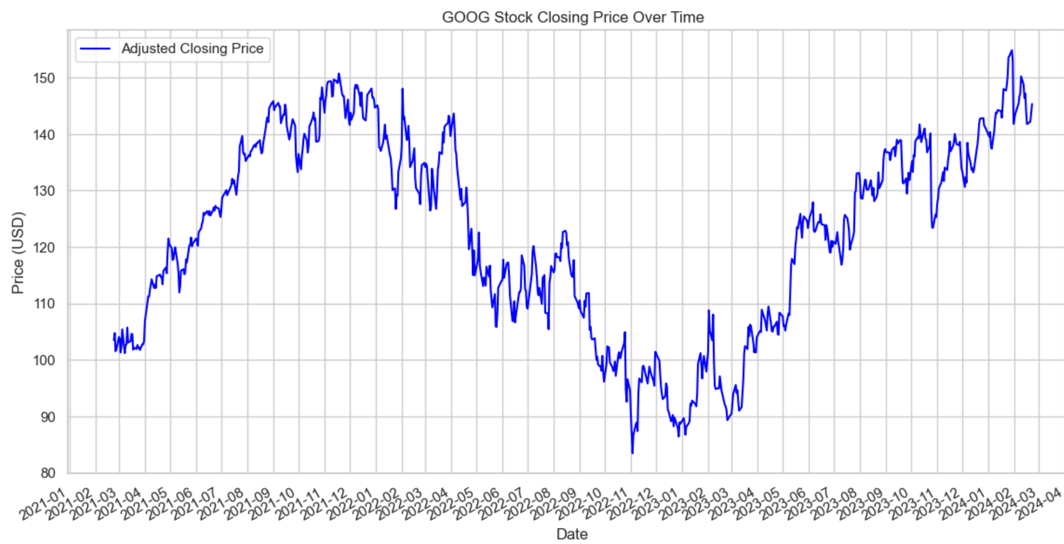

Figure 3: The plot shows the adjusted closing price of GOOG (Google) stock over time, depicting a volatile trend with a significant dip in the middle of the timeframe followed by a recovery. The stock price appears to experience sharp fluctuations but overall shows an upward trend in the period displayed.

Initialisation of Heston's model parameters

The paper by Cape et al. provided suggestions for prior distributions along with their respective parameters. Details regarding the parameters of the posterior distributions can be found in Section 4.3.

Parameter	Prior Distribution	Posterior Distribution
μ	$\mathcal{N}(0,1)$	$\mathcal{N}(\mu^*, \sigma^{*2})$
κ	$\mathcal{N}(0,1)$	$\mathcal{N}(\kappa^*, \sigma_\kappa^{*2})$
θ	$\mathcal{N}(0,1)$	$\mathcal{N}(\theta^*, \sigma_\theta^{*2})$
ψ	$\mathcal{N}\left(0, \frac{\Omega}{2}\right)$, where $\psi_0 = 0$ and $p_0 = 2$	$\psi\vert_\Omega \sim \mathcal{N}(\psi^*, \sigma_\psi^{*2})$
Ω	$\text{IG}(2, \frac{1}{200})$	$\mathcal{IG}(\alpha_*, \beta_*)$
$V_t^{(0)}$	$\mathcal{N}(0.0225, 0.005)1_{V_t>0}$	Provided in 4.3

Table 1: Prior and Posterior Distributions for Parameters μ, κ, θ, ψ and Ω,

In the Gibbs Sampling algorithm we assign initial values to μ, κ, θ, ψ and Ω. The values of the parameters are chosen arbitrarily to serve as a starting point for the algorithm and are initialised as follows:

$$\mu^{(0)} = 0.1,$$
$$\kappa^{(0)} = 0.2,$$
$$\theta^{(0)} = 0.2,$$
$$\psi^{(0)} = 0.1,$$
$$\Omega^{(0)} = 0.005.$$

Note that V_t is assumed to follow a truncated normal distribution, as the volatility must remain positive. Thus, the initial value for V_t is drawn from this distribution. This approach ensures that the starting condition for V_t is in alignment with our fundamental assumptions about the behavior of volatility in financial contexts.

Parameter Calibration Results

We conduct MCMC sampling on the posterior of the 5 paramters and the volatilities: Gibbs for the parameters and Metropolis-Hasting for the volatities, for 10,000 iterations. At implementation, we iteratively update the matrix that concatenate the initial values of the desired parameters and volatility terms. For each sample of each parameter, we calculate a posterior mean, which allows us to produce a trace plot of the posterior mean values over iterations. We can thereby diagnose whether the samples had converged to a stationary distribution. The result is in Figure 4.

- The trace plot of μ shows it wiggling around a steady average with a uniform spread, meaning that its average value might have settled.
- The trace plots of θ and κ shows a positive process, which align with what we defined. The trace plot exhibit sharp spikes, meaning the processes visit their distribution tails, then come back the realisations at high density.
- The trace plot of ψ varies within a soft range, showing some ups and downs, yet it seems to hover around an average value, pointing towards a possible settling.
- The trace plot of Ω shows a consistent spread around the value of the 0.02, indicating that the average value of Ω is stable.

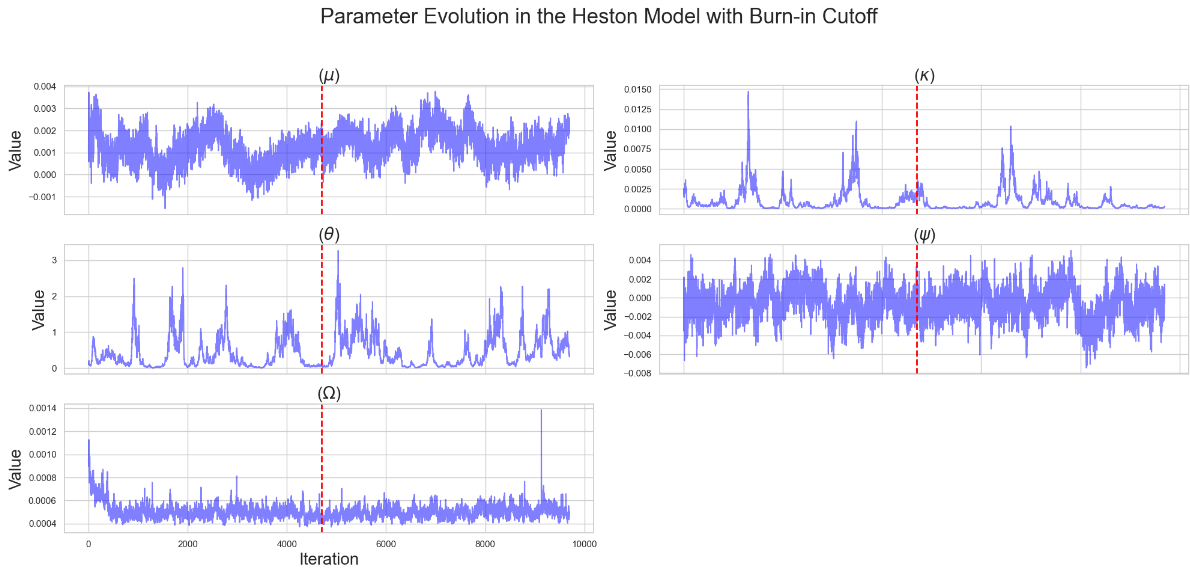

Figure 4: Trace plots for the parameters μ, θ, κ, ψ and Ω in the Heston model. Gibbs sampling was used to estimate these parameters, with the red dashed line representing the burn-in cutoff at iteration = 5,000.

Parameter	Mean	Std Dev
μ	0.001465	0.000668
κ	0.000786	0.001153
θ	0.495983	0.462522
ψ	-0.029971	0.241236
Ω	0.022337	0.007243

Table 2: Final Mean and Standard Deviation of Parameters Posterior Distribution: standard deviation of θ and κ and Ω are large relative to the mean, meaning the posterior did not converge well.

Simulation of Future Paths with Calibrated Parameters

Once we obtain the distribution of calibrated parameters, we use them to simulate future prices, as seen in Figure 5.

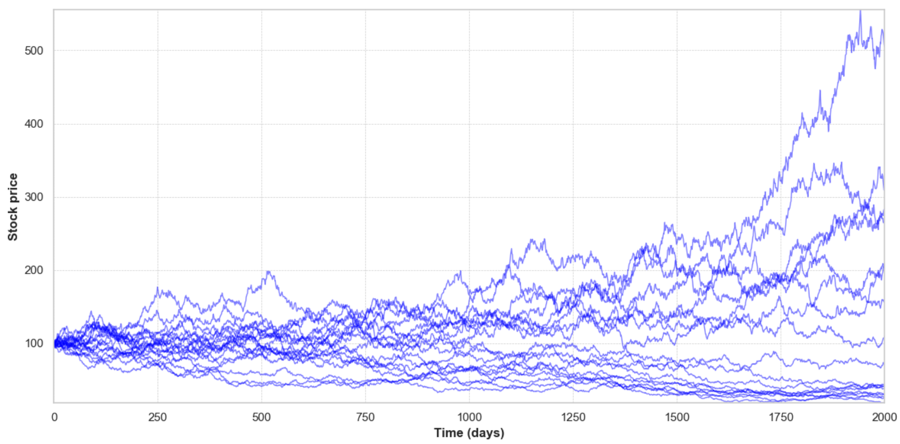

Figure 5: Using calibrated Heston's model, we simulate 18 possible paths of future prices of Google stock in the next 2,000 days.The stock prices shown are volatile with a general upward trend over time, showing significant growth particularly after day 1500. The spread between each paths appear to widen as time progresses, which is natural because we become more uncertain as we move further from historical data.

Note that from the simulation above, we used the posterior mean for each parameters. However, as we have the whole distribution over every parameter thanks to Bayesian inference, we can actually choose the estimates that are the most probable from the posterior. This is beyond the scope of this current work and can be explored in the future.

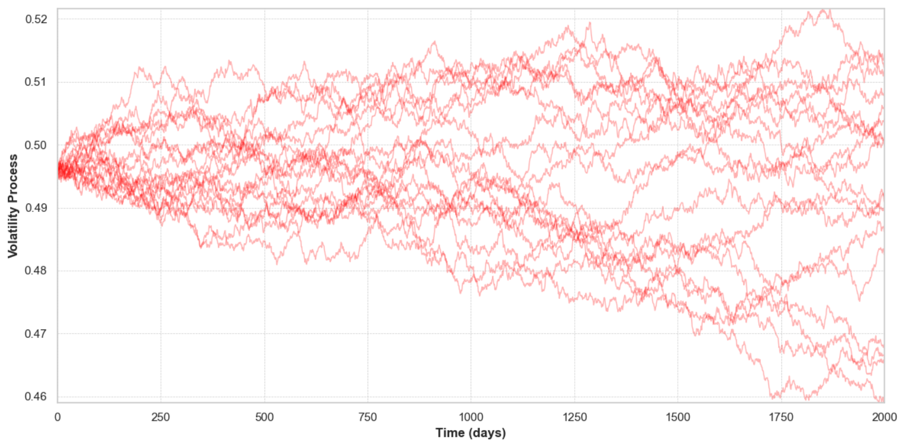

Figure 6: The corresponding volatility paths that were substituted into the price paths in Figure 5. The volatility process demonstrates a slight downward trend over the observed period, with individual paths converging slightly, indicating a small decrease in volatility estimates over time.

6 Conclusion and Future Work

In conclusion, we have explored the theory behind Wiener process and Heston's model. We understood that we can calibrate the parameters of these models according to the historical data of asset prices. However, this is at risk of overfitting in Heston's model, and we need to do Bayesian inference to have a distribution over each parameters. This can be effectively done with MCMC so that we can avoid calculating the marginal likelihood.

There are two main developments that we can explore in the future to improve MCMC parameter inference in Heston's model:

- **Prior setting**: We currently initialise prior parameters of the model's parameters using default setting for stock pricing in Cape et al.. These can be hand-picked specifically for each stock, which requires some knowledge in financial markets. Good initialisation will help with MCMC convergence and sampling efficiency.

- **Numerical stability management**: We currently did not incorporate any numerical instability management in place, which often leads to numerical overflow while calibrating the parameters. Implementing numerical techniques such as log transformations of the parameters, adaptive step sizes in the MCMC algorithm, or employing variance reduction techniques could mitigate these issues.

In addition, in the paper [1], there is also an extension to Heston's model called Bates' model, where it adds a 'jump' term that follows a Poisson distribution to the price SDE. This helps to account for major jumps in stock prices, which is also a frequently event property in finance.

References

[1] Joshua Cape, William Dearden, William Gamber, Jeffrey Liebner, Qin Lu, and M. Linh Nguyen. Estimating heston's and bates' models parameters using markov chain monte carlo simulation. pages 2295–2314, 2014. doi:http://dx.doi.org/10.1080/00949655.2014.926899. Published online: 11 Jun 2014.

[2] CFI. Adjusted closing price - overview, importance, example. https://corporatefinanceinstitute.com/resources/equities/adjusted-closing-price/, 2024. Accessed: [24th February 2024].

[3] Anna Hagenblad, Lennart Ljung, and Adrian Wills. Maximum likelihood identification of wiener models. *IFAC Proceedings Volumes*, 41(2):2714–2719, 2008. ISSN 1474-6670. doi:10.3182/20080706-5-KR-1001.00457. URL https://www.sciencedirect.com/science/article/pii/S1474667016393624.

[4] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 5:327–343, 1993.

[5] Kiyoshi Itô. On a formula concerning stochastic differentials. *Nagoya Math. J.*, 3:55–65, 1951.

[6] E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4(4):727–752, 1991. URL https://www.jstor.org/stable/2962156.

A Appendix

Derivation of equation (1)

To derive dX_t, we can start with the definition of X_t given by:

$$X_t = \mu t + \sigma W_t$$

Then, consider an infinitesimally small increment dt. The change in X_t over this increment is dX_t:

$$dX_t = X_{t+dt} - X_t$$

Substituting the definition of X_t into this, we get:

$$dX_t = (\mu(t + dt) + \sigma W_{t+dt}) - (\mu t + \sigma W_t)$$

Expanding this and simplifying gives us:

$$dX_t = \mu dt + \sigma(W_{t+dt} - W_t)$$

Recognising that $W_{t+dt} - W_t$ is the increment of the Wiener Process over dt, which is dW_t, we can write:

$$dX_t = \mu dt + \sigma dW_t$$

This is the derived form of the stochastic differential equation for X_t, which includes both the drift and the stochastic components.

Derivation of equation (3), using Ito's lemma

We aim to derive the stochastic differential equation (SDE) for the variance V_t from the given SDE of its square root. The starting point is the following SDE for $\sqrt{V_t}$, which follows an Ornstein-Uhlenbeck process (by Stein and Stein, 1991[6]):

$$d\sqrt{V_t} = -\beta\sqrt{V_t}\,dt + \delta\,dz_2(t)$$

Note that the Ornstein-Uhlenbeck (O-U) process is a type of Itô process with $X_t = \sqrt{V_t}$, $\mu(X_t, t) = -\beta X_t$, and $\sigma(X_t, t) = \delta$. Here, β and δ are constant parameters, dt represents a small increment in time, and W_t^V denotes a Wiener process.
We want to find the SDE for V_t, and to do so, we will use Ito's Lemma, which provides a way to compute the differential of a function of a stochastic process. Define $f(u) = u^2$, so $f(X_t) = X_t^2 = f(\sqrt{V_t}) = V_t$. Ito's Lemma requires the first and second derivatives of f with respect to x:

$$\frac{\partial f}{\partial X_t} = 2X_t, \quad \frac{\partial^2 f}{\partial X_t^2} = 2, \quad \frac{\partial f}{\partial t} = 0.$$

Substituting the derivatives into the equation given by Ito's Lemma:

$$\begin{aligned} df(X_t) &= \left(-\beta X_t \cdot 2X_t + \frac{1}{2}\delta^2 \cdot 2\right) dt + \delta \cdot 2X_t\, dW_t^V \\ &= \left(-2\beta X_t^2 + \delta^2\right) dt + 2\delta X_t\, dW_t^V \\ &= \left(-2\beta V_t + \delta^2\right) dt + 2\delta\sqrt{V_t}\, dW_t^V \end{aligned}$$

Simplifying and noting that $df(\sqrt{V_t}) = dV_t$, we obtain the SDE for the variance:

$$dV_t = (-2\beta V_t + \delta^2)\,dt + 2\delta\sqrt{V_t}\,dW_t^V$$

This is the final form of the SDE for the variance of the asset price. It shows that the variance follows a mean-reverting square-root process with drift and diffusion components.
Notice that the equation model the variance of the asset with a mean-reverting term and a stochastic term that depends on the square root of the variance. The parameters β and δ in the first equation can be related to κ, θ, and σ in the square root process proposed by Cox, Ingersoll and Ross (1985) (3).

Derivation of equation (5), using Ito's lemma

Applying Ito's Lemma to transform dS_t to $d(\ln S_t)$ Given the stochastic differential equation for S_t (2):

$$S_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^S$$

We apply Ito's Lemma to find the equation for a function $f(S_t, t)$, given by:

$$df(S_t, t) = \left(\frac{\partial f}{\partial t} + \mu S_t \frac{\partial f}{\partial S_t} + \frac{1}{2} V S_t^2 \frac{\partial^2 f}{\partial S_t^2} \right) dt + \sqrt{V} S_t \frac{\partial f}{\partial S_t} dW_t^S$$

For $f(S_t) = \ln(S_t)$, the partial derivatives are:

$$\frac{\partial f}{\partial S_t} = \frac{1}{S_t}, \quad \frac{\partial^2 f}{\partial S_t^2} = -\frac{1}{S_t^2}, \quad \text{and} \quad \frac{\partial f}{\partial t} = 0$$

Thus, applying these to Ito's Lemma gives us:

$$d(\ln S_t) = \left(\mu - \frac{1}{2} V_t \right) dt + \sqrt{V_t} dW_t^S$$

Which is the transformed equation under the logarithm.