

Article

Parameter Estimation of the Heston Volatility Model with Jumps in the Asset Prices

Jarosław Gruszka *  and Janusz Szwabiński 

Hugo Steinhaus Center, Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland

* Correspondence: jaroslaw.gruszka@pwr.edu.pl

Abstract: The parametric estimation of stochastic differential equations (SDEs) has been the subject of intense studies already for several decades. The Heston model, for instance, is based on two coupled SDEs and is often used in financial mathematics for the dynamics of asset prices and their volatility. Calibrating it to real data would be very useful in many practical scenarios. It is very challenging, however, since the volatility is not directly observable. In this paper, a complete estimation procedure of the Heston model without and with jumps in the asset prices is presented. Bayesian regression combined with the particle filtering method is used as the estimation framework. Within the framework, we propose a novel approach to handle jumps in order to neutralise their negative impact on the estimates of the key parameters of the model. An improvement in the sampling in the particle filtering method is discussed as well. Our analysis is supported by numerical simulations of the Heston model to investigate the performance of the estimators. In addition, a practical follow-along recipe is given to allow finding adequate estimates from any given data.

Keywords: Heston model; estimation; Bayesian inference; particle filtering; Monte Carlo Markov chains



Citation: Gruszka, Jarosław, and Janusz Szwabiński. 2023. Parameter Estimation of the Heston Volatility Model with Jumps in the Asset Prices. *Econometrics* 11: 15. <https://doi.org/10.3390/econometrics11020015>

Academic Editor: Fredj Jawadi

Received: 10 December 2022

Revised: 11 May 2023

Accepted: 29 May 2023

Published: 2 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The problem of the parameter estimation of mathematical models applied in the fields of economy and finance is of critical importance. In order to use most of the models, such as the ones for pricing financial instruments or finding an optimal investment portfolio, one needs to provide values for the model parameters, which are often not easily available. For example, a famous Nobel-prize winning Black–Scholes model for pricing European options [Black and Scholes \(1973\)](#) assumes that the dynamics of the underlying asset is what we now call the Geometric Brownian Motion (GBM)—a stochastic process, which has two parameters commonly called the drift and the volatility. Knowing the values of those parameters for a particular underlying instrument is required to make use of the model, as they need to be plugged into the formulas the model provides.

Over the last decades, mathematical models describing the behaviour of observed market quantities (e.g., prices of assets, interest rates, etc.) have become more complicated to be able to reflect some particular characteristics of their dynamics. For instance, the phenomenon called the volatility smile is widely observed across various types of options and in different markets; however, it is not possible to “configure” the classical Black–Scholes model to reproduce it [Meissner and Kawano \(2001\)](#). Similarly, financial markets occasionally experience sudden drops in the value of assets traded, which can be treated as discontinuities in their trajectories; yet, the GBM, as a model having time-continuous sample paths, would never display any kind of a jump in the value of the modelled asset. Therefore, a need for more complex models emerges, such as the ones of [Heston \(1993\)](#) and [Bates \(1996\)](#), which were designed to address those two specific issues, respectively. The problem is that more complicated models typically use more parameters, which need to be estimated; moreover, standard estimation techniques, such as the Maximum Likelihood

Estimators (MLE) or the Generalised Method of Moments (GMM), fail very often for them [Johannes and Polson \(2010\)](#). Apart from that, most existing methods for estimating the parameters of more complex financial models, such as the ones of Heston or Bates work in the context of derivative instruments only, and as such, they require options prices as an input, despite the fact that the models themselves actually describe the dynamics of the underlying instruments. This presents two major problems. The first one is that the mentioned models are not always used in the context of derivative instruments, as sometimes we are only interested in modelling stock price dynamics (e.g., in research related to stock portfolio management). The other problem is that the historical values of basic instruments such as indices, stocks, or commodities are much more easily found publicly on the Internet, compared to the options prices. Thus, from the data availability perspective, the estimation tools based on the values of underlying instrument outcompete the ones that require prices of derivatives as an input.

There is a wide range of methods that use the prices of the actual instruments, instead of the prices of the derivatives for parameter estimation, but the Bayesian approach [Lindley and Smith \(1972\)](#) seems to be especially effective in that field. Among the methods based on Bayesian inference, the ones using Monte Carlo Markov chains (MCMC) are the most prominent for complex financial models. In this group of methods, one assumes some distribution for the value of each of the parameters of a model (called the priordistribution) and uses it, along with the data, to produce what is called the posteriordistribution, samples that we can treat as possible values of our parameters (see [Johannes and Polson \(2010\)](#) for a great overview of the MCMC methods used for financial mathematics).

The MCMC concept can be applied in multiple ways and by utilising various different algorithms, including Gibbs sampling or the Metropolis–Hastings algorithm [Chib and Greenberg \(1995\)](#), depending on the complexity of the problem. Both are generally very useful for the effective estimation of “single” parameters, i.e., those parameters that only have one constant number as their value. However, some models assume that the directly observable dynamic quantities (e.g., prices) are dependent on other dynamically changing properties of the model. The latter are often called latent variables or state variables. In case of the Heston model, for example, the volatility process is a state variable. Estimation of the state variables is inherently more complicated than that of the regular parameters, as each value, which is observed directly, is partly determined by the value of the state variable at that particular point in time. A very elegant solution to this complication is a methodology called particle filtering. It is based on the idea of creating a collection of values (called particles), which are meant to represent the distribution of the latent variable at a given point in time. Each particle then has a probability assigned to it, which serves as a measure of how likely it is that a given value of the state variable generated the outcome observed at a given moment of time. For an overview of particle filtering methods, we recommend [Johannes et al. \(2009\)](#) and [Doucet and Johansen \(2009\)](#).

The methods outlined above have been studied quite thoroughly in recent years. However, the research articles and the literature focused on the theoretical aspect of the estimation process and often lacked precision and concreteness. This is, in fact, a serious issue, since applying the results of theoretical research in practice almost always requires estimation in one way or another. In our last paper, we studied the performance of various investment portfolios depending on the assets they contained, which were represented by trajectories of the Heston model [Gruszka and Szwabiński \(2021\)](#). The behaviour of those portfolios turned out to be dependent on some of the assets’ characteristics, which are captured by the values of certain parameters of the Heston model. Hence, an estimation scheme would allow us to determine whether a given strategy is suitable for a particular asset portfolio. This is just one example of how the estimation of a financial market model can be utilised.

In this paper, we present a complete setup for parameter estimation of the Heston model, using only the prices of the basic instrument one wants to study using the model (an index, a stock, a commodity, etc., no derivative prices needed). We provide the estimation

process for both the pure Heston model and its extended version, with the inclusion of Merton-style jumps (discontinuities), which is then known as the Bates model. In Section 2, we present the Heston model as well as its extension allowing for the appearance of jumps. We also present a way of changing the time-character of the model from a continuous in time one to a discrete one. In Section 3, we describe in detail the posterior distributions from which one can sample to obtain the parameters of the model. We also provide a detailed description of the particle filtering scheme needed to reconstruct the volatility process. The whole procedure is summarised in an easy-to-follow pseudo-code algorithm. An exemplary estimation, as well as the analysis of the factors that impact the quality of the estimation in general, is presented in Section 4. Finally, some conclusions are drawn in the last section.

2. Heston Model Without and with Jumps

2.1. Model Characterisation

The Heston model can be described using two stochastic differential equations, one for the process of prices and one for the process of volatility [Heston \(1993\)](#)

$$dS(t) = \mu S(t)dt + \sqrt{v(t)}S(t)dB^S(t), \quad (1)$$

$$dv(t) = \kappa(\theta - v(t))dt + \sigma\sqrt{v(t)}dB^v(t), \quad (2)$$

where $t \in [0, T]$. In Equation (1), the parameter μ represents the drift of the stock price. Equation (2) is widely known as the CIR (Cox–Ingersoll–Ross) model, featuring an interesting quality called mean reversion [Cox et al. \(1985\)](#). Parameter θ is the long-term average from which the volatility diverges and to which it then returns, and κ is the rate of those fluctuations (the larger the κ , the longer it takes to return to θ). Parameter σ is called the volatility of the volatility, and it is generally responsible for the “scale” of randomness of the volatility process.

Both stochastic processes are based on their respective Brownian motions— $B^S(t)$ and $B^v(t)$. The Heston model allows for the possibility of those two processes being correlated with an instantaneous correlation coefficient ρ ,

$$dB^S(t)dB^v(t) = \rho dt. \quad (3)$$

To complete the setup, deterministic initial conditions for S and v need to be specified.

$$S(0) = S_0 > 0, \quad (4)$$

$$v(0) = v_0 > 0. \quad (5)$$

It is worth highlighting that the description of the Heston model outlined above is expressed via the physical probability measure, often denoted by \mathbb{P} , which should not be confused with the risk-neutral measure (also called the martingale measure), often denoted by \mathbb{Q} [Wong and Heyde \(2006\)](#). Using the risk-neutral version is especially important when the model is used for pricing derivative instruments, as the goal is to make the discounted process of prices a martingale and hence eliminate arbitrage opportunities. Versions of the same model under those two measures usually differ in regard to the parameters that they feature. The classical example is the Geometric Brownian Motion, mentioned in the introduction, which has two parameters, the drift and the volatility. During the procedure of changing the measure, the drift variable is replaced by the risk-free interest rate, and the volatility parameter remains in place unchanged. In case of the Heston model, the interdependence between the parameters of the models under the \mathbb{P} and \mathbb{Q} measures are more subtle and depend on additional assumptions made during the measure-changing procedure itself; however, in most cases, there are explicit formulas to calculate the parameter values under \mathbb{Q} having them under \mathbb{P} and vice versa. However, the transformations often require some additional inputs, related to the particular derivative

instrument being priced (e.g., its market price of risk). Throughout this work, we are only interested in the values of the model parameters under the physical probability measure \mathbb{P} . For more in-depth analysis of the change in the measure problem in stochastic volatility models, we recommend [Wong and Heyde \(2006\)](#).

The trajectories coming from the Heston model are continuous, although the model itself can easily be extended to include discontinuities. The most common type of jump, which can easily be incorporated into the model, is called the Merton log-normal jump. To add it, one needs to augment Equation (1) with an additional term,

$$dS(t) = \mu S(t)dt + \sqrt{v(t)}S(t)dB^S(t) + (e^{Z(t)} - 1)S(t)dq(t), \tag{6}$$

where $Z(t)$ is a series of independent and identically distributed normally distributed random variables with mean μ^J and standard deviation σ^J , whereas $q(t)$ is a Poisson counting process with constant intensity λ . The added term turns the Heston model into the Bates model [Bates \(1996\)](#). The above extension has an easy real-life interpretation. Namely, $e^{Z(t)}$ is the actual (absolute) rate of the difference between the price before the jump at time t and right after it, i.e., $S(t-) \cdot e^{Z(t)} = S(t+)$. So if, for example, for a given t , $e^{Z(t)} \approx 0.85$, that means the stock experienced $\sim 15\%$ drop in value at that moment.

2.2. Euler–Maruyama Discretisation

In order to make the model applicable in practice, one needs to discretise it, that is, to rewrite the continuous (theoretical) equations in such a way that the values of the process are given in specific equidistant points of time. To this end, we split the domain $[0, T]$, into n short intervals, each of length Δt . Thus, $n \cdot \Delta t = T$. To properly transform the SDEs of the model into this new time domain, a discretisation scheme is necessary. We use the Euler–Maruyama discretisation scheme for that purpose [Kloeden and Platen \(1992\)](#). The stock price Equation (1) can be discretised as

$$S(k\Delta t) - S((k-1)\Delta t) = \mu S((k-1)\Delta t)\Delta t + S((k-1)\Delta t)\sqrt{v((k-1)\Delta t)}\varepsilon^S(k\Delta t)\sqrt{\Delta t}, \tag{7}$$

where $k \in \{1, \dots, n\}$, and ε^S is a series of n independent and identically distributed standard normal random variables.

To highlight the ratio between two consecutive values of the stock price, Equation (7) is often rewritten as

$$\frac{S(k\Delta t)}{S((k-1)\Delta t)} = \mu\Delta t + 1 + \sqrt{v((k-1)\Delta t)}\varepsilon^S(k\Delta t)\sqrt{\Delta t}. \tag{8}$$

The same discretisation scheme can be applied to Equation (2) to obtain:

$$v(k\Delta t) - v((k-1)\Delta t) = \kappa \left(\theta - v((k-1)\Delta t) \right) \Delta t + \sigma \sqrt{v((k-1)\Delta t)}\varepsilon^v(k\Delta t)\sqrt{\Delta t}. \tag{9}$$

If $\rho = 0$, then ε^v in the above formula is also a series of n independent and identically distributed standard normal random variables. However, if $\rho \neq 0$, then to ensure the proper dependency between S and v , we take

$$\varepsilon^v(k\Delta t) = \rho\varepsilon^S(k\Delta t) + \sqrt{1 - \rho^2}\varepsilon^{add}(k\Delta t), \tag{10}$$

where ε^{add} is an additional series of n independent and identically distributed standard normal random variables, which are “mixed” with the ones from ε^S and, hence, become dependent on them.

3. Estimation Framework

The estimation of the Heston model consists of two major parts. The first part is estimating the parameters of the model, i.e., $\mu, \kappa, \theta, \sigma,$ and $\rho,$ for the basic version of the model and, additionally, $\lambda, \mu^J,$ and σ^J after the inclusion of jumps. The second part is estimating the state variable, volatility $v(t).$ For all the estimation procedures we present here, we use the Bayesian inference methodology, in particular, Monte Carlo Markov chains (for parameter estimation within the base model) and particle filtering (for estimation of the volatility as well as the jump-related parameters).

3.1. Regular Heston Model

In order to estimate the Heston model with no jumps, we mainly use the principles of Bayesian inference, in particular, Bayesian regression [O’Hagan and Kendall \(1994\)](#).

3.1.1. Estimation of μ

We start by finding a way to estimate the drift parameter $\mu.$ First, Equation (8) is transformed to a regression form. To this end, we introduce several additional variables. The first, $\eta,$ is defined as

$$\eta = \mu\Delta t + 1. \tag{11}$$

Let $R(t)$ be a series of ratios between consecutive prices of assets,

$$R(k\Delta t) = \frac{S(k\Delta t)}{S((k-1)\Delta t)}, \tag{12}$$

for $k \in \{1, 2, \dots, n\}.$ Taking the above definitions into consideration, Equation (8) can be rewritten as:

$$R(k\Delta t) = \eta + \sqrt{v((k-1)\Delta t)} \varepsilon^S(k\Delta t) \sqrt{\Delta t}. \tag{13}$$

Now, let us divide both sides of this equation by $\sqrt{v((k-1)\Delta t)} \sqrt{\Delta t},$ as Δt is known, and at this stage, we consider $v(t)$ to be known too. Let us now introduce another two new variables, $y^S(t)$ as

$$y^S(k\Delta t) = \frac{1}{\sqrt{v((k-1)\Delta t)} \sqrt{\Delta t}} R(k\Delta t) \tag{14}$$

and $x^S(t)$ as

$$x^S(k\Delta t) = \frac{1}{\sqrt{v((k-1)\Delta t)} \sqrt{\Delta t}}. \tag{15}$$

Inserting them into Equation (13) gives

$$y^S(k\Delta t) = \eta x^S(k\Delta t) + \varepsilon^S(k\Delta t). \tag{16}$$

The last expression has the form of a linear regression with $y^S(t)$ explained by $x^S(t).$ We want to treat it with the Bayesian regression framework. To this end, we first collect all the discretised values of $y^S(t)$ and $x^S(t)$ into n -element column vectors— \mathbf{y}^S and $\mathbf{x}^S,$ respectively,

$$\mathbf{y}^S = \frac{1}{\sqrt{\Delta t}} \begin{bmatrix} \frac{R(\Delta t)}{\sqrt{v(0)}} & \frac{R(2\Delta t)}{\sqrt{v(\Delta t)}} & \dots & \frac{R(n\Delta t)}{\sqrt{v((n-1)\Delta t)}} \end{bmatrix}', \tag{17}$$

$$\mathbf{x}^S = \frac{1}{\sqrt{\Delta t}} \left[\frac{1}{\sqrt{v(0)}} \quad \frac{1}{\sqrt{v(\Delta t)}} \quad \cdots \quad \frac{1}{\sqrt{v((n-1)\Delta t)}} \right]', \tag{18}$$

where the prime symbol is used for the transpose.

Assuming a prior distribution for η to be normal with mean μ_0^η and standard deviation σ_0^η , it follows from the Bayesian regression general results [O’Hagan and Kendall \(1994\)](#) that the posterior distribution for η is also normal with precision (inverse of variance) τ^η , which can be calculated as

$$\tau^\eta = (\mathbf{x}^S)' \cdot \mathbf{x}^S + \tau_0^\eta. \tag{19}$$

Here, τ_0^η is the precision of the prior distribution, i.e., $\tau_0^\eta = \frac{1}{(\sigma_0^\eta)^2}$. The mean μ^η of the posterior distribution is of the following form

$$\mu^\eta = \frac{1}{\tau^\eta} \left(\tau_0^\eta \mu_0^\eta + (\mathbf{x}^S)' \cdot \mathbf{x}^S \hat{\eta} \right), \tag{20}$$

where $\hat{\eta}$ is a classical ordinary-least-square (OLS) estimator of η , i.e.,

$$\hat{\eta} = \left((\mathbf{x}^S)' \cdot \mathbf{x}^S \right)^{-1} (\mathbf{x}^S)' \mathbf{y}^S. \tag{21}$$

Hence, we can sample the realisations of η as follows:

$$\eta_i \sim \mathcal{N} \left(\mu^\eta, \frac{1}{\tau^\eta} \right), \tag{22}$$

where i indicates the i -th sample from the posterior distribution, which has been found for η . Having a realisation of η in form of η_i , we can quickly turn it into a realisation of the μ parameter itself by a simple transform, inverse to Equation (11)

$$\mu_i = \frac{\eta_i - 1}{\Delta t}. \tag{23}$$

3.1.2. Estimation of κ , θ , and σ

In order to estimate the parameters related to the volatility process, i.e., κ , θ , and σ , we conduct a similar exercise but this time using the volatility process. Let us first rewrite Equation (9) as

$$v(k\Delta t) = \kappa\theta\Delta t + (1 - \kappa\Delta t)v((k-1)\Delta t) + \sigma \sqrt{v((k-1)\Delta t)} \varepsilon^v(k\Delta t) \sqrt{\Delta t}. \tag{24}$$

Now, let us introduce two new parameters,

$$\beta_1 = \kappa\theta\Delta t \tag{25}$$

and

$$\beta_2 = 1 - \kappa\Delta t. \tag{26}$$

From Equations (24)–(26), we obtain

$$v(k\Delta t) = \beta_1 + \beta_2 v((k-1)\Delta t) + \sigma \sqrt{v((k-1)\Delta t)} \varepsilon^v(k\Delta t) \sqrt{\Delta t}. \tag{27}$$

In a fashion similar to the equation for the stock price, we can rewrite this last expression as

$$\frac{v(k\Delta t)}{\sqrt{\Delta t}\sqrt{v((k-1)\Delta t)}} = \frac{\beta_1}{\sqrt{\Delta t}\sqrt{v((k-1)\Delta t)}} + \frac{\beta_2 v((k-1)\Delta t)}{\sqrt{\Delta t}\sqrt{v((k-1)\Delta t)}} + \sigma \varepsilon^v(k\Delta t). \tag{28}$$

Introducing the following vectors,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \tag{29}$$

$$\mathbf{y}^v = \frac{1}{\sqrt{\Delta t}} \begin{bmatrix} \frac{v(2\Delta t)}{\sqrt{v(\Delta t)}} & \frac{v(3\Delta t)}{\sqrt{v(2\Delta t)}} & \dots & \frac{v(n\Delta t)}{\sqrt{v((n-1)\Delta t)}} \end{bmatrix}', \tag{30}$$

$$\mathbf{x}_1^v = \frac{1}{\sqrt{\Delta t}} \begin{bmatrix} \frac{1}{\sqrt{v(\Delta t)}} & \frac{1}{\sqrt{v(2\Delta t)}} & \dots & \frac{1}{\sqrt{v((n-1)\Delta t)}} \end{bmatrix}', \tag{31}$$

$$\mathbf{x}_2^v = \frac{1}{\sqrt{\Delta t}} \begin{bmatrix} \frac{v(\Delta t)}{\sqrt{v(\Delta t)}} & \frac{v(2\Delta t)}{\sqrt{v(2\Delta t)}} & \dots & \frac{v((n-1)\Delta t)}{\sqrt{v((n-1)\Delta t)}} \end{bmatrix}' = \frac{1}{\sqrt{\Delta t}} \begin{bmatrix} \sqrt{v(\Delta t)} & \sqrt{v(2\Delta t)} & \dots & \sqrt{v((n-1)\Delta t)} \end{bmatrix}', \tag{32}$$

allows us to rewrite the original volatility equation in form of a linear regression

$$\mathbf{y}^v = \mathbf{X}^v \boldsymbol{\beta} + \sigma \boldsymbol{\varepsilon}^v, \tag{33}$$

where

$$\mathbf{X}^v = [\mathbf{x}_1^v \quad \mathbf{x}_2^v], \tag{34}$$

and

$$\boldsymbol{\varepsilon}^v = [\varepsilon^v(\Delta t) \quad \varepsilon^v(2\Delta t) \quad \dots \quad \varepsilon^v((n-1)\Delta t)]. \tag{35}$$

Using the formulas for Bayesian regression and assuming a multivariate (two-dimensional) normal prior for $\boldsymbol{\beta}$ with a mean vector $\boldsymbol{\mu}_0^\beta$ and a precision matrix $\boldsymbol{\Lambda}_0^\beta$, we obtain the conjugate posterior distribution that is also multivariate normal with a precision matrix given by

$$\boldsymbol{\Lambda}^\beta = (\mathbf{X}^v)' \cdot \mathbf{X}^v + \boldsymbol{\Lambda}_0^\beta \tag{36}$$

and mean vector given by

$$\boldsymbol{\mu}^\beta = (\boldsymbol{\Lambda}^\beta)^{-1} (\boldsymbol{\Lambda}_0^\beta \boldsymbol{\mu}_0^\beta + (\mathbf{X}^v)' \cdot \mathbf{X}^v \hat{\boldsymbol{\beta}}), \tag{37}$$

where again $\hat{\boldsymbol{\beta}}$ is a standard OLS estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = ((\mathbf{X}^v)' \cdot \mathbf{X}^v)^{-1} (\mathbf{X}^v)' \mathbf{y}^v. \tag{38}$$

We can then use this posterior distribution of $\boldsymbol{\beta}$ for sampling

$$\beta_i \sim \mathcal{N}(\boldsymbol{\mu}^\beta, \sigma_{i-1}^2 (\boldsymbol{\Lambda}^\beta)^{-1}). \tag{39}$$

It is worth noting that the realisation of σ appears in Equation (39); however, we have not defined it yet. This is because the distribution of $\boldsymbol{\beta}$ is dependent on σ , and the distribution of σ is dependent on $\boldsymbol{\beta}$. Hence, we suggest taking the realisation of σ from the

previous iteration here (which is indicated by the $i - 1$ subscript). We address the order of performing calculations in more detail later in this article.

Obtaining realisations of the actual parameters is very easy; one simply needs to inverse the equations defining β_1 and β_2 :

$$\kappa_i = \frac{1 - \beta_i[2]}{\Delta t}, \tag{40}$$

and

$$\theta_i = \frac{\beta_i[1]}{\kappa_i \Delta t}, \tag{41}$$

where $\beta_i[1]$ and $\beta_i[2]$ are, respectively, the first and the second component of the β_i vector.

The most common approach for estimating σ is assuming the inverse-gamma prior distribution for σ^2 . If the parameters of the prior distribution are a_0^σ and b_0^σ , then the conjugate posterior distribution is also inverse gamma

$$(\sigma_i)^2 \sim \mathcal{IG}(a^\sigma, b^\sigma), \tag{42}$$

where

$$a^\sigma = a_0^\sigma + \frac{n}{2}, \tag{43}$$

and

$$b^\sigma = b_0^\sigma + \frac{1}{2} \left((\mathbf{y}^v)' \cdot \mathbf{y}^v + (\boldsymbol{\mu}_0^\beta)' \boldsymbol{\Lambda}_0^\beta \boldsymbol{\mu}_0^\beta - (\boldsymbol{\mu}^\beta)' \boldsymbol{\Lambda}^\beta \boldsymbol{\mu}^\beta \right). \tag{44}$$

3.1.3. Estimation of ρ

For the estimation of ρ , we follow an approach presented in [Jacquier et al. \(2004\)](#). We first define the residuals for the stock price equation.

$$e_1^\rho(k\Delta t) = \frac{R(k\Delta t) - \mu_i \Delta t - 1}{\sqrt{\Delta t} \sqrt{v((k-1)\Delta t)}}, \tag{45}$$

and for the volatility equation,

$$e_2^\rho(k\Delta t) = \frac{v(k\Delta t) - v((k-1)\Delta t) - \kappa_i \left(\theta_i - v((k-1)\Delta t) \right) \Delta t}{\sqrt{\Delta t} \sqrt{v((k-1)\Delta t)}}. \tag{46}$$

By calculating those residuals, we try to retrieve the error terms from Equations (1) and (2), $\varepsilon^S(t)$ and $\sigma \varepsilon^v(t)$, respectively, as we know they are tied with each other by a relationship given by Equation (10). Taking this fact into consideration, we end up with the following equation

$$e_2^\rho(k\Delta t) = \sigma \rho e_1^\rho(k\Delta t) + \sigma \sqrt{1 - \rho^2} \varepsilon^{add}(k\Delta t). \tag{47}$$

We now introduce two new variables, traditionally called $\psi = \sigma \rho$ and $\omega = \sigma^2(1 - \rho^2)$. It is not difficult to deduce that the relationship between ρ and the newly-introduced variables ψ and ω is

$$\rho = \frac{\psi}{\sqrt{\psi^2 + \omega}}. \tag{48}$$

Then, Equation (47) becomes

$$e_2^\rho(k\Delta t) = \psi e_1^\rho(k\Delta t) + \sqrt{\omega} \varepsilon^{add}(k\Delta t), \tag{49}$$

which is again a linear regression of $e_2^\rho(t)$ on e_1^ρ . Thus, we can use the exact same estimation scheme as in case of the previously described regressions. We first collect the values of e_1^ρ and e_2^ρ in two n -element vectors:

$$\mathbf{e}_1^\rho = [e_1^\rho(\Delta t) \quad e_1^\rho(2\Delta t) \quad \dots \quad e_1^\rho(n\Delta t)]', \tag{50}$$

$$\mathbf{e}_2^\rho = [e_2^\rho(\Delta t) \quad e_2^\rho(2\Delta t) \quad \dots \quad e_2^\rho(n\Delta t)]'. \tag{51}$$

Then we appose both vectors, forming them into an n -by-2 matrix:

$$\mathbf{e}^\rho = [\mathbf{e}_1^\rho \quad \mathbf{e}_2^\rho]. \tag{52}$$

Next, we define a 2-by-2 matrix \mathbf{A}^ρ as

$$\mathbf{A}^\rho = (\mathbf{e}^\rho)' \cdot \mathbf{e}^\rho. \tag{53}$$

If we assume a normal prior for ψ with mean μ_0^ψ and precision τ_0^ψ , the posterior distribution for ψ is also normal with the mean μ^ψ given by

$$\mu^\psi = \frac{\mathbf{A}_{12}^\rho + \mu_0^\psi \tau_0^\psi}{\mathbf{A}_{11}^\rho + \tau_0^\psi} \tag{54}$$

and the precision τ^ψ equal to

$$\tau^\psi = \mathbf{A}_{11}^\rho + \tau_0^\psi, \tag{55}$$

where \mathbf{A}_{11}^ρ , \mathbf{A}_{12}^ρ , and \mathbf{A}_{22}^ρ are the elements of the matrix \mathbf{A}^ρ on positions (1, 1), (1, 2), and (2, 2) respectively.

Assuming the inverse gamma prior with parameters a_0^ω and b_0^ω for ω , the conjugate posterior distribution is also the inverse gamma with parameters

$$a^\omega = a_0^\omega + \frac{n}{2} \tag{56}$$

and

$$b^\omega = b_0^\omega + \frac{1}{2} \left(\mathbf{A}_{22}^\rho - \frac{(\mathbf{A}_{12}^\rho)^2}{\mathbf{A}_{11}^\rho} \right). \tag{57}$$

Thus, sampling from the posterior distribution of ω can be summarised as

$$\omega_i \sim \mathcal{IG}(a^\omega, b^\omega), \tag{58}$$

while when it comes to sampling from ψ it is

$$\psi_i \sim \mathcal{N} \left(\mu^\psi, \frac{\sqrt{\omega_i}}{\sqrt{\tau^\psi}} \right). \tag{59}$$

To obtain ρ , we simply make use of Equation (48).

3.1.4. Estimation of $v(t)$ —Particle Filtering

For all the estimation procedures shown in the previous sections, we assumed $v(t)$ to be known. However, in practice, the volatility is not a directly observable quantity, it is “hidden” in the process of prices, to which we have access. Hence, we need a way to extract the volatility from the price process, and the particle filtering methodology is extremely useful for that purpose. Here, we only sketch the outline of the particle filtering logic, namely the SIR algorithm, which we utilise to obtain the volatility estimator. For a more in-depth review

of particle filtering, we suggest the works of Johannes et al. (2009) and Doucet and Johansen (2009). Here, we follow a procedure similar to the one presented in Christoffersen et al. (2007).

We start by fixing the number of particles N . In each moment of time $t = k\Delta t$, we produce N particles, which represent various possible values of the volatility at that point in time. By averaging out all of those particles, we obtain an estimate of the true volatility $v(t)$. The process of creating the particles is as follows: at the time $t = 0$, we create N initial particles, all with the initial value of the volatility, which we assume to be the long-term average θ . Denoting each of the particles by V_j , for $j \in \{1, 2, \dots, N\}$, we have

$$V_j(0) = \theta_i. \quad (60)$$

For any subsequent moment of time, except the last one $t = k\Delta t, k \notin \{0, n\}$, we define three sequences of size N . ε_j is a series of independent standard normal random variables

$$\varepsilon_j(k\Delta t) \sim \mathcal{N}(0, 1). \quad (61)$$

The series z_j contains residuals from the stock price process, where the past values of volatility are replaced by the values of the particles from the previous time step

$$z_j(k\Delta t) = \frac{R(k\Delta t) - \mu_i \Delta t - 1}{\sqrt{\Delta t} \sqrt{V_j((k-1)\Delta t)}}. \quad (62)$$

Finally, the series w_j , which incorporates the possible dependency between the stock process and the volatility particles, is

$$w_j(k\Delta t) = z_j(k\Delta t) \rho_i + \varepsilon_j(k\Delta t) \sqrt{1 - (\rho_i)^2}. \quad (63)$$

Having all that, the candidates for the new particles \tilde{V}_j are created as follows

$$\tilde{V}_j(k\Delta t) = V_j((k-1)\Delta t) + \kappa_i \left(\theta_i - V_j((k-1)\Delta t) \right) \Delta t + \sigma_i \sqrt{\Delta t} \sqrt{V_j((k-1)\Delta t)} w_j. \quad (64)$$

Each candidate for a particle is evaluated based on how probable it is that such a value of the volatility would generate the return that was actually observed. The measure of this probability \tilde{W}_j is a value of a normal distribution PDF function designed specifically for this purpose¹,

$$\tilde{W}_j(k\Delta t) = \frac{1}{\sqrt{2\pi \tilde{V}_j(k\Delta t) \Delta t}} \exp \left(-\frac{1}{2} \frac{\left(R((k+1)\Delta t) - \mu_i \Delta t - 1 \right)^2}{\tilde{V}_j(k\Delta t) \Delta t} \right). \quad (65)$$

To be able to treat the values of the proposed measure along with the values of particles as a proper probability distribution on its own, we normalise them, so that their sum is equal to 1,

$$\check{W}_j(k\Delta t) = \tilde{W}_j(k\Delta t) \left(\sum_{j=1}^N \tilde{W}_j(k\Delta t) \right)^{-1}. \quad (66)$$

Now, we combine the particles with their respective probabilities, forming two-element vectors \mathbf{U}_j

$$\mathbf{U}_j(k\Delta t) = (\tilde{V}_j(k\Delta t), \check{W}_j(k\Delta t)). \quad (67)$$

We now want to sample from the probability distribution described by U_j to obtain the true “refined” particles. Most sources suggest drawing from it, treating it as a multinomial distribution. However, this makes all the “refined” particles have the same values as the “raw” ones, with just the proportions changed (the same “raw” particle can be drawn several times, if it has a higher probability than the others). To address this problem, we conduct the sampling in a different way. We first need to sort the values of particles in ascending order. Mathematically speaking, we create another sequence and call it \tilde{V}_j^{sort} , ensuring that the following conditions are all met:

1. The particle with the smallest value is the first in the new sequence, i.e.,

$$\tilde{V}_1^{sort}(k\Delta t) = \min_{j \in \{1,2,\dots,N\}} \{ \tilde{V}_j(k\Delta t) \}, \tag{68}$$

2. The particle with the largest value is the last in the new sequence, i.e.,

$$\tilde{V}_N^{sort}(k\Delta t) = \max_{j \in \{1,2,\dots,N\}} \{ \tilde{V}_j(k\Delta t) \}, \tag{69}$$

3. For any $j \in \{2,3,\dots,N-1\}$, we have

$$\tilde{V}_{j-1}^{sort}(k\Delta t) < \tilde{V}_j^{sort}(k\Delta t) < \tilde{V}_{j+1}^{sort}(k\Delta t). \tag{70}$$

We also want to keep track of the probabilities of our sorted particles; so, we order the probabilities in the same way, by defining another probability sequence \check{W}_j^{sort} ,

$$\check{W}_j^{sort}(k\Delta t) = \check{W}_m(k\Delta t) \text{ for } m, \text{ such that } \tilde{V}_j^{sort}(k\Delta t) \in \mathbf{U}_m(k\Delta t). \tag{71}$$

This step is necessary to ensure that each element of the sorted sequence of particle values \tilde{V}_j^{sort} still has an original normalised probability \check{W}_j^{sort} assigned to it. This is why it was necessary to pair up the particles and their normalised probabilities into two-element vectors, under Equation (67). These pairs now help us approximate a continuous distribution, from which we can sample the refined particles. The “extreme” particles, \tilde{V}_1^{sort} and \tilde{V}_N^{sort} , will become the edges of the support of this new continuous distribution. The CDF function is given by the formula below (the time labels have been dropped for the sake of legibility, as all the variables are evaluated at $t = k\Delta t$):

$$F_{\tilde{V}^{sort}}(v) = \begin{cases} 0 & \text{if } v \leq \tilde{V}_1^{sort} \\ \frac{v - \tilde{V}_1^{sort}}{\tilde{V}_2^{sort} - \tilde{V}_1^{sort}} (\check{W}_1^{sort} + \frac{1}{2} \check{W}_2^{sort}) & \text{if } v \in (\tilde{V}_1^{sort}, \tilde{V}_2^{sort}] \\ \left(\sum_{m=1}^{j-1} \check{W}_m^{sort} + \frac{1}{2} \check{W}_j^{sort} \right) + \frac{v - \tilde{V}_j^{sort}}{\tilde{V}_{j+1}^{sort} - \tilde{V}_j^{sort}} \left(\frac{1}{2} \check{W}_j^{sort} + \frac{1}{2} \check{W}_{j+1}^{sort} \right) & \text{if } v \in (\tilde{V}_j^{sort}, \tilde{V}_{j+1}^{sort}] \\ & \text{for } j \in \{2,3,\dots,N-2\} \\ \left(\sum_{m=1}^{N-2} \check{W}_m^{sort} + \frac{1}{2} \check{W}_{N-1}^{sort} \right) + \frac{v - \tilde{V}_{N-1}^{sort}}{\tilde{V}_N^{sort} - \tilde{V}_{N-1}^{sort}} \left(\frac{1}{2} \check{W}_{N-1}^{sort} + \check{W}_N^{sort} \right) & \text{if } v \in (\tilde{V}_{N-1}^{sort}, \tilde{V}_N^{sort}] \\ 1 & \text{if } v > \tilde{V}_N^{sort} \end{cases} . \tag{72}$$

The formula might seem overwhelming, but there is a very easy-to-follow interpretation behind it (see Figure 1). The new “refined” particles can be generated by drawing from

the distribution given by $F_{\tilde{V}_{sort}}$; the simplest way to do so is to use the inverse transform sampling.

$$V_j(k\Delta t) \sim F_{\tilde{V}_{sort}}. \tag{73}$$

After following the described procedure for each $k \in \{1, 2, \dots, n - 1\}$, we can specify the actual estimate of the volatility process as the mean of the “refined” particles.

$$v(k\Delta t) = \frac{1}{N} \sum_{j=1}^N V_j(k\Delta t). \tag{74}$$

For $k = n$, we can simply assume $v(n\Delta t) = v((n - 1)\Delta t)$, which should not have any tangible negative impact on any procedure using the $v(t)$ estimate for a sufficiently dense time discretisation grid.

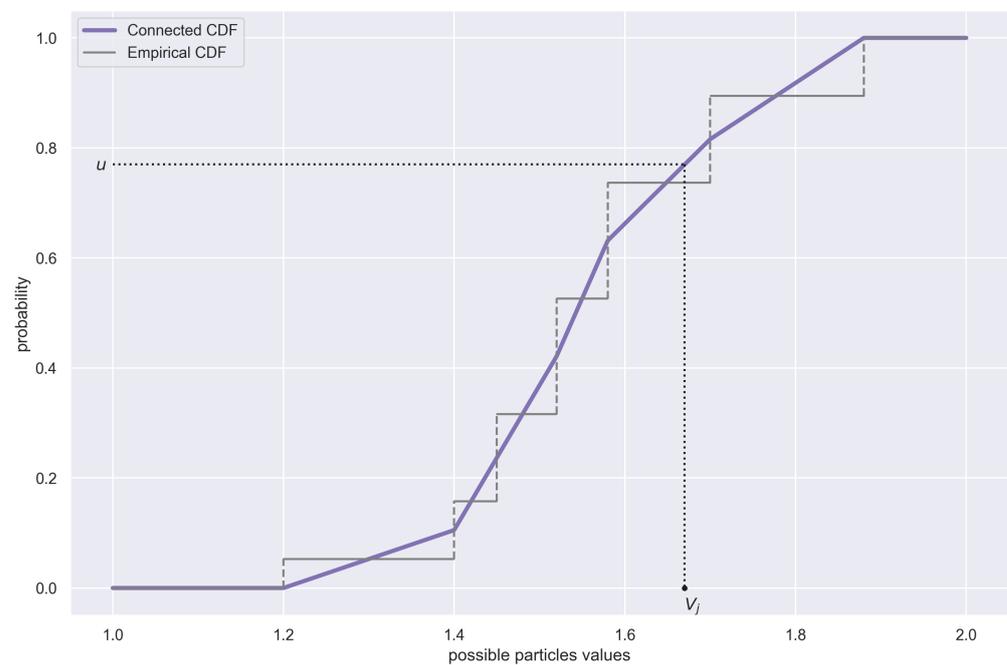


Figure 1. Visualisation of the process of resampling particles according to their probabilities. The values of the raw particles are the places, where the empirical cumulative distribution function (ECDF in short) jumps, and each jump size represents the probability of a respective raw particle. The connected CDF is a continuous modification of the ECDF, built according to Formula (72). In order to resample, a uniform random variable u is generated, and then its inverse through the connected CDF function becomes a new resampled particle, V_j .

3.2. Heston Model with Jumps

The above estimation framework can be used with minor changes to also estimate the Heston model with jumps. The model’s SDE is defined in Equation (6). After the incorporation of jumps, changes are needed particularly in the particle filtering part of the estimation procedure. The particles need to be created not only for various possible values of volatility $V_j(t)$ but also for the possibility of a jump at that particular moment in time, $J_j(t)$, and the size of that jump, $Z_j(t)$. So, one can think of a particle as of a three-element “tuple” (V_j, J_j, Z_j) . Generating “raw” values for J_j and Z_j is easy; for each $j \in \{0, 1, \dots, N\}$, J_j is simply a random variable from a Bernoulli distribution with parameter λ^{th} ,

$$\tilde{J}_j(k\Delta t) \sim \mathcal{B}(\lambda^{th}). \tag{75}$$

Parameter $\lambda^{th} \in [0, 1)$ can be thought of as a “threshold” value—a proportion of the number of particles that encodes the occurrence of a jump to all the particles. If the number of jumps is expected to be significant, it is good to increase the value of λ^{th} , hence increasing the number of particles suggesting the jump in each step. The values of λ^{th} , which we observe to work reasonably well for most datasets, are between 0.1 and 0.35. There is a multitude of ways in which one can assess a rational value of this parameter, and one of the most basic ones is visualising the returns of the asset in question and assessing the number of distinct downward spikes on the plot. The ratio of this number and the length of the sample should be a good indication of the region of the interval $[0.1, 0.35]$ from which λ^{th} should be taken.

The raw particles for Z_j are simply independent normal random variables with mean μ_0^J and standard deviation σ_0^J , which depict our “prior” beliefs about the size and variance of the jumps

$$\tilde{Z}_j(k\Delta t) \sim \mathcal{N}(\mu_0^J, \sigma_0^J). \tag{76}$$

Assigning probabilities to the particles is different as well, since the normal PDF function that we use is different when there is a jump. Hence, Equation (65) needs to be updated to

$$\tilde{W}_j(k\Delta t) = \begin{cases} \frac{1}{\sqrt{2\pi\tilde{V}_j((k-1)\Delta t)\Delta t}} \times \exp\left(-\frac{1}{2} \frac{(R(k\Delta t) - \mu_i\Delta t - 1)^2}{\tilde{V}_j((k-1)\Delta t)\Delta t}\right) & \text{if } \tilde{J}_j = 0 \\ \frac{1}{\exp(\tilde{Z}_j(k\Delta t))\sqrt{2\pi\tilde{V}_j((k-1)\Delta t)\Delta t}} \times \exp\left(-\frac{1}{2} \frac{(R(k\Delta t) - \exp(\tilde{Z}_j(k\Delta t))(\mu_i\Delta t + 1))^2}{\exp(2\tilde{Z}_j(k\Delta t))\tilde{V}_j((k-1)\Delta t)\Delta t}\right) & \text{if } \tilde{J}_j = 1 \end{cases}. \tag{77}$$

We then normalise \tilde{W}_j so that it sums to 1 and resample \tilde{V}_j , as in the case with no jumps; additionally, we resample \tilde{Z}_j in the exact same way as \tilde{V}_j , i.e., we sort the particles and draw from the distribution $F_{\tilde{Z}_{sort}}$ to obtain the “refined” particles Z_j ,

$$Z_j(k\Delta t) \sim F_{\tilde{Z}_{sort}}. \tag{78}$$

Finally, for the estimate of λ , for each $k \in \{1, 2, \dots, n\}$, one needs to sum the cumulative value of all particles declaring a jump. That way, we obtain a probability that a jump took place at the time $t = k\Delta t$,

$$\lambda(k\Delta t) = \sum_{j=1}^N J_j(k\Delta t)\tilde{W}_j(k\Delta t). \tag{79}$$

To obtain the actual estimate of λ , one needs to average $\lambda(t)$ across all the time points obtained for different values of k ,

$$\lambda_i = \frac{1}{T} \sum_{k=1}^n \lambda(k\Delta t). \tag{80}$$

Similarly, to obtain the estimate of μ^J and σ^J , for each k , one needs to first calculate the average size of a jump from the refined particles,

$$Z(k\Delta t) = \frac{1}{N} \sum_{j=1}^N Z_j(k\Delta t), \tag{81}$$

and then calculate the mean and standard deviation of the results, weighed by the probability of a jump at time moment t indicated by $\lambda(t)$. For the weighted mean of the jumps, we obtain

$$\mu_i^J = \left(\sum_{k=1}^n Z(k\Delta t) \lambda(k\Delta t) \right) \left(\sum_{k=1}^n \lambda(k\Delta t) \right)^{-1}, \tag{82}$$

and for the standard deviation, we obtain

$$\sigma_i^J = \sqrt{\left(\sum_{k=1}^n \lambda(k\Delta t) (Z(k\Delta t) - \mu_i^J)^2 \right) \left(\frac{n-1}{n} \sum_{k=1}^n \lambda(k\Delta t) \right)^{-1}}. \tag{83}$$

The presence of jumps also influences the estimation of other parameters; some of the procedures presented in the previous subsection are not equally applicable, as jumps added to the stock price will additionally increase or more likely decrease the returns. To improve that, a correction of the definitions of $R(t)$ is needed in order to “neutralise” the impact of the jumps on the parameters. In other words, Equation (12) should be replaced with

$$R(k\Delta t) = \frac{S(k\Delta t)}{S((k-1)\Delta t)} \left(1 - \lambda(k\Delta t) \left(1 - \exp(-Z(k\Delta t)) \right) \right). \tag{84}$$

Note that the added term has a value very close to 1 when $\lambda(t)$ is close to 0, which indicates there was no jump at time t ; so, the correction to the “original” value of $R(t)$ is very minor. However, if $\lambda(t)$ is close to 1, which means there was a jump, the value of the term becomes close to $\exp(-Z(t))$, which is an inverse of the jump factor (with the estimated jump size $Z(t)$). Multiplying by that inverse brings the value of $R(t)$ to a level as if there was no jump at time t (see Figure 2), and the estimation of the parameters of the model can be carried out as before.

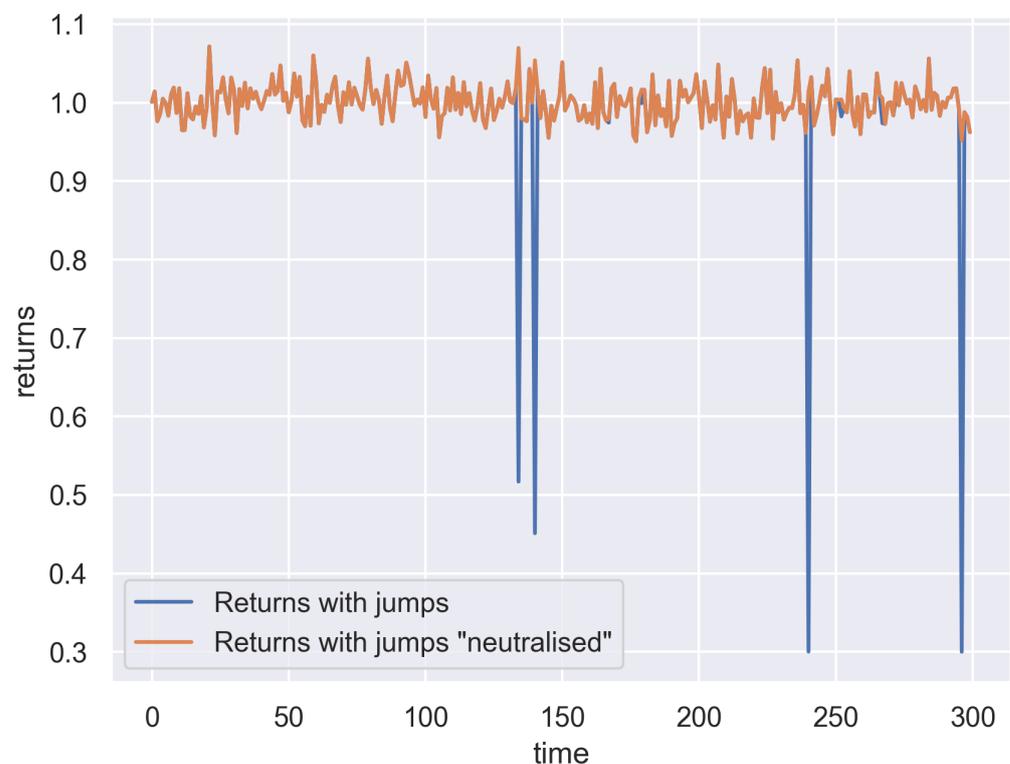


Figure 2. Comparison of the returns for a process with jumps calculated based on Formula (12) (blue line) and (84) (orange line). It can be clearly seen that the jumps have been “neutralised” in the latter case.

3.3. Estimation Procedure

The Bayesian estimation framework presented above relies on several parameters for the prior distributions that cannot be calculated within the procedure itself. They are often referred to as metaparameters. For example, for the estimation of the μ parameter, the values of two metaparameters are required, μ_0^η and τ_0^η (see Equations (19) and (20)). Their values should reflect our preexisting beliefs regarding the value of the parameter that we are trying to estimate, μ in this case. Let us say that for a given trajectory of the Heston model process, we assume the value of μ to be around 0.5. What values should we then choose for the metaparameters? First of all, we need to note that μ_0^η and τ_0^η are not the parameters of the prior distribution for μ directly. They are parameters of another random variable, which we introduced to utilise the Bayesian framework, namely η . The connection between μ and η is known and given by Equation (11). Hence, if we assume μ to be around a certain value, then using this relationship, we can deduce the value of η . In addition, since the prior distribution of η is normal, with the mean μ_0^η and the variance $\sigma_0^\eta = \frac{1}{\tau_0^\eta}$, we can propose the value of the mean of this distribution to be whatever η is for the supposed value of μ . Selecting a value for σ_0^η (and thus τ_0^η) is even more equivocal; it should reflect the level of confidence that we have for choosing a mean parameter. That is, if we feel that the value we chose for μ_0^η would bring us close to the true μ , we should choose a smaller value for σ_0^η . However, if we are not as sure about it, a larger value of σ_0^η should be used. A similar analysis can be repeated for choosing the values of the other metaparameters. We need to be aware that the prior used always in some way influences the final estimate of a given parameter. More detailed analysis on this topic is given in Section 4.2.

Another problem that emerges when applying Bayesian inference (especially for more complex models) is that estimating one parameter often requires knowing the value of some of the others and vice versa. Hence, there is not an obvious way of how to start the whole procedure. One way to address this problem is to provide the initial guesses for the values of all the parameters (as described in the paragraph above) and use them in the first round of samplings. A well-designed MCMC estimation algorithm should bring us closer to the true values of parameters with each new round of samplings. In Algorithms 1 and 2, our procedure for the Heston model without and with jumps is shown, respectively. The meaning of all symbols is briefly summarised in Appendix A.

Algorithm 1 Estimating the Heston model

Require:

- number of samples: n_s
 - time step Δt
 - maturity T
 - prices $S(k\Delta t)$ for $k \in \{0, 1, \dots, n\}$, so that $n\Delta t = T$
 - number of particles N
 - initial value of μ : μ_0
 - initial value of κ : κ_0
 - initial value of θ : θ_0
 - initial value of σ : σ_0
 - initial value of ρ : ρ_0
 - prior distribution parameters for η : μ_0^η and τ_0^η
 - prior distribution parameters for β : μ_0^β and Λ_0^β
 - prior distribution parameters for σ^2 : a_0^σ and b_0^σ
 - prior distribution parameters for ψ : μ_0^ψ and τ_0^ψ
 - prior distribution parameters for ω : a_0^ω and b_0^ω
-

Ensure:

estimate of parameter μ : $\hat{\mu}$
 estimate of parameter κ : $\hat{\kappa}$
 estimate of parameter θ : $\hat{\theta}$
 estimate of parameter σ : $\hat{\sigma}$
 estimate of parameter ρ : $\hat{\rho}$
 estimate of the volatility process: $v(t)$

for $k = 1 \rightarrow n$ **do**

set $R(k\Delta t)$, as shown in Equation (12)

end for

for $i = 0 \rightarrow n_s$ **do**

for $k = 1 \rightarrow n - 1$ **do**

▷ particle filtering procedure

for $j = 1 \rightarrow N$ **do**

obtain $V_j(k\Delta t)$, as shown in Equations (60)–(73)

end for

obtain $v(k\Delta t)$, as shown in Equation (74)

end for

obtain μ_i , as shown in Equations (13)–(23)

obtain κ_i , θ_i , and σ_i , as shown in Equations (25)–(44)

obtain ρ_i , as shown in Equations (45)–(59)

end for

set $\hat{\mu} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mu_i$

set $\hat{\kappa} = \frac{1}{n_s} \sum_{i=1}^{n_s} \kappa_i$

set $\hat{\theta} = \frac{1}{n_s} \sum_{i=1}^{n_s} \theta_i$

set $\hat{\sigma} = \frac{1}{n_s} \sum_{i=1}^{n_s} \sigma_i$

set $\hat{\rho} = \frac{1}{n_s} \sum_{i=1}^{n_s} \rho_i$

Algorithm 2 Estimating the Heston model with jumps**Require:**

number of samples: n_s

time step Δt

maturity T

prices $S(k\Delta t)$ for $k \in \{0, 1, \dots, n\}$, so that $n\Delta t = T$

number of particles N

initial value of μ : μ_0

initial value of κ : κ_0

initial value of θ : θ_0

initial value of σ : σ_0

initial value of ρ : ρ_0

prior distribution parameters for η : μ_0^η and τ_0^η

prior distribution parameters for β : μ_0^β and Λ_0^β

prior distribution parameters for σ^2 : a_0^σ and b_0^σ

prior distribution parameters for ψ : μ_0^ψ and τ_0^ψ

prior distribution parameters for ω : a_0^ω and b_0^ω

ratio of particle indicating jumps: λ^{th}

prior distribution parameters for Z : μ_0^Z and σ_0^Z

Ensure:

estimate of parameter μ : $\hat{\mu}$

estimate of parameter κ : $\hat{\kappa}$

estimate of parameter θ : $\hat{\theta}$

estimate of parameter σ : $\hat{\sigma}$

estimate of parameter ρ : $\hat{\rho}$

estimate of the volatility process: $v(t)$

```

for  $k = 1 \rightarrow n$  do
  set  $R(k\Delta t)$ , as shown in Equation (12)
end for
for  $i = 0 \rightarrow n_s$  do
  for  $k = 1 \rightarrow n$  do ▷ particle filtering procedure
    for  $j = 1 \rightarrow N$  do
      generate  $\tilde{J}_j(k\Delta t)$ , as shown in Equation (75)
      generate  $\tilde{Z}_j(k\Delta t)$ , as shown in Equation (76)
      obtain  $V_j((k-1)\Delta t)$ , as shown in Equations (61)–(73) and (77)
    end for
    obtain  $v(k\Delta t)$ , as shown in Equation (74)
    obtain  $Z(k\Delta t)$  and  $\lambda(k\Delta t)$ , as shown in Equations (78)–(79) and (81)
  end for
  for  $k = 1 \rightarrow n$  do
    update  $R(k\Delta t)$ , as shown in Equation (84)
  end for
  obtain  $\mu_i$ , as shown in Equations (13)–(23)
  obtain  $\kappa_i$ ,  $\theta_i$ , and  $\sigma_i$  as shown in Equations (25)–(44)
  obtain  $\rho_i$ , as shown in Equations (45)–(59)
  obtain  $\lambda_i$ , as shown in Equation (80)
  obtain  $\mu_i^J$ , as shown in Equation (82)
  obtain  $\sigma_i^J$ , as shown in Equation (83)
end for
set  $\hat{\mu} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mu_i$ 
set  $\hat{\kappa} = \frac{1}{n_s} \sum_{i=1}^{n_s} \kappa_i$ 
set  $\hat{\theta} = \frac{1}{n_s} \sum_{i=1}^{n_s} \theta_i$ 
set  $\hat{\sigma} = \frac{1}{n_s} \sum_{i=1}^{n_s} \sigma_i$ 
set  $\hat{\rho} = \frac{1}{n_s} \sum_{i=1}^{n_s} \rho_i$ 
set  $\hat{\lambda} = \frac{1}{n_s} \sum_{i=1}^{n_s} \lambda_i$ 
set  $\hat{\mu}^J = \frac{1}{n_s} \sum_{i=1}^{n_s} \mu_i^J$ 
set  $\hat{\sigma}^J = \frac{1}{n_s} \sum_{i=1}^{n_s} \sigma_i^J$ 

```

4. Analysis of the Estimation Results

4.1. Exemplary Estimation

We present here an exemplary estimation of the Heston model with jumps, to show the outcomes of the entire procedure. We assumed relatively noninformative prior distributions, with expected values shifted from the true parameters to make the task more challenging for the algorithm and to better reflect the real-life situation in which the priors used do not match the true parameters exactly most of the time, but they should be rather close to them. Table 1 lists all the values of the priors we used. Table 2 summarises the results obtained, and Figure 3 elaborates on those results by showing the empirical distributions of the samples for all parameters of the model.

Analysing the sample estimates for each of the parameters (presented in Figure 3), one can observe that for most of them (Figure 3a–e,h), the true value of the parameter was within the support of the distribution of all samples. However, in case of two parameters, λ and μ^J (Figure 3f and Figure 3g, respectively), the scope of the samples generated by the estimation procedure seemed not even to include the parameter's true value. This was due to the fact that those parameters were related to the intensity and size of the jumps, and for the simulation parameters we chose, those jumps did not occur frequently (as in case of real-life price falls). Hence, even though the procedure correctly identified the moments of the jumps and estimated their sizes, those estimates were relatively far from the true values, simply because there was very little source material for the estimation in the first

place. To be precise, the stock price simulated for our exemplary estimation experienced four jumps, and the times of those jumps were easily identified by our procedure with almost 100% certainty. Thus, since the length of time of the price observation (in years) was $T = 3$, the most probable value of the jump intensity λ was around $\frac{4}{3}$ (compared to the actual result in Table 2), although, obviously, the other values of λ (slightly smaller or larger) could also have led to four jumps, and this was exactly what happened in our case, as our true intensity was $\lambda = 1$ (again, see Table 2). Similarly, in the case of μ_J , the reason for the estimated average jump to be larger (in terms of magnitude) than the actual one was that the four jumps, which were simulated, all happened to be more severe than the true value of μ_J would suggest (by pure chance), and this “pushed” the procedure towards overestimating the (absolute) size of the jump.

Table 1. Priors for the exemplary estimation procedure.

Prior Parameter	Value
μ_0^{η}	1.00125
σ_0^{η}	0.001
Λ_0	$\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}$
μ_0	$\begin{bmatrix} 35 \times 10^{-6} \\ 0.988 \end{bmatrix}$
a_0^{σ}	149
b_0^{σ}	0.025
μ_0^{ψ}	−0.45
σ_0^{ψ}	0.3
a_0^{ω}	1.03
b_0^{ω}	0.05
λ^{th}	0.15
μ_0^J	−0.96
σ_0^J	0.3

Table 2. Results of the exemplary estimation procedure.

Parameter	True Value	Estimated Value	Relative Error [%]
μ	0.1	0.09829	1.77
κ	1	1.2190	21.90
θ	0.05	0.0493	1.92
σ	0.01	0.0108	8.55
ρ	−0.5	0.4379	12.40
λ	1	1.3349	33.49
μ_J	−0.8	−0.9651	20.64
σ_J	0.2	0.2298	14.88

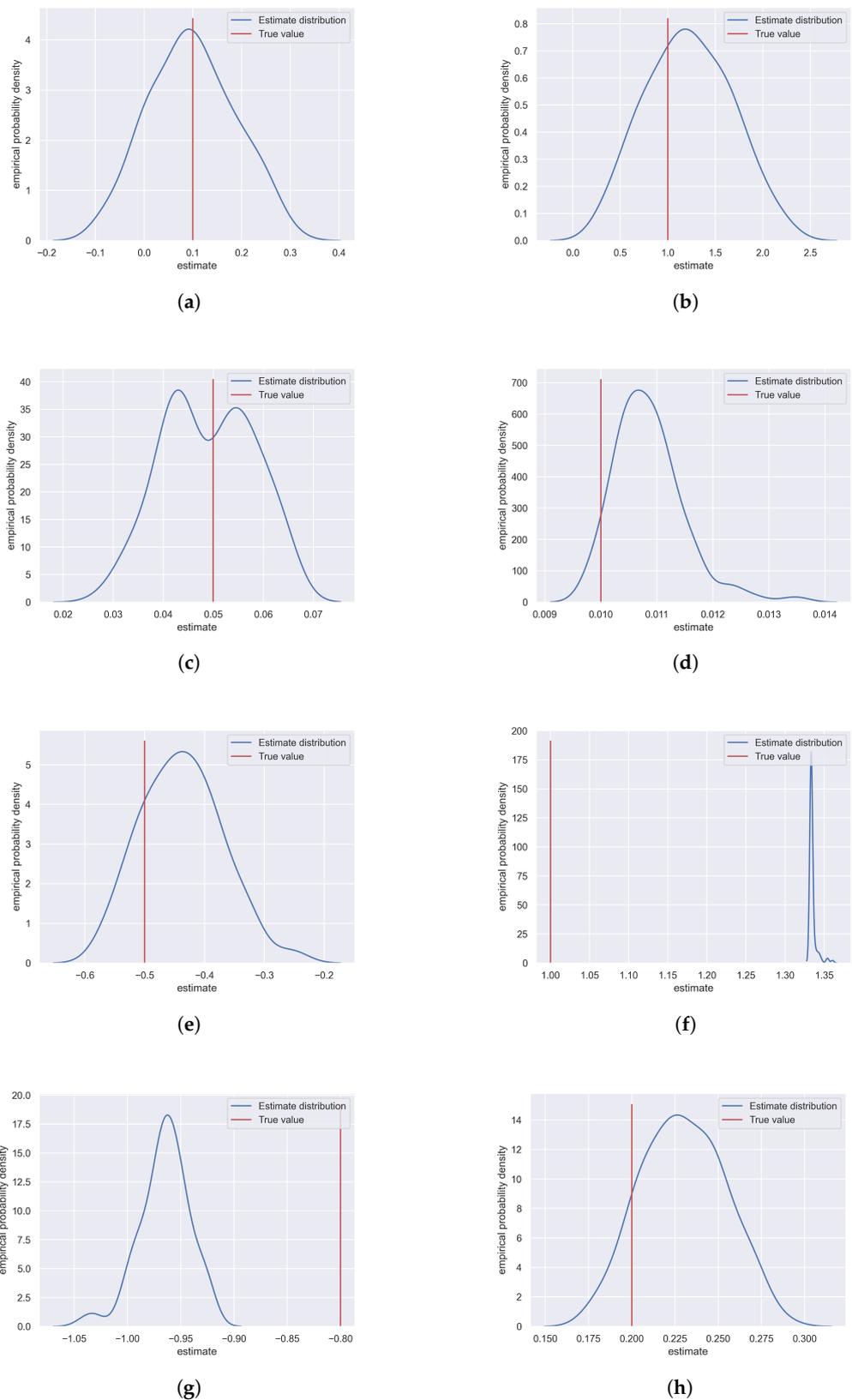


Figure 3. Empirical PDFs of the exemplary Heston parameter estimates. (a) parameter μ . (b) parameter κ . (c) parameter θ . (d) parameter σ . (e) parameter ρ . (f) parameter λ . (g) parameter μ_J . (h) parameter σ_J .

4.2. Important Findings

Although estimation through the joint forces of Bayesian inference, Monte Carlo Markov chains, and particle filtering is generally considered very effective [Johannes and Polson \(2010\)](#), it has several areas the user needs to be aware of while using this estimation scheme. One of the issues worth considering is the impact of the prior parameters. A Bayesian estimator of any kind needs to be fed with the parameters of the prior distribution, which should reflect our preexisting beliefs of what the value of the actual estimated parameter could be. The amount of information conveyed by a prior can be different, depending on several factors. One of these is the parameters of the prior distribution itself. Consider μ_0^η and σ_0^η , mentioned already in the previous section. They are the prior parameters for η , the predecessor for the μ estimates. The larger the σ_0^η we take, the more volatile the estimates of η and μ are going to be. This is a pretty intuitive fact, being a direct consequence of the Bayesian approach itself. A more subtle influence of priors is hidden in the alternation between the MCMC sampling and particle filtering procedures.

As mentioned in the previous section, the MCMC and particle filtering depend on one another. As can be seen in Algorithms 1 and 2, we took the approach that the particle filtering procedure should be conducted first, and the singular parameters that it needed should be the expected values of the prior distributions that we assumed. Having the volatility estimated this way, we can estimate the parameters and, based on them, re-estimate the volatility process, and so on. Although we can keep alternating that way as many times as we want, until the planned end of the estimation procedure, one might be tempted to perform the particle filtering procedure fewer times, as it is much more computationally expensive than the MCMC draws. The premise for this would be that after several trials, the volatility estimate becomes “good enough”, and from that point onward, one could solely generate more MCMC samples. A critical observation that we have made is that the quality of the initial volatility estimates depends very highly on the prior parameters that were used to initiate it. With a small number of particle filtering procedures followed by multiple MCMC draws, the entire scheme does not have “enough time” to properly calibrate, and the results tend to stick to the priors that are used. That means for a prior leading exactly to the true value of the parameter, the estimator returns almost error-free results; however, if one uses a prior leading to value of the true parameter, e.g., 20% larger than it really is, the estimate will probably be off by roughly 20%, which does not make the estimator very useful. A counterproposal can then be made to perform particle filtering as long as possible. This, however, is not an ideal solution either. Firstly, as we said, it is very computationally expensive, and secondly, a very long chain of samples increases the probability that the estimation procedure would at some point return an “outlier”, i.e., an estimate far away from the true value of the parameter, which is especially likely if we use metaparameters responsible for such a parameter’s larger variance (e.g., σ_0^η for η). The appearance of such “outliers” is especially unfavourable in case of the MCMC methods, since its nature is that each sample is directly dependent on the previous one; so, the whole procedure is likely to “stay” in the given “region” of the parameter space for some number of subsequent simulations, thus impacting the final estimate of the parameter (which is the mean of all the observed samples). Therefore, a clear tradeoff appears. If we believe strongly that the prior we use is rather correct and only needs some “tweaking” to adjust it to the particular dataset, a modest amount of particle filtering can be applied², followed by an arbitrary number of MCMC draws. If, however, we do not know much about our dataset and do not want to convey too much information through the prior, even at the cost of slightly worse final results, we should run particle filtering for a larger number of times. The visual interpretation of this rule is presented in Figures 4 and 5.

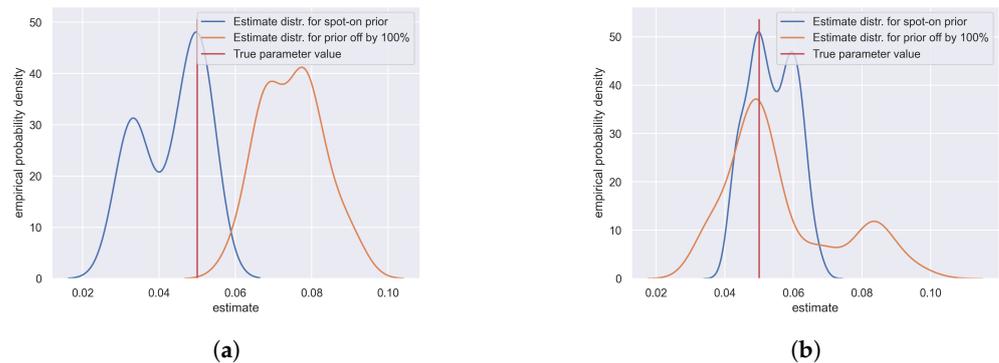


Figure 4. Empirical distributions of the estimate samples for the θ parameter in the case when the mean of the prior distribution exactly matches the true parameter and when it is twice as large. The distributions in (a) were based on 10 sampling cycles and in (b) on 500 cycles. One can observe that for the first figure, the distribution with the exact prior gives very good results, much better than the shifted one. In the second figure, both distributions are comparable.

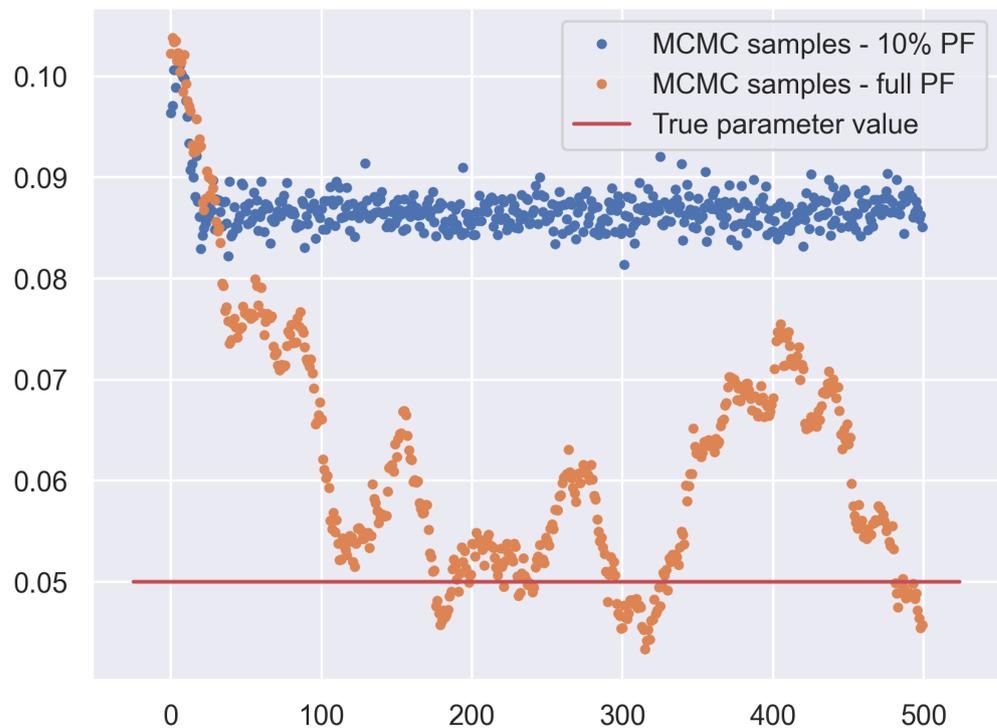


Figure 5. Sequences of estimate samples for a procedure in which particle filtering was conducted only for the first 5% of samplings and another one in which the particle filtering was conducted for all the samplings. In both cases, the mean of the prior distribution was shifted by 100% compared to its true value. It can be observed that the samples of the first procedure remained around the value close to the one dictated by the prior, whereas the samples of the other procedure converged to the true value of the parameter, which led to the better final result, less dependent on the prior parameters.

Another factor which should be taken into consideration using the Bayesian approach for estimating the Heston model is that the quality of results depends highly on the very parameters we try to estimate. The σ parameter seems to play a critical role for the Heston model, in particular. This can be observed in Figure 6. To produce it, an identical estimation procedure was performed for two sample paths (which we can think of as of two different stocks). They were simulated with the same parameters, except for σ . Path no. 1 was simulated with $\sigma = 0.01$ and path no. 2 with $\sigma = 0.1$, ten times larger. The histograms present the distribution of the estimated values of the κ parameter; the true κ was $\kappa = 1$,

and the red vertical line illustrates this true value. It is clearly visible that for the value of $\sigma = 0.01$, the samples were much more concentrated around the true value, while for a larger value of $\sigma = 0.1$, they were more dispersed, and the variance of the distribution was significantly larger. This incommodity cannot be easily resolved, as the true values of the parameters of the trajectories are idiosyncratic; they cannot be influenced by the estimation procedure itself. However, we wanted to sensitise the reader to the fact that the larger the value of σ , the less trustworthy the results of the estimation of the other parameters might be.

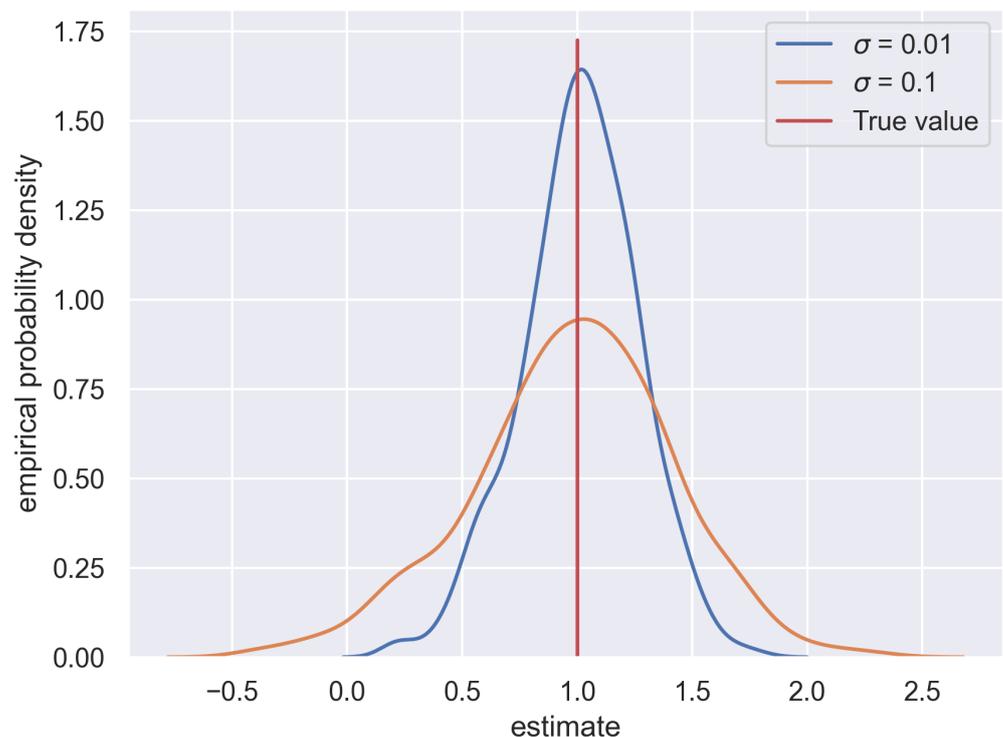


Figure 6. Empirical distributions of the estimate samples for the κ parameter of two different trajectories of the Heston model—one with $\sigma = 0.01$ and the other with $\sigma = 0.1$. The distribution of the estimate samples of the trajectory with a smaller value of σ is narrower and more concentrated; hence, it is likely to give less variable final estimates.

4.3. Towards Real-Life Applications

At the end, we wanted to emphasise the applicability of the methods described above to real market data. We briefly mentioned in the Introduction that the results yielded by some investment strategies are dependent on the character of the asset, which might be accurately captured by parameters of the models we used to model this asset (see [Gruszka and Szwabiński \(2021\)](#) for more details). Having a tool for obtaining the values of those parameters from the data opens up a new field of investigation relying on comparing synthetically generated stock price trajectories to the real ones and selecting optimal portfolio management strategies based on the results of such comparisons. We explore the possibility of utilising parameter estimates this exact way in our other work; see [Gruszka and Szwabiński \(2023\)](#).

5. Conclusions

In this paper, a complete estimation procedure of the Heston model without and with jumps in the asset prices was presented. Bayesian regression combined with the particle filtering method was used as the estimation framework. Although some parts of the procedure have been used in the past, our work provides the first complete follow-along recipe of how to estimate the Heston model for real stock market data. Moreover, we presented a novel approach to handle jumps in order to neutralize their negative impact on

the estimates of the key parameters of the model. In addition, we proposed an improvement of the sampling in the particle filtering method to obtain a better estimate of the volatility.

We extensively analysed the impact of the prior parameters as well as the number of MCMC samplings and particle filtering iterations on the performance of our procedure. Our findings may help to avoid several difficulties related with the Bayesian methods and apply them successfully to the estimation of the model.

Our results have an important practical impact. In one of our recent papers, [Gruszka and Szwabiński \(2021\)](#), we showed that the relative performance of several investment strategies within the Heston model varied with the values of its parameters. In other words, what turned out to be the best strategy in one range of the parameter values may have been the worst one in the others. Thus, determining which parameter of the model corresponds to a given stock market will allow one to choose the optimal investment strategy for that market.

Author Contributions: Conceptualization, J.G. and J.S.; methodology, J.G.; software, J.G.; validation, J.S.; formal analysis, J.S.; investigation, J.G.; resources, J.G.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, J.S.; visualization, J.G.; supervision, J.S.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All the data used throughout the article was synthetically simulated using models and methods outlined in the content of the article, no real-life data was used. Authors will respond to the individual requests for sharing the exact data sets simulated for the purpose of this article raised via e-mail to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Table of Symbols

Table A1. Table of symbols used throughout the article.

Quantity	Explanation
T	max time (i.e., $t \in [0, T]$)
$S(t)$	asset price (with $S(0) \equiv S_0$)
$v(t)$	volatility (with $v(0) \equiv v_0$)
$B^S(t)$	Brownian motion for the price process
$B^v(t)$	Brownian motion for the volatility process
μ	drift
κ	rate of return to the long-time average
θ	long-time average
σ	volatility of the volatility
ρ	correlation between prices and volatility
$Z(t)$	size of the jump
μ^J	mean of the jump size
σ^J	standard deviation of the jump size
$q(t)$	Poisson process counting jumps
λ	intensity of jumps
Δt	time step (also known as the discretisation constant)
n	number of time steps (also known as the length of data)
$\varepsilon^S(t)$	price process random component
$\varepsilon^v(t)$	volatility process random component
$\varepsilon^{add}(t)$	additional random component—see Equation (10))
η	regression parameter for drift estimation—see Equation (11)
$R(t)$	ratio between neighbouring prices—see Equations (12) and (84)
$y^S(t)$	series of dependent variables for the drift estimation—see Equation (14)
$x^S(t)$	series of independent variables for the drift estimation—see Equation (15)
\mathbf{y}^S	vector of the dependent variable for the drift estimation—see Equation (17)
\mathbf{x}^S	vector of the independent variable for the drift estimation—see Equation (18)
μ_0^η	mean of the prior distribution of η
σ_0^η	standard deviation of the prior distribution of η

Table A2. Table of symbols used throughout the article (continued).

Quantity	Explanation
τ_0^η	precision of the prior distribution of η
$\hat{\eta}$	OLS estimator of η —see (21)
μ^η	mean of the posterior distribution of η —see (20)
τ^η	precision of the posterior distribution of η —see (19)
η_i	i -th sample of η —see (22)
μ_i	i -th estimate of the drift—see (23)
β_1	regression parameter for volatility parameters estimation—see Equation (25)
β_2	regression parameter for volatility parameters estimation—see Equation (26)
β	vector of regression parameters for volatility parameters estimation—see Equation (29)
y^v	vector of the dependent variable for the volatility parameter estimation—see Equation (30)
x^v	vector of the independent variable for the volatility parameter estimation—see Equation (31)
x_2^v	vector of the independent variable for the volatility parameter estimation—see Equation (32)
X^v	matrix of the independent variable for the volatility parameter estimation—see Equation (34)
ε^v	noise vector of the volatility parameter estimation—see Equation (35)
μ_0^β	mean vector of the prior distribution of β
Λ_0^β	precision matrix of the prior distribution of β
μ^β	mean vector of the posterior distribution of β —see Equation (37)
Λ^β	precision matrix of the posterior distribution of β —see (36)
$\hat{\beta}$	OLS estimator of β —see (38)
β_i	i -th sample of β —see (39)
κ_i	i -th estimate of κ —see (40)
θ_i	i -th estimate of θ —see (41)
a_0^σ	shape parameter of the prior distribution of σ^2
b_0^σ	scale parameter of the prior distribution of σ^2
a^σ	shape parameter of the posterior distribution of σ^2
b^σ	scale parameter of the posterior distribution of σ^2
σ_i	i -th estimate of the σ —see (42)
$e_1^p(t)$	series of residuals of the price equation—see (45)
$e_2^p(t)$	series of residuals of the volatility equation—see (46)
ψ	regression parameter for ρ estimation, $\psi = \sigma\rho$ —see (47)
ω	regression parameter for ρ estimation, $\omega = \sigma^2(1\rho^2)$ —see (47)
$e_1^p(t)$	series of independent variables for the estimation of ρ —see (45)
$e_2^p(t)$	series of dependent variables for the estimation of ρ —see (46)
e_1^p	vector of the independent variables for the estimation of ρ —see (50)
e_2^p	vector of the dependent variables for the estimation of ρ —see (51)
e^p	matrix of residuals—see (52)
A^p	auxiliary matrix for solving ρ regression—see (53)
μ_0^ψ	mean of the prior distribution of ψ
τ_0^ψ	precision of the prior distribution of ψ
μ^ψ	mean of the posterior distribution of ψ —see (54)
τ^ψ	precision of the posterior distribution of ψ —see (55)
a_0^ω	shape parameter of the prior distribution of ω
b_0^ω	scale parameter of the prior distribution of ω
a^ω	shape parameter of the posterior distribution of ω —see (56)
b^ω	scale parameter of the posterior distribution of ω —see (57)
ψ_i	i -th sample of ψ —see (59)
ω_i	i -th sample of ω —see (58)
$\varepsilon_j(t)$	j -th sample of particle filtering independent errors—see (61)
$z_j(t)$	j -th sample of particle filtering residuals—see (62)
$w_j(t)$	j -th sample of particle filtering correlated—see (62)
$\tilde{V}_j(t)$	j -th raw volatility particle—see (64)
$\tilde{W}_j(t)$	j -th particle likelihood measure—see (65) and (77)
$\check{W}_j(t)$	j -th particle probability—see (66)
$U_j(t)$	j -th particle-probability vector—see (67)
$\tilde{V}_j^{sort}(t)$	j -th raw volatility particle in a sorted sequence—see (68), (69), and (70)
$\check{W}_j^{sort}(t)$	probability of the j -th particle in the sorted sequence—see (71)
$F_{\tilde{V}_j^{sort}}(v)$	continuous CDF of resampled particles—see (72)

Table A3. Table of symbols used throughout the article (continued).

Quantity	Explanation
$V_j(t)$	j -th final volatility particle—see (73)
λ^{th}	proportion of particles encoding a jump
$\tilde{J}_j(t)$	j -th raw moment-of-a-jump particle—see (75)
μ_0^J	mean of the raw size-of-a-jump particle
σ_0^J	standard deviation of the raw size-of-a-jump particle
$\tilde{Z}_j(t)$	j -th raw size-of-a-jump particle—see (76)
$Z_j(t)$	j -th resampled size-of-a-jump particle—see (78)
$\lambda(t)$	probability of a jump—see (79)
λ_i	i -th estimate of λ —see (80)
$Z(t)$	estimate of an average size of a jump—see (81)
μ_i^J	i -th estimate of μ^J —see (82)
σ_i^J	i -th estimate of σ^J —see (83)

Notes

- Equation (65) is the reason we cannot run this procedure for $k = n$, as we would not be able to obtain $R((n + 1)\Delta t)$, since the last available value is $R(n\Delta t)$.
- For applications in finance, this task is sometimes easier than for some other fields of science, as numerous works have been published already, presenting the results of the estimates of well-known stocks or market indices within various models. See, e.g., Eraker et al. (2003)

References

- Bates, David S. 1996. Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *The Review of Financial Studies* 9: 69–107. [CrossRef]
- Black, Fischer, and Myron Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]
- Chib, Siddhartha, and Edward Greenberg. 1995. Understanding the Metropolis-Hastings Algorithm. *The American Statistician* 49: 327–35. [CrossRef]
- Christoffersen, Peter, Kris Jacobs, and Karim Mimouni. 2007. *Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns and Option Prices*. Rochester: Social Science Research Network. [CrossRef]
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica* 53: 385–407. [CrossRef]
- Doucet, Arnaud, and Adam Johansen. 2009. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. *Handbook of Nonlinear Filtering* 12: 3.
- Eraker, Bjørn, Michael Johannes, and Nicholas Polson. 2003. The Impact of Jumps in Volatility and Returns. *The Journal of Finance* 58: 1269–300. [CrossRef]
- Gruszka, Jarosław, and Janusz Szewabiński. 2021. Advanced strategies of portfolio management in the Heston market model. *Physica A: Statistical Mechanics and Its Applications* 574: 125978. [CrossRef]
- Gruszka, Jarosław, and Janusz Szewabiński. 2023. Portfolio optimisation via the heston model calibrated to real asset data. *arXiv* arXiv:2302.01816.
- Heston, Steven. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies* 6: 327–43. [CrossRef]
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122: 185–212. [CrossRef]
- Johannes, Michael, and Nicholas Polson. 2010. CHAPTER 13—MCMC Methods for Continuous-Time Financial Econometrics. In *Handbook of Financial Econometrics: Applications*. Edited by Yacine Aït-Sahalia and Lars Peter Hansen. San Diego: Elsevier, vol. 2, pp. 1–72. [CrossRef]
- Johannes, Michael, Nicholas Polson, and Jonathan Stroud. 2009. Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies* 22: 2559–99. [CrossRef]
- Kloeden, Peter, and Eckhard Platen. 1992. *Numerical Solution of Stochastic Differential Equations*, 1st ed. Stochastic Modelling and Applied Probability. Berlin/Heidelberg: Springer.
- Lindley, Dennis V., and Adrian F. M. Smith. 1972. Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society: Series B (Methodological)* 34: 1–18. [CrossRef]
- Meissner, Gunter, and Noriko Kawano. 2001. Capturing the volatility smile of options on high-tech stocks—A combined GARCH-neural network approach. *Journal of Economics and Finance* 25: 276–92. [CrossRef]

O'Hagan, Anthony, and Maurice George Kendall. 1994. *Kendall's Advanced Theory of Statistics: Bayesian Inference. Volume 2B*. London: Arnold.

Wong, Bernard, and Chris C. Heyde. 2006. On changes of measure in stochastic volatility models. *International Journal of Stochastic Analysis* 2006: 018130. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.