

ROBUSTNESS AND SENSITIVITY ANALYSES FOR STOCHASTIC VOLATILITY MODELS UNDER UNCERTAIN DATA STRUCTURE

Jan Pospíšil^{*1}, Tomáš Sobotka¹, and Philipp Ziegler²

¹NTIS - New Technologies for the Information Society, Faculty of Applied Sciences,
University of West Bohemia, Univerzitní 2732/8, 301 00 Plzeň, Czech Republic,

²Department of Mathematics, University of Rostock, Ulmenstraße 69, 18057 Rostock, Germany

Received 3 March 2017

Abstract

In this paper we perform robustness and sensitivity analysis of several continuous-time stochastic volatility (SV) models with respect to the process of market calibration. The analyses should validate the hypothesis on importance of the jump part in the underlying model dynamics. Also an impact of the long memory parameter is measured for the approximative fractional SV model. For the first time, the robustness of calibrated models is measured using bootstrapping methods on market data and Monte-Carlo filtering techniques. In contrast to several other sensitivity analysis approaches for SV models, the newly proposed methodology does not require independence of calibrated parameters - an assumption that is typically not satisfied in practice. Empirical study is performed on data sets of Apple Inc. equity options traded in April and May 2015.

Keywords: robustness analysis; sensitivity analysis; stochastic volatility models; bootstrapping; Monte-Carlo filtering

MSC classification: 62F35; 62F40; 91G20; 91G70

JEL classification: C52; C58; C12; G12

*Corresponding author, honik@kma.zcu.cz

1 Introduction

Stochastic volatility (SV) models are common tools for evaluating option prices at financial markets and are of the interest of both academics and practitioners. For the applicability in practice, one obtains model parameters by means of calibration or by using filtering estimation techniques as in [Creel and Kristensen \(2015\)](#). We consider calibration to European option prices, as they are widely traded and therefore a sufficient number of contracts is available. As there exists a lot of different SV models, one has to choose an appropriate model for this process. The main assumption of any option pricing model is the structure of modelled dynamics of the underlying. Several empirical studies of various price processes have been analysed in the literature.

Authors [Carr and Wu \(2003\)](#) found the presence of continuous and jump components of modelled market dynamics for the SPX 500 index data. This was done by analysing out-of-the money and at-the-money options' decays in the price for time to maturity reaching zero. As in our case, the authors did not examine prices of the underlying directly which would require extremely high-frequency data sets that could be effected by market microstructure (for time-series tests see e.g. [Barndorff-Nielsen and Shephard \(2006\)](#); [Hwang and Shin \(2014\)](#)). In [Campolongo, Cariboni, and Schoutens \(2006\)](#) the use of stochastic volatility models with jumps was recommended because the uncertainty in the estimated option prices mostly came from jump parameters of the considered model. This statement was derived from a study with fictional data and model parameters. We test the hypotheses of [Carr and Wu \(2003\)](#); [Campolongo, Cariboni, and Schoutens \(2006\)](#) in the case of real market data and we also show that [Campolongo, Cariboni, and Schoutens \(2006\)](#) method is not suitable for practice (at least for our data sets). A different approach, where a model robustness to varying data structure plays a crucial role, is proposed and applied to real market data sets including Apple Inc. equity options traded in April and May 2015.

In this paper, three subclasses of SV models are considered alongside their representative models: standard [Heston \(1993\)](#), jump-diffusion [Bates \(1996\)](#) and so-called approximative fractional jump diffusion (FSV) model, which outperformed the [Heston \(1993\)](#) model in terms of in-sample calibration errors in [Pospíšil and Sobotka 2016](#)). Unlike the case of [Bayer, Friz, and Gatheral \(2016\)](#) and many other very recent manuscripts, the FSV model is considered only in the long-memory regime ($H > 0.5$). This is due to restrictions on the pricing solution and also in this case we can use the same unifying pricing approach for all three models ([Baustian, Mrázek, Pospíšil, and Sobotka 2017](#)). Some comments on the rough volatility regime are to be found in the conclusion.

The considered approaches are tested under uncertainty in the option price structure and are compared with sensitivity and uncertainty analysis tools. [Saltelli, Tarantola, Campolongo, and Ratto \(2004\)](#) defined sensitivity analysis as *“the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input”*. We want to know how sensitive calibration errors are with respect to the changing data structure and also how the calibrated parameters are effected. This is done by performing an uncertainty analysis. According to [Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli et al. \(2008\)](#), *“uncertainty and sensitivity analyses should be run in tandem, with uncertainty analysis preceding in current practice”*. The method of Sobol indices is the most common approach for global sensitivity analysis. An application of Sobol indices in option pricing can be found in the paper of [Bianchetti, Kucherenko, and Scoleri \(2015\)](#), where the impact of uncertainty in prices and greeks is measured. However, for real market data one cannot assume independence of the input parameters for calculating Sobol indices. In this paper we use different global sensitivity analysis methods which are discussed e.g. in [Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli et al. \(2008\)](#): on a bootstrapped data structure we visualize a dependence of calibrated parameters by scatterplots and hypotheses of jump-presence and long-memory persistence were tested by Monte-Carlo filtering techniques.

Considered models are calibrated from markets (or bootstrapped data) comprising European call options. A European call is a contract that gives the buyer a right to buy a share of the underlying asset for a fixed (strike) price K at some future time T . If the buyer observes a stock price at maturity lower than K , he or she doesn't utilize her right to buy the asset for K . Vice

versa, the buyer is exercising the right as long as $S_T \geq K$. This translates into the following pay-off function,

$$P(x) = \max(x - K, 0), \quad \text{where } x = S_T.$$

To answer the question – what is the fair price to pay for this contract – one has to build up a set of assumptions on the market that drives $(S_t)_{0 < t \leq T}$. Since the Nobel prize winning [Black and Scholes \(1973\)](#) model one usually considers the stock market to be a stochastic process and the fair price is then obtained using arbitrage-free arguments¹. Main differences between the consider models are in the process that drives evolution of the stock price. All approaches in this paper not only assume that the stock price process is of random nature, but also it is assumed that the variance thereof is a stochastic process itself. Hence, a stochastic volatility model can be viewed as a natural extension to the Black-Scholes approach.

Purpose of this article is to help practitioners in the daily calibration process of option pricing models. For quantitative tasks beyond the Black-Scholes model, one might face a decision call of choosing a suitable model for the current situation. Different criteria have to be considered, for example the *in-sample / out-of-sample errors*, the ability to model the *volatility smile* etc. We compare the *robustness* of different models with respect to a given option structure. This is important, because the traded equity options can be very different every day, and so their structure (amount of options, strike prices, expiration times, ask-bid-spread, etc.) is another source of uncertainty for the model choice - models might perform differently with respect to different market structures. We show how to analyse this uncertainty, measure it's impact on the predicted option prices and use this as a criteria for choosing a suitable option pricing model. In doing so, we use bootstrapping of the option data and we also introduce several measures of robustness.

The structure of the paper is as follows. In [Section 2](#) we introduce the studied stochastic volatility models and the process of calibration of these models to real market data. In [Section 3](#) we describe the methodology, in particular the bootstrapping of option prices, as well as we detail the uncertainty and sensitivity analyses. In [Section 4](#) we present the results, we compare all models in terms of variation in model parameters and in bootstrapped option prices. We also provide the results of the Monte-Carlo filtering trials, showing us the importance of the jump intensity for the Bates model and the importance of the long memory parameter for the approximative fractional model. We conclude all obtained results in [Section 5](#).

2 Stochastic volatility models

In this paper we focus on a class of stochastic volatility models that captured attention of both practitioners and academics. Modelling approaches of this type are not restricted by the constant volatility assumption. Many of these models are tractable for a wide range of applications including market calibration that will be described at the end of this section.

We consider risk-neutral jump-diffusion dynamics of the equity market, where a stock price S_t evolves in time according to the following stochastic differential equations

$$dS_t = rS_t dt + \sqrt{v_t} S_t d\widetilde{W}_t^S + S_t dJ_t, \quad (1)$$

$$dv_t = p(v_t) dt + q(v_t) d\widetilde{W}_t^v, \quad (2)$$

$$d\widetilde{W}_t^S d\widetilde{W}_t^v = \rho dt, \quad S_0, v_0 \in \mathbb{R}^+, \quad (3)$$

where $p, q \in C^\infty(0, \infty)$ are general coefficient functions for the volatility process and ρ is the correlation between standard Wiener processes \widetilde{W}_t^S and \widetilde{W}_t^v . To get market dynamics postulated by [Heston \(1993\)](#) we set $dJ_t \equiv 0$, $p(v_t) = \kappa(\theta - v_t)$ and $q(v_t) = \sigma\sqrt{v_t}$. The set of model parameters Θ^H is then defined as $\Theta^H := \{v_0, \kappa, \theta, \sigma, \rho\}$. For the [Bates \(1996\)](#) model, functions p, q remain the same only dJ_t corresponds to the compensated compound Poisson process with log-normal

¹For more detail on the arbitrage pricing see, for instance, [Shreve \(2004\)](#).

jump sizes – jumps occur with intensity λ and their sizes are log-normal with parameters μ_J and σ_J . The set of parameters, in the Bates model case, consists of $\Theta^B := \{v_0, \kappa, \theta, \sigma, \rho, \lambda, \mu_J, \sigma_J\}$. Due to more degrees of freedom, the model should provide a better market fit and as was shown in [Duffie, Pan, and Singleton \(2000\)](#) adding a second jump process to (1) might not improve the fit any more. Instead of considering a stochastic volatility model with jumps in both underlying and variance dynamics, we use an approximative fractional process as in [Baustian, Mrázek, Pospíšil, and Sobotka \(2017\)](#); [Pospíšil and Sobotka \(2016\)](#). Under the approximative fractional model one assumes the same type of jumps as in the previous case, but $p(v_t) = [(H - 1/2)\psi_t\sigma\sqrt{v} + \kappa(\theta - v_t)]$ and $q(v_t) = \varepsilon^{H-1/2}\sigma\sqrt{v}$, where $\varepsilon > 0$ is an approximating factor and ψ_t is an Itô integral:

$$\psi_t = \int_0^t (t - s + \varepsilon)^{H-3/2} dW_s^\psi.$$

The set of parameters $\Theta^F := \{v_0, \kappa, \theta, \sigma, \rho, \lambda, \mu_J, \sigma_J, H\}$ also includes the Hurst exponent H . As was shown by [Lewis \(2000\)](#) and [Baustian, Mrázek, Pospíšil, and Sobotka \(2017\)](#) respectively, all three models attain a semi-closed form solution not only for plain European options, but also for more complex structured derivatives - this is crucial for our experiments, a single trial will involve 200 calibrations of each model to different data sets. We also did not perform analyses of models with time dependent parameters which were studied by [Mikhailov and Nögel \(2003\)](#), [Osajima \(2007\)](#), [Elices \(2008\)](#), [Benhamou, Gobet, and Miri \(2010\)](#) etc. As mentioned in [Bayer, Friz, and Gatheral \(2016\)](#), the general overall shape of the volatility surface, with respect to equities, does not change in time significantly.

To use the aforementioned models in practice one has to calibrate them to a given market beforehand². The calibration process can be viewed as an optimization problem of finding the best fit to the given option price surface. Let the surface consist of N options, each with a different strike price (K) and time to maturity (T) combination. Then typically one uses a utility function defined by the following weighted least-square criterion:

$$\begin{aligned} \hat{\Theta} &= \arg \inf_{\Theta} G(\Theta), \\ G(\Theta) &= \sum_{j=1}^N w_j (C_j^\Theta(T_j, K_j) - C_j^*)^2, \end{aligned} \quad (4)$$

where $C_j^\Theta(T_j, K_j)$ is a model price calculated using the parameter set Θ and C_j^* represents the j^{th} quoted option price. Weights w_j are commonly represented as a function of the ask-bid price spread – based on the results of [Mrázek, Pospíšil, and Sobotka \(2016\)](#), we consider the weight function

$$w_j = \frac{1}{(C_j^{\text{ask}} - C_j^{\text{bid}})^2} \quad (5)$$

for $j = 1, 2, \dots, N$.

For the analyses we utilize a data corresponding to European call options on Apple Inc. stock. Four data sets from slightly different time periods (1/4/2015, 15/4/2015, 1/5/2015 and 15/5/2015.) and a structure of the newest data set is depicted by [Figure 1](#). The structure of Apple Inc. stock options should illustrate that we did not restrict our trials to only specific time-to-maturities nor to a specific moneyness range.

²Alternatively one can estimate the parameters from time-series data.

(a) Structure of the Apple Inc. call options (15/5/2015). (b) Parameter bounds for all considered calibration trials.

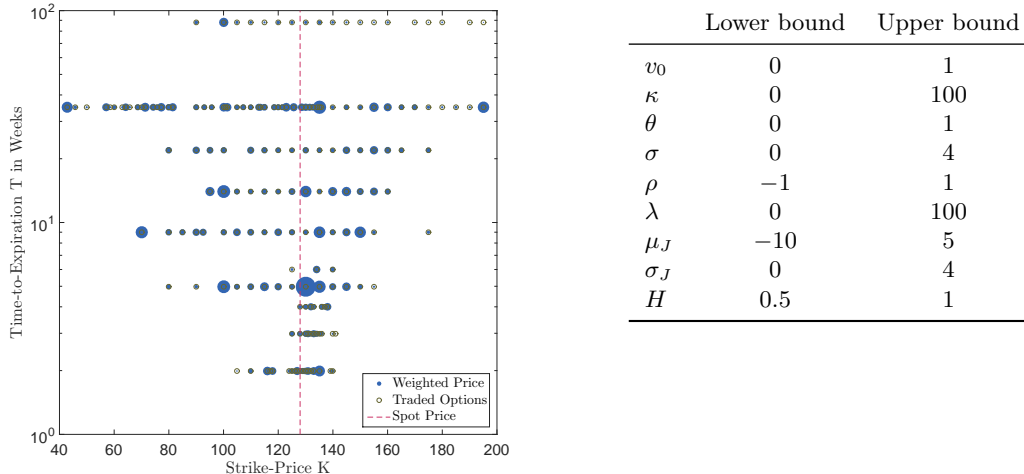


Figure 1: Data structure and bounds for calibrated parameters. On the left, we depict weighted call prices $w_j C_j^*$ by a ball centred in the $K - T$ plane. The diameter of each filled ball relates to the weighted call price and its center corresponds to the pair K_j, T_j .

3 Methodology

In this section we introduce a methodology to analyse a model robustness with respect to uncertain option price structure. This is done by using bootstrapping technique to estimate unobserved samples from the probability density function of the data structure. Also we introduce several measures of robustness that will be later used to compare the models.

We also define what we mean by a sensitivity analysis. In particular, we would like to analyse whether calibrated values of the jump-intensity parameter λ (for the Bates model) and of the Hurst exponent H (for the FSV model) can significantly effect market fit of the model. As for the measures of robustness, we take advantage of the bootstrapped samples and we use a Monte-Carlo filtering technique to quantify the importance of the mentioned parameters. Both λ and H have important consequences for the model choice - by setting $\lambda = 0$ and $H = 0.5$ we obtain the standard diffusion dynamics postulated by [Heston \(1993\)](#).

3.1 Bootstrapping Option Prices

The daily option prices are given as a set of ask- and bid-prices with different strike-price K and time to maturity T . This data is new for every day, as the behaviour and the value of the underlying stock, the reference value for the option prices, changes, and it is not statistical in the sense that we only have one dataset for every time instance. Although the option prices usually have some similarities with prices of former days, the focus of traders can change to different options and therefore not all $K \times T$ combinations have to be the same as well as ask- and bid-prices can differ strongly. This can significantly impact calibration results. Therefore we focus on uncertainty in the options structure $K \times T$. Let X be a random variable representing the pair (K, T) . Then currently observed strike prices and maturities $(K_j, T_j), j = 1, \dots, N$, as mentioned in (4), are samples of X and to each pair we can attach the quoted market option price C_j^* .

To measure the impact of the uncertain option structure on the model calibration, common methods for uncertainty analysis need statistical data which is not available in practice. In fact, option pricing models are typically recalibrated daily and only to current available and suitable data sets [see e.g. [Mikhailov and Nögel \(2003\)](#), [Yekutieli \(2004\)](#)]. In the following we will apply

the bootstrapping method to our data set. Since the original paper by Efron (1979) and especially his monograph (Efron 1982), research activity on the bootstrap method grew dramatically and we refer the reader for example to the book by Chernick (2008) and the comprehensive literature review therein. In what follows we apply the standard non-parametric i.i.d. bootstrap method.

We will perform bootstrapping on the set of observed structure (K_j, T_j) , $j = 1, \dots, N$, i.e. we obtain a new set \mathbf{X}^\dagger by sampling N times with replacement³. Obviously, for each element of \mathbf{X}^\dagger we can assign a market option price that corresponds to the strike and maturity combination. This provides us with the bootstrap option prices $\mathbf{C}^\dagger = (C_j^\dagger)_{j=1}^N$. This bootstrap procedure is then repeated M times and hence we get M bootstrap samples $\mathbf{C}^{\dagger,1}, \mathbf{C}^{\dagger,2}, \dots, \mathbf{C}^{\dagger,M}$, each of size N .

Let $\Theta^{\dagger,i}$ denote the outcome of the calibration procedure (4) applied to the bootstrap sample. The bootstrap estimate of the mean of the bootstrap replications is

$$\bar{\Theta} = \frac{1}{M} \sum_{i=1}^M \Theta^{\dagger,i}. \quad (6)$$

3.2 Model Comparison

With the bootstrap method, we estimate the calibration parameters M times. As we want to compare different models based on their robustness, we want to analyze

- ... the variation of the bootstrap replications $\Theta^{\dagger,i}$ and the following
- ... variation of the predicted option prices $C^{\Theta^{\dagger,i}}$, based on the bootstrap replications.

In this paper the comparison of the three SVJD models is widely supported by diagrams for exploring their inner structure and performance. We analyze the bootstrapped calibration parameters with two different diagrams: For the variation in the bootstrapped calibration parameters Θ_i we use *scatterplot matrices*. For the variation in option prices, we visualize the errors and variations of $C^{\Theta^{\dagger,i}}$ in the $K \times T$ -plane.

Variation in $\Theta^{\dagger,i}$

To study the calibration parameter's variation, one can use a variety of methods. First of all, we want to analyze the variation in the bootstrap replications $\Theta^{\dagger,i}$ to derive informations about the model – e.g. if one really can state that the volatility is fast mean reverting. Additionally, we want to study connections between individual calibration parameters – e.g. if the strength of mean reversion varies for different correlations between the Wiener processes. To gather the information, we analyze the variation in $\Theta^{\dagger,i}$ with the help of scatterplot matrices, square matrices with size equal to the number of model parameters. On the diagonal, histograms of the individual calibration parameters are plotted, while the other entries are occupied by scatterplots.

If one further wants to analyze $\Theta^{\dagger,i}$ with statistical methods, normality of the bootstrapped calibration parameters is an important property – e.g. if we want to calculate confidence intervals for the bootstrap estimate $\bar{\Theta}$ of the calibration parameters. To support such analyses, quantile-quantile plots with respect to a normal distribution are suitable visualization tools.

Variation in $C^{\Theta^{\dagger,i}}$

The bootstrapped calibration parameters $\Theta^{\dagger,i}$ contain the uncertainty of the option pricing model with respect to the available options for the calibration. As the bootstrapping was motivated by the option pricing structure, vice versa it is helpful to know how this uncertainty affects the model price predictions $C^{\Theta^{\dagger,i}}(K_j, T_j)$ of an individual option C_j^* . For this purpose, two measures are introduced:

³For instance, if $N = 6$ one might obtain $\mathbf{X}^\dagger = (X_2, X_1, X_4, X_4, X_3, X_2)$ where $X_j = (K_j, T_j)$.

Firstly, the *bootstrap relative error* for the j -th option with market price C_j^* is calculated by:

$$BRE_j = \frac{|\bar{C}_j - C_j^*|}{C_j^*}, \quad (7)$$

for $j = 1, \dots, N$, where \bar{C}_j is defined as

$$\bar{C}_j = \frac{1}{M} \sum_{i=1}^M C^{\Theta^{\dagger,i}}(K_j, T_j).$$

The measure indicates an individual price prediction error of the bootstrap estimation \bar{C} normalized with the market option price. Using bootstrap relative errors, we should be able to detect systematic prediction errors, which can come from the specific option pricing structure.

We are also interested in the variance of prediction error $|C^{\Theta^{\dagger,i}}(K_j, T_j) - C_j^*|$ for the bootstrapped parameters $\Theta^{\dagger,i}$ with respect to bootstrap trials $i = 1, \dots, M$. To be able to compare variances of predictions for options with different prices C_j^* , we use relative errors as before to get the variance error measure V_j for the j -th option:

$$V_j = \text{Var} \left(\frac{|C^{\Theta^{\dagger,i}}(K_j, T_j) - C_j^*|}{C_j^*} \right), \quad (8)$$

This measure is evaluated for all traded options $j = 1, \dots, N$.

The error and variance measures are visualized with diagrams in the $K \times T$ plane. Each traded option is marked with a circle which is centred according to (K, T) of the contract. For a clear arrangement, the T -axis is in logarithmic scale, because there are many traded options with short but slightly different time to maturity. Current asset price, the reference for the option prices, is plotted as a dashed line. Finally, the average relative error respectively the variance of the bootstrapped prices are visualized as balls, where the balls area is scaled with the error/variance.

3.3 Sensitivity analysis

According to Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli et al. (2008), the scatterplots can be used as a tool for sensitivity analysis to measure the impact of input parameters on the model output. Additionally, in this paper we would like to know, if fractionality of stochastic volatility and jumps are important for the robustness of option pricing models. Fractionality is represented by the Hurst parameter H , while jumps are represented by the intensity parameter λ (which is linked to the parameters σ_J and μ_J). Therefore, the importance of jumps and fractionality can be translated into the question, if H and λ have an impact on the quality of the calibration result. This question will be addressed by the Monte-Carlo filtering technique, which analyzes if a distribution of values of a chosen parameter effects significantly some specific quality measure.

In our context, we have chosen the following Monte-Carlo filtering technique⁴: To each set of calibrated model parameters, obtained from the bootstrapped data, we assign average absolute relative error (AARE) with respect to the whole set of traded options as a quality measure for the parameter set. This enables us to divide the sets of parameters into a behavioural (well fitting) group and a non-behavioural (poor fitting) group with respect to the AARE measure. As a behavioural set of parameters we consider parameters for which AARE is in the lower 3/8 quantile. A non-behavioural set, on the other hand, consists of parameter sets that lead to the worst 37.5% of the AARE values (upper 3/8 quantile). The rest (1/4 of the results) is considered as a “grey zone” and is not taken into account for the comparison⁵. For the considered parameter set groups we

⁴For more details on Monte-Carlo filtering approaches see, for instance Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli et al. (2008).

⁵In this case, we will not be able to decide whether the parameters lead to a good or bad description of the modelled market

perform a two-sample Kolmogorov-Smirnov test to verify the null hypothesis whether both sets are sampled from the same distribution function. According to [Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli et al. \(2008\)](#), by rejecting the null hypothesis at a reasonable level of significance⁶ we show that the parameters are important with respect to the calibration procedure. However, if we are not able to reject the hypothesis then we cannot judge the importance of the selected parameter.

For the Bates model we would like to answer whether the jumps are worth implementing to fit the observed market or if one should stay within the Heston model framework ($\lambda \neq 0$). We also judge the importance of the Hurst parameter in the fractional stochastic volatility case.

4 Results

In our trials, the bootstrap calibration was performed $M = 200$ times. In the following text we discuss the results based on the four data sets mentioned above⁷. For example the data set from 15th May consists of 197 options and the most weight is typically assigned to the at-the-money contracts allocated near spot price in Figure 1. Apart from that, weights are almost evenly distributed in the $K \times T$ plane.

We start the model comparison by examining the overall calibration errors of all three models as seen in Table 1. We note that the additional model features of the Bates model (jumps) and the FSV model (jumps and approximative fractional Brownian motion) usually lead to a better market fit. The best average relative errors were obtained on the 15/4/2015 data set (FSV model reached 2.16% error) and the worst market fit in terms of the consider measure was w.r.t. 1/5/2015 data and the Heston model (6.58%). We conclude that using the Heston model we were able to retrieve similar error measures to the Bates and FSV model only for the data set from 15/5/2015.

Table 1: Overall calibration errors of the three models for Apple Inc. stock on all four datasets.

Trading day	1/4/2015	15/4/2015	1/5/2015	15/5/2015
Heston model	5.15%	3.79%	6.58%	3.39%
Bates model	3.73%	3.57%	5.77%	3.41%
FSV model	2.21%	2.16%	5.89%	3.20%

Variation in $\Theta^{\dagger,i}$

In Figures 2 and 3, the scatterplot matrices of the parameters $\Theta_{\text{Bates}}^{\dagger,i}$ and $\Theta_{\text{FSV}}^{\dagger,i}$ are depicted (the results for $\Theta_{\text{Heston}}^{\dagger,i}$ are similar to the ones for the Bates model, we discuss them shortly at the end of the section). First and foremost, we inspect if we reached the lower and upper bounds for the calibration parameters (the bounds are listed in Figure 1b). This can indicate

- ... for a zero bound (e.g. $\kappa \geq 0$), that a model parameter (e.g. mean reversion κ) could be dropped,
- ... for non-zero bounds (e.g. $\kappa \leq 100$), that they should be reselected if it is not in contradiction with the parameter interpretation and if it does not breach model restrictions.

For correlation ρ , the natural limits at -1 and 1 indicate that only one Brownian motion can model both, the random movement of the asset price and its volatility. Additionally, ρ and μ_J include zero in the interior of their calibration range, which should be considered during the exploration of scatterplots. E.g. for the uncorrelated Heston model [De Marco and Martini \(2012\)](#) showed an explicit formula which not even needs numerical integration and, possibly, even the

⁶For all trials we use “standard” $\alpha = 5\%$ level of significance. In most of the trials we could have even lower α and still we would reject the null hypothesis.

⁷All results and data are available in supplementary materials.

other models might be simplified. On the contrary, the value $\mu_J = 0$ has no model reducing consequences. Last but not least, a dependence structure between the calibration parameters can be obtained from a single scatterplot and we are able to compare the bootstrap mean $\bar{\Theta}$ (red star) and the parameters Θ from the overall calibration (black cross).

Starting with the Bates model and the last criteria, one cannot observe a noticeable accumulation of ρ and μ_J at zero in the histograms at the diagonal of Figure 2. Further on, the histograms show that the parameters $v_0, \kappa, \theta, \sigma, \rho$ and μ_J have no concentration at their limits. However, the small values of λ (mostly between 10^{-3} and 10^{-4}) and the accumulation of σ_J at zero are noticeable. Moreover, if one looks at the scatterplot between λ and μ_J , λ and σ_J , one observes that *either* λ is nearly zero *or* σ_J and μ_J are close to zero. For the model this means, we have two possible cases: either the Bates model has very rare jump, or it has many very small jumps of the same intensity. If the jump-frequency λ approaches zero, then the average jumps sizes μ_J are almost for all calibrations negative. This statistical connection between λ , μ_J and σ_J should be considered at the calibration by a general modelling decision. One option would be to fix the jump intensity parameter λ beforehand.

Furthermore, the scatterplots depict that from all parameters κ is the one with the most correlations. One can see from the scatterplots that the stronger the mean reversion is,

- ... the higher is the volatility of volatility σ ,
- ... the lower is the initial volatility v_0 and the long-run volatility θ ,
- ... the more negative the correlation ρ between the two Brownian motions is.

The difference between $\bar{\Theta}_{\text{Bates}}$ and Θ_{Bates} for most parameters is not very large, apart from λ and σ_J : the overall calibration resulted in a model with many small jumps of the same size, while the bootstrap resulted in the mean in a model with rare jumps of different size.

In Figures 2 and 3 we can see that the histograms of FSV and Bates model are quite similar for $v_0, \theta, \rho, \lambda, \mu_J$ and σ_J . The rate of mean-reversion κ in FSV model is slightly different to Bates model, as in some of the trials we can get a fast mean-reversion rate, $\kappa \gtrsim 50$. Furthermore, σ in FSV model is significantly higher than in Bates model. This can be explained by the scaling with $\varepsilon^{H-1/2}$ (see equation (2) and the definition of $q(v_t)$). The Hurst parameter H is positively correlated with v_0, θ and ρ and negatively with κ and σ . In Tables 2 and 3 we provide pairwise correlation coefficients for both models. Note that these are all stochastic volatility parameters, a connection of the Hurst parameter with the jump parameters is not obvious in the scatterplots, see also Table 3. Finally, in FSV model $\bar{\Theta}_{\text{FSV}}$ and Θ_{FSV} are very close together which is a desirable result.

Table 2: Pairwise correlation coefficients of calibrated parameters for the Bates model and the data from 15/5/2015.

	v_0	κ	θ	σ	ρ	λ	μ_J	σ_J
λ	-0.7456	-0.0356	-0.7940	-0.5483	0.7235	--	0.0012	-0.2984
μ_J	0.0866	-0.0356	0.0938	0.0526	-0.1065	0.0012	--	-0.3553
σ_J	0.2647	0.3807	0.3726	0.5504	-0.0207	-0.2984	-0.3553	--

For the Heston model, the scatterplot matrix showed similar results as the upper left 5×5 submatrix of Figure 2. There were no accumulations of $\Theta_{\text{Heston}}^{\dagger, i}$ at the bounds and the correlation seemed nearly linear. Independence of the calibration parameters, necessary for the sensitivity analysis method proposed in Campolongo, Cariboni, and Schoutens (2006), cannot be assumed for any of the models. Thus, this method is not suitable in our context.

One can notice from the Q-N plots in Figure 4, that normality of the bootstrapped parameters and of the resulting calibration error can be assumed for the Heston model – for Bates and FSV this was not the case. Therefore, statistical techniques which assume normality of the data could be used to further analyze the Heston model, but not for comparison of all three models.

Scatterplot Matrix of Calibration Parameters in Bates Model

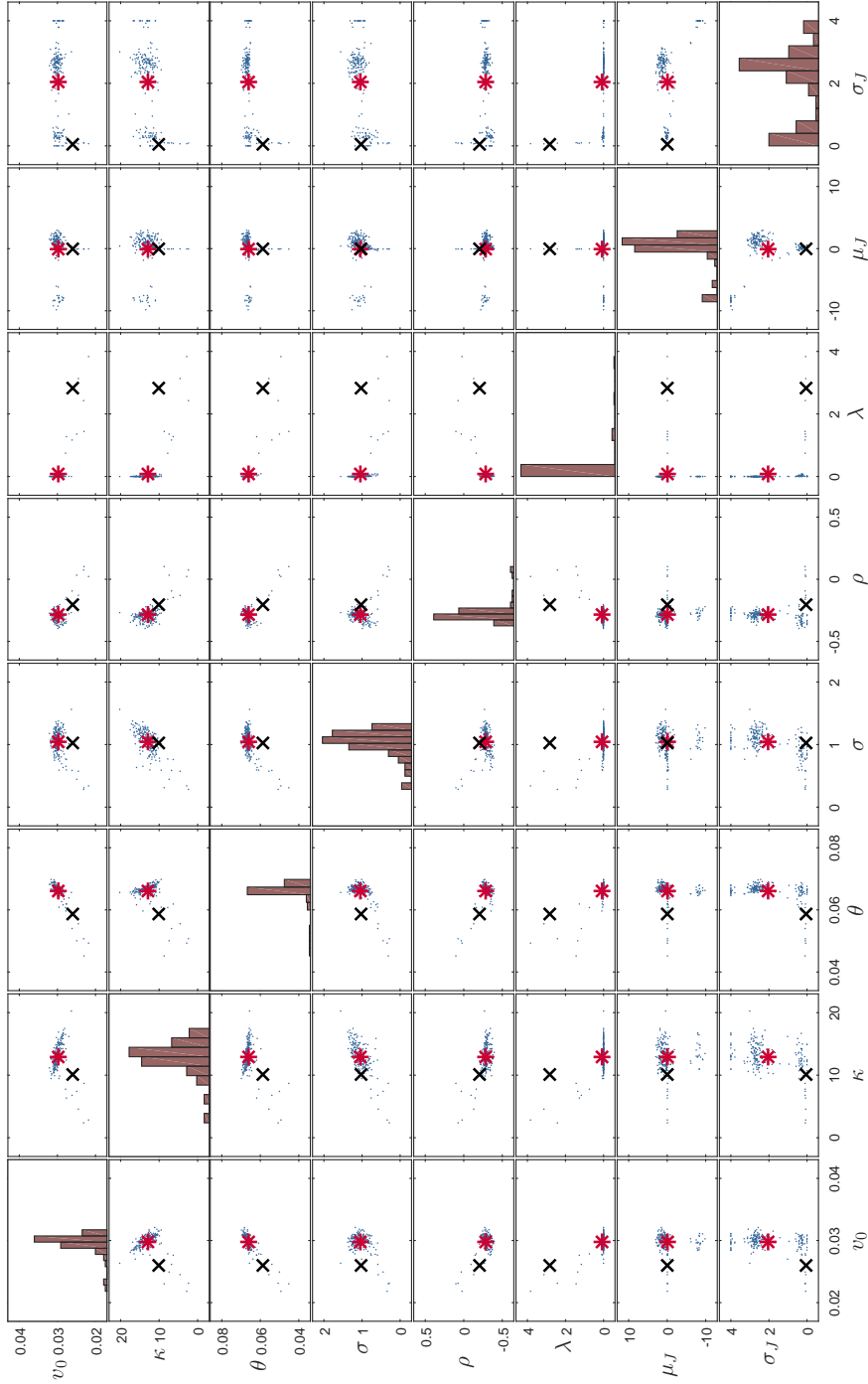


Figure 2: Scatterplot matrix for the Bates model (15/5/2015, for results on other data sets see the supplementary materials). Diagonal elements depict histograms of parameter values obtained by bootstrap calibrations (e.g. the first histogram corresponds to the values of v_0). Off-diagonal elements illustrate a dependence structure for each parameter pair. In those figures, a black cross represents the reference value of the specific parameter (obtained from calibration to the whole data set) and by a red star we depict the bootstrap estimate of the value - the closer the two are, the better.

Scatterplot Matrix of Calibration Parameters in FSV Model

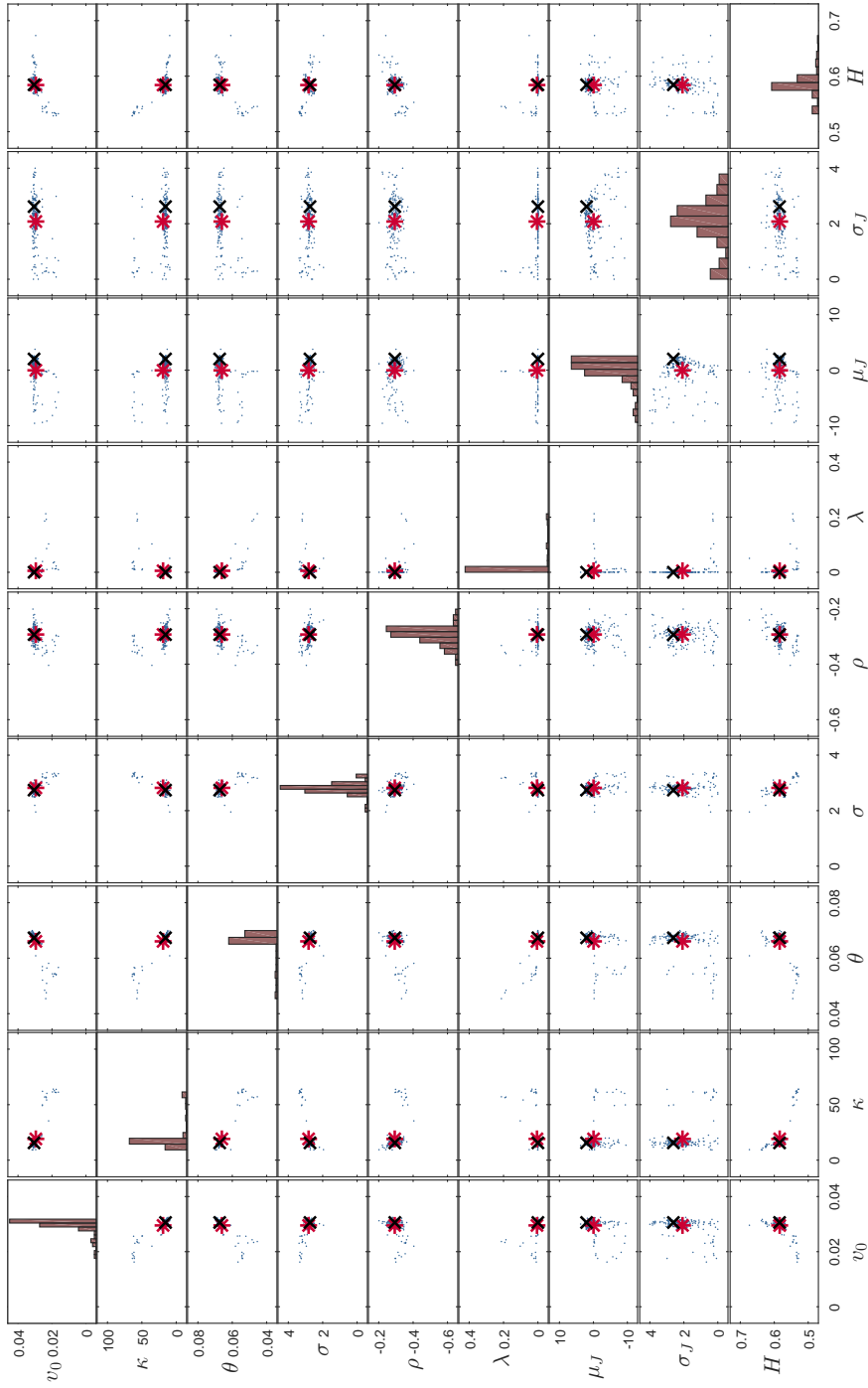


Figure 3: Scatterplot matrix for the FSV model (15/5/2015, for results on other data sets see the supplementary materials). Diagonal elements depict histograms of parameter values obtained by bootstrap calibrations (e.g. the first histogram corresponds to the values of v_0). Off-diagonal elements illustrate a dependence structure for each parameter pair. In those figures, a black cross represents the reference value of the specific parameter (obtained from calibration to the whole data set) and by a red star we depict the bootstrap estimate of the value - the closer the two are, the better.

Table 3: Pairwise correlation coefficients of calibrated parameters for the FSV model and the data from 15/5/2015.

	ν_0	κ	θ	σ	ρ	λ	μ_J	σ_J	H
λ	-0.4622	0.5778	-0.7648	0.3080	-0.3574	--	-0.0637	-0.3983	-0.3600
μ_J	0.3347	-0.3603	0.3055	-0.3131	-0.0725	-0.0637	--	-0.1080	0.2063
σ_J	0.4027	-0.3974	0.4286	-0.1420	0.0986	-0.3983	-0.1080	--	0.2894
H	0.7134	-0.7859	0.6600	-0.6059	0.4673	-0.3600	0.2063	0.2894	--

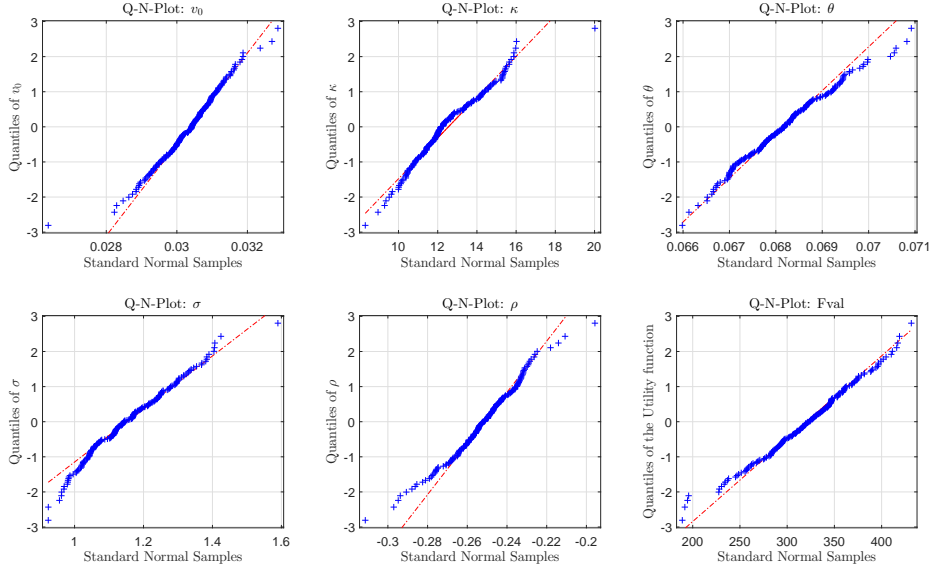


Figure 4: Q-N plots for Heston model parameters ($\nu_0, \kappa, \theta, \sigma, \rho$) and corresponding values $Fval$ of the calibration utility function (15/5/2015).

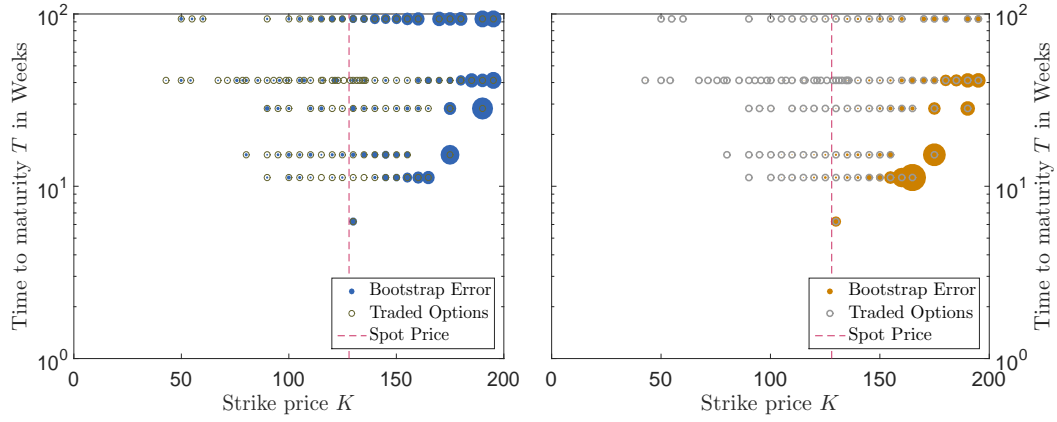
Variation in $C^{\Theta^\dagger, i}$

In Figure 5, the bootstrap relative errors (7) and the bootstrap variances (8) for every call option are shown. The structure of the errors appears similar for all three models – a result that fits very good to the overall calibration errors for the considered data set. The highest errors appear for all three models for the OTM options, especially if the strike price is greater than 150 USD. For all data sets the lowest values of the bootstrap relative error were obtained by the FSV model.

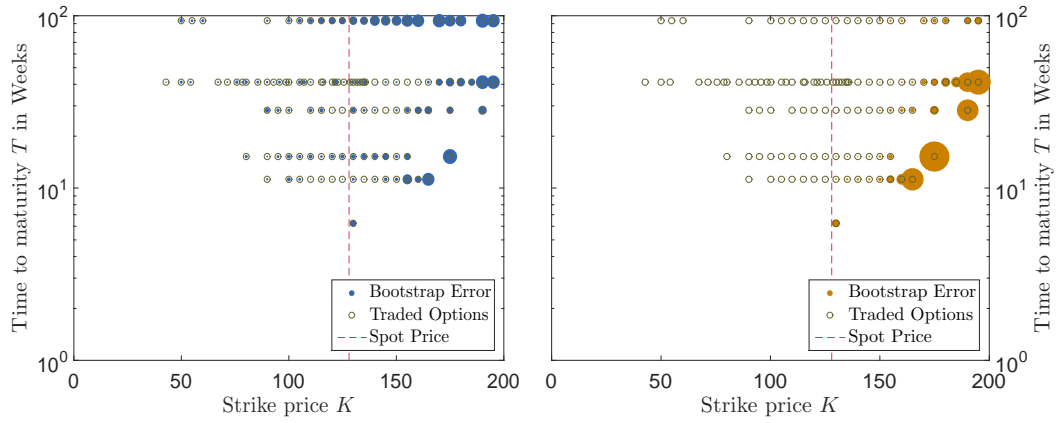
The variance error measure V_j shows for all three models the same structure in the $K \times T$ plane, but values differ strongly. In the data set from 1/4/2015, the measured variances were the lowest for the FSV model again, whereas the Bates model provided us with the worst values. Surprisingly, Bates model was clearly outperformed for all considered data sets. On the other hand results for Heston and FSV slightly differed for the other data sets. We refer a reader to the corresponding figures in supplementary materials.

4.1 Sensitivity analysis

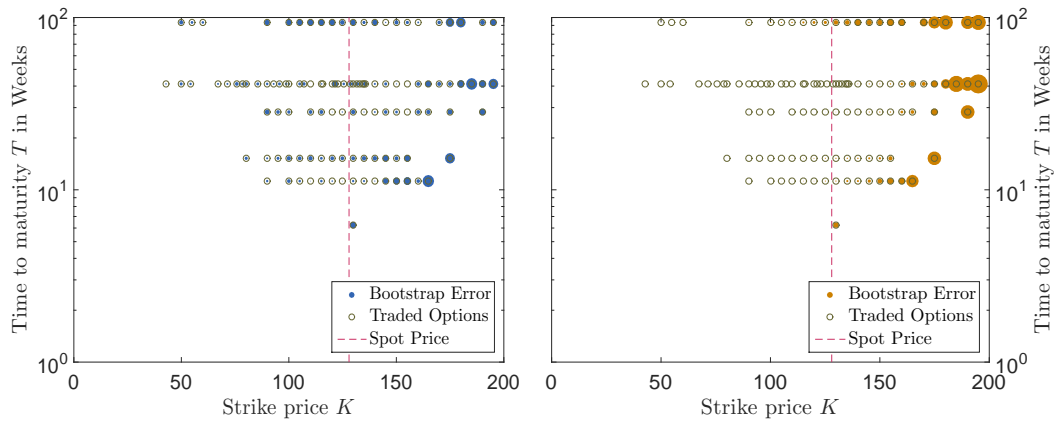
In this section we would like to inspect parameter reducing possibilities for the Bates and the FSV model. We check whether the jump-intensity λ plays a crucial role in obtaining good error measures for the Bates model calibration. If we fix $\lambda = 0$ we would obtain the standard Heston model. Similarly we proceed with the FSV model, where we inspect if we can profit from setting $H > 0.5$, unlike formally fixing $H = 0.5$ to obtain the Bates model.



(a) Heston model



(b) Bates model



(c) FSV model

Figure 5: Calibration errors and variance of the obtained option price surface in $K \times T$ plane. On the left, we depict bootstrap relative errors BRE_j for all bootstrap calibrations w.r.t. 1/4/2015 data set by filled circles with diameter proportionate to the error. On the right, the variance V_j of each option price is illustrated - as before, a diameter of a specific filled circle is proportionate to the option price variance. For results on other data sets see the supplementary materials.

Table 4: Importance of λ for calibrations Apple Inc. stock on all four datasets. Hypothesis 0 denotes we were unable to reject the null hypothesis and vice versa for 1.

Data sets	1/4/2015	15/4/2015	1/5/2015	15/5/2015
Hypothesis	1	1	1	1
p-value	1.30%	0.43%	8.45e-12%	3.56%

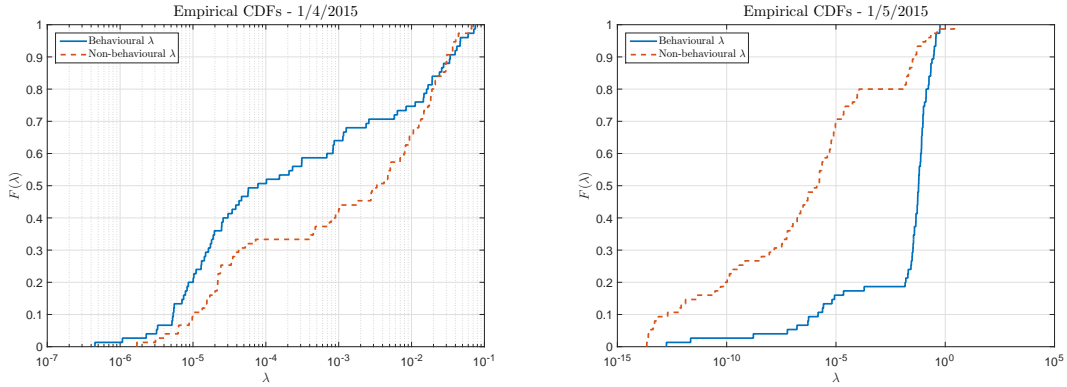


Figure 6: Empirical cumulative distribution functions of both sets (λ).

Importance of jumps

For all available datasets we managed to reject the null hypothesis that both the behavioural and non-behavioural sets of λ are from the same distribution with 5% level of significance. In Table 4 we also display the p-values obtained from the 2-sample Kolmogorov-Smirnov test. These are the maximal levels of significance that would lead to not rejecting the null hypothesis. Hence, we are able to conclude the similar result as in Campolongo, Cariboni, and Schoutens (2006) - the jump term is of significant help for calibration trials. In our case the conclusion is drawn from the real market data and using the Monte-Carlo filtering technique introduced in Section 3. However, it is worth mentioning that this technique identifies input parameters which influence extremes in the output (quality of calibration fit) and hence slightly differs from the classical variance-based sensitivity analysis.

We observe that the calibrated λ 's can take quite small values, but as was shown in Campolongo, Cariboni, and Schoutens (2006), even in that case, the jumps might effect option prices significantly, especially for out-of-the-money contracts.

Sensitivity of the calibration with respect to the Hurst parameter

Following the procedure of Monte-Carlo filtering for jump intensity λ we are interested in the importance of the Hurst parameter. Since for $H = 0.5$ one gets a standard stochastic volatility model with jumps, if we are able to conclude that calibration of H is crucial to obtain a good market fit, then we get a justification of the approximative fractional model which is in-line with the long-memory phenomenon of realized volatility time series.

We were able to reject the null hypothesis for data sets from 1st April and May and also from 15th May. P-values were quite small (see Table 5) for these data sets, unlike for the data from 15th April. In this case were not able to reject the null hypothesis at any reasonable level of significance and hence we cannot make any conclusion regarding this data set.

Table 5: Importance of H for calibrations Apple Inc. stock on all four datasets. Hypothesis 0 denotes we were unable to reject the null hypothesis and vice versa for 1.

Data sets	1/4/2015	15/4/2015	1/5/2015	15/5/2015
Hypothesis	1	0	1	1
p-value	2.78e-10%	30.00%	5.08e-03%	2.17%

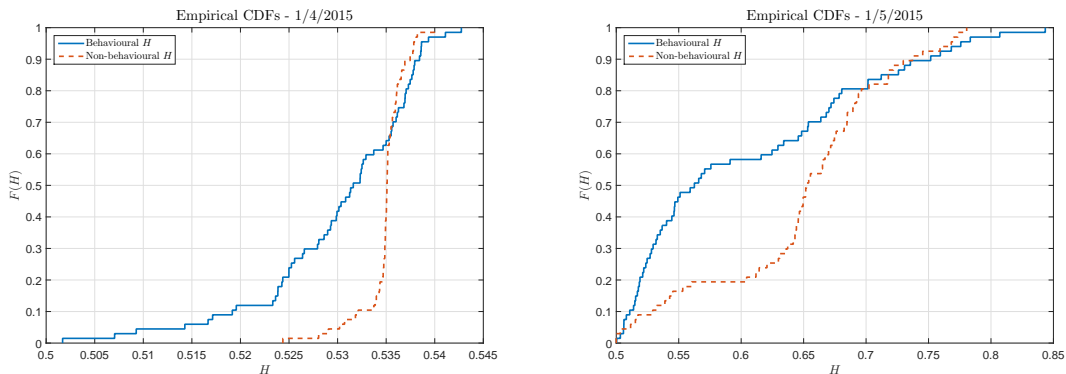


Figure 7: Empirical cumulative distribution functions of both sets (H).

5 Conclusion

In this paper we performed the robustness and sensitivity analysis of several continuous-time stochastic volatility models (Heston, Bates and FSV model) with respect to market calibration. Using the bootstrap method we calibrated the model parameters 200 times and we compared all three models with respect to the variation in model parameters and in bootstrapped option prices.

The bootstrap relative errors of all three models (Figure 5, data from 1/4/2015) are qualitatively similar – the best errors are achieved by FSV model and the worst results by Heston model for all data sets. One can observe higher errors for OTM options ($K > S_0$). As for the bootstrap variances, the structure remains similar for all three models, but qualitatively the results differ strongly. Option prices (and hence market errors) obtained by Bates model have the largest variance with respect to changing data structure. Therefore, the Bates model appears to be the least robust. For 1/4/2015 data set, we retrieved the best bootstrap errors and lowest variances thereof by the FSV model. The Heston model can achieve lower variances (e.g. 15/5/2015), but bootstrap errors were greater compared to the FSV approach. From scatterplot matrices depicted in Figures 2 and 3 we can observe that the histograms of Bates and FSV model differ especially for parameters κ and σ . It is worth to mention that considering different parameter bounds (cf. Figure 1b) may lead to different calibration results, with values of some of the calibrated parameters close to the boundary. Since κ is the parameter with the most correlations, we performed all the tests with relatively high upper bound ($\kappa \leq 100$). In the scatterplot matrices one can further see non-statistical connections of jump parameters, especially for the Bates model. To avoid this, one could fix one jump parameter for the calibration process (e.g. λ).

In Figure 4 we can observe that the calibrated parameters for the Heston model are almost normally distributed unlike for the other models. One should be careful especially with normality assumptions of Θ . Additionally the calibrated parameters cannot be modelled as independent random variables (see Figures 2 and 3), therefore standard sensitivity analysis tests are not suitable in this context. For this reason we used the Monte-Carlo filtering technique to show the importance of the jumps intensity λ in the Bates model and the importance of long memory parameter H in the FSV model. As for the jumps, in all four considered data sets we were able to conclude that considering jumps (non-zero λ) in a model plays a significant role in calibration to real market data. Even small values of λ can effect the call prices, especially for the out-of-the-money contracts. We

could say that calibration of the fractionality parameter H is important only in three cases out of four.

Recently, [Mrázek, Pospíšil, and Sobotka \(2016\)](#) studied the calibration task for FSV model and compared it to the Heston case with respect to in- and out-of-sample errors on equity index data sets. Our study confirms that the approximative fractional model can outperform other studied SV models (see [Table 1](#)). Moreover, we show that this approach is more robust with respect to the uncertainty in the data structure, especially when compared to the other jump-diffusion model. However, it is surprising that an additional parameter (Hurst parameter H) lead to smaller bootstrap variance. We are also able to draw the conclusion that jumps can also lead to decreased robustness (Bates model), so the importance of jump terms discussed in [Campolongo, Cariboni, and Schoutens \(2006\)](#) can effect model performance in a negative manner as well.

5.1 Further issues

As mentioned in the introduction, we have considered only a long-memory regime ($H > 0.5$) of the FSV model, due to technical restrictions of the pricing solution. [Bayer, Friz, and Gatheral \(2016\)](#) have shown that a simple rough paths volatility model can perform surprisingly well even for short maturities, unlike the standard diffusion volatility models without jumps. This observation was also supported by [Fukasawa \(2011\)](#), who has shown a jump-like behaviour of the rough volatility model. Incorporating a rough volatility regime ($H < 0.5$) could also improve robustness of the model in terms of criteria introduced in this paper. Verification of this hypothesis is still due to a further research. In fact, the proposed methodology can be successfully applied to a rough volatility model as soon as one has an efficient pricing solution.

Funding

This work was supported by the GACR Grant 14-11559S Analysis of Fractional Stochastic Volatility Models and their Grid Implementation.

Acknowledgements

Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures".

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2006). *Econometrics of testing for jumps in financial economics using bipower variation*. J. Financ. Econometrics 4(1), 1–30. ISSN 1479-8409. DOI 10.1093/jffinec/nbi022.
- BATES, D. S. (1996). *Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options*. Rev. Financ. Stud. 9(1), 69–107. DOI 10.1093/rfs/9.1.69.
- BAUSTIAN, F., MRÁZEK, M., POSPÍŠIL, J., AND SOBOTKA, T. (2017). *Unifying pricing formula for several stochastic volatility models with jumps*. Appl. Stoch. Models Bus. Ind. 33(4), 422–442. ISSN 1524-1904. DOI 10.1002/asmb.2248.
- BAYER, C., FRIZ, P., AND GATHERAL, J. (2016). *Pricing under rough volatility*. Quant. Finance 16(6), 887–904. ISSN 1469-7688. DOI 10.1080/14697688.2015.1099717.
- BENHAMOU, E., GOBET, E., AND MIRI, M. (2010). *Time dependent Heston model*. SIAM J. Finan. Math. 1(1), 289–325. ISSN 1945-497X. DOI 10.1137/090753814.
- BIANCHETTI, M., KUCHERENKO, S., AND SCOLERI, S. (2015). *Pricing and risk management with high-dimensional quasi Monte Carlo and global sensitivity analysis*. Wilmott 2015(78), 46–70. ISSN 1541-8286. DOI 10.1002/wilm.10434.
- BLACK, F. S. AND SCHOLES, M. S. (1973). *The pricing of options and corporate liabilities*. J. Polit. Econ. 81(3), 637–654. ISSN 0022-3808. DOI 10.1086/260062.
- CAMPOLONGO, F., CARIBONI, J., AND SCHOUTENS, W. (2006). *The importance of jumps in pricing European options*. Reliab. Eng. Syst. Safe. 91(10), 1148–1154. ISSN 0951-8320. DOI 10.1016/j.res.2005.11.016.
- CARR, P. AND WU, L. (2003). *What type of process underlies options? A simple robust test*. The Journal of Finance 58(6), 2581–2610. ISSN 1540-6261. DOI 10.1046/j.1540-6261.2003.00616.x.
- CHERNICK, M. R. (2008). *Bootstrap methods: a guide for practitioners and researchers*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, second edn. ISBN 978-0-471-75621-7.
- CHRISTOFFERSEN, P., HESTON, S., AND JACOBS, K. (2009). *The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well*. Manage. Sci. 55(12), 1914–1932. ISSN 0025-1909. DOI 10.1287/mnsc.1090.1065.
- CREEL, M. AND KRISTENSEN, D. (2015). *ABC of SV: Limited information likelihood inference in stochastic volatility jump-diffusion models*. J. Empir. Financ. 31, 85–108. ISSN 0927-5398. DOI 10.1016/j.jempfin.2015.01.002.
- DE MARCO, S. AND MARTINI, C. (2012). *The term structure of implied volatility in symmetric models with applications to Heston*. Int. J. Theor. Appl. Finance 15(04), 1250026. ISSN 0219-0249. DOI 10.1142/S0219024912500264.
- DETLEFSEN, K. AND HÄRDLE, W. K. (2007). *Calibration risk for exotic options*. J. Derivatives 14(4), 47–63. ISSN 1074-1240. DOI 10.3905/jod.2007.686422.
- DUFFIE, D., PAN, J., AND SINGLETON, K. (2000). *Transform analysis and asset pricing for affine jump-diffusions*. Econometrica 68(6), 1343–1376. ISSN 0012-9682. DOI 10.1111/1468-0262.00164.
- EFRON, B. (1979). *Bootstrap methods: another look at the jackknife*. Ann. Statist. 7(1), 1–26. ISSN 0090-5364.
- EFRON, B. (1982). *The jackknife, the bootstrap and other resampling plans*, vol. 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa. ISBN 0-89871-179-7.
- ELICES, A. (2008). *Models with time-dependent parameters using transform methods: application to Heston’s model*. Available at arXiv: <https://arxiv.org/abs/0708.2020>.
- FUKASAWA, M. (2011). *Asymptotic analysis for stochastic volatility: martingale expansion*. Finance Stoch. 15(4), 635–654. ISSN 1432-1122. DOI 10.1007/s00780-010-0136-6.
- HÁJEK, J., ŠIDÁK, Z., AND SEN, P. K. (1999). *Theory of rank tests*. Probability and Mathematical Statistics. Academic Press, Inc., San Diego, CA, second edn. ISBN 0-12-642350-4.
- HESTON, S. L. (1993). *A closed-form solution for options with stochastic volatility with applications to bond and currency options*. Rev. Financ. Stud. 6(2), 327–343. ISSN 0893-9454. DOI 10.1093/rfs/6.2.327.
- HWANG, E. AND SHIN, D. W. (2014). *A bootstrap test for jumps in financial economics*. Econom. Lett. 125(1), 74–78. ISSN 0165-1765. DOI 10.1016/j.econlet.2014.08.024.
- KIM, P. J. (1976). *The smirnov distribution*. Ann. Inst. Stat. Math. 28(1), 267–275. ISSN 0020-3157. DOI 10.1007/BF02504745.
- LEWIS, A. L. (2000). *Option Valuation Under Stochastic Volatility: With Mathematica code*. Finance Press, Newport Beach, CA. ISBN 9780967637204.
- MIKHAILOV, S. AND NÖGEL, U. (2003). *Heston’s stochastic volatility model - implementation, calibration and some extensions*. Wilmott magazine 2003(July), 74–79.
- MRÁZEK, M., POSPÍŠIL, J., AND SOBOTKA, T. (2016). *On calibration of stochastic and fractional stochastic volatility models*. European J. Oper. Res. 254(3), 1036–1046. ISSN 0377-2217. DOI 10.1016/j.ejor.2016.04.033.
- OSAJIMA, Y. (2007). *The asymptotic expansion formula of implied volatility for dynamic SABR model and FX hybrid model*. DOI 10.2139/ssrn.965265. Available at SSRN: <http://ssrn.com/abstract=965265>.
- POSPÍŠIL, J. AND SOBOTKA, T. (2016). *Market calibration under a long memory stochastic volatility model*. Appl. Math. Finance 23(5), 323–343. ISSN 1350-486X. DOI 10.1080/1350486X.2017.1279977.
- SALTELLI, A., RATTO, M., ANDRES, T., CAMPOLONGO, F., CARIBONI, J., GATELLI, D., SAISANA, M., AND TARANTOLA, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley, Chichester. ISBN 9780470725177.
- SALTELLI, A., TARANTOLA, S., CAMPOLONGO, F., AND RATTO, M. (2004). *Sensitivity analysis in practice: a*

- guide to assessing scientific models. Wiley, Chichester.
- SENGER, O. AND CELIK, A. K. (2013). *A Monte Carlo simulation study for Kolmogorov-Smirnov two-sample test under the precondition of heterogeneity: upon the changes on the probabilities of statistical power and type I error rates with respect to skewness measure*. J. Stat. Econometric Methods 2(4), 1–16. ISSN 2241-0384.
- SHREVE, S. E. (2004). Stochastic calculus for finance. II. Springer Finance. Springer-Verlag, New York.
- YEKUTIELI, I. (2004). *Implementation of the Heston model for the pricing of FX options*. Tech. rep., Bloomberg LP.