An Empirical Study of S&P 500 Index Option Pricing

Nan Yang
Institute of Statistics and Big Data
Renmin University of China

Date: May 9, 2022

1 Introduction

Options were first introduced on the Chicago Options Exchange (CBOE) in the United States. In 1973, with the development of standardized option trading contracts by CBOE, options have experienced a rapid development stage. Nowadays, options trading has become more frequent and become an important risk hedging tool in the capital market.

In terms of the nature of options, rights and obligations are not equal. For example, European call option, the option buyer can purchase certain underlying assets at a specific time (T) in the future. If the asset price at the maturity does not reach the strike price specified in the option, then the option is no longer valuable, then his loss is the cost of purchasing the option at the beginning. However, if the price of the underlying asset exceeds the strike price specified in the option on the maturity date, he can exercise the option, and then the income obtained from the exercise of the option is the difference between the price of the underlying asset and the strike price minus the option price.

In the Black-Scholes world, the volatility is assumed constant. But in reality, options of different strike require different volatilities to match their market prices. This is called the volatility smile. I first calibrate the Heston model by the real option market data. Then, applying the Fast Fourier Transform method, I use the calibrated model and the Black-Scholes model to calculate option prices separately. Next, I leverage the Rooting and Minimize Object Function methods to compute the implied volatilities and visualize them accordingly. Our verification shows that compared with the Black-Scholes model, the Heston model can indeed show the shape of the volatility more accurately.

2 Data Description

2.1 Data Information

IvyDB is a comprehensive database of historical price, implied volatility, and sensitivity information for the entire US listed index and equity options markets. IvyDB includes historical data for all US listed equities and market indices and all US listed index and equity options from 1996 till present.

OptionMetrics compiles the IvyDB data from price information. Interest rate curves, dividend projections, and option implied volatilities and sensitivities are calculated by OptionMetrics using their proprietary algorithms, which are based on standard market conventions.

2.2 Exploratory Data Analysis

First, as for the price of the underlying asset, I obtain data from "IndexPrice1996_2020.csv". The "SP500OptionPrice1996_2020_2" file contains information on standardized (interpolated) options. Currently, this file contains information on at-the-money-forward options with expirations of 10, 30, 60, 91, 122, 152, 182, 273, 365, 547 and 730 calendar days. A standardized option is only included if there exists enough option price data on that date to accurately interpolate the required values. The data sample consists of 6295 daily observations. The price of the underlying asset can be shown as in figure 1.

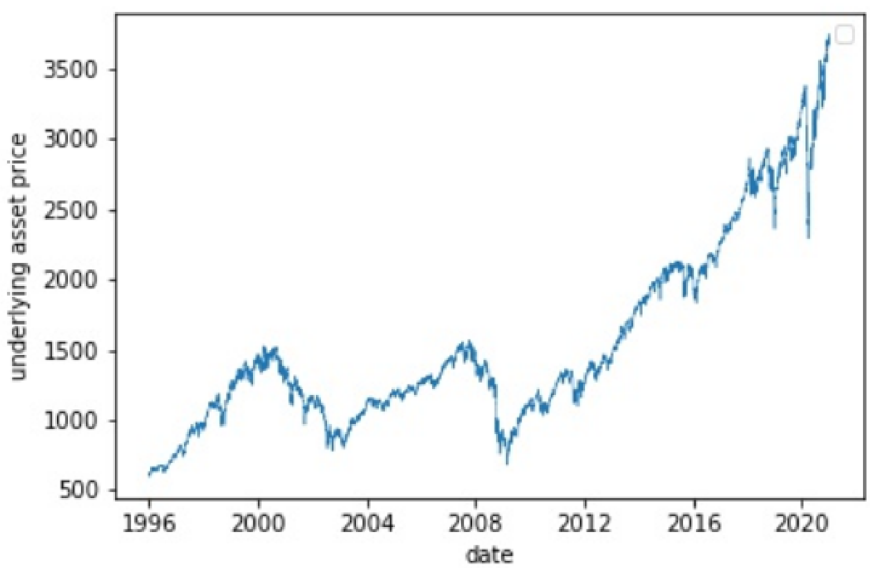

Figure 1: Underlying Asset Price

Secondly, The Volatility_Surface file contains the interpolated volatility surface for each security on each day, using a methodology based on a kernel smoothing algorithm. This file contains information on standardized options, both calls and puts, with expirations of 10,30, 60, 91, 122, 152, 182, 273, 365, 547, and 730 calendar days, at deltas of 0.10, 0.15,0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75,0.80, 0.85, 0.90 (negative deltas for puts). A standardized option is only included if there exists enough option price data on that date to accurately interpolate the required values. After data preprocessing, it contains 2353582 observations. Relationship between volatility and the time to maturity can be seen in figure 2 preliminarily, and relationship between volatility and date can be seen in figure 3.

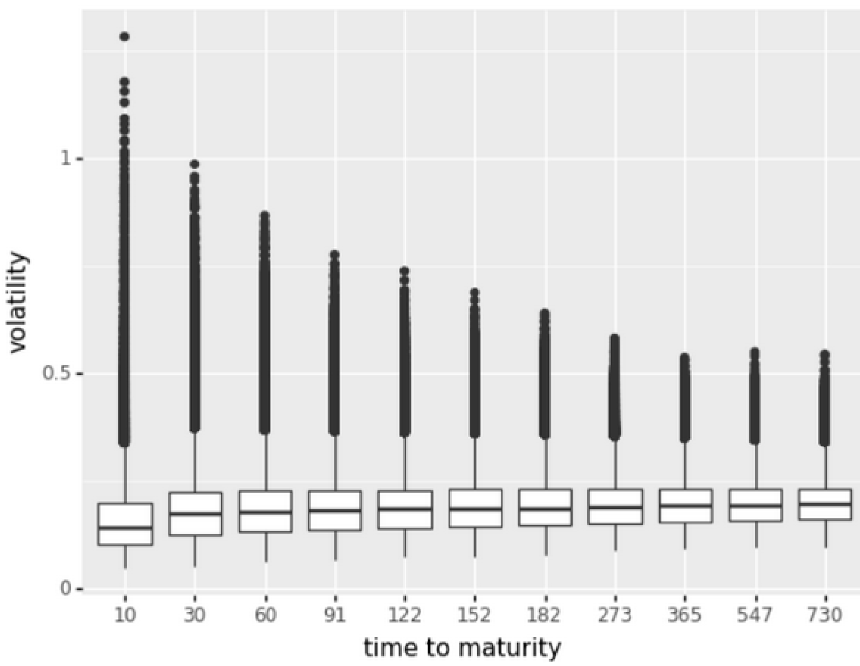

Figure 2: Relationship between volatility and the time to maturity

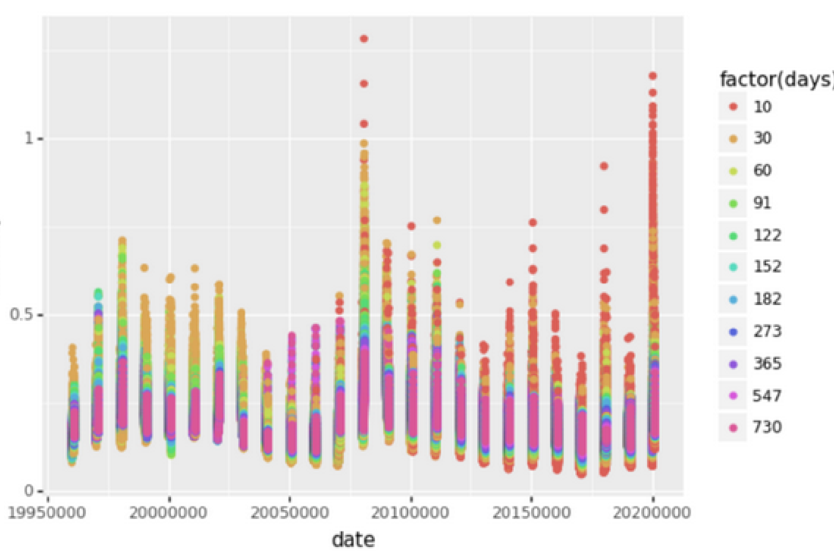

Figure 3: Relationship between volatility and the time to maturity

Set the date to 2020-06-05, and we can get the implied volatility curve as in figure 4.

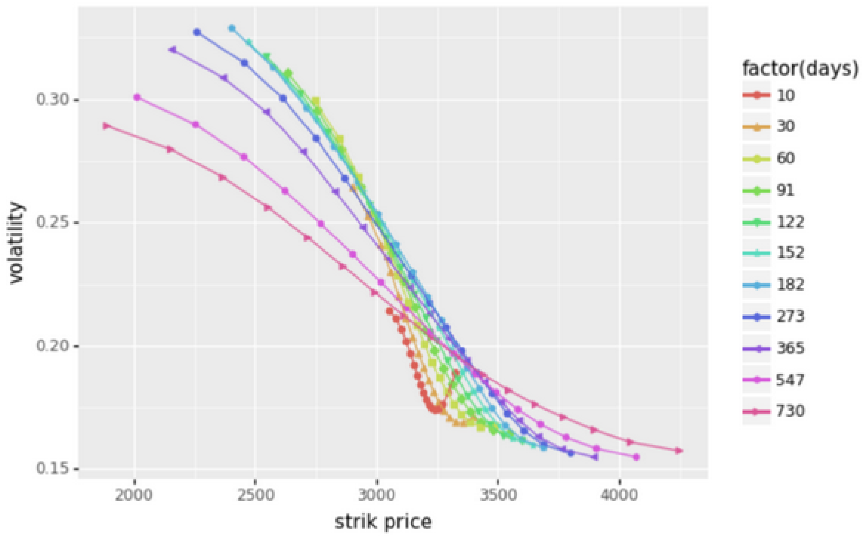

Figure 4: implied volatility curve

Each of the option pricing models used by IvyDB requires a continuously compounded interest rate as input. The "InterestRate1996_2020.csv" file contains the interest rate data for each day and there are 283103 observations.

At last, The"SP500OptionPrice1996_2020_2.csv" file contains the historical price of 21738936 observations, implied volatility, and sensitivity information for the options on an underlying security.

Use "SP500OptionPrice1996_2020_2" data as main option data. It contains useful columns: 'date', 'exdate', 'cp_flag', 'strike_price', 'impl_volatility', 'delta', 'gamma', 'theta', 'vega', 'best_bid', 'best_offer', 'volume'.

From "InterestRate1996_2020" data, obtain interest rate data of every day and every maturity.

"IndexPrice1996_2020" data includes underlying asset price during 1996-2020.

"show_data.ipynb" shows the exploratory data analysis.

"real_data_preprocessing.ipynb" consolidates all data into a dataframe "option_clean_all_df".

3 Models Description

3.1 Black-Scholes Model

In 1973, Black and Scholes proposed the famous B-S model, which initiated a revolution in option pricing. Different from the previous researches, they mainly analyzed the random process of the stock price, obtained a partial differential equation through a series of assumptions and deduction, and then solved the partial differential equation to obtain the pricing formula of European call options.

In the BS model, the dynamics of the stock price S_t under the risk-neutral measure is given by:

$$dS_t = (r - \delta)S_t dt + \sigma S_t dW_t$$

for some constants $r \in \mathbb{R}, \delta \geq 0$, and $\sigma > 0$, and where W_t is a standard Brownian motion.

Let V_t denotes the price of a derivative at time t, then it satisfies the following partial differential equation:

$$0 = \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \delta) S \frac{\partial V}{\partial S} - rV$$

The call option payoff $(x - K)^+$ with strike K and maturity T.

Solving that PDE with boundary condition $V_T = (S_T - K)^+$ with $S_t = S$ gives the following formula for the time-t call option price C:

$$C(S, \sigma, r, \delta, K, \tau) = S^{-\delta\tau}\Phi(d_+) - e^{-r\tau} K \Phi(d_-)$$

where $\tau = T - t$, $d_\pm = (\log(S/K) + (r - \delta)\tau)/(\sigma\sqrt{\tau}) \pm (1/2)\sigma\sqrt{\tau}$ and Φ is the standard Gaussian CDF.

3.2 Heston Model

Heston model assumes that the volatility of the underlying asset is not constant, but a random mean reversion process. This process includes a long-term mean θ of the volatility and a reversion rate κ of the volatility. If the previous volatility is lower than the long-term mean, the model can be adjusted upward at a certain rate.

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{t1}$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_{t2}$$

$$dW_{t1} \cdot dW_{t2} = \rho dt$$

- μ: drift of the stock process
- σ: volatility coefficient of the variance process
- ρ: correlation between W_{t1} and W_{t2} i.e.

If the parameters obey the following condition $2\kappa\theta > \sigma^2$(known as the Feller condition) then the process v_t is strictly positive.

There are several ways to calculate the option price such as Lewis method, Fourier-inversion method, etc. However, when we want to compute the price of a large number of options with the same maturity, the previous methods are no more efficient. A solution is to take advantage of the FFT (Fast Fourier Transform) algorithm to reduce the computational cost.

3.2.1 FFT Method

Recall that the real part of a complex number is linear, i.e. for $a, b \in \mathbb{R}$ we have $Re[az_1 + bz_2] = aRe[z_1] + bRe[z_2]$, and the real part of an integral is the integral of the real part.

Thanks to this property, the integral pricing formula can be written in the following form:

$$C_0 = S_0 - \frac{\sqrt{S_0 K}e^{-rT}}{\pi} Re\left[\int_0^\infty e^{iuk}\phi_T\left(u - \frac{i}{2}\right)\frac{1}{u^2+\frac{1}{4}}du\right].$$

At this point we can discretize the integral. we use the Simpson rule to deal with it.

The domain of integration is truncated in $[A, B] = [0, B]$, and divided in N steps of size $\Delta x = \frac{B}{N}$.

We have that $x_0 = 0$ and $x_n = x_0 + n\Delta x$, for $n = 0, 1, 2, ..., N$. The integral is evaluated in $N + 1$ points $f(x_n) = f_n$.

The Simpson rule is a 3 points rule that approximates the integral as:

$$\int_{x_0}^{x_2} f(x)\,dx \approx \frac{\Delta x}{3}\left[f_0 + 4f_1 + f_2\right].$$

If we sum over the integration domain $[x_0, x_{N-1}]$ we get

$$\int_{x_0}^{x_{N-1}} f(x)\,dx \approx \frac{\Delta x}{3}\sum_{n=0}^{N-1} w_n\, f_n.$$

with $w_n = 1$ for $n = 0$ and $n = N - 1$, and $w_n = 4$ for n odd, and $w_n = 2$ for n even. Notice that we are not considering the last point!

Define a set of N values $k_j \in [-b, b]$, such that $k_j = -b + j\Delta k$, for $j = 0, 1, 2, ..., N - 1$. We define the step Δk as

$$\Delta k := \frac{2\pi}{B} = \frac{1}{\Delta x}\frac{2\pi}{N}$$

such that we can obtain the value $b = \frac{N\Delta k}{2}$. The integral in the pricing function is:

$$\begin{aligned} I(k_k) &= \int_0^\infty e^{ixk_j}\phi_T\left(x - \frac{i}{2}\right)\frac{1}{x^2+\frac{1}{4}}dx \quad \approx \quad \frac{\Delta x}{3}\sum_{n=0}^{N-1} w_n\, e^{ik_j x_n}\phi_T\left(x_n - \frac{i}{2}\right)\frac{1}{{x_n}^2+\frac{1}{4}} \\ &= \frac{\Delta x}{3}\sum_{n=0}^{N-1} w_n\, e^{i(-b+j\Delta k)n\Delta x}\phi_T\left(x_n - \frac{i}{2}\right)\frac{1}{{x_n}^2+\frac{1}{4}}. \\ &= \frac{\Delta x}{3}\sum_{n=0}^{N-1} e^{i2\pi j\frac{n}{N}}\; w_n e^{-ibn\Delta x}\phi_T\left(x_n - \frac{i}{2}\right)\frac{1}{{x_n}^2+\frac{1}{4}}. \end{aligned}$$

4 Reslut

4.1 Black-Scholes Model

"BS_estimation.ipynb" calculates implied volatility by Newton's method. Through empirical research, we can see that the actual implied volatility curve is not a straight line, but a curve with a curved or inclined shape as shown in figure 5.

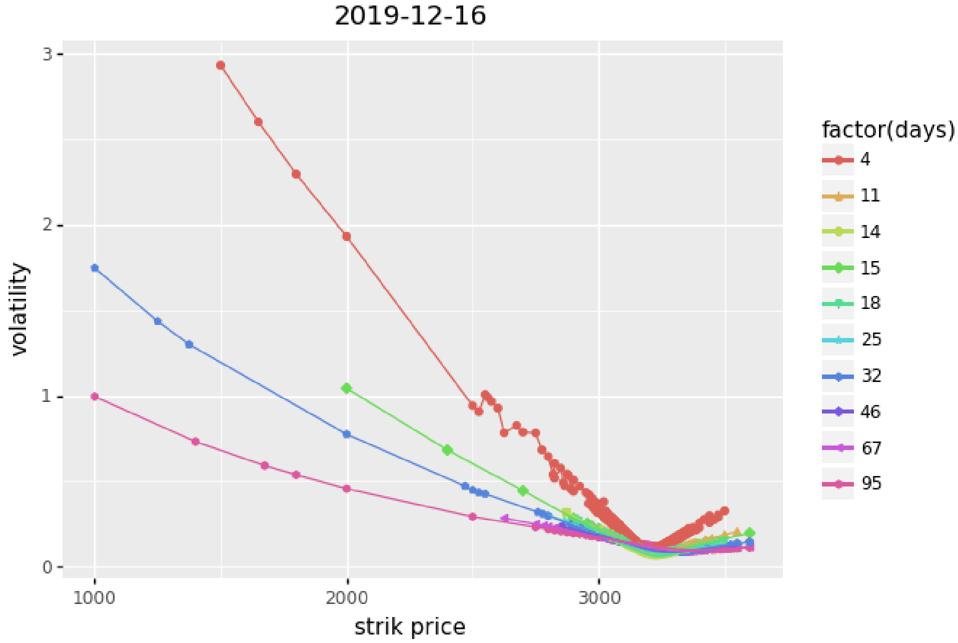

Figure 5: implied volatility curve 2019-12-16

4.2 Heston Model

If we call Θ the set of parameters, the goal is to find the optimal parameters Θ^* that minimize the following objective function:

$$\sum_{i=1}^{N} w_i \Big(P_i - f(K_i|\Theta) \Big)^2$$

where P_i are the market prices, f is the pricing function and w_i are weights.

The Feller condition is hardly satisfied in the market, mainly because having $\kappa\theta > \frac{1}{2}\sigma^2$ implies a high mean reversion which reduces the volatility of the stochastic volatility σ^2. The vol is responsible for the convexity of the smile. In order to increase the convexity to match the empirical smile, we have to increase σ^2 violating the Feller condition. However this is not really an issue, at the moment the v_t reaches 0 we have a positive drift $dv_t = \kappa\theta dt$ that will push instantly the process away from 0.

If we ignore the feller condition, the calibration result can be seen in figure 6.

	rho	sigma	theta	kappa	v0
Initial guess	-0.6000	1.0000	0.0400	2.5000	0.0400
Calibrated	-0.5292	3.0542	0.4019	8.3793	0.0002

Figure 6: calibrated parameters: ignore feller condition

If we take feller condition into account, the calibration result can be seen in figure 7.

	rho	sigma	theta	kappa	v0
Initial guess	-0.4000	1.1000	0.1000	0.6000	0.020
Calibrated	-0.4587	1.4455	0.8294	0.5845	0.027

Figure 7: calibrated parameters: with feller condition

These two cases' comparison is in figure 8, figure 9 and figure 10.

	Mean Absolute (Pricing) Error
Heston	0.0325
Heston with Feller.	0.0581

Figure 8: error

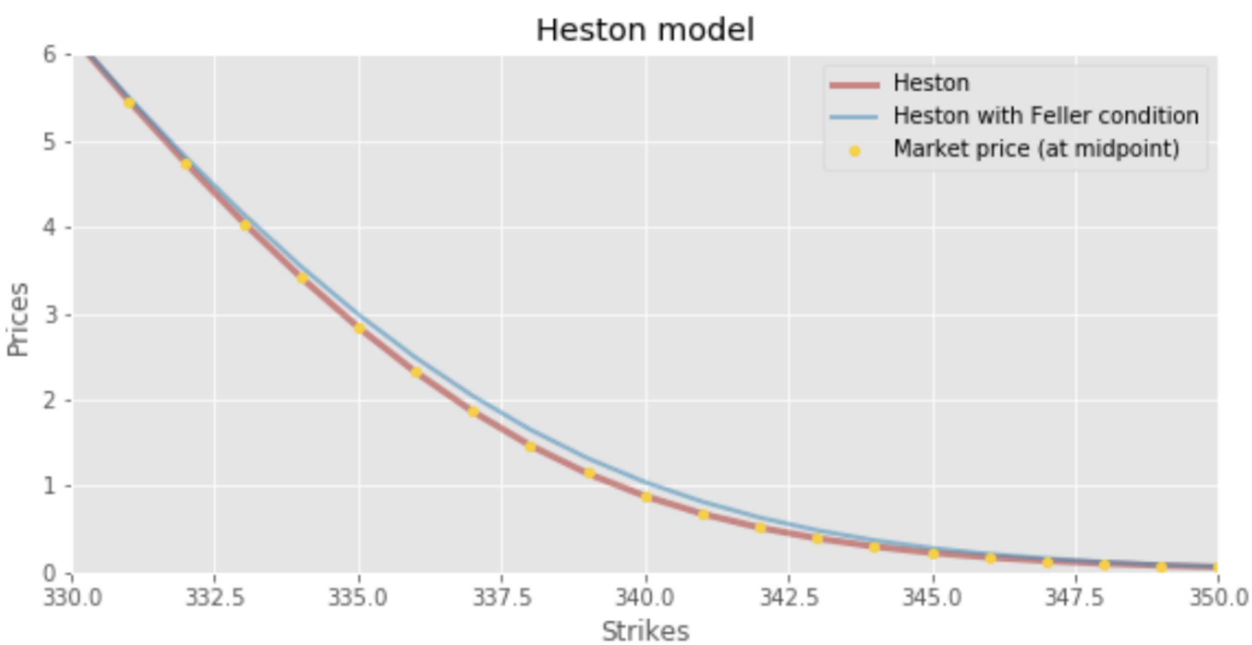

Figure 9: Comparison of the Two Cases

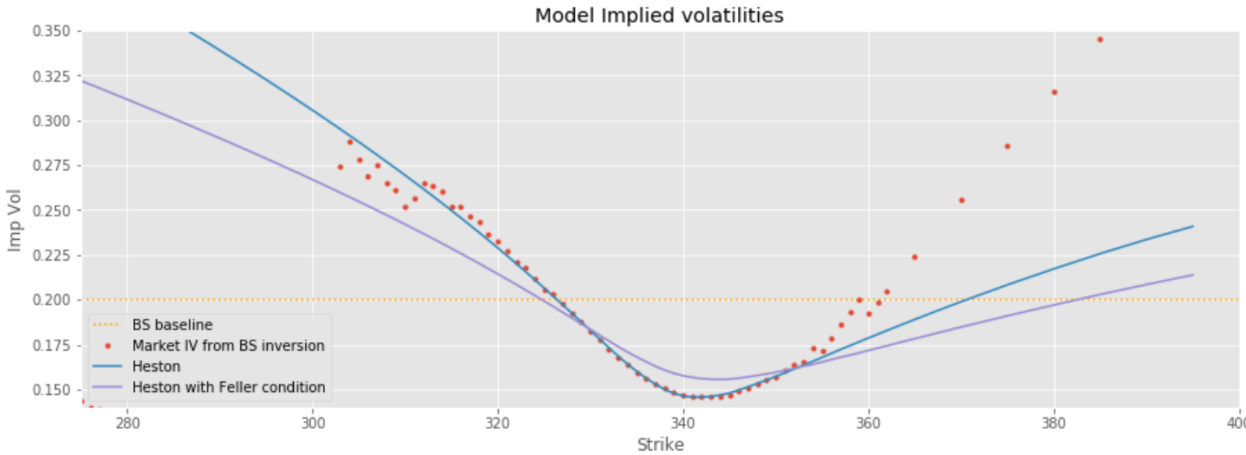

Figure 10: Implied Volatilityurves

In general, Heston model is indeed more advantageous than the classic BS model in showing the real smile curve because of its more reasonable assumptions.

For the two types of Heston models, one is with Feller condition, the other is without Feller condition. In the pricing aspect and smaller intervals, the performance of Heston model without Feller condition is better than the other. Besides, if we consider the perspective of implied volatility, we also notice that the Heston model without Feller has a better reduction of the real situation.

References

Ackerer, Damien, Natasa Tagasovska, and Thibault Vatter, "Deep smoothing of the implied volatility surface," *Advances in Neural Information Processing Systems*, 2020, *33*, 11552–11563.

Bali, Turan G, Nusret Cakici, Fousseni Chabi-Yo, and Scott Murray, "The risk-neutral distribution of option returns," *Georgetown McDonough School of Business Research Paper*, 2017, (2902209).

Fabozzi, Frank J, Radu Tunaru, and George Albota, "Estimating risk-neutral density with parametric models in interest rate markets," *Quantitative Finance*, 2009, *9* (1), 55–70.

Heston, Steven L, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *The review of financial studies*, 1993, *6* (2), 327–343.