VALUING EXOTIC OPTIONS AND ESTIMATING MODEL RISK

Jay Cao, Jacky Chen, John Hull, Zissis Poulos*

Joseph L. Rotman School of Management

University of Toronto

January 2021

This Version: March 2021

Abstract

A common approach to valuing exotic options involves choosing a model and then determining its parameters to fit the volatility surface as closely as possible. We refer to this as the model calibration approach (MCA). This paper considers an alternative approach where the points on the volatility surface are features input to a neural network. We refer to this as the volatility feature approach (VFA). We conduct experiments showing that VFA can be expected to outperform MCA for the volatility surfaces encountered in practice. Once the upfront computational time has been invested in developing the neural network, the valuation of exotic options using VFA is very fast. VFA is a useful tool for the estimation of model risk. We illustrate this using S&P 500 data for the 2001 to 2019 period.

Key words: exotic options, volatility surfaces, neural networks

*We are grateful to Ryan Ferguson and Ivan Sergienko for helpful comments.

VALUING EXOTIC OPTIONS AND ESTIMATING MODEL RISK

Jay Cao, Jacky Chen, John Hull, Zissis Poulos

1. Introduction

For many underlying assets, there is little uncertainty about the pricing of plain vanilla European and American options. Quotes and trades by market participants provide points on the volatility surface. Interpolating between these points as necessary, a trader can derive a reasonable estimate of the implied volatility appropriate for any new plain vanilla European or American option that is of interest.

The pricing of exotic options is more challenging. All derivatives dealers want their prices to be consistent with the underlying asset's volatility surface, but the model that is assumed in conjunction with this volatility surface is liable to vary from dealer to dealer. This gives rise to model risk which has assumed increasing importance for both dealers and their regulators. SR 11-7, which was published by the U.S. Board of Governors of the Federal Reserve System in 2011, and is widely used by other regulators throughout the world, requires financial institutions to identify the sources of model risk and carry out frequent model validation exercises.[1]

A natural approach for using a chosen model to value an exotic option is to search for parameters of the model that fit the current volatility surface as closely as possible and then proceed to price the option using those model parameters. This approach, which we refer to as the "model calibration approach" or MCA, has been used by researchers such as Horvath et al (2021) and Liu et al (2019). A drawback of the approach is that some of the points on the volatility surface are likely to be more important than others for any particular exotic option that is considered. It is of course possible to vary the weights assigned to different points on the volatility surface according to the instrument being valued. However, it is difficult to determine in advance what these weights should be. As a result, the points are usually given equal weight when model parameters are determined. This potential drawback of MCA has led us to investigate an alternative approach.

[1] See Federal Reserve System (2011).

We will refer to the alternative approach as the "volatility feature approach" or VFA. It involves randomly generating many sets of parameters for (a) the model under consideration and (b) the exotic option under consideration. For each set of parameters, a volatility surface and a price for the exotic option are calculated. A neural network is then constructed. The features input to the network are those defining the volatility surface and the exotic option parameters. The target is the exotic option price. The model parameters are not inputs to the neural network. The MCA and VFA valuation methods are illustrated in Figures 1 and 2.

Figure 1: The MCA approach to the valuation of exotic options

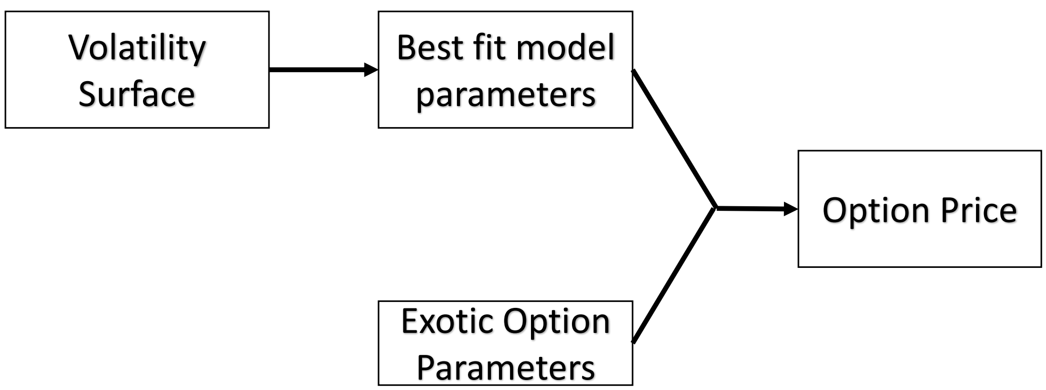
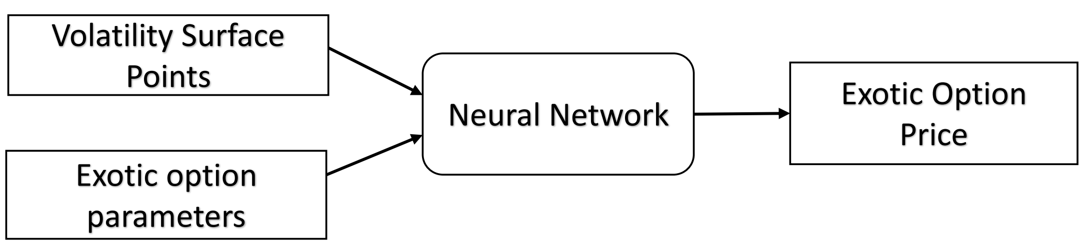


Figure 2: The VFA approach to the valuation of exotic options



A potential advantage of VFA over MCA is that all information about the current shape of the volatility surface is retained. The neural network can learn how the importance of different points on the volatility surface depend on an exotic option's parameters. To avoid losing volatility information an analyst can use the implied volatility function model. This is a one-factor model that is designed to exactly fit the volatility surface by making volatility a function of the asset price and time.[2] (VFA then has no potential advantages over MCA.) However, it has been shown that the model can give poor prices is for some exotic options in some circumstances.[3] This is because the volatility surface at time *t* conditional on the asset price is liable to be quite different from that at time zero and can have unrealistic properties. As a result, most practitioners use models with stationary parameters that involve more than one factor or jumps (or both). The initial fit to the volatility surface is then imperfect but volatility surface movements are plausible. For these models, some information about the volatility surface is inevitably lost when MCA is used.

Many quite different shapes for the volatility surface for an asset are observed in practice. MCA is liable to lead to the same (or very similar) best fit model parameters for two quite different volatility surfaces. VFA has the potential to improve MCA by retaining all points on the volatility surface as model inputs. VFA also provides fast valuation. As a first step, it is necessary to devote computational resources to creating a data set and training the neural network. But, once this has been done, VFA's processing time is several orders of magnitude faster than methods involving the use of Monte Carlo simulation because it involves nothing more than working forward through the network. This point has been made by authors such as Ferguson and Green (2018).

In this paper we first carry out experiments to compare the performance of MCA and VFA. We assume that the world is governed by a particular model, A, and that the model used for pricing a particular type of exotic option is an alternative model, B. The "true" prices of a panel of exotic options are determined using model A. The average pricing errors of VFA and MCA are about the same in these experiments. However, a key point is that the performance of VFA relative to MCA increases significantly as the MCA calibration error increases. For the volatility surfaces encountered in practice, the calibration error is often much greater than in our idealized experiments. This provides evidence suggesting that VFA is a better tool than MCA.

We propose a methodology for quantifying model risk. This involves using VFA with a number of different models in conjunction with volatility surfaces that have been observed in the past. The differences between the prices obtained are measures of model risk. We illustrate the approach with two models that are extensions of Heston (1993). The first model was proposed by Bates (1996) and adds jumps in the asset price to the process assumed by Heston. The second is a version of the rough volatility model, known as rough Heston. We use S&P 500 volatility surfaces between 2001 and 2019 to illustrate our approach and consider three different types of exotic call options:

[2] See, for example, Dupire (1994) for a description of the model.
[3] See, for example, Hull and Suo (2002)

i. Barrier options where the barrier level is greater than either the initial asset price or the strike price and the option ceases to exist if the barrier is reached.
ii. Asian options where the payoff is the excess of the average asset price during the life of the option over the strike price if this is positive and zero otherwise.
iii. Lookback options where the payoff is the excess of the highest price during the life of the option over the strike price if this is positive and zero otherwise.

The rest of the paper is organized as follows. Section 2 reviews the models that will be used. Section 3 presents the results from our controlled experiments comparing MCA and VFA. Section 4 shows that points on the volatility surface do not have equal importance when exotic options are being valued. Section 5 explains the approach we used to generate standard points on the S&P 500 volatility surfaces from data on the implied volatilities of the options that trade. Section 6 illustrates our procedure for estimating model risk. Conclusions are in Section 7.

2. The Models

In this section we review the models that we use in our analyses. Define

S_t: Asset price at time t

r: Risk-free rate (assumed constant)

δ: Yield on asset (assumed constant)

The Heston (1993) model is

$$dS_t = (r - \delta)S_t dt + S_t\sqrt{V_t}dz_t$$

$$dV_t = a(V_L - V_t)dt + \xi\sqrt{V_t}\, dw_t$$

The variance rate, V_t, reverts to a level V_L at rate a. Superimposed on this mean reversion is a "volatility of volatility" variable, ξ. The dz_t and dw_t are Wiener processes with correlation ρ.

One extension of the Heston model is Bates (1996). This incorporates jumps in the process for S_t. The model is:

$$dS_t = \left(r - \delta - \lambda\bar{k}\right)S_t dt + S_t\sqrt{V_t}dz_t + kdq$$

$$dV_t = a(V_L - V_t)dt + \xi\sqrt{V_t}\,dw_t$$

The jumps are assumed to occur randomly at rate λ per unit time. The percentage jump size, k, is assumed to have the property that ln(1+k) is normal with mean μ and standard deviation σ. The average jump size $\bar{k}$ is $e^{(\mu+\sigma^2)} - 1$. Merton's (1976) model can be viewed as a particular case of Bates (1996) where V_t is constant.

The rough Heston model, as described by for example El Euch et al (2019), is another extension of the basic Heston model. In this, the process for the variance, V_t , involves a fractional Brownian motion rather than a regular Brownian motion. This model can be written as

$$dS_t = (r - \delta)S_t dt + S_t\sqrt{V_t}dz_t$$

$$V_t = V_0 + \frac{1}{\Gamma(H+0.5)}\int_0^t (t-s)^{H-0.5}\left((a(V_L - V_s)ds + \xi\sqrt{V_s}dw_s\right)$$

where the correlation between dz_t and dw_t is ρ, H is the Hurst exponent, and Γ is the gamma function. When H = 0.5 the model reduces to Heston (1993). When H < 0.5 the correlation between volatility movements in successive periods of time is negative while when H > 0.5 this correlation is positive. Gatheral et al (2018) produce results for indices showing that the model reflects market data well when H is between 0.06 and 0.20.

We use a close approximation to rough Heston: the lifted Heston model proposed by Jaber (2019). This has the advantage that it involves only regular Brownian motions and, as a result, is much faster for the computation of exotic option prices. The version of the Jaber (2019) model we use is

$$dS_t = (r - \delta)S_t dt + S_t\sqrt{V_t}dz_t$$

$$V_t = V_0 + aV_L\sum_{i=1}^{n}\frac{c_i}{x_i}(1 - e^{-x_i t}) + \sum_{i=1}^{n} c_i U_{i,t}$$

$$dU_{i,t} = -x_i U_{i,t}dt + \xi\sqrt{V_t}dw_t$$

where the correlation between the Wiener processes, dz_t and dw_t , is ρ. The $U_{i,t}$ are initially zero and revert to zero at different speeds x_i.

As recommended by Jaber, we set n = 20 and

$$c_i = \frac{(\beta^{1-\alpha} - 1)\beta^{(\alpha-1)(1+\frac{n}{2})}}{\Gamma(\alpha)\Gamma(2-\alpha)}\beta^{(1-\alpha)i}$$

$$x_i = \left(\frac{1-\alpha}{2-\alpha}\right)\left(\frac{\beta^{2-\alpha} - 1}{\beta^{1-\alpha} - 1}\right)\beta^{i-1-\frac{n}{2}}$$

with $\alpha = H + 0.5$ and $\beta = 2.5$.

3. Controlled Experiments

In this section, we describe two different controlled experiments we have carried out to compare the results from MCA and VFA. We assumed that an asset price evolves according to a particular model A and that an alternative model B is used for pricing a particular type of exotic option dependent on the asset price. The "true" prices of a panel of exotic options were determined using model A. The prices using both MCA and VFA in conjunction with model B were then computed and the accuracies of the models compared.

We defined the volatility surface by considering five different times to maturity (one month, three months, six months, one year, and two years) and five different strike prices (0.7S, 0.85S, S, 1.15S and 1.3S, where S is the asset price). When combined, these maturity/moneyness combinations give 25 "standard" points on the volatility surface.

The steps in carrying out the controlled experiment were as follows:

(a) Parameters for model A were sampled to provide "true" scenarios.

(b) European call options corresponding to the 25 standard points on the volatility surface were priced using model A for each "true" scenario.

(c) The implied volatilities for the 25 standard points were calculated for each "true" scenario from the prices in (b) using Black–Scholes–Merton.

(d) For each "true" scenario, parameters of model B that fit the 25 volatilities calculated in (c) as closely as possible were determined using an iterative procedure that minimizes the mean absolute error:

$$\frac{1}{25}\sum_{j=1}^{25}\left|\sigma_j - \sigma_j^*\right|$$

where σ_j is implied volatility at the jth point on the volatility surface given by model B and σ_j^* is the "true" implied volatility calculated in (c).

(e) For each "true" scenario, one set of parameters for the exotic option was sampled. Both model A and the calibrated model B were used to value the option defined by the sampled parameters. The MCA error is the absolute value of the difference between the two model prices.

(f) Sets of parameters for model B were generated randomly. For each set, parameters describing an exotic option of the type being considered were also generated randomly. The sampled model B parameters were used to determine the 25 standard points on the volatility surface. The sampled model B parameters and the exotic option parameters were jointly used to value the exotic option.

(g) The data from step (f) were used to construct a neural network where the features were the volatility surface points and the exotic option parameters. The target was the model B exotic option price.

(h) The neural network was used to provide VFA prices for the same "true" scenarios as those considered for MCA. The points on the volatility surface were those in (c) and the exotic option parameters were those in (e). The VFA error for each option was the absolute value of the difference between the VFA price and the model A price.

The neural networks we used in the experiments consisted of three hidden layers with 30 neurons per layer, and used the sigmoid activation function. We used Adam as the optimizer and mean absolute error to measure training loss. In the training process, we saved the weights of the neural network every time when the validation loss reached a new low. We stopped the training when there was no improvement in the validation loss for 50 epochs and used the last saved set of weights as the final weights of the model.

3.1 First Experiment

In the first controlled experiment, model A was Heston (1993) while model B was Merton (1976). The "exotic" option being considered was a plain vanilla European call option with non-standard strike price and maturity.

A plain vanilla call option would not in practice be valued using either MCA or VFA. However, this controlled experiment provides a convenient first test because the prices of the options considered can be calculated analytically using either model A or model B. The MCA calculations are therefore accurate. The only source of error lies in the construction of the neural network (which affects only the accuracy of VFA).

We assumed an initial asset price, S_0, of 100, a yield on the asset, δ, of zero, and sampled other parameters from uniform distributions. We will denote the ranges considered by parentheses. For example, [1, 2] indicates a uniform distribution between 1 and 2.

The model A (Heston) parameters used to determine the scenarios in (a) were sampled as follows:

Risk-free rate, r: [0.01, 0.05]
Initial variance, V_0: [0.01, 0.25]
Mean reversion parameter, a: [0.1, 3]
Long-term variance, V_L: [0.01, 0.25]
Volatility of volatility, ξ: [0.1, 0.8]
Correlation, ρ: [–0.9, 0]

The model B (Merton) parameters in (f) were sampled as follows.

Risk-free rate, r: [0.01, 0.05]
Volatility parameter: [0.1, 0.5]
Jump intensity, λ: [1, 5]
Jump size parameter, μ: [–0.05, 0.05]
Jump size parameter, σ: [0, 0.05]

The option parameters (both in (e) and (f)) were sampled as follows

Time to maturity (yrs): [0.05, 2]
Strike price: [80, 120]

For the construction of the neural network, sets of model B parameters and option parameters were sampled to create 100,000 instances. A total of 1000 option parameters were sampled to compare prices given by VFA and MCA.

The mean absolute errors for VFA and MCA were 0.925 and 0.965. VFA therefore performed slightly better than MCA on average. Define Y as the MCA pricing error minus the VFA pricing error and X as the MCA calibration error. Results from a linear regression, with t-statistics in parentheses, were

$$\underset{(-9.08)}{Y = -0.44} + \underset{(11.19)}{16.86X} \tag{1}$$

This shows that MCA tends to outperform VFA when the MCA calibration error is low and the reverse is true when the MCA calibration error is large. The expected performance of VFA, relative to MCA, improves significantly as the MCA calibration error increases.

3.2 Second Experiment

In the second controlled experiment, model A was Bates (1996) and model B was Heston (1993). The exotic options we considered were the barrier, Asian, and lookback options described earlier. We used Monte Carlo simulation with 50,000 paths to calculate the prices of these options as required in steps (e) and (f).

We sampled parameters similarly to the first experiment.[4] The barrier level was set equal to C times the maximum of the asset price and strike price with C sampled from [1.05, 1.30]. The neural networks were constructed using 200,000 instances and 1000 option parameters were sampled to compare prices given by VFA and MCA.

We found that VFA gave lower average errors for barrier and lookback options than MCA, but higher errors for Asian options.[5] The overall average errors, as in the case of the first experiment, are slightly higher for MCA. Table 1 provides similar results to equation (1), showing the relationship between Y, the excess of the MCA error over the VFA error, and X, the MCA calibration error.

[4] The volatility parameters were chosen as indicated in Section 6 so that the volatility surfaces had similar characteristics to those those observed for the S&P 500.
[5] The relatively good performance of MCA for Asian options can perhaps be explained by the fact that these options are most dependent on early asset price movements and these are similar for the Bates model and the calibrated Heston model.

Table 1: Results of regressing the MCA error minus the VFA error (Y) against the MCA calibration error (X). t-statistics are in parentheses.

Exotic Option	Relationship
Barrier	$Y = -0.10 + 16.35X$ $(-8.30) \quad (15.64)$
Asian	$Y = -0.21 + 10.37X$ $(-17.90) \quad (9.96)$
Lookback	$Y = -0.42 + 58.42X$ $(-11.41) \quad (18.04)$

3.3 Implications for MCA and VFA Applications

In both controlled experiments, we might expect MCA to perform fairly well because the "true" volatility surfaces come from a model and are therefore "well behaved."[6] It is therefore encouraging that VFA performs slightly better than MCA on average for both experiments. The regression results we have presented show a significant improvement in the performance of VFA relative to MCA as the MCA calibration error increases for all the cases we considered.

We investigated the MCA calibration errors encountered in practice by fitting the Merton (1976) and Heston (1993) to daily S&P 500 volatility surfaces between 2001 and 2019. The way the volatility surfaces were determined is described in Section 5. As described in Section 6, the S&P 500 volatility surfaces we used were described by 19 points rather than 25 points. Table 2 compares the calibration errors for the fit to the S&P 500 with the calibration errors in the experiments. Calibration errors are much higher on average for the S&P 500 data. Given that the relative performance of VFA improves as the calibration error increases, our results strongly suggest that

[6] This is particularly true in the second experiment where the model generating the "true" volatility surface is an extension of the model used for valuation.

VFA is likely to be a better tool than MCA in practice.[7] Furthermore, once the upfront computational time to create the neural network has been invested, VFA is much faster pricing tool than MCA as it has been traditionally implemented.[8]

Table 2: Comparison of average calibration errors in experiments with calibration errors when models are fitted to S&P 500 volatility surfaces

	Mean Absolute Calibration Error	
	Experiment (25 points)	S&P 500 (19 points)
Merton	2.7%	6.4%
Heston	0.9%	2.0%

4. The Importance of Volatility Surface Points

We implement MCA by minimizing

$$\frac{1}{25}\sum_{j=1}^{25}\left|\sigma_j - \sigma_j^*\right|$$

where σ_j is the jth point on the volatility surface given by model B and σ_j^* is the "true" volatility at the jth on the volatility surface. This objective function (in common with other similar objective functions that might be used such as mean squared error) assumes that the 25 points are equally important. In this section we outline analyses we have carried out showing that the points are not equally important for any particular option or when averaged across all the options that are considered. This suggests that, for the valuation of any given option, some relevant information is lost when MCA is used and may be a reason why VFA tends to give more accurate results as the MCA calibration error increases.

The sensitivity of the value of exotic options to the positions of the 25 standard points on the volatility surface can be estimated from the gradient of the function that the VFA neural network represents. Letting y represent the exotic option price estimated by the neural network, we define

[7] The fact that 25 points had to be fitted in the experiments and only 19 points had to be fitted to the S&P 500 data reinforces our conclusions.
[8] However, this is not an advantage of VFA over MCA. As shown by Horvath et al (2021), neural networks can be used to speed up MCA computations.

the sensitivity, $s(v_i)$, of the price to the implied volatility at the *i*th point on the volatility surface as the absolute value of

$$\frac{\left|\frac{\partial y}{\partial v_i}\right|}{E\left(\left|\frac{\partial y}{\partial v_i}\right|\right)} - 1$$

where $\frac{\partial y}{\partial v_i}$ is the partial derivative of the prediction function with respect to volatility surface point v_i.

In our sensitivity analysis we aim to measure the rate of change irrespective of sign, thus we only consider the absolute values of partial derivatives. Further, in the equation above, expectation is taken over the entire set of volatility surfaces that are used. Computing the average partial derivative with respect to each volatility surface point is necessary so that the baseline sensitivity of the neural network is taken into account and used as reference. A detailed discussion on the use of baselines for gradient-based sensitivity analysis can be found in the work of Sundararajan et al (2017). Intuitively, the average gradient corresponds to how sensitive the neural network output is when the "average" surface is passed as input. We therefore define sensitivity to v_i using the absolute relative difference between $\left|\frac{\partial y}{\partial v_i}\right|$ and its average: a larger sensitivity implies a larger impact on how the predicted value changes relative to the expected output of the network.

The gradients are computed by applying the back-propagation algorithm, which was first introduced by Rumelhart et al (1986) and is typically used to train neural networks. Specifically, we apply back-propagation on the final trained network and take an additional step in the chain rule to obtain input gradients. Note that during this process we do not update the neural network weights.

We randomly sampled 200 volatility surfaces and calculated $s(v_i)$ for barrier, Asian, and lookback options for *i*=1 to 25. Table 3 reports the standard deviations of the $s(v_i)$.

From the table it can be seen that different volatility surface points exhibit different dispersions, meaning that there is variability in how a particular point contributes to the price prediction across different input surfaces. These and other more detailed results allow us to conclude that a) different volatility surface points have different importance when valuing a single exotic option, and b) the same volatility point has varying importance when valuing different exotic options. The results corroborate our hypothesis that the VFA approach, by virtue of using a neural network, can exploit the structural information of the volatility surface when valuing an exotic option and can therefore be advantageous compared to MCA.

Table 3: Standard deviation of $s(v_i)$ for barrier, Asian and lookback options. T is the maturity for volatility surface point considered, K is strike price and S is asset price.

	Barrier Options				
	T = 1 month	**T = 3 months**	**T = 6 months**	**T = 1 year**	**T = 2 years**
K = 0.7S	0.77	0.59	0.22	0.47	0.37
K = 0.85S	0.37	0.52	0.42	0.37	0.48
K = S	0.58	0.45	0.33	0.34	0.47
K = 1.15S	0.49	0.40	0.38	0.40	0.32
K = 1.3S	0.36	0.42	0.38	0.46	0.25
	Asian Options				
K = 0.7S	0.67	0.65	0.40	0.39	0.40
K = 0.85S	0.26	0.44	0.34	0.43	0.54
K = S	0.37	0.23	0.35	0.38	0.47
K = 1.15S	0.56	0.44	0.31	0.38	0.34
K = 1.3S	0.30	0.35	0.27	0.48	0.31
	Lookback Options				
K = 0.7S	0.40	0.22	0.34	0.44	0.57
K = 0.85S	0.40	0.22	0.34	0.31	0.41
K = S	0.51	0.23	0.29	0.28	0.41
K = 1.15S	0.58	0.23	0.24	0.29	0.40
K = 1.3S	0.33	0.26	0.27	0.43	0.52

It is clear from the analysis in this section that it is at best an approximation to give all points on the volatility surface equal weights when an exotic option is valued. The weights appropriate for different points vary according to the parameters of the exotic being considered. When MCA is used the volatility surface is fitted with some error. It is possible to vary the weights used for different points on the volatility surface so that some parts of the volatility surface are fitted more accurately than other parts. However, it is difficult to know ex ante what the appropriate weights are. This is what motivates us to experiment with the VFA approach. The neural network that is used in VFA has the potential to better reflect inter-relationships between points on the volatility surface and the parameters of the exotic option being valued.

5. Volatility Surface Generation

For the tests which will be reported in the rest of the paper, we collected data on call options on the S&P 500 between June 2001 and June 2019 from OptionMetrics. We cleaned the data in a number of ways. In particular, we only kept options with open interest greater than 0 and time to maturity between 1 month and 2 years. Options with no implied volatility reported by the database were removed. We defined moneyness as the strike price divided by the index level and removed all options with moneyness smaller than 0.7 and greater than 1.3. This led to an average number of options each day of 500.4.

Options trade on any given day with nonstandard maturity/moneyness combinations. Each day we used a search algorithm to determine the implied volatilities for standard points that, with interpolation, gave the best fit to the options in our data set. Consider a particular option in the data set has a strike price of K and time to maturity T. Define K_u and K_d as the standard strike prices that are closest to K with the property that $K_u \geq K \geq K_d$. Similarly, define T_u and T_d as the standard times to maturity that are closest to T with the property that $T_u \geq T \geq T_d$. For any given trial set of standard implied volatilities, the implied volatility for the $\{K, T\}$ option was determined using a bivariate linear interpolation between the $\{K_u, T_u\}$, $\{K_u, T_d\}$, $\{K_d, T_u\}$, and $\{K_d, T_d\}$ implied volatilities. The best-fit implied volatilities for the standard maturities and strike prices were those that minimized the sum of squared differences between the interpolated volatilities and the reported implied volatilities.

This procedure seems to produce implied volatility estimates for standard maturities and strike prices that are reasonably consistent with the market. The daily mean absolute difference between the interpolated implied volatilities and the actual implied volatilities for the options in the data set averaged less than 0.4%.

6. Model Risk

The VFA model is a useful tool for estimating model risk. The first stage is to identify a range of different models that can be used to price a particular exotic option. A neural network is then constructed for each model and the price differences between the models are evaluated. We illustrate the approach by using it for barrier, Asian, and lookback options on the S&P 500.

It is necessary to define standard points on the volatility surface. We chose the standard points on the volatility surface to be those used in Sections 3 and 4 except that six points were eliminated. These were the one-month points that correspond to strike prices of $0.7S$, $0.85S$, $1.15S$, and $1.3S$ and three-month points that correspond to strike prices of $0.7S$ and $1.3S$, where S is the index level. This was because S&P 500 implied volatilities were often either unreliable or nonexistent for these extreme maturity/moneyness combinations.

The steps used to produce results for each of the models and exotic options considered are as follows:

(a) Model parameters are generated randomly. For each set of model parameters, a set of parameters describing the exotic option are also generated randomly. For each of the resulting data sets, the standard points on the volatility surface are calculated and the exotic option is valued.
(b) Data from step (a) are used to construct a neural network. The features are the volatility surface points and the exotic option parameters while the target is the exotic option price.
(c) The neural network in (b) is used to price a panel of exotic options using historical data on the asset of interest (the S&P 500 in our case)

The price differences when different models are used to price the same panel of exotic options provide statistics to quantify model risk

We illustrate the approach using the Bates model and Jaber (2019) version of the rough Heston model described in Section 2. The historical data was used to generate standard points on the volatility surface each day between June 2001 and June 2019 as outlined in Section 5. For Asian and lookback options, the maturities considered each day for the panel of options in step (c) were 0.25, 0.5, 1, and 2 years. The strike prices considered were 90%, 95%, 100%, 105%, 110%, and 115% of the index. In the case of barrier options, we considered the same set of maturities. Three strike prices (90%, 100% and 110% of the index) and two barriers (110% and 130% of the maximum of the strike price and index level) were considered. The neural network was constructed similarly to the neural networks used in Section 3.

It is important that the parameters sampled for a model in (a) lead to volatility surfaces that have the same characteristics as those observed in the past for the underlying asset being considered. In the case of the S&P 500, the at-the-money one-month implied volatility, X, ranged from 0.054 to 0.7589. The value of $\ln(X-0.05)$ had a mean of -2.41 and a standard deviation of 0.69. This led us to sample the initial volatility for both Bates and rough Heston as $0.05+\exp(u)$ where u has a mean of -2.5 and a standard deviation of 1.0. The best fit relationship between the two-year implied volatility, W, and the one-month implied volatility, Z, for the S&P 500 was

$$W=0.12 + 0.46Z$$

with a standard error of 0.02. We sampled the long-term variance rate so that the two-year volatility was in a 0.08 range that was approximately centered on this. Other parameters were sampled similarly to the controlled experiments.

We verified that the sampled volatility surfaces produced in this way and the historical S&P volatility surfaces had similar characteristics by using mutual information measures.[9] Specifically, for each set of Bates or rough Heston parameters we created a balanced dataset where 50% of the data consisted of sampled volatility surfaces based on these parameters, and 50% of the data consisted of S&P volatility surfaces. We then performed a k-means cluster analysis, constraining the algorithm to partition the dataset into two distinct clusters. Finally, we used the adjusted mutual information (AMI) to measure the quality of the produced clusters. A large AMI indicates good clustering, where sampled and S&P surfaces are mostly placed in separate clusters, i.e., it is easier for the algorithm to discriminate between the two types of surfaces. The opposite is also true. If AMI is small then it is harder to discriminate between the two types of surfaces, and thus one can conclude that they share more similar characteristics. Based on this rationale we validated whether the sampled surfaces were similar enough to the S&P surfaces using the following rules:

(a) If the average AMI of clustering was distinctly larger than the average AMI of a random partition (both processes were repeated several times) then the sampled surfaces were deemed to be unsatisfactory for capturing the data distribution of S&P volatility surfaces.

[9] See Vinh et al. (2010) for a discussion of mutual information measures.

In other words, sampled surfaces were easy to identify and the model parameters had to be adjusted.

(b) If the average AMI of clustering was close to the average AMI of a random partition then the sampled surfaces were considered valid for our purposes and the model parameters were accepted.

Note that this process was not used to explicitly search for the best model parameters (clustering results did not dictate how model parameters should be sampled), but rather it was applied as a validation step to ensure the quality of the dataset

Table 4 shows the means and standard deviations of the difference between the prices given by rough Heston and Bates for the three exotic options considered. It can be seen that rough Heston tends to produce higher prices than Bates for barrier and Asian options and lower prices for lookback options.[10]

Table 4 Statistics for the difference between rough Heston (RH) and Bates prices, as a percentage of the index level

	Mean RH Price	Mean Bates Price	Mean Price Diff: RH minus Bates	S.D. of Price Difference
Barrier	4.494	4.286	0.208	0.271
Asian	4.880	4.835	0.045	0.154
Lookback	11.631	11.946	−0.315	0.484

We also tested whether pricing differences depend on the calibration error when Heston is fitted to the S&P 500 volatility surfaces. For all three exotic options we found that price differences increase significantly as the calibration error increases. This suggests (unsurprisingly) that model risk increases as the deviation of the volatility surface from the volatility surfaces given by models commonly used increases.

7. Conclusions

[10] A reason for this may be that the jumps in Bates make lookbacks more valuable and the barrier more likely to be hit.

In this paper, we have considered a new approach to pricing exotic options. This involves constructing a neural network relating prices of an exotic option to the volatility surface and the option parameters directly. We refer to this as the VFA approach. Tests show that the approach is likely to give better results in practice than the alternative approach (referred to as MCA) of implying model parameters from the observed volatility surface. Once the neural network has been developed, the pricing of exotic options using VFA is very fast.

An important task for risk managers is coming up with measures of model risk. We have illustrated how such measures can be produced using VFA by comparing the way two different models (both extensions of the well-known Heston model) price three different types of exotic options. An interesting extension might be to use Bayesian neural networks to provide probability distributions for the prices given by different models rather than point estimates. Possibly, a distribution of market prices can be estimated using VFA with a training set created by randomly sampling from alternative models in an appropriate way.

References

Bates DS (1996) Jumps and stochastic volatility: Exchange rate processes implicit in deutschemark options. *Rev. Financ. Stud.* 9(1):69-107.

Dupire B (1994) Pricing with a smile. *Risk*. 7(January):18-20.

El Euch O, Gatheral J, Rosenbaum M (2019) Roughening Heston. *Risk* (May): 84-89.

Federal Reserve System (2011) SR Letter 11-7: Supervisory guidance on model risk management, Washington, DC.

Ferguson R, Green A (2018) Deeply learning derivatives. Working paper, arXiv: 1809.02233.

Gatheral J, Jaisson T, Rosenbaum M (2018) Volatility is rough. *Quant. Financ.* 18(6):933-949.

Heston SL (1993) A closed form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* 6(2):327-343.

Horvath B, Muguruza A, Tomas M (2021) Deep learning volatility: A deep neural network perspective on pricing and calibration in (rough) volatility models. *Quant. Financ.* 21(1):11-27.

Hull JC., Suo W (2002) A methodology for the assessment of model risk and its application to the implied volatility function model. *J. Financ. Quant. Anal.*, 37(2):297-318.

Jaber EA (2019) Lifting the Heston model. *Quant. Financ.* 19(12):1995-2013.

Liu S, Borovykh A, Grzelak LA, Oosterlee CW (2019), A neural network-based framework for financial model calibration. *J. Math. Industry* 9(9), https://doi.org/10.1186/s13362-019-0066-7.

Merton RC (1976) Option pricing when underlying stock returns are discontinuous*. J. Financ. Econ.*, 3(1-2):125-144.

Sundararajan M, Taly A, and Yan Q (2017) Axiomatic attributions for deep networks. *Proc. 34th International Conference on Machine Learning*, 3319-3328.

Vinh, N. X., Epps, J., and J. Bailey (2010) Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res*. 11: 2837–2854.