Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis (Quantitative Finance FEM61008)

Forecasting the implied volatility surface using deep learning to improve parametric models

Author:

Name student: Tim Hack

Student ID Number: 508048

Supervisor: Dr. Gustavo Freire

Second assessor: Dr. Onno Kleen

Version date: July 2022

Abstract

Options are widely used by financial institutions for hedging or trading activities. An important key in option pricing is the implied volatility surface. Next to pricing options, the implied volatility surface is often used to manage option positions. It is therefore important to correctly forecast the implied volatility surface. This thesis examines a two-step approach for predicting the implied volatility surface. First, five different parametric models are implemented after which a feedforward neural network is used to correct for the errors made by these models. The Black-Scholes model, ad-hoc Black Scholes model, local volatility model, SABR model and Heston model are implemented. The research finds that the SABR model is the best performing model for predicting the implied volatility, considering the forecasting with and without a feedforward neural network. The feedforward neural network has a positive effect on all parametric models, decreasing the prediction error. The correction effect of the neural network is the most present in the case where the original error is large and the prediction exercise is not far away in the future.

Keywords: Implied Volatility, Deep learning, Stochastic Volatility, Local Volatility, Correction Model

Contents

1 Introduction

Options are commonly used by banks, hedge funds or pension funds for hedging purposes or other trading activities. A key role for pricing options is in the implied volatility surface. The implied volatility surface is given as the implied volatility of a set of options for different moneyness and tenors. Besides pricing options, the implied volatility surface is often used to manage option positions, see e.g. Kim and Lee (2018). The implied volatility surface is widely used since it is easy to compare options with different strike prices and maturities to each other. Thus, precisely forecasting the implied volatility surface is important to make the correct trading decisions when buying and selling options for all kinds of investors.

The most classic approach to obtain the implied volatility surface is to use the Black-Scholes model (Black and Scholes, 1973). Here the assumption is made that the volatility of the underlying asset price is constant and therefore, the implied volatility surface will be flat. Considering the market however, the implied volatility surface shows a skew or smile for different strikes. Besides, over time the surface shows a highly non-linear behaviour, see e.g Homescu (2011) where properties of the implied volatility surface are discussed. Therefore, the assumption of a constant volatility of the underlying is a misspecification of the Black-Scholes model. Over the years, many other models have been developed to model the implied volatility. Dumas et al. (1998) propose an ad-hoc correction for the Black-Scholes model. Here they argue that their model exceeds local volatility models (Rubinstein, 1994, Dupire et al., 1994).

Yet, according to Ren et al. (2007) local volatility models are still widely used. Besides, some papers conclude that local volatility models are well-suited to forecast the implied volatility surface. For example, Homescu (2014) give a couple of advantages of a local volatility model. One of them is that local volatility models (LVM) always fit today's market prices due to the way it is constructed. Also, Dupire (1994) shows that in some European option pricing context, local volatility models work better compared to stochastic volatility models. For a more broad overview between different methods to model the implied volatility smile, see e.g. Skiadopoulos (2001). Thus, after the paper of Dumas et al. (1998) there have been multiple papers that don't conclude that local volatility models should no longer be used because they are outperformed. Therefore, it will be interesting to see how the conclusion of Dumas et al. (1998) fits in the current timeframe with data of a more recent period.

Another type of model are the stochastic volatility models. For example, Heston model (Heston,

1993), where the volatility follows a mean-reverting process. Further, there is the Stochastic-Alpha-Beta-Rho stochastic volatility (SABR-SV) model, Hagan et al. (2002). This model is an extension of the constant-elasticity-of-variance model (Cox, 1975). There are a couple of advantages to choose the SABR model over the Heston model. First of all, it is much simpler to calibrate the parameters, making it easy to fit the model to the implied volatility smile (Hagan et al., 2002, West, 2005). This is because an analytical approach exists in the form of Hagan's formula, while for the Heston model a numerical approach has to be chosen. Second, the SABR parameters are easy to interpret and all have their unique impact on the skewness, level and curvature of the implied volatility smile. While for the Heston model certain parameters have the same effect on the smile. Gauthier and Rivaille (2009) notice that the model is overparameterized. Finally, it is known that the Heston model is unable to fit the smile to options with a short maturity while this is not an issue for the SABR model (Lemaire et al., 2021). A more thorough comparison between both models is discussed later in this thesis.

Although in the research of Dumas et al. (1998) local volatility models and the ad-hoc Black-Scholes model are compared, no comparison is made with a stochastic volatility model. It would be interesting to examine how the SABR model compares to the ad-hoc Black-Scholes model, the Heston model and a local volatility model. Thus, the first problem this research will examine is

How does the SABR model perform compared to a local volatility model, the ad-hoc Black-Scholes model and the Heston model for forecasting the implied volatility surface?

All mentioned models until now are parametric models. Here, the implied volatility surface is forecasted as a function of some parameters. Recently, Almeida et al. (2022) proposed a two-step method to correct for the prediction error of the parametric model. In the paper first a parametric model is fitted after which a feedforward neural network is trained to refine the original prediction. An advantage of this approach is that a feedforward neural network is able to capture the non-linear behaviour of the implied volatility surface over the time-to-maturity cross-section and the smile behaviour over the moneyness cross-section.

In their paper, four models were considered to which this method was applied. For the stochastic volatility model the choice for the Heston model was taken. It would be interesting to consider how the SABR model performs using this correction approach over the parametric fit. First of all, to investigate how much it improves the original fit, since a better model for predicting the

implied volatility surface is important for pricing and managing options. Second, to compare it to the results of the Heston stochastic volatility model of the analysis.

The SABR model is a commonly used stochastic volatility model, see e.g. Floc'h et al. (2014). The advantage of the simple calibration of the parameters together with the improvement of the feedforward neural network could make it a good alternative to predict the implied volatility surface. Therefore, the second problem this research will consider is

What is the prediction power of the SABR model when applying a feedforward neural network on the error surface.

For the research a dataset of S&P 500 option data is used. The different parametric models that will be compared are the Black-Scholes model, where a constant implied volatility surface is assumed. The ad-hoc Black-Scholes model, where a correction is applied to the classic Black-Scholes model. A local volatility model and finally, the SABR model which is a type of stochastic volatility model.

The main contribution of this research lies in the fact that the proposed two-step approach for the SABR model could potentially be an useful method to predict the implied volatility surface. It combines the main advantages of two different methods. First of all, the SABR stochastic volatility model, which is easily calibrated and could fit very well to the implied volatility smile or skew as seen in the market. Second, the combination with a feedforward neural network which is able to model the non-linear behaviour over the cross-section of the implied volatility surface could result in an even better performance for this parametric model. This combination could make it a very strong model to forecast the implied volatility surface, which hasn't been investigated yet.

This thesis has been organized in the following way. First, a section with some background information is presented. In section 3, relevant literature is discussed. Then, the data which will be used for the research is considered. Section 5 gives an overview of the different parametric models and the feedforward neural network which are all used for this research. In the sixth chapter the results of the parametric models and the feedforward neural network are presented. Finally, section 7 gives the conclusion of the research and some future research work which could be performed.

2 Literature review

This section gives an overview of the relevant literature related to the forecasting of implied volatility surfaces. The basic approach to obtain the implied volatility is to use the Black-Scholes model (Black and Scholes, 1973). However, the assumption of a constant volatility is rejected by the data, see e.g. Rubinstein (1994) or Dumas et al. (1998). Over the years plenty of other models have been developed to account for this flaw. The assumption that the volatility could be a function of the underlying level and time, but not by a stochastic process, was introduced by Dupire et al. (1994) and Rubinstein (1994). These models are known as local volatility models. After the local volatility models, Dumas et al. (1998) introduced the ad-hoc Black-Scholes model. In their paper they show that this method beats the local volatility methods. The ad-hoc Black-Scholes model smoothes the implied volatilities over time-to-maturity and strikes given by the classic Black-Scholes model. Besides, a drawback from local volatility models is that in some cases they are not consistent with the market. When the underlying asset decreases in value, local volatility models would report a higher volatility smile. This is in contrast of what actually happens in the market, see e.g. Hagan et al. (2002). Nevertheless, according to Ren et al. (2007) and Homescu (2014) local volatility models are still used to price options. Crépey (2004) even conclude that local volatility models are the solution for hedging when using index options. Another type of model are the different jump diffusion models, such as Merton's model (Merton, 1976) and Kou's model (Kou and Wang, 2004). Here a jump process is taken into account for the underlying asset. The difference in the models is in the choice for the jump distribution of the asset path. There is also more recent literature into the forecasting of the implied volatility surface. There is Fengler et al. (2007) where a semi-parametric factor model is used. Further, Chen et al. (2016) implement a Kalman filter in order to forecast the implied volatility surface.

This research also focuses on stochastic volatility models. There is a lot of literature related to the research of stochastic volatility models. Some of the popular ones are, Heston's model (Heston, 1993) where the volatility follows a mean-reverting process or the stochastic-alpha-beta-rho (SABR) model (Hagan et al., 2002). A simpler form of the Heston model is known as the Schöbel-Zhu stochastic volatility model (Schöbel and Zhu, 1999). Combinations of a local volatility model with a stochastic volatility model also exist. These models were developed by Jex and Wang (1999) and Lipton (2002). The advantage of this type of model is that it combines the power of both types of models. They are easy to calibrate, which is harder for stochastic volatility models.

However, they also have the property of stochastic volatility models that they fit very well to the market implied volatility curve, see Oosterlee and Grzelak (2019). A disadvantage lies in the fact that the models can become complicated and hard to implement. For more information on stochastic local volatility models the paper of Homescu (2014) is a good start.

There has also been research in pricing options and predicting the behaviour of the implied volatility surface using neural networks. Malliaris and Salchenberger (1996) use a neural network to forecast the implied volatility for the S&P 100 index. In Hutchinson et al. (1994) a non-parametric approach using learning networks is applied for the pricing and hedging of options. Other examples are Yao et al. (2000) and Carverhill and Cheuk (2003) where a neural network is used for option pricing and hedging. Chen and Zhang (2019) and Medvedev and Wang (2022) both use a LSTM deep-learning approach to forecast the implied volatility surface. Finally the paper of Almeida et al. (2022) is of great importance for this research. The proposed two-step technique in their research will be followed and is used on the SABR model for this thesis.

3 Background information

This chapter covers topics regarding the fundamentals of the research. Only the basic definitions are discussed. Starting with the definition of options, second the Black-Scholes model and finally, the implied volatility parameter. Extra information on these topics can be found in Appendix A or in Oosterlee and Grzelak (2019).

3.1 Basic definitions

Definition 3.1 (Option contract)**.** An option contract gives the right but not the obligation to buy or sell an underlying asset for a predetermined price at a predetermined date.

The price at which the asset can be bought is known as the strike price, K. The option matures at time T and the time that is left till the option expires, τ, is known as the time-to-maturity ($\tau = T - t$). The payoff function at maturity for a call option is given as

$$V_c(T, S) = max(S(T) - K, 0). \tag{3.1}$$

Definition 3.2 (Put-call parity)**.** The put-call parity gives the relation between a call and a put option with the same strike price and time-to-maturity.

Consider an underlying asset S, a call option C, put option P, strike price K, dividend yield δ, time till expiry τ and interest rate r, the put-call parity with a continuous dividend yield is given as

$$P = C + Ke^{-r\tau} - Se^{-\delta\tau}. \tag{3.2}$$

The relationship also holds in a dividend free environment.

Next, we consider the Black-Scholes model. The Black-Scholes equation is a closed form solution to price European options under some assumptions. One of the assumptions is that the underlying asset follows a geometric Brownian motion (GBM). This means that the path of the underlying $S(t)$ is in line with the following stochastic differential equation

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(t_0) = S_0. \tag{3.3}$$

Here μ is the drift parameter, σ is known as the volatility parameter and $W(t)$ is the Brownian motion. The larger σ the more variation is present in the path of the underlying.

The SDE can be solved using either a martingale approach or using the Feynman-Kac theorem. This results in a closed-form solution to price an option. The value of an European call option at time t with strike K, expiry T and interest rate r is given as

$$\begin{aligned} V_c(t, S(t)) &= S(t)\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \\ d_1 &= \frac{\log(\frac{S(t)}{K}) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned} \tag{3.4}$$

Here $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable.

Lastly, the implied volatility parameter is considered, which is an important input parameter in the Black-Scholes equation.

Definition 3.3 (Implied volatility)**.** The implied volatility is an input that makes the Black-Scholes model deliver the observed option price in the market and is the markets view on how the underlying asset will move in the future.

There are multiple ways to obtain the implied volatility parameter. This thesis will use Newton-Rhapson method, see e.g. Oosterlee and Grzelak (2019). In order to acquire the parameter the option price of the Black-Scholes equation should equal to the market price.

$$V(t_0, S, K, \tau, \sigma_{imp}, r) = V^{mkt}(K, T). \tag{3.5}$$

Figure 1 shows an example of the implied volatility surface which can be observed in the market. The implied volatility of option data on a single day for different maturities can be seen in the figure. The x-axis is the underlying price (S) over the strike price (K). The left side of the figure represents deep out-of-the-money call options while the right side constitute to deep out-of-the-money put options. We see that the figure exhibits the smile behaviour which is usually seen in the market. The implied volatility rises for more deep out-of-the-money options while it is the lowest around at-the-money options. Out-of-the-money options are considered to be more risky and the underlying has to move more in order to make a profit off the option.

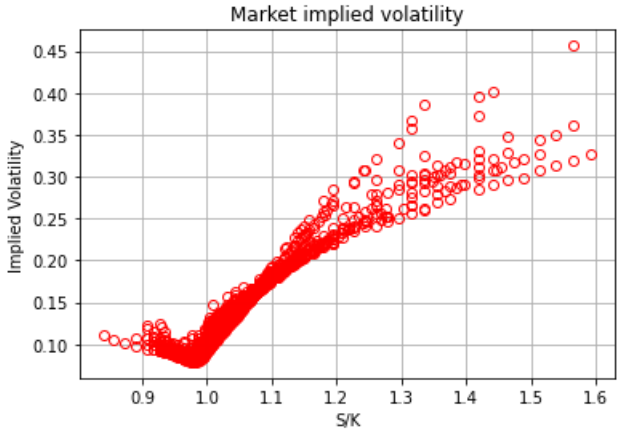

Figure 1: **Implied volatility smile.** This figure shows the implied volatility smile of S&P 500 index options on January 3, 2017. The implied volatility smile for different maturities is plotted all in the same figure. On the x-axis the underlying price over the strike price is visible.

4 Data

For this research S&P 500 index options traded on the Chicago Board Options Exchange (CBOE) will be considered. Only European style options are taken into account. The choice for this dataset is made since these type of options are the most widely traded ones and are therefore the most liquid options. The option data is obtained through OptionMetrics which was accessed through the Wharton Research Data Services.[1] The data ranges from January 2, 2017 till December 31 2021. This means that the most recent five years of available data are taken. We don't consider a longer time period since this increases the computational time and a shorter horizon isn't chosen either since we don't want the results to be to specific for a certain date range. For example, taking a shorter horizon would mean that the data becomes focused on the Covid-19 crisis. Only options with an expiry between 20 and 240 days are considered to make the data consistent with Almeida et al. (2022).

Before applying the models to the data, preprocessing has to be performed. First of all, the option prices are taken as the average of the bid price and ask price at the end of the day. The time-to-maturity for the options is calculated by taking the difference between the trading date and the maturity date. It is taken into account if the option settles at the start or at the end of the day. Finally, options with zero volume or a price below 0.125 are dropped out of the database.

The S&P 500 also produces a dividend which has to be obtained to calculate the implied volatilities. In order to calculate this dividend, the put-call parity relation adjusted for dividends is used. For the put and call options the closest at-the-money options are considered for each day and time-to-maturity. Thus, for every day t in the dataset we find the put and call option price with the same strike price which are the closest to the current S&P500 level. With this information we can fill in equation (3.2). The only unknown in this equation will be the dividend yield which can then be acquired for day t.

The value of the S&P 500 index is obtained from yahoo finance. Finally, a representative for the risk-free rate is acquired from the St. Louis Federal Reserve Economic Data (FRED) database.[2] The proxy that is taken is the 3-month Treasury bill rate. Figure 2 shows the level of the S&P 500 of our dataset over the entire period. We clearly see that overall the index was increasing in value. However, there is a large drop in the level in March 2020 due to the covid-19 crisis. The same can be seen in the value of the risk-free-rate in figure 3. During the crash there was a massive drop in

[1] https://wrds-www.wharton.upenn.edu/
[2] https://fred.stlouisfed.org/

Figure 2: **Level of the S&P 500.** The sample ranges from January 2017 - December 2021.

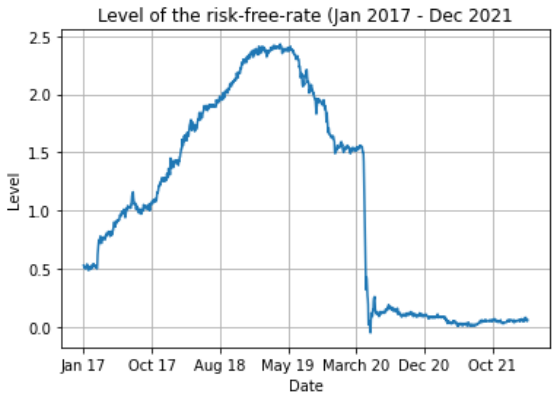

Figure 3: **Level of the risk-free-rate.** The sample ranges from January 2017 - December 2021.

the 3-month Treasury bill rate.

The moneyness ($\frac{S_t}{K_{i,t}}$) of the options is taken to be in the range of $[0.8, 1.6]$. Besides, it is common practice to only use out-of-the money options for the research, see Almeida et al. (2022). This choice is made since OTM options are considered more liquid compared to the ITM options. This means that if the moneyness is below one, only call options are considered and if the moneyness is above one only put options are taken into consideration. A larger moneyness range could be chosen however, this means that the dataset contains more deep out-the-money options with very low prices. As discussed in Dumas et al. (1998), these options are illiquid and do not contribute to information about the volatility surface. Therefore this specific range for the moneyness is kept.

Finally, in order to give a clear and convenient overview of the results later in this thesis the options are put into different brackets depending on their time-to-maturity and moneyness. Again, this split is chosen accordingly with Almeida et al. (2022). Regarding the days to expiry, we consider short-term options, $\tau \leq 60$, and long-term options, $\tau > 60$. For the moneyness, six splits are chosen. We consider the following thresholds: [0.8,0.9) (deep out-of-the-money calls), [0.9,0.97) (out-of-the-money calls), [0.97, 1.03) (at-the-money options), [1.03,1.10) (out-of-the-money puts) and [1.10, 1.60] (deep out-of-the-money puts).

Table 1 presents an overview of the data. The table provides the number of observations, the average implied volatility and the standard deviation of the implied volatility for short and long-term options in each of the moneyness classes. The implied volatilities were obtained by inverting the Black-Scholes equation from the options prices using Newtons method. First of all, our dataset contains 2,340,031 observations of which the largest part are out-of-the-money put options. Besides,

options with a shorter time to expiry are more present compared to long-term options. Interestingly enough, the smile shape of the implied volatility curve is present but, the lowest point is not for at-the-money options. This is common for S&P 500 index options. Immediately we also see that the implied volatility is not constant for the options, disproving the prediction coming from the Black-Scholes model. The implied volatility is the highest for deep OTM put options. A possible explanation for this is that banks and pension funds need to hedge their positions and achieve this by buying put options. More demand for these options leads to higher option prices and the only variable that changes in the Black-Scholes equation (3.4) is the implied volatility. In order to obtain higher option prices the implied volatility parameter should go up.

A final observation that can be made is the fact that the standard deviation and the average implied volatility is higher compared to the sample period of Almeida et al. (2022). A possible reason for this is that our dataset contains periods where there was much more stress in the market. Take the Covid-19 crisis and the final part of 2018, when there was a big sell-off in the S&P 500 (figure 2). This results in a higher implied volatility for options.

Table 1: **Summary statistics of S&P 500 options**

	# Observations		Mean IV		Std. Dev.	
Days to expiry / Moneyness	Short	Long	Short	Long	Short	Long
[0.8,0.9)	30818	41192	23.8	16.7	10.4	6.5
[0.9,0.97)	240545	140970	13.4	13.5	6.7	5.5
[0.97,1.03)	419641	168507	14.3	16.3	6.9	6.1
[1.03,1.10)	368805	154772	20.9	21.1	7.0	6.0
[1.10,1.60]	440699	334082	34.3	30.1	11.4	8.1

This table provides the summary statistics of the S&P 500 index options. The sample ranges from January 2017 to December 2021. The columns represent the number of observations, the average implied volatility (%) and the standard deviation (%) for short and long-term options. These are shown for different classes of options depending on their moneyness S/K.

5 Methodology

This section examines the main methods that will be used for the research. First, the ad-hoc Black-Scholes model will be considered. Second, the local volatility model is discussed. Third, the SABR model is reviewed together with how the parameters should be fitted to the implied volatility surface. Then, the Heston model is examined and finally, the feedforward neural network is investigated. It is also discussed how the neural network approach should be applied to the parametric models.

5.1 Ad-hoc Black-Scholes

The ad-hoc Black-Scholes model (AHBS) was first introduced in Dumas et al. (1998). A main difference between the BS model and the AHBS model is in the fact that the volatility parameter is no longer assumed to be constant. For the AHBS model the volatility in equation (3.3) changes from being constant to depending on the moneyness and time-to-expiry ($\sigma(m, \tau)$). Due to this dependency the forecasts of the implied volatility surfaces will no longer be flat, but change over the strike cross-section and the tenor cross-section. The ad-hoc approach basically smoothes the implied volatility surface over both cross sections.

For all options, $i = 1, ..., N$, on a given day t the volatility parameter for the AHBS is, as in Dumas et al. (1998), given by

$$\sigma_{i,t} = a_{0,t} + a_{1,t}m_{i,t} + a_{2,t}m_{i,t}^2 + a_{3,t}\tau_{i,t} + a_{4,t}\tau_{i,t}^2 + a_{5,t}m_{i,t}\tau_{i,t} + \epsilon_{i,t}. \tag{5.1}$$

Here m is the moneyness and τ is the time-to-expiry for the given option. When the parameters are estimated a prediction for the implied volatility can be made, this is done as follows

$$\hat{\sigma}_{i,t} = \max(0.01, \hat{a}_{0,t} + \hat{a}_{1,t}m_{i,t} + \hat{a}_{2,t}m_{i,t}^2 + \hat{a}_{3,t}\tau_{i,t} + \hat{a}_{4,t}\tau_{i,t}^2 + \hat{a}_{5,t}m_{i,t}\tau_{i,t}). \tag{5.2}$$

The maximum function is in order to guarantee that the implied volatility doesn't become too small or even negative. There were a few rare occasions where the forecasted implied volatility parameter would become very small or even negative. This happens since the parameters are estimated with different moneyness and time-to-maturity values compared to the prediction parameters. The equation is of a quadratic form, this is chosen since the implied volatility surface for S&P 500 index options usually show quadratic behaviour (Dumas et al., 1998). In Dumas et al. (1998) three

other volatility functions are specified where there is less dependency on the moneyness or expiry's. These representations are not considered in this thesis. This is motivated by the fact that their paper was made with much less liquid data. Thus, there were cases with almost no options over the expiry or strike cross-section. This meant that the other models, with less focus on the strike or maturity might performed better. Since this research works with much more recent data and has a liquid option dataset, this is no longer necessary.

In order to estimate the parameters $\mathbf{a}_t$, an ordinary least square approach is implemented. Therefore, the following equation has to be minimized to obtain the parameters

$$\frac{1}{n}\sum_{i=1}^{n}\left[\sigma_{i,t} - \sigma_{AHBS}\left(\boldsymbol{a}_t, m_{i,t}, \tau_{i,t}\right)\right]^2, \tag{5.3}$$

where $\sigma_{i,t}$ are the actual implied volatilities. Hence, for predicting the implied volatility surface with the ad-hoc Black-Scholes model the following procedure is followed:

1. For day t obtain the option cross-section including the expiries and moneyness of all options.
2. Estimate the parameters in equation (5.1) by minimizing equation (5.3).
3. For day $t+h$ obtain the option cross-section including the true implied volatilities by inverting the Black-Scholes equation.
4. Fill in equation (5.2) for all options $i = 1, ..., N$ on this future day, using the parameters obtained from the OLS approach. This results in the forecasted implied volatility surface.
5. Repeat the previous steps for all days t in the dataset.

This chapter of the report doesn't contain a separate section for solely the Black-Scholes model. The reason for this is that the same procedure can be used. In equation (5.1) all parameters are set to 0 except for a_0. Note that the volatility parameter becomes constant, $\sigma_{i,t} = a_{0,t}$, implying the Black-Scholes model. Applying ordinary least square to this simplified expression results in calculating the average implied volatility over day t, which is then used as forecasted implied volatility surface for day $t + h$. The Black-Scholes prediction doesn't take into account any shape or form for the implied volatility surface, it only forecasts a flat surface.

5.2 Local Volatility Model

Local volatility models are another type of model that offer a solution to reproducing the implied volatility smile. They were first presented by Dupire et al. (1994) and Rubinstein (1994). In local

volatility models the volatility process is a function of the underlying asset and time $(S(t), t)$. The difference with a stochastic volatility model is that the volatility parameter is not driven by a stochastic process. For the local volatility model, equation (3.3) changes to

$$dS(t) = r(t)S(t)dt + \sigma(t, S)S(t)dW(t), \qquad S(t_0) = S_0. \tag{5.4}$$

It turns out that the value of the local volatility parameter can be extracted from the market using call options. This representation is known as Dupire's local volatility term. The expression for the local volatility is given as

$$\sigma^2_{LV}(T, K) = \frac{\frac{\partial V_c(t_0,S_0;K,T)}{\partial T} + rK\frac{\partial V_c(t_0,S_0;K,T)}{\partial K}}{\frac{1}{2}K^2\frac{\partial^2 V_c(t_0,S_0;K,T)}{\partial K^2}}. \tag{5.5}$$

The derivation for this equation starts from the Feynman-Kac theorem which provides a partial differential equation (PDE). Then using the Fokker-Planck PDE the above formula can be acquired. It is a very tidy derivation and for more information one could look at Oosterlee and Grzelak (2019). Hence from equation (5.5) the value of the local volatility for different strikes and maturities can be calculated. These local volatility values can then be used in order to price options. From these priced options the implied volatility parameter can be extracted by inverting the Black-Scholes equation. However, from the equation it is also clear that in order to obtain the local volatility parameter an option surface is necessary for different calls at different expiries and strikes. The market doesn't present a smooth call option surface for every possible maturity and strike, that is, the option price surface is not a continuous process. In order to calculate the second derivative in equation (5.5) you would need four different option values to compute a finite difference approximation. With insufficient market data this could result in large approximation errors for the local volatility surface. For a more thorough explanation of this problem see Oosterlee and Grzelak (2019).

A solution to this problem is to rewrite equation (5.5) into terms of the implied volatility. This results in the following equation for the local volatility value

$$\begin{aligned}&\sigma^2_{LV}(T, K)\\ &= \frac{\frac{\partial b}{\partial T} + rK\frac{\partial b}{\partial K}}{1 + K\frac{\partial b}{\partial K}\left(\frac{1}{2} - \frac{a}{b}\right) + \frac{1}{2}K^2\frac{\partial^2 b}{\partial K^2} + \frac{1}{2}K^2\left(\frac{\partial b}{\partial K}\right)^2\left(-\frac{1}{8} - \frac{1}{2b} + \frac{a^2}{2b^2}\right)}.\end{aligned} \tag{5.6}$$

Here $a = \log(\frac{K}{S_0}) - r(T - t_0)$, $b = \sigma_{imp}^2 \tau$ and

$$\begin{aligned}
\frac{\partial b}{\partial T} &= \sigma_{\text{imp}}^2 + 2\,(T - t_0)\,\sigma_{\text{imp}} \frac{\partial \sigma_{\text{imp}}}{\partial T} \\
\frac{\partial b}{\partial K} &= 2\,(T - t_0)\,\sigma_{\text{imp}} \frac{\partial \sigma_{\text{imp}}}{\partial K} \\
\frac{\partial^2 b}{\partial K^2} &= 2\,(T - t_0) \left(\frac{\partial \sigma_{\text{imp}}}{\partial K} \right)^2 + 2\,(T - t_0)\,\sigma_{\text{imp}} \frac{\partial^2 \sigma_{\text{imp}}}{\partial K^2}.
\end{aligned} \tag{5.7}$$

The proof can be found in Appendix B.

Thus, the local volatility can be expressed in terms of the implied volatility. Since the implied volatility surface can be obtained from option prices and then be smoothed in order to obtain the implied volatility for every maturity and strike it becomes easier to obtain the local volatility surface to price options.

Thus, in order to forecast the implied volatility surface for day $t + h$, the following steps are followed.

1. For day t obtain all quoted option prices i for $i = 1, ..., N$.
2. Gather the implied volatilities for all these options by inverting the Black-Scholes equation.
3. Perform an interpolation technique on the implied volatility surface in order to collect the IV values for all expiries and strikes.
4. Use equation (5.6) to obtain the local volatility surface.
5. Implement a Monte Carlo simulation for equation (5.4) using the obtained local volatility surface and level of the underlying on day t. With this Monte Carlo simulation, the asset path will be simulated.
6. Using the simulated asset paths, price the different call and put options with different strikes and maturities for day $t + h$.
7. Again invert the Black-Scholes equation to obtain the predicted implied volatility surface from the option prices.

There are many interpolation techniques for the implied volatility surface and these approaches can become complicated very quickly. A couple of examples are given in Brecher (2006), Fengler et al. (2007) or Kim and Lee (2018). Since the focus on this thesis isn't on the best interpolation of the implied volatility surface, a simpler approach will be chosen as in Zhang et al. (2021). In their research, first the implied volatility values for the given option cross-section is obtained. Then the

following parametric model is assumed for the implied volatility

$$\sigma(m,\tau) = \max(0.01, a_{0,t} + a_{1,t}m + a_{2,t}m^2 + a_{3,t}\tau + a_{4,t}\tau^2 + a_{5,t}m\tau). \tag{5.8}$$

The max function makes sure that the implied volatility doesn't become too small. Notice how the equation resembles the ad-hoc Black-Scholes model. An advantage of this is that we can investigate if the local volatility model will add anything to the ad-hoc Black-Scholes approach. It can be seen as a two step approach. Here the AHBS model used equation (5.8) as a forecast for the implied volatility, but the LVM uses this implied volatility to obtain the local volatility parameter. From here, options are priced and the implied volatility is backed-out. The different parameters can be estimated using OLS with the given implied volatilities from the option data.

A disadvantage of the local volatility model is given in Hagan et al. (2002). The obtained implied volatility smile from local volatility models exhibits opposite behaviour from what is seen in the actual market. This means that when the price of the underlying goes up, local volatility models suggest a lower implied volatility. In practice the opposite is actually true.

5.3 Stochastic Alpha Beta Rho model

The Stochastic-Alpha-Beta-Rho stochastic volatility (SABR-SV) model was first introduced in the paper of Hagan et al. (2002). The main difference with the Black-Scholes model is that the volatility has its own stochastic process. This means that the implied volatility is no longer constant but could be fitted to the the market implied volatility curve. The dynamics of the model are denoted as follows

$$\begin{aligned} dS(t) &= \sigma(t)S(t)^{\beta}dW(t), \\ d\sigma(t) &= \nu\sigma(t)dZ(t), \qquad\qquad \sigma(0) = \alpha. \end{aligned} \tag{5.9}$$

The following restrictions must hold: First of all, $0 \leq \beta \leq 1$. Besides, $\alpha, \nu > 0$. If the value of β is set to one and the value of ν is set to zero, the classic Black-Scholes model dynamics are obtained. Finally, the two Brownian motions are correlated as $E[dW(t)dZ(t)] = \rho dt$, where $-1 \leq \rho \leq 1$. The fact that the volatility and underlying asset level could be negatively correlated is an advantage of the stochastic volatility model (Dumas et al., 1998). The negative correlation is also often visible in the market.

Another advantage of the SABR model over other stochastic volatility models is that there exists an approximation for an analytical equation for the volatility parameter. This means that

options are easily priced, contrary to other stochastic volatility models where a numerical approach has to be implemented to price options. The value of a European option can be obtained by filling in Black's equation where the volatility parameter is given by Hagan's approxmation as

$$\begin{aligned}\sigma(K,S) = &\frac{\alpha}{(SK)^{(1-\beta)/2}\left\{1+\frac{(1-\beta)^2}{24}\log^2 S/K+\frac{(1-\beta)^4}{1920}\log^4 S/K+\cdots\right\}}\cdot\left(\frac{z}{x(z)}\right)\\ &\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(SK)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\nu\alpha}{(SK)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2\right]\tau+\cdots\right.\end{aligned} \tag{5.10}$$

Here

$$z=\frac{\nu}{\alpha}(SK)^{(1-\beta)/2}\log S/K,$$

and

$$x(z)=\log\left\{\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right\}.$$

The derivation to obtain this equation is extremely complicated and far out of the scope of this thesis. For someone who is interested, the paper of Fahrner (2009) is recommended. Notice how the volatility parameter depends on the strike price, this is not the case for the implied volatility in the Black-Scholes model. Therefore, the implied volatility parameter is able to vary in level over the strike price and will not be a flat line. We can see that equation (5.10) could be further expanded. However, there is no need for any further extension according to Hagan et al. (2002). Any more terms will only results in marginal better results which can be disregarded. The expression simplifies for ATM options, see Hagan et al. (2002).

As can been seen from equation (5.9), the SABR model has four different parameters. There is β which controls the skewness of the implied volatility curve, this is usually estimated by traders and not calibrated. The second parameter is ν also known as the volatility of volatility parameter, this controls the curvature of the curve. Then there is α, this sets the initial level of the curve. Finally, there is the correlation parameter, controlling the correlation between the underlying level and the volatility. This parameter has the same effect on the volatility curve as β (Hagan et al., 2002). The reason that ρ is calibrated and β is chosen from the market and not the otherway around has to do with the fact that β also tells you something about the distribution of the underlying asset. This is easier to see in the market and therefore it is simpler to obtain β from market data. Besides, it has been shown by Hagan et al. (2002) that no matter the choice of β the SABR model fits very well to the implied volatility surface. Other examples where the SABR parameters are calibrated

without calibrating the β parameter are West (2005), Nilsson (2008), Gauthier and Rivaille (2009) and Floc'h et al. (2014). All papers give the same reason, the effect of β is the same as ρ, by calibrating both parameters the model becomes overparameterized and the results do not improve or change significantly. Thus, it is not necessary to calibrate the β parameter. For $\beta = 0$ the underlying asset is assumed to follow a normal process, while for $\beta = 1$ the underlying is assumed to follow a log-normal distribution.

For the calibration of the model parameters, the method of the original paper is chosen. Three more parameters are left to fit to the market smile. The other three parameters are obtained by minimizing

$$(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \underset{\alpha,\rho,\nu}{\arg\min} \sum_{(K,\tau)\in\Omega} \left(\sigma^{\text{mkt}}(K,\tau) - \sigma_{Hagan}\left(K,\tau;\alpha,\rho,\nu\right)\right)^2. \tag{5.11}$$

Thus, every day the best parameters are chosen to fit the implied volatilties over the entire cross-section of options. The SABR parameters remain constant over time. A more common approach to fit the SABR parameters is to actually calibrate them per time-to-maturity cross section. Hence, equation (5.11) for a specified τ simplifies to

$$(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \underset{\alpha,\rho,\nu}{\arg\min} \sum_{(K)\in\Omega} \left(\sigma^{\text{mkt}}(K) - \sigma_{Hagan}\left(K;\alpha,\rho,\nu\right)\right)^2. \tag{5.12}$$

Note that equation (5.12) could be simplified by extracting α from the implied volatility from the ATM level, see (West, 2005). This is common to do and reduces the equation from three variables to two variables. However, this is not possible if the SABR parameters are fitted over the entire volatility surface, only if they are fitted to a cross-section for a single expiry time. The advantage of equation (5.12) over equation (5.11) is in the fact that the SABR parameters are no longer constant over the maturity cross-section and could thus better fit to the implied volatility surface. The disadvantage is in the fact that only specific tenors and strikes are available to which the parameters could be fitted, resulting in cross-section with a low number of data points. Besides, an interpolation technique has to be chosen in order to forecast the implied volatility for strike and maturity couples to which the parameters are not calibrated. This is all pointed out in Bogatyreva et al. (2019). The calibration over the entire implied volatility surface could be very ineffective and the calibration per maturity is a preferred method. Considering the literature there are also many more papers where the parameters are fitted per expiry (West, 2005, Gauthier and Rivaille (2009), Floc'h et al. (2014)) instead of over the entire surface (Nilsson, 2008).

A minimization method to solve the equations is for example the Levenberg-Marquardt algorithm (Levenberg, 1944) or the Nelder-Mead method Nelder and Mead (1965). In thesis research the Nelder-Mead algorithm will be chosen. A faster calibration approach can be found in Floc'h et al. (2014).

Considering the calibration procedure of the SABR model there are multiple possibilities, which will all be tested out in this thesis. In the end four different calibration techniques will be tested out.

1. For the β parameter it is either possible to calibrate it or to choose it based on historical data.

2. For all parameters it is either possible to calibrate them over the entire volatility surface or to make them variable over different maturities.

5.4 Heston model

The second type of stochastic volatility model that is examined is the Heston model. The model was developed by Heston (1993) and the stochastic process for the underlying and volatility are given as

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{v(t)}S(t)dW_1(t), \\ dv(t) &= \kappa(\bar{v} - v(t))dt + \gamma\sqrt{v(t)}dW_2(t). \end{aligned} \tag{5.13}$$

Here, the two different Brownian motions are correlated as $dW_1(t)dW_2(t) = \rho dt$. Besides, the following restrictions hold for the parameters, $\kappa \geq 0, \bar{v} \geq 0$ and $\gamma > 0$. The difference with the SABR model is that the Heston model follows a mean reverting process. Thus, the volatility parameter will always revert back to its given mean. The parameter κ determines how fast the process returns to the mean, $\bar{v}$ is the long-term mean of the process and γ is the volatility of the volatility parameter.

Options can be priced in different ways for the Heston model. A simple approach would be to use a discretization technique, where the path of the underlying asset would be split up into many small steps. However, this will take up a lot of computational time and is infeasible when we also need to calibrate the Heston parameters. A faster numerical technique is to implement the COS method by Fang and Oosterlee (2009). Here a Fourier-cosine expansion is implemented to compute the option price.[3]

[3]The code can be found on: https://compfinance.ddns.net/wordpress/book/

For the purpose of pricing options, the different parameters of the Heston model have to be calibrated. A disadvantage of the model over the SABR model is in the fact that there is no closed-form solution for the option price. Therefore, it is more complicated to calibrate the parameters. For the calibration five parameters, $\rho, \bar{v}, v(t_0), \kappa, \gamma$, have to be calibrated to match the market option prices as close as possible. A similar approach regarding the SABR model can be used for this calibration. For day t the following expression can be minimized

$$(\hat{\rho}, \hat{\bar{v}}, \hat{v}(t_0), \hat{\kappa}, \hat{\gamma}) = \underset{\rho,\bar{v},v(t_0),\kappa,\gamma}{\arg\min} \sum_{(K,\tau)\in\Omega} \left(\sigma_{\text{mkt}}(K,\tau) - \sigma_{Heston}\left(K,\tau;\rho,\bar{v},v(t_0),\kappa,\gamma\right)\right)^2. \tag{5.14}$$

Again the Nelder-Mead algorithm can be used for this. After the model variables are calibrated, the parameters can be used to correctly price options for a future day $t+h$. With these options prices the forecast for the implied volatility surface can be obtained.

To summarize, the following steps are taken to predict the implied volatility surface using the Heston model:

1. Obtain the option cross-section data for day t.
2. Minimize equation 5.14 to acquire the Heston parameters, using the Nelder-Mead algorithm. Here the implied volatility for the Heston model is obtained using the COS method.
   2.1. Options are priced using the parameters and the COS method.
   2.2. These option prices are used to invert the Black-Scholes equation.
3. Gather the option cross-section data for day $t+h$.
4. Price all options for this day. The Heston parameters from step 2 are used and filled into the COS method.
5. Repeat the above process for all days t

For this research, the Heston parameters were also bounded above. The reason for this is that it sometimes happened that the optimization algorithm wouldn't work due to strange results coming from the COS method. Another solution that could be implemented was to instead calibrate the Heston parameters to the option price surface contrary to the implied volatility surface. However, according to Christoffersen and Jacobs (2004) it is important to take loss function for the optimization the same as the number we are trying to predict. Therefore, the decision is made to set a maximum for the Heston parameters and minimize equation (5.14) rather than minimizing to the option price surface.

5.4.1 Heston vs SABR

We briefly discuss a couple of differences between the Heston model and SABR model. The obvious similarity is that both models are stochastic volatility models where the volatility follows a stochastic process. This means that both models are able to reproduce the implied volatility curve seen in the market. First we note some (dis)advantages of the SABR model.

Advantages SABR:

1. The SABR model still has an analytical approximation for the option price. Therefore, it is extremely easy to calibrate the model parameters to the market data.
2. The SABR parameters are easily interpretable on what the effect is on the implied volatility curve. This makes the model easy to understand

Disadvantage SABR:

1. An issue with the SABR model is that equation (5.10) is not arbitrage free. The problem is that the implied probability density function can become negative using Hagan's approximation. The issue is discussed in for example Hagan et al. (2014). Pricing options with the COS method for SABR discussed in Grzelak and Oosterlee (2014) is a possible solution to this inconsistency.

For the Heston model we have the following (dis)advantages.

Advantages Heston:

1. The fact that the volatility process is a mean-reverting process in the Heston model is an advantage. The reason is that the volatility in the market also shows mean-reverting behaviour.

Disadvantage Heston:

1. The Heston model doesn't have a closed form solution for option prices. This makes it more complicated to calibrate the parameters to the option cross-section. A fast approach to calibrate the variables is to use the COS method introduced in Fang and Oosterlee (2009). However, this calibration technique still takes much more computational time compared to the SABR model.
2. According to Lemaire et al. (2021), the Heston model doesn't fit the implied volatility smile well for short time-to-maturity options.

5.5 Feedforward Neural Network

The feedforward neural network is a type of neural network where the nodes do not form a loop. Thus, data enters the input layers and then moves through all the hidden layers leading to the output layer. Contrary to a recurrent neural network where the nodes have different connections with each other. The choice for a neural network is due to their universal approximation property, see e.g. Cybenko (1989) or Hornik et al. (1989). This tells us that the neural network can model any kind of function, even non-linear relationships. As has been discussed before, the behaviour of the implied volatility surface is highly non-linear, hence a model that could capture this behaviour is favorable. Due to this property neural networks are a good solution to forecast the pricing error surface.

The feedforward neural network in this thesis tries to find a function $\hat{f}(\mathbf{x})$ which maps the input data, moneyness and time-to-expiry, to the output which is the error surface between the true implied volatility surface and the forecasted one from the parametric models. Using deep learning it is possible to obtain a function f which is able to explain some highly non-linear relationship in the error surface. The basics of a feedfoward neural network are very well explained in Fan et al. (2021). The training of the model is performed by minimizing the loss function.

To give a bit more of mathematical background. Consider option i on day t, the vector $\mathbf{x}_{i,t} \in \mathbb{R}^2$ represents the moneyness and expiry of this option. The feedforward neural network, $f : \mathbb{R}^2 \to \mathbb{R}$, is built in the following way:

$$\underset{d_l \times 1}{\mathbf{z}_l} = h\left(\underset{d_l \times d_{l-1}}{\mathbf{A}_{l-1}} \underset{d_{l-1} \times 1}{\mathbf{z}_{l-1}} + \underset{d_l \times 1}{\mathbf{b}_{l-1}}\right), \text{ for } l = 1, \cdots, L, \quad \underset{1 \times 1}{f(\mathbf{x}_{i,t})} = \underset{1 \times d_L}{\mathbf{A}_L} \underset{d_L \times 1}{\mathbf{z}_L} + \underset{1 \times 1}{\mathbf{b}_L}, \tag{5.15}$$

where $\mathbf{z}_0 = \mathbf{x}_{i,t}$ and the equation is performed iteratively. Here $\mathbf{A}_{l-1}$ is known as the transformation matrix, $\mathbf{b}_{l-1}$ as the intercept or bias vector and the function $h(\cdot)$ is the activation function. Thus we start with the vector $\mathbf{x}_{i,t}$ and iteratively this vector is transformed and then put into the activation function. In the end we have a function f which hopefully is able to explain the error surface coming from the parametric model. Finally, L is the number of hidden layers in the neural network and d_l is the number of neurons in this hidden layer.

Figure 4 shows an example of how a feedforward neural network is built. In the figure we see the input layer with two neurons which are the predictors. For this thesis it will be the moneyness and time-to-maturity for different options. Then the figure shows two hidden layers with each three neurons. All neurons are connected with each other and a function $\hat{f}$ will modify the moneyness

and expiry parameters to get to the prediction value. The arrows represent the different weights the neural network will give. Finally, we have one output layer which will be a prediction for the error between the predicted and true implied volatility surface. The choice of the number of hidden layers is a complicated problem. A shallow network is easier to train, since less parameters are present, however a more deep network could be able to find the highly non-linear relation for the implied volatility surface, see Gu et al. (2020).

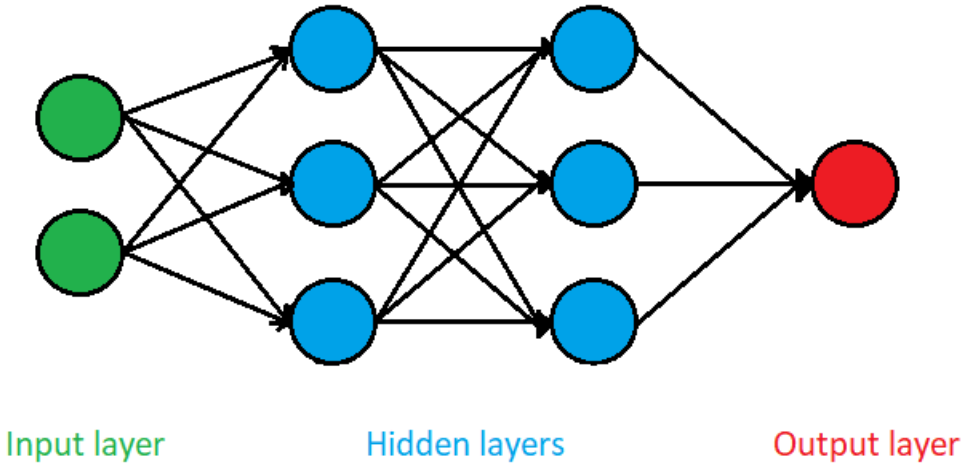

Figure 4: Example of a feedforward neural network

A feedforward neural network will be used to improve the forecasting power of the implied volatility surface. For this approach the paper of Almeida et al. (2022) is carried out. The method from Almeida et al. (2022) applied to our parametric model works as follow. We have the implied volatility surface for options with different maturities (τ) and the moneyness (m). The implied volatility surface on day t is estimated by the parametric model. This results in different implied volatility values $\hat{\sigma}_p(m_{i,t}, \tau_{i,t})$ where i are the different options in the dataset. The fitting of the parametric model leads to a difference between the true volatility surface and the predicted one. This is called the error surface and is given by

$$\hat{\epsilon}_p(m_{i,t}, \tau_{i,t}) = \sigma(m_{i,t}, \tau_{i,t}) - \hat{\sigma}_p(m_{i,t}, \tau_{i,t}).$$

The next step is to obtain an estimation of the error surface by training the feedforward neural network. This is done by minimizing over all options

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_p(m_{i,t}, \tau_{i,t}) - f(m_{i,t}, \tau_{i,t}))^2. \tag{5.16}$$

Thus, the feedforward neural network will train on a set of error surfaces and comes up with the best function $\hat{f}$ to represent the error surface. Finally, the two-step approach should give a better approximation for the implied volatility surface, given by $\hat{\sigma}_p(m_{i,t}, \tau_{i,t}) + \hat{f}(m_{i,t}, \tau_{i,t})$.

Just as in Almeida et al. (2022), two types of forecasts could be investigated. First of all, the implied volatility for different options could be predicted based on data of options on the same day. This means that every day, the option data is split up into two sets. The first set is used to train the neural network, where the second set is used as testing set. Therefore, part of the options on day t is used to predict a second part of the options on this same day. Second, the neural network model could be trained on all the options on day t and then a prediction for the implied volatility surface could be made, h days ahead. For h different values could be taken into account in order to investigate how far ahead the model could predict the surface.

Note that contrary to Almeida et al. (2022), this thesis also focus on the comparison of our parametric models. The (ad-hoc) Black-Scholes model, the local volatility model, the Heston model and the SABR model. This means that the results section also contains the prediction results of solely these models without applying the two-step approach.

In this research the implied volatility will be used as the error measure to comment on the prediction power of the different models. The error measure which will be used is the implied-volatility-root-mean-squared-error and is given as

$$IVRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\sigma_{i,t} - \hat{\sigma}_{i,t})^2} \tag{5.17}$$

Here $\sigma_{i,t}$ are the true implied volatilities on a given day and $\hat{\sigma}_{i,t}$ are the forecasted implied volatilities. For single model all the IVRMSE's are averaged over all the prediction days to obtain the final forecast measure.

5.5.1 Architecture of the neural network

There are many different possibilities in choosing the set-up for the neural network. A couple of examples where choices have to be made are

1. The number of hidden layers.
2. The number of neurons in each layer.
3. The activation function for each layer.

Even if these would be optimised, other hyperparameters as the number of epochs or the batch size for each epoch has to be determined. An approach to this could be to use some sort of grid search, where all sorts of combinations are tested out and the best one is kept in the end. However, this is so time consuming that it would unreasonable. Following Gu et al. (2020) it is also noted that an infinite mixture of choices could be tested out and that it becomes an unrealistic task.

The choice is made to build the feedforward neural network comparable with Gu et al. (2020) and Almeida et al. (2022). In their papers the pyramid rule introduced in Masters (1993) is used. With this method, five different neural networks are built. Each new network will increase in the number of hidden layers, but the number of neurons in this layer will decrease. Thus, the first network architecture will have a single hidden layer with 32 neurons. The second network has two hidden layers with 32 and 16 neurons in each layer respectively. This continues till the final fifth network which has five hidden layers, where each layer has 32, 16, 8, 4, 2 and 1 neuron. The advantage of this approach is considered in Almeida et al. (2022), it creates a good overview of how well a deep neural network improves the forecasting of the implied volatility surface. It will also be interesting to see if this new dataset would create similar results. Meaning, if the best network architecture overlaps with Almeida et al. (2022) for more recent data and the SABR or LVM model.

The next decision is regarding the activation function for the layers. As in Almeida et al. (2022) the sigmoid activation function ($h(x) = 1/(1 + e^{-}x)$) is used for all different layers. Finally, the minimization algorithm has to be chosen which will be used in the brackpropogation. A lot of neural networks use a stochastic gradient descent algorithm in their architecture. In this research this decision is not made. It doesn't make sense to train the feedforward neural network on just a part of the implied volatility surface since it could overfit on this specific part but give bad correction results for other parts of the surface. It would be better to use a algorithm which takes into account the entire dataset and train the neural network on this. The approach we use for this is to implement the feedforward neural network with the ADAM algorithm. By choosing the batch size in such a way that all data is used as an input for this algorithm. This makes us step away from a SGD approach and we implement a batch gradient algorithm taking into account all datapoints each epoch. Every time the neural network is trained, 15 epochs are used, this is chosen since we found that after this amount the loss function almost doesn't decrease anymore.

6 Results

This section discusses the results of the different models that were implemented. First, we briefly go over how the models were implemented. In section 6.2 some results of the forecasting without a neural network are presented. In the final section the results with the feedforward neural network are presented.

6.1 Implementation

For every model in the results section, four prediction exercises are tested. For the first three, the models are calibrated using a day-to-day estimation after which the implied volatility surface for the future is forecasted. Thus, for every day t in the dataset the five models (BS, AHBS, LVM, SABR, Heston) are trained on the entire option cross-section of that day. The second step is to predict the implied volatility surface, which will be done for $t + h$ days ahead. In this thesis three values for h are considered, $h = 1$, $h = 5$ or $h = 21$. Representing 1 business day ahead, 1 week ahead and 1 month in the future. Considering these values, we will test if the different models are able to pull out the correct details from today's implied volatility surface to predict the future surface.

The fourth prediction exercise is a same-day problem. For every day t in the option data the option set is split up into two parts, a training and testing part. The five models are fitted on the training data and then the prediction performance for the testing data on the same day is assessed. The splitting of the data is done in line with Almeida et al. (2022). All options with a strike divisible by 10 are used for training purposes and the rest for the testing. Just as in their paper, around 60% of the daily option data is assigned to the training data and 40% to the testing data.

6.2 Parametric model performance

Table 2 shows the IVRMSE (in %) of the model forecasts for the implied volatility surface. The four different columns show the results for 1, 5, 21 or same-day prediction. The table shows two rows for the SABR model. For the first row (SABR) the model parameters were fitted for every maturity cross-section on day t and β was fixed. For this thesis a value of $\beta = 0.5$ is chosen. This means that we assume that the returns are not fully normally or log-normally distributed, but they follow a constant-elasticity of variance process. This is by far the most widely used method for

picking the value of β. Besides, Hagan et al. (2002) mentions that the results won't be greatly impacted for different β values. The second row (SABR (beta)) calibrated the model parameters including β to every maturity cross-section on day t. The results for the SABR model where the parameters are fitted over the entire option cross-section without considering the different expiry cross-sections can be found in appendix C. These are less important since calibrating the model to every time-to-maturity cross-section is always superior. The bold numbers represent the best performing model for every prediction exercise.

A couple of things stand out in the table. It shows that the Black-Scholes model, by far, is the worst performer for all four different forecasts. The IVRMSE is much higher compared to the other models however, the contrast between the different predictions is smaller compared to the other models. For example, the IVRMSE for the AHBS model almost triples predicting 1 day or 21 days ahead. For the BS model this is definitely not the case. What also is interesting about the data in this table is that a small correction on the BS model in the form of the AHBS model already greatly improves the forecasting results.

The results for the SABR models, as shown in Table 2, indicate that not calibrating the β parameter outperforms the model where this parameter is fitted, except for the same-day prediction. This is intuitive, as previously mentioned when calibrating the β parameter the model could be overfitted and this comes back in the results. The SABR model gives an IVRMSE (%) of 1.89 for $h = 1$, of 3.20 for $h = 5$, of 5.80 for $h = 21$ and of 0.71 for the same-day prediction, these are the best forecast results.

Another interesting aspect of the table are the results for the local volatility model. We see that this model is actually outperformed by the AHBS model. Therefore, smoothing the implied volatility surface with the AHBS model and then using a LVM to predict the implied volatility surface is not beneficiary for the forecasting power. However, it could be that using a different smoothing technique would show better results.

The final observation we make from the table is the fact that the stochastic volatility models are clearly the best model performers in predicting the implied volatility surface. Besides, the SABR model outperforms the Heston model on all four prediction exercises. Although, for forecasting the future implied volatility surfaces it is not by much. A more striking result from the table is the fact that the same-day prediction result for the Heston model is very poor. Even the AHBS model gives a better result. A possible explanation might be that the Heston model does not perform well predicting the implied volatility when the market is extremely volatile. Figure 5 gives

a breakdown of the IVRMSE of the same-day prediction per day of the Heston model. The blue part of the figure represents the COVID-19 crash, where the market was very volatile. Clearly, the model performs much worse compared to the rest of the same-day predictions. Leaving out these days would already reduce the IVRMSE of the Heston same-day prediction to 0.83, resembling the results of the SABR model. This inconsistency may be due to the overfitting of the model parameters. The model was also fitted where the parameters were calibrated without any upper bound restrictions. However, this did not result in a better IVRMSE.

Table 2: **IVRMSE of the model predictions for different prediction exercises.**

Models prediction (IVRMSE)	h = 1	h = 5	h = 21	Same-day
BS	9.30	9.57	10.70	8.38
AHBS	2.50	3.81	6.80	1.48
LVM	2.79	4.11	7.05	1.65
SABR	**1.89**	**3.20**	**5.80**	0.71
SABR (Beta)	2.00	3.35	6.16	**0.66**
Heston	2.26	3.64	5.94	1.63

This table provides the IVRMSE (%) of the four predictions exercises: same-day, 1, 5 and 21 days ahead of the implied volatility surface. The rows represent the five parametric models that were considered. The SABR model is shown twice, one where β is calibrated (SABR) and one where it is not (SABR (beta)). The bold numbers show the best preforming model depending on the prediction exercise. The sample ranges from January 2017 to December 2021.

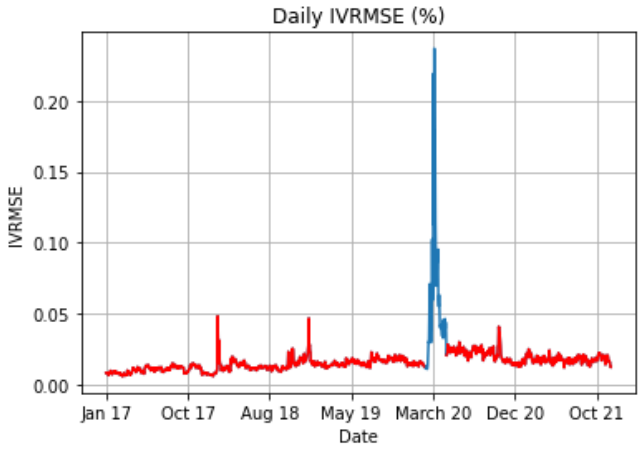

Figure 5: **Daily IVRMSE for the Heston same-day prediction.** This figure shows the daily IVRMSE of the Heston model for the same-day prediction exercise. The sample period runs from January 2017 to December 2021. The blue part of the figure is the COVID-19 crash during March 2020.

For the Black-Scholes, ad-hoc Black-Scholes, SABR and Heston model it is also possible to

examine the parameter values over the lifespan of the dataset. Figure 6 shows the intercept value when the Black-Scholes model is fitted, in equation (5.1), to the option data on every day t. The Black-Scholes model assumes a constant volatility therefore, the parameter value a_0 is just the average implied volatility for every day in our option cross-section. Clearly, during more uncertain times the predicted implied volatility goes up. Take for example the Covid-19 crisis in March 20202 where the implied volatility parameter is extremely high.

Figure 7 is already much more interesting to consider. Here the six different parameters of the AHBS model are plotted over time. We see that three parameters, almost always, have a negative impact on the implied volatility parameter (a_0, a_2 and a_5) and three parameters have a positive influence (a_1, a_3 and a_4). Considering equation (5.1), it is clear that the moneyness has the greatest effect on forecasting the implied volatility surface, while the intercept also has a great impact. Note that the intercept is not multiplied with m or τ. We also see that during March 2020 the estimation of the parameters shoots up and down with very large jumps. This is because the volatility during that period was much more unstable. This could be a reason that the results of the model prediction are worse compared to Almeida et al. (2022) where a different time period is used. The figures where the parameters are calibrated to the same-day option cross-section can be found in Appendix C.

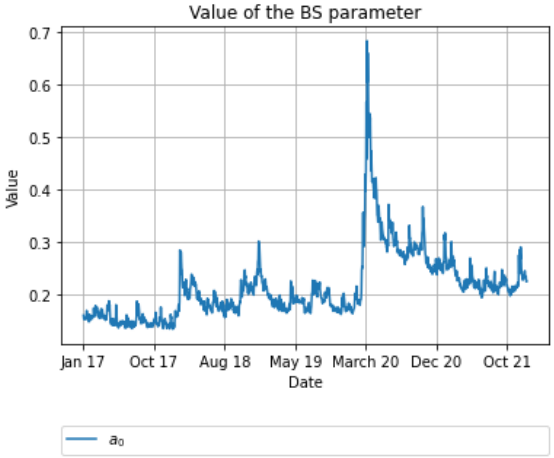

Figure 6: **BS parameters.**

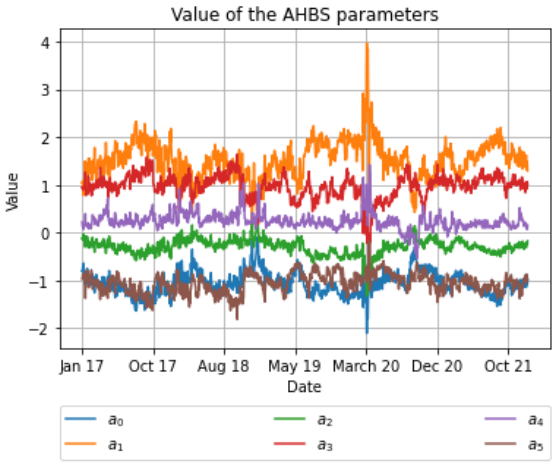

Figure 7: **AHBS parameters.**

These two figures show the calibrated model parameters for the Black-Scholes and ad-hoc Black-Scholes model. The sample period ranges from January 2017 to December 2021.

Figure 8 provides an overview of the different SABR parameters fitted to the option data. The results presented below are for the case where β was set to 0.5 and the parameters were calibrated

over the entire option cross-section for day t and not per maturity cross-section. Besides, these parameters were calibrated to all options for day t, the parameters for the same-day prediction differ a bit an can be found in Appendix C. They don't deviate much from the showed figures here. In those figures it also turns out that when β is calibrated, the value is almost always equal to one.

From the chart below, a couple of things can be seen. First of all, the value for α which sets the initial level of the implied volatility curve shoots up during March 2020. It also follows an extremely similar behaviour compared to Figure 6, mimicking the average volatility. Second, we consider the level of ν which was the volatility of volatility and controls the curvature of the implied volatility surface. We see in the figure that in times of stress the value decreases, this would result in a flatter implied volatility curve. A reason for this could be that in times of extreme market behaviour, the volatility is high for all the options. It doesn't matter if the options are deep in-the-money or out-the-money. This would results in a much flatter implied volatility curve. Finally, ρ is examined which was the correlation between the underlying asset and the volatility. Clearly, when there is much tension in the market asset values go down while the volatility goes up. This result is definitely visible in the figure, during the COVID-19 crisis the value of ρ even reaches -1.

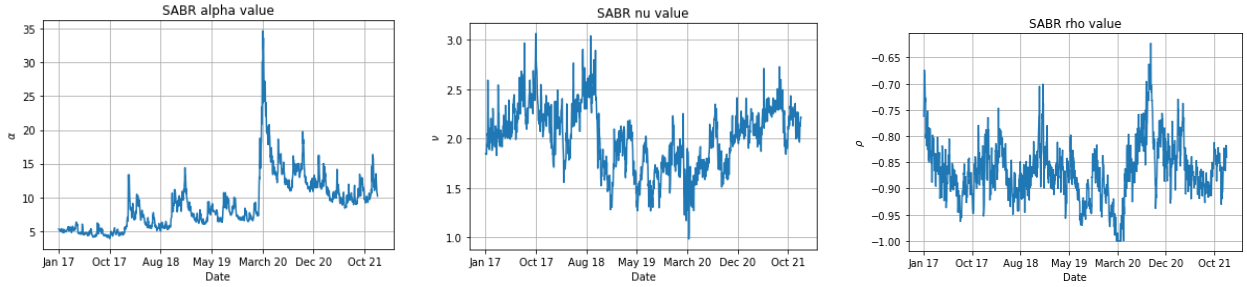

Figure 8: **SABR parameters**. These three figures show the calibrated SABR parameters. It shows the value of α, ν and ρ over the sample period ranging from January 2017 to December 2021.

Figure 9 shows the calibration of the Heston parameters over the dataset. The figures look a bit off. For example, the plot of γ and κ start with some large values for these parameters after which they are almost always equal to zero. More interestingly is the fourth plot representing v_0. This sets the initial value of the volatility parameter. Just as for the SABR model the level of this parameter follows the average implied volatility over the time period. Another interesting figure is the third one representing ρ. Surprisingly, the volatility and the underlying asset become less correlated during the very volatile time period. This is exactly the opposite as we saw for the SABR model and not a clear explanation can be found for this.

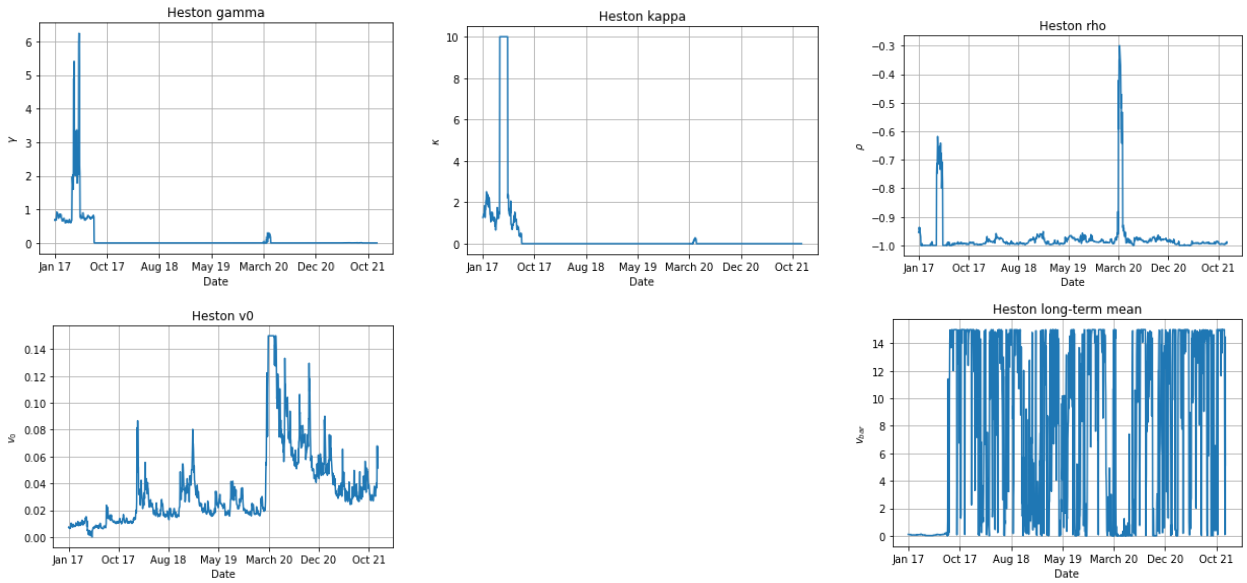

Figure 9: **Heston parameters.** These five figures show the calibrated Heston parameters bounded from above. It shows the value of γ, κ, ρ, v_0 and $\bar{v}$ over the sample period ranging from January 2017 to December 2021.

6.3 Feedforward neural network performance

Here we present the IVRMSE results for the feedforward neural network in table 3, 4, 5 and 6 for the different prediction exercises. In all tables, the model forecast results are again reported for a clear overview. The bold numbers represent the neural network which had the best forecasting power per parametric model. The results for the SABR model calibrated to the entire implied volatility surface instead of every maturity cross-section can be found in appendix D.

What stands out in the tables is that the feedforward neural network has a great positive impact on the forecasting results for the parametric models. For example, the IVRMSE of the Black-Scholes for the 1-day ahead prediction reduces with 75% and for the same-day prediction with 79%. The correction power is weaker for the models where the original prediction was already much better. For instance, the IVRMSE of the SABR model 1-day ahead prediction only reduces with 4% and for the same-day prediction with 21%.

Closer inspection of the tables also shows that the effect of feedforward neural network is less and less present when predicting the implied volatility surface further away in the future. Take for instance the best predictions of the BS model. The reduction in the error is 79% for the same-day forecast, 75% for predicting 1-day ahead, 62% for the 5-day ahead prediction, and 'only' 43% for the 21-day ahead forecast. This decline in the reduction of the IVRMSE is present for all parametric

models over all the prediction exercises.

A second observation which can be made from the tables is that the two stochastic volatility models are still the best performers for forecasting the implied volatility surface. However, the results are much closer this time. Such as the 5-day and 21-day ahead prediction where there is almost no difference between the BS, AHBS and LVM model and less difference between the SABR and Heston model. Although, the error of the feedforward neural network converged after 15 epochs, it might be that increasing the number of epochs would put all models even closer to each other. This is however not reachable considering computational time.

The tables also shows that the neural network with three or four (sometimes two) hidden layers results in the best forecasting outcomes. Especially, considering table 5, where almost all best results are in the 3-hidden layers column. Therefore, a too shallow neural network seems unable to capture the highly non-linear behaviour of the implied volatility surface. While a too deep neural network seems to be overfitted on the training data. These results are inline with Almeida et al. (2022).

Table 3: **IVRMSE of the model predictions for 1-day ahead forecast.**

1-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
BS	9.30	2.46	2.32	2.33	**2.30**	2.33
AHBS	2.50	2.29	**1.89**	1.90	**1.89**	2.29
LVM	2.79	2.65	2.05	**2.04**	2.05	2.64
SABR	1.89	1.88	**1.82**	1.87	1.85	1.87
SABR (Beta)	2.00	1.99	1.98	**1.95**	**1.95**	1.98
Heston	2.26	1.93	1.92	1.89	**1.87**	**1.87**

Table 4: **IVRMSE of the model predictions for 5-day ahead forecast.**

5-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
BS	9.57	3.72	**3.60**	**3.60**	**3.60**	3.61
AHBS	3.81	3.71	3.60	**3.59**	3.60	3.67
LVM	4.11	4.10	3.61	3.61	**3.60**	3.65
SABR	3.20	3.18	**3.14**	3.17	3.17	3.17
SABR (Beta)	3.35	3.34	**3.31**	3.34	3.34	3.34
Heston	3.64	3.42	3.41	**3.38**	**3.38**	**3.38**

Table 5: **IVRMSE of the model predictions for 21-day ahead forecast.**

21-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
BS	10.70	6.36	6.16	**6.15**	**6.15**	6.18
AHBS	6.80	6.25	6.22	**6.15**	6.28	6.27
LVM	7.05	6.26	6.25	**6.18**	6.20	6.22
SABR	5.80	5.71	**5.67**	**5.67**	5.68	5.68
SABR (Beta)	6.16	6.09	**6.05**	6.07	6.06	6.06
Heston	5.94	5.85	5.84	**5.79**	**5.79**	5.82

Table 6: **IVRMSE of the model predictions for same-day forecast.**

same-day prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
BS	8.38	1.92	**1.75**	1.77	1.82	1.90
AHBS	1.48	0.97	**0.90**	0.92	**0.90**	0.95
LVM	1.65	0.99	0.93	**0.92**	**0.92**	0.94
SABR	0.71	0.66	0.57	**0.56**	0.59	0.70
SABR (Beta)	0.66	0.62	**0.53**	**0.53**	**0.53**	0.65
Heston	1.63	1.02	0.99	0.97	0.97	**0.95**

These four tables provide the IVRMSE (%) of the four predictions exercises: same-day, 1, 5 and 21 days ahead. The columns indicate whether or not a feedforward neural network was used and how many hidden layers were used for the neural network. The rows represent the five parametric models that were considered. The SABR model is shown twice, one where β is calibrated (SABR) and one where it is not (SABR (beta)). The bold numbers show what architecture has to be used for every parametric model to obtain the best forecasting results. The sample ranges from January 2017 to December 2021.

Table 7 and table 8 give an overview of how the parametric models perform over the different option classes. The tables shows the performance of predicting the implied volatility surface using only a parametric model or the model with the neural network correction. Only the combination of a parametric model together with the 3-hidden layer feedforward neural network is presented, this was the best performing neural network. The bold numbers represent the model which had the best prediction power for the certain option class.

We first start with table 7. Only considering the Black-Scholes model, it performs the best for OTMP options. This means that the average volatility of the dataset is most similar to the volatility of these type of options. The SABR model with the neural network correction is the best performing model for OTMC, ATM, OTMP, DOTMP options and options with a longer expiry. However, the Heston model performs better for DOTMC options and options with a short time-to-expiry. Considering the market, these options are usually the most illiquid ones, this generally means that the price of these options say less about the implied volatility curve as they are set by

a smaller set of traders. This also makes it harder to forecast these categories of options. It seems that the Heston model does this better compared to the SABR model.

Table 8 presents the results of the same-day prediction. Just as for the overall results in the tables above, the SABR model with the neural network and calibrating the β parameter outperforms all the other models. We also see that the poor performance of the Heston model for the same-day prediction is not due to a certain option category, instead the Heston model is a weak model over all different option classes.

Table 7: **IVRMSE of the model predictions for different option classes.**

1-day ahead prediction	DOTMC	OTMC	ATM	OTMP	DOTMP	Long	Short
BS	10.55	9.41	6.81	2.69	12.87	8.04	9.92
BS + NN3	3.72	2.47	2.35	2.08	2.20	1.89	2.50
AHBS	5.43	2.51	2.08	1.71	2.33	2.08	2.48
AHBS + NN3	3.45	2.18	1.70	1.62	1.79	1.55	2.04
LVM	5.60	2.62	2.28	1.86	2.37	2.15	2.53
LVM + NN3	3.47	2.17	1.71	1.64	1.77	1.55	2.05
SABR	4.22	1.96	1.74	1.64	1.75	1.53	2.06
SABR + NN3	3.99	**1.92**	**1.64**	**1.57**	**1.72**	**1.40**	2.02
SABR (Beta)	4.57	2.11	1.83	1.71	1.85	1.63	2.18
SABR (Beta) + NN3	4.36	2.08	1.74	1.66	1.84	1.54	2.14
Heston	3.71	2.49	1.77	1.76	2.59	2.34	2.21
Heston + NN3	**3.50**	2.20	**1.64**	1.59	1.86	1.79	**1.92**

Table 8: **IVRMSE of the model predictions for different option classes.**

Same-day prediction	DOTMC	OTMC	ATM	OTMP	DOTMP	Long	Short
BS	13.03	10.07	7.33	2.42	10.66	7.72	8.68
BS + NN3	3.77	1.92	1.81	1.49	1.59	1.56	1.85
AHBS	3.43	1.80	1.57	0.71	1.45	1.38	1.53
AHBS + NN3	2.66	1.43	0.68	0.53	0.67	0.74	0.97
LVM	3.71	1.86	1.61	0.72	1.49	1.41	1.59
LVM + NN3	2.71	1.48	0.69	0.55	0.67	0.77	0.99
SABR	1.39	0.78	0.74	0.65	0.62	0.80	0.67
SABR + NN3	1.37	0.65	0.51	0.50	0.51	0.52	0.58
SABR (Beta)	1.26	0.71	0.71	0.61	0.56	0.74	0.62
SABR (Beta) + NN3	**1.22**	**0.60**	**0.48**	**0.46**	**0.49**	**0.50**	**0.54**
Heston	3.87	2.10	1.28	1.38	2.20	2.09	1.69
Heston + NN3	3.6	1.74	1.10	1.16	1.42	1.35	1.44

These two tables provide the IVRMSE (%) of two predictions exercises: same-day and 1-day ahead. The columns show the different option categories for different moneyness (S/K) and time-to-maturities. The rows represent the five parametric models that were considered or the parametric model correct with the neural network. The SABR model is shown twice, one where β is calibrated (SABR) and one where it is not (SABR (beta)). The bold numbers show which model performs the best for each option class. The sample ranges from January 2017 to December 2021.

Finally, Figure 10 gives a visual presentation of the worst performing model (Black-Scholes) and best performing model (SABR). The figures show the 1-day ahead forecasting results. The figures show the true implied volatility curves and predicted curves for a single day of the dataset over all maturities on that day. The top two figures are the actual implied volatility and the parametric model forecast and the bottom two plots are the true implied volatilities and the parametric model including the correction of the 3-hidden layer neural network.

Solely the Black-Scholes model predicts a flat implied volatility surface, as is expected. The neural network correction has a significant impact on the forecasting power of the model. For the SABR model it is a bit harder to see the correction effect of the neural network. However, considering the DOTMC and DOTMP options we see that the neural network corrects the predicted implied volatility such that the true and predicted values are closer to each other. From the figures it is also understandable that the correction effect on the Black-Scholes model is much greater in relative terms compared to a parametric model, as SABR, which is already able to give a good forecast for the implied volatility surface. Table 3 also showed that the implied volatility of DOTMC options is harder to predict compared to DOTMP options. This clearly comes back in the Black-Scholes figure. Here we see that the neural network correction is much worse for the DOTMC

options and it doesn't reproduce the smile shape we see in the market.

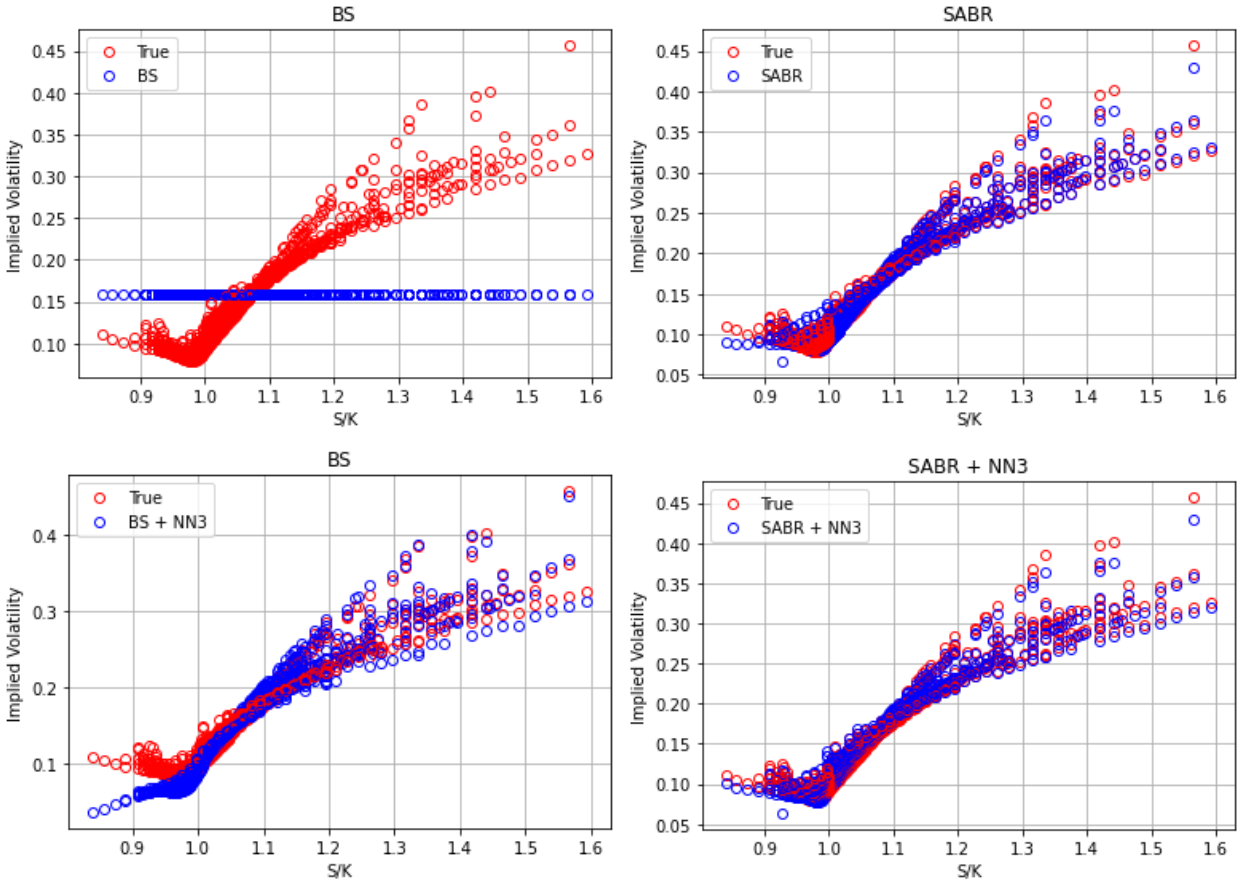

Figure 10: **Implied volatility surface 1-day ahead forecast.** The top two figures show the implied volatility smiles from the data in red. In blue the 1-day ahead predicted implied volatility smiles from the parametric models (BS and SABR) are plotted. The bottom two figures show the same in red and in blue the correct implied volatility smiles for the parametric models with NN3 are presented. The x-axis represents the moneyness (S/K). The sample data is from January 3, 2017.

7 Conclusion and future research

In this final section of the report we present the conclusion of the thesis and answer the research questions. Besides, an overview of some future work that could be conducted is presented.

7.1 Conclusion

Options are used by all sort of financial institutions for hedging or trading purposes. An important factor for option pricing is the implied volatility surface. Besides pricing options, the implied volatility surface is often used to manage option positions. Precisely forecasting the implied volatility surface is important to make the correct trading decisions. This thesis examined five different parametric models and their prediction power for the implied volatility surface. As a second step, all models were corrected using a feedforward neural network. This neural network was first trained on the error surfaces obtained from the five different parametric models and then applied a correction to improve the forecast of the implied volatility surface. For this research, option data ranging from January 2017 till December 2021 was used.

The first considered model was the classic Black-Scholes model. The model always predicts a flat implied volatility surface but is a useful benchmark model. Second, the ad-hoc Black Scholes model was presented. Here the volatility parameter is no longer constant, it depends on the moneyness and maturity of the option. The third approach was to use a local volatility model to predict the implied volatility surface. The volatility parameter is controlled by the underlying asset path and time. In order to obtain the local volatility parameter a smoothing technique had to be applied to the implied volatility surface. For this thesis a simple approach in the form of the AHBS was chosen. The advantage was that it would be an interesting comparison between the AHBS and local volatility model. The disadvantage is that the forecast using a simple smoothing technique could be much worse. The final two models were both stochastic volatility models. For these type of models the volatility has its own stochastic process. A benefit for these models is that the underlying asset and the volatility could now be negatively correlated, which is often visible in the market. The first stochastic volatility model was the SABR model. This model is easy to understand and has a closed-form solution for the option prices, making it incredibly simple to calibrate it to market data. The second model was the Heston model. An advantage of this model is that the volatility follows a mean-reverting process. However, it is harder to calibrate, in this thesis the COS method was used for the fitting of the model parameters.

The parametric models were tested on four different prediction exercises. Three of them were to fit the models on all the option data on day t and forecast the implied volatility 1, 5 or 21 days ahead. The fourth variant was to calibrate the model parameters on part of the data on day t and perform a same-day forecast for the volatility surface. We found that the benchmark Black-Scholes model clearly performs the worse as it only predicts a constant volatility parameter. The local volatility model did not outperform the AHBS model however, this might be due to the simple smoothing process for the implied volatility surface. Besides, the two stochastic volatility models performed the best, where the SABR model outperformed the Heston model for all prediction exercises. With these results we are also able to answer the first research question which stated how the SABR model performed against the other parametric models.

The SABR model outperforms other parametric models in predicting the implied volatility surface. Only the Heston model is able to slightly compete with the SABR model. It is important to keep in mind that for the SABR model the best performance is achieved when not calibrating β but it is calibrated per time-to-maturity cross-section. However, for the same-day prediction exercise it is beneficiary to calibrate the β parameter.

The next step was to implement a feedforward neural network to make a correction for the errors coming from the different parametric models. The feedforward neural network should improve the forecasting results due to its universal approximation property. Most hyperparameters for the network were chosen accordingly with Almeida et al. (2022). Five different neural networks were tested, each with a different amount of hidden layers and a different amount of neurons in these layers. These followed a pyramid scheme, where shallow layers would have more neurons and deeper layers would have less.

The study showed that the feedforward neural network has a positive effect on the forecasting of the implied volatility surface for all four prediction exercises. The impact is the biggest for the parametric models that perform the worst and reduces for the stochastic volatility models. Besides, the research has also shown that the effect of the neural network is less present for predictions further away in the future. The feedforward neural network also lowers the error of the BS, AHBS and LVM model to the same level. Finally, the results showed that overall a neural network with three of four hidden layers performs the best. The 1-layer neural network cannot capture the behaviour of the implied volatility surface well and the 5-hidden layer network is likely overfitted. The second research question stated how the feedforward neural network affects the forecasting power of the SABR model and how this compares to the other models.

Applying a correction on the error surface of the SABR model using a feedforward neural network has a positive effect on the prediction performance. After the correction, the SABR model still outperforms the other parametric models with their corrections. It should be considered that the best performance is reached if a neural network with three or four hidden layers is used.

The final observation which was made, was regarding the forecasting of the implied volatility for the different option categories. We saw that the models perform the worst for DOTMC options and in lesser extent DOTMP options, while for closer at-the-money options the IVRMSE was much lower. This was the case for the parametric models as for the models including the correction. Besides, the implied volatility was easier to forecast for options with a longer maturity (> 60 days) than options with a short time-to-expiry. A possible explanation for this is that deep out-of-the-money and options with a short time-to-maturity are less liquid and present in the market. Regarding the forecast of future implied volatility surfaces, the SABR model almost always performed the best out of all the models except for the DOTMC case and options with a short expiry date, where the Heston model outperformed all the different choices. Therefore, it seems that the Heston model should be chosen for forecasting less liquid options and the SABR model for the other option classes. For the same-day forecast of the implied volatility surface the SABR model where β was calibrated clearly outperformed all the different possibilities.

7.2 Future research

In this final section a couple of research points are discussed which could help in establishing a better overview of the research. First of all, the local volatility model can be changed in order to likely obtain better results. The smoothing process for the implied volatility process for this thesis was a really simple approach, possibly leading to some poor results. A more advanced smoothing technique as in Brecher (2006) or in Kim and Lee (2018) could be implemented. This will likely result in a better prediction performance for the local volatility model.

Further research might also explore different option dataset across different periods. Comparing the results of this research to the paper of Almeida et al. (2022), we have seen that our forecasting results are worse compared to theirs. A possible explanation for this might be the more volatile period that this thesis considered. Future modelling work will have to be conducted in order to determine if this is the actual reason and if some models perform better compared to other depending on the type of time period.

Future work is also necessary to figure out if a different architecture for the feedforward neural

network would lead to better performance. Currently, most hyperparameters were set accordingly with Almeida et al. (2022). Although this is favourable in the comparison of some results, it might be that other values could work better. A couple of examples are the number of nodes in every layer, the number of hidden layers or the number of epochs for each run.

There is a wide variety and literature about other parametric or semi-parametric models that forecast the implied volatility surface. It could be very interesting to examine some of these approaches in future research. For example, according to Homescu (2014) the use of stochastic-local volatility models are a better approach for predicting the implied volatility surface. They are however, much more complicated to implement. More research in these type of models, or other approaches could be a fruitful area for future work. More information on what type of models performs the best could be seen as a great advantage.

Appendices

A Extra information on the basic topics

A.1 Options

Options are a type of financial contracts used mainly for hedging or speculation purposes. An option contract is made by the writer and bought by the buyer for a certain option premium which the option is worth at time t. There are two type of options, the call option which gives the right to buy the underlying product and the put option giving the right to sell the underlying. Call options can be in-the-money, which means that the strike price is below the underlying level for the option. They can also be out-of-the-money or at-the-money. When out-of-the-money, the option is worthless as the underlying asset is cheaper than the strike price. The opposite is true for put options.

Next to European option contracts, other type of contract exists. The most commonly know is the American option type. Here the buyer of the contract can buy or sell the underlying product over the entire lifespan and not only at maturity.

A.2 Black-Scholes model

The following assumptions are presumed under the Black-Scholes model

1. The underlying asset follows a geometric Brownian motion (GBM).
2. There are no transaction costs in buying / selling options.
3. The risk-free-rate is known and constant.
4. The asset doesn't pay any dividend over it's lifetime.
5. The underlying product follows a log-normal distribution.

A.3 Implied volatility

The implied volatility parameter is the value which can be put into the Black-Scholes equation to obtain the market price of an option. Therefore, it is the markets view of how the underlying asset will move in the future. Usually, when the volatility in the market increases the implied volatilities increase. One reason for this is that with more market volatility, options are more likely to end up in-the-money increasing their prices. The only way to achieve this is by increasing the implied

volatility parameter. Options are usually quoted in terms of implied volatility in order to get a better overview of how they relate to each other.

The Black-Scholes equation presumes a constant volatility for the asset, the result follows that the implied volatility should be constant for an option over different strikes. Yet, when considering the market option prices and deriving the implied volatilities a clear skew or smile shaped pattern is visible for the implied volatility parameter. This means that the implied volatility will be higher for ITM and OTM options, while being lower for ATM options. Considering figure 1, the implied volatility parameter is not constant, contradicting the Black-Scholes result.

B Proof: Local volatility parameter written in implied volatility

Proof. The same setup as in Oosterlee and Grzelak (2019) will be used for the proof. We will work from equation (5.5) towards equation (5.6). Remember that the value of a call option, at t_0, is given as

$$V_c(t_0, S(t_0)) = S(t_0)\Phi(d_1) - Ke^{-r\tau}\Phi(d_2),$$

where the values of d_1 and d_2 can be found in chapter 2 and $\tau = T - t_0$. Using the fact that $a = \log(\frac{K}{S_0}) - r(T - t_0)$ and $b = \sigma_{imp}^2\tau$ the above equation can be rewritten to

$$V_c(t_0, S(t_0)) = S(t_0)\left(\Phi(d_1) - e^a\Phi(d_2)\right) = y(a, b). \tag{B.1}$$

Here $d_1 = \frac{1}{2}\sqrt{b} - \frac{a}{\sqrt{b}}$ and $d_2 = d_1 - \sqrt{b}$. The goal is now to write the respective derivatives in equation 5.5 as functions of the implied volatility. Starting with the value of a call option with respect to the strike price K, using equation (B.1)

$$\frac{\partial V_c}{\partial K} = \frac{\partial y}{\partial a}\frac{1}{K} + \frac{\partial y}{\partial b}\frac{\partial b}{\partial K}. \tag{B.2}$$

The second derivative with respect to the strike is then given as

$$\frac{\partial^2 V_c}{\partial K^2} = \frac{1}{K^2}\left(\frac{\partial^2 y}{\partial a^2} - \frac{\partial y}{\partial a}\right) + \frac{2}{K}\frac{\partial b}{\partial K}\frac{\partial^2 c}{\partial b \partial a} + \frac{\partial^2 w}{\partial K^2}\frac{\partial y}{\partial b} + \left(\frac{\partial b}{\partial K}\right)^2\frac{\partial^2 y}{\partial b^2}. \tag{B.3}$$

Finally, the derivative with respect to maturity is given as

$$\frac{\partial V_c}{\partial T} = -r\frac{\partial y}{\partial a} + \frac{\partial y}{\partial b}\frac{\partial b}{\partial T}. \tag{B.4}$$

Equation (5.7) easily follows by taking the derivative of $b = \sigma_{imp}^2\tau$. What remains is to obtain equation (5.6). By filling in the different derivative expression into equation (5.5)

$$\sigma_{LV}^2(T, K) = \frac{\frac{\partial y}{\partial b}\frac{\partial b}{\partial T} + rK\frac{\partial y}{\partial b}\frac{\partial b}{\partial K}}{\frac{1}{2}\left(\frac{\partial^2 y}{\partial y^2} - \frac{\partial y}{\partial a}\right) + K\frac{\partial b}{\partial K}\frac{\partial^2 y}{\partial b \partial a} + \frac{1}{2}K^2\left[\frac{\partial^2 b}{\partial K^2}\frac{\partial y}{\partial b} + \left(\frac{\partial b}{\partial K}\right)^2\frac{\partial^2 y}{\partial b^2}\right]}. \tag{B.5}$$

We proceed using the fact that

$$\frac{\partial^2 y}{\partial b^2} = \frac{\partial y}{\partial b}\left(-\frac{1}{8} - \frac{1}{2b} + \frac{a^2}{2b^2}\right), \quad \frac{\partial^2 y}{\partial b \partial a} = \frac{\partial y}{\partial b}\left(\frac{1}{2} - \frac{a}{b}\right), \quad \frac{\partial^2 y}{\partial a^2} = \frac{\partial y}{\partial a} + 2\frac{\partial y}{\partial b}.$$

Filling these expression into equation B.5 concludes the proof.

□

C Parametric models extra results

C.1 (Ad-hoc) Black-Scholes model

Parameters fitted on options with strikes divisible by 10.

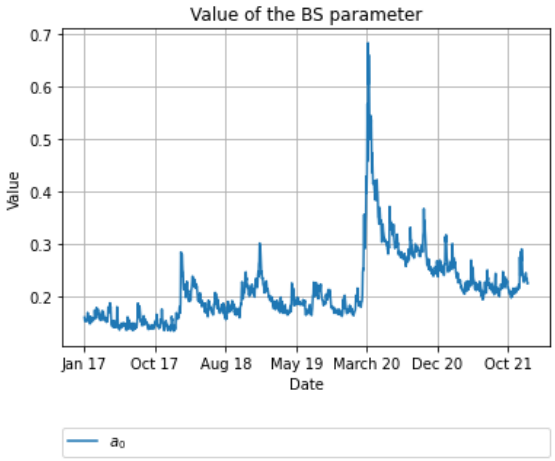

Figure 11: BS parameters.

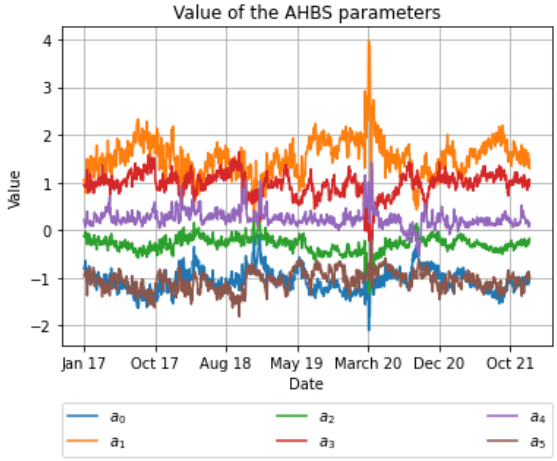

Figure 12: AHBS parameters.

C.2 SABR model

Parameters fitted for the entire option panel on day t, including the fitting of β.

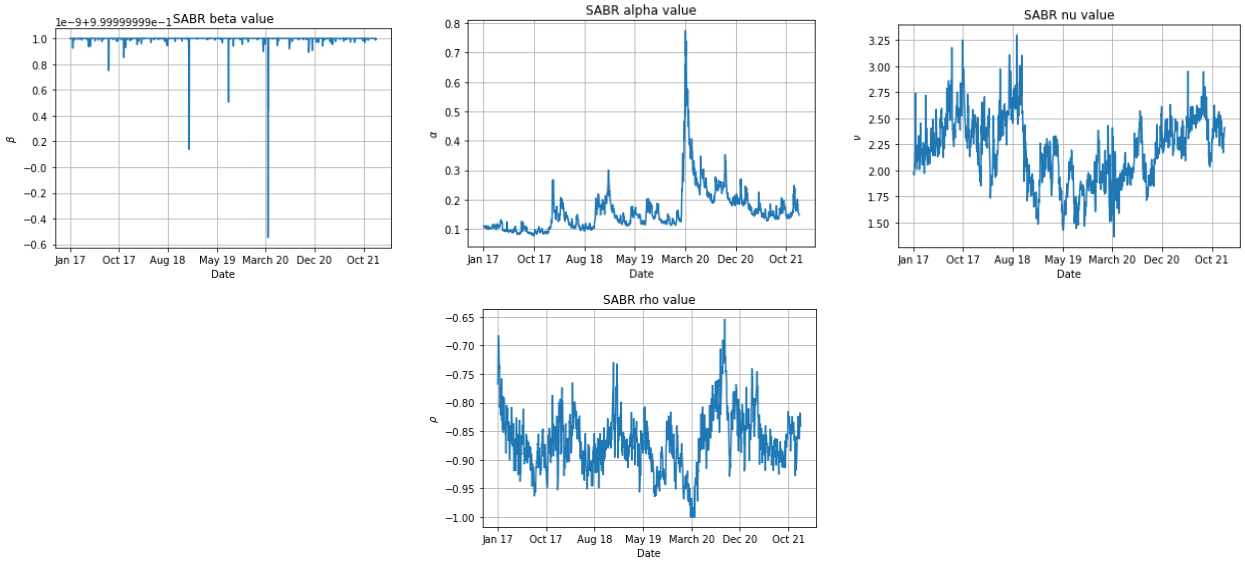

Figure 13: SABR parameters

Parameters fitted on options with strikes divisible by 10, excluding the fitting of β.

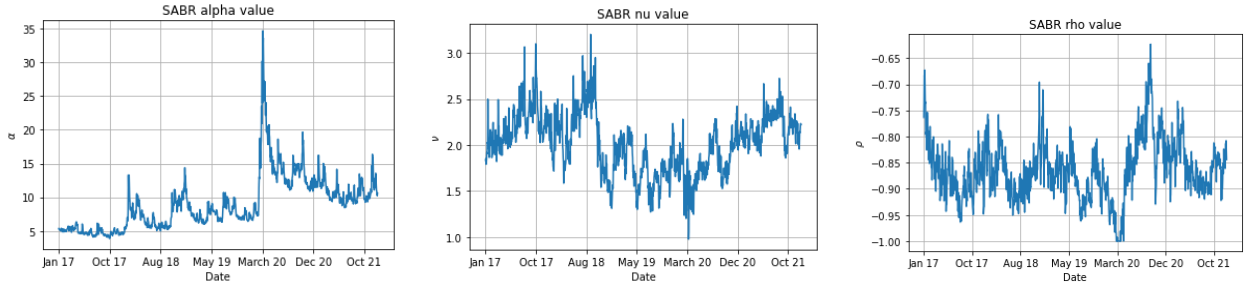

Figure 14: SABR parameters

Parameters fitted on options with strikes divisible by 10, including the fitting of β.

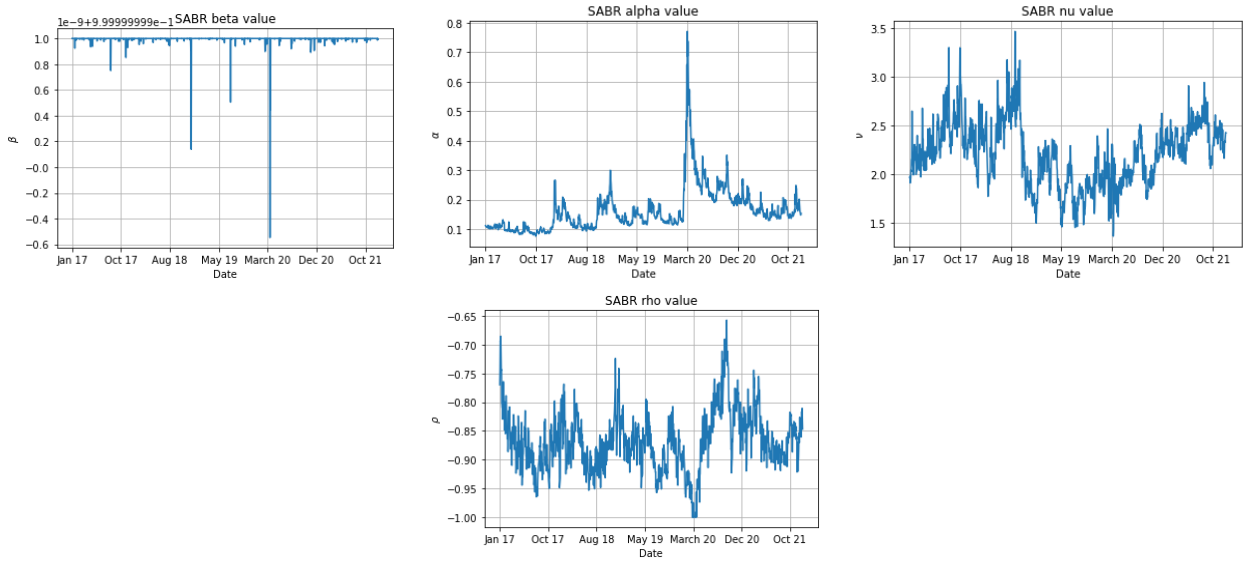

Figure 15: SABR parameters

C.3 Heston model

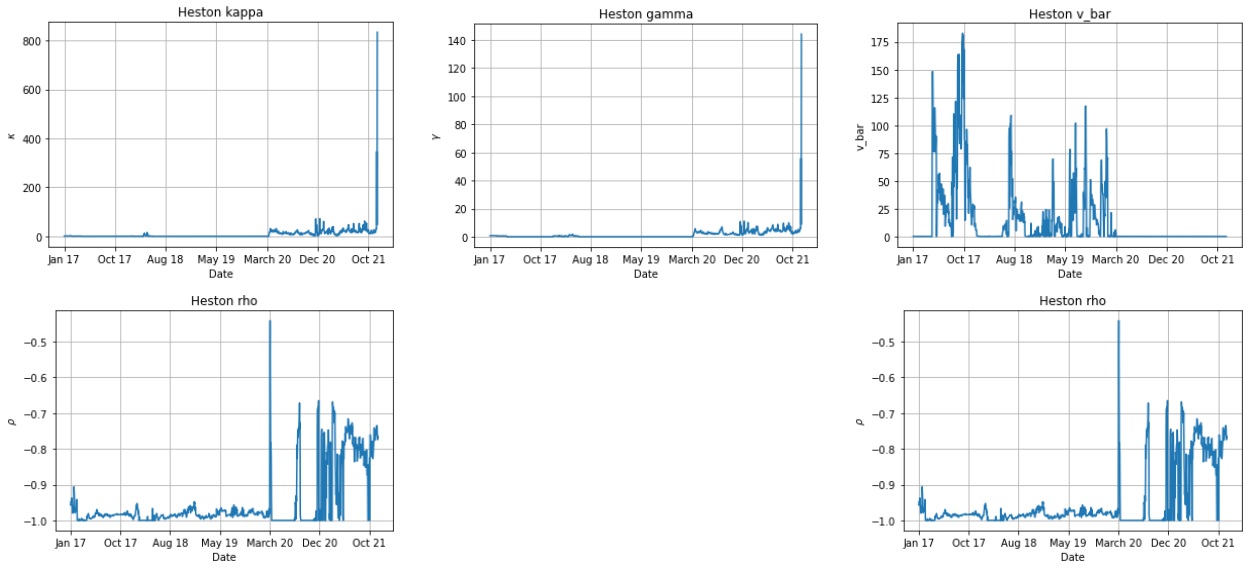

Figure 16: Heston parameters, unbounded calibration

D SABR model calibrated to the entire implied volatility surface

The SABR row represents the calibration to all of the option data on day t with a fixed β. The SABR (beta) row also took into consideration the fitting of β.

Table 9: IVRMSE (%) of the model predictions for 1-day ahead forecast.

1-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
SABR	3.04	2.17	2.11	2.10	**2.09**	**2.09**
SABR (beta)	3.02	2.26	2.67	**2.18**	2.19	2.19

Table 10: IVRMSE (%) of the model predictions for 5-day ahead forecast.

5-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
SABR	3.99	3.34	3.29	**3.27**	3.29	3.28
SABR (beta)	4.10	3.55	3.49	3.49	**3.48**	3.49

Table 11: IVRMSE (%) of the model predictions for 21-day ahead forecast.

21-day ahead prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
SABR	6.31	5.82	5.75	5.74	**5.73**	5.75
SABR (beta)	6.69	6.35	**6.24**	6.25	**6.24**	6.25

Table 12: IVRMSE (%) of the model predictions for intra-day forecast.

Intra-day prediction	No FNN	1-layer FNN	2-layer FNN	3-layer FNN	4-layer FNN	5-layer FNN
SABR	2.28	1.37	1.27	**1.25**	1.26	1.27
SABR (beta)	2.12	1.27	1.25	**1.16**	1.17	**1.16**

References

Almeida, C., Fan, J., Freire, G., and Tang, F. (2022). Can a machine correct option pricing models? *Available at SSRN 3835108*.

Black, F. and Scholes, M. (1973). The valuation of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bogatyreva, N., Grandez, R., Rodríguez Apolinar, S., and Soldevilla, A. (2019). SABr: a stochastic volatility model in practice.

Brecher, D. (2006). Pushing the limits of local volatility in option pricing. *Wilmott Magazine, September*, 6:15.

Carverhill, A. P. and Cheuk, T. H. (2003). Alternative neural network approach for option pricing and hedging. *Available at SSRN 480562*.

Chen, S. and Zhang, Z. (2019). Forecasting implied volatility smile surface via deep learning and attention mechanism. *Available at SSRN 3508585*.

Chen, S., Zhou, Z., and Li, S. (2016). An efficient estimate and forecast of the implied volatility surface: A nonlinear kalman filter approach. *Economic Modelling*, 58:655–664.

Christoffersen, P. and Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291–318.

Cox, J. (1975). Notes on option pricing I: constant elasticity of diffusions, draft of a paper, standford university.

Crépey, S. (2004). Delta-hedging vega risk? *Quantitative Finance*, 4(5):559–579.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6):2059–2106.

Dupire, B. (1994). A unified theory of volatility. Technical report, Working Paper, Paribas Capital Markets. Reprinted in Derivative Pricing: The Classic Collection, edited by Peter Carr, Risk Books.

Dupire, B. et al. (1994). Pricing with a smile. *Risk*, 7(1):18–20.

Fahrner, I. (2009). Differential geometry,(m, λ)-SABR and a formula by pierre-henry labordere. *Available at SSRN 1717676*.

Fan, J., Ma, C., and Zhong, Y. (2021). A selective overview of deep learning. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 36(2):264.

Fang, F. and Oosterlee, C. W. (2009). A novel pricing method for european options based on fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848.

Fengler, M. R., Härdle, W. K., and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics*, 5(2):189–218.

Floc'h, L., Kennedy, G. J., et al. (2014). Explicit SABR calibration through simple expansions. *Available at SSRN 2467231*.

Gauthier, P. and Rivaille, P.-Y. H. (2009). Fitting the smile, smart parameters for SABR and heston (October 30, 2009). *Available at SSRN: https://ssrn.com/abstract=1496982*.

Grzelak, L. A. and Oosterlee, C. W. (2014). From arbitrage to arbitrage-free implied volatilities (March 7, 2016). *Journal of Computational Finance 20(3), 1-19, 2016*.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hagan, P. S., Kumar, D., Lesniewski, A., and Woodward, D. (2014). Arbitrage-free SABR. *Wilmott*, 2014(69):60–75.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *The Best of Wilmott*, 1:249–296.

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

Homescu, C. (2011). Implied volatility surface: Construction methodologies and characteristics. *Available at SSRN 1882567*.

Homescu, C. (2014). Local stochastic volatility models: calibration and pricing. *Available at SSRN 2448098*.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The journal of Finance*, 49(3):851–889.

Jex, H. and Wang (1999). Pricing exotics under the smile. *Risk Magazine*, 12(11):72–75.

Kim, N. and Lee, Y. (2018). Estimation and prediction under local volatility jump–diffusion model. *Physica A: Statistical Mechanics and its Applications*, 491:729–740.

Kou, S. G. and Wang, H. (2004). Option pricing under a double exponential jump diffusion model. *Management science*, 50(9):1178–1192.

Lemaire, V., Montes, T., and Pagès, G. (2021). Stationary heston model: calibration and pricing of exotics using product recursive quantization. *Quantitative Finance*, pages 1–19.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.

Lipton (2002). The vol smile problem. *Risk Magazine*, 15(2):61–66.

Malliaris, M. and Salchenberger, L. (1996). Using neural networks to forecast the s&p 100 implied volatility. *Neurocomputing*, 10(2):183–195.

Masters, T. (1993). *Practical neural network recipes in C++*. New York: Academic Press Professional, Inc.

Medvedev, N. and Wang, Z. (2022). Multistep forecast of the implied volatility surface using deep learning. *Journal of Futures Markets*, 42(4):645–667.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

Nilsson, E. (2008). Calibration of the volatility surface. *PhD Thesis*.

Oosterlee, C. W. and Grzelak, L. A. (2019). *Mathematical Modeling and Computation in Finance: With Exercises and Python and Matlab Computer Codes*. World Scientific.

Ren, Y., Madan, D., and Qian, M. Q. (2007). Calibrating and pricing with embedded local volatility models. *Risk magazine limited*, 20(9):138.

Rubinstein, M. (1994). Implied binomial trees. *The journal of finance*, 49(3):771–818.

Schöbel, R. and Zhu, J. (1999). Stochastic volatility with an Ornstein–Uhlenbeck process: an extension. *Review of Finance*, 3(1):23–46.

Skiadopoulos, G. (2001). Volatility smile consistent option models: a survey. *International Journal of Theoretical and Applied Finance*, 4(03):403–437.

West, G. (2005). Calibration of the SABR model in illiquid markets. *Applied Mathematical Finance*, 12(4):371–385.

Yao, J., Li, Y., and Tan, C. L. (2000). Option price forecasting using neural networks. *Omega*, 28(4):455–466.

Zhang, W., Li, L., and Zhang, G. (2021). A two-step framework for arbitrage-free prediction of the implied volatility surface. *arXiv preprint arXiv:2106.07177*.