ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

THESIS QUANTITATIVE FINANCE [FEM61008]

# Nonparametric IVS Modelling of Equity Options

Thierry Hopman (604543)

| | |
|---|---|
| Supervisor: | dr. M. Grith |
| Second assessor: | dr. G. Freire |
| Date final version: | 8th August 2023 |

# Nonparametric IVS Modelling of Equity Options

Thierry Hopman

8th August 2023

**Abstract**

We provide a valuable extension of the original model as proposed by Almeida et al. (2022), this model finds a nonlinear mapping of the errors implied by any fitted parametric option pricing model. It does so by training a feedforward neural network on this error surface, using a cross-section of options. We consider a similar model that differs in using an ADAM gradient descent method. We apply this model to an identical data set and extend the research on this methodology by using this model-guided approach for equity options. We find promising results on the use of this method for prediction exercises on the cross-sections of equity options, as for a short enough horizon it always increases the performance and sometimes by a large extent. Contrary to the results of Almeida et al. (2022) for the S&P 500 option data on an identical sample period, our model does not outperform a neural network fitted directly to the implied volatility surface for equity options. Even though the performance increase for the use of equity options is not as large as found for the use of index options (nearly twice as large performance increase for index options), the nonparametric correction model still offers a valuable extension to any parametric option pricing model. Empirical results show that boosting a Black-Scholes model using equity options can offer up to 361, 138 and 45 % performance increase, for a 1-, 5- and 21-day ahead prediction exercise, respectively. When the equity options are ordered according to their prediction error for this model over the whole sample period from smallest to largest error, we find the following order Exxon Mobil, JP Morgan, Apple, Google, Amazon and Tesla. Remarkably, this order remains exactly the same over all prediction exercises, this might be indicative of the model's stability.

# Acknowledgement

I could not have undertaken this journey without the help of my supervisor dr. Maria Grith. First, the course Financial Derivatives given by Maria, inspired me to pursue a thesis in option pricing. Whenever I was stuck in the modelling process, her fast and thorough help steered me into the right direction. Furthermore, without her feedback the report would not have been what it is now. I am also thankful for dr. Gustavo Freire as a second assessor, his research inspired me to pursue this exciting topic. Lastly, I'd like to mention the employees of OptionMetrics, as their correspondence helped me guide through their database.

# Contents

# 1 Introduction

Traditionally, parametric option pricing models use the price of any underlying asset together with its volatility dynamics, to derive option prices for varying tenors (time-to-maturities) and strike prices. Pioneers in the field of option pricing are Fischer Black and Myron Scholes, who came up with the famous Black-Scholes model (Black & Scholes, 1973). Under the assumptions imposed by the Black-Scholes model, the implied volatility should remain constant across moneyness, time and tenors. In reality, it is observed that the implied volatility of S&P 500 options in the cross-section displays a smirk pattern (Rubinstein, 1994) and this smirk pattern increases for longer time-to-maturities. This inability of their model to incorporate the volatility is due to some strong assumptions that it imposes, e.g.: stock returns are log-normally distributed with known mean and variance, and the assumption of a fixed and constant short-term interest rate. It is even more remarkable considering that the market for S&P 500 options, is a market where the common conditions required for the Black-Scholes formula are likely to be approximated the most in practice. Especially in comparison to individual equity options, as pointed out by Rubinstein (1994). Evidently, the Black-Scholes model cannot be relied upon to perfectly calculate or predict option prices. As a result, there is extensive, ever-growing literature trying to relax the strong parametric assumptions imposed by the Black-Scholes model. Examples include the implied binomial tree model Rubinstein (1994), the stochastic-volatility (Heston) model Heston (1993), stochastic-interest-rate model Merton (1973). Unfortunately, none of these models have been successful in perfectly reproducing the implied volatility surface. The Black-Scholes formula for European call options is among the most popular instruments used to make many important economic decisions (Oh & Patton, 2021). A reason for this given by Heston (1993), is that most options are traded near-the-money, but it is mainly at options deep at-the-money and out-the-money where the mispricing occurs.

Not only academic researchers are interested in finding a model that accurately fits the implied volatility model. Market makers and option traders use the implied volatility surface to compare different options instead of using the option prices of different options (Carr & Wu, 2016). Therefore traders and market makers rely on a good estimation of the implied volatility such that they can minimise their risk. Furthermore, implied volatility can be used as a forward-looking metric as it represents the market expectations on future price movements of the underlying asset. Hence, the shape of the implied volatility surface provides insights into the risk-neutral distribution of the underlying asset returns (Almeida et al., 2022). Therefore, also institutional investors and policy-makers rely on a good estimation of the implied volatility surface. Given the fact that parametric models are still popular in practice, together with the fact that their performance is limited, motivates the use of a hybrid model which nonparametrically 'boosts' the performance of a parametric model. To add to this argument, Fang and Oosterlee (2009) point out several advantages of parametrically guided nonparametric models.

Almeida et al. (2022) propose such a method for S&P 500 options and find very promising results: nonparametrically corrected models always outperform the original parametric models and often by a large extent. In addition, this model always outperforms a neural network fitted directly to the implied volatility surface. They conduct a comparative analysis using different parametric methods and a feedforward neural network. They design their research such that their results can be interpreted as a lower bound on the performance of their model-guided approach.

A conclusion that can be drawn from the previously stated literature: the focus is on the performance of models considering options on S&P 500 data. This is logical as options on S&P 500 are among the most liquid options in the world, and it is valuable to know whether a model performs well if sufficient data is available. Together with the fact that the market for S&P 500 options is where the common conditions required for the Black-Scholes formula are likely to be approximated the most in practice. But a gap in the literature remains in the use of new models for equity options, such as the model of Almeida et al. (2022). This model is only tested on S&P 500 option data for which they find promising results. A valuable extension would be to test its performance on equity options, where fewer data is available and options could have higher implied volatility. We contribute to the literature, by using a similar method as described by Almeida et al. (2022) on equity options. Where we diverge with respect to their research in the use of a stochastic gradient descent from (Kingma & Ba, 2014) together with a ReLu activation function for the nonparametric correction. The motivation comes from the promising results of Gu et al. (2020). The main focus is to find whether the model performs equally or different among options originating from underlying stocks from different industries. Furthermore, we want to find the difference in results of this model between options on the S&P 500 and equity options. This leads to the following research question:

*In general, what is the performance increase of the model-guided approach with respect to the underlying parametric model for individual equity options?*

To address this research question we formulate the following sub-questions:

- *What is the relative performance increase of the model-guided approach of S&P 500 options with respect to the relative performance increase of the model-guided approach using individual equity options?*
- *Does this model-guided approach outperform a neural network fitted directly to the implied volatility surface for equity options? In other words: do the less misspecified corrected models outperform the corrected Black-Scholes model?*
- *Are there significant differences visible in relative performance increase/overall perform-*

*ance of the model-guided approach between different individual equity options (industries) and how can these differences be explained?*

We contribute to the literature by diverging from Almeida et al. (2022) in the use of both a cross-section of equity options together with a cross-section of S&P500 options. The equity options chosen in this research stem from the companies in the S&P 500 index with approximately the largest market capitalization per industry. Six industries are considered. The data on options range from January 4, 2016, to June 28, 2019. We use data available on day $t$ to fit the model and make predictions for day $t+1$, $t+5$ and $t+21$. We evaluate these predictions in terms of implied volatility root mean squared error (IVRMSE).

Our main findings are that this semi-parametric method is also promising for predicting the implied volatility of equity options. Even though the performance increase is not as convincing as when the model is used on index options, for a short enough prediction horizon it does always offer performance increase and sometimes by a large extent, similar to the findings of Almeida et al. (2022). Empirical evidence suggests that its performance increase does vary per equity options. Furthermore, a certain consistency in the order of the prediction performance of the model on equity options is found over the prediction exercises. This can be interpreted as evidence that the model can predict certain equity options better than others.

This paper is organized as follows: In section 2, we present relevant literature regarding this research. Section 3 expands on the origin and descriptive statistics of the data set. In Section 4, the use of parametric models in combination with the nonparametric correction model is discussed extensively. The experimental results are highlighted in Section 5. Finally, the main findings are summarised and discussed in Section 6.

# 2 Literature

In the derivation of their formula, Black and Scholes (1973) assume some 'ideal conditions' in the market for the stock and for the option. Some of these assumptions are: the distribution of stock prices at the end of any finite interval is log-normal, the variance rate of the return on the stock is constant, the short-term interest rate is known and is constant through time, there are no transaction costs and the stock pays no dividends. Obviously, some of these assumptions are too strict for the real world, e.g. Schwert (1989) shows that financial asset returns display time-varying volatility. Also, it is a well-known fact that historically stock prices display excess kurtosis and skewness. Rubinstein (1994) found that before the stock market crash of October 1987 options that are out-the-money (OTM) or deep in-the-money (ITM), have much higher implied volatilities than at-the-money (ATM) options. More specifically, he found that prior to the crash, S&P 500 option-implied volatilities formed a "smile" pattern. After the market crash, he observed a "smirk" pattern, more concretely he observed that for an increasing strike price, the implied volatility of the calls (puts) decreases monotonically with

an increasing rate of decrease for calls (puts) that hence go deeper OTM (ITM). Dumas et al. (1998), have found that this observed "smirk" pattern is a typical implied volatility pattern ever since the crash of 1987. They also found evidence for a general increase in implied volatility as the tenor increases.

As a result, an extensive variety of option pricing models have been proposed trying to relax some of the above-mentioned assumptions. Bakshi et al. (1997) point out that the list of models trying to relax the restrictive assumptions of the Black-Scholes model is exhaustive. It contains stochastic-interest-rate option models e.g. Merton (1973); Markovian (implied binomial tree) models e.g. Rubinstein (1994); stochastic-volatility models e.g. Heston (1993); stochastic-volatility and stochastic-interest-rates models e.g. Amin and Ng (1993).

Even though the literature on using machine learning to model the implied volatility is relatively limited in comparison to the literature on parametric models, there are some examples. One of the first papers to propose the use of machine learning methods such as the neural network as an estimation model for option prices is Hutchinson et al. (1994). Where they consider a variety of machine learning methods to model a mapping of the moneyness and tenor to the observed S&P 500 option prices. They find that the machine learning methods all significantly outperform the Black-Scholes model.

Another example of a direct nonparametric model for option pricing is given by Amilon (2003). In this paper, a neural network is used to model a mapping of strike price, tenor, risk-free rate and some additional inputs onto the observed option prices. The author considers options on the Swedish OMX index as data. The author finds that the model outperforms the two benchmark models (two Black-Scholes models with historical and implicit volatility estimates) both in terms of pricing and hedging exercises, however, the results are often insignificant.

Audrino and Colangelo (2010) propose a somewhat similar semi-parametric model as considered in Almeida et al. (2022), they consider different parametric models and use a regression tree as the nonparametric correction algorithm. They find that their model improves the performance of the starting (parametric) model in forecasting future implied volatilities (up to 60 days). Furthermore, they find that using a regression tree as an initial model and then boosting the performance by using a regression tree to explain the residuals, is the best-performing model.

A different use of neural networks in option-pricing is considered by De Spiegeleer et al. (2018), Horvath et al. (2021) and Liu et al. (2019), which focuses on the reduction of computational time for normally time-expensive procedures such as the estimation of stochastic-volatility models.

Almeida et al. (2022) consider the use of four parametric option pricing models: Black-Scholes model, Ad-hoc Black-Scholes Model, Heston model and the Carr and Wu model. First, they fit these four parametric models to the observed implied volatilities to obtain model parameter estimates. These estimates are then used to calculate the model-predicted implied volatility. The observed implied volatility data is calculated from a cross-section of options

with tenors of 20 to 240 months and moneyness between 0.8 and 1.60. The next step is to evaluate the model predicted implied volatility with the observed implied volatility, they use the difference of these two for the whole range of options to obtain a so-called pricing error surface. This pricing error surface is then approximated by a feedforward neural network. This nonparametric approximation is combined with the parametric model to end up with a 'boosted' model. They test their models by splitting the option data into a training and test for each day $t$.
Additionally, they test their models by training them on all data available on day $t$ and test their performance by predicting the IVS for 1-, 5- and 21-days ahead keeping the parameters fixed. They evaluate the models based on the out-of-sample implied volatility root mean square error (IVRMSE), which is calculated from the prediction errors obtained from the test sets. They compare the 'base' parametric models with the neural network corrected parametric models and find that across all exercises every boosted model always outperforms the pure parametric model itself and often by a large extent. Another interesting result is that the corrected Black-Scholes model is always outperformed by correcting a less-misspecified model, the difference is the largest in the same-day and 21-day ahead prediction exercises. They argue that correcting the Black Scholes model is equivalent to employing a stand-alone neural network, as the IVS predicted by the Black-Scholes formula is flat. These results show that the model-guided approach yields better results than direct nonparametric estimation in the case of a feedforward neural network in this setup. They also find that regardless of the misspecification of the original model, every correct model brings down the error to the same magnitude for all types of options regarding their moneyness (ATM/OTM/ITM).

Bakshi et al. (2003) highlighted some significant differences between the market of equity options and index options among other important findings. They motivate their research by arguing that the skewness and kurtosis of stock returns have been a prominent topic of research. Furthermore, it is a well-known fact that higher-order risk-neutral moments influence the implied volatility surface (IVS). But no literature has tried to establish a relationship between the physical return density and the risk-neutral distribution. More specifically, they derive restrictions on the physical density and pricing kernel process such that the risk-neutral distributions are shifted to the right. In their research options on S&P 100 index and its 30 largest individual equity components are considered. The authors find that the slope of individual equity volatility smiles is much less negative than the index volatility smile. Additionally, they find that the more left-skewed the return distribution, the steeper its volatility smile and the higher the risk-neutral kurtosis, the more flattened the volatility smile. A significant difference between the index and individual equities is that: the returns on individual stocks are mildly left-skewed, whereas the index return distributions are heavily left skewed. Finally, they relate the skewness of the risk-neutral index to higher-moments of the physical distributions. More specifically, they relate the differences in magnitudes of the risk-neutral and physical skew to risk-aversion and long-tailed physical distribution.

Duan and Wei (2009) continued this study by considering an identical data set. In their research empirical results show a clear relationship between the implied volatility of individual equity options and the amount of systematic risk in the underlying equities. They find that higher systemic risk leads to a steeper implied volatility curve and higher level of implied volatility.

Finally, Bernales and Guidolin (2014) consider daily data on 150 sets of equity and S&P 500 index option prices. They find that regardless of the moneyness nor the tenor of the options, the equity options have a higher mean and standard deviation implied volatility. By proposing a two-step method that first fits the IVS using a deterministic function for both the equity and index options and then using the historical estimates of these functions in a VARX-type model. They find evidence for cross-sectional and dynamic relationships between the IVS of equity and index options. Moreover, when they forecast the IVS of equity options using this VARX model, they find it outperforms a VAR model using only equity IVS dynamics, an option GARCH(1,1) model and a random walk using coefficients of the deterministic IVS function.

# 3 Data

We consider the European-style S&P 500 options as a "benchmark dataset", the reason to see this as a benchmark is because it is among the most liquid options in the world. Moreover, options on this stock have been used in the seminal papers considered in section 2. We use OptionMetrics to obtain data on this option, where the sample period is from January 4, 2016, to June 28, 2019. This option data exists of crucial information such as the bid and ask quotes, trading volume, implied volatility and the trade and expiration date. But also the spot price, dividend yield of the security and risk-free rates are obtained from OptionMetrics.

We follow the common practice of taking the midpoint of the end-of-day bid and ask quotes as the option prices and considering only out-of-the-money (OTM) options. The latter design choice is because these options are more liquid and reliable than their in-the-money (ITM) counterparts (Andersen et al., 2015). For every PM (AM) settled option, the time to maturity is equal to the number of days between the trade date and the expiration date (minus one day). OptionMetrics provides a risk-free rate for every date and specific tenors on this date, we interpolate this data to obtain risk-free rates for corresponding time-to-maturities from the option data. In addition, they provide a continuous dividend yield for each day and tenor, which we use to discount the spot price. We filter the data by dropping observations with zero trading volume, an option price $<$ \$ 0.125 and violating the standard no-arbitrage conditions. Furthermore, on any date $t$ for option $i$, We consider only options with a moneyness $\left(m_{i,t} \equiv \frac{S_t}{K_{i,t}}\right)$ between 0.80 and 1.60 and with time to maturities between 20 and 240 calendar days. We follow Almeida et al. (2022) by grouping the options according to their moneyness: deep OTM call if $m_{i,t} \in [0.80.0.90)$, OTM call if $m_{it} \in [0.90, 0.97)$, ATM option if $m_{it} \in [0.97, 1.03)$, OTM put if $m_{it} \in [1.03, 1.10)$ and deep OTM put if $m_{it} \in [1.10, 1.60]$.

We define short-term options, as options with a time-to-maturity between 20 and 40 days and long-term options as options with a time-to-maturity between 40 and 240 days. Hence options are categorized short-term if $\tau_i \in [20, 40]$, and long-term if $\tau_i \in (20, 240]$.

Besides considering options on S&P 500, we consider equity options for various industries. The equity options chosen in this research stem from the companies in the S&P500 index with approximately the largest market capitalization per industry. This ensures liquidity of the stocks and allows us to find differences in model performance between options originating from different industries. The companies selected as underlying stocks for the equity options are chosen by their liquidity (or data available from OptionMetrics), and their market capitalisation, and every underlying stock should represent a different industry. One of these equity options is Apple, associated with the computer and consumer electronics industry (tech-industry). The second tech-industry company is Google (Alphabet Class A), which is associated with the computer software industry. Options on Amazon are associated with the online retail industry. The automotive industry is represented by options on Tesla, and the oil and gas industry is associated with Exxon Mobil options. The last industry considered is financial services, represented by JP Morgan Chase. Google, Tesla and Amazon do not pay dividends hence the stock price does not have to be adjusted for the dividends. Whereas Apple, Exxon Mobil and JP Morgan do pay discrete dividends. We assume the dividend yield of a security on any given date to stay constant over the remaining term of the option. Together with additional assumptions of constant average security return and constant earnings growth, this constant dividend yield assumption is consistent with dividend-based equity pricing models such as the Gorden growth model (Gordon, 1959). This dividend yield on day $t$ is defined as the most recently announced dividend payment on day $t - i \; \forall \; i \geq 0$, divided by the most recent closing price for the underlying stock.

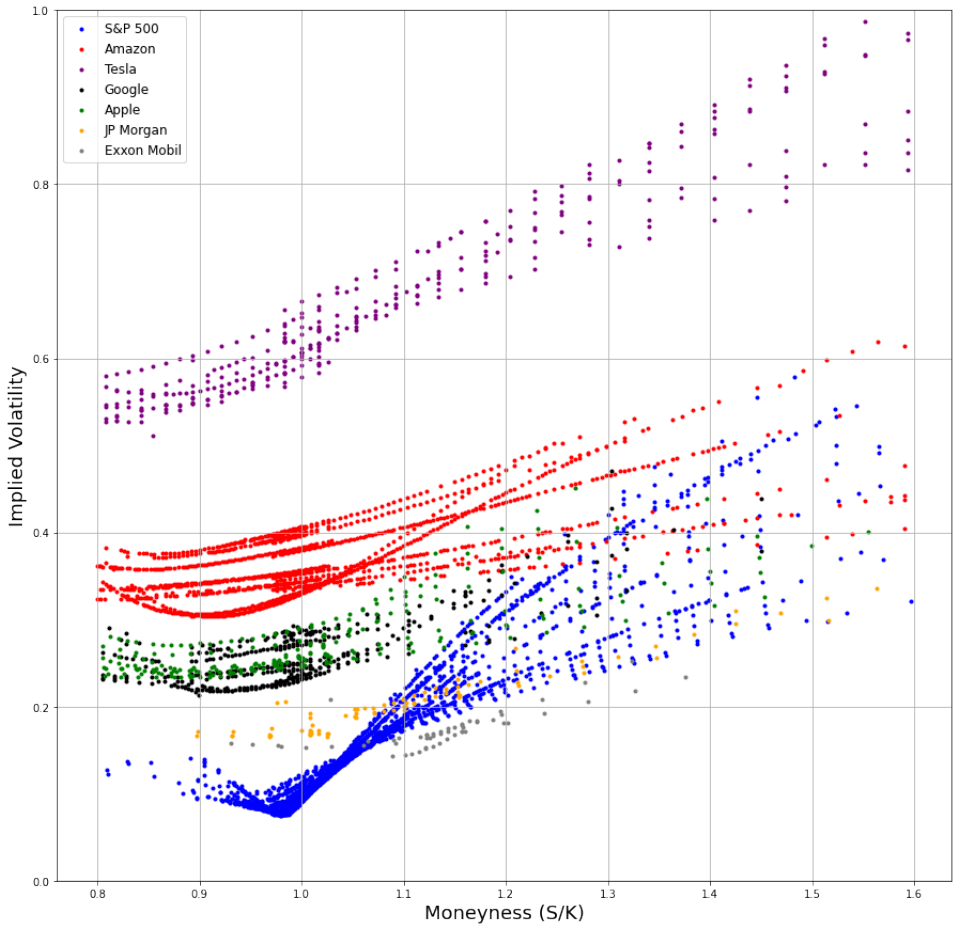


Figure 1: Observed implied volatility surface of S&P 500, Amazon, Tesla, Google, Apple, JP Morgan and Exxon Mobil as observed on September 17, 2018. These implied volatilities are observed for all options with moneyness inside the considered range of [0.8, 1.6] and time-to-maturities between 20 and 240 days.

Figure 1 shows that, especially for ATM options, equity options are relatively more dispersed around their mean than the index options. It is also visible that on average the index option has a lower implied volatility than (almost) any equity-based option. However, it is visible that for DOTM puts the implied volatilities of the index options somewhat converge to all (except Tesla) equity options. This figure also clearly displays the larger average observed volatility for Tesla relative to the other options, which are somewhat more clustered together. The typical implied volatility "smirk" is present for nearly all options, however for the index option the "smirk" is the most prominently visible. It is not observable for the options on JP Morgan and Exxon Mobil, but that is likely to be due to the absence of data.

Some of the previously made statements can be confirmed when inspecting Table 1. Namely: the standard deviation of ATM S&P 500 options is relatively smaller than nearly all ATM equity options. It can also be seen that the implied volatility of DOTM puts (except for Tesla) are relatively close to one another, whereas the implied volatility D(OTM) calls and ATM S&P 500 options are relatively smaller than nearly all other D(OTM) calls and ATM options. The higher average implied volatility as well as its higher standard deviation for Tesla is well visible. The total standard deviation of the observed implied volatility is relatively smaller

for JP Morgan, Exxon Mobil and Google.

| | Number | Mean IV | Std. dev. IV | Number | Mean IV | Std. dev. IV |
|---|---|---|---|---|---|---|
| Moneyness | Panel A: S&P500 | | | Panel B: Amazon | | |
| [0.80, 0.90) | 21,643 | 14.26 | 3.42 | 66,138 | 30.23 | 6.41 |
| [0.90, 0.97) | 195,643 | 11.10 | 3.05 | 81,073 | 28.37 | 7.21 |
| [0.97, 1.03) | 543,828 | 12.02 | 3.92 | 170,830 | 28.48 | 7.29 |
| [1.03, 1.10) | 303,290 | 17.30 | 3.81 | 79,464 | 29.65 | 6.97 |
| [1.10, 1.60] | 418,110 | 27.91 | 7.59 | 137,804 | 36.31 | 8.48 |
| Total | 1,482,514 | 17.49 | 8.56 | 535,309 | 30.87 | 8.14 |
| | Panel C: Tesla | | | Panel D: Google | | |
| [0.80, 0.90) | 45,072 | 44.43 | 8.45 | 25,424 | 23.71 | 4.59 |
| [0.90, 0.97) | 35,610 | 44.12 | 9.54 | 49,916 | 22.21 | 4.96 |
| [0.97, 1.03) | 56,003 | 46.77 | 10.45 | 104,600 | 22.49 | 5.25 |
| [1.03, 1.10) | 27,655 | 49.89 | 10.79 | 45,501 | 24.25 | 4.99 |
| [1.10, 1.60] | 84.160 | 60.11 | 14.55 | 52,793 | 30.29 | 6.86 |
| Total | 248,500 | 50.83 | 13.49 | 278,324 | 24.32 | 6.21 |
| | Panel E: Apple | | | Panel F: JP Morgan | | |
| [0.80, 0.90) | 37,046 | 23.48 | 4.33 | 2,738 | 19.93 | 3.11 |
| [0.90, 0.97) | 32,809 | 23.12 | 4.93 | 8,929 | 20.15 | 3.79 |
| [0.97, 1.03) | 46,412 | 25.11 | 4.97 | 32,078 | 19.47 | 3.91 |
| [1.03, 1.10) | 20,478 | 27.88 | 4.92 | 22,719 | 20.81 | 4.12 |
| [1.10, 1.60] | 32,216 | 32.52 | 6.54 | 30,810 | 26.58 | 5.72 |
| Total | 168,961 | 26.11 | 9.20 | 97,233 | 22.11 | 5.52 |
| | Panel G: Exxon Mobil | | | | | |
| [0.80, 0.90) | 754 | 20.00 | 4.65 | | | |
| [0.90, 0.97) | 4,928 | 18.08 | 4.76 | | | |
| [0.97, 1.03) | 22,586 | 17.35 | 4.51 | | | |
| [1.03, 1.10) | 19,281 | 17.74 | 4.70 | | | |
| [1.10, 1.60] | 30,990 | 21.83 | 5.76 | | | |
| Total | 78,539 | 19.29 | 5.50 | | | |

Table 1: This table shows the global statistics for options on six different underlying stocks. The sample of the data ranges from January 4, 2016, to June 28, 2019, for time-to-maturities between 20 and 240 days. The number of observations, mean implied volatility and standard deviation of the implied volatility (both in percent) for options on every underlying stock are given by the columns: Number, Mean IV and Std. dev. IV, respectively.

# 4 Methodology

First, the parametric models used to evaluate the model implied volatility surface are introduced. Then the nonparametric correction method to allow for performance boosting is intro-

duced, also possible extensions are discussed. Finally, we elaborate on the model evaluation and prediction exercises.

## 4.1 Parametric option pricing models

### 4.1.1 Black-Scholes Model

The derivation done by Black and Scholes (1973) imposes some strong parametric assumptions, this has as a result that it has an analytical solution of the price of a European option. This price of a European option is a function of the spot price $S_t$, the strike price $K$, the time to maturity $\tau = (T - t)$ and the constant instantaneous volatility of the underlying asset $\sigma$. Its closed-form solution for a European call is given by:

$$\begin{aligned} C_{BS}(S_t, K, \tau, \sigma) &= \Phi(d_1)S_t - \Phi(d_2)K\exp(-r\tau), \\ d_1 &= \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau\right], \\ d_2 &= d_1 - \sigma\sqrt{\tau}, \end{aligned} \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. The short-term risk-free rate $r$ is assumed to be constant. The analytical solution given by eq. (1) is derived by assuming that the spot price $S_t$ follows a geometric Brownian motion, denoted by the following partial differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{2}$$

where $\mu$ is the expected return of the asset, the constant, instantaneous volatility is given by $\sigma$, and $W_t$ is a standard Wiener process. The strong parametric assumptions imposed by the model lead to the mispricing of options. This is observable when evaluating the implied volatility, which is the volatility implied by the market prices of the options. The Black-Scholes predicts constant implied volatility across all tenors and strike prices. The implied volatility surface can be calculated by evaluating $\sigma_i = C_{BS}^{-1}(C_i, S_t, K_i, \tau_i, r)$, where $C_{Bs}^{-1}(\cdot)$ is the inverse of the Black-Scholes formula, across tenors and strike prices on a given day $t$. There is no analytical solution of the inversion of the Black-Scholes formula however, typically numerical estimation methods are used such as the Newton-Rhapson, Bisection or Brent method. The Black-Scholes predicted volatility is a constant value $c$ on day $t$ for a cross-section of options, i.e. the volatility surface is flat for all time-to-maturities and strike prices. It is common practice to use the mean implied volatility of the most recent observations. We set the constant volatility for every day $t$ equal to the mean implied volatility of the options considered between 4, January 2016 and 28, June 2019, such that:

$$\sigma_{BS} = \frac{1}{N}\sum_{i=0}^{N}\sigma_i \quad \forall\ i = 1, ..., N \tag{3}$$

where $N$ is equal to the total amount of options over the sample period and $\sigma_i$ is the implied volatility for option $i$. Though this is incorrect for a prediction exercise, as we are using future information. We consider this method as using a constant volatility through time, might help us uncover the performance of the correction model when its parametric estimate is relatively good and when it is far off. More specifically, it could help us highlight the time-dynamics of the nonparametric correction method. To add to this, we know from previous literature that the Black-Scholes model is always far off.

### 4.1.2 Ad-hoc Black Scholes Model

Besides performing empirical tests to assess the performance of the implied tree approach, Dumas et al. (1998) propose an Ad-hoc Black Scholes as a benchmark for performance comparison. This Ad-hoc model is inspired by market makers who smooth the implied volatility relation across strike prices and tenors, to value options using this smoothed relation. They propose a two-step procedure where they first fit the Black-Scholes implied volatilities using 6 parameters accounting for a constant, the (squared) strike price, the (squared) date of expiration and their cross-term. The term Ad-hoc arises from the fact that this model is internally inconsistent, as it is applying the Black-Scholes formula for varying volatility. Whereas the Black-Scholes formula assumes constant volatility in the cross-section. We follow the slightly different Ad-hoc model as proposed by Almeida et al. (2022), where on a given day $t$ with a cross-section $i = 1, ...n$ options, the implied volatility $\sigma_{i,t}$ is regressed using:

$$\begin{aligned}\sigma_{AHBS}(\mathbf{a_t}, m_{i,t}, \tau_{i,t}) \equiv \sigma_{i,t} = a_{0,t} + a_{1,t}m_{i,t} + a_{2,t}m_{i,t}^2 + a_{3,t}\tau_{i,t} \\ + a_{4,t}\tau_{i,t}^2 + a_{5,t}m_{i,t}\tau_{i,t} + \epsilon_{i,t},\end{aligned} \tag{4}$$

where moneyness and time to maturity of option $i$ are defined as $m_{i,t} = S_t/K_{i,t}$ and $\tau_{i,t}$, respectively. The corresponding exercise and spot price are denoted as $K_{i,t}$ and $S_t$, respectively. The parameters $a_{j,t},\ \forall\ j = 1, ..., 5;\ t \geq 0$, are estimated using ordinary least squares for a range of options with varying tenors and moneyness and their given implied volatility for any day $t$. Thus obtaining a vector of the estimated parameter $\mathbf{a}_t$. This is equivalent to minimizing:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\sigma_{i,t} - \sigma_{AHBS}(\mathbf{a}_t, m_{i,t}, \tau_{i,t})\right]^2, \tag{5}$$

where $\sigma_{i,t}$ is the observed implied volatility on day $t$ for option $i$ in the cross-section $i = 1, ..., n$. These estimated parameters can then be used to calculate the model implied volatility surface by evaluating $\hat{\sigma}_{AHBS} = (\hat{\mathbf{a}}_t, m_i, \tau_i)$, in the cross-section of options available on day $t$. More details on the theoretical foundation of the Ad-Hoc Black Scholes model can be found in section A.0.1.

### 4.1.3 Heston Model

Heston (1993) generalizes the assumption of a constant volatility model, by deriving a stochastic volatility model. He derives an analytical solution in an analogous way to the Black-Scholes formula but with some different assumptions. The most crucial difference is the assumption that the price of the underlying asset has stochastic volatility $v_t$ under the risk-neutral measure:

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dZ_1(t), \tag{6}$$

where $S_t$ is the spot price at time $t$, the drift rate is denoted by $\mu$ and $dZ_1(t)$ is a standard Wiener process. The stochastic differential equation for the volatility is given by:

$$dv_t = \kappa(\theta - v_t) + \sigma\sqrt{v_t} dZ_2(t), \tag{7}$$

where the speed of the mean reversion is denoted as $\kappa$, the unconditional mean is $\theta$, the volatility of the volatility is denoted by $\sigma$ and $dZ_2(t)$ is a standard Wiener process correlated with $dZ_1(t)$ via $\rho$. In his paper he provides a quasi-closed form solution, we follow Almeida et al. (2022) and numerically compute the option prices using the Fourier-cosine series expansion method of Fang and Oosterlee (2009). For this, we use a python script, which we modified to handle large datasets, obtained from Oosterlee and Grzelak (2019). To predict the model implied volatility surface, the model parameters $\boldsymbol{\xi}_t = (\nu_t, \theta, \kappa, \sigma, \rho)$ for each day $t$ are approximated by minimizing:

$$\sum_{i=1}^{n} [\sigma_{i,t} - \sigma_H(\boldsymbol{\xi}_t, S_t, K_{i,t}, \tau_{i,t}, \mu_t)] \tag{8}$$

using non-linear least squares. Where $\sigma_{i,t}$ is the observed implied volatility, $K_{i,t}$ and $\tau_{i,t}$ denotes the strike prices and times to maturity on day $t$ for a cross-section of $i = 1, ..., n_t$ options, respectively. The fitted values of the Heston model $\sigma_H(\hat{\boldsymbol{\xi}}_t, S_t, K_{i,t}, \tau_{i,t}, \mu_t)$ are calculated by first numerically computing the option prices under the Heston model as described before, after which these are translated to implied volatilities to minimize the loss function. The estimated parameters can then be used to calculate a model implied volatility surface by evaluating $\hat{\sigma}_H(\hat{\boldsymbol{\xi}}_t, S_t, K_{i,t}, \tau_{i,t}, \mu_t)$ over a cross-section of options available on day $t$. More details on the theoretical foundations of the Heston model can be found in section A.0.2.

### 4.1.4 Carr and Wu Model

It is a well-known fact that institutional investors manage their positions in options by looking at implied volatility as computed from the Black-Scholes formula, as opposed to looking at the dynamics of the underlying assets and options prices. That is the reason why Carr and Wu (2016) came up with a new framework for option pricing, that tries to incorporate this

practicality. Instead of a full specification of the instantaneous variance of the underlying asset, this framework models the near-term dynamics of the Black-Scholes implied volatility surface and derives no-arbitrage constraints directly on its shape. This model casts the shape of the implied volatility surface as a solution of a quadratic equation under the assumption of the implied volatility dynamics. As a foundation of their framework they specify the following specification for the dynamics of the underlying stock price $S_t$ and option implied volatility $\sigma_t(K,\tau)$ under the risk-neutral measure:

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{v_t} dW_t, \\ \frac{d\sigma_t(K,\tau)}{\sigma_t(K,\tau)} &= e^{-\eta_t \tau}(m_t dt + w_t dZ_t) \end{aligned} \tag{9}$$

where $v_t$ is the instantaneous variance of the underlying asset log returns. The parameter $dW_t$ is a standard Wiener process and correlated to the standard Wiener process $dZ_t$ via $\rho_t$, which is a stochastic process on the interval [-1,1]. The stochastic processes $m_t, w_t$ and $\eta_t$ do not depend on strike price $K$ & maturity date $T$. The stochastic drift is defined as $m_t$, the volatility of the implied volatility $w_t$ is constrained to be a strictly positive process. The exponential dampening $e^{-\eta_t(T-t)}$ is used to accommodate the empirical observation that implied volatilities for options with longer tenors tend to have less volatility. The end result of their derivation is obtained by using the previously described dynamics in an algebraic equation that imposes dynamic no-arbitrage restrictions, resulting in a quadratic equation. This quadratic equation links the squared implied volatility $\sigma_t^2(k,\tau)$ with the relative strike price $k = \log(K/S_t)$:

$$\begin{aligned} &\frac{1}{4} e^{-2\eta_t \tau^2} w_t^2 \tau \sigma_t^4 + (1 - 2e^{-\eta_t \tau} m_t \tau - e^{-\eta_t \tau} w_t \rho_t \sqrt{v_t} \tau) \sigma_t^2 \\ &\quad - (v_t + 2e^{-\eta_t \tau} w_t \rho_t \sqrt{v_t} k + e^{-2\eta_t \tau} w_t^2 k^2) = 0, \end{aligned} \tag{10}$$

this equation has a neat result that the no-arbitrage constraint depends on the current levels of the dynamic processes $(m_t, w_t, \eta_t, v_t, \rho_t)$, but not on the exact dynamics. Confirming that this model relies on the near-term dynamics of the Black-Scholes implied volatility surface. Again, we follow Almeida et al. (2022) who estimate the following parameters $\hat{\boldsymbol{\theta}}_t = (v_t, m_t, w_t, \eta_t, \rho_t)$ by minimizing:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t = \arg\min_{\hat{\boldsymbol{\theta}}_t} \sum_{i=1}^{n} \Big[ &\frac{1}{4} e^{-2\eta_t \tau_{i,t}^2} w_t^2 \tau_{i,t} \sigma_{i,t}^4 + (1 - 2e^{-\eta_t \tau_{i,t}} m_t \tau_{i,t} - e^{-\eta_t \tau_{i,t}} w_t \rho_t \sqrt{v_t} \tau_{i,t}) \sigma_{i,t}^2 \\ &- (v_t + 2e^{-\eta_t \tau_{i,t}} w_t \rho_t \sqrt{v_t} k_{i,t} + e^{-2\eta_t \tau_{i,t}} w_t^2 k_{i,t}^2) \Big]^2, \end{aligned} \tag{11}$$

given a cross-section of options on day $t$, relative strike price $k_{i,t}$, with implied volatility $\sigma_{i,t}$ and tenor $\tau_{i,t}$. Equation (11) is minimized by using nonlinear least squares. The obtained estimated parameters can then be used to solve eq. (10) for $\sigma_{CW}^2(\hat{\boldsymbol{\theta}}_t, k, \tau)$ for a cross-section

of options available on day $t$ to evaluate the model implied volatility surface accordingly. More details on this methodology can be found in section A.0.3.

## 4.2 Model Boosting

### 4.2.1 Nonparametric Correction

Following Almeida et al. (2022), a nonparametric method is used to correct the model implied volatility surface $\sigma_p(m, \tau)$, which is a volatility surface defined for varying moneyness (m) and time to maturity ($\tau$) for any model $p$. The model implied volatility surface comes from any of the previously described parametric option pricing models which are then cast in terms of its own implied mapping, with the model parameters being suppressed. The simplest example is the Black-Scholes implied volatility surface which predicts a flat volatility surface, i.e. $\sigma_{BS}(m, \tau) = c$. As none of the parametric option pricing is able to perfectly reproduce the observed implied volatility surface, the misspecification of these models can be expressed in quantitative terms. To quantify this misspecification a pricing error surface is defined as $\epsilon_p(m, \tau) = \sigma(m, \tau) - \sigma_p(m, \tau)$. This pricing error surface is then minimized using a nonparametric method to boost the accuracy of the corresponding parametric option pricing model. Correcting the Black-Scholes implied volatility surface is equivalent to using solely a neural network to estimate the pricing error surface, as the BS implied volatility surface is constant. A two-step approach is used to boost the accuracy of any parametric option pricing model. First, the observed implied volatilities $\sigma(m_{i,t}, \tau_{i,t})$ are fitted to any parametric model $p$ for a given cross-section $i = 1, ..., n$ options on day $t$. This leads to the estimation of $\hat{\sigma}_p(m_{i,t}, \tau_{i,t})$ and the corresponding model-implied pricing errors $\hat{\epsilon}_p(m_{i,t}, \tau_{i,t})$. The second step is to minimize the following function:

$$\min_f \frac{1}{n} \sum_{i=1}^{n} [\hat{\epsilon}_p(m_{i,t}, \tau_{i,t}) - f(m_{i,t}, \tau_{i,t})]^2 \tag{12}$$

where $\hat{f}(m, \tau)$ is nonparametrically estimated to approximate the pricing error surface. Subsequently, the boosted model implied volatility surface is then obtained by summing $\hat{\sigma}_p(m_{i,t}, \tau_{i,t})$ and $\hat{f}(m_{i,t}, \tau_{i,t})$ for every $1 \leq i \leq n$.

### 4.2.2 Feedforward Neural Networks

We follow Almeida et al. (2022) in using the feedforward neural network as a non-parametric estimation method to estimate the pricing error surface. This neural network consists of an input layer of predictors, denoted as $\mathbf{x}_{i,t} = (m_{i,t}, \tau_{i,t})' \in \mathbb{R}^2$ containing the moneyness and tenor for option $i$ on day $t$. This predictor input is then transformed nonlinearly via one or more hidden layers, where each hidden layer is made up of neurons. Each of these neurons delivers an

output which is a nonlinear transformation of a linear combination of the input variables. The nonlinear transformation is determined by the so-called activation function, which typically is a simple nonlinear function such as the sigmoïd or Rectified Linear Unit (ReLu) function (Hasti et al., 2009). Finally, the neurons are aggregated to an output layer that contains predicted (trained) or fitted (untrained) values, which in our case is the estimated pricing error. Hence the model performs a nonlinear mapping $f : \mathbb{R}^2 \to \mathbb{R}$, where the following recursive formulas are initialized with $\mathbf{z}^{(0)} = \mathbf{x}_{i,t}$:

$$\begin{aligned} \mathbf{z}^{(l)} &= h\left(A^{(l-1)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l-1)}\right), \text{ for } l = 1, ..., L, \\ f(\mathbf{x}_{i,t}) &= A^{(L)}\mathbf{z}^{(L)} + b^{(L)}, \end{aligned} \tag{13}$$

and the number of hidden layers is denoted by $L$, which is also called the depth of the neural network. The nonlinear activation function is denoted as $h(\cdot)$. For hidden layer $l$, the number of neurons is equal to the dimension of $\mathbf{z}^l$ and $A^{(l)}$ denotes the weight matrix, containing the weights connecting the neurons between every hidden layer. The intercepts are stacked in $\mathbf{b}^{(l)}$. We elaborate on the details of a feedforward neural network in section A.1.

There are certain aspects that have to be taken into consideration when designing a neural network, such as the number of hidden layers, the number of neurons and their corresponding activation function in every hidden layer. As pointed out by Gu et al. (2020) there are infinitely many design choices and therefore they proposed a framework with a set of five network architectures, where the number of neurons in each layer decreases following the geometric pyramid rule proposed by Masters (1993). The reason they use this architecture is that it allows them to generate a lower bound of their results. This comes down to the following network architectures: A neural network with one hidden layer that consists of 32 neurons, denoted as NN1. A neural network with two hidden layers with 32 and 16 neurons respectively, denoted as NN2. A neural network with three hidden layers with 32, 16 and 8 neurons respectively, denoted as NN3. This design choice allows the authors to infer the trade-offs of network depth. In this paper, we follow Almeida et al. (2022) to use this structure, but we stop at the third neural network as Gu et al. (2020) found in their research that this was the optimal structure, and in early-stage testing we also found similar results when considering the same structure as Almeida et al. (2022). We diverge from Almeida et al. (2022) experimental model by using a ReLu function as an activation function and using it for all neurons. This ReLu function is defined as:

$$ReLu(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ otherwise,} \end{cases} \tag{14}$$

and this activation function allows for faster derivative evaluation. It also promotes sparsity in the number of active neurons with respect to more traditional activation functions such as the sigmoïd function Gu et al. (2020). The training of a feedforward neural network is done

via a gradient descent algorithm. This algorithm is initialized by using random values for the weight parameters which are all set close to zero, then these weights are recursively updated in the direction of the steepest descent. Backpropagation is used to compute the gradient, where in a two-pass algorithm the parameters are updated by calculating gradients through the network in a backward fashion. We differ from Almeida et al. (2022) using the stochastic gradient descent algorithm ADAM (Kingma & Ba, 2014), as Gu et al. (2020) found promising results in their research where they used this algorithm. Almeida et al. (2022) consider the scaled conjugate gradient (SCG) algorithm by Møller (1993). A key difference between these two gradient descent algorithms is that ADAM is a first-order stochastic optimization method whereas the SCG is a non-stochastic second-order optimization method. Hence, the SCG algorithm requires both the gradient and the Hessian of the cost function, as opposed to only gradient information for the ADAM algorithm. It is a well-known fact that second-order derivatives can be computationally expensive, to add to this: a stochastic gradient method is known for its ability to speed up the optimization process. We use a batch size of 32 and evaluate over 640 epochs. To read more about general guidelines to follow when training a neural network please refer to section A.1.2.

## 4.3 Prediction Exercises in the Option Cross-Section

To evaluate the predictive power of the models a prediction exercise will be used for every cross-section of any equity option. The first prediction exercise trains the models on day $t$ using the full option data set available and tests their performance in predicting the implied volatility surface $h$ days ahead. The implied volatility surface is directly predicted for one-day ($h = 1$), one-week ($h = 5$) and one-month ahead ($h = 21$). To assess the quality of the implied volatility predictions, we will use the average implied volatility root mean squared error (IVRMSE) overall predictions. The IVRMSE is determined for all time horizons, this prediction exercise gives us insight into the stability of the models over time. In addition to this, the mean implied volatility per industry gives us a general insight into the volatility of the traded options and what effect this might have on the performance of the model. To verify whether the difference in forecasts between the parametric and boosted parametric model is significant, we consider using the Diebold-Mariano test statistic as proposed by Diebold and Mariano (1995). More specifically, we test the null hypothesis of equal forecast accuracy between the two models, under the assumption that these models are non-nested. The test statistic is defined as:

$$DM = \frac{\bar{d}_{A,B}}{\hat{\sigma}_{A,B}} \sim N(0,1) \tag{15}$$

with

$$\hat{\sigma}_{A,B} = \frac{1}{P-1} \sum_{t=T}^{T+P-1} \left(d_{t+1} - \bar{d}\right)^2$$
$$d_{t+1} = e_{A,t+1}^2 - e_{B,t+1}^2 \tag{16}$$

where $e_{A,t+1}$ indicates the forecast error of model A at time $t+1$ and the sample mean of the loss is given by $\bar{d}$. We set model A equal to the parametric model and model B equal to the boosted parametric model. In this case, a significant and positive test statistic tells us that the boosted parametric model performs significantly better than the parametric model. We test the null hypothesis under a two-sided test and set the significance level equal to 1%. Furthermore, the mean implied volatility for options categorized in options with different moneyness and tenors visualizes the presence of a volatility smile in the cross-section.
The prediction exercises are conducted for a total of four parametric models (BS, AHBS, Heston and CW) and twelve parametrically guided nonparametric models (the three neural network structures in combination with the four parametric models).

# 5 Empirical Results

First, the 1-, 5- and 21-day ahead predictions for the four different (nonparametrically corrected) parametric models using the options on S&P 500 can be found in Table 2. Note that we found nearly the same IVRMSE values for the parametric predictions of the S&P 500 index as Almeida et al. (2022). The IVRMSE values for the BS model are the only results that are relatively different, likely due to a different definition of what the constant volatility should be. The IVRMSEs of the semi-parametric models are a bit higher than the values found in Almeida et al. (2022) for the 1-day ahead prediction exercise. For the 5- and 21-day ahead prediction exercise the results of the best semi-parametric models are very similar again. It is also noteworthy that as opposed to the findings of Almeida et al. (2022): correcting the BS model is not always outperformed by correcting the other parametric models. For two of the three prediction exercises it is the best performing and in the other exercise it is 0.01 % off from being the best performing model. Another significant difference is the under-performance of the nonparametric correction of the Heston model. Only for the 21-day ahead prediction exercise the results are similar, for the other exercises the IVRMSEs we found are a bit higher.

Table 2: S&P 500 Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 8.56 | 1.23 | **1.20** | 1.21 | 8.56 | 2.23 | **2.22** | **2.22** | 8.52 | 3.29 | 3.29 | 3.29 |
| AHBS | 1.93 | 1.23 | 1.21 | **1.20** | 2.74 | 2.24 | 2.23 | **2.22** | 3.77 | 3.30 | 3.30 | 3.29 |
| Heston | 1.50 | 1.27 | 1.24 | 1.24 | 2.35 | 2.28 | 2.27 | 2.27 | 3.34 | 3.36 | 3.34 | 3.34 |
| CW | 2.06 | 1.23 | 1.21 | 1.23 | 2.78 | **2.22** | **2.22** | 2.23 | 3.75 | **3.28** | **3.28** | 3.29 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, using options on S&P 500. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Second, the results for prediction exercises of the equity options are given in Tables 3, 4, 5, 6, 7 and 8. When comparing Table 2 with the previously mentioned Tables, it comes to notice that the variability between the prediction errors of the nonparametric correction of different parametric models is relatively larger for equity options than for the index option. This increased variability is visible for all three prediction exercises. Especially when comparing the nonparametric correction of the different parametric methods of Exxon Mobil, Tesla and Google this variability of the prediction errors is visible.

Moreover, it becomes visible that the models are not able to decrease the errors to similar values as obtained in Table 2 for all three prediction exercises. This is remarkable as for all equity options, the prediction error obtained by the BS model is lower than the prediction error obtained for the BS model used on index options. To add to this: the prediction error of the AHBS model used on options on JP Morgan and Exxon Mobil are lower in comparison to the prediction error of the AHBS model but then used for index options. The prediction errors obtained using a nonparametric correction (using this AHBS model) are nevertheless always larger for the latter equity options. Stated differently, the performance increase for these cases significantly deteriorates, especially in comparison to the BS nonparametric correction on these options. Nevertheless, the errors obtained using the best performing semi-parametric method for the predictions exercises on Apple, JP Morgan and especially Exxon Mobil come close to the errors obtained for the best-performing model used on index options.

Table 3: Amazon Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 8.13 | 2.04 | **2.00** | 2.01 | 8.10 | 4.03 | **4.02** | **4.02** | 7.98 | 7.76 | 7.74 | **7.73** |
| AHBS | 2.78 | 2.07 | 2.01 | 2.03 | 4.26 | 4.07 | 4.04 | 4.05 | 7.74 | 7.80 | 7.77 | 7.76 |
| Heston | 6.14 | 2.59 | 2.46 | 2.42 | 6.55 | 4.47 | 4.42 | 4.40 | 7.80 | 8.06 | 8.02 | 8.00 |
| CW | 4.60 | 2.05 | **2.00** | **2.00** | 5.60 | 4.04 | **4.02** | **4.02** | 7.97 | 7.79 | 7.75 | 7.74 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, for options on Amazon. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Table 4: Tesla Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 13.50 | 2.96 | 2.88 | 2.93 | 13.51 | **5.63** | 5.64 | 5.67 | 13.55 | 9.30 | 9.31 | 9.32 |
| AHBS | 3.53 | 3.01 | 2.92 | 2.94 | 6.00 | 5.74 | 5.73 | 5.72 | 9.47 | 9.37 | 9.37 | 9.37 |
| Heston | 20.45 | 3.57 | 3.31 | 3.31 | 20.54 | 6.03 | 5.91 | 5.91 | 20.80 | 9.69 | 9.60 | 9.60 |
| CW | 4.53 | 2.95 | 2.88 | **2.87** | 6.39 | **5.63** | 5.64 | 5.64 | 9.57 | 9.30 | **9.29** | 9.30 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, for options on Tesla. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Table 5: Google Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 6.20 | 1.72 | **1.67** | 1.69 | 6.17 | 3.14 | 3.13 | 3.13 | 5.99 | **5.61** | 5.57 | 5.56 |
| AHBS | 2.34 | 1.82 | 1.73 | 1.72 | 3.51 | 3.29 | 3.25 | 3.25 | 5.80 | 5.73 | 5.67 | 5.67 |
| Heston | 3.13 | 2.22 | 2.15 | 2.15 | 3.88 | 3.56 | 3.51 | 3.54 | 5.72 | 5.97 | 5.93 | 5.92 |
| CW | 3.53 | 1.75 | 1.70 | 1.70 | 4.21 | 3.14 | **3.12** | 3.13 | 5.92 | 5.62 | 5.58 | 5.57 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, for options on Google. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Table 6: Apple Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 6.20 | 1.62 | **1.55** | 1.56 | 6.16 | **2.92** | **2.92** | 2.94 | 6.04 | 4.76 | **4.74** | **4.74** |
| AHBS | 2.09 | 1.65 | 1.56 | 1.56 | 3.17 | 2.95 | 2.93 | 2.94 | 4.90 | 4.77 | **4.74** | 4.76 |
| Heston | 2.21 | 1.72 | 1.63 | 1.60 | 3.47 | 3.32 | 3.30 | 3.23 | 4.88 | 5.21 | 5.11 | 5.09 |
| CW | 3.16 | 1.62 | **1.55** | **1.55** | 3.82 | 2.94 | 2.93 | 2.93 | 5.13 | 4.76 | 4.75 | 4.75 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, for options on Apple. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Table 7: JP Morgan Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | |
| BS | 5.52 | 1.59 | **1.54** | 1.55 | 5.51 | 2.72 | **2.68** | **2.68** | 5.43 | 4.05 | **4.00** | **4.00** |
| AHBS | 1.74 | 1.62 | 1.56 | 1.55 | 2.82 | 2.74 | 2.70 | 2.69 | 4.14 | 4.07 | 4.02 | 4.03 |
| Heston | 1.96 | 1.75 | 1.69 | 1.68 | 3.07 | 2.95 | 2.89 | 2.87 | 4.46 | 4.35 | 4.27 | 4.27 |
| CW | 2.12 | 1.58 | 1.55 | **1.54** | 3.05 | 2.71 | **2.68** | **2.68** | 4.27 | 4.04 | **4.00** | **4.00** |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019. for options on JP Morgan. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

Table 8: Exxon Mobil Prediction Exercises Results

| | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 | No NN | NN1 | NN2 | NN3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Panel A: 1-day ahead | | | | Panel B: 5-day ahead | | | | Panel C: 21-day ahead | | | |
| BS | 5.50 | 1.59 | **1.51** | 1.52 | 5.46 | 2.44 | **2.39** | **2.39** | 4.96 | 3.70 | 3.65 | **3.63** |
| AHBS | 1.72 | 1.63 | 1.57 | 1.57 | 2.52 | 2.48 | 2.46 | 2.45 | 3.78 | 3.75 | 3.69 | 3.68 |
| Heston | 3.48 | 2.33 | 2.08 | 2.02 | 4.04 | 3.07 | 2.84 | 2.79 | 5.08 | 4.25 | 4.05 | 3.96 |
| CW | 3.05 | 1.84 | 1.79 | 1.77 | 3.51 | 2.45 | 2.42 | 2.40 | 4.53 | 3.79 | 3.76 | 3.76 |

The IVRMSE (in %) for a 1-, 5- and 21-day ahead prediction exercise (as indicated by the panels) over a sample range from January 4, 2016, to June 28, 2019, for options on Exxon Mobil. The rows represent the parametric option pricing models, and the columns correspond to the nonparametric correction model (feedforward neural networks with a geometric pyramid structure). The best-performing model is indicated by a bold number.

To detect significant differences in relative performance increase/overall performance of the model-guided approach between different individual equity options and the index option we use the following method. We compare the increase in percentage between the IVRMSE of the parametric model and the IVRMSE of its nonparametric correction with a neural network of three hidden layers. This correction model is chosen as it is roughly on average the best-performing correction model. The relative performance increase for the BS, AHBS, Heston and Carr and Wu model for all types of options can be found in Tables 9, 10, 11 and 12, respectively. The best performing correction models among the equity options are given by the bold numbers. The Diebold Mariano tests show that all models reject the null hypothesis of equal forecast accuracy at a 1 % significance for all underlying securities and prediction exercises. Tables 9 and 10 show that the model-guided approach has the most significant performance increase for the index options for all three prediction exercises. For tables 11 and 12, this is interesting enough not the case. When this nonparametric correction is used for index options, it has the least performance increase if the original model is the Heston model, this is not necessarily the case for equity options. The least performance increase for equity options (as well as variability in performance increase) is roughly when the original model is the AHBS model.

| | BS+NN3 (h=1) | BS+NN3 (h=5) | BS+NN3 (h=21) |
|---|---|---|---|
| S&P 500 | $607^*$ | $286^*$ | $159^*$ |
| Amazon | $307^*$ | $101^*$ | $3^*$ |
| Tesla | $\mathbf{361}^*$ | $\mathbf{138}^*$ | $\mathbf{45}^*$ |
| Google | $267^*$ | $97^*$ | $8^*$ |
| Apple | $297^*$ | $110^*$ | $27^*$ |
| JP Morgan | $256^*$ | $106^*$ | $36^*$ |
| Exxon Mobil | $262^*$ | $128^*$ | $37^*$ |

Table 9: Performance increase (%) of the nonparametrically corrected BS model with respect to the original BS model. The nonparametrically corrected BS model is a feedforward neural network with a NN3 structure. The performance increase is displayed for the three different prediction exercises, being h= 1, 5 and 21 day(s) ahead. The best-performing correction models among the equity options are given by the bold numbers. A number with a * indicates a p-value $< 0.01$, corresponding to significance level 1 %, for the Diebold-Mariano test.

| | AHBS+NN3 (h=1) | AHBS+NN3 (h=5) | AHBS+NN3 (h=21) |
|---|---|---|---|
| S&P 500 | $61^*$ | $23^*$ | $15^*$ |
| Amazon | $\mathbf{37}^*$ | $5^*$ | $0^*$ |
| Tesla | $20^*$ | $5^*$ | $1^*$ |
| Google | $36^*$ | $\mathbf{8}^*$ | $2^*$ |
| Apple | $34^*$ | $\mathbf{8}^*$ | $\mathbf{3}^*$ |
| JP Morgan | $12^*$ | $5^*$ | $\mathbf{3}^*$ |
| Exxon Mobil | $10^*$ | $3^*$ | $\mathbf{3}^*$ |

Table 10: Performance increase (%) of the nonparametrically corrected Ad-Hoc BS model with respect to the original Ad-Hoc BS model. The nonparametrically corrected Ad-Hoc BS model is a feedforward neural network with a NN3 structure. The performance increase is displayed for the three different prediction exercises, being h= 1, 5 and 21 day(s) ahead. The best-performing correction models among the equity options are given by the bold numbers. A number with a * indicates a p-value $<$ 0.01, corresponding to significance level 1 %, for the Diebold-Mariano test.

| | Heston+NN3 (h=1) | Heston+NN3 (h=5) | Heston+NN3 (h=21) |
|---|---|---|---|
| S&P 500 | $21^*$ | $4^*$ | $0^*$ |
| Amazon | $154^*$ | $49^*$ | $-3^*$ |
| Tesla | $\mathbf{518}^*$ | $\mathbf{248}^*$ | $\mathbf{117}^*$ |
| Google | $46^*$ | $10^*$ | $-4^*$ |
| Apple | $38^*$ | $7^*$ | $-4^*$ |
| JP Morgan | $17^*$ | $7^*$ | $4^*$ |
| Exxon Mobil | $72^*$ | $45^*$ | $28^*$ |

Table 11: Performance increase (%) of the nonparametrically corrected Heston model with respect to the original Heston model. The nonparametrically corrected Heston model is a feedforward neural network with a NN3 structure. The performance increase is displayed for the three different prediction exercises, being h= 1, 5 and 21 day(s) ahead. The best-performing correction models among the equity options are given by the bold numbers. A number with a * indicates a p-value $<$ 0.01, corresponding to significance level 1 %, for the Diebold-Mariano test.

| | CarrWu+NN3 (h=1) | CarrWu+NN3 (h=5) | CarrWu+NN3 (h=21) |
|---|---|---|---|
| S&P 500 | $67^*$ | $25^*$ | $14^*$ |
| Amazon | $\mathbf{130}^*$ | $39^*$ | $3^*$ |
| Tesla | $58^*$ | $13^*$ | $3^*$ |
| Google | $108^*$ | $35^*$ | $6^*$ |
| Apple | $104^*$ | $30^*$ | $8^*$ |
| JP Morgan | $37^*$ | $14^*$ | $7^*$ |
| Exxon Mobil | $72^*$ | $\mathbf{46}^*$ | $\mathbf{20}^*$ |

Table 12: Performance increase (%) of the nonparametrically corrected Carr and Wu model with respect to the original Carr and Wu model. The nonparametrically corrected Carr and Wu model is a feedforward neural network with a NN3 structure. The performance increase is displayed for the three different prediction exercises, being h= 1, 5 and 21 day(s) ahead. The best-performing correction models among the equity options are given by the bold numbers. A number with a * indicates a p-value $<$ 0.01, corresponding to significance level 1 %, for the Diebold-Mariano test.

Third, we compare the differences between the prediction errors of the different models used on equity options as showcased by Tables 3, 4, 5, 6, 7 and 8. For all prediction exercises,

the boosted BS model has the lowest error in 5 out of the 6 cases. Put differently for any given prediction exercise, the boosted BS model achieves the lowest error for 5 out of the 6 underlying equity options. Another interesting result is that the 1-day ahead prediction results for Apple, JP Morgan and Exxon Mobil are fairly similar. For the other prediction exercises the results of these three equity options seem to slightly diverge, it is however clear that for all the prediction exercises, Amazon, Google and Tesla always have significantly higher IVRMSEs than the latter three. If we rank the prediction error of the best-performing model per equity option we get from largest to smallest error: Tesla, Amazon, Google, Apple, JP Morgan and finally Exxon Mobil. This order is the same for all three prediction exercises, which seems to be evidence for a certain consistency in the model as it can price certain equity options better than others for different prediction exercises.

Furthermore, it is remarkable that even though options on Tesla have a large mean and standard deviation of implied volatility, the nonparametric model can significantly improve the model, as highlighted in Table 9. The second best option in terms of performance increase is Amazon and after that Apple both with an increase of around 300 percent. The other options all have a performance increase of approximately 2,5 times compared to the original model.

For the five days ahead prediction exercise, Tesla still has the highest performance increase with 138 percent, Exxon Mobil is the second best with 128 percent and the rest of the options roughly have a performance increase of 100 percent.

For the 21 days ahead prediction exercise, the performance increase varies more across the different equity option types. Tesla still has the largest performance increase for the 21-days ahead prediction exercise, which is approximately a fourth of the performance increase of the S&P 500 corrected model. Both JP Morgan, Exxon Mobil (and Apple to a certain degree) have similar increases for all three prediction exercises, which fits with the previously observed results that the model for these equity options leads to similar results. Google and Amazon have relatively low performance increases when it comes to a 21-day ahead forecast. A general observation is that: the performance increase appears to be the largest when the error of the parametric model is the largest. However, even though the variability of the results for a given prediction exercise and option is relatively large, it still brings down the error to a similar magnitude regardless of the original parametric model.

Additionally, it is noticeable that besides the BS corrected model also the Carr and Wu corrected model has a fairly good performance for the one-day ahead prediction exercise as it is beating the corrected AHBS and Heston model for all panels except for the Exxon Mobil panel. A similar observation can be made for the five and 21-day ahead prediction exercises. However, it is important to state the following nuance: the AHBS model reaches relatively similar IVRMSE values. The only deviating model in terms of results is the nonparametrically corrected Heston model, which systematically underperforms among all prediction exercises relative to the other corrected models. This is remarkable as its parametric model often comes out as the second or third best for all exercises. However for some reason to be further re-

searched the feedforward neural network finds it difficult to estimate the pricing error surface as given by the parametric Heston model.

A visualisation of the nonparametric correction of the BS model for a 1-,5- and 21-day ahead prediction of a cross-section of options on Apple on 17 September 2018 can be seen in Figure 2. From this graph, it appears that as the tenor gets smaller and the more the option moves OTM, the less accurate the neural network correction. The neural network correction does appear to capture the curvature of the implied volatility surface fairly well, but seems to underestimate the magnitude and this underestimation increases as the prediction-horizon increases.

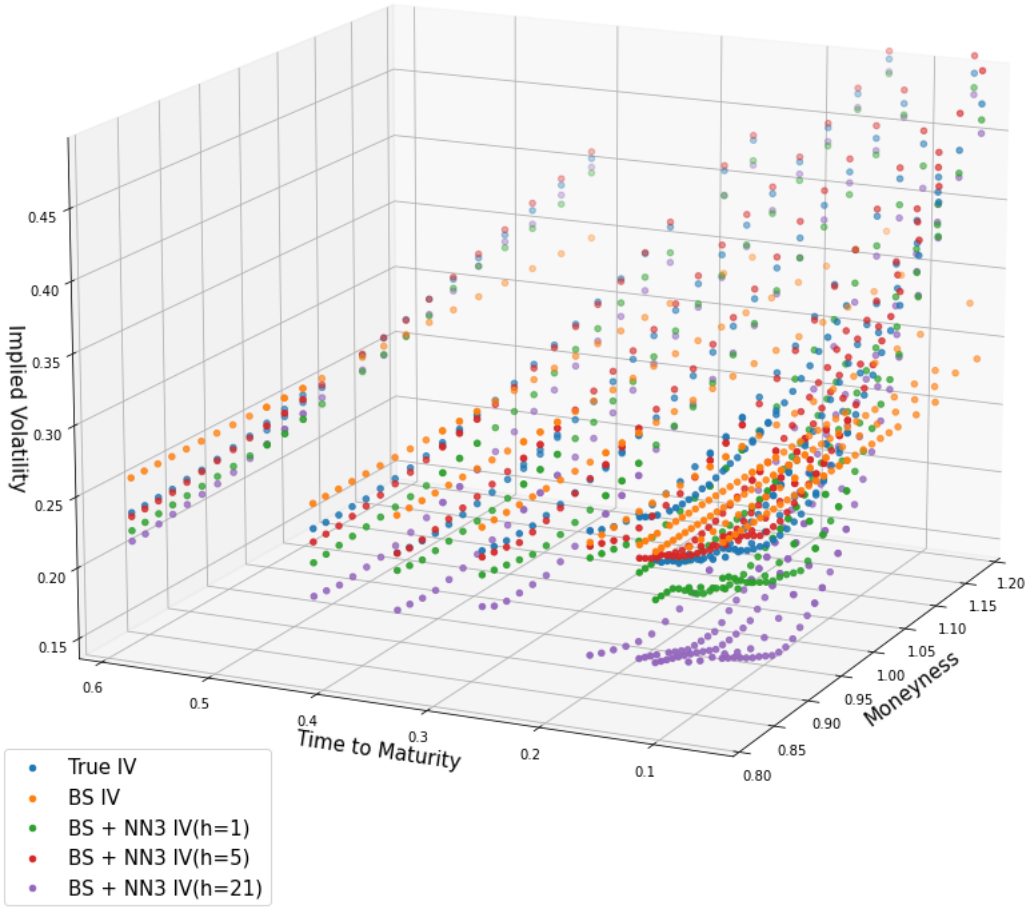


Figure 2: An implied volatility surface plot for a cross-section of options on Apple on 17 September 2018. This model shows the observed implied volatility, the parametric Black-Scholes 1-, 5- and 21-day ahead prediction of the implied volatility surface and a nonparametric correction of these models.

To further examine the latter suspicions, we will examine Table 13 which shows the IVRMSEs for the equity options categorized by their tenor and moneyness properties as described in

section 3. To come back to the suspicions stated in the previous paragraph, it is visible that for the 1-day ahead prediction exercise, the performance increase is nearly always larger for long-term equity options than for short-term options. To add to this, the evidence becomes even stronger if we look at Table 14 and 15. There also appears to be evidence for the ability of the model to capture the curvature of the implied volatility surface as the nonparametric correction roughly brings down the error to a similar magnitude for all options and prediction exercises. However, there does not appear to be convincing evidence for the fact that the neural network correction becomes less accurate as the option moves more OTM for the 1-day ahead prediction exercise. If we look at the performance increase across the different moneyness categories for multiple equity options (e.g. JP Morgan) and models there does not appear to be convincing evidence. Moreover, looking at Table 14 and 15 the evidence against this hypothesis is increasing. If we focus on the best-performing models per column for all the equity options, it comes to notice that options that are ATM always have the highest error in comparison to the other moneyness categories. Whereas the best predictions of DOTMC options nearly always have the smallest prediction error. It is visible that on average the best prediction errors increase as the options move from DOTMC towards ATM and decrease again as the options move from ATM towards DOTMP. It is immediately clear that for all equity options, the nonparametrically corrected short-term options have much higher IVRMSEs than the long-term options. Remarkably, the parametric Carr and Wu model is the only model that can bring the error to the same magnitude for both the short- and long-term options. However, if this parametric model is corrected with a neural network, the difference is again noticeable with on average a two times smaller IVRMSE for the long-term options. For Amazon, Tesla, Google and Apple the nonparametrically corrected Carr and Wu model, has the smallest error for nearly all categories. Whereas for JP Morgan the corrected BS model has the smallest error for nearly all categories.

| | DOTMC | OTMC | ATM | OTMP | DOTMP | Short | Long | DOTMC | OMTC | ATM | OTMP | DOTMP | Short | Long |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Panel A: Amazon | | | | | | | Panel B: Tesla | | | | | | |
| BS | 5.49 | 6.16 | 5.73 | 5.32 | 8.75 | 7.59 | 4.5 | 5.85 | 6.70 | 8.47 | 10.92 | 21.74 | 16.09 | 11.94 |
| BS+NN3 | **1.30** | 1.64 | **1.68** | 1.59 | 1.49 | 1.85 | 0.96 | 2.45 | 2.59 | 2.70 | 2.63 | 2.67 | 3.10 | 1.84 |
| AHBS | 2.57 | 2.09 | 2.16 | 2.04 | 2.28 | 2.53 | 1.58 | 3.06 | 3.01 | 3.07 | 3.05 | 3.22 | 3.57 | 2.41 |
| AHBS+NN3 | 1.33 | **1.63** | **1.68** | **1.58** | **1.39** | **1.82** | 0.96 | 2.43 | 2.58 | 2.69 | 2.62 | 2.65 | 3.08 | 1.83 |
| Heston | 6.03 | 3.85 | 3.82 | 3.65 | 5.91 | 5.70 | 2.30 | 27.74 | 24.94 | 25.99 | 26.68 | 28.86 | 28.01 | 26.43 |
| Heston+NN3 | 2.27 | 1.66 | 1.70 | 1.62 | 2.89 | 2.57 | 0.96 | 3.28 | 2.59 | 2.68 | 2.63 | 2.76 | 3.38 | 1.90 |
| CW | 4.04 | 3.18 | 3.20 | 3.16 | 4.95 | 3.12 | 4.69 | 4.17 | 3.88 | 3.77 | 3.51 | 4.47 | 4.05 | 4.17 |
| CW+NN3 | **1.30** | 1.64 | **1.68** | **1.58** | **1.39** | 1.83 | **0.93** | **2.40** | **2.54** | **2.64** | **2.57** | **2.60** | **3.04** | **1.77** |
| | Panel C: Google | | | | | | | Panel D: Apple | | | | | | |
| BS | 3.10 | 3.82 | 3.49 | 3.34 | 8.59 | 5.43 | 3.59 | 3.53 | 3.87 | 4.29 | 5.55 | 10.13 | 7.03 | 5.19 |
| BS+NN3 | 1.11 | 1.44 | 1.57 | 1.46 | 1.27 | 1.68 | **0.87** | 1.39 | 1.69 | 1.80 | 1.77 | 1.55 | 1.99 | **0.79** |
| AHBS | 2.65 | 1.90 | 1.94 | 1.77 | 2.39 | 2.28 | 1.65 | 2.26 | 2.08 | 2.15 | 2.00 | 2.29 | 2.60 | 1.42 |
| AHBS+NN3 | 1.19 | 1.46 | 1.58 | 1.45 | 1.35 | 1.67 | 1.00 | 1.36 | 1.65 | 1.78 | 1.77 | 1.50 | 1.94 | 0.99 |
| Heston | 3.16 | 1.77 | 1.88 | 1.83 | 3.42 | 2.79 | 1.30 | 2.39 | 1.96 | 2.25 | 2.22 | 2.16 | 2.66 | 1.36 |
| Heston+NN3 | 2.25 | 1.43 | 1.58 | 1.46 | 2.69 | 2.24 | 0.87 | 1.51 | **1.62** | 1.76 | **1.72** | 1.40 | 1.97 | 0.86 |
| CW | 3.09 | 2.38 | 2.39 | 2.40 | 3.70 | 2.56 | 1.65 | 3.40 | 2.94 | 3.02 | 3.19 | 4.64 | 3.58 | 3.60 |
| CW+NN3 | **1.08** | **1.43** | **1.57** | **1.45** | **1.24** | **1.66** | 1.00 | **1.33** | 1.75 | **1.75** | **1.72** | **1.39** | **1.91** | 0.86 |
| | Panel E: JP Morgan | | | | | | | Panel F: Exxon Mobil | | | | | | |
| BS | 4.47 | 4.57 | 4.40 | 3.65 | 4.40 | 4.19 | 4.38 | 1.68 | 2.54 | 3.97 | 3.78 | 4.04 | 4.08 | 1.91 |
| BS+NN3 | **0.75** | **0.65** | **1.69** | 1.45 | **1.19** | **1.55** | **0.76** | 1.24 | 0.99 | 2.61 | 2.41 | **1.24** | **1.93** | **0.77** |
| AHBS | 1.27 | 1.19 | 1.80 | 1.55 | 1.37 | 1.67 | 1.05 | 1.35 | 0.98 | 3.29 | 2.52 | 1.47 | 2.20 | 1.19 |
| AHBS+NN3 | 0.85 | 0.75 | 1.71 | 1.45 | 1.20 | **1.55** | 0.85 | 0.82 | **0.81** | 2.92 | 2.55 | 1.29 | 2.01 | 1.03 |
| Heston | 1.11 | 1.03 | 2.09 | 1.67 | 2.00 | 2.18 | 1.19 | **0.62** | 1.46 | 3.84 | 3.50 | 6.66 | 8.11 | 2.22 |
| Heston+NN3 | 0.92 | 0.73 | 1.70 | 1.45 | 1.49 | 1.74 | 0.81 | 1.03 | 0.92 | 3.01 | 2.67 | 2.34 | 3.11 | 1.15 |
| CW | 3.13 | 1.82 | 1.87 | 1.60 | 1.79 | 1.73 | 1.91 | 4.59 | 3.45 | 2.86 | 2.58 | 3.33 | 2.96 | 3.18 |
| CW+NN3 | 0.98 | 0.67 | **1.69** | **1.43** | 1.20 | 1.56 | 0.77 | 1.60 | 1.90 | **2.06** | **1.94** | 1.40 | 2.16 | 0.95 |

Table 13: IVRMSEs of the 1-day ahead prediction exercise in the option cross-section across categories, the models considered are the four (boosted) parametric models. The models are boosted by nonparametrically estimating the error surface with a feedforward neural network with three hidden layers. The different panels indicate the underlying equities, the columns indicate the moneyness or tenor category and the rows indicate the model used.

To get more insights on the long-term performance of the nonparametric models, we examine tables 14 and 15, which show the IVRMSE values of the (boosted) BS model for a 5- and 21-day ahead prediction exercise. It is again visible that for nearly all equity options the best prediction errors increase as the options move from DOTMC towards ATM and decrease again as the options move from ATM towards DOTMP, for the two prediction exercises. ATM options again nearly always have the highest IVRMSE values. Furthermore, for all equity options, the boosted model applied to short-term options leads to much higher IVRMSEs in comparison to when applied to long-term options. A visible difference is that in contrast to the results of the 1-day ahead prediction exercise, the boosted model is not always the best-performing model for the 5- and 21-day ahead prediction exercises. Especially, for ATM and OTMP options the AHBS or the Heston model sometimes comes out as the best-performing model for the 5-day ahead prediction exercise. For the 21-day ahead prediction exercise, this fact becomes even more evident as even the BS model sometimes comes out as the best-performing model and parametric models come even more out as the best performing model. Another observation to make from Tables 13, 14 and 15, is that as the prediction horizon increases, the less obvious it becomes which model is the best performing.

| | DOTMC | OTMC | ATM | OTMP | DOTMP | Short | Long | DOTMC | OMTC | ATM | OTMP | DOTMP | Short | Long |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Panel A: Amazon | | | | | | | Panel B: Tesla | | | | | | |
| BS | 6.39 | 7.55 | 7.61 | 7.03 | 10.08 | 9.45 | 5.53 | 10.70 | 11.76 | 11.29 | 10.86 | 17.19 | 14.55 | 11.55 |
| BS+NN3 | **3.68** | **4.39** | 4.42 | 4.07 | **3.33** | 4.85 | **2.27** | 5.16 | 5.28 | 5.95 | 5.55 | **5.92** | 6.48 | 3.91 |
| AHBS | 4.62 | 4.45 | **4.32** | **4.06** | 4.00 | 4.99 | 2.84 | 5.78 | 5.57 | 6.02 | 5.75 | 6.34 | 6.58 | 4.85 |
| AHBS+NN3 | 3.74 | 4.42 | 4.44 | 4.10 | 3.39 | 4.85 | 2.40 | 5.17 | 5.32 | 5.98 | 5.61 | 6.02 | **6.46** | 4.19 |
| Heston | 8.64 | 6.11 | 5.46 | 4.88 | 7.63 | 7.88 | 3.77 | 21.95 | 17.96 | 18.83 | 19.38 | 22.19 | 21.0 | 19.75 |
| Heston+NN3 | 4.68 | 4.41 | 4.43 | 4.09 | 4.38 | 5.36 | 2.30 | 5.85 | 5.24 | **5.92** | 5.55 | 6.31 | 6.82 | 3.91 |
| CW | 5.73 | 5.39 | 5.24 | 5.00 | 6.39 | 5.54 | 5.70 | 6.23 | 5.96 | 6.20 | 5.75 | 6.93 | 6.59 | 6.03 |
| CW+NN3 | **3.68** | **4.39** | 4.42 | 4.07 | **3.33** | **4.84** | 2.30 | **5.07** | **5.23** | 5.93 | **5.54** | **5.92** | 6.47 | **3.85** |
| | Panel C: Google | | | | | | | Panel D: Apple | | | | | | |
| BS | 4.61 | 5.35 | 5.52 | 4.95 | 9.06 | 7.11 | 4.28 | 5.05 | 5.77 | 5.06 | 5.18 | 9.04 | 6.82 | 5.14 |
| BS+NN3 | **2.61** | **3.23** | 3.35 | 3.15 | 2.75 | 3.76 | **1.71** | 2.72 | 3.22 | 3.19 | 2.97 | 2.47 | 3.60 | 1.71 |
| AHBS | 4.08 | 3.51 | 3.44 | 3.26 | 3.57 | 3.90 | 2.79 | 3.19 | 3.19 | **3.16** | 3.16 | 3.12 | 3.81 | 2.01 |
| AHBS+NN3 | 2.98 | 3.35 | 3.43 | 3.22 | 2.92 | **3.75** | 2.25 | 2.73 | 3.22 | 3.18 | **2.95** | 2.48 | 3.59 | 1.74 |
| Heston | 5.05 | 3.37 | **3.30** | **3.13** | 5.14 | 4.66 | 2.13 | 4.27 | **3.15** | 3.18 | 3.11 | 3.35 | 4.21 | 2.07 |
| Heston+NN3 | 4.01 | 3.27 | 3.37 | 3.15 | 4.13 | 4.28 | 1.84 | 3.56 | 3.21 | 3.18 | **2.95** | 2.58 | 3.88 | **1.69** |
| CW | 4.64 | 3.87 | 3.79 | 3.77 | 5.31 | 4.14 | 4.32 | 3.49 | 3.39 | 3.44 | 3.79 | 4.97 | 3.93 | 3.67 |
| CW+NN3 | 2.64 | 3.24 | 3.36 | 3.14 | **2.72** | **3.75** | 1.72 | **2.70** | 3.21 | 3.18 | 2.96 | **2.44** | **3.58** | **1.69** |
| | Panel E: JP Morgan | | | | | | | Panel F: Exxon Mobil | | | | | | |
| BS | 3.79 | 4.25 | 4.70 | 4.29 | 7.26 | 5.50 | 5.53 | 4.77 | 4.87 | 4.87 | 4.89 | 6.24 | 5.80 | 4.90 |
| BS+NN3 | 2.42 | 2.83 | **2.79** | 2.67 | 2.54 | **3.16** | 1.69 | 3.10 | **2.55** | **2.48** | **2.25** | 2.35 | **2.83** | **1.49** |
| AHBS | 3.17 | 3.29 | 2.85 | **2.66** | 2.73 | 3.30 | 1.87 | 3.58 | 2.86 | 2.64 | 2.26 | 2.50 | 2.95 | 1.69 |
| AHBS+NN3 | 2.66 | 2.91 | **2.79** | 2.67 | 2.54 | **3.16** | 1.74 | 3.20 | 2.67 | 2.56 | 2.28 | 2.41 | 2.88 | 1.59 |
| Heston | 2.79 | 3.09 | 2.95 | 2.72 | 3.44 | 3.62 | 1.96 | 4.50 | 3.02 | 2.81 | 2.50 | 5.45 | 4.94 | 2.04 |
| Heston+NN3 | **2.41** | **2.75** | 2.80 | 2.67 | 3.13 | 3.43 | **1.66** | 3.70 | 2.60 | 2.54 | 2.28 | 3.22 | 3.37 | 1.55 |
| CW | 3.88 | 3.58 | 2.93 | **2.66** | 3.19 | 3.30 | 2.61 | 5.10 | 3.88 | 3.29 | 2.90 | 3.88 | 3.55 | 3.44 |
| CW+NN3 | 2.51 | 2.86 | 2.80 | 2.68 | **2.51** | 3.17 | **1.66** | **3.03** | 2.59 | 2.52 | 2.26 | **2.33** | 2.84 | **1.49** |

Table 14: IVRMSEs of the 5-day ahead prediction exercise in the option cross-section across categories, the models considered are the four (boosted) parametric models. The models are boosted by nonparametrically estimating the error surface with a feedforward neural network with three hidden layers. The different panels indicate the underlying equities, the columns indicate the moneyness or tenor category and the rows indicate the model used.

| | DOTMC | OTMC | ATM | OTMP | DOTMP | Short | Long | DOTMC | OMTC | ATM | OTMP | DOTMP | Short | Long |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Panel A: Amazon | | | | | | | Panel B: Tesla | | | | | | |
| BS | **6.27** | 7.46 | 7.53 | 6.92 | 9.89 | 9.29 | 5.47 | 10.75 | 11.81 | 11.33 | 10.88 | 17.22 | 14.60 | 11.57 |
| BS+NN3 | 6.44 | 8.23 | 8.78 | 8.26 | **6.10** | 9.48 | **3.88** | **8.51** | 9.20 | 9.73 | 9.70 | 9.38 | 10.90 | 5.75 |
| AHBS | 7.44 | 8.07 | 8.36 | 7.98 | 6.67 | 9.39 | 4.23 | 9.02 | 9.29 | 9.62 | 9.58 | 9.64 | 10.85 | 6.52 |
| AHBS+NN3 | 6.52 | 8.27 | 8.79 | 8.28 | 6.16 | 9.49 | 4.00 | 8.58 | 9.25 | 9.77 | 9.74 | 9.41 | 10.87 | 6.02 |
| Heston | 9.33 | **7.38** | **7.29** | **6.65** | 8.46 | 9.45 | 4.31 | 22.21 | 18.13 | 18.92 | 19.45 | 22.63 | 21.30 | 19.88 |
| Heston+NN3 | 7.49 | 8.26 | 8.78 | 8.27 | 6.79 | 9.83 | 3.89 | 9.12 | 9.20 | 9.72 | 9.70 | 9.91 | 11.27 | 5.80 |
| CW | 7.69 | 7.89 | 8.12 | 7.83 | 8.02 | **8.79** | 6.55 | 9.19 | 9.34 | **9.58** | **9.39** | 9.91 | **10.52** | 7.71 |
| CW+NN3 | 6.47 | 8.24 | 8.77 | 8.27 | 6.13 | 9.47 | 3.95 | **8.51** | **9.18** | 9.72 | 9.68 | **9.33** | 10.88 | **5.73** |
| | Panel C: Google | | | | | | | Panel D: Apple | | | | | | |
| BS | 4.49 | **5.24** | **5.39** | **4.74** | 8.78 | 6.89 | 4.22 | 4.97 | 5.65 | 4.88 | 5.04 | 8.93 | 6.67 | 5.08 |
| BS+NN3 | **4.37** | 5.71 | 6.14 | 5.67 | 4.50 | 6.71 | **2.94** | **4.41** | 5.20 | 5.16 | 4.66 | 3.87 | 5.80 | **2.76** |
| AHBS | 6.00 | 5.89 | 6.02 | 5.65 | 5.27 | 6.76 | 3.83 | 4.95 | 5.11 | 4.99 | 4.83 | 4.53 | 5.93 | 3.02 |
| AHBS+NN3 | 4.70 | 5.83 | 6.21 | 5.75 | 4.66 | 6.74 | 3.37 | 4.48 | 5.31 | 5.17 | 4.67 | 3.82 | 5.80 | 2.82 |
| Heston | 6.75 | 5.25 | 5.58 | 5.13 | 6.35 | 6.85 | 3.23 | 6.34 | **4.66** | **4.50** | **4.13** | 4.10 | 5.98 | 2.83 |
| Heston+NN3 | 6.08 | 5.73 | 6.16 | 5.68 | 5.75 | 7.19 | 2.99 | 5.89 | 5.30 | 5.17 | 4.65 | 3.95 | 6.30 | 2.78 |
| CW | 6.04 | 5.77 | 5.80 | 5.61 | 6.46 | **6.36** | 5.15 | 4.86 | 4.93 | 4.90 | 5.09 | 5.93 | **5.72** | 4.22 |
| CW+NN3 | 4.43 | 5.73 | 6.15 | 5.68 | **4.48** | 6.72 | 2.98 | 4.42 | 5.31 | 5.18 | 4.67 | **3.81** | 5.80 | 2.77 |
| | Panel E: JP Morgan | | | | | | | Panel F: Exxon Mobil | | | | | | |
| BS | 3.85 | 4.26 | 4.67 | 4.21 | 7.10 | 5.41 | 5.45 | 4.15 | 4.38 | 4.46 | 4.49 | 5.63 | 5.18 | 4.63 |
| BS+NN3 | **3.12** | 3.83 | 4.17 | 4.08 | 3.88 | 4.71 | 2.57 | 4.32 | 3.77 | **3.66** | **3.46** | **3.67** | **4.23** | **2.50** |
| AHBS | 3.98 | 4.44 | 4.23 | 4.06 | 4.04 | 4.83 | 2.77 | 4.79 | 4.16 | 3.92 | 3.50 | 3.78 | 4.37 | 2.69 |
| AHBS+NN3 | 3.46 | 3.99 | 4.18 | 4.08 | 3.88 | 4.71 | 2.66 | 4.47 | 3.91 | 3.74 | 3.49 | 3.69 | 4.26 | 2.59 |
| Heston | 3.16 | 4.13 | 4.30 | 4.07 | 5.04 | 5.27 | 2.79 | 5.01 | 4.22 | 3.91 | 3.54 | 6.53 | 6.11 | 2.97 |
| Heston+NN3 | 2.93 | **3.70** | 4.17 | 4.08 | 4.72 | 5.08 | 2.57 | **4.07** | **3.70** | 3.68 | 3.51 | 4.43 | 4.68 | 2.57 |
| CW | 4.62 | 4.68 | **4.15** | **3.95** | 4.47 | **4.70** | 3.51 | 6.67 | 5.15 | 4.39 | 3.95 | 4.82 | 4.63 | 4.39 |
| CW+NN3 | 3.27 | 3.94 | 4.17 | 4.07 | **3.85** | 4.72 | **2.55** | 4.31 | 3.84 | 3.89 | 3.64 | 3.71 | 4.39 | 2.57 |

Table 15: IVRMSEs of the 21-day ahead prediction exercise in the option cross-section across categories, the models considered are the four (boosted) parametric models. The models are boosted by nonparametrically estimating the error surface with a feedforward neural network with three hidden layers. The different panels indicate the underlying equities, the columns indicate the moneyness or tenor category and the rows indicate the model used.

Finally, figure 3 graphically displays the dynamics of the performance of boosted BS model applied to all equity options. More specifically, the performance of a BS-model corrected by a 3-layer neural network for a 1-day ahead prediction exercise. The dynamic forecasting performance of the model on Amazon and Tesla option data is displayed in Figure 3a; Google and Apple option data is visualized in Figure 3b; JP Morgan and Exxon Mobil option data is showcased in Figure 3c. It is important to keep in mind that the Y-axis of Figure 3a is twice as large as the latter two figures. A local smoother with a window set to 10 days, is used to better visualize the trend. It becomes clear that the performance of the model applied to Tesla options deteriorates in the last year of the sample period, nevertheless, its stability is varying a lot over time. Moreover, Figure 3b clearly shows some type of interaction between the performance of Google and Apple. Whenever the error of one of the two increases for a specific day, the other follows either instantly or with some delay, to a similar magnitude. However, on average the error of the model applied on Google options is over time, larger than the error for Apple. There are some specific points in time, such as March, 2018 where the performance of the model on nearly all options deteriorates. This might be indicative of a sudden increase in systemic risk that influences the volatility of the equity options.

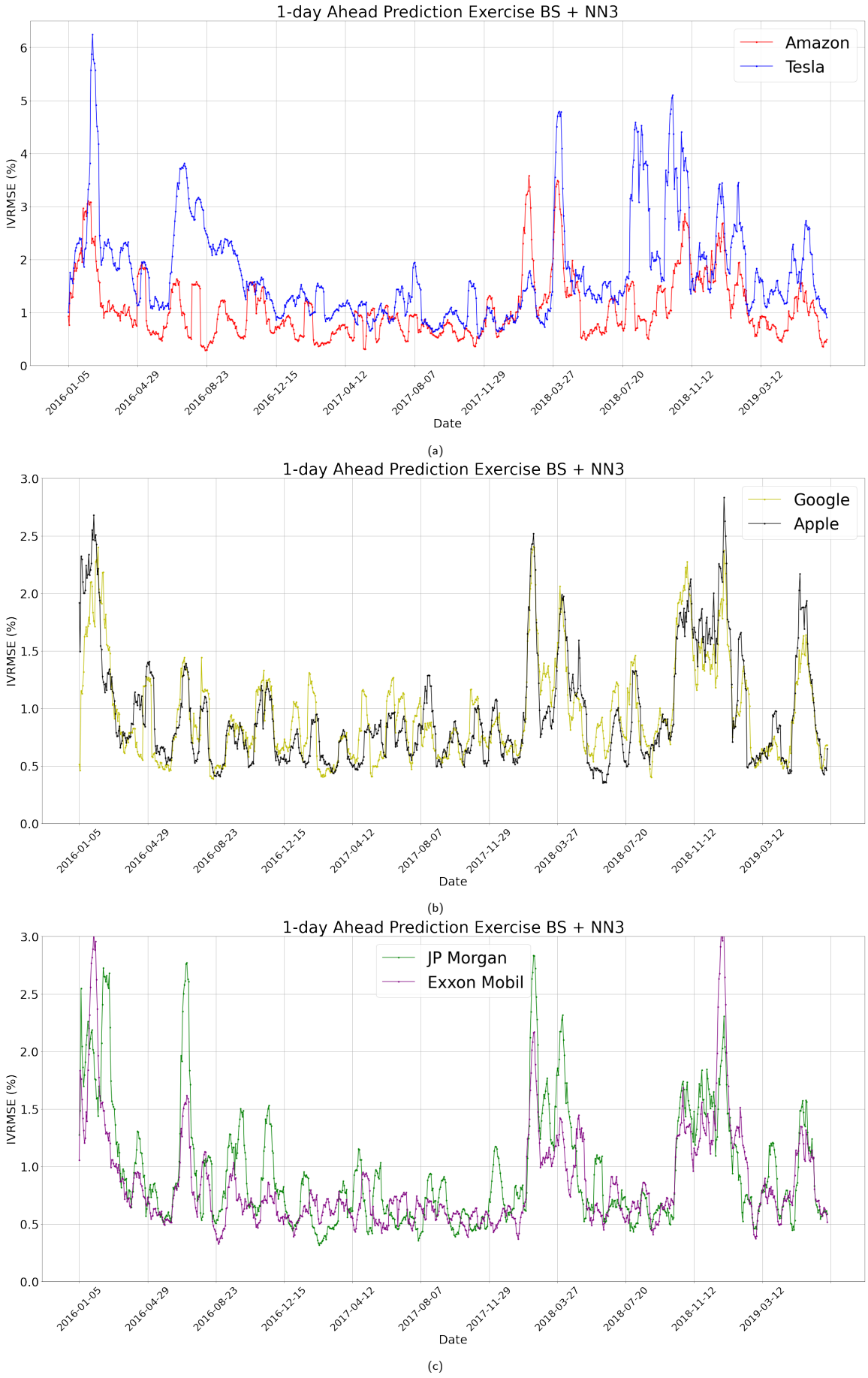


Figure 3: Daily 1-day ahead prediction results (IVRMSE) for the BS model corrected by a 3-layer neural network. The results are displayed for the six underlying equity options as indicated by the legend. The sample period ranges from 4 January 2016 to 28 June 2019

# 6 Conclusion

To conclude this research paper we reiterate the main research question:

*In general, what is the performance increase of the model-guided approach with respect to the underlying parametric model for individual equity options?*

We found that the best predictions for options on S&P 500 had the smallest prediction errors, among all different type of options. As well as the largest performance increase (often twice as large as the best performance increase of equity options). However, the best prediction errors for certain equity options (JP Morgan and Exxon Mobil) come fairly close to those of the options on the S&P 500. Moreover, for equity options with a large mean and standard deviation of implied volatility such as Tesla, the hybrid model delivers a relatively large performance increase. Furthermore, we found that the variability of the prediction errors is larger for equity options than the prediction errors for options on S&P 500.

In the previous section we found that for all prediction exercises and nearly all options, the nonparametric correction of the Black-Scholes model led to the smallest IVRMSE. It is important to keep in mind that this nonparametric correcting model diverges from the model from Almeida et al. (2022), as this model considers a different gradient descent method (more details on this can be found in section 4) and also different hyperparameter settings for the neural network (e.g. different number of epochs). But given these results, we can conclude that we found evidence that this specific model-guided approach does not outperform a neural network fitted directly to the implied volatility surface for equity options.

Furthermore, it is visible that for the one-day ahead prediction exercise the performance increase is the most significant for Tesla, the second best option in terms of performance increase is Amazon and after that Apple. The performance increase for Tesla is the best for all three prediction exercises, however, the performance increase for Amazon (and to some extent Apple) decreases relative to the other options. Even though Tesla (and to some extent Amazon) has the most performance increase, the best predictions for (Amazon) Tesla have the (second to last) largest prediction errors for all three types of prediction exercises relative to the other equity and index options. This might not be surprising as it is a well-known fact that the stock of Tesla is very volatile and overvalued. Its stock value is incredibly high in comparison to competitors that are much larger in terms of sales/revenue. Therefore, it should not come as a surprise that the options on Tesla might be relatively hard to predict, as the underlying stock is not performing as it should be. The best predictions of Exxon Mobil and JP Morgan have the lowest prediction errors (relative to the other options) and after that Apple. It is also not surprising that the performance of the model on Exxon Mobil and JP Morgan has the lowest error. In general investments in the oil/gas and financial sector tend to perform fairly well and stable. Hence, the options on the underlying stock should also be much more stable

and easy to predict. The performance increase for equity options appears to be the largest when the errors of the parametric models are the largest. We found evidence of interaction between the options on Google and Apple. This could also have a logical explanation as these companies are both active in the technical sector. Also, we found that there appear to be some systemic events that increase the error of the hybrid model. If we order the option to thr prediction errors of their best predictions, it comes to notice that this order is the same for all three prediction exercises. Ranked from smallest to largest error this order is Exxon Mobil, JP Morgan, Apple, Google, Amazon and Tesla. This consistent order indicates a certain stability of the model as it can predict certain options better than other options.
We found that for all equity options, the short-term equity options lead to larger prediction errors than the long-term prediction errors. Furthermore, options that are ATM always lead to the largest prediction error relative to the other moneyness categories, whereas the best predictions of DOTMC options nearly always have the smallest prediction error. On average the prediction errors increase as the options move from DOTMC towards ATM and decrease again if they move from ATM towards DOTMP. This is indicative of the fact that options that are ATM are the hardest to predict. This is interesting as the cross-section of options is often the largest for ATM options.

All in all, the results indicate that this parametrically guided nonparametric method can be promising for predicting the implied volatility of equity options. However, its performance may be limited for certain options. The nonparametric correction model brings down the errors to a similar magnitude regardless of the original parametric model. But there, still appears to be more variability between the boosted results of different parametric models on equity options, in comparison to the results on the index option. Therefore an interesting extension would be to find out whether this variability is due to hyperparameters of the nonparametric method or if there is another reason for this variability. Another valuable extension would be to find out whether the relatively poor performance of the boosted Heston model is due to the hyperparameters of the nonparametric model.

# References


Almeida, C., Fan, J., Freire, G., & Tang, F. (2022). Can a machine correct option pricing models? *Journal of Business & Economic Statistics*, 1–14.

Amilon, H. (2003). A neural network versus black–scholes: A comparison of pricing and hedging performances. *Journal of Forecasting*, *22*(4), 317–335.

Amin, K. I., & Ng, V. K. (1993). Option valuation with systematic stochastic volatility. *The Journal of Finance*, *48*(3), 881–910.

Andersen, T. G., Fusari, N., & Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, *117*(3), 558–584.

Audrino, F., & Colangelo, D. (2010). Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing*, *20*, 421–434.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of finance*, *52*(5), 2003–2049.

Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, *16*(1), 101–143.

Bernales, A., & Guidolin, M. (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance*, *46*, 326–342.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637–654.

Carr, P., & Wu, L. (2016). Analyzing volatility risk and risk premium in option contracts: A new theory. *Journal of Financial Economics*, *120*(1), 1–20.

De Spiegeleer, J., Madan, D. B., Reyners, S., & Schoutens, W. (2018). Machine learning for quantitative finance: Fast derivative pricing, hedging and fitting. *Quantitative Finance*, *18*(10), 1635–1643.

Diebold, F. X., & Mariano, R. S. (1995). Com paring predictive accu racy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Duan, J.-C., & Wei, J. (2009). Systematic risk and the price structure of individual equity options. *The Review of Financial studies*, *22*(5), 1981–2006.

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

Fang, F., & Oosterlee, C. W. (2009). Pricing early-exercise and discrete barrier options by fourier-cosine series expansions. *Numerische Mathematik*, *114*(1), 27.

Gordon, M. J. (1959). Dividends, earnings, and stock prices. *The review of economics and statistics*, 99–105.

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

Hasti, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

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.

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

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, S., Oosterlee, C. W., & Bohte, S. M. (2019). Pricing options and computing implied volatilities using neural networks. *Risks*, *7*(1), 16.

Masters, T. (1993). *Practical neural network recipes in c++*. Academic Press Professional, Inc.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, 141–183.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, *6*(4), 525–533.

Oh, D. H., & Patton, A. J. (2021). Better the devil you know: Improved forecasts from imperfect models.

Oosterlee, C. W., & Grzelak, L. A. (2019). *Mathematical modeling and computation in finance: With exercises and python and matlab computer codes*. World Scientific.

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

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance*, *44*(5), 1115–1153.

# A Modelling Implied Volatility

### A.0.1 Ad-Hoc Black-Scholes Model

Dumas et al. (1998) asses the time-series validity of the implied tree approach such as the implied binomial tree developed by Rubinstein (1994) by using options on the S&P500 index. More specifically, they asses the time-series validity of the volatility function used in this implied tree approach as this approach assumes that the volatility function is stable through time. This function is a local volatility rate that is a flexible but deterministic function of the asset price and time. They motivate their research by arguing that the implied tree approach is prone to overfitting as a cross-section of option prices is used to fit the tree which achieves an exact fit of the observed option prices. This is possible as the tree can be defined with the same number of degrees of freedom as the observed option prices. They find that the two most parsimonious volatility functions deliver the best fit. Furthermore, they find that the in-sample parameter estimates are not stable through time (every week), where they find evidence that this instability is not entirely due to economic factors but may be the result of overfitting. For the out-of-sample prediction, they use a benchmark model called the ad-hoc Black-Scholes model. This benchmark model is based on the strategy of market-makers who fit the implied volatility pattern across exercise prices and tenors, after this they value options using the obtained relation and the Black-Scholes formula. The name Ad-Hoc arises from the fact that this model is theoretically inconsistent as the Black-Scholes model assumes a constant volatility. However, the out-of-sample results show that the local volatility models do not outperform this theoretically inconsistent benchmark model. In addition to this, the Ad-Hoc Black-Scholes model displays more stable (time-series) parameter estimates.

### A.0.2 Heston Model

In his famous paper, Heston (1993) tries to generalize the assumptions of constant volatility. He does so by providing a quasi-closed-form solution for the price of a European call option which is not based on the Black-Scholes formula. This solution uses the fact that the spot asset is correlated with volatility, more specifically the Wiener process of the volatility process of an asset is correlated with the Wiener process of that same asset denoted by $\rho$. The solution assumes stochastic volatility but still with constant interest rate. This derivation is based on the following SDE

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dZ_1(t), \tag{17}$$

where $S_t$ is the stock price at time $t$, $\mu$ denotes the drift rate of the stock price, $v_t$ is the stochastic volatility and $dZ_1(t)$ is a Wiener process. The SDE for the volatility is then given by the following Ornstein-Uhlenbeck (mean-reverting) process:

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t} dZ_2(t), \tag{18}$$

where $\kappa$ denotes the speed of the mean reversion, $\theta$ is the unconditional mean of the volatility, the volatility of the volatility is given by $\sigma$ and $dZ_2$ is a second Wiener process, correlated with $dZ_1(t)$ via $\rho$. Then using standard arbitrage arguments, as shown by Black and Scholes (1973), it follows that the value of any European call $C(S_t, \nu_t, t)$ must satisfy the following partial differential equation:

$$\begin{aligned} \frac{1}{2}\nu S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \rho\sigma\nu_t S_t \frac{\partial^2 C}{\partial S_t \partial \nu_t} + \frac{1}{2}\sigma^2 \nu_t \frac{\partial^2 C}{\partial \nu_t^2} + r S_t \frac{\partial C}{\partial S_t} \\ + (\kappa[\theta - \nu_t] - \lambda(S_t, \nu_t, t)) \frac{\partial C}{\partial \nu} - rC_t + \frac{\partial C}{\partial t} = 0 \end{aligned} \tag{19}$$

where $\lambda(S_t, \nu_t, t)$ represents the price of volatility risk, that is independent of the underlying asset. He then considers a European call for which he specifies standard boundary conditions

$$\begin{aligned} C(S_T, \nu_T, T) &= \max(0, S_T - K), \\ C(0, \nu_t, t) &= 0, \\ \frac{\partial C}{\partial S_t}(\infty, \nu_t, t) &= 1, \\ r\frac{\partial C}{\partial S}(S_t, 0, t) + \kappa\theta \frac{\partial U}{\partial \nu}(S_t, 0, t) - rC(S_t, 0, t) + C(S_t, 0, t) &= 0, \\ C(S, \infty, t) &= S_t. \end{aligned} \tag{20}$$

After which he makes an educated guess that the solution of the European call must be of the form:

$$C(S_t, \nu_t, t) = S_t P_1 - K \exp(-r\tau) P_2, \tag{21}$$

where the first term represents the discounted expected value of the stock if the option is exercised and the second term is the present value of the strike price. Hence, $P_1$ can be seen as the conditional probability that given the option is in-the-money, what the expected discount value of the underlying asset is. And $P_2$ can be interpreted as the probability for the option to end up in-the-money. It is convenient to use the transformation: $x_t = log(S_t)$, such that eq. (21) can be substituted into eq. (19). Now it follows that $P_1$ and $P_2$ must satisfy the following partial differential equations:

$$\begin{aligned} \frac{1}{2}\nu \frac{\partial^2 P_j}{\partial x_t^2} + \rho\sigma\nu_t \frac{\partial^2 P_j}{\partial x_t \partial \nu_t} + \frac{1}{2}\sigma^2 \nu_t \frac{\partial^2 P_j}{\partial \nu_t^2} + (r + u_j \nu) \frac{\partial P_J}{\partial x_t} \\ + (a - b_j \nu) \frac{\partial P_j}{\partial \nu} + \frac{\partial P_j}{\partial t} = 0, \end{aligned} \tag{22}$$

for $j = 1, 2$ where

$$u_1 = \frac{1}{2},\ u_2 = -\frac{1}{2},\ a = \kappa\theta,\ b_1 = \kappa + \lambda - \rho\sigma,\ b_2 = \kappa + \lambda$$

These partial differential equations are subject to the terminal condition:

$$P_j(x, \nu, T; log[K]) = \mathbb{1}_{[x \geq \log(K)]}, \tag{23}$$

such that the option price satisfies the terminal condition imposed by eq. (20). He then proves that if $x$ follows the stochastic process as the one in equations (17) and (18), then $P_j$ is the conditional probability that the option expires in the money:

$$P_j(x, \nu, T; log(K)) = Pr\left[x(T) \geq log[K] \mid x_t = x, \nu_t = \nu\right]. \tag{24}$$

Implying that $P_j$ can be used in eq. (21) to find the price of a European call option. He argues that these probabilities $P_j$ are not immediately available in closed form, but he derives their characteristic functions as:

$$f_j(x, \nu, T; \phi) = \exp(C(\tau, \phi) + D(\tau; \phi)\nu + i\phi x) \tag{25}$$

where

$$\begin{aligned} C(\tau; \phi) &= r\phi i\tau + \frac{a}{b}\left\{(b_j - \rho\sigma\phi i + d)\tau - 2\ln\left[\frac{1 - ge^{d\tau}}{1 - g}\right]\right\}, \\ D(\tau, \phi) &= \frac{b_j - \rho\sigma\phi i + d}{\sigma^2}\left[\frac{1 - e^{d\tau}}{1 - ge^{d\tau}}\right] \end{aligned} \tag{26}$$

and

$$\begin{aligned} g &= \frac{b_j - \rho\sigma\phi i + d}{b_j - \rho\sigma\phi i - d}, \\ d &= \sqrt{(\rho\sigma\phi i)^2 - \sigma^2(2u_j\phi i - \phi^2)}. \end{aligned} \tag{27}$$

These characteristic functions can then be used in the COS method as developed by Fang and Oosterlee (2009). The COS method is used to obtain the probabilities $P_1$ and $P_2$ and consequently, using eq. (21) the option price under the Heston model can be computed.

### A.0.3 Carr And Wu Model

Carr and Wu (2016) propose a new framework for option pricing that is based on the fact that institutional investors manage their positions in options by looking at implied volatility as computed from the Black-Scholes formula. They do this rather than looking at the dynamics of the underlying asset and options prices. Whereas traditional option pricing theory requires the full specification of the instantaneous variance of the underlying asset in both the short and long run. They choose to model the near-term dynamics of the Black-Scholes implied volatility surface and derive no-arbitrage constraints directly on its shape. The shape of

the implied volatility surface can be cast as the solution of a quadratic equation under the assumption of the implied volatility dynamics. In contrast to traditional asset pricing models that guarantee type I bounds, which are derived based on no-arbitrage arguments between European options and underlying stocks, by quoting an option in terms of positive implied volatility. They choose to derive the no-arbitrage constraint directly on the current shape of the implied volatility surface. They start their derivation by specifying the dynamics of the underlying asset and the implied volatility of each option:

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{v_t} dW_t \\ d\sigma_t(K,T) &= \mu_t dt + \omega_t dZ_t, \ |\log(K/S_t)| < \kappa, \ 0 < T-t < \tau \\ E[dW_t dZ_t] &= \rho_t dt \end{aligned} \tag{28}$$

where $v_t$ is the instantaneous variance rate and follows a positive, real-valued stochastic process. The standard Brownian motions $dW_T$ and $dZ_t$ have a stochastic correlation $\rho_t$ that takes values on [-1,1]. The drift and volatility of the implied volatility are denoted as $\mu_t$ and $\omega_t$, respectively. Then they propose the following partial differential equation which under the previously specified dynamics guarantees no static arbitrage between options at (K, T) and the underlying stock. Furthermore, it requires no dynamic arbitrage to be allowed on any option (K, T) relative to the basis option $(K_0, T_0)$ and the stock:

$$-B_t = \mu_t B_\sigma + \frac{1}{2} v_t S_t^2 B_{SS} + \rho_t \omega_t \sqrt{v_t} S_t B_{S\sigma} + \frac{1}{2} \omega_t^2 B_{\sigma\sigma} \tag{29}$$

where $B$ denotes the standard Black-Scholes formula for the price of a European call option and the subscript denotes their partial derivatives, e.g. $B_t$ is the partial derivative of the Black-Scholes formula w.r.t $t$ and hence being the theta. This fundamental partial differential equation is then translated, by plugging in the partial derivatives. This results in an algebraic equation linking the implied volatility dynamics to the current shape of the implied volatility surface. Then they consider the following specification for the implied volatility dynamics:

$$\frac{d\sigma_t(K,T)}{\sigma_t(K,T)} = e^{-\eta_t(T-t)}(m_t dt + w_t dZ_t), \tag{30}$$

where the stochastic processes $m_t, w_t$ and $\eta_t$ do not depend on strike price $K$ & maturity date $T$. The stochastic drift is defined as $m_t$, the volatility of the implied volatility $w_t$ is constrained to be a strictly positive process. The exponential dampening $e^{-\eta_t(T-t)}$ is used to accommodate the empirical observation that implied volatilities for options with longer tenors tend to have less volatility. Finally, they apply the previously described dynamics for the underlying stock and implied volatility to the algebraic equation that guarantees no dynamic arbitrage, which

leads to the following quadratic equation:

$$0 = \frac{1}{4}e^{-2\eta_t\tau^2}w_t^2\tau\sigma_t^4 + (1 - 2e^{-\eta_t\tau}m_t\tau - e^{-\eta_t\tau}w_t\rho_t\sqrt{v_t}\tau)\sigma_t^2 - (v_t + 2e^{-\eta_t\tau}w_t\rho_t\sqrt{v_t}k + e^{-2\eta_t\tau}w_t^2k^2), \tag{31}$$

where $\tau = T - t$ is the time to maturity. From eq. (31) it can be seen that no-arbitrage constraint depends on the current levels of the dynamic processes $(m_t, w_t, \eta_t, v_t, \rho_t)$, but not on the exact dynamics. Hence, this model relies on the near-term dynamics of the Black-Scholes implied volatility surface as opposed to traditional option pricing models.

## A.1 Neural Network

### A.1.1 Asset Pricing with Machine Learning

Gu et al. (2020) conduct a comparative analysis of the performance of machine learning methods for predicting equity returns. Although the literature on equity risk premia is quite extensive, the literature of estimating equity risk premia using machine learning methods is still relatively small. The prediction for equity returns is done in the cross-section and time series. As a benchmark model, they consider a panel regression of individual stock returns onto size, book-to-market and momentum, hence following traditional asset pricing literature. They find that random forest and neural networks perform the best as they in general have the highest out-of-sample $R^2$ among all considered methods. Neural networks achieve the highest performance when three hidden layers are considered (with a geometric pyramid rule structure), out of a maximum of five hidden layers. Furthermore, statistical tests (Diebold-Mariano tests) show a clear rejection of the OLS benchmark in favor of nonparametric machine learning tools. All methods share the objective functions which minimizes the mean squared predictions error.

The neural networks considered by Gu et al. (2020) are feed-forward networks, that consist of an input layer of predictors, one or more hidden layers that nonlinearly transform the predictors and an output layer that aggregates the hidden layers into a prediction. The number of units in the input layer is equal to the number of predictors, these inputs are connected to the first hidden layer via weights. The hidden layer consists of a specific amount of intermediary outputs, which are called neurons, these neurons are connected to the other hidden layers and finally to the output layer. In the simplest case of no hidden layers, all predictor signals, e.g. $(z_1, ..., z_4)$, are connected via a weight (denoted $\theta_1, ..., \theta_4$) and an intercept (denoted $\theta_0$) to the output layer. The output layer is therefore the aggregation of the weighted signals and the intercept; $\theta_0 + \sum_{k=1}^{4} z_k\theta_k$. Hence, this neural network is a linear mapping $\mathbb{R}^n \to \mathbb{R}$, with $n$ equal to the dimension of the predictors.

A more advanced example with $L$ hidden layers, $N$ predictors defined as $x^{(0)} = (1, z_1, ..., z_N)'$

and let $K^{(l)}$ denote the number of neurons for hidden layer $l$ . For each layer $l = 1, ..., L$ there are $K^{(L)}$ neurons $x^{(l)} = (1, x_1^{(l)}, ..., x_{K^{(L)}}^{(l)})'$. Each neuron in the hidden layer draws information from all predictors in a linear way using the weights as before. After which each neuron applies an activation function $f$ to its aggregated signal. Mathematically for neuron $i$ and hidden layer $l$ the recursive output formula is defined as:

$$x_k^{(l)} = f(x'^{(l-1)}\theta_k^{(l-1)}) \tag{32}$$

where $x_k^{(l)}$ is the output of the $k$th neuron in layer $l$ and $f$ denotes any nonlinear activation function. The network is initialized using an input layer with the raw predictors, $x^{(0)} = (1, z_1, ..., z_N)'$ The final outputs are then aggregated into an ultimate output forecast

$$g(z, \theta) = x'^{(L-1)}\theta^{(L-1)}, \tag{33}$$

where for clarification, $x'^{(L-1)}$ are the outputs of the neurons of the last hidden layer. It becomes obvious that certain design choices, such as the number of hidden layers, neurons and activation function have a significant influence on the outcome of the neural network. Gu et al. (2020) point out that there are infinitely many design choices to make and there are arguments in favor and against deep layer networks. It is however infeasible to consider them all, therefore they suggest a certain type of structure such that they can achieve a lower bound on the performance of machine learning methods. They consider the following networks: A neural network with one hidden layer that consists of 32 neurons, denoted as NN1. A neural network with two hidden layers with 32 and 16 neurons respectively, denoted as NN2. They continue with this logic until NN5, which has five hidden layers with 32, 16, 8, 4 and 2 neurons, respectively. This design choice allows the authors to infer the trade-offs of network depth. Regarding the design choice for the nonlinear activation function, there are many options such as a sigmoïd, hyperbolic or softmax function. The activation function is the component that makes the parametric network nonparametric. They use the same activation function for every neuron, this activation function is the rectified linear unit (ReLu) function and is defined as

$$ReLu(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ otherwise,} \end{cases} \tag{34}$$

literature shows that it allows for faster derivative evaluation and promotes sparsity in the number of active neurons in comparison to more traditional activation functions e.g. sigmoïd function. They estimate the weights of the neural network by minimizing the following loss function.

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{35}$$

where $y$ and $\hat{y}$ denote the true and predicted value (excess stock return) respectively. The model is fitted to the training data using a stochastic gradient descent method where the learning rate is shrunk using Kingma and Ba (2014)'s stochastic gradient descent algorithm. The authors use $l_1$ regularization to ensure early stopping and an ensemble method to decrease the variance of the final model (discussed in the next section).

### A.1.2 Guidelines for training a Neural Network

Some aspects are important to keep in mind when making design choices for a neural network as described by Hasti et al. (2009).
The most stable way to initialize the neural network is to randomly set the starting values to a value near zero. If the weights are set to zero, the derivatives will also be zero which leads to a non-moving algorithm. Starting with large weights often leads to poor solutions.
To prevent overfitting an early stopping rule is used, which boils down to stopping the algorithm well before the global minimum is reached. To determine when to stop this algorithm a validation set is used. A method similar to ridge regression is weight decay. Where a penalty is added to the error function $R(\theta) + \lambda J(\theta)$, where

$$J(\theta) = \sum_{k,m} \beta_{km}^2 + \sum_{m,l} \alpha_{ml}^2 \tag{36}$$

and $\lambda \geq 0$ is defined as a tuning parameter, with large values shrinking the weights to zero.
Standardization of input data to have mean zero and standard deviation one is crucial for the regularization process.
The authors claim that it is better to have too many hidden layers than too few, they advise setting the number of hidden layers somewhere in the range of 5 to 100.
Finally, it is possible for a nonconvex error function $R(\theta)$ to have many local minima. To decrease the variance of the predictions, it is important to try several random starting configurations and use the average predictions of these configurations as a final prediction (ensemble method). Another option would be to use bagging to decrease the variance, this uses randomly perturbed versions of the training data.