Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ubes20

Can a Machine Correct Option Pricing Models?

Caio Almeida, Jianqing Fan, Gustavo Freire & Francesca Tang

To cite this article: Caio Almeida, Jianqing Fan, Gustavo Freire & Francesca Tang (2023) Can a Machine Correct Option Pricing Models?, Journal of Business & Economic Statistics, 41:3, 995-1009, DOI: 10.1080/07350015.2022.2099871

To link to this article: https://doi.org/10.1080/07350015.2022.2099871

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 3, 995–1009
https://doi.org/10.1080/07350015.2022.2099871

Can a Machine Correct Option Pricing Models?

Caio Almeida[a], Jianqing Fan[b], Gustavo Freire[c], and Francesca Tang[b]

[a]Department of Economics, Princeton University, Princeton, NJ; [b]Operations Research and Financial Engineering, Princeton University, Princeton, NJ; [c]Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, Netherlands

ABSTRACT

We introduce a novel two-step approach to predict implied volatility surfaces. Given any fitted parametric option pricing model, we train a feedforward neural network on the model-implied pricing errors to correct for mispricing and boost performance. Using a large dataset of S&P 500 options, we test our nonparametric correction on several parametric models ranging from ad-hoc Black–Scholes to structural stochastic volatility models and demonstrate the boosted performance for each model. Out-of-sample prediction exercises in the cross-section and in the option panel show that machine-corrected models always outperform their respective original ones, often by a large extent. Our method is relatively indiscriminate, bringing pricing errors down to a similar magnitude regardless of the misspecification of the original parametric model. Even so, correcting models that are less misspecified usually leads to additional improvements in performance and also outperforms a neural network fitted directly to the implied volatility surface.

ARTICLE HISTORY
Received May 2021
Accepted July 2022

KEYWORDS
Boosting; Deep learning; Implied volatility; Model correction; Stochastic volatility

1. Introduction

Option prices observed on a particular day are commonly represented by the implied volatility surface.[1] This one-to-one mapping is useful as implied volatilities of different options are easier to compare than option prices, both in the cross-section and across time. The shape of the implied volatility surface also conveys information about the risk-neutral distribution of the underlying asset returns, allowing for a more concise identification of market expectations about future price movements. In fact, it is well-known that institutional investors manage their option positions through implied volatility (Carr and Wu 2016), which also plays a key role in the estimation of structural option pricing models (Andersen, Fusari, and Todorov 2015; Ait-Sahalia, Li, and Li 2021). Therefore, accurately fitting and predicting the implied volatility surface is critical for both researchers and traders in the option market.

It is a well-established stylized fact that observed implied volatility surfaces strongly contradict the seminal Black and Scholes (1973) model, which predicts a constant implied volatility across time, moneyness and maturity. In particular, the implied volatilities of S&P 500 options display a pronounced "skew" or "smile" in the cross-section (Rubinstein 1994), besides changing over time following a highly nonlinear dynamics (Andersen, Fusari, and Todorov 2015). Given the severe misspecification of the Black–Scholes model, a large spectrum of option pricing models has been proposed to capture the observed implied volatility patterns. While substantial progress has been made in this direction, any parametric model still incurs pricing errors when fitting the implied volatility surface.

In this article, we propose a simple two-step approach to predict implied volatility surfaces that is based on a nonparametric correction of any given parametric option pricing model. First, we fit the parametric model to the observed implied volatilities. Then, we estimate nonparametrically the model-implied pricing error function. Our method can be seen as a generalization of a purely nonparametric estimation of the implied volatility surface. Indeed, a direct nonparametric fit is equivalent to correcting the Black–Scholes model, which does not provide any information about the shape of the implied volatility surface. This motivates correcting a more sensible parametric model that can capture some of the patterns in implied volatilities, as its flattened pricing error surface should be easier to estimate nonparametrically.[2]

Our approach is also motivated by the fact that investors trade options on a daily basis based on a variety of option pricing models. These models in general provide misspecified but useful views about the dynamics of the underlying asset as well as the evolution of the implied volatility surface. From a broader sense, Oh and Patton (2021) point out that many important economic decisions are taken based on a parametric forecasting model that

[1]The implied volatility surface represents the implied volatility of each option as a function of its moneyness and time to maturity. The implied volatility of a given option is the volatility parameter that makes the Black and Scholes (1973) formula deliver the observed option price.

[2]See Glad (1998), Fan and Ullah (1999), and Fan, Wu, and Feng (2009) for the advantages of a parametrically guided nonparametric approach.

CONTACT Caio Almeida *calmeida@princeton.edu* Department of Economics, Princeton University, Princeton, NJ.

is known to be good but imperfect. Hence, introducing an easily adaptable methodology based on boosting the accuracy of parametric option pricing models may lead to superior decisions in terms of option pricing, hedging and risk management.

We use feedforward neural networks to nonparametrically correct option pricing models. This is a natural choice as they possess a "universal approximation" property, being able to approximate virtually any function with arbitrary precision (Cybenko 1989; Hornik, Stinchcombe, and White 1989; Poggio et al. 2017). In addition, they incorporate many statistical and optimization techniques that help avoid the curse of dimensionality with good generalization properties (Fan, Ma, and Zhong 2021). Neural networks also have a long tradition in option pricing due to their ability to model nonlinear relationships nonparametrically (Hutchinson, Lo, and Poggio 1994; Garcia and Gençay 2000), and superior performance amongst machine learning methods for prediction tasks in finance (Gu, Kelly, and Xiu 2020).

Using a large dataset of S&P 500 options, we test our neural network correction on a diverse set of parametric option pricing models: the Black and Scholes (1973) model, which does not provide information about the shape of the implied volatility surface; the Heston (1993) structural stochastic volatility model; the ad-hoc correction of Black–Scholes proposed by Dumas, Fleming, and Whaley (1998), which expresses the surface as a quadratic polynomial of moneyness and time to maturity; and the Carr and Wu (2016) model, which provides a parametric specification for the dynamics of the implied volatility surface. We consider out-of-sample prediction exercises in the cross-section for different horizons and in the option panel.

As a whole, we find strong empirical support for our model-guided nonparametric approach. The nonparametrically corrected models consistently outperform the respective original ones for all prediction exercises, and in many cases by a large extent. Moreover, the neural network is flexible enough to bring pricing errors down to a similar magnitude regardless of the misspecification of the parametric model. Even so, correcting models that are able to capture some of the features of the implied volatility surface usually leads to additional improvements in performance. In particular, doing so outperforms a direct nonparametric fit, due to the flattened pricing error surface as previously explained.

For the prediction exercise in the option panel, we further demonstrate the benefits of including time-varying covariates in estimation, which allow the neural network to learn the shape of the misspecification of a model as a function of different states of the economy. Our results show that the VIX index, which is a popular benchmark for market risk, and the measure of market jump risk of Bollerslev, Todorov, and Xu (2015) are the most relevant covariates in the nonparametric correction. This indicates that information about the evolution of volatility and jump risks is important for a more accurate fit of the option panel. In contrast, macroeconomic variables do not seem to contribute significantly to the correction of the parametric models.

The remainder of the article is organized as follows. After a brief review of the related literature, Section 2 presents the considered parametric option pricing models. Section 3 describes our model-guided nonparametric approach, while Section 4 explains the option data. Sections 5 and 6 discuss the empirical results of the prediction exercise in the cross-section and in the option panel, respectively. Finally, Section 7 concludes the article.

1.1. Related Literature

This article is related to the extensive option pricing literature. Given the severe misspecification of the Black and Scholes (1973) model and its constant volatility assumption, several extensions and more general specifications have been proposed. Dupire (1994) and Rubinstein (1994) consider that local volatility is a deterministic function of the underlying asset price and time, estimating it as a binomial lattice. Dumas, Fleming, and Whaley (1998) show that an ad-hoc correction of Black–Scholes that smooths implied volatilities outperforms local volatility models. A different strand of research develops structural continuous-time option pricing models incorporating additional sources of risk such as stochastic volatility and jumps (Heston 1993; Bates 2000; Duffie, Pan, and Singleton 2000; Andersen, Fusari, and Todorov 2015). Another important category consists of models that use approximation theory to express implied volatilities as a function of parameters of stochastic volatility models (Medvedev and Scaillet 2007; Gatheral and Jacquier 2014; Carr and Wu 2016).[3] Considering all of these advancements, we provide an easily adaptable methodology to correct any given parametric option pricing model and boost its predictive accuracy.

Alternative methods have been proposed to nonparametrically estimate the state price density or risk-neutral distribution from option prices, borrowing from the insights of Ross (1976) and Breeden and Litzenberger (1978). This is useful as the option price can be obtained by taking the expectation under the risk-neutral distribution of its payoff. Prominent examples include Jackwerth and Rubinstein (1996), Stutzer (1996), Ait-Sahalia and Lo (1998), Ait-Sahalia and Duarte (2003), and Fan and Mancini (2009). In our case, we nonparametrically estimate the implied volatility surface, instead of the state price density, using a parametric model as a template.

Finally, a number of papers have employed neural networks for option pricing and hedging purposes due to their flexibility in modeling nonlinear relations. Hutchinson, Lo, and Poggio (1994), Garcia and Gençay (2000), and Amilon (2003) use neural networks to nonparametrically estimate the option pricing function. Taking it a step further, Dugas et al. (2009) and Ackerer, Tagasovska, and Vatter (2020) incorporate no-arbitrage restrictions in the network design. Buhler et al. (2019) and Ruf and Wang (2021) apply neural networks for the hedging of options, while Liu, Oosterlee, and Bohte (2019) use them to numerically solve option pricing models with the aim of reducing computational time. Along a similar vein but adapted to our specific purposes, we combine useful parametric model-based information with the power of neural networks to predict implied volatility surfaces.

[3] More recently, Ait-Sahalia, Li, and Li (2021) reverse this logic by linking observed shape characteristics of the implied volatility surface to the coefficients of stochastic volatility models, while Bandi, Fusari, and Renò (2021) expand the characteristic function of the underlying asset process to price short-maturity options and study the relation between equity characteristics and sources of structural risk.

2. Parametric Option Pricing Models

2.1. Black–Scholes Model

The seminal Black and Scholes (1973) model allows for a closed-form solution to the price of a European option, which is possible due to strong parametric assumptions. More specifically, its original pricing formula is derived by assuming that the price of the underlying asset S_t follows a geometric Brownian motion under the physical measure:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{1}$$

where μ is the expected asset return, σ is the constant instantaneous volatility and W_t is a standard Wiener process. As can be seen, the only source of risk is given by the underlying price process.

The Black–Scholes option pricing formula of a European call with a strike price K and time to maturity $\tau = (T - t)$ is given by:

$$C_{\text{BS}}(S_t, K, \tau, r, \sigma) = \Phi(d_1)S_t - \Phi(d_2)Ke^{-r\tau}, \tag{2}$$

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}\left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau\right],$$

$$d_2 = d_1 - \sigma\sqrt{\tau},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and r is the risk-free rate.[4] Given an option price C_i observed in the market, the so-called implied volatility of the option solves $\sigma_i = C_{\text{BS}}^{-1}(C_i; S_t, K_i, \tau_i, r)$, where $C_{\text{BS}}^{-1}(\cdot)$ is the inverse of the Black–Scholes pricing formula (2) as a function of σ.[5] In other words, the implied volatility is simply the volatility parameter σ_i that makes $C_{\text{BS}}(S_t, K_i, \tau_i, r, \sigma_i)$ equal to the observed price C_i.

The Black–Scholes model predicts that implied volatility should be constant across time, strike prices and maturities. That is, at any given point in time, the model predicts that the implied volatility surface is perfectly flat. However, it is well-known that empirical data strongly reject these predictions. In particular, for the S&P option market, which comes closest to satisfying the model assumptions, there is a pronounced "skew" or "smile" in implied volatility as a function of strike prices and maturities (Rubinstein 1994). Moreover, the implied volatility surface changes significantly over time, displaying highly nonlinear dynamics (Andersen, Fusari, and Todorov 2015). Given the severe misspecification of the Black–Scholes model, different option pricing models have been proposed to capture the persistent patterns observed in implied volatility, some of which we consider in the next sections.

2.2. Ad-hoc Black–Scholes Model

Dumas, Fleming, and Whaley (1998) propose as a benchmark model the so-called ad-hoc or practitioner Black–Scholes (AHBS). The AHBS is based on the fact that practitioners are known to fit the implied volatility surface and then translate it to options prices using the Black–Scholes formula with the specific implied volatility of each option. This procedure is ad-hoc in the sense that it is inconsistent with the original Black–Scholes model and its constant volatility assumption, providing instead a reduced-form approximation to observed implied volatility patterns.

The AHBS model is operationalized by specifying implied volatility as a quadratic function of the strike price (or moneyness) and time to maturity.[6] Defining moneyness as $m_{i,t} = S_t/K_{i,t}$, we estimate the AHBS model on a given day t with a cross-section of $i = 1, \ldots, n$ options using the following regression:[7]

$$\begin{aligned}\sigma_{i,t} = {} & a_{0,t} + a_{1,t}m_{i,t} + a_{2,t}m_{i,t}^2 + a_{3,t}\tau_{i,t} \\ & + a_{4,t}\tau_{i,t}^2 + a_{5,t}m_{i,t}\tau_{i,t} + \epsilon_{i,t}, \quad i = 1, \ldots, n,\end{aligned} \tag{3}$$

where $\sigma_{i,t}$, $m_{i,t}$, and $\tau_{i,t}$ are the observed implied volatilities, moneyness and times to maturity of the options, respectively. The parameters in (3) are estimated via ordinary least squares. This amounts to minimizing the implied volatility mean squared error (IVMSE):

$$\frac{1}{n}\sum_{i=1}^{n}[\sigma_{i,t} - \sigma_{\text{AHBS}}(\boldsymbol{a}_t, m_{i,t}, \tau_{i,t})]^2, \tag{4}$$

where $\sigma_{\text{AHBS}}(\boldsymbol{a}_t, m_{i,t}, \tau_{i,t}) = a_{0,t} + a_{1,t}m_{i,t} + a_{2,t}m_{i,t}^2 + a_{3,t}\tau_{i,t} + a_{4,t}\tau_{i,t}^2 + a_{5,t}m_{i,t}\tau_{i,t}$ and $\boldsymbol{a}_t$ is the vector of parameters to be estimated. Given the estimated parameters $\hat{\boldsymbol{a}}_t$, the implied volatility predicted by the AHBS model for an option with moneyness m and time to maturity τ is $\sigma_{\text{AHBS}}(\hat{\boldsymbol{a}}_t, m, \tau)$.

The AHBS model nests the volatility function implied by the Black–Scholes model as a particular case. By setting all parameters in (3) to zero except for $a_{0,t}$, we obtain a constant implied volatility model: $\sigma_{\text{BS}}(a_{0,t}, m_{i,t}, \tau_{i,t}) = a_{0,t}$. When estimating the Black–Scholes model (BS) in the cross-section, we run regression (3) only with the intercept, which amounts to calculating the average implied volatility observed on day t. The prediction of the BS model is then simply $\hat{a}_{0,t}$ for any option.

2.3. Heston Model

Stochastic volatility models generalize the constant volatility assumption of the BS model in a structural way that assumes the volatility of the underlying asset price to be itself a random process. Thus, stochastic volatility represents an additional source of risk. The Heston (1993) model falls into this category and specifies the following dynamics for the underlying asset under the risk-neutral measure:

$$\frac{dS_t}{S_t} = rdt + \sqrt{V_t}\, dW_{1,t}, \tag{5}$$

[4] Essentially, the formula is obtained by noticing that the option can be dynamically hedged by buying and/or selling the underlying security financing with a risk-free bond.

[5] There are several iterative methods available to solve for C_{BS}^{-1}, including the Newton–Raphson method, the bisection method and the Brent method.

[6] Dumas, Fleming, and Whaley (1998) choose a quadratic specification due to the parabolic shape of implied volatilities in the cross-section and to favor a parsimonious model.

[7] Specifying the regression as a function of the strike price instead of moneyness leads to essentially the same results.

$$dV_t = \kappa(\bar{v} - V_t)dt + \sigma_v\sqrt{V_t}\,dW_{2,t},$$

where V_t is the spot variance, κ governs the rate at which V_t reverts to the long-run variance $\bar{v}$, σ_v is the volatility of the volatility generating process and $W_{1,t}$ and $W_{2,t}$ are Wiener processes with correlation ρ. As part of the affine class of parametric option pricing models (Duffie, Pan, and Singleton 2000), the Heston model allows for a quasi-closed form solution for the price of a European option. We numerically compute option prices under the Heston model using the Fourier-cosine series expansion method of Fang and Oosterlee (2009).

Aligned with our goal of predicting the implied volatility surface, we estimate the parameters of the Heston model by minimizing pricing errors in terms of implied volatility. That is, we first compute the model-implied option prices, and then translate them to implied volatilities to minimize the loss function. We estimate the Heston model in two different prediction exercises. In the cross-section exercise of Section 5, we estimate for each day t the structural parameters and the spot variance, $\boldsymbol{\xi}_t = (V_t, \bar{v}, \kappa, \sigma_v, \rho)$, by minimizing $\sum_{i=1}^{n}[\sigma_{i,t} - \sigma_{\mathrm{H}}(\boldsymbol{\xi}_t, S_t, K_{i,t}, \tau_{i,t}, r_t)]^2$, where $\sigma_{\mathrm{H}}(\boldsymbol{\xi}_t, S_t, K_{i,t}, \tau_{i,t}, r_t)$ are the fitted values of the Heston model. As for the option panel exercise in Section 6, we take a sequence of cross-sections over $t = 1, \ldots, T$ days, with $j = 1, \ldots, n_t$ implied volatilities $\sigma_{j,t}$, strikes $K_{j,t}$ and times to maturity $\tau_{j,t}$ for each day t; use the nonparametric estimator of Todorov (2019) for the spot variance $\hat{V}_t$ in each day; and estimate the structural parameters of the Heston model $\boldsymbol{\xi} = (\bar{v}, \kappa, \sigma_v, \rho)$ by minimizing $\sum_{t=1}^{T}\sum_{j=1}^{n_t}[\sigma_{j,t} - \sigma_{\mathrm{H}}(\boldsymbol{\xi}, \hat{V}_t, S_t, K_{j,t}, \tau_{j,t}, r_t)]^2$. In other words, in the panel exercise the structural parameters are the same over time. For both minimizations described above we use nonlinear least squares.[8]

2.4. Carr and Wu Model

Carr and Wu (2016) propose a new option pricing framework considering that institutional investors manage option positions mainly by looking at implied volatilities as opposed to the underlying asset dynamics and option prices. Instead of fully specifying the dynamics of the instantaneous variance of the underlying asset, they model the near-term dynamics of the implied volatility surface, deriving no-arbitrage restrictions directly on its shape. Under the assumptions of the model, the implied volatility surface can be obtained by solving a simple quadratic equation.

More specifically, for an option with strike K and time to maturity τ, Carr and Wu (2016) consider the following specification for the dynamics of the underlying asset price S_t and the option implied volatility $\sigma_t(K, \tau)$ under the risk-neutral measure, respectively:

$$\begin{aligned} dS_t/S_t &= \sqrt{v_t}dW_t, \\ d\sigma_t(K,\tau)/\sigma_t(K,\tau) &= e^{-\eta_t\tau}(m_t dt + w_t dZ_t), \end{aligned} \tag{6}$$

where v_t is the instantaneous variance of the underlying asset log-returns, m_t is the average drift of the implied volatility, w_t is the volatility of the implied volatility and $e^{-\eta_t}$ is a dampening parameter that accommodates the empirical observation that implied volatilities for long-dated options tend to move less. The Wiener processes W_t and Z_t have correlation ρ_t, where ρ_t is a stochastic process that takes values from the interval $[-1, 1]$. In addition, m_t, w_t and η_t are stochastic processes independent of K, τ and $\sigma_t(K, \tau)$.

Imposing dynamic no-arbitrage restrictions, they then show that the squared implied volatility $\sigma_t^2(k, \tau)$ depends on S_t through the relative strike $k = \ln(K/S_t)$ and must satisfy the following quadratic equation:

$$\begin{aligned} &1/4e^{-2\eta_t\tau}w_t^2\tau^2\sigma_t^4 + (1 - 2e^{-\eta_t\tau}m_t\tau - e^{-\eta_t\tau}w_t\rho_t\sqrt{v_t}\tau)\sigma_t^2 \\ &\quad - (v_t + 2e^{-\eta_t\tau}w_t\rho_t\sqrt{v_t}k + e^{-2\eta_t\tau}w_t^2k^2) = 0. \end{aligned} \tag{7}$$

An important feature of the solution to (7) is that the no-arbitrage constraint depends on the current levels of the five dynamic processes $(v_t, m_t, w_t, \eta_t, \rho_t)$ but does not depend directly on the exact dynamics of these processes. Hence, one can fit the implied volatility surface on a given day t by treating the levels of the dynamic states at t as parameters. This results in a parametric family $\sigma_{\mathrm{CW}}^2(\boldsymbol{\theta}_t, k, \tau)$, where $\boldsymbol{\theta}_t = (v_t, m_t, w_t, \eta_t, \rho_t)$. Given the cross-section of options on day t with implied volatility $\sigma_{i,t}$, relative strike $k_{i,t}$ and time to maturity $\tau_{i,t}$, we estimate $\boldsymbol{\theta}_t$ by minimizing the following equation via nonlinear least squares:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta}_t} \sum_{i=1}^{n} \Big[&1/4e^{-2\eta_t\tau_{i,t}}w_t^2\tau_{i,t}^2\sigma_{i,t}^4 + (1 - 2e^{-\eta_t\tau_{i,t}}m_t\tau_{i,t} \\ &- e^{-\eta_t\tau_{i,t}}w_t\rho_t\sqrt{v_t}\tau_{i,t})\sigma_{i,t}^2 - (v_t + 2e^{-\eta_t\tau_{i,t}}w_t\rho_t\sqrt{v_t}k_{i,t} \\ &+ e^{-2\eta_t\tau_{i,t}}w_t^2k_{i,t}^2)\Big]^2. \end{aligned} \tag{8}$$

With the estimated parameters $\hat{\boldsymbol{\theta}}_t$, the implied volatility $\sigma_{\mathrm{CW}}^2(\hat{\boldsymbol{\theta}}_t, k, \tau)$ predicted by the Carr and Wu (CW) model for a given option is obtained by solving the quadratic equation (7) using $\hat{\boldsymbol{\theta}}_t$ as inputs as well as the option relative strike and time to maturity.

3. Nonparametric Neural Network Correction

3.1. A Model-Guided Nonparametric Approach

The implied volatility surface on a particular day is by definition the mapping between its cross-sectional dimensions, moneyness (m) and time to maturity (τ), and observed implied volatilities, which we denote by $\sigma(m, \tau)$. Any parametric option pricing model p can be cast in terms of its own implied mapping for the implied volatility surface $\sigma_p(m, \tau)$, with the model parameters suppressed. The simplest specification possible comes from the Black–Scholes model, which predicts a constant function $\sigma_{\mathrm{BS}}(m, \tau) = a_0$, that is, a flat implied volatility surface. While much progress has been made in the literature in developing more general parametric models capturing stylized facts in the option data, they are still misspecified in the sense of not perfectly reproducing the implied volatility surface in quantitative terms. In other words, any model p implies a pricing error surface $\epsilon_p(m, \tau) = \sigma(m, \tau) - \sigma_p(m, \tau)$.

We propose a simple two-step approach to boost the accuracy of any parametric option pricing model. Given a cross-section of

[8] Since our main goal is prediction, we do not impose the Feller condition in the estimation.

$i = 1, \ldots, n$ options on day t, we first use the parametric model p to fit the observed implied volatility surface $\sigma(m_{i,t}, \tau_{i,t})$, obtaining $\hat{\sigma}_p(m_{i,t}, \tau_{i,t})$ as fitted values and $\hat{\epsilon}_p(m_{i,t}, \tau_{i,t}) = \sigma(m_{i,t}, \tau_{i,t}) - \hat{\sigma}_p(m_{i,t}, \tau_{i,t})$ as model-implied pricing errors. Then, we estimate the pricing error surface $\epsilon_p(m, \tau)$ nonparametrically by minimizing the following objective function:

$$\frac{1}{n}\sum_{i=1}^{n}[\hat{\epsilon}_p(m_{i,t}, \tau_{i,t}) - f(m_{i,t}, \tau_{i,t})]^2, \tag{9}$$

for an appropriate space of functions $f(\cdot)$. That is, we learn the function $\hat{f}(m, \tau)$ that best approximates the pricing error surface. The implied volatility surface fitted by our method is thus given by the sum of the fitted value of the model and its nonparametric correction: $\hat{\sigma}_p(m_{i,t}, \tau_{i,t}) + \hat{f}(m_{i,t}, \tau_{i,t})$.

This approach can be treated as a generalization of a purely nonparametric estimation of the implied volatility surface. To see that, note that the following minimization problems are equivalent:[9]

$$\begin{aligned}&\min_f \frac{1}{n}\sum_{i=1}^{n}[\sigma(m_{i,t}, \tau_{i,t}) - f(m_{i,t}, \tau_{i,t})]^2 \\ &\quad\equiv \min_f \frac{1}{n}\sum_{i=1}^{n}[\sigma(m_{i,t}, \tau_{i,t}) - c - f(m_{i,t}, \tau_{i,t})]^2,\end{aligned} \tag{10}$$

where c is any constant. This implies that a direct nonparametric fit to the implied volatility surface can be seen as a correction of the BS model, which predicts a flat surface, where $c = \hat{a}_0$. Since the BS model does not provide any information about the shape of the implied volatility surface, correcting it is the same as fitting the surface directly.

The above reasoning provides motivation for our model-guided approach: if we consider a parametric model p that, in contrast to Black–Scholes, captures the main curvatures of the true implied volatility surface $\sigma(m, \tau)$, its pricing error surface $\epsilon_p(m, \tau)$ will vary less or be relatively flat compared to $\sigma(m, \tau)$, and hence easier to estimate nonparametrically. The advantages of a parametrically guided nonparametric approach over a direct nonparametric model have been theoretically documented by Glad (1998), Fan and Ullah (1999), and Fan, Wu, and Feng (2009).

3.2. Feedforward Neural Networks

Empirically, we use feedforward neural networks to nonparametrically correct the parametric models described in Section 2. We first briefly describe the model and then motivate this choice. While we focus here on the description of our method for capturing the implied volatility surface on a given day, we discuss the analogous case of fitting the option panel in Appendix A.

Feedforward neural networks consist of an input layer of explanatory variables, one or more intermediate or "hidden" layers that nonlinearly transform the variables, and a final output layer, aggregating the hidden layers to a fitted or predicted value. Letting $\mathbf{x}_{i,t} = (m_{i,t}, \tau_{i,t})' \in \mathbb{R}^2$ denote the vector containing the moneyness and time to maturity for option i on day t, the neural network model $f : \mathbb{R}^2 \to \mathbb{R}$ is constructed as follows. Starting from $\mathbf{z}_0 = \mathbf{x}_{i,t}$, define iteratively:

$$\begin{aligned}\underset{d_l\times 1}{\mathbf{z}_l} &= \overset{\circ}{h}\Big(\underset{d_l\times d_{l-1}}{\mathbf{A}_{l-1}}\ \underset{d_{l-1}\times 1}{\mathbf{z}_{l-1}} + \underset{d_l\times 1}{\mathbf{b}_{l-1}}\Big), \quad \text{for } l = 1, \ldots, L, \\ f\big(\underset{1\times 1}{\mathbf{x}_{i,t}}\big) &= \underset{1\times d_L}{\mathbf{A}_L}\ \underset{d_L\times 1}{\mathbf{z}_L} + \underset{1\times 1}{b_L},\end{aligned} \tag{11}$$

where: L is the number of hidden layers, each containing d_l output units or "neurons" stacked in vector $\mathbf{z}_l$; $\mathbf{A}_{l-1}$ is a transformation or "weight" matrix; $\mathbf{b}_{l-1}$ is the intercept or "bias" vector; and $\overset{\circ}{h} : \mathbb{R}^{d_l} \to \mathbb{R}^{d_l}$ applies a scalar "activation" function $h(\cdot)$ to each element of the vector $\mathbf{A}_{l-1}\mathbf{z}_{l-1} + \mathbf{b}_{l-1}$.[10] The neurons play the role of a basis expansion, where the unknown parameters $\{\mathbf{A}_l, \mathbf{b}_l\}_{l=0}^{L}$ defining the basis functions are estimated from the data. The activation function is usually a simple given nonlinear function such as the sigmoid $h(x) = 1/(1 + e^{-x})$ or ReLU (Rectified Linear Unit) $h(x) = \max(0, x)$. The number of hidden layers L is called the "depth" of a neural network, while the number of neurons d_l at the lth hidden layer is called the "width" of that layer.[11]

We now motivate the use of feedforward neural networks to approximate the pricing error surface of a given parametric option pricing model. Neural networks are a natural choice as they possess a "universal approximation" property, in the sense of being able to approximate any Borel measurable function with arbitrary precision, as shown by Cybenko (1989), Hornik, Stinchcombe, and White (1989), and Poggio et al. (2017), among others. In particular, motivated by the empirical success of deep learning, recent work further demonstrates that several classes of functions can be approximated by deeper networks more efficiently (i.e., with fewer neurons) than by shallower ones (Mhaskar, Liao, and Poggio 2016; Rolnick and Tegmark 2018).

The efficiency of deep learning is also related to its ability to avoid the curse of dimensionality when approximating compositional functions (Poggio et al. 2017). In the context of nonparametric regression, Bauer and Kohler (2019) show that deep neural network estimates can achieve a rate of convergence that does not depend on the number of regressors, outperforming other competing approaches in simulations. Furthermore, Fan, Ma, and Zhong (2021) discuss the practical and theoretical benefits of deep learning, which incorporates many statistical and optimization techniques such as implicit regularization and gradient descent that help avoid the curse of dimensionality in practice. This property is important to our implementations in the option panel using additional time-varying covariates.

Finally, feedforward neural networks have a long tradition in the option pricing literature due to their flexibility in modeling nonlinear relations nonparametrically (Hutchinson, Lo, and Poggio 1994; Garcia and Gençay 2000), besides being among the best performing machine learning techniques for prediction tasks in finance (see, e.g., Gu, Kelly, and Xiu 2020).

[9] Provided that we allow $f(.)$ to be composed by a constant plus other functions of moneyness and maturity, which is generally the case.

[10] Note that, for $l = 1$, the previous layer $l - 1$ is the input layer 0, where $\mathbf{z}_0 = \mathbf{x}_{i,t}$ and $d_0 = 2$.

[11] In this sense, "deep learning" refers to the use of a deep neural network, where "deep" usually means that $L \geq 2$.

3.3. Implementation Details

There are many decisions to make when implementing a neural network, such as the number of hidden layers, the number of neurons in each hidden layer and the activation functions. As discussed by Gu, Kelly, and Xiu (2020), attempting to select an optimal specification or "architecture" by searching over infinitely many combinations of architectures is in general unrealistic. Instead, we follow their approach of fixing a set of five network architectures with an increasing number of hidden layers (NN1-NN5), where the number of neurons in each layer decreases according to the geometric pyramid rule proposed by Masters (1993). With that, we can compare the performance of the five neural networks and assess the tradeoff of network depth in the nonparametric correction of the option pricing models.

More specifically, the first neural network NN1 has one hidden layer of 32 neurons; NN2 has two hidden layers with 32 and 16 neurons, respectively; NN3 has three hidden layers with 32, 16, and 8 neurons, respectively; NN4 has four hidden layers with 32, 16, 8, and 4 neurons, respectively; and NN5 has five hidden layers with 32, 16, 8, 4, and 2 neurons, respectively. That is, NN1 can be written as

$$\underset{1\times 1}{f\left(\mathbf{x}_{i,t}\right)} = \underset{1\times 32}{\mathbf{A}_1}\ \overset{\circ}{h}\Big(\underset{32\times 2}{\mathbf{A}_0}\ \underset{2\times 1}{\mathbf{z}_0} + \underset{32\times 1}{\mathbf{b}_0}\Big) + \underset{1\times 1}{b_1}; \tag{12}$$

NN2 can be written as

$$\underset{1\times 1}{f\left(\mathbf{x}_{i,t}\right)} = \underset{1\times 16}{\mathbf{A}_2}\ \overset{\circ}{h}\Big[\underset{16\times 32}{\mathbf{A}_1}\ \overset{\circ}{h}\Big(\underset{32\times 2}{\mathbf{A}_0}\ \underset{2\times 1}{\mathbf{z}_0} + \underset{32\times 1}{\mathbf{b}_0}\Big) + \underset{16\times 1}{\mathbf{b}_1}\Big] + \underset{1\times 1}{b_2}; \tag{13}$$

and so on.

For the nonlinear activation function, we choose the sigmoid function and use it for all neurons. We estimate the neural network parameters for the correction of a given parametric model by minimizing (9) with $f(m_{i,t}, \tau_{i,t})$ given by (11). The minimization is usually done via gradient descent algorithms, which start with an initial value and then iteratively update the parameters in the direction in which the loss function decreases most rapidly: the negative gradient. The gradient is computed via backpropagation, performing calculations through the network in a backward fashion based on the chain rule. In particular, we adopt the scaled conjugate gradient algorithm of Moller (1993), which is fully-automated and includes no critical user-dependent hyperparameters.[12] As one could still try to optimize over different choices of network architecture, activation function and optimization algorithm, our results can be seen as a lower bound on the performance of our model-guided approach coming from sensible specifications that are easily reproducible.

4. Implied Volatility Data

We consider European-style S&P 500 options traded at the Chicago Board Options Exchange (CBOE). These options are among the most actively traded derivatives in the world and have been the main focus of the option pricing literature.[13] The data is obtained from OptionMetrics, for the sample ranging from January 4, 2016 to June 28, 2019. We take the midpoint of the end-of-day bid and ask quotes as the option prices and apply standard filters to the data.[14] The time to expiration of each option is calculated taking into account if the contract settlement is at the open or close.[15] Since the S&P 500 index typically pays a dividend, for each day and available time to maturity we estimate the dividend yield from the put-call parity relation using the pair of call and put options that are closest to at-the-money (ATM).[16] Finally, as a proxy for the risk-free rate we download the 3-month treasury bill rate from the St. Louis Federal Reserve Economic Data (FRED) database, while data for the S&P 500 index is obtained from Bloomberg.

We focus on options with time to expiration between 20 and 240 calendar days and moneyness ($m_{i,t} \equiv S_t/K_{i,t}$) between 0.80 and 1.60, where strike price intervals are 5 and 25 points. We follow the common practice of using only out-of-the-money (OTM) options, which are considerably more liquid and reliable than their in-the-money (ITM) counterparts.[17] In fact, given the put-call parity, ITM option prices can be discarded without any loss of information. From the prices of OTM options, we back out the corresponding implied volatilities by inverting the Black–Scholes formula. Given the well-known fact that implied volatility varies substantially across moneyness and maturity, we divide the option data into different categories. With respect to moneyness, we group options according to five intervals: deep OTM call (DOTMC) if $m_{i,t} \in [0.80, 0.90)$, OTM call (OTMC) if $m_{i,t} \in [0.90, 0.97)$, ATM if $m_{i,t} \in [0.97, 1.03)$, OTM put (OTMP) if $m_{i,t} \in [1.03, 1.10)$ and deep OTM put (DOTMP) if $m_{i,t} \in [1.10, 1.60]$. As for maturities, we classify options as short-term ($[20, 60]$ days to expiration) and long-term ($(60, 240]$ days to expiration). Our final sample consists of 1,267,191 observations, with an average of 1443 options in the cross-section per trading day.

Table 1 depicts the summary statistics of our S&P 500 implied volatility data. In particular, we report the total number, mean and standard deviation of implied volatilities in each of the moneyness and maturity categories. As can be seen, short-term options represent 67.56% of the sample, with ATM options, DOTM puts and OTM puts being the most actively traded. Average implied volatilities present the usual U-shaped pattern or smile across moneyness, decreasing from DOTM puts to OTM calls and then increasing to DOTM calls. The smile is more pronounced for short-term options, but is also clearly evident for the long-term category. This is evidence of the severe misspecification of the Black–Scholes model. Implied volatilities tend to be more dispersed around the mean for options with

[12] We use the implementation of this method available in Matlab, with its default values.

[13] See, for instance, Hutchinson, Lo, and Poggio (1994), Rubinstein (1994), Bakshi, Cao, and Chen (1997), Ait-Sahalia and Lo (1998), Dumas, Fleming, and Whaley (1998), Duffie, Pan, and Singleton (2000), Garcia and Gençay (2000), Ait-Sahalia and Duarte (2003), Fan and Mancini (2009), Andersen, Fusari, and Todorov (2015), Carr and Wu (2016), and Ait-Sahalia, Li, and Li (2021).

[14] More specifically, observations with zero volume, with price lower than 1/8 or violating the usual no-arbitrage conditions are dropped.

[15] For PM settled options, the time to expiration is the number of days between the trade date and the expiration date, while for AM settled options, it is the number of days between the dates less one.

[16] For the rare case that there is no such pair of ATM call and put options for a given time to maturity, we use the dividend yield provided by OptionMetrics.

[17] See, for instance, Ait-Sahalia and Lo (1998), Fan and Mancini (2009), and Andersen, Fusari, and Todorov (2015).

Table 1. Summary statistics of S&P 500 implied volatility data.

	Number		Mean IV		Std. dev.	
Time to maturity Moneyness	Short	Long	Short	Long	Short	Long
[0.80, 0.90)	8,813	14,691	16.72	12.44	3.44	2.43
[0.90, 0.97)	135,224	70,079	11.17	10.92	3.41	2.72
[0.97, 1.03)	250,283	81,325	11.65	13.13	4.08	3.32
[1.03, 1.10)	222,245	77,681	17.50	17.13	4.00	3.11
[1.10, 1.60]	239,514	167,336	29.28	25.87	8.31	5.45
Total	856,079	411,112	18.08	18.67	9.23	7.54

NOTE: This table presents summary statistics of our S&P 500 index implied volatility data. The sample ranges from January 4, 2016 to June 28, 2019. The columns Number, Mean IV and Std. dev. report the number of options, the average implied volatility (in %) and the standard deviation of the implied volatilities (in %), respectively, for each moneyness ($S_t/K_{i,t}$) and time to maturity category.

time to maturity between 20 and 60 days, especially for DOTM puts.

5. Prediction in the Option Cross-Section

In this section, we conduct prediction exercises in the option cross-section based on a day-by-day model estimation, which is common practice in the option pricing literature.[18] In fact, the AHBS and CW models are designed to fit the implied volatility surface on a given day. While the daily re-estimation approach is theoretically inconsistent with structural models such as Black and Scholes (1973) and Heston (1993), which assume that parameters are constant over time, these models are often also implemented on a day-by-day basis (see Bakshi, Cao, and Chen 1997). Therefore, it is informative to analyze whether neural networks can improve performance in this setting. In Section 6, we consider the Heston model estimation in the option panel and the corresponding nonparametric correction.

5.1. Implementation

We consider two types of prediction exercises in the cross-section. In the first exercise, for each day t in our sample, we split the option data into training and testing sets according to the strike price: the models are trained on options with strikes divisible by 10 and tested on options with strikes that are not divisible by 10. Such procedure tends to allocate roughly 60% of the data to training and the remaining to testing while guaranteeing that options along the whole spectrum of moneyness and time to maturity are represented in both datasets. This can be seen as a same-day interpolation exercise, where we test the performance of models in accurately capturing the implied volatility surface after learning from a thicker grid. This would mimic the scenario where one observes only a fraction of the option cross-section and seeks to use this information to price other unobserved options on the same day t.

In the second exercise, for each day t in our sample, we train the models on all options available on day t and test their performance in predicting the implied volatility surface $t + h$ business days ahead. We consider three different horizons: one-day ($h = 1$), one-week ($h = 5$) and one-month ahead ($h = 21$). It is worth emphasizing that model parameters are estimated at time t and kept fixed to predict implied volatilities at $t + h$, such that their dynamics is not modeled. Rather, the goal of this exercise is to assess the stability of each parametric model and its correction over time, that is, to what extent the model estimated at t generalizes to cross-sections at future dates. In fact, incorporating dynamics in the AHBS or CW would require an additional step of specifying a time series model for their parameters.[19] By keeping the model fixed, we can isolate and focus on the effect that model correction has on performance. While Heston could take dynamics into account by updating the spot variance $\hat{V}_t$ with its expectation for day $t + h$, we maintain $\hat{V}_t$ fixed in the prediction to be consistent with the implementation of the other models.[20]

For both exercises, we evaluate models based on the out-of-sample implied volatility root mean squared error (IVRMSE), which is calculated using the prediction errors pooled over the testing samples. We consider this error metric as it is defined in terms of implied volatilities, making it easily interpretable, comparable across different option categories and aligned with the estimation of the models and our implied volatility prediction analysis. We have also tested the mean absolute deviation of the prediction errors, but these results are omitted for brevity as they are qualitatively similar.

We investigate the performance of two sets of models. The first set consists of the parametric option pricing models, namely BS, AHBS, Heston, and CW. The second set is comprised of the five neural networks (NN1-NN5) correcting each of the parametric models as a function of moneyness and maturity, the cross-sectional dimensions indexing the implied volatility surface. With that, our main goal is to answer whether neural networks can successfully correct and improve upon the parametric models by learning their pricing error surface. Secondary but also important questions we ask include which parametric model performs best before and after the nonparametric correction and what is the tradeoff of network depth in our prediction problem.

5.2. Empirical Results

Table 2 reports the IVRMSE of each model for the four prediction exercises in the option cross-section. As can be seen across all exercises, the neural network correcting the parametric model always outperforms the model by itself, usually by a significant amount. Improvements are greater for the same-day exercise and can be as large as 3457% and 850% for the NN3 correcting the BS and CW model, respectively. The relative enhancement of the nonparametric correction decreases with the forecast horizon. Taking CW+NN3 and CW as an example, the improvement is 71%, 24%, and 15% for 1-, 5- and 21-day ahead predictions, respectively. This is expected as

[18] See, among others, Bakshi, Cao, and Chen (1997), Dumas, Fleming, and Whaley (1998), and Fan and Mancini (2009).

[19] For instance, Gonçalves and Guidolin (2006) propose a vector autoregressive approach to model the dynamics of the parameters of the AHBS model and predict their values in the future.

[20] In unreported results available upon request, we also test Heston using the model-implied expectation at t of future variance at $t+h$, $\mathbb{E}_t[V_{t+h}] = \hat{\bar{v}} + e^{-\hat{\kappa}h/252}(\hat{V}_t - \hat{\bar{v}})$, instead of $\hat{V}_t$. We find that the predictive performance improves only marginally, without changing the qualitative nature of our results.

Table 2. IVRMSE of prediction in the option cross-section.

	No NN	NN1	NN2	NN3	NN4	NN5	No NN	NN1	NN2	NN3	NN4	NN5
	Panel A: Same-day						Panel B: 1-day ahead					
BS	7.47	0.36	0.23	0.21	0.20	0.23	8.17	1.23	1.19	1.19	1.18	1.19
AHBS	1.35	0.22	0.17	0.16	0.16	0.20	1.92	1.18	1.18	**1.17**	1.18	1.19
Heston	0.90	0.24	0.16	**0.15**	**0.15**	0.21	1.49	1.18	**1.17**	**1.17**	**1.17**	1.18
CW	1.52	0.22	0.17	0.16	0.16	0.19	2.00	1.18	**1.17**	**1.17**	**1.17**	1.18
	Panel C: 5-day ahead						Panel D: 21-day ahead					
BS	8.32	2.26	2.24	2.23	2.25	2.23	8.62	3.40	3.36	3.64	3.32	3.33
AHBS	2.75	2.22	2.22	2.22	2.22	2.23	3.79	3.32	3.33	3.29	3.29	3.30
Heston	2.39	2.23	2.23	2.23	2.22	2.23	3.41	3.34	3.33	3.35	3.32	3.33
CW	2.76	**2.21**	**2.21**	2.22	**2.21**	2.21	3.75	3.30	3.28	**3.27**	**3.27**	3.29

NOTE: This table presents the IVRMSE (in %) of each model for the same-day and h-day ahead (h = 1, 5, 21) prediction exercises as indicated by the panels. The rows indicate the parametric option pricing model, while the columns refer to the neural network architecture (NN1-NN5) or to the case of no model correction (No NN). Bold numbers indicate the best performing model (or models) in a given panel. The sample ranges from January 4, 2016 to June 28, 2019.

the implied volatility surface changes over time, such that the predictive signal is weaker for longer horizons. Moreover, the neural network correction brings the error down to a similar magnitude independent of the original parametric model and its misspecification, which is robust across all exercises and network architectures.

Comparing performance among corrected models, we can see that correcting the BS model is always outperformed by correcting less misspecified parametric models, again across all exercises and network depths. The difference is reasonably large for the same-day and 21-day ahead exercises, and negligible for 1-day and 5-day ahead predictions. For instance, CW+NN3 performs 32% and 12% better than BS+NN3 in panels A and D, respectively, but only 2% and 0.5% better in panels B and C, respectively. Since correcting the BS model is equivalent to employing a standalone neural network,[21] these results show the additional gains that can be obtained by fitting a flattened pricing error surface instead of the implied volatility surface directly. In fact, the best performing model is always among the corrected AHBS, Heston, and CW models.

Table 3 provides a closer look at performance across option categories, focusing for brevity sake on the same-day and 1-day ahead predictions and neural networks with three hidden layers.[22] Interestingly, the nonparametric correction brings the error down to the same magnitude for each option category, irrespective of the misspecification of the original model for that category. As a result, corrected models do not display significant biases across moneyness and maturity. To illustrate this, Figure 1 compares the implied volatility smiles predicted by the parametric models and the corrected models for a representative date of the sample. While there are clear differences between the shapes of the implied volatility surface generated by each model, it is hard to distinguish between the surfaces implied by the corrected models, which overall provide a better and more balanced fit.

To put these results in perspective, it is useful to think of a neural network as a universal approximator for an unknown function, which, in our case, is the pricing error surface of a given parametric model. Perhaps not surprisingly, we see that for the same-day prediction exercise, which resembles an interpolation, the neural networks are very successful in approximating these surfaces from a thicker grid and generate huge gains compared to parametric models. More surprisingly so, for h-day ahead prediction exercises where the surfaces are changing over time, neural network corrections still lead to considerably large improvements. This suggests that there is a strong signal about future implied volatilities in the option data at time t that is captured by the complex nonlinearities embedded in neural networks but missed by parametric models. Furthermore, we find that neural networks are flexible enough to correct different parametric models to a similar degree. Even so, correcting models with flatter pricing error surfaces that are easier to estimate nonparametrically can still lead to additional improvements.

Turning the attention to the parametric option pricing models, Tables 2 and 3 make clear the well-known severe misspecification of the BS model, which predicts a flat implied volatility surface. The remaining parametric models perform orders of magnitude better, as expected. In particular, the Heston model consistently beats the AHBS and CW models and provides a more balanced performance across option categories. For instance, AHBS and CW perform poorly for DOTM calls as they fail to generate the U-shaped pattern of the implied volatility smile, which the Heston model is able to reproduce (see Figure 1). However, this does not imply that correcting the Heston model always leads to the best prediction among the corrected models. For 21-day ahead predictions, correcting the CW model sightly outperforms the correction of the Heston model. This indicates that, for longer horizons where the predictive signal is weaker, correcting the least misspecified model can lead to small overfitting.

A final comment applies to the tradeoffs of network depth. Even though there are no significant differences in magnitude between the pricing errors of the considered network architectures, we find that neural networks with three or four hidden layers tend to perform slightly better in predicting option prices. While one hidden layer may not be enough to fully capture the nonlinear features of the implied volatility surface, five hidden layers may lead to overfitting. This is in line with neural network applications in the finance literature such as Gu, Kelly, and Xiu (2020), who find that the performance of a standalone neural network in forecasting stock returns peaks at a specification with three hidden layers.

In sum, we find strong empirical support for our proposed methodology of nonparametrically correcting and boosting option pricing models to predict the implied volatility surface. The corrected models outperform the original ones by a large extent, for all option categories and across predictions of different horizons. The correction by the neural network is relatively indiscriminate, bringing pricing errors down to a similar magnitude regardless of the parametric model misspecification. Even so, correcting less misspecified models usually leads to additional improvements in performance, highlighting the usefulness of a model-guided neural network approach.

[21] As a sanity check, we also conducted our tests with a standalone neural network and the results are identical to those of correcting the BS model. To avoid redundancy and save space, we omit these results.

[22] Results for other network architectures and 5- and 21-day ahead predictions are qualitatively similar.

Table 3. IVRMSE of prediction in the option cross-section across categories.

	DOTMC	OTMC	ATM	OTMP	DOTMP	Short	Long	DOTMC	OTMC	ATM	OTMP	DOTMP	Short	Long
	Panel A: Same-day							Panel B: 1-day ahead						
BS	8.95	8.59	6.67	2.32	10.03	7.68	6.94	7.81	8.00	6.16	2.33	11.72	8.57	7.27
BS + NN3	0.30	0.15	0.17	0.12	0.32	0.18	0.28	1.20	**1.03**	**1.16**	**1.21**	1.26	1.30	0.90
AHBS	3.55	1.66	1.32	0.62	1.46	1.40	1.20	5.13	2.23	1.58	1.30	2.04	2.02	1.68
AHBS + NN3	**0.29**	**0.12**	0.15	**0.09**	0.22	**0.15**	0.17	1.27	**1.03**	**1.16**	**1.21**	1.22	**1.29**	0.88
Heston	1.25	0.69	0.88	0.72	1.12	0.97	0.69	1.89	1.19	1.39	1.37	1.74	1.65	1.07
Heston + NN3	0.30	0.15	**0.13**	**0.09**	**0.18**	**0.15**	**0.13**	1.36	1.04	**1.16**	1.22	**1.21**	1.30	**0.85**
CW	2.94	1.63	1.30	0.89	1.96	1.31	1.94	3.89	2.12	1.60	1.45	2.38	1.92	2.15
CW + NN3	**0.29**	**0.12**	0.14	**0.09**	0.21	**0.15**	0.17	**1.15**	**1.03**	**1.16**	**1.21**	1.22	**1.29**	0.87

NOTE: This table presents the IVRMSE (in %) of each model for specific moneyness and time to maturity categories for the same-day and 1-day ahead prediction exercises as indicated by the panels. The rows indicate the model, while the columns refer to the moneyness or maturity category. Bold numbers indicate the best performing model (or models) in a given column. The sample ranges from January 4, 2016 to June 28, 2019.

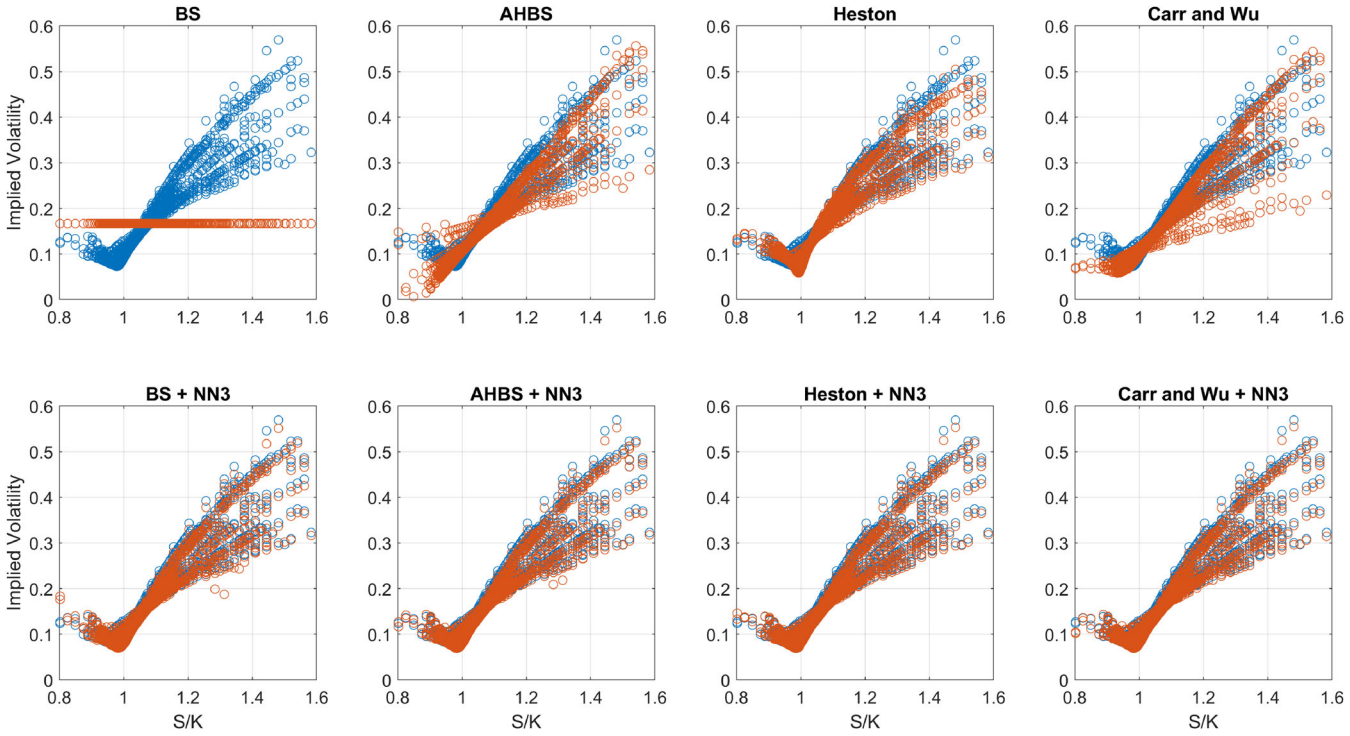

Figure 1. Implied volatility surface—1-day ahead prediction. This figure plots, in each panel, the implied volatility smiles across moneyness observed from the data (in blue) and predicted by a model (in red) for all times to maturity on September 17, 2018. The prediction is based on the information available from the previous day. For interpretation of the references to color in this figure, see the online version of this article.

6. Prediction in the Option Panel

As previously mentioned, even though we can use the Heston model on a day-by-day basis to fit the implied volatility surface, the daily re-estimation is theoretically inconsistent with such structural models. In this section, we follow a similar approach as Andersen, Fusari, and Todorov (2015) in fitting the Heston model in an option panel, that is, in a sequence of implied volatility surfaces. We are interested in assessing the misspecification of the Heston model and investigating the extent to which neural networks can correct it. In particular, given the option panel structure, we can test whether incorporating information from time-varying observable covariates in the neural network estimation helps to learn the "shape" of such misspecification as a function of different states of the economy.[23]

6.1. Implementation

For this exercise, we consider as the training set the first two years of our sample, while the remaining sample constitutes the testing set.[24] The inputs of the Heston model consist of the parameters $\bar{v}$, κ, σ_v, and ρ, which are fixed over time, and the state variable V_t representing the spot variance for each day t. We proceed by first setting V_t equal to the spot variance estimator of Todorov (2019), for which daily estimates are made available by Torben Andersen and Viktor Todorov on their website.[25] In other words, V_t is regarded as a daily observable variable. Then, we estimate the fixed parameters using the whole training panel dataset as described in Section 2.3. When evaluating the Heston model in the out-of-sample testing set, we keep

[23] We thank an anonymous referee for this suggestion.

[24] The in-sample training set consists of 604,749 options over 503 days, while the out-of-sample testing set consists of 662,442 options over 375 days.

[25] *https://tailindex.com/index.html*.

the fixed parameters estimated in-sample and use the available estimates of the spot variance for the corresponding dates.

We train the neural network on the pricing errors of the fitted Heston model pooled over the whole in-sample option panel, as described in Appendix A. Given the results from the previous section that architectures with three hidden layers usually work better, we consider this specification for our panel data analysis.[26] To assess the information content of observable covariates, we fit two different neural networks. The first one (NN3) contains as features only the cross-sectional dimensions of the implied volatility surface, moneyness and time to maturity, as in the previous section. The second one (NN3F) takes advantage of the option panel structure by including a set of time-varying features that can potentially provide useful conditioning information. For comparison, we also fit the neural networks correcting the BS model, which is equivalent to fitting them directly on the implied volatility surfaces, since the BS model generates a constant implied volatility over time, moneyness and maturity.[27]

We consider a set of eight features observed daily that can be relevant for the problem at hand. The first is the VIX index from CBOE, which is a popular benchmark for market risk and sentiment. Since the VIX captures two distinct types of risk, namely diffusive and jump risk, Bollerslev, Todorov, and Xu (2015) propose a method to construct a separate measure of market jump risk. Based on that, we obtain from Torben Andersen and Viktor Todorov's website the Left Tail Volatility (LTV), which estimates the expected return volatility that stems from large negative price jumps, and the Left Tail Probability (LTP), which estimates the probability that the S&P 500 index drops by 10% or more over the subsequent week. These measures are derived from securities priced under the risk-neutral measure. To include a risk variable under the physical measure, we construct the realized volatility (RVOL) for each day using high-frequency 5-minute S&P 500 returns from Tick Data.

We also include measures of uncertainty and macroeconomic conditions. We obtain the Economic Policy Uncertainty (EPU) index of Baker, Bloom, and Davis (2016) from *https://www.policyuncertainty.com/* and the Aruoba, Diebold, and Scotti (2009) (ADS) business conditions index from the Federal Reserve Bank of Philadelphia. Finally, we include the time series of the first differences of the term spread (TMS) and first differences of the credit spread (CRS), where both features are downloaded from the FRED database. The term spread is the difference between the 10-year Treasury rate and 3-month Treasury rate while the credit spread is the difference between the Moody's Seasoned Baa Corporate Bond yield and 10-year Treasury yield.

To identify the covariates that play an important role in capturing implied volatility surfaces and correcting the Heston model, we use a simple notion of feature importance in the spirit of Gu, Kelly, and Xiu (2020). The importance of feature j (FI_j) is defined as the increase in the IVRMSE arising from setting all values of feature j to zero, while keeping the remaining model estimates and features the same. If a feature is negligible, FI_j should be close to zero. In contrast, FI_j should be high for covariates that help predict implied volatilities.

Table 4. Prediction in the option panel.

	Heston	Heston + NN3	Heston + NN3F	BS	BS + NN3	BS + NN3F
			Panel A: IVRMSE			
All	2.47	1.87	**1.39**	9.20	3.97	1.49
DOTMC	**1.69**	1.78	1.78	4.29	2.41	2.57
OTMC	1.73	1.78	**1.17**	5.94	3.56	1.68
ATM	2.63	2.00	**1.25**	5.42	4.61	1.31
OTMP	1.69	1.69	1.41	4.16	4.14	**1.32**
DOTMP	3.11	1.93	1.56	14.48	3.58	**1.52**
Short	2.78	1.99	**1.42**	9.90	4.37	1.52
Long	1.70	1.61	**1.33**	7.67	3.05	1.42
			Panel B: IV surface characteristics RMSE			
Level	2.61	2.03	**0.80**	6.33	4.37	0.94
TS	0.89	**0.59**	0.61	1.58	1.29	0.72
Skew	1.41	1.22	1.05	13.35	1.82	**0.98**
Skew TS	1.04	0.80	0.73	1.76	1.01	**0.72**

NOTE: This table presents results for the prediction exercise in the option panel for each model as indicated by the columns. Panel A depicts the out-of-sample IVRMSE of each model considering all options (All) and options within different moneyness and maturity categories. Panel B depicts the RMSE over the whole sample for the Level, Term Structure (TS), Skew and Skew Term Structure (Skew TS) of the implied volatility surface (see footnote 28 for definitions). Bold numbers indicate the best performing model (or models) in a given row. The sample ranges from January 4, 2016 to June 28, 2019, where the in-sample period ranges from January 4, 2016 to December 29, 2017 and the out-of-sample period from January 2, 2018 to June 28, 2019.

We evaluate the ability of the models in fitting the option panel both by analyzing their pricing errors, in terms of the IVRMSE, and by investigating to what extent they can reproduce salient characteristics of the implied volatility surface over time, which consist of the level, term structure, skew and skew term structure of the surface.[28] As discussed in Andersen, Fusari, and Todorov (2015), these shape characteristics help summarize the nonlinear dynamics of the implied volatility surface and can be useful to assess model fit, even more so because the models are not directly trained to explain the time variation in the characteristics.[29] We compute the root mean squared error (RMSE) between the model-implied and data-implied option characteristics.

6.2. Empirical Results

Table 4 contains the results for the prediction exercise in the option panel. Focusing on the first row of Panel A, we can see that our nonparametric correction leads to considerably large improvements in out-of-sample pricing performance compared to the structural option pricing models BS and Heston. In particular, exploiting the panel structure by including time-varying covariates in the neural network estimation further reduces

[26] That is, we use neural networks with three hidden layers with 32, 16, and 8 neurons, respectively.

[27] The BS model is estimated as the constant implied volatility that minimizes the IVMSE in the in-sample option panel. The predictions of the BS model out-of-sample for any option are simply the estimated constant implied volatility.

[28] More specifically, the level of the implied volatility surface is the average implied volatility for short-term ATM options, the term structure is the difference between the average implied volatility of long- and short-term ATM options, the skew is defined as the difference between the average implied volatility of short-term OTM put and OTM call options, and the skew term structure is the difference between the long- and short-term skew, where the long-term skew is defined analogously to the short-term skew.

[29] More recently, Ait-Sahalia, Li, and Li (2021) propose stochastic volatility models designed to fit directly the shape characteristics of the implied volatility surface.

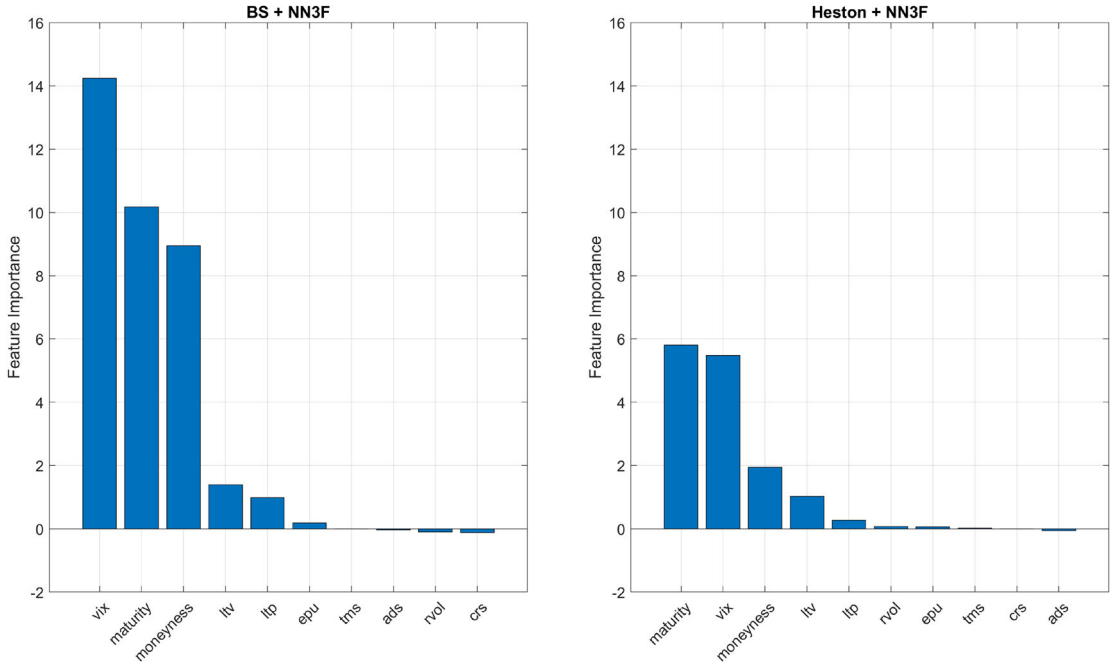

Figure 2. Prediction in the option panel—Feature importance. This figure reports the importance of each feature in the neural network correcting the BS model and correcting the Heston model in the option panel prediction exercise. The importance of feature *j* is the increase in the out-of-sample IVRMSE (in %) arising from setting all values of feature *j* to zero, while keeping the remaining model estimates and features the same. The sample ranges from January 4, 2016 to June 28, 2019, where the in-sample period ranges from January 4, 2016 to December 29, 2017 and the out-of-sample period from January 2, 2018 to June 28, 2019.

pricing errors and makes corrected models more comparable. For instance, while Heston+NN3 performs 112% better than BS+NN3, Heston+NN3F outperforms BS+NN3F only by 7%. The remaining rows of Panel A show similar improvements when looking at specific option categories, where the greatest enhancement from including the covariates is seen for ATM options.

Focusing now on Panel B of Table 4, the neural networks improve considerably upon the parametric models in capturing the dynamics of the option panel, even though they were not directly trained to do so. The inclusion of time-varying covariates as features further enhances the explanatory power for the option characteristics, especially for the level of the implied volatility surface. Naturally, the relative improvements of the nonparametric correction are larger for the BS model, but are also substantial for the Heston model. Despite the fact that the misspecification of the original models differs significantly, the performance of the models corrected by NN3F is similar.

The results above are fairly intuitive. In the option panel, the nonparametric correction solely as a function of moneyness and time to maturity does not differentiate between distinct days in the sample, thus, only being able to learn the "average" misspecification of the parametric model across days. In contrast, the inclusion of time-varying covariates in the estimation provides valuable conditioning information: the neural network can now learn how the parametric model is misspecified in reproducing the implied volatility surface as a function of the state of the economy, as measured by the covariates (e.g., when the VIX is relatively high or low). In the former case, the misspecification of the parametric model matters more, while in the latter, the neural network is flexible enough to capture the time-varying "shape" of the misspecification to a similar extent regardless of the original bias. This explains why Heston+NN3 performs so much better than BS+NN3, while BS+NN3F is comparable to Heston+NN3F.[30]

We analyze the feature importance in the correction of the parametric models in Figure 2. Remarkably, the VIX index is the most important time-varying covariate by a large difference, doing justice to its status as a benchmark measure of market risk and sentiment. For the correction of the BS model, the VIX is even more important than moneyness and maturity, highlighting the key role that conditioning information plays in fitting the option panel. LTV and LTP are also relevant, indicating that separate measures of market jump risk help in identifying the time-varying shape of the misspecification of the parametric models. The remaining covariates are negligible, suggesting that macroeconomic conditions are not very informative for correcting the models. In particular, RVOL, which is a risk variable under the physical measure, is redundant given the risk-neutral expected volatility captured by the VIX. Moreover, we can see that features are more important for the BS model, which is more severely misspecified.

We now assess in more detail the misspecification of the Heston model, which is more realistic than BS and provides a basis for more general specifications that incorporate stochastic volatility. Figure 3 compares the data-implied and Heston model-implied option characteristics. Focusing first on the former, the observed level reflects the general level of volatility, with occasional spikes that are also visible in the other characteristics. The term structure is positive, apart from negative outliers,

[30]The same comment would apply to a standalone neural network, which is equivalent to correcting the BS model. This further highlights the usefulness of our model-guided approach.

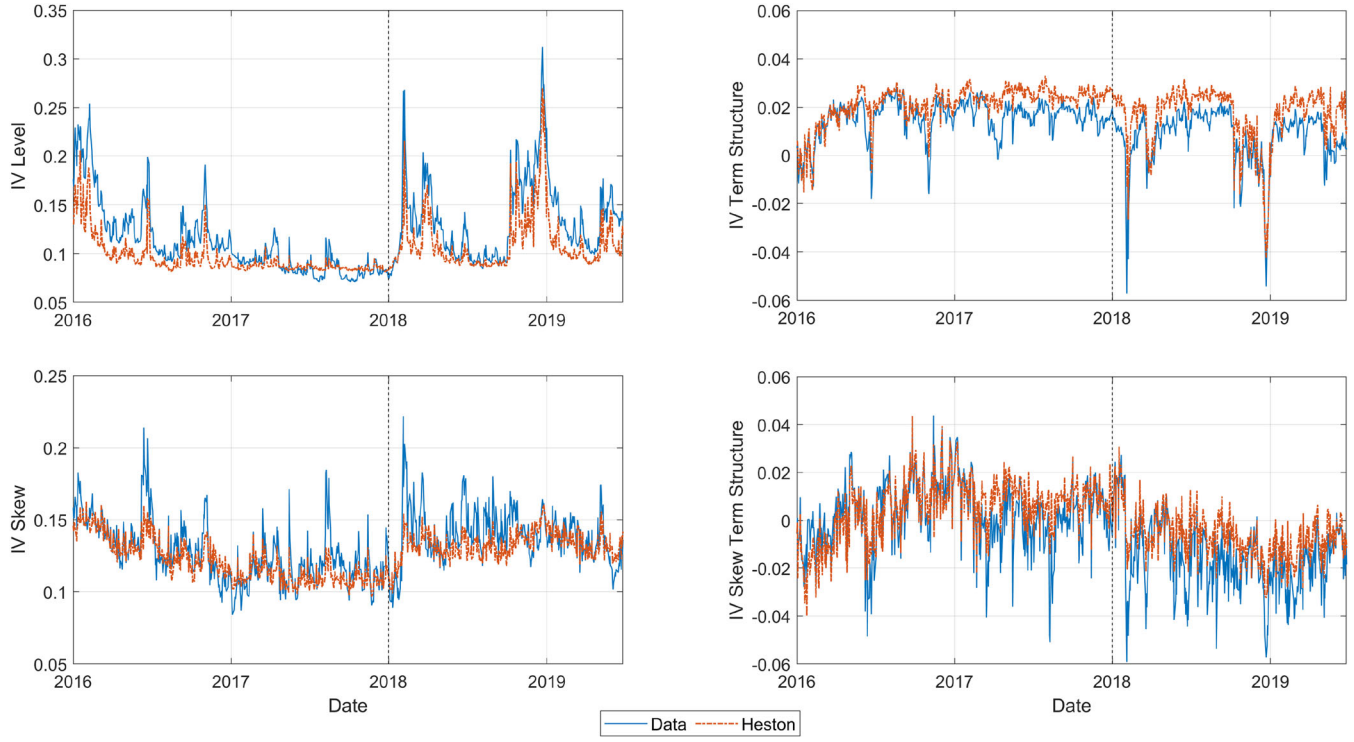

Figure 3. Implied volatility surface characteristics—Heston. This figure plots the time series of the Level, Term Structure, Skew and Skew Term Structure of the implied volatility surface implied by the data and by the Heston model. The sample ranges from January 4, 2016 to June 28, 2019, where the dashed vertical line splits the sample between the in-sample period (January 4, 2016 to December 29, 2017) and the out-of-sample period (January 2, 2018 to June 28, 2019).

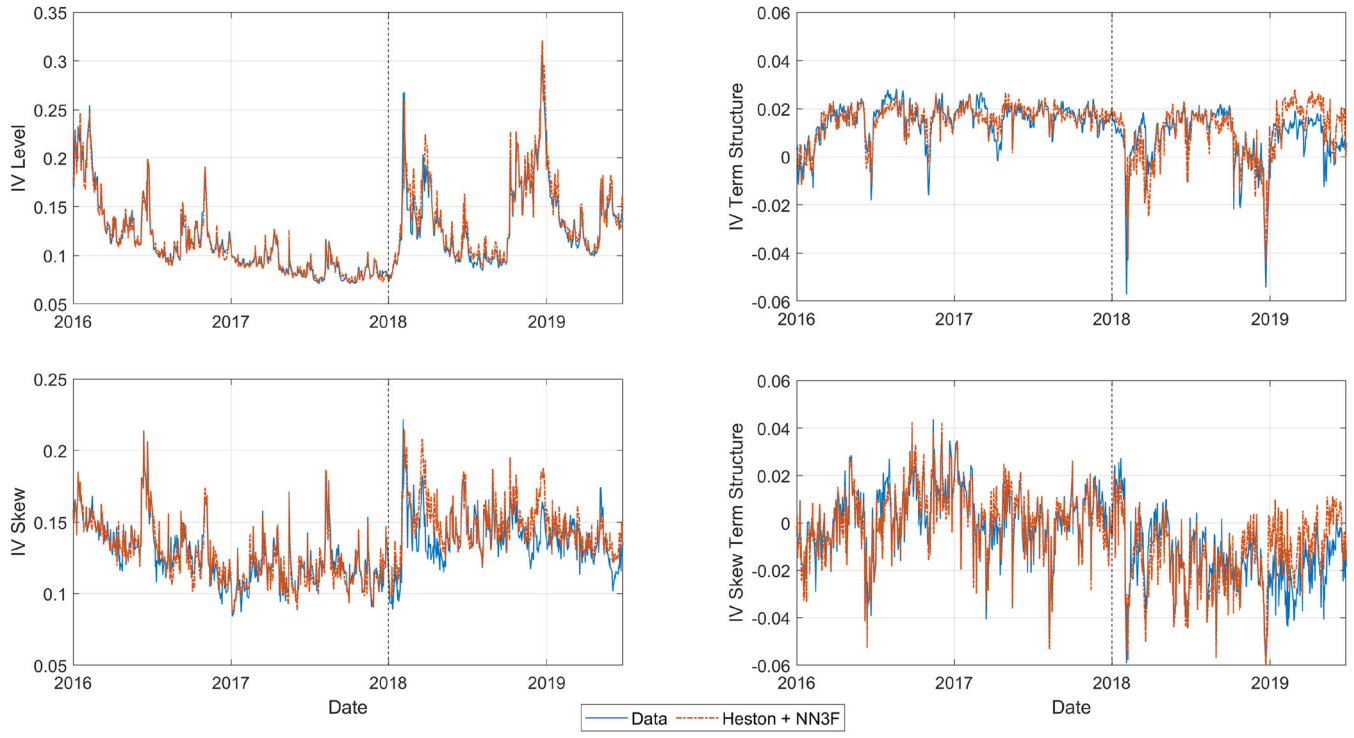

Figure 4. Implied volatility surface characteristics—Heston + NN3F. This figure plots the time series of the Level, Term Structure, Skew, and Skew Term Structure of the implied volatility surface implied by the data and by the Heston model corrected by the neural network with features. The sample ranges from January 4, 2016 to June 28, 2019, where the dashed vertical line splits the sample between the in-sample period (January 4, 2016 to December 29, 2017) and the out-of-sample period (January 2, 2018 to June 28, 2019).

while the skew is always positive and particularly large during periods of higher volatility. The skew term structure tends to be negative throughout the sample.

Figure 3 shows that the Heston model does a decent job in capturing the overall trends of the option characteristics, but often misses their levels and spikes. As discussed by Andersen,

Fusari, and Todorov (2015), additional volatility and jump factors would be necessary to more adequately capture the dynamics dependencies between the option characteristics. Alternatively, we investigate whether a nonparametric correction of the Heston model can better reproduce the time-varying patterns of the implied volatility surface. As seen in Figure 4, the corrected model provides an almost perfect fit for the level and substantial improvements for all other characteristics. As indicated by Table 4, the inclusion of the conditioning covariates is fundamental for that performance. In particular, given the importance of the VIX, LTV and LTP in Figure 2, our results confirm that information about the evolution of volatility and jump risks is instrumental for an accurate fit of the option panel.

In sum, the results in this section provide further evidence of the strength and reliability of our nonparametric correction methodology. Neural networks substantially improve structural parametric option pricing models designed to fit the option panel. The option panel structure allows one to incorporate observable time-varying covariates in the neural network estimation. This provides valuable conditioning information that helps to learn the shape of the parametric model misspecification and capture the nonlinear dynamics of the implied volatility surface. Important covariates include measures of volatility and jump risk such as the VIX, LTV and LTP. With conditioning information, neural networks correct parametric models to a similar extent, but correcting a less misspecified model tends to lead to the best performance.

7. Conclusion

In this article, we introduce an easily adaptable methodology based on boosting the accuracy of parametric option pricing models with a feedforward neural network. Our two-step approach first estimates the chosen parametric model to fit the observed implied volatility surface and then trains the neural network on the model-implied pricing errors. Using a large dataset of S&P 500 options, we test our nonparametric correction on the Black and Scholes (1973), Heston (1993), ad-hoc Black–Scholes (Dumas, Fleming, and Whaley 1998), and Carr and Wu (2016) models, considering out-of-sample prediction exercises in the cross-section for different horizons and in the option panel.

We find that the nonparametrically corrected models always outperform the original ones for all prediction exercises, and often by a large extent. The correction by the neural network is relatively indiscriminate, bringing pricing errors down to a similar magnitude regardless of the misspecification of the parametric model. Even so, correcting less misspecified models usually leads to additional improvements in performance and also outperforms a neural network fitted directly to the implied volatility surface. This highlights the usefulness of our model-guided nonparametric approach.

This article provides a number of implications for industry practices and future research. The versatility of our approach demonstrates that financial institutions relying on parametric models to price options can boost performance by employing the nonparametric correction that we propose. Moreover, our results suggest that neural networks (either correcting a parametric model or fitted directly to the implied volatility surface) should be considered as benchmarks for testing new option pricing models, as they offer a stronger test than quadratic models as in Dumas, Fleming, and Whaley (1998). Finally, our methodology can be easily adapted to correct existing frameworks in other areas of financial research such as volatility forecasting and risk management.

Appendix A: Nonparametric Correction in the Option Panel

The option panel consists of a sequence of implied volatility surfaces over $t = 1, \ldots, T$ days. As such, the mapping of interest is now a function not only of moneyness and time to maturity, but also of time: $\sigma(t, m, \tau)$. Again, a parametric option pricing model p can be cast in terms of its implied mapping for the panel $\sigma_p(t, m, \tau)$, with model parameters suppressed, where the simplest specification comes from the Black–Scholes constant volatility function: $\sigma_{BS}(t, m, \tau) = a_0$. Given any misspecified parametric model, we have the pricing error function $\epsilon_p(t, m, \tau) = \sigma(t, m, \tau) - \sigma_p(t, m, \tau)$ for the option panel.

How is our two-step approach implemented in the option panel? Given a sequence of cross-sections over $t = 1, \ldots, T$ days, with $j = 1, \ldots, n_t$ moneyness and times to maturity for each day t, we first estimate the parametric model p to fit the observed option panel $\{\sigma(t, m_{j,t}, \tau_{j,t}), j = 1, \ldots, n_t\}_{t=1}^{T}$, obtaining $\hat{\sigma}_p(t, m_{j,t}, \tau_{j,t})$ as fitted values and $\hat{\epsilon}_p(t, m_{j,t}, \tau_{j,t}) = \sigma(t, m_{j,t}, \tau_{j,t}) - \hat{\sigma}_p(t, m_{j,t}, \tau_{j,t})$ as model-implied pricing errors.[31] Then, there are two possibilities. The first is to estimate nonparametrically the pricing error function using moneyness and time to maturity as explanatory variables, as in (9), by minimizing:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{j=1}^{n_t}[\hat{\epsilon}_p(t, m_{j,t}, \tau_{j,t}) - f(m_{j,t}, \tau_{j,t})]^2, \tag{A.1}$$

where the corrected model fitted value is $\hat{\sigma}_p(t, m_{j,t}, \tau_{j,t}) + \hat{f}(m_{j,t}, \tau_{j,t})$. In this case, the nonparametric correction does not differentiate between distinct days t, thus, ignoring the time-varying dynamics of the pricing error surface and only being able to learn the "average" misspecification of the parametric model across days.

The second possibility is to include as explanatory variables in the correction time-varying covariates $\mathbf{y}_t = (y_{1,t}, \ldots, y_{q,t})$ that may be relevant to condition the misspecification of the parametric model to specific types of days. The estimation of the nonparametric correction then obtains by minimizing:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{j=1}^{n_t}[\hat{\epsilon}_p(t, m_{j,t}, \tau_{j,t}) - f(\mathbf{y}_t, m_{j,t}, \tau_{j,t})]^2, \tag{A.2}$$

with fitted values given by $\hat{\sigma}_p(t, m_{j,t}, \tau_{j,t}) + \hat{f}(\mathbf{y}_t, m_{j,t}, \tau_{j,t})$. Now the nonparametric correction can potentially learn the time-varying shape of the misspecification of the parametric model as a function of $\mathbf{y}_t$. We implement both possibilities above in the prediction exercise in the option panel in Section 6.

Note that the interpretation of our model-guided approach as a generalization of a purely nonparametric estimation also carries out to the option panel setting. This is due to the same argument that correcting the Black–Scholes constant volatility function is the same as fitting directly the panel of implied volatilities. Again, this suggests that it should be easier to estimate nonparametrically the flattened pricing error function of a parametric model that captures the main dynamics of the implied volatility surface.

[31] See, for instance, the estimation of the Heston model in the option panel described in Section 2.3, where we estimate the structural parameters $\boldsymbol{\xi} = (\bar{v}, \kappa, \sigma_v, \rho)$ which are fixed over time.

The definition of the neural network model for $f(.)$ is analogous to that laid out in Section 3.2, with the differences that now estimation is conducted in the option panel and explanatory variables in the input layer can include time-varying covariates. That is, in the first case using only moneyness and time to maturity as inputs, we have $\mathbf{x}_{i,t} = (m_{i,t}, \tau_{i,t})'$, $f(\mathbf{x}_{i,t})$ as in (11), and parameters estimated by minimizing (A.1). In the second case including q conditioning time-varying covariates, we have $\mathbf{x}_{i,t} = (\mathbf{y}_t, m_{i,t}, \tau_{i,t})'$, $f(\mathbf{x}_{i,t})$ as in (11) but with $f : \mathbb{R}^{(q+2)} \to \mathbb{R}$ and $d_0 = q + 2$, and parameters estimated by minimizing (A.2). Other than these differences, implementation details are the same as detailed in Section 3.3.

In the option panel exercise, we do not consider the AHBS and CW models since they are not designed to fit an option panel, but rather only the implied volatility surface on a given day, as their parameters are time-varying by definition. This is in contrast to structural models such as Black–Scholes and Heston that specify the underlying asset dynamics as a function of parameters that are fixed over time.

Acknowledgments

We would like to thank the Associate Editor, two anonymous referees and conference participants at Econometric and Big Data Analyses of Global Economy, Financial Markets and Economic Policies for useful comments and suggestions.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Fan's research was supported by NSFC grant No.71991470/71991471.

References

Ackerer, D., Tagasovska, N., Vatter, T. (2020), "Deep Smoothing of the Implied Volatility Surface," *Advances in Neural Information Processing Systems*, 33, 11552–11563. [996]

Amilon, H. (2003), "A Neural Network Versus Black–Scholes: A Comparison of Pricing and Hedging Performances," *Journal of Forecasting*, 22, 317–335. [996]

Ait-Sahalia, Y., and Lo, A. W. (1998),"Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53, 499–547. [996,1000]

Ait-Sahalia, Y., and Duarte, J. (2003), "Nonparametric Option Pricing Under Shape Restrictions," *Journal of Econometrics*, 116, 9–47. [996,1000]

Ait-Sahalia, Y., Li, C., and Li, C. X. (2021), "Implied Stochastic Volatility Models," *The Review of Financial Studies*, 34, 394–450. [995,996,1000,1004]

Andersen, T. G., Fusari, N., and Todorov, V. (2015), "The Risk Premia Embedded in Index Options," *Journal of Financial Economics*, 117, 558–584. [995,996,997,1000,1003,1004,1007]

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27, 417–427. [1004]

Baker, S. R., Bloom, N., and Davis, S. J. (2016), "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636. [1004]

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical Performance of Alternative Option Pricing Models," *The Journal of Finance*, 52, 2003–2049. [1000,1001]

Bandi, F. M., Fusari, N., and Renò, R. (2021), "Structural Stochastic Volatility," Unpublished working paper. [996]

Bates, D. (2000), "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238. [996]

Bauer, B., and Kohler, M. (2019), "On Deep Learning as a Remedy for the Curse of Dimensionality in Nonparametric Regression," *Annals of Statistics*, 47, 2261–2285. [999]

Black, F., and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654. [995,996,997,1001,1007]

Bollerslev, T., Todorov, V., and Xu, L. (2015), "Tail Risk Premia and Return Predictability," *Journal of Financial Economics*, 118, 113–134. [996,1004]

Breeden, D., and Litzenberger, R. H. (1978), "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651. [996]

Buhler, H., Gonon, L., Teichmann, J., and Wood, B. (2019), "Deep Hedging," *Quantitative Finance*, 19, 1271–1291. [996]

Carr, P., and Wu, L. (2016), "Analyzing Volatility Risk and Risk Premium in Option Contracts: A New Theory," *Journal of Financial Economics*, 120, 1–20. [995,996,998,1000,1007]

Cybenko, G. (1989), "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, 2, 303–314. [996,999]

Duffie, D., Pan, J., and Singleton, K. (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–137. [996,998,1000]

Dugas, C., Bengio, Y., Belisle, F., Nadeau, C., and Garcia, R. (2009), "Incorporating Functional Knowledge in Neural Networks," *Journal of Machine Learning Research*, 10, 1239–1262. [996]

Dumas, B., Fleming, J., Whaley, R. E. (1998), "Implied Volatility Functions: Empirical Tests," *The Journal of Finance*, 53, 2059–2106. [996,997,1000,1001,1007]

Dupire, B. (1994), "Pricing with a Smile," *Risk*, 7, 18–20. [996]

Fan, J., Ma, C., and Zhong, Y. (2021), "A Selective Overview of Deep Learning," *Statistical Science*, 36, 264–290. [996,999]

Fan, J., and Mancini, L. (2009), "Option Pricing with Model-Guided Nonparametric Methods," *Journal of the American Statistical Association*, 104, 1351–1372. [996,1000,1001]

Fan, J., Wu, Y., and Feng, Y. (2009), "Local Quasi-Likelihood with a Parametric Guide," *Annals of Statistics*, 37, 4153–4183. [995,999]

Fan, Y., and Ullah, A. (1999), "Asymptotic Normality of a Combined Regression Estimator," *Journal of Multivariate Analysis*, 71, 191–240. [995,999]

Fang, F., and Oosterlee, C. W. (2009), "A Novel Pricing Method for European Options based on Fourier-Cosine Series Expansions," *SIAM Journal on Scientific Computing*, 31, 826–848. [998]

Garcia, R., and Gençay, R. (2000), "Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint," *Journal of Econometrics*, 94, 93–115. [996,999,1000]

Gatheral, J., and Jacquier, A. (2014), "Arbitrage-Free SVI Volatility Surfaces," *Quantitative Finance*, 14, 59–71. [996]

Glad, I. K. (1998), "Parametrically Guided Non-parametric Regression," *Scandinavian Journal of Statistics*, 25, 649–668. [995,999]

Gonçalves, S., and Guidolin, M. (2006), "Predictable Dynamics in the S&P 500 Index Options Implied Volatility Surface," *Journal of Business*, 79, 1591–1635. [1001]

Gu, S., Kelly, B., and Xiu, D. (2020), "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33, 2223–2273. [996,999,1000,1002,1004]

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, 327–343. [996,997,1001,1007]

Hornik, K., Stinchcombe, M., and White, H. (1989), "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366. [996,999]

Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994), "A Nonparametric Approach to the Pricing and Hedging of Derivative Securities via Learning Networks," *The Journal of Finance*, 49, 851–889. [996,999,1000]

Jackwerth, J. C., and Rubinstein, M. (1996), "Recovering Probability Distributions from Option Prices," *The Journal of Finance*, 51, 1611–1631. [996]

Liu, S., Oosterlee, C. W., and Bohte, S. (2019), "Pricing Options and Computing Implied Volatilities using Neural Networks," *Risks*, 7, 1–22. [996]

Masters, T. (1993), "Practical Neural Network Recipes in C++," New York: Academic Press. [1000]

Medvedev, A., and Scaillet, O. (2007), "Approximation and Calibration of Short-Term Implied Volatilities under Jump-Diffusion Stochastic Volatility," *The Review of Financial Studies*, 20, 427–459. [996]

Mhaskar, H., Liao, Q., and Poggio, T. (2016), "Learning Functions: When is Deep Better Shallow," Center for Brains, Minds and Machines (CBMM) Memo 45. [999]

Moller, M. F. (1993), "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks*, 6, 525–533. [1000]

Oh, D. H. and Patton, A. J. (2021), "Better the Devil You Know: Improved Forecasts from Imperfect Models," Unpublished working paper. [995]

Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., and Liao, Q. (2017), "Why and When can Deep-but not Shallow-Networks Avoid the Curse of Dimensionality: A Review," *International Journal of Automation and Computing*, 14, 503–519. [996,999]

Rolnick, D., and Tegmark, M. (2018), "The Power of Deeper Networks for Expressing Natural Functions," in *International Conference on Learning Representations*. [999]

Ross, S. A. (1976), "Options and Efficiency," *The Quarterly Journal of Economics*, 90, 75–89. [996]

Rubinstein, M. (1994), "Implied Binomial Trees," *The Journal of Finance*, 49, 771–818. [995,996,997,1000]

Ruf, J., and Wang, W. (2021), "Hedging with Linear Regressions and Neural Networks," *Journal of Business and Economic Statistics* (forthcoming). [996]

Stutzer, M. (1996), "A Simple Nonparametric Approach to Derivative Security Valuation," *The Journal of Finance*, 51, 1633–1652. [996]

Todorov, V. (2019), "Nonparametric Spot Volatility from Options," *Annals of Applied Probability*, 29, 3590–3636. [998,1003]