A Framework for Real-Time Modeling and Forecasting of Large Unbalanced Option Implied Volatility Surfaces*

Arnaud Dufays, Kris Jacobs and Jeroen Rombouts

September 12, 2024

Abstract

Forecasting the option implied volatility (IV) surface is difficult with standard time series models because of its time-varying granularity. We propose a two-step real-time sequential forecasting framework. This framework can accommodate any underlying model for option prices and IVs, including dynamic option pricing models with latent variables, nonparametric models, and machine learning methods. The first step constructs a fitted IV surface for a given model. The second step estimates a dynamic model on summary statistics of the entire history of the fitted IV surfaces. It is implemented sequentially using an updating rule, which allows the method to handle very large datasets and high data frequencies. In an empirical application, we estimate a heterogeneous autoregressive (HAR) type model on the fitted surfaces and evaluate its performance in forecasting S&P 500 IV surfaces. We find that the model yields improved forecasts regardless of the model used to fit the IV surface, and significantly outperforms random walk forecasts. Nonparametric and machine learning approaches typically outperform state-of-the-art dynamic option pricing models.

JEL Classification: G12

Keywords: Implied volatility surface; Option pricing; Machine learning; Big data; Real-time estimation; Forecasting; Surface heterogeneous autoregressive model.

*Dufays: EDHEC Business School; Jacobs: University of Houston; Rombouts: ESSEC Business School; Correspondence to Arnaud Dufays, EDHEC, arnaud.dufays@edhec.edu. We are very grateful to Thorben Eilers for research support.

1 Introduction

The rapid growth of the global derivatives markets over the past few decades continues unabated. Option markets are especially dynamic, with a record 108.2 billion contracts traded and/or cleared in 2023, a 98.4% increase from 2022, and 90% of this trading volume consists of equity and index options.[1]

New and improved techniques for modeling and forecasting equity and index options are therefore badly needed. However, the literature on option pricing is increasingly fragmented. On the one hand, researchers have studied dynamic models with stochastic volatility and jumps that are extensions of the seminal Black-Scholes model. While these models offer valuable insights, they are notoriously difficult to implement and their estimation is time-consuming. It is therefore extremely challenging to implement them recursively in real time. Because of this complexity, the option literature has occasionally relied on parametric and nonparametric techniques that directly model the implied volatility (IV) surface, see for instance the ad-hoc Black-Scholes method of Dumas et al. (1998) and the kernel smoother used by OptionMetrics (2022). Recently, more sophisticated machine learning methods have been proposed to model the IV surface. These techniques use option characteristics as features and differ with respect to the nonlinear functions used for predicting IVs.

This paper proposes a new framework that makes several contributions to the option literature. First, our proposed approach facilitates comparisons between very different option pricing models, as well as any newly proposed model. Second, we contribute to the study of the *forecasting* performance of the various models and model classes, and we use the entire history of the volatility surface to construct these forecasts. Our focus on forecasting provides an interesting complement to in-sample comparisons, which favor more complex models. Our forecasting setup is novel and differs from existing approaches that use a random walk

[1] See FIA (2024) for more details on historical option volumes.

assumption (Almeida et al., 2023). We instead use the full history of the IV surfaces. We find that this yields substantial improvements in forecasting accuracy. Estimation is done in closed form and can be implemented sequentially using an updating rule, that is, additional information can be added to the estimator without re-estimating using the entire history of IV surfaces. This ensures that our approach can be implemented and continuously updated at high frequencies, i.e. in real time.

Our proposed modeling and forecasting setup overcomes several well-known important challenges. Because the number of option contracts and their characteristics change on a daily basis, option surfaces are characterized by time-varying granularity. Many recent papers (Medvedev and Wang, 2022; Kelly et al., 2023; Shang and Kearney, 2022) tackle this issue by transforming daily option panels into a fixed-grid implied volatility surface, defined by specific maturity and moneyness categories. The advantage of this approach is that it allows the use of standard time series methods due to the stable grid dimension. However, it also has significant drawbacks. First, since actual option panels are unbalanced, some surface areas have to be constructed by interpolation or extrapolation. Second, a fixed grid does not reflect the distribution of the number of available options across the surface, often over-representing long-maturity options. This can bias model training to minimize errors in surface areas with infrequent trading, impacting forecast performance when error metrics are computed on the truly observed option panels. Third, because the number of contracts in daily option surfaces has been increasing exponentially, the grid needs to be refined over time. Our approach demonstrates that a fixed-grid implied volatility surface is not required for using time series methods. Our approach takes the implied volatility surface as observed and fits standard realized variance time series type models on the *unbalanced* option panel over time. This allows for improved forecast evaluation for increasingly dense option IV surfaces.

Our proposed forecasting setup for implied volatility surfaces can be used with any type

of option pricing model. The literature has proposed very different approaches to price options and fit volatility surfaces, including dynamic option pricing models with latent state variables, as well as nonparametric and machine learning models. However, there is a lack of studies that compare the (forecasting) performance of these various modeling approaches. While dynamic option pricing models provide many useful insights, estimation is complex and computationally demanding, even on small option panels over short periods, because they typically contain multiple latent stochastic processes.[2] Recursive implementation and forecasting option prices with these models is therefore extremely challenging, and the benefits from including stochastic volatility factors and/or jumps for forecasting purposes have not yet been extensively studied. Partly in response to this, more pragmatic machine learning alternatives for fitting IV surfaces have been developed. These are based on polynomial functions (Zhang and Xiang, 2008), spline functions (Fengler, 2009) or artificial neural networks (Ackerer et al., 2020; Zhang et al., 2023). Nonparametric approaches based on kernel estimations to model daily surfaces have bee used by OptionMetrics (2022) and Ulrich et al. (2023).

To compare the forecasting performance of the various modeling approaches, one can fit each model separately to the daily option surface and assume that the forecasted surface is the same as today's surface, i.e. a random walk (RW) assumption. This assumption is convenient and significantly reduces computing time, because the estimation based on daily option data can be carried out in parallel using nonlinear least squares. This method is used in the comparative forecasting study by Almeida et al. (2023). Given the well-known strong persistence in volatility, the RW assumption for predicting daily option surfaces is a good starting point.[3] However, because (risk-neutral) volatility is mean-reverting, it stands to reason that it may be possible to improve IV surface forecasts by exploiting the history

[2]For contributions to this literature, see for instance Heston (1993), Bates (2000), Christoffersen et al. (2009), Christoffersen et al. (2010), Andersen et al. (2015), and Gruber et al. (2021).

[3]For example, Bollerslev et al. (2009) uses this assumption to forecast realized variance when computing the variance risk premium.

of the IV surfaces and the rich information in past implied volatility surfaces, simply relying on the past fitted surfaces. Moreover, such a setup can draw from a variety of existing time series models for forecasting unbalanced option panels. We illustrate the approach by implementing a model inspired by the well-known heterogenous autoregressive (HAR) model of Corsi (2009) that has been shown to be successful for realized volatility forecasting. Specifically, we linearly combine generated past surfaces as in the HAR model. The proposed model is computationally efficient, because it can be estimated sequentially in closed form.

Our empirical exercise uses daily S&P 500 equity-index implied volatility option surfaces for the 2016-2021 sample period. We compare the performance of dynamic option pricing models with that of a simple Black-Scholes model, the so-called ad-hoc Black-Scholes model (Dumas et al., 1998), which models the IV surface using a second degree polynomial in moneyness and maturity, and three techniques from the nonparametrics and machine learning literature. The first of these is an artificial neural network with an implementation similar to that in Almeida et al. (2023). The second approach is a kernel smoother used by OptionMetrics (2022) to model the volatility surface. The third approach uses random forests.

For each of these models and a given horizon, we evaluate the forecast performance in IV root mean squared errors for three different forecasts. The first forecast follows Almeida et al. (2023) and uses a random walk assumption. The second forecast uses the autoregressive SHAR approach which uses the entire history of the implied volatility surface. The third forecast uses a robust version of this approach, which we refer to as SHAR-R. We evaluate the quality of these forecasts one day, one week, and one month ahead. Finally, we follow Almeida et al. (2023) and investigate if correcting models using an artificial neural network leads to superior forecasts.

We find significant forecasting performance gains compared to the random walk approach,

We report three main novel empirical findings. Our first conclusion is that the proposed

SHAR and SHAR-R forecasts outperform the random walk for most horizons, regardless of the model used to fit the option surface, i.e. dynamic option pricing models, nonparametric methods, or machine learning techniques. While the estimated parameters of the dynamic surface model indicate strong persistence, which is required for the random walk model to be useful, this persistence extends to longer horizons, up to one month. Second, our comparison of the models used to fit the daily implied volatility surface indicates that machine learning techniques and nonparametric methods yield smaller forecast errors than the dynamic option pricing models. Third, combining our SHAR approach with the artificial neural network error correction, as advocated by Almeida et al. (2023), yields the most accurate forecasts.

Several additional findings are also noteworthy. First, within the class of dynamic option pricing models, stochastic volatility models with jumps outperform multi-factor stochastic volatility models. Second, we also compare our approach to the method of Goncalves and Guidolin (2006) and find that our approach compares favourably. Third, we demonstrate that our SHAR-R robust model implementation can lead to improved forecasts. We find the highest occurrence of sequentially abnormal implied volatility surfaces occurs in 2018 and 2020. Fourth, we find additional forecasting gains from implementing the robust version of the Surface HAR model that mitigates the impact of significant shifts in the implied volatility surface, due to events such as the Covid pandemic crisis for example. Finally, all models experience difficulties when forecasting implied volatilities for short maturities and large moneyness.

Our results are related to various strands of literature. The most closely related paper is Almeida et al. (2023), who find that neural networks perform particularly well at fitting daily option surfaces, both directly from data and from the error surfaces generated by parametric models like Heston (1993) and Carr and Wu (2016). We confirm that these findings. However, our objective is very different because our main focus is on the *forecasting* models, not the models used to fit the implied volatility surface. Specifically, while Almeida et al. (2023) use

a random walk assumption to forecast the implied volatility surface, our proposed SHAR and SHAR-R approaches exploit the entire history of the surfaces. To the best of our knowledge, this modeling approach is novel, and its sequential implementation allows the method to handle very large datasets and high data frequencies. We find that it leads to substantial improvements in forecast accuracy.

We also contribute to the extensive literature on dynamic option pricing models with latent factors that builds on the seminal work of Black and Scholes (1973) and Heston (1993). This literature has given rise to increasingly complex models. It is well-understood that more complex models may be at a disadvantage in forecasting, but this dimension of these models has not been sufficiently explored due to the models' nonlinearities and the computational burdens associated with recursive implementations of these models. While the use of an autoregressive structure of the implied volatility surface may admittedly be inconsistent with the model assumptions, our two-step approach provides a convenient framework to evaluate the forecasting performance of these models and compare them with competitors.[4] It is especially difficult to evaluate the forecasting performance of models with jumps, and their relatively good performance using our approach is especially encouraging.

For parametric option pricing models, an alternative technique to forecast implied volatility surfaces in the absence of the RW assumption models the dynamics of the model parameters, rather than modeling the surface directly. This approach is suggested by Goncalves and Guidolin (2006).[5] This implies a mutivariate time series of coefficient estimates, which are subsequently modeled using a vector autoregressive (VAR) model. The forecasted coefficients are then used to reconstruct the implied volatility surfaces. This two-step method is particularly effective when the number of parameters is low and the parameter space is

[4]Note that the use of random walk forecasts for the IV surface of these models suffers from the same inconsistency criticism.

[5]Goncalves and Guidolin (2006) use a polynomial regression model based on moneyness and maturity to fit the daily implied volatility surface. For similar polynomial functions, see also Dumas et al. (1998) and Fengler et al. (2007).

unrestricted as it allows for quick and sequential forecast updates. However, in scenarios involving large parameter sets, such as with artificial neural networks, nonparametric models, or option pricing models with bounded coefficients, this method becomes infeasible using standard VAR models. In contrast, our method directly models the dynamics of the implied volatility surface and is compatible with any fitting model.

Finally, our findings also contribute to the literature on nonparametrics and machine learning. Our finding that these methods yield better forecasts than state-of-the-art dynamic option pricing models with latent states is very encouraging. We believe that our framework is well suited to develop improved machine learning techniques and to test them out-of-sample.

The rest of the paper is structured as follows. Section 2 describes the option surface data and the option data filters. Section 3 discusses the model setup and the proposed dynamic surface SHAR model. Section 4 summarizes the various models used to fit the daily IV surface. Section 5 presents the empirical evidence on the fit of the various surface models and the dynamics of the daily surfaces. Section 6 reports the results of the forecasting exercise. Finally, Section 7 concludes and proposes directions for future research.

2 Data

We use daily observations of Black and Scholes implied volatilities from S&P 500 equity-index (SPX) options traded at the Chicago Board Options Exchange covering the period from 2016 to 2021. The data is filtered to only include option data with implied volatilities between 0.05 and 1.5, with a volume and open interest above 4. Following Almeida et al. (2023), options must have a positive bid price, a higher offer than bid price as well as an average price (of bid and ask) above 0.5, moneyness between 0.8 and 1.6, maturity between

20 and 240 days.[6] Finally, we exclude options which violate the no-arbitrage put-call-parity condition. Post filtering, the dataset contains 2,332,617 options.

Table 1 provides standard descriptive statistics. Apart from the well known characteristics of implied volatility such as a long term average around 20% and large variation over time, we observe a strong increase in average number of options per day, from respectively 898 in 2016 to 1871 in 2021. Given our option filters, average moneyness and maturity is stable over time. Forecasting IVs given the recent high market activity regarding ultrashort maturities, i.e. the so-called zerot options as studied in Bandi et al. (2023), is subject for future research.

Table 1: Summary statistics of IV surfaces

Year	2016	2017	2018	2019	2020	2021
IV mean	0.18	0.13	0.19	0.18	0.29	0.22
IV standard deviation	0.07	0.06	0.09	0.07	0.13	0.10
IV min	0.05	0.05	0.05	0.05	0.07	0.06
IV max	0.69	0.92	0.94	0.87	1.12	0.90
Moneyness mean	0.95	0.95	0.95	0.95	0.95	0.94
Maturity mean (days)	59	57	58	62	63	63
# contracts/day	898	1175	1481	1464	1661	1871

Notes: Summary statistics for implied volatility surfaces over the considered years. The last line measures the average number of option contracts per day in each year.

Figure 1, panel (a), displays the implied volatility surface on December 30, 2019. On this typical day, the implied volatilities are displaying the characteristic asymmetric smile for a short maturity levels. For longer maturities, the smile is less strong. For a given moneyness level, the term structure is relatively flat. We can also clearly see that there are much more options for short maturities. Figure 1, panel (b), shows the implied volatility surface on March 16, 2020, an atypical day at the height of the initial Covid market turmoil. The shape of the implied volatility surfaces inverses along the maturity axis with extremely high

[6]Moneyness is defined as strike price divided by the index level, maturity as the number of days between now and the expiration date.

short-term and overall implied volatility which decreases for larger maturities. Given the strong uncertainty about near-term economic developments during that period, the surface adapts to investor sentiment. The typical smile behaviour appears absent with the slope being visibly flatter. The term structure is downward sloping with long maturity IVs still substantially higher than their long term averages. In a standard sequential forecasting procedure such an atypical surface impacts both forecast performance and sequential estimators for several days.

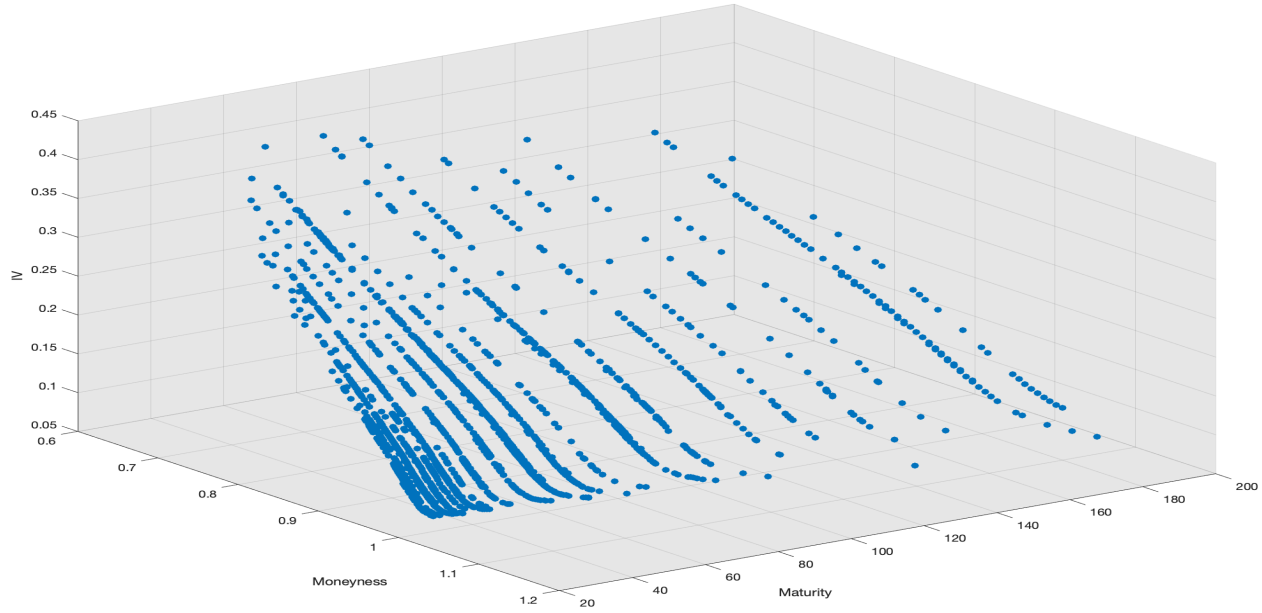
Figure 2 plots in Panel (a) average daily implied volatility between 2016 and 2021. We split the options in short maturity (less than 60 days) and long maturity (more than 150 days). While both time series are in general highly persistent, we also observe several peaks, the onset of COVID-19 crisis in particular, where implied volatility jumps drastically. The long maturity implied volatilities are more stable than the short maturity counterparts, and are also higher in about 80 percent of the days in our sample. However in crisis periods, the short maturities' implied volatilities increase substantially, which can be seen in the term structure plot in Panel (b) of Figure 2 where term structure is computed as the difference between the short and long maturity implied volatility time series.

3 A Dynamic Surface Model

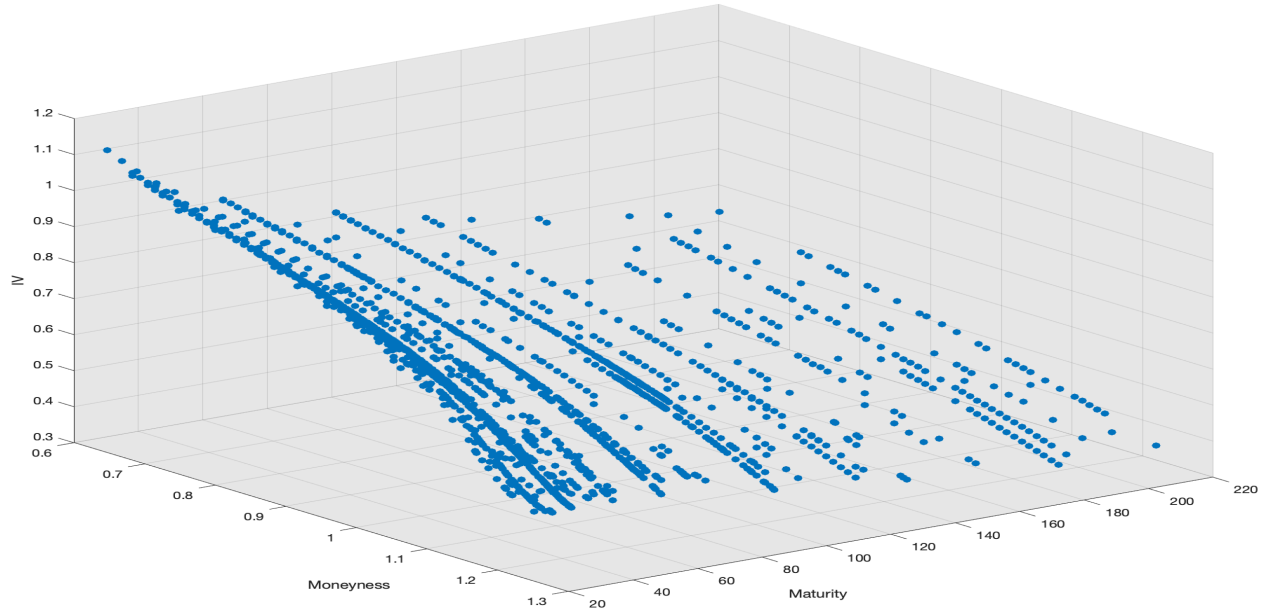
In Section 3.1, we first introduce the notations and settings. Section 3.2 lays out the general framework for using a time series model without a fixed-grid panel and we define a dynamic surface model. Next, we highlight in Section 3.3 how to estimate the model sequentially to keep computations as efficient as possible. Section 3.4 robustifies the estimator with an automatic abnormal surface detection algorithm.

Figure 1: Implied volatility surfaces - raw data

(a) December 30, 2019



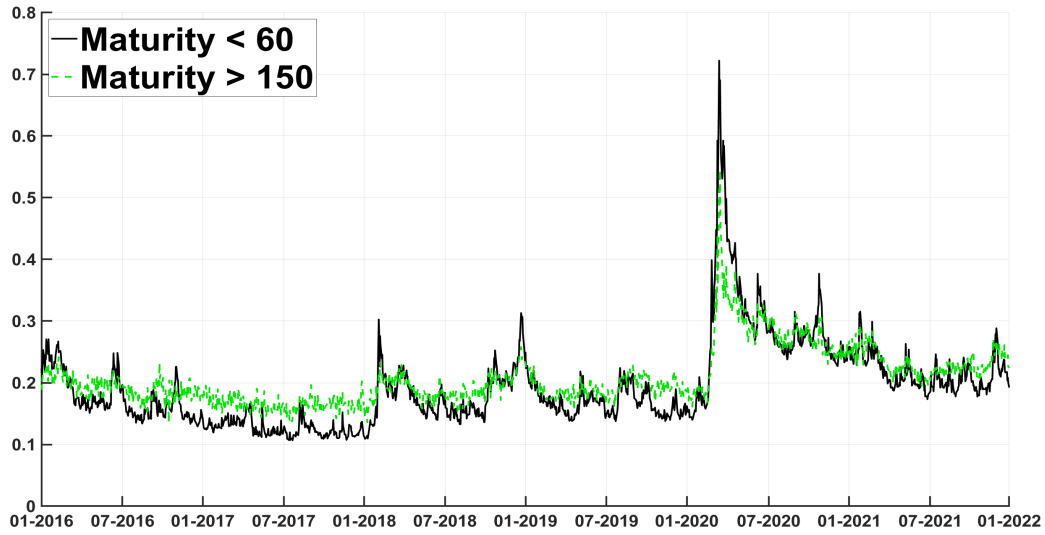
(b) March 16, 2020



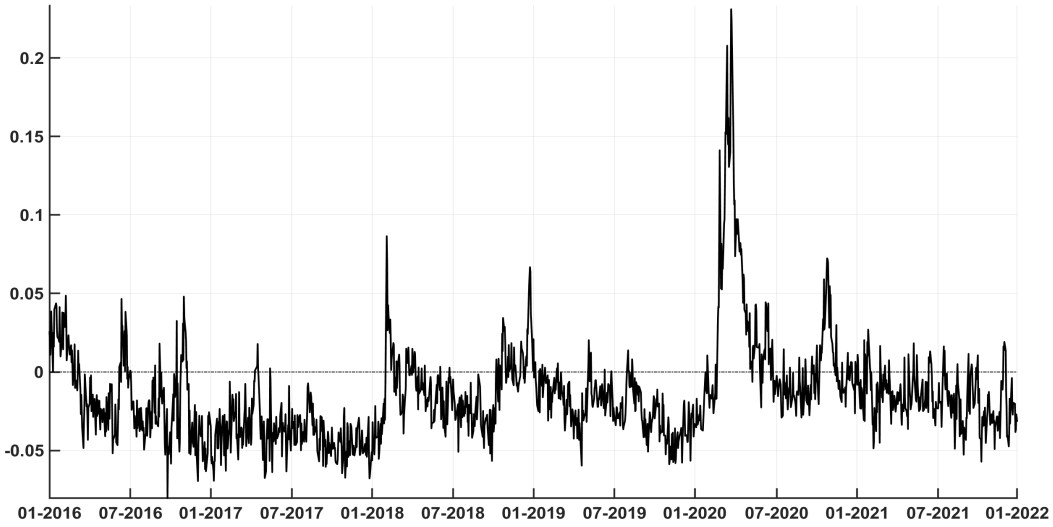
Notes: Panel (a) displays the option surface observed on December 30, 2019, Panel (b) for March 16, 2020. Maturity is measured in days, moneyness as strike divided by index level. IV are Black?Scholes implied volatilities.

Figure 2: Daily average implied volatilities

(a) Daily average implied volatilities



(b) Daily term structure



Notes: Panel (a) displays daily average implied volatilities for options with maturity below 60 days (blue) and above 150 days (magenta). Panel (b) displays the daily term structure, measured as the difference between the two time series in panel (a), short minus long maturity.

3.1 Setting

Before introducing the model for IV surfaces, let us fix some notation and the general setting. From an initial day onwards[7], we observe at daily frequency new IV surfaces which yields at day t a stream of IV surfaces $IV(O_{i,1}), IV(O_{i,2}), \ldots, IV(O_{i,t})$, each surface consisting of N_l option implied volatilities with their characteristics $O_{i,l}$, $l = 1, \ldots, t$. For instance on day l, we observe the surface $IV(O_{i,l})$ with options $i = 1, \ldots, N_l$, and option characteristics $O_{i,l} = \{m_{il}, \tau_{il}\}$ in which $m_{il} \equiv \frac{K_{il}}{S_l}$ measures the moneyness with K_{il} the strike price and S_l the index level, and τ_{il} is the maturity of the option measured in calendar days. The main question is how to forecast at day t the IV surface h days ahead, i.e. at day $t+h$, which is not trivial given that the option characteristics are time varying. A standard approach starts by fitting the implied volatility surface as follows:

$$IV(O_{i,l}) \quad = \quad IV^M(O_{i,l}, \Theta_l) + \varepsilon_{il}, \tag{1}$$

in which $IV^M(\cdots)$ emphasizes that the model M generates the IV for an option with characteristics $O_{i,l}$ given the set of model parameters Θ_l. For instance, the model parameters Θ_l stands for to the spot variance for the BS model and to $\Theta_l = \{V_l, \kappa, \theta, \sigma, \rho\}$ for the Heston model, see Section 4.4 for its specification. Note that Θ_l can be highly dimensional, e.g. a neural network, and is not necessarily a finite dimensional parameter. In fact, it can also represent a nonparametric function for example in the case of a random forest implied decision tree.

An easy procedure for forecasting at day t the IV surface at day $t + h$ consists then of the following two steps. First, fit (1) using least squares, that is estimate the parameters Θ_t given the N_t options observed at day t. This yields $\hat{\Theta}_t$. Second, forecast the IV surface at day $t + h$ as $\widehat{IV}(O_{i,t+h}) = IV^M(O_{i,t+h}, \hat{\Theta}_t)$. The second step is called the random walk

[7] The setup works for any frequency, but we work with daily IV surfaces in the application.

(RW) forecasting approach because it assumes that the fitted surface at day t is expected to be identical to any future surface to be forecasted. This procedure is computationally fast because it only requires estimating a model on a single daily surface.

3.2 Dynamic Surface Model

The daily fit and forecast approach explained above relies on one day of option data only, ignoring the information set observed up to that day, consisting on day t of the stream of IV surfaces which can be summarized by the parameter estimates stream $\hat{\Theta}_{1:t} = \{\hat{\Theta}_1, \ldots, \hat{\Theta}_t\}$. The RW forecast assumption implies forecasting the IV surface at day $t+h$ as todays fitted surface generated by the model parameter estimate $\hat{\Theta}_t$. However, the financial econometrics literature has consistently shown that autoregressive fractionally integrated moving average and heterogeneous autoregressive models (see, Baillie et al., 1996; Corsi, 2009), among other processes, typically outperform random walk processes for modeling realized variances and the VIX.

While modelling directly the dynamics of the IV surfaces is computationally very expensive and difficult to update, we use instead the parameter estimates stream $\hat{\Theta}_{1:t}$ to build a dynamic IV surface model that can be estimated in real time. Computational speed is important given that a daily surface is represented by a large number of option contracts. Our key idea is that, for any moneyness-maturity couple on the to be forecasted surface, one can exploit the parameter estimates stream for generating a time series of length t composed of past model implied volatilities. Then we can apply a dynamic time series model for forecasting the implied volatility surface at time $t+h$.

More precisely, we model the IV surface at day $l = 1, \ldots, t$ for horizon h as follows:

$$IV(O_{i,l}) \quad = \quad \beta_h^{(0)} + \sum_{k=1}^{K} \beta_h^{(k)} f_k(O_{i,l}, \hat{\Theta}_{1:l-h}) + \varepsilon_{i,l} \qquad i = 1, \ldots, N_l, \tag{2}$$

with f_k an arbitrary function of past model implied IV surfaces. This general specification allows polynomial type structures and interaction effects. An important case is a linear function f_k, allowing an autoregressive process with K lags: $\beta_h^{(0)} + \sum_{k=1}^{K} \beta_h^{(k)} IV^M(O_{i,l}, \hat{\Theta}_{l-h-k+1})$. Note that the latter linear specification is different from a standard autoregressive type process that uses linear combinations of previous IV values to predict the current IV value. In fact, the specification predicts the IV value using past IV values generated by a specific model. This is possible because we use the IV values observed at options characteristics $O_{i,l}$ from model surfaces given by $IV^M(O_{i,l}, \hat{\Theta}_{l-h-k+1})$ with $k = 0, \ldots, K$. Note also that the random walk forecasting approach is a special case of Model (2) for $K = 1$, $f_1(O_{i,l}, \hat{\Theta}_{1:l-h}) = IV^M(O_{i,l}, \hat{\Theta}_{l-h})$, $\beta_h^{(0)}$ equal to zero and $\beta_h^{(1)}$ equal to one.

Model (2) can be written in matrix notation by stacking all the implied volatilities observed at day l into $\boldsymbol{y}_l = \big(IV(O_{1,l}), \ldots, IV(O_{N_l,l})\big)'$ and using the $l-h$ model implied volatilities parameter estimates $\hat{\Theta}_{1:l-h}$ into $\boldsymbol{x}_{l|1:l-h}^{(k)} = \big(f_k(O_{1,l}, \hat{\Theta}_{1:l-h}), \ldots, f_k(O_{N_l,l}, \hat{\Theta}_{1:l-h})\big)'$:

$$\begin{aligned} \boldsymbol{y}_l &= \beta_h^{(0)} + \sum_{k=1}^{K} \beta_h^{(k)} \boldsymbol{x}_{l|1:l-h}^{(k)} + \boldsymbol{\varepsilon}_{l+h}, \\ &= \boldsymbol{X}_{l|l-h} \boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_l, \end{aligned} \tag{3}$$

where $\boldsymbol{\beta}_h = (\beta_h^{(0)} \quad \beta_h^{(1)} \quad \ldots \quad \beta_h^{(K)})'$ and $\boldsymbol{X}_{l|l-h} = (\mathbf{1}_{N_l}, \boldsymbol{x}_{l|1:l-h}^{(1)}, \ldots, \boldsymbol{x}_{l|1:l-h}^{(K)})$, a matrix of dimension $(N_l \times K)$.

To use all available information at day t to forecast the IV surface at $t+h$, we pool Model (3) for days $l = h+K, \ldots, t$ and denote the vector $\boldsymbol{Y}_t = (\boldsymbol{y}_{h+K}, \ldots, \boldsymbol{y}_t)'$ the available IV surfaces and the corresponding matrix $\boldsymbol{X}_{h|t}$ stacking the matrices $\boldsymbol{X}_{h+K|K} \ldots \boldsymbol{X}_{t|t-h}$ which are the model-predicted implied volatilities according to the model structure. Using a squared Euclidean loss function, the model parameters can be estimated using the ordinary least squares (OLS) formula as follows: $\hat{\boldsymbol{\beta}}_{h|t} = (\boldsymbol{X}_{h|t}' \boldsymbol{X}_{h|t})^{-1} \boldsymbol{X}_{h|t}' \boldsymbol{Y}_t$, where $\hat{\boldsymbol{\beta}}_{h|t}$ is indexed by t to highlight that the estimation will be sequentially repeated for every new day t. Given

the option characteristics $O_{i,t+h}$ for $i = 1, \ldots, N_{t+h}$, the direct h day horizon forecast of the IV surface at day t is then given by:

$$\widehat{IV}(O_{i,t+h}) \quad = \quad \hat{\beta}^{(0)}_{h|t} + \sum_{k=1}^{K} \hat{\beta}^{(k)}_{h|t} f_k(O_{i,t+h}, \hat{\Theta}_{1:t-j+1}) \quad i = 1, \ldots, N_{t+h}. \tag{4}$$

Extensions of the model outlined above can be made, for example according to Bollerslev et al. (2016), Cipollini et al. (2021) who put forward new models incorporating measurement error in realized volatility models.

3.3 Real-time estimation

In a daily updating setup, the matrix $\boldsymbol{X}_{h|t}$ and vector $\boldsymbol{y}_t$ involved in the OLS formula increases rapidly in size since the row dimension grows every day t with N_t observations. To save computational resources, we implement a sequential estimation of the parameter vector $\boldsymbol{\beta}_{h|t}$. Specifically, given the OLS estimates at time t for the h horizon model and denoting the matrix inverse $\boldsymbol{\Omega}^{-1}_{h|t} = (\boldsymbol{X}'_{h|t}\boldsymbol{X}_{h|t})^{-1}$, the parameter estimated are updated at time $t+1$ as follows:

$$\begin{aligned}
\boldsymbol{\Omega}^{-1}_{h|t+1} &= \boldsymbol{\Omega}^{-1}_{h|t} + (\boldsymbol{\Omega}^{-1}_{h|t}\boldsymbol{X}'_{h|t+1})(\boldsymbol{I}_{N_{t+1}} + \boldsymbol{X}_{h|t+1}\boldsymbol{\Omega}^{-1}_{h|t}\boldsymbol{X}'_{h|t+1})^{-1}(\boldsymbol{X}_{h|t+1}\boldsymbol{\Omega}^{-1}_{h|t}), \\
\hat{\boldsymbol{\beta}}_{h|t+1} &= \boldsymbol{\Omega}^{-1}_{h|t+1}\big(\boldsymbol{\Omega}_{h|t}\hat{\boldsymbol{\beta}}_{h|t} + \boldsymbol{X}'_{h|t+1}\boldsymbol{y}_{t+1}\big),
\end{aligned} \tag{5}$$

where $\boldsymbol{I}_{N_{t+1}}$ denotes the identity matrix of size N_{t+1}.

3.4 Robust estimation to abnormal surfaces

The Model in (2) forecasts the IV surface for day $t+h$ using a linear combination of the predicted IV surface of the model at time t, as well as average predicted IV surfaces from periods $t-4$ to t and $t-21$ to t. Due to the high persistence of IV surfaces, the predicted

IV surface at time t generally carries the most weight in this combination, indicated by a $\hat{\beta}_h^{(1)}$ value close to one. However, this prediction relies heavily on the parameter estimate $\hat{\Theta}_t$, which is derived from options observed at time t. If an exogenous shock, such as the COVID pandemic or a flash crash, significantly disrupts the IV surface at time t, the resulting predicted IV surface could be distorted. To identify those abnormal surfaces, we apply the outlier detection method proposed by Rousseeuw and Croux (1993). Let us denote the average squared residuals from day 1 to t by $\text{ASR}_{1:t}$ where $\text{ASR}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}\big(IV(O_{i,t}) - IV^M(O_{i,t},\hat{\Theta}_t)\big)^2$. A surface is considered abnormal at time t according to the following steps:

1. Compute the median of the logarithm of the series, $m = \text{Median}_{i\in[1,t-1]}\big(\ln(\text{ASR}_i)\big)$.

2. Compute a robust standard deviation estimate following Rousseeuw and Croux (1993), $\hat{\sigma} = 1.1926\text{Median}_{i\in[1,t-1]}\big(\text{Median}_{j\in[1,t-1]}\big(\ln(\text{ASR}_i) - \ln(\text{ASR}_j)\big)\big)$.

3. The surface at time t is considered as abnormal if $\frac{\ln(\text{ASR}_t)-m}{\hat{\sigma}} > 3$.

This outlier detection procedure offers two significant advantages within our framework. First, its computation is fast and does not require extensive memory resources. Secondly, the method is applicable across any IV model. As a result, the procedure can be universally applied to any model, and the outliers detected will vary depending on the specific IV model used. This versatility is appealing, as every model exhibits its own flexibility for capturing the shape of the IV surface.

We use this outlier detection procedure to make our OLS parameter estimates robust to abnormal surfaces. Specifically, we do not update the matrix inverse $\mathbf{\Omega}_{h|t+1}^{-1}$ and parameter estimates $\hat{\boldsymbol{\beta}}_{h|t+1}$ (see Subsection 3.3) when an outlier is detected at time t. In addition, since the current surface is abnormal, it does not seem optimal to use it as the most important predictor (see $IV^M(O_{i,t+h},\hat{\Theta}_t)$ in the forecast equation (4)). Instead, we replace $IV^M(O_{i,t+h},\hat{\Theta}_t)$ with the average over the last two surfaces, i.e. $\frac{1}{2}(IV^M(O_{i,t+h},\hat{\Theta}_t)+IV^M(O_{i,t+h},\hat{\Theta}_{t-1}))$, when an outlier is detected. This model is referred to as the robust version in the following.

4 Fitting the option surface

In this section, we discuss the different families of IV surface fit models that we consider in our empirical application below. Sections 4.1 and 4.2 summarize the classical ad-hoc Black-Scholes and nonparametric approaches. Section 4.3 discusses two popular machine learning approaches. Finally, Section 4.4 discusses various models from the literature on dynamic option pricing models with latent factors.

4.1 The Ad-Hoc Black-Scholes Model

A very practical and fast to fit model is the ad-hoc Black-Scholes (AHBS) specification, used by Dumas et al. (1998) and Seo and Wachter (2019) among others, which fits the IV surface using moneyness and maturity with a second degree polynomial:[8]

$$IV(O_{it}) = \beta_0^{\text{LR}} + \beta_1^{\text{LR}} m_{it} + \beta_2^{\text{LR}} m_{it}^2 + \beta_3^{\text{LR}} \tau_{it} + \beta_4^{\text{LR}} \tau_{it}^2 + \beta_5^{\text{LR}} m_{it} \tau_{it} + \varepsilon_{it}. \tag{6}$$

The AHBS model provides a smooth approximation of the IV surface.

4.2 Nonparametric Models

We consider a standard kernel density smoother, as implemented in OptionMetrics (2022) for example. We compute the IV surface with a separate kernel smoother for calls ($I_{it} = 1$) and puts ($I_{it} = 0$), computed on the option's delta Δ_{it} rather than moneyness m_{it}, therefore $O_{it} = \{\text{rf}_t, \text{div}_t, \Delta_{it}, \tau_{it}, I_{it}\}$. It takes the form:

$$IV^M(O_{it}, b) = \sum_{j=1}^{N_t} \frac{\text{Vega}(\tau_{jt}, K_{jt}) \phi(\Delta_{jt} - \Delta_{it}, \ln(\tau_{jt}) - \ln(\tau_{it}), I_j - I_i)}{\sum_{j=1}^{N_t} \text{Vega}(\tau_{jt}, K_{jt}) \phi(\Delta_{jt} - \Delta_{it}, \ln(\tau_{jt}) - \ln(\tau_{it}), I_j - I_i)} IV(O_{jt}) \tag{7}$$

[8] The Black-Scholes model assumes a constant spot variance which is equivalent to a linear regression with constant only model in this framework.

where $\phi(x, y, z) = \exp\left(-(x^2/2b_1 + y^2/2b_2 + z^2/2b_3)\right)$ and with $b = (b_1, b_2, b_3)$ the bandwidth vector that determines the level of smoothing. OptionMetrics (2022) uses $b = (0.05, 0.005, 0.001)$. For consistency with the other models used in this study, we have modified the kernel smoother in Equation (7) by incorporating the moneyness dimension in place of the option's delta. Additionally, as we transform each put option into a call option using Put-Call parity, we use a 2-dimensional Gaussian kernel. The adapted optionMetrics kernel smoother is computed as follows:

$$IV^M(O_{it}, b) = \sum_{j=1}^{N_t} \frac{\text{Vega}(\tau_{jt}, K_{jt})\phi(m_{jt} - m_{it}, \ln(\tau_{jt}) - \ln(\tau_{it}))}{\sum_{j=1}^{N_t} \text{Vega}(\tau_{jt}, K_{jt})\phi(m_{jt} - m_{it}, \ln(\tau_{jt}) - \ln(\tau_{it}))} IV(O_{jt}), \tag{8}$$

in which $\phi(x, y) = \exp\left(-(x^2/2b_1 + y^2/2b_2)\right)$. The bandwidth parameters b_1 and b_2 are estimated on a daily basis using cross-validation, where the test set exhibits 30% of the daily options data.

4.3 Machine Learning Methods

To allow for additional non-linearities, we also consider machine learning methods. We first consider artificial neural networks (ANNs). For simplicity, we follow the same geometric rule as Almeida et al. (2023) with an ANN that exhibits five hidden layers based on sigmoid activation functions with 32, 16, 8, 4 and 2 neurons, respectively:

$$\begin{aligned}
z^{(1)} &= \sigma(W_1^{(1)} m_{it} + W_2^{(1)} \tau_{it} + b^{(1)}), \text{ with } z^{(1)} \in \Re^{32\times 1} \\
z^{(2)} &= \sigma(W^{(2)} z^{(1)} + b^{(2)}) \in \Re^{16\times 1}, \text{ with } z^{(2)} \in \Re^{16\times 1} \\
&\vdots \\
z^{(5)} &= \sigma(W^{(5)} z^{(4)} + b^{(5)}), \text{ with } z^{(5)} \in \Re^{2\times 1} \\
IV(O_{it}) &= W^{(\text{IV})} z^{(5)} + b^{(\text{IV})} + \varepsilon_{it},
\end{aligned}$$

in which $\sigma(\cdot)$ denotes the sigmoid activation function. This configuration implies 811 parameters to estimate for each daily surface.

Our second machine learning application fits the option surface with random forests. Random forests are proposed by Breiman (2001) and produce a surface fit by combining regression trees. A regression tree is a nonparametric method that partitions the feature space to compute local averages as forecasts, see Efron and Hastie (2016) for a textbook treatment and Medeiros et al. (2021) for a comparison of several machine learning techniques to forecast inflation data. The random forest tuning parameters are the number of trees that are used in the forecast combination, the number of features to randomly select when constructing each regression tree split, and the minimum number of observations in each terminal node to compute the local forecasts. We use the standard implementation of the `RandomForestRegressor` function of the `sklearn.ensemble` package in Python with hyperparameters selected by grid search.

4.4 Dynamic Option Pricing Models

We consider a general class of affine option pricing models specified under the risk-neutral measure as

$$\frac{dS_t}{S_t} = (r_t - \delta_t - \lambda\bar{\mu}_s)dt + \sum_{i=1}^{N} \sqrt{V_{it}} dZ_{it} + (e^{J_t^s} - 1)dN_t, \tag{9}$$

$$dV_{it} = \kappa_i(\theta_i - V_{it})dt + \sigma_i \sqrt{V_{it}} dW_{it}, \tag{10}$$

$$\text{Corr}(dW_{it}, dZ_{it}) = \rho_i dt. \tag{11}$$

where S_t is the index level, r_t is the risk-free rate, δ_t is the dividend yield. For each variance factor i, κ_i denotes the speed of mean reversion, θ_i the unconditional mean variance, and σ_i determines the variance of variance. dZ_{it} and dW_{it} are Brownian motions with $corr(dZ_{it}, dZ_{jt}) = 0$ and $corr(dW_{it}, dW_{jt}) = 0, i \neq j$. N_t is a Poisson process with

constant jump intensity λ and J_t^s is the jump size parameter related to returns. We assume $J_t^s \sim N(\mu_s, \sigma_s^2)$. The term $\lambda\bar{\mu}_s$ is the compensation of the jump component, with $\bar{\mu}_s = e^{(\mu_s+\sigma_s^2/2)} - 1$. We also pursue an alternative parameterization of the jump factor, the double exponential (DE) distribution $J_t^s \sim p_l\eta_l \exp(\eta_l y)1_{y\leq 0} + (1-p_l)\eta_r \exp(-\eta_r y)1_{y>0}$.

Denoting $V_t = (V_{1t}, \ldots, V_{Nt})$, $\Theta = (\kappa_1, \theta_1, \sigma_1, \rho_1, \ldots, \kappa_N, \theta_N, \sigma_N, \rho_N, \lambda, \mu_s, \sigma_s^2)$, then the model price of a European call option $C^M(\tau_{it}, K_{it}, V_t, \Theta)$ with maturity τ_{it} and strike price K_{it} is given by:

$$C^M(\tau_{it}, K_{it}, V_t, \Theta) = e^{-r_t\tau_{it}} E[\max(S_{t+\tau} - K_{it}, 0)]. \tag{12}$$

The model exhibits a closed-form expression of the conditional characteristic function for the log index level which makes the integral given in (13) numerically tractable. To optimize the parameters Θ and the spot variances V_t, we minimize the following loss function:

$$\hat{V}_t, \hat{\Theta}_t = \text{argmin}_{V_t,\Theta_t} \sum_{i=1}^{N_t} \left(\frac{C(\tau_{it}, K_{it}) - C^M(\tau_{it}, K_{it}, V_t, \Theta_t)}{\text{Vega}(\tau_{it}, K_{it})} \right)^2, \tag{13}$$

where $C(\tau_{it}, K_{it})$ is the quoted price of the contract with maturity τ_{it} and strike K_{it} on day t and $\text{Vega}(\tau_{it}, K_{it})$ stands for the Black-Scholes sensitivity of the option computed using the implied volatility from the market price of the option $C(\tau_{it}, K_{it})$. The loss function given in (13) can be understood as the first-order approximation of the difference between the observed and the model implied volatilities as outlined in Christoffersen et al. (2009).

From the general specification given by (9)-(10), we consider the following option pricing models:

1. The Heston model shortly written as SV(1). When our model specification exhibits one volatility factor (i.e. $N = 1$) with no jump in return, the model becomes the Heston (1993) model. The process has five unknown parameters to estimate, i.e., the spot variance V_t and $\Theta_{\text{SV}(1)} = \{\kappa_1, \theta_1, \sigma_1, \rho_1\}$.

2. A stochastic volatility model with two factors, shortly written as SV(2). This model was proposed by Christoffersen et al. (2009) and is nested in our model specification by setting $N = 2$ and no jump in returns. This process exhibits two spot variances $V_t = \{V_{1t}, V_{2t}\}$ and 8 parameters, $\Theta_{\text{SV(2)}} = \{\kappa_1, \theta_1, \sigma_1, \rho_1, \kappa_2, \theta_2, \sigma_2, \rho_2\}$.

3. A stochastic volatility model with three factors, written as SV(3). This specification is found to improve upon one and two factor models by Dufays et al. (2024). The number of model parameters amount to 15, i.e. three spot variances $V_t = \{V_{1t}, V_{2t}, V_{3t}\}$ and 12 parameters
$\Theta_{\text{SV(3)}} = \{\kappa_1, \theta_1, \sigma_1, \rho_1, \kappa_2, \theta_2, \sigma_2, \rho_2, \kappa_3, \theta_3, \sigma_3, \rho_3\}$.

4. A stochastic volatility model with jump in returns, shortly written as SVJR. This is the model of Bates (2000) and corresponds to $N = 1$ in our model specification. The number of parameters to estimate each day amounts to 8 and are given by V_t and $\Theta_{\text{SVJR}} = \{\kappa_1, \theta_1, \sigma_1, \rho_1, \lambda, \mu_s, \sigma_s^2\}$.

5. A stochastic volatility model with double exponential jumps in returns, corresponding to $N = 1$ in our model specification, and written as SVDE. There are 9 parameters to estimate and are given by V_t and $\Theta_{\text{SVJR}} = \{\kappa_1, \theta_1, \sigma_1, \rho_1, \lambda, \eta_l, \eta_r, p_l\}$. Double exponential jumps are for example used in Andersen et al. (2015); Dufays et al. (2024).

5 Daily Surface fit

In this Section, we report the results of fitting the available IV surface stream. Section 5.1 defines the dynamic specification used. Section 5.2 summarises the real time estimates and provides information about the detection of abnormal IV surfaces.

5.1 A Surface HAR specification

The specification of the dynamic surface model in (2) depends on the number of regressors (K) and the regressor functionals ($f_1, \ldots, f_K$). We have explored several values of K, functions, parameter restrictions and it turns out that a heterogenous autoregressive model type model of Corsi (2009) is particularly successful and for which we report the results in the next sections. The latter specification takes takes all linear functions $K = 22$ and restricts parameters such that the IV surface only depends on the previous day, over the week and month aggregated fitted surfaces. This yields a parsimonious model with only four parameters.

More specifically, we define the surface HAR (SHAR) to model the IV at day $l = 1, \ldots, t$ for horizon h as follows:

$$\begin{aligned} IV(O_{i,l}) &= \beta_h^{(0)} + \beta_h^{(1)} IV^M(O_{i,l}, \hat{\Theta}_{l-h}) + \beta_h^{(2)} \sum_{j=0}^{4} \frac{IV^M(O_{i,l}, \hat{\Theta}_{l-h-j})}{5} \\ &\quad + \beta_h^{(3)} \sum_{j=0}^{21} \frac{IV^M(O_{i,l}, \hat{\Theta}_{l-h-j})}{22} + \varepsilon_{i,l} \qquad i = 1, \ldots, N_l. \end{aligned} \tag{14}$$

The model parameters are estimated sequentially with OLS (5) and used for forecasting. The robust version is referred to as SHAR-Robust.

5.2 Real Time Estimates

Our approach consists of daily fitting and dynamic modelling of the IV surface. This is done sequentially and we check for abnormal surfaces when new surfaces become available at daily frequency. While our forecasting exercise runs daily from the year 2016 until the end of 2021, we use data from 2015 to guarantee stability of the dynamic model. Doing this, the sequential estimator (5) has been running through the days in 2015 already so that forecasting in 2016 is not impacted by initial conditions.

Figure 3 displays the daily variance estimates of the stochastic volatility process in (10) for the considered option pricing models. The stochastic volatility models without jumps estimates are close to each other and ressemble Figure 2 as expected. The SVJR and SVDE models with jumps yield slightly smaller variance estimates because the jump process captures some of the large movements in the option surfaces.

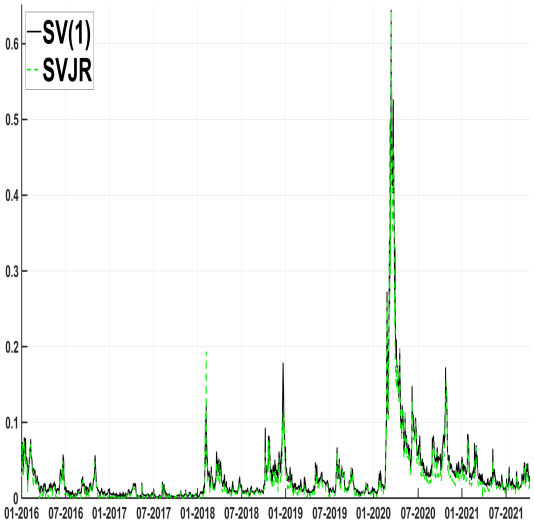
Figure 4 shows the daily implied volatility root mean squared errors (IVRMSEs) expressed in percentages for the fitted SV(1) model. The IVRMSEs are frequently below one percent pre-COVID and systematically larger than one afterwards. The vertical bars in the figure indicate the dates when IV surfaces are classified as outliers according to the procedure described in Section 3.4. We see a clear dependence between fit and detected abnormal surfaces. In fact, forecasting abnormal surfaces is complicated as expected, and the IVRMSEs in such cases can mount to above seven percent. Unsurprisingly, the spikes in implied volatility in Figure 2 and IVRMSE in Figure 4 are highly correlated.

Table 2 gives more details on the number of detected abnormal IV surfaces for the considered models. The year 2020 is marked by the highest number of abnormal surfaces because of the onset of the COVID-19 crisis around March. For example in the class of option pricing models, the SV(1) model finds 39 abnormal surfaces, though this reduces to 26 for the more flexible SVDE model, highlighting the limited flexibility of the former model. In 2018, the emerging market turmoil and interest rate hikes of the Federal Reserve caused several strong fluctuations in the options market which led to multiple abnormal IV surfaces. In particular, the option pricing models fits detect almost 20 abnormal surfaces. However, for the other approaches, there are much less abnormal surfaces. Apart from 2018 and 2020, the other years in the sample have only very few abnormal surfaces for all considered surface fit models.

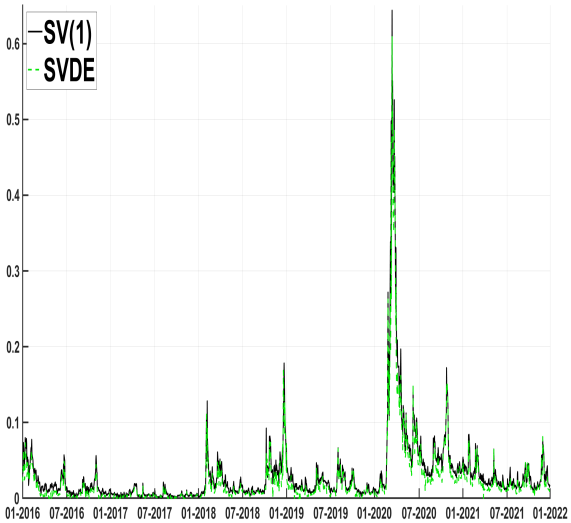
Figure 5 highlights the value of more complex option pricing models by comparing in two ways the daily fit from the one factor stochastic volatility model SV(1) with the same

Figure 3: Spot variance of option pricing models

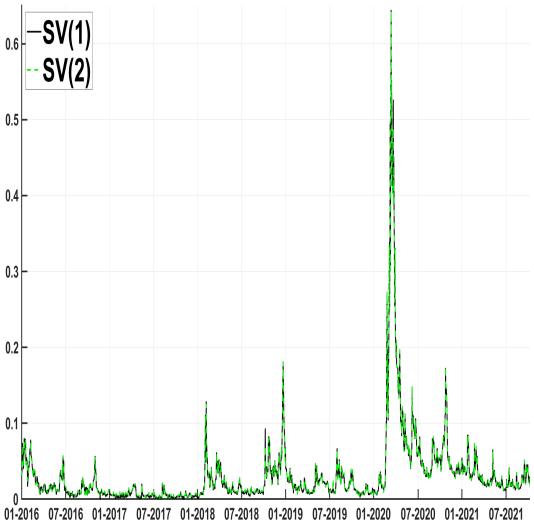
(a) SV(1) - SVJR



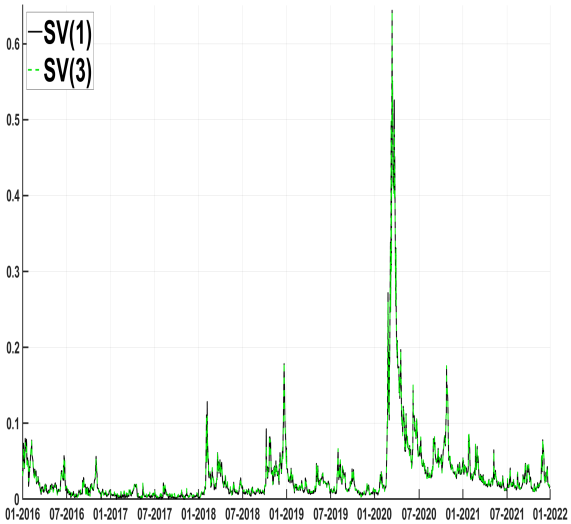
(b) SV(1) - SVDE



(c) SV(1) - SV(2)

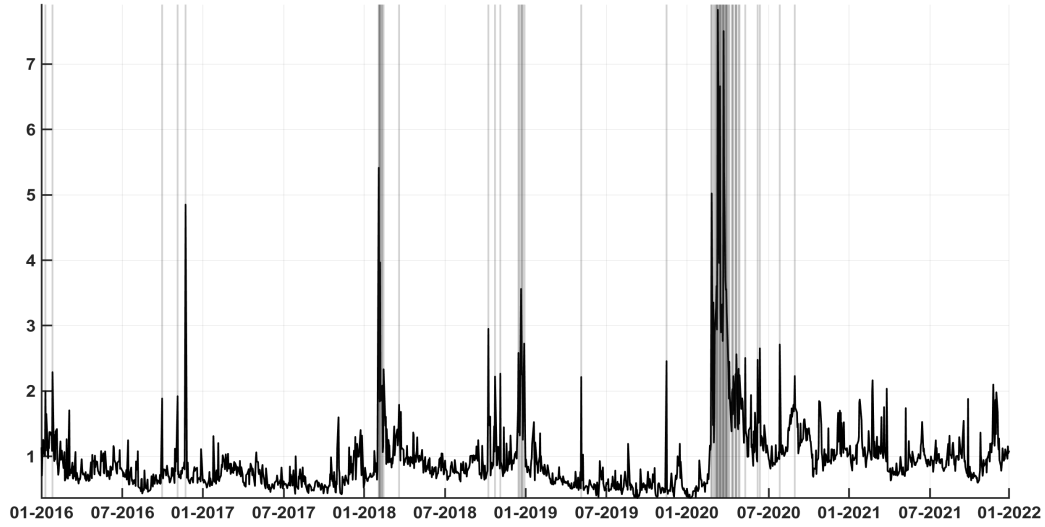


(d) SV(1) - SV(3)



Notes: We show the spot variance of the option pricing models computed on a daily basis. For multi-factor models, the spot variance is given by the sum of the latent factors.

Figure 4: Daily IVRMSE for the SV(1) model.



Notes: Daily IVRMSEs computed for day l as $100\sqrt{\sum_{i=1}^{N_l}\left(IV(O_{i,l})-\widehat{IV}(O_{i,l})\right)^2/N_l}$. Detected abnormal surface dates are displayed using vertical gray bars.

Table 2: Number of abnormal surfaces per year for various models.

Model	2016	2017	2018	2019	2020	2021
BS	1	0	20	0	65	2
AHBS	1	0	3	0	14	0
ANN(3)	1	1	5	2	23	0
OptionMetrics	1	0	6	1	27	1
Random Forest	1	0	6	1	23	1
SV(1)	5	0	18	2	39	0
SV(2)	3	1	18	2	33	0
SV(3)	5	1	18	2	35	0
SVJR	4	1	19	2	28	0
SVDE	5	1	18	2	26	1

For each year, we count the number of in-sample mean squared errors that are larger than the outlier threshold (see Section 3.4).

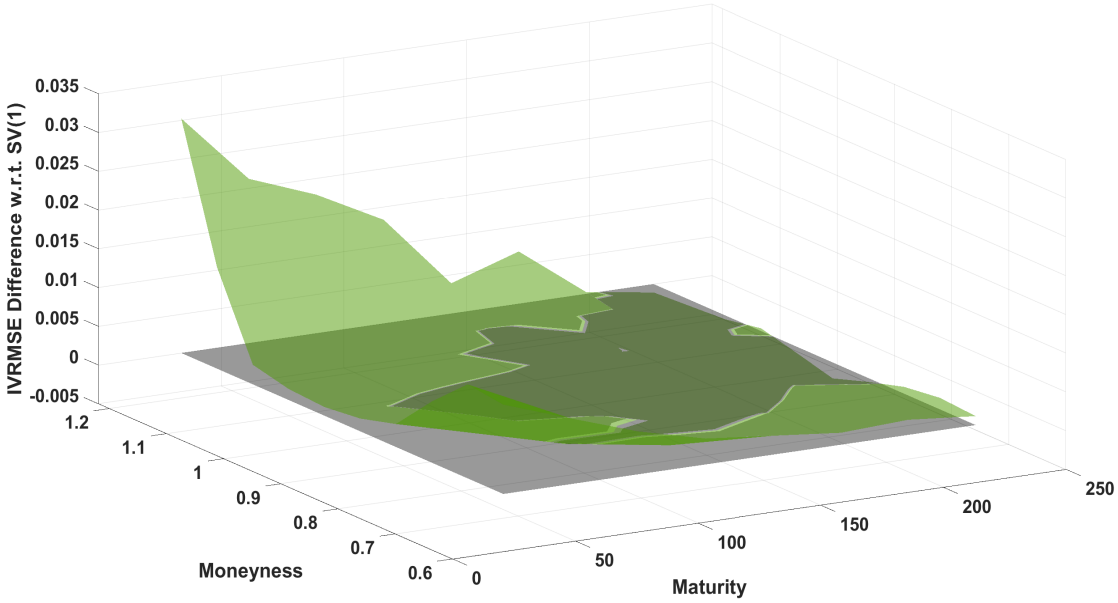
model enriched with double exponential jumps SVDE. First, the 3-dimensional figure in Panel (a) plots the difference in IVRMSE fit between SV(1) and SVDE for all days between 2016 and 2021 as a function of moneyness and maturity. We can clearly see that the jump model better fit the option surface for high moneyness levels for most maturities but also to a lesser extent for low moneyness levels. Jumps contribute little for at-the-money options, especially for longer maturities which is in line with the findings of the existing option pricing literature. Second, Panel (b) of Figure 5 displays the daily IVRMSE overall fit ratio between SV(1) and SVDE between 2016 and 2021. Apart from a few days, the ratio is above one and when abnormal IV surfaces are detected, the IVRMSE ratio can be twice as high for the SV(1) model. This confirms again that jumps can better adjust to sudden changes in the IV surface. The performance of the SV(1) model deteriorates post COVID-19 when we observe IVRMSE ratios frequently above two.

Having discussed the daily surface fit, we next turn to the dynamic modelling of the surface which is used for forecasting. The SHAR type surface model has four forecast horizon dependent parameters. Table 3 illustrates the magnitude of these parameters when estimated at the last available day in 2021, i.e. the full sample period, for the OptionMetrics model to fit the daily surface. The one and five day ahead surface is highly related to the most recent available surface with parameter estimates for $\hat{\beta}^{(1)}_{h|t}$ of about 0.8. The one day ahead surface is impacted by the average surface of the most recent week with $\hat{\beta}^{(2)}_{1|t} = 0.18$, but not by the average surface over the most recent month ($\hat{\beta}^{(3)}_{1|t} = 0$). Instead, for the twenty day ahead horizon, this average surface has a significant impact with $\hat{\beta}^{(3)}_{20|t} = 0.35$ while $\hat{\beta}^{(1)}_{20|t} = 0.69$ remains high which compensated by a negative $\hat{\beta}^{(2)}_{20|t}$ estimate.[9] Overall, similar parameter estimates are obtained for other surface fit models than the OptionMetrics approach which we omit here to save space.

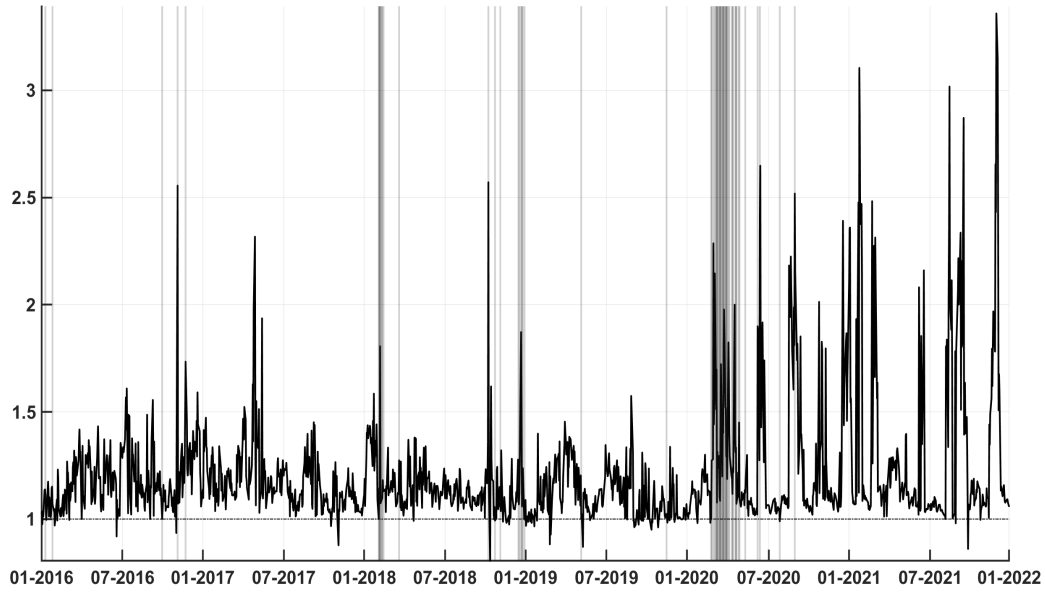
[9]Unlike the HAR model, the SHAR model relates current and past IV surfaces in a nonlinear way and therefore the sufficient stationarity condition of the sum of parameters smaller than one does not necessarily apply here.

Figure 5: Daily IVRMSE fit - SV(1) versus SVDE model

(a) Difference between SV(1) and SVDE for all days



(b) Daily SV(1)/SVDE IVRMSE fits



Notes: We show the IVRMSE computed per bucket.

Table 3: Parameter estimates of the SHAR model for the OptionMetrics method.

Horizon (h)	$\hat{\beta}_{h\|t}^{(0)}$	$\hat{\beta}_{h\|t}^{(1)}$	$\hat{\beta}_{h\|t}^{(2)}$	$\hat{\beta}_{h\|t}^{(3)}$
1	0.00	0.81	0.18	0.00
5	0.01	0.82	0.08	0.06
20	0.03	0.69	-0.18	0.35

Estimated parameters of the SHAR model over the full sample period.

Figure 6 shows for the OptionMetrics approach the sequentially fitted parameters of the SHAR (black) and SHAR-Robust (dashed green) model related to the most recent available surface, $\hat{\beta}_{h|t}^{(1)}$, for one day (panel a) and one month (panel b) ahead horizons. Regarding the SHAR model in Panel (a), the parameter estimate $\hat{\beta}_{1|t}^{(1)}$ is just below one early 2016, slightly declines until early 2018 where it drops below 0.9. Then it stays almost constant until the onset of the COVID-19 crisis when it drops to around 0.8 and where it stays until the end of 2021. The SHAR-Robust parameter estimates are close to the SHAR parameter estimates until the COVID-19 crisis when the multitude of abnormal surfaces heavily impact the SHAR estimates. Regarding Panel (b) of Figure 6, the parameter estimate $\hat{\beta}_{20|t}^{(1)}$ of the SHAR model, the parameter estimates hover around 0.5 between 2016 and 2018, then drop to just above 0.4, and jump to values around 0.7 from COVID-19 crisis start (March 2020) onwards. The SHAR-Robust estimates are similar until March 2020 when they increase until just below 0.6 until the end of 2021. These differences in parameter estimates impact the forecasts of the IV surfaces as we document in the next section.

6 Surface forecasting results

In this section, we report the forecasting performance of our approach and do two robustness checks. Section 6.1 discusses the performance for the different surface fit models at several forecast horizons. Section 6.2 investigates if the error correction method of Almeida et al. (2023) still provides improvements when applied on top of our approach. Section 6.3

Figure 6: Parameter estimate over time for the OptionMetrics model

(a) $\hat{\beta}^{(1)}_{1|t}$



(b) $\hat{\beta}^{(1)}_{20|t}$



Notes: At each day t, we estimate the SHAR linear regression parameters using all options from 2015 up to day t and predicted surfaces up to $t-h$, where the horizon is set to $h=1$ for this graph. The graphics show the parameter estimates related to the lagged predicted IV surface over time. The detected outliers are displayed using vertical gray bars.

compares our forecasting approach against an approach proposed by Goncalves and Guidolin (2006) in which the dynamics of the IV surfaces are modelled through the parameters of the surface fit models rather than through the surfaces themselves, an approach that works for unconstrained and in a low dimensional parameter settings only.

6.1 Main results

We start the forecasting exercise in 2016 and run it until the end of 2021. Every day, a new IV surface becomes available, it is fitted with each of the models we consider, for each horizon the sequential estimator for the dynamic surface SHAR model is updated, and forecasts are produced. Forecast performance is evaluated by the implied volatility root-mean-square-error (IVRMSE) defined from day 1 to t as $100\sqrt{\sum_{l=1}^{t}\sum_{i=1}^{N_l}\left(IV(O_{i,l}) - \widehat{IV}(O_{i,l}\right)^2 / \sum_{l=1}^{t} N_l}$ and reported per year to better understand the performance of our approach over time.

Table 4 summarises the results for one day ahead IV surface forecasts. We focus first on the left part of the table *All periods* which evaluates the forecast performance for all days in each year, including the detected abnormal surface days. For each of the considered models to fit the daily IV surface, we compare the random walk (RW) with the SHAR and SHAR-Robust (SHAR-R.) approaches. For the year 2016, the SHAR approaches outperform RW except for the ANN and random forest machine learning type models. The differences are small though, for example in the case of the SV(3) option pricing model, RW and SHAR have IVRMSEs of 1.29 and 1.27 respectively, amounting to almost two percent. After 2016, the differences between RW and SHAR approaches are more pronounced. For example in 2018, the difference in random forest IVRMSEs are 1.81 and 1.68 respectively, which is almost a difference of eight percent. Comparing the SHAR and SHAR-Robust approaches, it turns out that the latter typically reduces IVRMSEs meaning that updating the sequential estimator only with normal IV surfaces is beneficial.

Among the considered surface fit models, machine learning approaches and the nonpara-

metric kernel OptionMetrics method yield the lowest IVRMSEs. For example in 2021 and the SHAR-Robust approach, ANN(3) and the best option pricing model SVDE have IVRMSEs equal to 1.45 and 1.57, respectively. Among the option pricing models, while the SV(2) model which adds an additional stochastic volatility factor to the SV(1) adds value systematically, the additional gain of a third factor is tiny. However, the addition of jumps, in particular double exponential jumps, outperforms multi-factor stochastic volatility models for 2018, 2020 and 2021.

We next focus next on the right panel *Normal periods* of Table 4 which reports the forecast performance leaving out the detected abnormal IV surface days. As expected given the evidence in Table 2, this lowers the IVRMSEs in the years 2018 and 2020, but the overall conclusions described before remain. For example for the SV(1) option pricing model in 2020, the IVRMSE for SHAR-Robust drops from 3.89 to 2.01 when evaluating forecast performance without abnormal IV surfaces. Overall, these results show that years with higher overall implied volatility levels are more difficult to forecast.

Table 5 reports the results for one week ahead IV surface forecasts. Given the longer horizon, these surfaces are more difficult to forecast and therefore the IVRMSEs are higher than the next day surface forecasts in Table 4. For the year 2016 in the left *All periods* panel, the SHAR approaches now systematically outperform RW, with in particular the SHAR-Robust frequently having a slightly lower IVRMSE than SHAR. The random forest approach yields the lowest IVRMSE of 2.16. The neural network ANN(3), multi-factor stochastic volatility and jump models have an IVRMSE slightly higher than 2.20. After 2016, the SHAR approaches continue to dominate except for the year 2020 when RW performs better for option pricing models, AHBS, and the OptionMetrics estimator. Leaving out the abnormal IV surfaces in the forecast evaluation in the right *Normal periods* panel, causes significant drops in IVRMSE for the years 2018 and 2020 but the conclusion that the SHAR approaches dominate RW is the same. We refer to Table A.1 in the Appendix for one month

Table 4: Out-of-sample IVRMSE of option pricing models over time (horizon 1).

	All periods						**Normal periods**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
					Model - BS						*Model - BS*	
RW	6.94	6.24	8.05	6.46	**9.97**	9.44	6.94	6.24	7.97	6.46	**9.55**	9.44
SHAR	**6.94**	6.24	**8.04**	6.45	9.99	9.44	**6.94**	6.24	7.97	6.45	9.57	9.44
SHAR-R.	6.94	**6.24**	8.04	**6.45**	9.98	**9.44**	6.94	**6.24**	**7.97**	**6.45**	9.56	**9.44**
					Model - AHBS						*Model - AHBS*	
RW	1.52	1.35	2.23	1.52	3.89	2.01	1.52	1.35	2.02	1.52	3.33	2.01
SHAR	**1.52**	1.35	**2.22**	**1.51**	3.93	2.00	**1.52**	1.35	**2.01**	**1.51**	3.41	2.00
SHAR-R.	1.53	**1.34**	2.24	1.52	**3.64**	**1.99**	1.53	**1.34**	2.02	1.52	**3.20**	**1.99**
					Model - ANN(3)						*Model - ANN(3)*	
RW	**1.30**	**0.94**	1.91	1.26	3.60	1.49	**1.29**	**0.93**	1.67	1.26	3.11	1.49
SHAR	1.41	1.01	**1.78**	**1.19**	3.17	1.48	1.41	1.00	**1.53**	**1.19**	**2.70**	1.47
SHAR-R.	1.41	1.01	1.78	1.19	**3.12**	**1.47**	1.41	1.00	1.53	1.19	2.72	**1.46**
					Model - OptionMetrics						*Model - OptionMetrics*	
RW	1.21	0.85	1.79	1.16	**3.51**	1.48	1.21	0.85	1.49	1.14	2.94	1.48
SHAR	**1.21**	0.83	**1.77**	**1.15**	3.55	**1.46**	**1.20**	0.83	**1.46**	**1.14**	2.95	**1.46**
SHAR-R.	1.21	**0.83**	1.78	1.15	3.59	1.49	1.20	**0.83**	1.49	1.14	**2.76**	1.48
					Model - Random Forest						*Model - Random Forest*	
RW	**1.25**	**0.86**	1.81	1.18	3.58	1.49	**1.24**	**0.86**	1.52	1.16	3.04	1.49
SHAR	1.33	0.96	1.68	1.13	3.15	1.46	1.33	0.96	1.40	1.12	**2.65**	1.45
SHAR-R.	1.33	0.96	**1.68**	**1.13**	**3.11**	**1.45**	1.33	0.96	**1.39**	**1.12**	2.68	**1.45**
					Option pricing model - SV(1)						*Option pricing model - SV(1)*	
RW	1.41	0.97	2.01	1.22	3.95	1.73	1.38	0.97	1.49	**1.22**	2.11	1.73
SHAR	1.38	0.95	1.96	1.22	3.95	1.71	1.36	0.95	1.44	1.22	2.08	1.71
SHAR-R.	**1.36**	**0.94**	**1.95**	**1.22**	**3.89**	**1.70**	**1.34**	**0.94**	**1.43**	1.22	**2.01**	**1.70**
					Option pricing model - SV(2)						*Option pricing model - SV(2)*	
RW	1.29	0.89	1.96	1.18	3.81	1.69	1.25	0.89	1.44	1.18	2.16	1.69
SHAR	1.27	0.87	1.92	1.18	3.83	1.68	1.23	0.87	**1.40**	1.18	2.15	1.68
SHAR-R.	**1.27**	**0.86**	**1.91**	**1.18**	**3.74**	**1.67**	**1.23**	**0.86**	1.41	**1.18**	**2.08**	**1.67**
					Option pricing model - SV(3)						*Option pricing model - SV(3)*	
RW	1.28	0.88	1.95	1.18	3.81	1.69	1.25	0.88	1.45	1.18	2.16	1.69
SHAR	1.27	0.87	1.92	1.18	3.83	1.68	1.23	0.87	**1.41**	1.18	2.15	1.68
SHAR-R.	**1.27**	**0.85**	**1.91**	**1.18**	**3.71**	**1.67**	**1.23**	**0.85**	1.41	**1.18**	**2.09**	**1.67**
					Option pricing model - SVJR						*Option pricing model - SVJR*	
RW	1.36	0.93	1.97	1.21	3.75	1.68	1.35	0.92	1.45	1.21	2.28	1.68
SHAR	1.34	0.91	**1.93**	1.20	3.77	1.66	1.33	0.91	**1.41**	1.21	2.27	1.66
SHAR-R.	**1.33**	**0.89**	1.94	**1.20**	**3.66**	**1.65**	**1.32**	**0.89**	1.43	**1.20**	**2.23**	**1.65**
					Option pricing model - SVDE						*Option pricing model - SVDE*	
RW	1.32	0.91	1.93	1.19	3.71	1.61	1.29	0.91	1.45	1.19	2.23	1.61
SHAR	1.31	0.89	1.90	1.19	3.73	1.59	1.27	0.89	**1.41**	1.19	2.25	1.59
SHAR-R.	**1.31**	**0.88**	**1.89**	**1.19**	**3.57**	**1.57**	**1.27**	**0.88**	1.42	**1.19**	**2.20**	**1.58**

For each day, we estimate each model using the options of the day and we forecast the implied volatility surface of the day after. Note that the option pricing models are trained using the vega loss function.

ahead IV surface forecasts, which reports similar findings.

Figure 7 plots the IVRMSE surface aggregated from 2016 to 2021 implied by the RW (blue) and SHAR (orange) for the SV(1) and SVDE option pricing models in Panel (a) and (b) respectively. For the SV(1) model in Panel (a), the surface is mostly blue outside the centre of the IV surface, meaning that the SHAR IVRMSEs are lower for moneyness levels far from one and for all maturities. For moneyness levels between 0.8 and one, there are spots where both SHAR and RW perform similarly, and some small areas where RW is best. The picture in Panel (b) for the SVDE model is slightly different. The SHAR approach dominates as before in large parts of the surface, with equal performance for RW in the moneyness levels ranging between 0.6 and 0.7. Note that irrespective of the surface fit model or forecasting method, a large part of the total IVRMSE is due to the difficulty in forecasting IVs for short maturity large moneyness options.

Table 6 reports the average computing time in seconds to fit one daily IV surface with the different models we consider. The SHAR computing time is not incorporated because the model parameters are estimated sequentially in closed form, and the estimator is identical for all surface fit models. Taking into account the forecast performance discussed above, the non-parametric model OptionMetrics and random forest are very fast compared to the neural network and especially to the option pricing models. For example, for random forest it takes about 3 seconds to fit the IV surface, compared to respectively 20 seconds and 4 minutes for the ANN(3) and SVDE models. For the implementation of high frequency, e.g. five minute, streaming IV surface forecasting, the option pricing models are prohibitively slow.

6.2 Enhancing surface fit with deep learning

Our proposed methodology can be thought as a two-step procedure. The first step involves fitting an IV model on a daily basis. The second step adds dynamics by forecasting

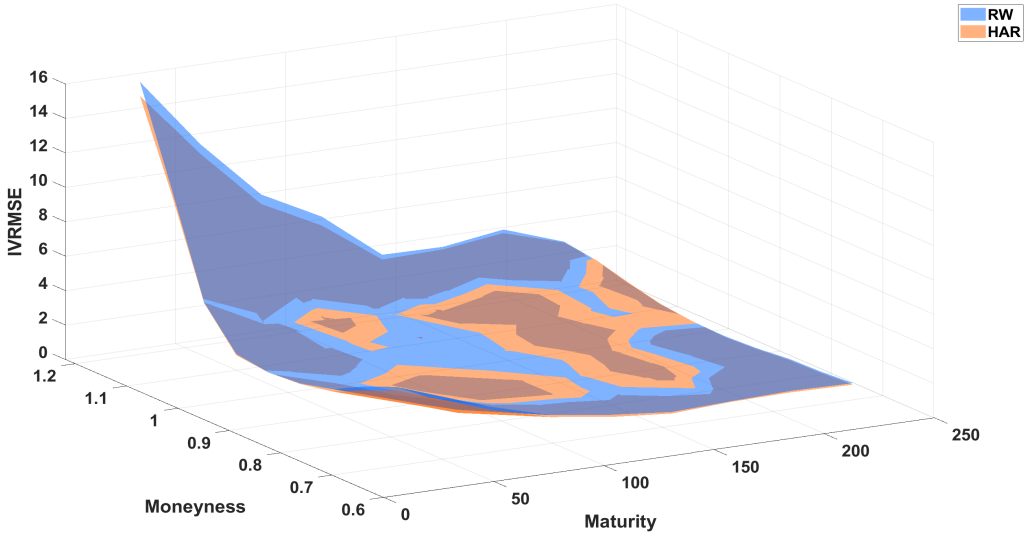
Table 5: Out-of-sample IVRMSE of option pricing models over time (horizon 5).

	All periods						**Normal periods**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
					Model - BS						*Model - BS*	
RW	7.12	**6.31**	8.31	6.63	**10.90**	9.58	7.12	**6.31**	8.20	6.63	**10.10**	9.58
SHAR	7.12	6.33	**8.31**	**6.59**	11.23	9.57	7.12	6.33	**8.20**	**6.59**	10.27	9.57
SHAR-R.	**7.11**	6.31	8.31	6.59	11.05	**9.56**	**7.11**	6.31	8.20	6.59	10.19	**9.56**
					Model - AHBS						*Model - AHBS*	
RW	2.48	1.63	3.49	2.35	**6.66**	2.86	2.48	1.63	3.25	2.35	**5.32**	2.86
SHAR	2.44	1.61	**3.44**	2.27	7.28	2.85	2.44	1.61	**3.16**	2.27	5.65	2.85
SHAR-R.	**2.42**	**1.59**	3.45	**2.26**	7.02	**2.81**	**2.42**	**1.59**	3.18	**2.26**	5.62	**2.81**
					Model - ANN(3)						*Model - ANN(3)*	
RW	2.41	**1.35**	3.16	2.15	6.30	2.60	2.41	**1.35**	2.94	2.15	5.04	2.60
SHAR	2.23	1.42	**2.80**	1.93	5.86	2.27	2.23	1.42	**2.51**	1.92	**4.45**	2.28
SHAR-R.	**2.22**	1.42	2.81	**1.92**	**5.72**	**2.25**	**2.22**	1.42	2.51	**1.91**	4.55	**2.25**
					Model - OptionMetrics						*Model - OptionMetrics*	
RW	2.30	1.26	3.14	2.09	**6.25**	2.52	2.30	1.26	2.82	2.09	**4.75**	2.50
SHAR	2.25	**1.19**	3.05	2.03	6.90	2.48	2.25	**1.19**	2.67	2.03	5.03	2.47
SHAR-R.	**2.25**	1.19	**3.04**	**2.03**	6.75	**2.46**	**2.25**	1.19	**2.65**	**2.03**	5.15	**2.45**
					Model - Random Forest						*Model - Random Forest*	
RW	2.33	**1.27**	3.15	2.11	6.29	2.52	2.33	**1.27**	2.82	2.10	4.98	2.51
SHAR	**2.16**	1.38	**2.78**	1.89	5.83	2.20	**2.16**	1.38	2.38	1.89	**4.37**	2.20
SHAR-R.	2.16	1.38	2.78	**1.89**	**5.68**	**2.18**	2.16	1.38	**2.38**	**1.89**	4.48	**2.18**
					Option pricing model - SV(1)						*Option pricing model - SV(1)*	
RW	2.34	1.30	3.13	2.10	**6.43**	2.63	2.35	1.30	2.30	2.09	**3.41**	2.63
SHAR	2.27	1.24	**3.04**	2.03	6.95	2.60	2.27	1.24	2.15	2.03	3.42	2.60
SHAR-R.	**2.26**	**1.23**	3.05	**2.00**	6.86	**2.52**	**2.25**	**1.23**	**2.12**	**1.99**	3.50	**2.52**
					Option pricing model - SV(2)						*Option pricing model - SV(2)*	
RW	2.28	1.24	3.12	2.08	**6.32**	2.61	2.28	1.24	2.32	2.07	**3.47**	2.61
SHAR	2.21	1.19	**3.04**	2.01	6.91	2.58	2.20	1.19	2.18	2.00	3.50	2.58
SHAR-R.	**2.20**	**1.17**	3.05	**1.98**	6.74	**2.50**	**2.18**	**1.17**	**2.15**	**1.97**	3.56	**2.50**
					Option pricing model - SV(3)						*Option pricing model - SV(3)*	
RW	2.27	1.24	3.14	2.07	**6.32**	2.61	2.27	1.24	2.35	2.07	**3.51**	2.61
SHAR	2.21	1.18	**3.05**	2.01	6.92	2.57	2.20	1.19	2.20	2.00	3.55	2.57
SHAR-R.	**2.20**	**1.16**	3.07	**1.98**	6.71	**2.51**	**2.18**	**1.17**	**2.18**	**1.97**	3.67	**2.51**
					Option pricing model - SVJR						*Option pricing model - SVJR*	
RW	2.33	1.27	3.14	2.10	**6.32**	2.62	2.32	1.27	2.32	2.09	**3.72**	2.62
SHAR	2.25	1.22	**3.05**	2.02	6.91	2.57	2.24	1.22	2.17	2.02	3.75	2.57
SHAR-R.	**2.23**	**1.19**	3.07	**1.99**	6.70	**2.51**	**2.23**	**1.20**	**2.14**	**1.98**	3.88	**2.51**
					Option pricing model - SVDE						*Option pricing model - SVDE*	
RW	2.29	1.26	3.11	2.08	**6.26**	2.57	2.29	1.26	2.36	2.08	**3.73**	2.56
SHAR	2.22	1.21	**3.03**	2.01	6.86	2.53	2.21	1.21	2.22	2.00	3.74	2.52
SHAR-R.	**2.21**	**1.19**	3.04	**1.98**	6.63	**2.47**	**2.20**	**1.19**	**2.19**	**1.97**	3.88	**2.46**

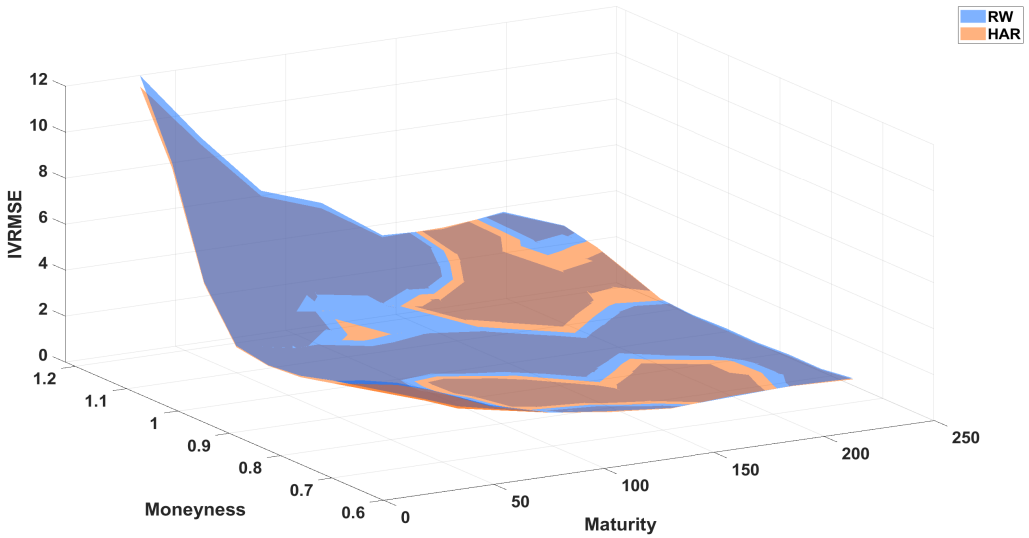
For each day, we estimate each model using the options of the day and we forecast the implied volatility surface of the day after. Note that the option pricing models are trained using the vega loss function.

Figure 7: IVRMSE surface - Comparison of the RW and the SHAR results for next day surface forecasts

(a) SV(1)



(b) SVDE



Notes: We plot the IVRMSE surface aggregated from 2016 to 2021 implied by the RW (blue) and SHAR (orange) for the SV(1) and SVDE models.

Table 6: Computing time for the option fit models.

Parametric Models		**Non-parametric Models**		**Option Pricing Models**			
AHBS	ANN(3)	Opt-Met	RF	SV(1)	SV(2)	SVJR	SVDE
0.21	19.7	1.94	2.71	90.92	220.71	249.52	244.26

This table reports the average time expressed in seconds for estimating the models on one day of options. We compute the average over daily options spanning from 2015 to 2021. Opt-Met and RF stand for OptionMetrics and random forest, respectively.

IVs using the SHAR model idea. An alternative two-step approach for enhancing the fit of the IV surface has been proposed by Almeida et al. (2023). It proceeds as follows:

1. Given a model, estimate the parameters Θ_t given the IV surface at day t, and compute the model residuals: $\hat{\varepsilon}_{it} = IV(O_{i,t}) - IV^M(O_{i,t}, \hat{\Theta}_t)$ for $i = 1, \ldots, N_t$.

2. Estimate an ANN on $\hat{\varepsilon}_{it}$ using the option characteristics $O_{i,t}$ as inputs, denoted as $\text{ANN}(O_{i,t})$, and obtain the h-step ahead forecast as $\hat{IV}(O_{i,t+h}) = IV^M(O_{i,t+h}, \hat{\Theta}_t) + \text{ANN}(O_{i,t+h})$.

Interestingly, this two-step procedure can be considered as another IV model within our context. In fact, our SHAR model, which extends the daily surface fits to account for temporal dynamics, can be applied in conjunction with this approach as a third step. We investigate the complementarity of these approaches, and assess whether the IVRMSE reductions differ from the non-linearities captured by the ANN applied to the residuals. The neural network step has been implemented using an ANN(3) consisting of 3 hidden layers with 32, 16 and 8 neurons, each activated by sigmoid function.

Table 7 shows the annual one-step ahead IVRMSEs for various models for a one-step ahead prediction given by the RW approach, the two-step neural network error correction procedure described here, and our SHAR approach applied in conjunction with this two-step procedure. There are two important findings. First, the results confirm the findings of Almeida et al. (2023) since the two-step approach improves the predictive performance of any model for each year. Even the advanced option pricing models such as the SV(2)

and the SVDE models show significant improvements in predictive accuracy when combined with ANN(3). Second, the three-step procedure or the dynamic modelling of the surfaces with the SHAR approach lowers further the IVRMSEs. For each year and all the models except the AHBS model in 2016, the SHAR approach yields additional gains. We conclude that both approaches lead to IVRMSE improvements by capturing different stylized facts, i.e. additional non-linearities for the ANN model and the temporal persistence for the SHAR process. This conclusion is strengthened given the five day ahead IV surface forecast results in Table 8. In fact, the gains from dynamic surface modelling on top of deep learning combined with a surface fit model are larger than the next day ahead surface forecast results, except for the year 2020.

6.3 Forecasting surfaces versus forecasting parameters

In this paper, we show that IV surfaces, when fitted daily and used as predictors in the SHAR model, significantly outperform those forecasted by the RW approach. As explained above, the RW forecasting approach is based on the following model:

$$IV(O_{i,t+h}) \quad = \quad IV^M(O_{i,t+h}, \hat{\Theta}_t) + \varepsilon_{i,t+h}. \tag{15}$$

Note that from (1) the optimal model prediction is $IV^M(O_{i,t+h}, \hat{\Theta}_{t+h})$. However, since at day t we do not have the information to compute $\hat{\Theta}_{t+h}$, the RW approach defaults to using the last estimated values, i.e. $\hat{\Theta}_{t+h} = \hat{\Theta}_t$. Since the information set at time t includes all the parameter estimates $\hat{\Theta}_{1:t} = \{\hat{\Theta}_1, \ldots, \hat{\Theta}_t\}$, the RW model can also be understood as a RW process applied to the model parameters themselves, i.e. $\Theta_{t+h} = \Theta_t + \eta_{t+h}$. Goncalves and Guidolin (2006) relax this RW assumption on the model parameters and suggest forecasting the model parameters using a VAR(p) model: $\hat{\Theta}_{t+h} = \gamma_0 + \sum_{i=1}^{p} \Gamma_i \hat{\Theta}_{t+1-i} + \eta_{t+h}$, where the number of lags p is determined using the Bayesian Information Criterion (BIC).

Table 7: IVRMSEs for the two and three step procedures (horizon 1).

Model	2016	2017	2018	2019	2020	2021
			Model - AHBS			
RW	1.52	1.35	2.23	1.52	3.89	2.01
ANN(3)+RW	**1.51**	1.32	2.06	1.46	3.74	1.75
ANN(3)+SHAR	1.51	1.31	**2.04**	**1.45**	3.77	1.73
ANN(3)+SHAR-R.	1.52	**1.30**	2.06	1.45	**3.49**	**1.72**
			Model - SV(1)			
RW	1.41	0.97	2.01	1.22	3.95	1.73
ANN(3)+RW	1.34	0.94	1.89	1.22	3.67	1.65
ANN(3)+SHAR	**1.33**	0.92	1.85	**1.21**	3.67	1.63
ANN(3)+SHAR-R.	1.33	**0.92**	**1.85**	1.21	**3.61**	**1.62**
			Model - SV(2)			
RW	1.31	0.90	1.97	1.20	3.81	1.70
ANN(3)+RW	1.29	0.86	1.92	1.19	3.65	1.65
ANN(3)+SHAR	**1.28**	0.85	**1.88**	**1.18**	3.66	1.62
ANN(3)+SHAR-R.	1.28	**0.84**	1.88	1.19	**3.50**	**1.61**
			Model - SVDE			
RW	1.32	0.91	1.93	1.19	3.71	1.61
ANN(3)+RW	1.29	0.87	1.86	1.16	3.60	1.58
ANN(3)+SHAR	**1.28**	0.85	1.82	**1.15**	3.61	1.55
ANN(3)+SHAR-R.	1.30	**0.85**	**1.81**	1.16	**3.45**	**1.55**
			Model - SVJR			
RW	1.36	0.93	1.97	1.21	3.75	1.68
ANN(3)+RW	1.29	0.90	1.89	1.20	3.66	1.63
ANN(3)+SHAR	**1.28**	0.89	**1.86**	**1.18**	3.67	1.60
ANN(3)+SHAR-R.	1.29	**0.88**	1.88	1.19	**3.50**	**1.59**

This table reports IVRMSEs aggregated per year. RW means fitting the model daily and using it as the next day forecast. ANN(3)+RW means fitting a model daily, training a neural network on the implied errors, and the next day forecast is model fit plus error fit. ANN(3)+SHAR (SHAR-R.) means fitting a model daily, training a neural network on the implied errors, and estimating the SHAR (SHAR-R.) type model to produce the next day forecast.

Table 8: IVRMSEs for the two and three step procedures (horizon 5).

Model	2016	2017	2018	2019	2020	2021
			Model - AHBS			
RW	2.48	1.63	3.49	2.35	6.66	2.88
ANN(3)+RW	2.46	1.60	3.38	2.29	**6.51**	2.69
ANN(3)+SHAR	2.42	1.57	**3.33**	2.21	7.16	2.68
ANN(3)+SHAR-R.	**2.41**	**1.55**	3.34	**2.20**	6.90	**2.63**
			Model - SV(1)			
RW	2.34	1.30	3.13	2.10	6.43	2.63
ANN(3)+RW	2.29	1.27	3.07	2.12	**6.35**	2.60
ANN(3)+SHAR	2.22	1.22	**2.99**	2.02	6.92	2.53
ANN(3)+SHAR-R.	**2.22**	**1.21**	2.99	**2.00**	6.71	**2.47**
			Model - SVDE			
RW	2.29	1.26	3.11	2.08	6.26	2.57
ANN(3)+RW	2.27	1.22	3.07	2.09	**6.26**	2.57
ANN(3)+SHAR	2.21	1.18	**3.00**	2.00	6.85	2.51
ANN(3)+SHAR-R.	**2.20**	**1.16**	3.01	**1.98**	6.59	**2.46**
			Model - SVJR			
RW	2.33	1.27	3.14	2.10	6.32	2.62
ANN(3)+RW	2.30	1.22	3.10	2.11	**6.31**	2.59
ANN(3)+SHAR	2.23	1.18	**3.02**	2.02	6.90	2.53
ANN(3)+SHAR-R.	**2.22**	**1.16**	3.04	**2.00**	6.64	**2.48**

This table reports IVRMSEs aggregated per year. RW means fitting the model daily and using it as the next day forecast. ANN(3)+RW means fitting a model daily, training a neural network on the implied errors, and the next day forecast is model fit plus error fit. ANN(3)+SHAR (SHAR-R.) means fitting a model daily, training a neural network on the implied errors, and estimating the SHAR (SHAR-R.) type model to produce the next day forecast.

Consequently, the forecast of the IV surface at day t for horizon h is given by:

$$\widehat{IV}(O_{i,t+h}) = IV^M\big(O_{i,t+h}, \mathbb{E}(\Theta_{t+h}|\hat{\Theta}_t, \ldots, \hat{\Theta}_1)\big). \tag{16}$$

Our SHAR approach offers two advantages over fitting a time series process to model parameters. First, our approach is more versatile, as it is applicable across a wide range of IV surface models. Specifically, the VAR method requires predicting the model parameters, which is not feasible for highly parametrized models like neural networks or for non-parametric methods such as the random forest. The VAR method is also difficult to apply to option pricing models as several of the parameters are constrained within compact supports (e.g. leverage correlations). Second, the conditional expectation from the VAR model does not provide the optimal prediction for non-linear IV models in terms of mean squared errors, since $\mathbb{E}(IV^M\big(O_{i,t+h}, \Theta_{t+h}\big)|\hat{\Theta}_t, \ldots, \hat{\Theta}_1) \neq IV^M\big(O_{i,t+h}, \mathbb{E}(\Theta_{t+h}|\hat{\Theta}_t, \ldots, \hat{\Theta}_1)\big)$. While these limitations make the VAR approach impractical for most IV models, it remains suitable for the AHBS model, which is linear and not highly parametrized.

Table 9 shows IVRMSEs of the VAR approach for the AHBS model across various forecast horizons. For ease of comparison, we also replicate the RMSE performance of the RW and the SHAR methods. The results indicate that the VAR(p) method leads to out-of-sample improvements in only 5 out of 18 instances. However, this approach performs particularly well in 2019, when the differences in IVRMSEs are substantial for 5 and 20 day ahead horizons. However, in certain years, such as 2017 and 2018, the difference in favor of the SHAR model is notable, especially at horizons 5 and 20. Note that the SHAR method has substantially better IVRMSE performance than the AHBS model when using the more sophisticated models, see Tables 4 to A.1 above.

The VAR method only works when the modelled parameter estimates exhibit persistence. Figure 8 illustrates a limitation of this method by displaying daily estimates of the leverage correlation coefficient ρ for the SV(1) model. First, the parameters are confined within

Table 9: IVRMSE of AHBS predicting the model parameters with a VAR(p) process.

Model	2016	2017	2018	2019	2020	2021
			Horizon - 1			
RW	1.52	1.35	2.23	1.52	3.89	2.01
SHAR	**1.52**	**1.35**	**2.22**	1.51	3.93	**2.00**
VAR(p)	1.54	1.35	2.23	**1.50**	**3.89**	2.03
			Horizon - 5			
RW	2.48	1.63	3.49	2.35	**6.66**	2.86
SHAR	**2.44**	**1.61**	**3.44**	2.27	7.28	**2.85**
VAR(p)	2.56	1.73	3.49	**2.18**	7.59	2.87
			Horizon - 20			
RW	3.36	**1.87**	4.97	3.46	13.06	**3.64**
SHAR	3.44	1.91	**4.64**	3.21	**12.36**	3.75
VAR(p)	**3.33**	2.71	4.92	**2.81**	16.63	3.73

For each day, we estimate the AHBS model using the options of the day and we forecast the implied volatility surface of the day by predicting the AHBS parameter using a VAR(p) model. The number of lags is determined using the BIC with a maximum number equal to 6.

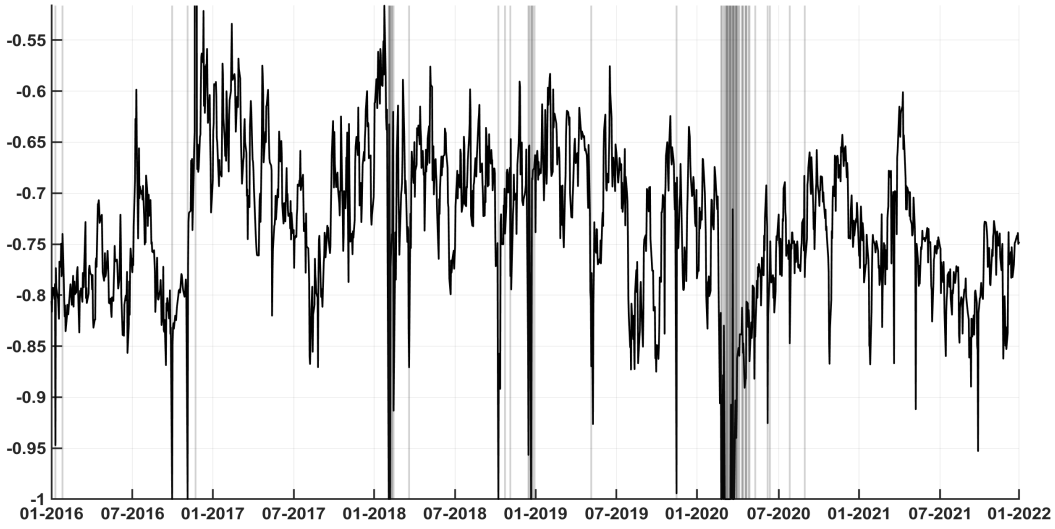
the bounded support $(-1, 1)$, potentially complicating their forecasting with a VAR model. Second, daily estimations do not necessarily guarantee persistence over time. Specifically, significant fluctuations of the parameter estimates can occur, particularly when abnormal option surfaces happen or limited information is available. For instance, the estimate for ρ is close to -1 on specific dates, leading to a significant breakdown in persistence of the estimated parameter series.

7 Conclusion

The option IV surface has been extensively studied in the literature. Most of the focus has been on fitting the surface using option pricing models, nonparametric and, more recently, machine learning approaches. While the fit of some of these models is impressive, their comparative performance in forecasting future IV surfaces remains less explored.

Given that IV surfaces are subject to daily changes due to fluctuations in economic

Figure 8: Daily estimates of ρ from the SV(1) model.



Notes: At each day t, we estimate the SV(1) model using daily options and we plot the leverage correlation parameter estimates. The detected outliers are displayed using vertical gray bars.

conditions as well as in the moneyness and maturity of available option contracts, a trend that has significantly increased in recent years, most models are fitted on a daily basis. This fit is then used as forecast for any future horizon surface, similar to a random-walk model. While this approach is practical, this paper puts forward the Surface HAR model which includes a time-dynamic step to better exploit all historically available IV surface information. This requires minimal retrofitting of IV surfaces and allows the model to be estimated sequentially in closed form using ordinary least squares.

We test our approach on the S&P 500 equity-index implied volatility option surfaces from 2016 to 2021 and find strong evidence that relying only on the last available day's IV surface typically yields inferior forecasts. Given the occasional sharp shifts in the IV surface from one day to the next, we introduce a robust version of the model to mitigate the impact of such abnormal surfaces on model estimates and subsequent surface forecasts. We show that our robust model version also provides excellent forecast performance.

Our proposed time-dynamic approach is based on the HAR process, popular in the realized volatility literature. We are convinced that more recent contributions from this literature could lead to better exploiting historical IV surface information. For example, our approach could potentially be extended with a joint modeling of returns and realized variances (e.g, Hansen et al., 2012; Hansen and Huang, 2016), time-varying parameters (e.g., Dufays and Rombouts, 2018; Bekierman and Manner, 2018) and the integration of higher order moments (Bollerslev et al., 2016; Cipollini et al., 2021). Another line of research, given the high trading activity, is extending the impact of the IV surface forecasts when including also short and ultrashort maturities.

References

Ackerer, Damien, Natasa Tagasovska, and Thibault Vatter, 2020, Deep smoothing of the implied volatility surface, in H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, eds., *Advances in Neural Information Processing Systems*, volume 33, 11552–11563 (Curran Associates, Inc.).

Almeida, Caio, Jianqing Fan, Gustavo Freire, and Francesca Tang, 2023, Can a machine correct option pricing models?, *Journal of Business & Economic Statistics* 41, 995–1009.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015, The risk premia embedded in index options, *Journal of Financial Economics* 117, 558–584.

Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3–30.

Bandi, Federico M., Nicola Fusari, and Roberto Ren, 2023, 0dte option pricing, Research Paper 2023-03, ESSEC Business School.

Bates, David S, 2000, Post-87 crash fears in the s&p 500 futures option market, *Journal of Econometrics* 94, 181–238.

Bekierman, Jeremias, and Hans Manner, 2018, Forecasting realized variance measures using time-varying coefficient models, *International Journal of Forecasting* 34, 276–287.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg, 2016, Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics* 192, 1–18.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* 22, 4463–4492.

Breiman, L., 2001, Random forests, *Machine Learning* 45, 5–32.

Carr, Peter, and Liuren Wu, 2016, Analyzing volatility risk and risk premium in option contracts: A new theory, *Journal of Financial Economics* 120, 1–20.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* 55, 1914–1932.

Christoffersen, Peter, Kris Jacobs, and Chayawat Ornthanalai, 2010, Option valuation with long-run and short-run volatility components, *Journal of Financial Economics* 97, 100–119.

Cipollini, Fabrizio, Giampiero M. Gallo, and Edoardo Otranto, 2021, Realized volatility forecasting: Robustness to measurement errors, *International Journal of Forecasting* 37, 44–57.

Corsi, Fulvio, 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

Dufays, Arnaud, Kris Jacobs, and Jeroen Rombouts, 2024, Modeling higher moments and risk premiums for s&p 500 returns, *Working Paper* -, –.

Dufays, Arnaud, and Jeroen VK Rombouts, 2018, Sparse change-point har models for realized variance, *Econometric Reviews* 857–880.

Dumas, Bernard, Jeff Fleming, and Robert E. Whaley, 1998, Implied volatility functions: Empirical tests, *The Journal of Finance* 53, 2059–2106.

Efron, B., and T. Hastie, 2016, *Computer Age Statistical Inference* (Cambridge University Press, Cambridge).

Fengler, Matthias R, 2009, Arbitrage-free smoothing of the implied volatility surface, *Quantitative Finance* 9, 417–428.

Fengler, Matthias R., Wolfgang K. Hrdle, and Enno Mammen, 2007, A semiparametric factor model for implied volatility surface dynamics, *Journal of Financial Econometrics* 5, 189–218.

FIA, 2024, Etd tracker, https://www.fia.org/etd-tracker, Accessed: 2024-05-06.

Goncalves, Silvia, and Massimo Guidolin, 2006, Predictable dynamics in the s&p 500 index options implied volatility surface, *The Journal of Business* 79, 1591–1635.

Gruber, Peter H., Claudio Tebaldi, and Fabio Trojani, 2021, The Price of the Smile and Variance Risk Premia, *Management Science* 67, 4056–4074.

Hansen, Peter Reinhard, and Zhuo Huang, 2016, Exponential GARCH modeling with realized measures of volatility, *Journal of Business & Economic Statistics* 34, 269–287.

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek, 2012, Realized GARCH: a joint model for returns and realized measures of volatility, *Journal of Applied Econometrics* 27, 877–906.

Heston, Steven 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.

Kelly, Bryan T., Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu, 2023, Deep learning from implied volatility surfaces, *Swiss Finance Institute Research Paper* .

Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, lvaro Veiga, and Eduardo Zilberman, 2021, Forecasting inflation in a data-rich environment: The benefits of machine learning methods, *Journal of Business & Economic Statistics* 39, 98–119.

Medvedev, Nikita, and Zhiguang Wang, 2022, Multistep forecast of the implied volatility surface using deep learning, *Journal of Futures Markets* 42, 645–667.

OptionMetrics, 2022, *IvyDB US file and data reference manual, Version 5.2*, 1776 Broadway, Suite 1800, New York, NY 10019, Computer software manual.

Rousseeuw, Peter J, and Christophe Croux, 1993, Alternatives to the median absolute deviation, *Journal of the American Statistical association* 88, 1273–1283.

Seo, Sang Byung, and Jessica A. Wachter, 2019, Option prices in a model with stochastic disaster risk, *Management Science* 65, 3449–3469.

Shang, Han Lin, and Fearghal Kearney, 2022, Dynamic functional time-series forecasts of foreign exchange implied volatility surfaces, *International Journal of Forecasting* 38, 1025–1049.

Ulrich, Matthias, Lukas Zimmer, and Christian Merbecks, 2023, Implied volatility surfaces:

a comprehensive analysis using half a billion option prices, *Review of Derivatives Research* 26, 135–169.

Zhang, Jin E, and Yi Xiang, 2008, The implied volatility smirk, *Quantitative Finance* 8, 263–284.

Zhang, Wenyong, Lingfei Li, and Gongqiu Zhang, 2023, A two-step framework for arbitrage-free prediction of the implied volatility surface, *Quantitative Finance* 23, 21–34.

Appendix

Table A.1: Out-of-sample IVRMSE of option pricing models over time (horizon 20).

	All periods						Normal periods					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	7.49	**6.38**	8.88	6.98	14.29	9.75	7.49	**6.38**	8.71	6.98	12.64	9.75
HAR	7.47	6.62	**8.74**	**6.76**	**14.07**	**9.67**	7.46	6.62	**8.59**	**6.76**	**12.12**	**9.67**
HAR-R.	**7.45**	6.57	8.75	6.77	14.10	9.68	**7.44**	6.57	8.60	6.77	12.15	9.68
				Model - AHBS						*Model - AHBS*		
RW	**3.36**	**1.87**	4.97	3.46	13.06	**3.64**	**3.36**	**1.87**	4.69	3.46	10.93	**3.64**
HAR	3.44	1.91	**4.64**	3.21	**12.36**	3.75	3.45	1.91	**4.35**	3.21	**9.78**	3.75
HAR-R.	3.41	1.87	4.68	**3.20**	12.56	3.65	3.41	1.87	4.39	**3.20**	10.00	3.65
				Model - ANN(3)						*Model - ANN(3)*		
RW	3.31	1.71	4.51	3.33	12.26	3.36	3.31	1.71	4.07	3.34	8.97	3.36
HAR	3.30	1.69	**3.96**	2.82	10.74	2.97	3.30	1.69	**3.50**	2.83	7.42	2.97
HAR-R.	**3.29**	**1.69**	3.98	**2.81**	**10.66**	**2.93**	**3.29**	**1.69**	3.52	**2.81**	**7.19**	**2.93**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	**3.20**	1.55	4.54	3.24	12.24	3.33	**3.20**	1.55	4.15	3.22	8.69	3.33
HAR	3.30	1.51	**4.19**	3.04	**11.83**	3.32	3.30	1.51	**3.78**	3.03	**8.19**	3.33
HAR-R.	3.30	**1.51**	4.19	**3.01**	12.02	**3.22**	3.30	**1.51**	3.78	**3.01**	8.32	**3.22**
				Model - Random Forest						*Model - Random Forest*		
RW	3.23	**1.55**	4.54	3.26	12.28	3.34	3.23	**1.55**	4.16	3.24	9.34	3.34
HAR	3.21	1.67	**3.96**	2.78	10.76	2.97	3.21	1.67	**3.53**	2.78	7.84	2.98
HAR-R.	**3.20**	1.67	3.97	**2.77**	**10.65**	**2.92**	**3.20**	1.67	3.54	**2.77**	**7.59**	**2.92**
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	**3.19**	1.53	4.41	3.19	11.91	3.49	**3.19**	1.53	3.49	3.19	6.19	3.49
HAR	3.25	1.52	**4.10**	2.97	**11.63**	3.47	3.25	1.52	**3.12**	2.97	**5.78**	3.47
HAR-R.	3.23	**1.50**	4.14	**2.92**	11.74	**3.41**	3.23	**1.50**	3.14	**2.91**	5.83	**3.41**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	**3.14**	1.48	4.42	3.17	11.97	3.48	**3.10**	1.48	3.52	3.17	6.07	3.48
HAR	3.21	1.48	**4.11**	2.96	**11.63**	3.45	3.17	1.48	**3.14**	2.95	**5.81**	3.45
HAR-R.	3.17	**1.45**	4.16	**2.89**	11.79	**3.39**	3.12	**1.45**	3.16	**2.89**	5.93	**3.39**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	**3.15**	1.48	4.43	3.17	11.98	3.48	**3.11**	1.48	3.53	3.17	6.43	3.48
HAR	3.21	1.48	**4.12**	2.96	**11.63**	3.45	3.17	1.48	**3.16**	2.95	**5.99**	3.45
HAR-R.	3.18	**1.45**	4.17	**2.89**	11.79	**3.37**	3.13	**1.45**	3.18	**2.89**	6.08	**3.37**
				Option pricing model - SVJR						*Option pricing model - SVJR*		
RW	**3.18**	1.50	4.46	3.19	11.99	3.49	**3.19**	1.50	3.53	3.19	7.25	3.49
HAR	3.24	1.51	**4.13**	2.97	**11.63**	3.44	3.24	1.52	**3.15**	2.96	**6.11**	3.44
HAR-R.	3.21	**1.48**	4.18	**2.91**	11.79	**3.37**	3.21	**1.48**	3.17	**2.90**	6.19	**3.37**
				Option pricing model - SVDE						*Option pricing model - SVDE*		
RW	**3.15**	1.49	4.41	3.18	11.99	3.45	**3.11**	1.49	3.53	3.18	7.92	3.44
HAR	3.21	1.50	**4.10**	2.95	**11.61**	3.41	3.18	1.50	**3.15**	2.95	**6.68**	3.40
HAR-R.	3.18	**1.47**	4.15	**2.90**	11.77	**3.33**	3.13	**1.47**	3.18	**2.89**	6.75	**3.32**

For each day, we estimate each model using the options of the day and we forecast the implied volatility surface of the day after. Note that the option pricing models are trained using the vega loss function.