Amundi
Investment Solutions

Trust must be earned

Investment Institute

WORKING PAPER #189 | MARCH 2026

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

ESSEC
BUSINESS SCHOOL

ESSEC-Amundi Chair on Asset & Risk Management

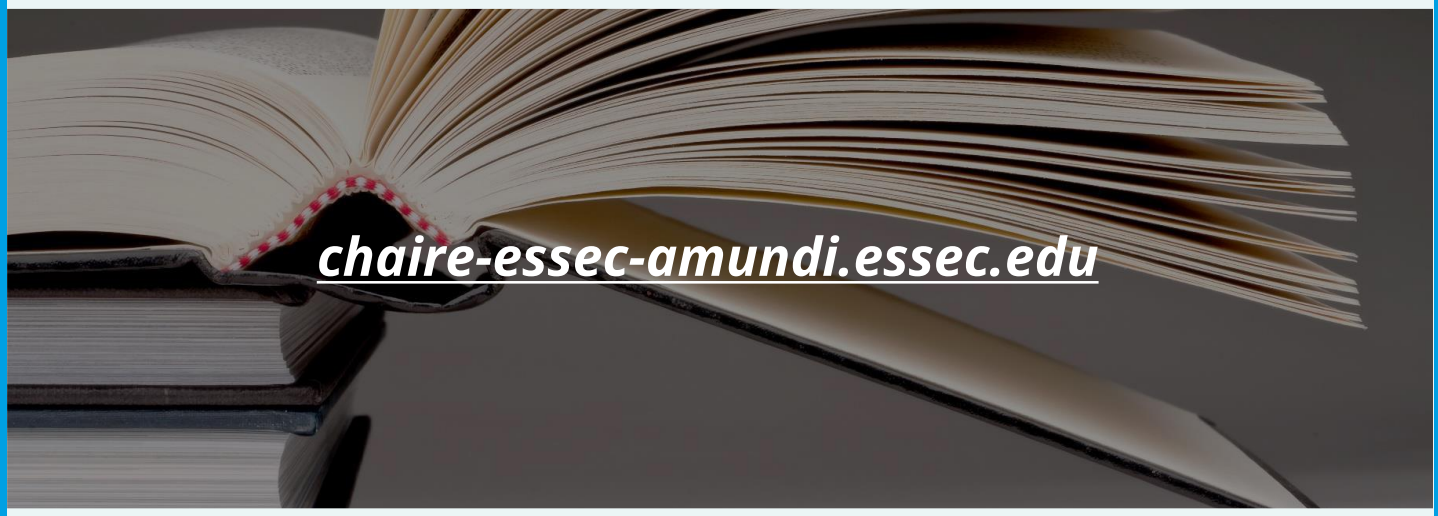
WORKING PAPER #189 | **MARCH 2026**

The Chair was created in 2016 with the objective to encourage and stimulate scientific cooperation between the research teams of ESSEC and Amundi on topics linked to asset and risk management with a special emphasis on responsible (ESG) investment. The Chair sponsors research projects, organizes an annual workshop and a series of webinars on asset management and ESG issues, prepares a series of research letters on ESG related issues for the Amundi research team and management and offers a series of research seminars for Amundi collaborators. Finally, it sponsors several scholarships for ESSEC Finance PhD students working in areas closely related to the Chair. As part of our collaboration with ESSEC, we are pleased to share this working paper authored by ESSEC researchers. This is an ad hoc release, outside our usual publications, reflecting the partnership agreement between our institutions. The papers cover key topics in machine learning and portfolio allocation.

Marie Brière
Head of Investor Intelligence & Academic Partnership (Amundi Investment Institute)
Co-head of the Chair

Elise Gourier
Associate Professor in Finance (ESSEC Business School)
Co-head of the Chair

Sofia Brito-Ramos
Professor of Finance (ESSEC Business School)
Co-head of the Chair



ESSEC BUSINESS SCHOOL

Amundi Investment Solutions

Trust must be earned

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

Arnaud Dufays, Kris Jacobs and Jeroen Rombouts

January 16, 2026

Abstract

Forecasting the option implied volatility (IV) surface is difficult with standard time series models because of its time-varying granularity. We propose a new two-step real-time sequential forecasting framework. The first step fits the daily surface and can accommodate any underlying specification for option prices or IVs, including dynamic option pricing models, nonparametric methods, and machine learning techniques. In the second step, we sequentially estimate a dynamic IV model using an updating rule. Our framework can accommodate large datasets and high data frequencies. An empirical application on S&P 500 IV surfaces shows that our approach significantly outperforms random walk forecasts.

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, arnaud.dufays@edhec.edu; Jacobs (Corresponding Author): University of Houston, kjacobs@bauer.uh.edu; Rombouts: ESSEC Business School, b00469813@essec.edu. We are very grateful to Hyung Joo Kim, Dong Hwan Oh, an anonymous referee, the editor, and seminar participants at the 2024 CFE-CMStatistics conference, the 2025 SoFiE meeting in Paris and the Federal Reserve Board for helpful comments, and 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 option prices 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 many 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, an alternative approach relies 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. Meaningful comparisons of these very different models are therefore a high priority, as well as frameworks that allow us to compare existing and new models.

This paper proposes a new framework that allows for these types of comparisons. We also contribute to the study of the *forecasting* performance of the various models and model classes. Forecasting implied volatility surfaces is critical for a variety of applications. First and foremost, market makers and risk managers require reliable forecasts to manage their exposure and optimize their hedging strategies. But forecasts are equally important for im-

[1] See FIA (2024) for more detailed information on trends in option trading volumes.

plementing trading strategies that exploit deviations between implied and realized volatility or that construct volatility arbitrage portfolios. Better out-of-sample predictions can also improve market efficiency and liquidity because market participants use the most recent surface to quote and price new illiquid contracts. Better forecasts therefore reduce pricing errors, thereby improving liquidity and reducing capital requirements. Our explicit focus on forecasting provides an interesting complement to in-sample comparisons, which by design favor more complex models. Our forecasting setup is novel and differs from existing approaches that rely on a random walk assumption (Almeida et al., 2023). We instead use the full history of the IV surfaces to construct our forecasts. We find that this results in 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; Chen et al., 2025) 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. It 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 been used by OptionMetrics (2022) and Ulrich et al. (2023). Note however that, in contrast to dynamic models with stochastic volatility and jumps, machine learning approaches are not arbitrage-free by construction.

To compare the forecasting performance of the various modeling approaches, one can fit

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

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 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. The dynamic surface model that we propose is general in terms of functional form, and we illustrate the approach by implementing the surface heterogenous autoregressive (SHAR) model inspired by the well-known 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 investigate the performance of dynamic multi-factor option pricing models, the 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

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

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 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-Robust. 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 report three main novel empirical findings. Our first conclusion is that the proposed SHAR and SHAR-Robust 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 the proposed 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 are outperformed by multi-factor stochastic volatility models. Second, SHAR forecasts can be improved by applying the approach to partitions of the option surface. Third, we also compare our approach to the method of Goncalves and Guidolin (2006) and find that our approach compares favourably.[4] Fourth, we demonstrate that our SHAR-Robust model implementation can lead to improved

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

forecasts. We find the highest occurrence of sequentially abnormal implied volatility surfaces occurs in 2018 and 2020. Fifth, 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 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 traditional random walk assumption to forecast the implied volatility surface, our proposed approach, illustrated by the SHAR and SHAR-Robust models, exploits 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 has not been sufficiently explored due to the models' nonlinearities and the computational burdens associated with recursive implementations.[5] It is especially difficult to evaluate the forecasting performance of models with jumps, and their relatively good performance using our approach is encouraging.

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

Finally, our paper also contributes 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 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. Section 7 compares our method with other existing approaches. Finally, Section 8 concludes and proposes directions for future research.

2 Data

We use daily observations of Black-Scholes implied volatilities from S&P 500 equity index (SPX) options traded on the Chicago Board Options Exchange (CBOE) for 2016-2021. We retain option contracts with implied volatilities between 0.05 and 1.5, and with volume and open interest of at least four contracts. Following standard practice, we also require that: the option price is positive; the offer price exceeds the bid price; the average of the bid and ask prices exceeds 50 cents; the moneyness is between 0.85 and 1.15; maturity is between 20 and 240 days.[6] Finally, we exclude options which violate put-call parity. After imposing these filters, the dataset contains 2,077,768 options. Table 1 presents descriptive statistics. Implied volatility has a long-term average of approximately 20% but it is highly-time-varying. The average number of option contracts per trading day increases from 790 in 2016 to 1674

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

in 2021. Due to the option filters, the average moneyness and maturity are relatively stable over the sample period.[7]

Table 1: IV Surfaces: Summary Statistics

Year	2016	2017	2018	2019	2020	2021
IV mean	0.16	0.12	0.16	0.16	0.26	0.19
IV standard deviation	0.05	0.05	0.06	0.05	0.11	0.06
IV min	0.05	0.05	0.05	0.05	0.07	0.05
IV max	0.41	0.39	0.57	0.36	0.98	0.49
Moneyness mean	0.98	0.97	0.98	0.98	0.99	0.98
Maturity mean (days)	55	53	55	57	59	61
# contracts/day	790	1105	1315	1329	1392	1674

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.

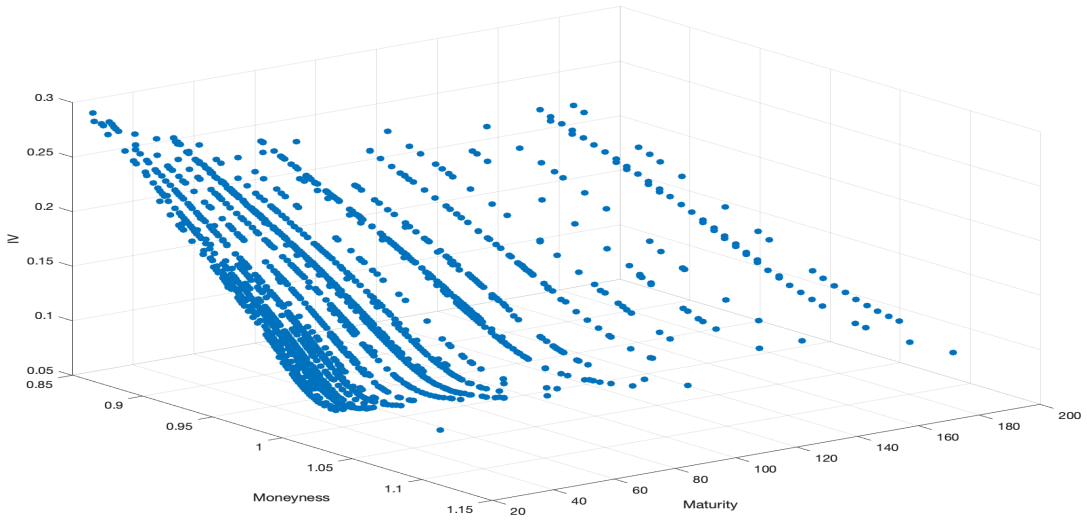
Figures 1 and 2 highlight some important stylized facts regarding the IV surfaces. The sample includes relatively more options with shorter maturities. Panel A of Figure 1 displays the implied volatility surface on a typical day, December 30, 2019. Implied volatilities display the characteristic asymmetric smile (smirk) at short maturities. For longer maturities, the smirk is less pronounced. For a given moneyness level, the term structure is relatively flat.

Panel B in Figure 1 displays the implied volatility surface on March 16, 2020. This is an atypical day at the start of the Covid period, when markets were in turmoil. The shape of the implied volatility surface inverses along the maturity axis, with extremely high implied volatilities for shorter maturities. This reflects elevated uncertainty about the near-term economic outlook. The smirk pattern is noticeably less pronounced. While the term structure is downward sloping, long maturity IVs are still substantially higher than their long-term averages. It stands to reason that in a standard sequential forecasting procedure, these atypical surfaces will impact both forecasting performance and sequential estimators for several days.

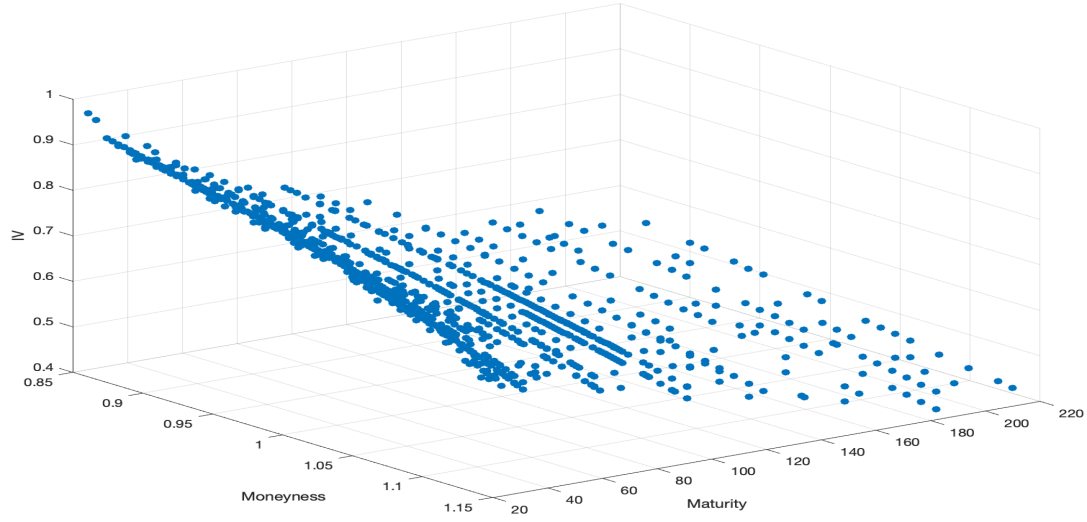
[7]While option volume has surged in recent years, this is mainly due to trading in ultrashort maturities, the so-called 0DTE options (Bandi et al., 2023; Almeida et al., 2024; Dim et al., 2024). We keep the application of our forecasting setup to these options with very short maturities for future research.

Figure 1: Implied Volatility Surfaces

(a) Panel A: December 30, 2019



(b) Panel B: March 16, 2020



Notes: We plot the option surface for two days in the sample. Panel A plots the surface for December 30, 2019, Panel B for March 16, 2020. Maturity is measured in days, moneyness is defined as the strike divided by the index level. IV is the Black-Scholes implied volatility.

Panel A of Figure 2 plots average daily implied volatility between 2016 and 2021. We report separately on options with short maturities (less than 60 days) and long maturities (more than 150 days). While both time series are highly persistent, we also observe several sharp spikes when implied volatility jumps up, see for example the onset of COVID-19 crisis. The long-maturity implied volatilities are clearly more stable than their short-maturity counterparts. While they are also higher on approximately 70 percent of the days in our sample, the short-maturity implied volatilities increase substantially in crisis periods. These stylized facts are illustrated by the volatility term structure plot in Panel B of Figure 2, where the term structure is computed as the difference between the short and long maturity implied volatility time series.

3 The Dynamic Surface Model

We start by introducing the notation and setting in Section 3.1. Section 3.2 outlines the random walk assumption, which is most often used when forecasting IV surfaces. Section 3.3 lays out the general framework for using a time series model without a fixed-grid panel and defines the dynamic surface model. Section 3.4 highlights the model used in the application. Next, Section 3.5 highlights how to sequentially estimate the model, with special attention given to efficient implementation. Section 3.6 robustifies the estimator with the help of an automatic abnormal surface detection algorithm.

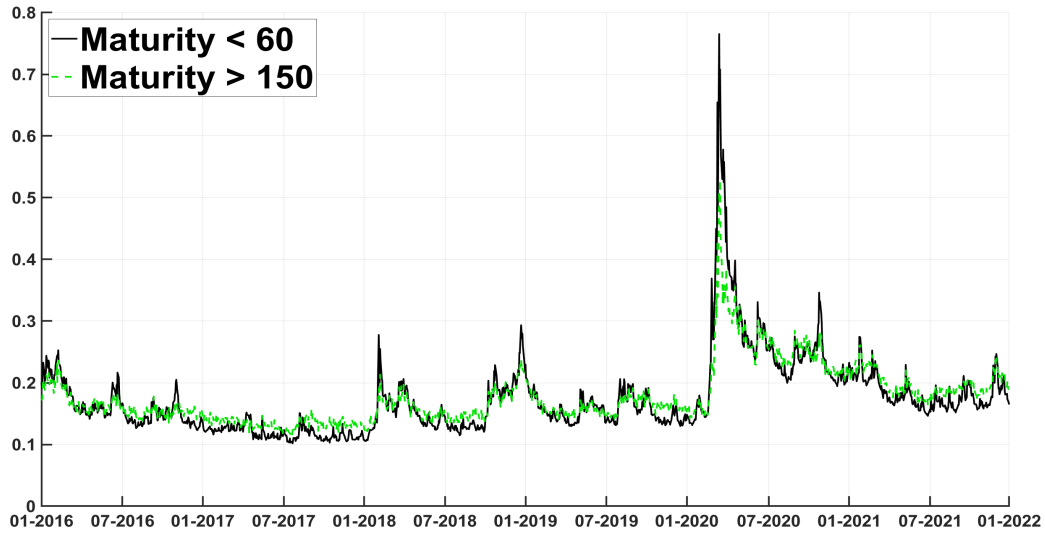
3.1 Preliminaries

Before introducing the IV surface model and the forecasting setup, we fix notation and the general model setting. From an initial day onwards, we observe new IV surfaces at the daily frequency.[8] This yields on every day t a stream of IV surfaces $IV(O_{i,1}), IV(O_{i,2}), \ldots, IV(O_{i,t})$, where each surface consists of N_l, $l = 1, \ldots, t$, option-implied volatilities with characteristics

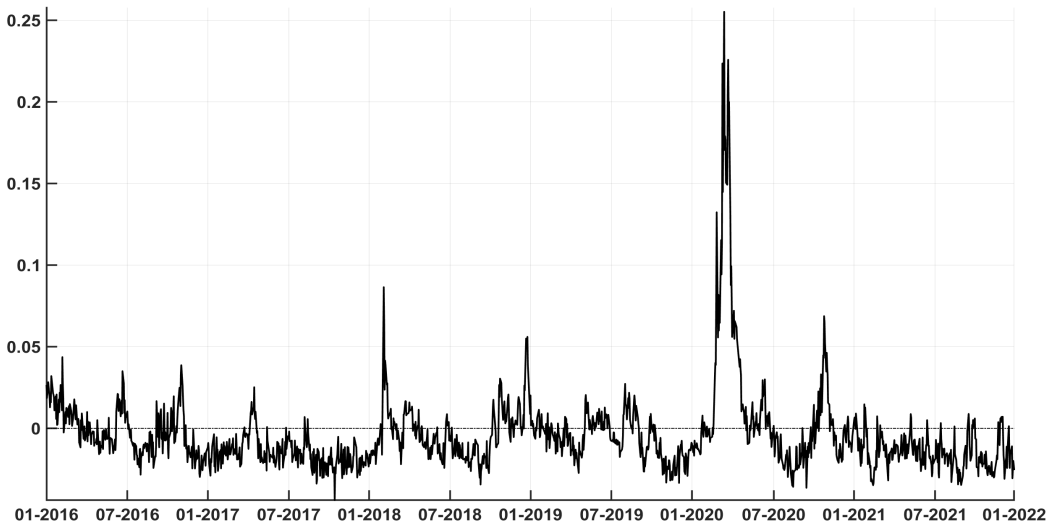
[8] Our setup is applicable to any frequency, but our implementation uses daily IV surfaces.

Figure 2: Time Series of Daily Implied Volatility

(a) Panel A: Daily Average Implied Volatility



(b) Panel B: Daily Term Structure



Notes: Panel A displays daily average implied volatilities for options with maturity below 60 days (solid black) and above 150 days (dashed green). Panel B displays the daily term structure, measured as the difference between the two time series in Panel A (short maturity minus long maturity).

$O_{i,l}$. For instance, on day l we observe the surface $IV(O_{i,l})$ with options $i = 1, \ldots, N_l$ and option characteristics $O_{i,l} = \{m_{i,l}, \tau_{i,l}\}$, where $m_{i,l} \equiv \frac{K_{i,l}}{S_l}$ measures the moneyness with $K_{i,l}$ the strike price, S_l the index level, and $\tau_{i,l}$ is the maturity of the option measured in calendar days. Our main research question is how to generate a day-t forecast of the IV surface h days ahead, i.e. at day $t + h$. Note that this is not a trivial question, mainly because the option characteristics are time-varying.

A standard approach starts by fitting the implied volatility surface, as follows:

$$IV(O_{i,l}) \quad = \quad IV^M(O_{i,l}, \mathbf{\Theta}_l) + \varepsilon_{i,l}, \tag{1}$$

where the notation $IV^M()$ emphasizes that model M generates IV surface given the set of model parameters in the vector $\mathbf{\Theta}_l$. For instance, the model parameters $\mathbf{\Theta}_l$ are comprised of the spot variance for the BS model and of the parameter vector $\mathbf{\Theta}_l = \{V_l, \kappa, \theta, \sigma, \rho\}$ for the Heston stochastic volatility model, which is defined in Section 4.4. Note that $\mathbf{\Theta}_l$ can be very high-dimensional, as in a neural network for example, and is not even necessarily a finite dimensional parameter. Specifically, it can also represent a nonparametric function, as in the case of a random forest composed of decision trees.

3.2 The Random Walk Surface Forecast

Our approach is very general because it combines any model to fit equation (1) with the researcher's favorite autoregressive volatility model. To highlight how this differs from the existing literature, note that available procedures for forecasting the surface for day $t + h$ on day t consists of two steps. First, fit equation (1) using least squares, that is, estimate the parameter vector $\mathbf{\Theta}_t$ given the N_t options observed at day t. This yields $\hat{\mathbf{\Theta}}_t$. Second, forecast the IV surface at day $t + h$ using $\widehat{IV}(O_{i,t+h}) = IV^M(O_{i,t+h}, \hat{\mathbf{\Theta}}_t)$. This second step is often referred to as the random walk (RW) forecasting step, because it assumes that the

fitted surface on day t is the best forecast of any future surface. Note that this procedure is computationally convenient and fast, because it suffices to estimate any given model on a single daily surface. This procedure is used in the academic literature (Almeida et al., 2023) as well as in practitioner approaches, see for example the platform documentation of SpiderRock (2024), which states "*Most of the time, but not always, the best view of the shape of market volatility in the future is what it is like right now*".

3.3 Dynamic Surface Modeling

We now propose our generalization of this forecasting approach, which replaces the random walk forecast with one's favorite autoregressive volatility model. This is a conceptually simple idea, but the implementation is not straightforward. The procedure explained in Sections 3.1 and 3.2 relies on one day of option data only, and ignores the information observed prior to that day. In general, on a given day t the available information consists of a sequence of IV surfaces summarized by the sequence of parameter vector estimates $\hat{\boldsymbol{\Theta}}_{1:t} = \{\hat{\boldsymbol{\Theta}}_1, \ldots, \hat{\boldsymbol{\Theta}}_t\}$. The RW forecast of the IV surface at day $t+h$ consists of the day-t fitted surface generated by the model parameter estimate $\hat{\boldsymbol{\Theta}}_t$. This RW assumption is unsatisfactory given the findings of the financial econometrics literature, which suggests that autoregressive processes, including fractionally integrated moving averages and heterogeneous autoregressive models (see, Baillie et al., 1996; Corsi, 2009), reliably outperform the random walk assumption for the purpose of modeling and forecasting realized variances and the VIX (or VIX^2).

In general, directly modeling the dynamics of the IV surfaces is computationally very expensive and updating is difficult. We therefore instead use the sequence of parameter estimates $\hat{\boldsymbol{\Theta}}_{1:t}$ to build a dynamic IV surface model that can be estimated in real time. Computational speed is a critical concern, because the daily surface contains a large number of option contracts. The key idea underlying our approach is that, for any moneyness-

maturity combination on the surface we are attempting to forecast, we exploit the entire parameter sequence to generate a time series of length t composed of past model-implied volatilities. Subsequently we apply a dynamic time series model for forecasting the implied volatility surface at time $t+h$.

More precisely, given a forecast horizon h, we model the IV surface on days $l = 1, \ldots, t$ as follows:

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

where f_k is an arbitrary (but known) function of past model-implied IV surfaces.[9] This general specification allows for polynomial-type structures and interaction effects. An important case is a linear function f_k, which includes autoregressive processes with K lags: $\beta_h^{(0)} + \sum_{k=1}^{K} \beta_h^{(k)} IV^M(O_{i,l}, \hat{\mathbf{\Theta}}_{l-h-k+1})$. Note however that this latter linear specification differs from a standard autoregressive specification that uses linear combinations of previous implied volatility values to predict the current implied volatilities. Specifically, our specification predicts the IV surface using past IV surfaces *generated by the chosen model*. This is possible in this framework because we use the implied volatility values observed at options characteristics $O_{i,l}$ from model surfaces given by $IV^M(O_{i,l}, \hat{\mathbf{\Theta}}_{l-h-k+1})$ with $k = 1, \ldots, K$. The random walk forecasting approach is of course a special case of equation (2), where $K = 1$, $f_1(O_{i,l}, \hat{\mathbf{\Theta}}_{1:l-h}) = IV^M(O_{i,l}, \hat{\mathbf{\Theta}}_{l-h})$, $\beta_h^{(0)}$ is equal to zero and $\beta_h^{(1)}$ is equal to one.

The model in equation (2) can then be written in matrix notation by stacking all the implied volatilities observed at day l into the vector $\boldsymbol{y}_l = \big(IV(O_{1,l}), \ldots, IV(O_{N_l,l})\big)'$. By using the $l-h$ model-implied volatility parameter estimates $\hat{\mathbf{\Theta}}_{1:l-h}$ into the following vector

[9] Equation (2) can be further generalized by allowing the model parameters to depend on option characteristics, i.e., $\beta_{h,i}^{(k)}$ for $k = 0, ..., K$. This is equivalent to considering (2) only for subsets of the IV surface. While this extension is more computationally expensive and our objective is to develop a real-time method, it is explored in Section 6.3.

$\boldsymbol{x}_{l|1:l-h}^{(k)} = \left(f_k(O_{1,l}, \hat{\boldsymbol{\Theta}}_{1:l-h}), \ldots, f_k(O_{N_l,l}, \hat{\boldsymbol{\Theta}}_{1:l-h})\right)'$, we get:

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

with parameter vector $\boldsymbol{\beta}_h = (\beta_h^{(0)}, \beta_h^{(1)}, \ldots, \beta_h^{(K)})'$, residual vector $\boldsymbol{\varepsilon}_l = \left(\varepsilon_{1,l}, \ldots, \varepsilon_{N_l,l}\right)'$ 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 $(N_l \times K)$.

Next, to use all available information at day t to forecast the IV surface at $t+h$, we pool equation (3) for days $l = h+K, \ldots, t$ and denote the vector $\boldsymbol{Y}_t = (\boldsymbol{y}_{h+K}, \ldots, \boldsymbol{y}_t)'$ which contains the available stacked IV surfaces. The corresponding matrix $\boldsymbol{X}_{h|t}$ stacks the matrices $\boldsymbol{X}_{h+K|K} \ldots \boldsymbol{X}_{t|t-h}$, which are the model-predicted implied volatilities. Using a squared Euclidean loss function, the model parameters can then be estimated using the ordinary least squares (OLS) formula, $\hat{\boldsymbol{\beta}}_{h|t} = (\boldsymbol{X}_{h|t}'\boldsymbol{X}_{h|t})^{-1}\boldsymbol{X}_{h|t}'\boldsymbol{Y}_t$, where $\hat{\boldsymbol{\beta}}_{h|t}$ is indexed by t to highlight that the estimation is 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 on day t is then given by:

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

This completely characterizes our estimation and forecasting setup. It is relatively straightforward to extend this setup, for instance according to Bollerslev et al. (2016) or Cipollini et al. (2021), who propose novel methods to account for and incorporate measurement error in realized volatility models. Another potential extension is to guarantee absence of arbitrage in the IV surfaces, see for instance Fengler (2009). This is straightforward for some of the option pricing models we consider, but not for machine learning approaches.[10]

[10] Zhang et al. (2023) include no arbitrage constraints while training a neural network on the implied volatility surface.

Moreover, forecasting IV surfaces that are ex-ante arbitrage free is not straightforward and requires constraints on the parameter estimates $\hat{\boldsymbol{\beta}}$, functional forms f_k, and the underlying surface model characterized by the vector $\hat{\boldsymbol{\Theta}}$. Because our main focus is on demonstrating that our setup can be used to study a variety of surface and forecasting models, we keep this extension for future work.

3.4 The Surface HAR Specification

The specification of the dynamic surface model in equation (2) depends on the number of regressors (K) and the regressor functionals ($f_1, \ldots, f_K$). We explored several values of K, functions, and parameter restrictions. In our empirical implementation, we found that the heterogenous autoregressive model of Corsi (2009) is particularly successful. This specification amounts to specifying all linear functions, $K = 22$, and it restricts the parameters such that the IV surface depends on the fitted surface from the previous day as well as fitted surfaces aggregated over the previous week and month respectively. This results in a parsimonious model with only four parameters.

We refer to the resulting forecasting approach as the Surface HAR approach, or SHAR for short. It models the IV surface on day $l = 1, \ldots, t$ for horizon h as follows:

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

Note that in our daily updating context, unlike models with fixed regressors, any SHAR approach requires a significant amount of calculations for re-evaluating the lagged surfaces at the current option's IV characteristics using each of the past 22 parameter vectors $\hat{\boldsymbol{\Theta}}_{l-h-j}$.

3.5 Real-Time Estimation

In a daily updating setup, the matrix $\boldsymbol{X}_{h|t}$ and vector $\boldsymbol{y}_t$ in the OLS formula increase rapidly in size since the row dimension increases by N_t observations on every day t. To save computational resources, we therefore 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 $\boldsymbol{\Omega}_{h|t}^{-1} = (\boldsymbol{X}_{h|t}'\boldsymbol{X}_{h|t})^{-1}$, we can update the parameter estimates at time $t+1$ according to:

$$\begin{aligned} \boldsymbol{\Omega}_{h|t+1}^{-1} &= \boldsymbol{\Omega}_{h|t}^{-1} + (\boldsymbol{\Omega}_{h|t}^{-1}\boldsymbol{X}_{h|t+1}')(\boldsymbol{I}_{N_{t+1}} + \boldsymbol{X}_{h|t+1}\boldsymbol{\Omega}_{h|t}^{-1}\boldsymbol{X}_{h|t+1}')^{-1}(\boldsymbol{X}_{h|t+1}\boldsymbol{\Omega}_{h|t}^{-1}), \\ \hat{\boldsymbol{\beta}}_{h|t+1} &= \boldsymbol{\Omega}_{h|t+1}^{-1}\big(\boldsymbol{\Omega}_{h|t}\hat{\boldsymbol{\beta}}_{h|t} + \boldsymbol{X}_{h|t+1}'\boldsymbol{y}_{t+1}\big), \end{aligned} \tag{6}$$

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

3.6 Accounting for Abnormal Volatility Surfaces

The model in equation (2) forecasts the IV surface for day $t+h$ using a linear combination of past model-implied IV surfaces. Due to the high persistence of IV surfaces, the predicted IV surface at time t generally carries the most weight in this combination, evidenced by $\hat{\beta}_h^{(1)}$ values close to one. This prediction therefore in turn heavily relies on the parameter estimate $\hat{\boldsymbol{\Theta}}_t$, which is derived from options observed at time t. However, if the IV surface at time t is disrupted by an exogenous shock, such as the Covid pandemic or the flash crash, the resulting predicted IV surface can be distorted. To identify such abnormal surfaces and further improve forecasts, we apply the outlier detection method proposed by Rousseeuw and Croux (1993). Denote the average squared residuals from day 1 to t by $\text{ASR}_{1:t}$, where $\text{ASR}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}\left(IV(O_{i,t}) - IV^M(O_{i,t}, \hat{\boldsymbol{\Theta}}_t)\right)^2$. A surface is determined to be 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. Second, the method is applicable to any IV model. Consequently, 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 robustify the OLS parameter estimates to abnormal surfaces. When an outlier is detected at time t, we do not update the matrix inverse $\mathbf{\Omega}_{h|t+1}^{-1}$ and parameter estimates $\hat{\boldsymbol{\beta}}_{h|t+1}$ according to equation (6). In addition, if an outlier is detected and the current surface is determined to be abnormal, it does not seem optimal to use it as the most important predictor in the forecast equation (4). We therefore replace $IV^M(O_{i,t+h}, \hat{\mathbf{\Theta}}_t)$ by the average of the last two surfaces, i.e. $\frac{1}{2}(IV^M(O_{i,t+h}, \hat{\mathbf{\Theta}}_t) + IV^M(O_{i,t+h}, \hat{\mathbf{\Theta}}_{t-1}))$. Henceforth we refer to this forecasting implementation as the robust version.

4 Modeling 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 models. Section 4.3 discusses two popular machine learning techniques. 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 model that is easy to implement is the ad-hoc Black-Scholes (AHBS) specification, used by Dumas et al. (1998) and Seo and Wachter (2019) among others. This approach fits the IV surface using a second degree polynomial in moneyness and maturity:[11]

$$IV(O_{i,t}) = \beta_0^{\text{LR}} + \beta_1^{\text{LR}} m_{i,t} + \beta_2^{\text{LR}} m_{i,t}^2 + \beta_3^{\text{LR}} \tau_{i,t} + \beta_4^{\text{LR}} \tau_{i,t}^2 + \beta_5^{\text{LR}} m_{i,t} \tau_{i,t} + \varepsilon_{i,t}. \tag{7}$$

The AHBS model provides a smooth approximation to the IV surface. As a simple benchmark, we also report the Black-Scholes (BS) model that consists of Equation (7) without explanatory variables (i.e., only the constant β_0^{LR}).

4.2 Nonparametric Models

We consider a standard kernel density smoother, as implemented in OptionMetrics (2022) for example. It computes the IV surface with a separate kernel smoother for calls ($I_{i,t} = 1$) and puts ($I_{i,t} = 0$), using the option's delta $\Delta_{i,t}$ rather than moneyness $m_{i,t}$, therefore $O_{i,t} = \{\Delta_{i,t}, \tau_{i,t}, I_{i,t}\}$. It takes the form:

$$IV^M(O_{i,t}, b) = \sum_{j=1}^{N_t} \frac{\text{Vega}_{j,t}\phi(\Delta_{j,t} - \Delta_{i,t}, \ln(\tau_{j,t}) - \ln(\tau_{i,t}), I_{j,t} - I_{i,t})}{\sum_{k=1}^{N_t} \text{Vega}_{k,t}\phi(\Delta_{k,t} - \Delta_{i,t}, \ln(\tau_{k,t}) - \ln(\tau_{i,t}), I_{k,t} - I_{i,t})} IV(O_{j,t}), \tag{8}$$

where Vega stands for the Black-Scholes sensitivity of the option, $\phi(x, y, z) = \exp\left(-\frac{1}{2}(x^2/b_1 + y^2/b_2 + z^2/b_3)\right)$, with $\boldsymbol{b} = (b_1, b_2, b_3)$ the bandwidth vector that determines the level of smoothing. OptionMetrics (2022) uses $\boldsymbol{b} = (0.05, 0.005, 0.001)$. For consistency with the other models used in this study, we have modified the kernel smoother in equation (8) by incorporating the moneyness dimension instead of the option's delta. Additionally, as we transform each put option into a call option using put-call parity, we use a two-

[11] The Black-Scholes model assumes a constant spot variance. It is equivalent to a linear regression with only a constant.

dimensional Gaussian kernel. The adapted OptionMetrics kernel smoother is thus computed as follows:

$$IV^M(O_{i,t}, b) = \sum_{j=1}^{N_t} \frac{\text{Vega}_{j,t}\phi(m_{j,t} - m_{i,t}, \ln(\tau_{j,t}) - \ln(\tau_{i,t}))}{\sum_{k=1}^{N_t} \text{Vega}_{k,t}\phi(m_{k,t} - m_{i,t}, \ln(\tau_{k,t}) - \ln(\tau_{i,t}))} IV(O_{j,t}), \tag{9}$$

where $\phi(x, y) = \exp\left(-\frac{1}{2}(x^2/b_1 + y^2/b_2)\right)$. The bandwidth parameters b_1 and b_2 are estimated on a daily basis using cross-validation, where the test set includes 30% of the daily options data.

4.3 Machine Learning Methods

To allow for additional non-linearities, we also consider machine learning methods. Similar to the nonparametric models, these are not arbitrage-free by construction. 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 three hidden layers based on ReLu activation functions with 32, 16 and 8 neurons, respectively:

$$\begin{aligned}
\boldsymbol{z}^{(1)} &= \text{RL}(W_1^{(1)} m_{i,t} + W_2^{(1)} \tau_{i,t} + \boldsymbol{b}^{(1)}), \text{ with } \boldsymbol{z}^{(1)} \in \Re^{32\times 1} \\
\boldsymbol{z}^{(2)} &= \text{RL}(W^{(2)} \boldsymbol{z}^{(1)} + \boldsymbol{b}^{(2)}) \in \Re^{16\times 1}, \text{ with } \boldsymbol{z}^{(2)} \in \Re^{16\times 1} \\
\boldsymbol{z}^{(3)} &= \text{RL}(W^{(3)} \boldsymbol{z}^{(2)} + \boldsymbol{b}^{(3)}), \text{ with } \boldsymbol{z}^{(3)} \in \Re^{8\times 1} \\
IV(O_{i,t}) &= W^{(\text{IV})} \boldsymbol{z}^{(3)} + \boldsymbol{b}^{(\text{IV})} + \varepsilon_{it},
\end{aligned}$$

where $\text{RL}(\cdot)$ denotes the ReLu activation function. This configuration implies 769 parameters to estimate for each daily surface.

Our second machine learning application fits the option surface with random forests. Random forests were 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 with latent variables, specified under the risk-neutral measure as

$$\frac{dS_t}{S_t} = (r_t - \delta_t - \lambda\bar{\mu}_s)dt + \sum_{i=1}^{N} \sqrt{V_{i,t}} dZ_{i,t} + (e^{J_t^s} - 1)dN_t, \tag{10}$$

$$dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \sigma_i \sqrt{V_{i,t}} dW_{i,t}, \tag{11}$$

$$\text{Corr}(dW_{i,t}, dZ_{i,t}) = \rho_i dt. \tag{12}$$

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. $dZ_{i,t}$ and $dW_{i,t}$ are Brownian motions with $corr(dZ_{i,t}, dZ_{j,t}) = 0$ and $corr(dW_{i,t}, dW_{j,t}) = 0$, $i \neq j$. 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 $J_t^s \sim p_l \eta_l \exp(\eta_l y) 1_{y \leq 0} + (1 - p_l)\eta_r \exp(-\eta_r y) 1_{y>0}$. In such a case, the compensation amounts to $\bar{\mu}_s = p_l \frac{\eta_l}{1+\eta_l} + (1 - p_l)\frac{\eta_r}{\eta_r - 1} - 1$. The jump param-

eters, denoted by the vector $\mathbf{\Lambda}_J$, are given by $\mathbf{\Lambda}_J = (\lambda, \mu_s, \sigma_s)$ for the Normal specification and by $\mathbf{\Lambda}_J = (\lambda, p_l, \eta_l, \eta_r)$ for the DE alternative.

Denoting $\boldsymbol{V}_t = (V_{1,t}, \ldots, V_{N,t})$, $\mathbf{\Theta} = (\kappa_1, \theta_1, \sigma_1, \rho_1, \ldots, \kappa_N, \theta_N, \sigma_N, \rho_N, \Lambda_J)$, the model price of a European call option $C^M(\tau_{i,t}, K_{i,t}, \boldsymbol{V}_t, \mathbf{\Theta})$ with maturity $\tau_{i,t}$ and strike price $K_{i,t}$ is given by:

$$C^M(\tau_{i,t}, K_{i,t}, \boldsymbol{V}_t, \mathbf{\Theta}) = e^{-r_t \tau_{i,t}} E[\max(S_{t+\tau} - K_{i,t}, 0)]. \tag{13}$$

The model yields a closed-form expression for the conditional characteristic function of the log index level, which makes the integral given in (13) numerically tractable. To optimize the parameters $\mathbf{\Theta}$ and the spot variances $\boldsymbol{V}_t$, we minimize the following loss function:

$$\hat{\boldsymbol{V}}_t, \hat{\mathbf{\Theta}}_t = \text{argmin}_{V_t, \Theta_t} \sum_{i=1}^{N_t} \left(\frac{C(\tau_{i,t}, K_{i,t}) - C^M(\tau_{i,t}, K_{i,t}, \boldsymbol{V}_t, \mathbf{\Theta}_t)}{\text{Vega}_{i,t}} \right)^2, \tag{14}$$

where $C(\tau_{i,t}, K_{i,t})$ is the quoted price of the contract with maturity $\tau_{i,t}$ and strike $K_{i,t}$ on day t and $\text{Vega}_{i,t}$ stands for the Black-Scholes sensitivity of the option computed using the implied volatility from the market price of the option $C(\tau_{i,t}, K_{i,t})$. The loss function given in (14) 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).

Within the general specification given by (10)-(12), we consider the following option pricing models:

1. The Heston model, which we refer to as SV(1). When the model contains one volatility factor (i.e. $N = 1$) and no jump in return, this amounts to the Heston (1993) model. The process has five unknown parameters that need to be estimated, the spot variance V_t and $\mathbf{\Theta}_{\text{SV(1)}} = (\kappa_1, \theta_1, \sigma_1, \rho_1)$.

2. A stochastic volatility model with two factors, 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 $\boldsymbol{V}_t = (V_{1,t}, V_{2,t})$ and 8 parameters, $\boldsymbol{\Theta}_{\text{SV(2)}} = (\kappa_1, \theta_1, \sigma_1, \rho_1, \kappa_2, \theta_2, \sigma_2, \rho_2)$.

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

4. A stochastic volatility model with a jump in returns based on a Normal distribution, denoted by SVJR. This is the model proposed by Bates (2000) and corresponds to $N = 1$ in our model specification. The number of parameters to be estimated on each day amounts to 8: V_t and $\boldsymbol{\Theta}_{\text{SVJR}} = (\kappa_1, \theta_1, \sigma_1, \rho_1, \lambda, \mu_s, \sigma_s^2)$.

5. A stochastic volatility model with double exponential jumps in returns, corresponding to $N = 1$ in our model specification, referred to as SVDE. In this case there are 9 parameters to be estimated, given by V_t and $\boldsymbol{\Theta}_{\text{SVDE}} = (\kappa_1, \theta_1, \sigma_1, \rho_1, \lambda, \eta_l, \eta_r, p_l)$. This model is used, among others, by Bates (2012), Andersen et al. (2015a), Andersen et al. (2015c), Andersen et al. (2020) and Dufays et al. (2024).

5 Empirical Results: Daily In-sample Fit

Our approach consists of daily fitting and dynamic modeling of the IV surface. This is done sequentially and we check for abnormal surfaces when new surfaces become available at the daily frequency. In this section, we report the results from fitting the sequence of IV surfaces. Section 5.1 discusses the daily surface fit and the detection of abnormal IV surfaces. Section 5.2 summarizes the real-time estimates of the dynamic surface model. Section 5.3 takes a closer look at the estimates of the dynamic option pricing models.

5.1 Daily Surface Fit

Table 2 reports the in-sample performance of the various surface fit models for each year from 2016 to 2021, assessed using Implied Volatility Root Mean Squared Errors (IVRMSEs) expressed in percentages. The nonparametric and machine learning models achieve very good fits, with IVRMSEs as low as 0.25 percent for the random forest in 2021. This is expected, as these methods are entirely data-driven and thus subject to the risk of overfitting. In contrast, the dynamic option pricing models, characterised by relatively few parameters, exhibit IVRMSEs approximately twice as high for all years in the sample. Interestingly, the SV(3) model slightly outperforms the SVJR and SVDE jump models in terms of IVRMSEs in the pre-Covid period. Overall, the largest errors occur in 2020, the year of the Covid outbreak.

Table 2: In-Sample Surface Fit for Various Models.

Model	2016	2017	2018	2019	2020	2021
BS	4.67	4.42	4.67	4.40	5.89	5.18
AHBS	0.78	0.88	1.03	0.82	1.30	0.95
ANN(3)	0.54	0.49	0.66	0.47	1.01	0.45
OptionMetrics	0.37	0.34	0.52	0.33	0.88	0.31
Random Forest	**0.31**	**0.28**	**0.43**	**0.28**	**0.71**	**0.25**
SV(1)	0.64	0.62	0.82	0.54	1.23	0.54
SV(2)	0.53	0.52	0.73	0.46	1.16	0.49
SV(3)	0.52	0.50	0.72	0.45	1.14	0.48
SVJR	0.60	0.56	0.78	0.51	1.14	0.50
SVDE	0.58	0.54	0.75	0.49	1.09	0.47

We report the in-sample implied volatility root mean squared errors for each year.

Figure 3 presents the time series of the daily IVRMSEs for the fitted OptionMetrics and SV(1) models in Panels (a) and (b), respectively. The OptionMetrics model consistently exhibits smaller and more stable errors than the SV(1) model throughout the entire period. For the SV(1) model, the IVRMSEs are frequently below one percent before Covid but remain consistently above one percent after the pandemic. The vertical bars in both panels

indicate the days when the IV surfaces are identified as abnormal based on the procedure outlined in Section 3.6. There is a clear relation between poor model fit and the detected abnormal surfaces. Fitting these abnormal surfaces is difficult and depends on the model. Finally, note that the spikes in implied volatility in Figure 2 and IVRMSE in Figure 3 are highly correlated.

Table 3 provides more details on the number of detected abnormal IV surfaces for the various models. We find high numbers of abnormal surfaces in 2020 and 2018. The highest number of abnormal surfaces is obtained in 2020, because of the start of the COVID-19 crisis in March of that year. The SV(1) model results in 36 abnormal surfaces in 2020, whereas the more flexible SVDE model results in 33 abnormal surfaces. In 2018, the emerging market turmoil and Federal Reserve interest rate hikes impacted option markets, which also resulted in multiple abnormal IV surfaces.

Table 3: Number of Abnormal Surfaces for Various Models.

Model	2016	2017	2018	2019	2020	2021
BS	2	0	5	0	64	4
AHBS	4	0	6	1	24	0
ANN(3)	4	2	6	1	29	2
OptionMetrics	1	0	5	0	19	0
Random Forest	1	0	5	0	15	0
SV(1)	4	3	17	2	36	1
SV(2)	4	4	16	1	39	2
SV(3)	4	2	14	1	37	0
SVJR	4	3	17	2	34	2
SVDE	2	3	19	1	33	1

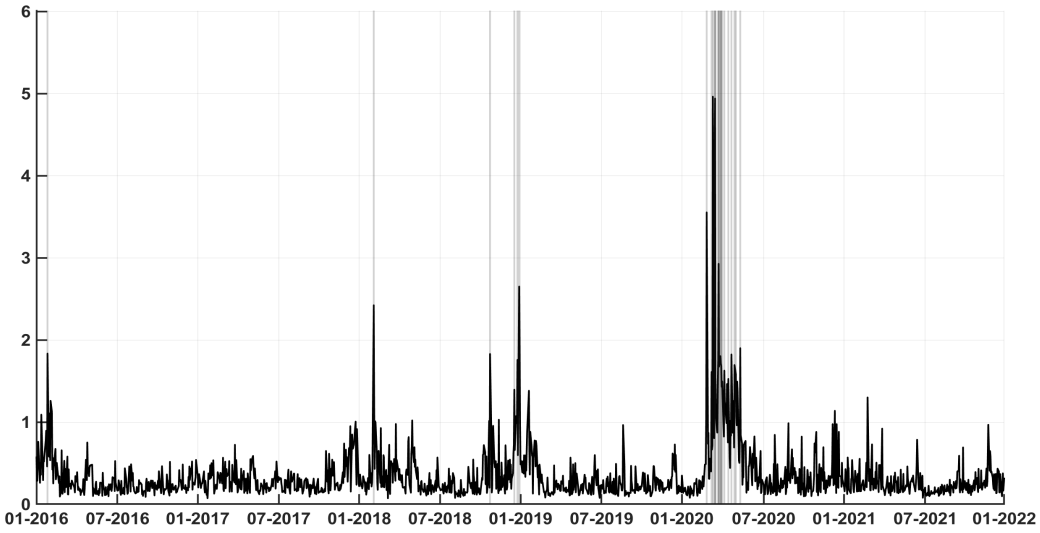
We report the number of in-sample mean squared errors that are larger than the outlier threshold for each year in the sample (see Section 3.6). Data from 2015 is used as initial sample.

5.2 Real-Time Dynamic Model Estimates

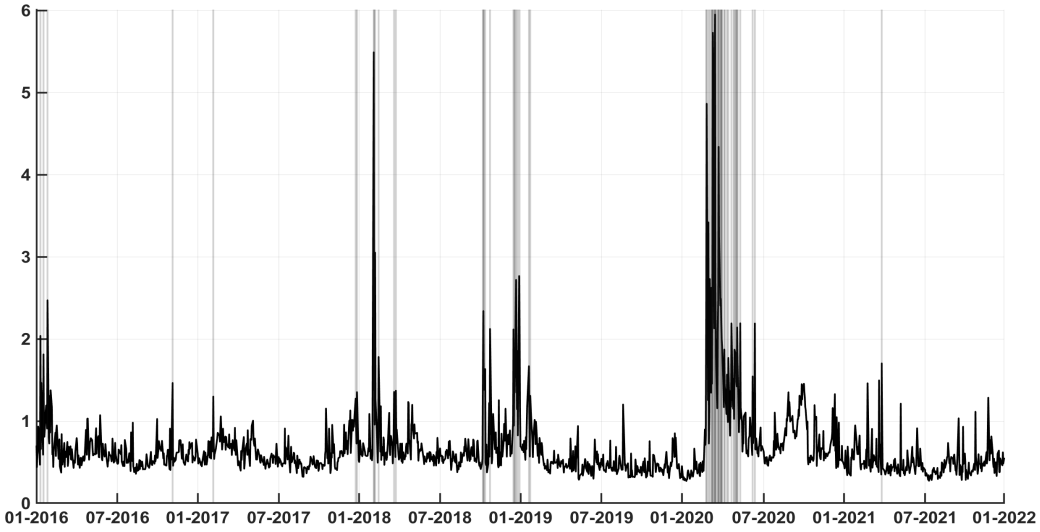
We next provide more insights into the dynamic properties of the surface used for forecasting. The SHAR type surface model has four parameters that depend on the forecast

Figure 3: Daily IVRMSE for the SV(1) Model.

(a) *OptionMetrics*



(b) *SV(1)*



Notes: Daily IVRMSEs for day l are computed 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 indicated by vertical gray bars.

horizon. While the forecasting exercise is implemented daily from 2016 until the end of 2021, we use data from 2015 to guarantee the stability of the dynamic model. That is, the sequential estimator (6) uses the 2015 data to ensure that the 2016 forecasts are not impacted by initial conditions.

Table 4 reports in the top panel the estimates of these parameters for the OptionMetrics model on the last day of 2021, i.e. based on the full sample period. The one- and five-day ahead surface is highly related to the most recent available surface, with $\hat{\beta}^{(1)}_{h|t}$ estimates of approximately 0.8. The one-day ahead surface is impacted by the average surface of the most recent week ($\hat{\beta}^{(2)}_{1|t} = 0.18$), but not by the average surface of the most recent month ($\hat{\beta}^{(3)}_{1|t} = 0$). However, for the twenty-day ahead horizon, this average surface has a significant impact ($\hat{\beta}^{(3)}_{20|t} = 0.35$), while $\hat{\beta}^{(1)}_{20|t} = 0.69$ remains high and $\hat{\beta}^{(2)}_{20|t}$ is negative. The bottom panel in Table 4 reports estimates for the SV(1) model and are similar to the OptionMetrics model. This holds true for the other surface fit models that we omit here to save space.

Table 4: SHAR Parameter Estimates (OptionMetrics and SV(1) Model)

Horizon (h)	$\hat{\beta}^{(0)}_{h\|t}$	$\hat{\beta}^{(1)}_{h\|t}$	$\hat{\beta}^{(2)}_{h\|t}$	$\hat{\beta}^{(3)}_{h\|t}$
		Model - OptionMetrics		
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
		Model - SV(1)		
1	0.00	0.78	0.22	0.00
5	0.01	0.84	0.06	0.07
20	0.03	0.67	-0.15	0.35

We report parameter estimates for the SHAR forecasting approach, using the OptionMetrics and SV(1) models, based on the full sample period.

Figure 4 shows the sequentially fitted parameters of the SHAR (solid black) and SHAR-Robust (dashed green) models related to the most recent available surface $\hat{\beta}^{(1)}_{h|t}$, again based on the OptionMetrics model. Panel A reports on the one-day horizon and Panel B on the

one-month horizon. For the SHAR model in Panel A, the parameter estimate $\hat{\beta}^{(1)}_{1|t}$ is just below one in early 2016, and slightly declines until early 2018, when 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 remains there until the end of 2021. The SHAR-Robust parameter estimates are close to the SHAR parameter estimates until the COVID-19 crisis; subsequently the multitude of abnormal surfaces heavily impacts the SHAR estimates.

In Panel B of Figure 4, the parameter estimate $\hat{\beta}^{(1)}_{20|t}$ in the SHAR model hovers around 0.5 between 2016 and 2018, then drop to just above 0.4, and jumps to approximately 0.7 at the start of the COVID-19 crisis (March 2020). In contrast, the SHAR-Robust estimates are stable over the entire period. In Section 6, we investigate how these differences impact the forecasts of the IV surfaces.

5.3 A Closer Look at Dynamic Option Pricing Models

Dynamic models with stochastic volatility and jumps take a central place in the option pricing literature. We now take a closer look at the relative performance of these models, the role of different model components, and their computing times. Figure 5 highlights the value of more complex option pricing models by comparing the daily fit of the SV(1) model with that of the SVDE model, which also includes double exponential jumps. The three-dimensional figure in Panel A plots the difference in IVRMSE fit between SV(1) and SVDE for every day between 2016 and 2021 as a function of moneyness and maturity. The jump model clearly performs better 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. These findings are consistent with the existing option pricing literature.

Panel B of Figure 5 plots the daily IVRMSE overall fit ratio between the SV(1) and SVDE models for 2016-2021. Apart from a few days, the ratio exceeds one. When abnormal

Figure 4: Time Series of SHAR Parameter Estimates (OptionMetrics model)

(a) Panel A: First-Lag Parameter, One-Day Horizon ($\hat{\beta}^{(1)}_{1|t}$)

(b) Panel B: First-Lag Parameter, Twenty-Day Horizon ($\hat{\beta}^{(1)}_{20|t}$)

Notes: 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 in this figure is set to $h=1$ (top panel) and $h=20$ (bottom panel). The graphics show the parameter estimates related to the lagged predicted IV surface over time. The detected outliers are indicated by vertical gray bars.

IV surfaces are detected, the IVRMSE ratio for the SV(1) model is sometimes twice that of the SVDE model. This confirms that jumps can capture sudden changes in the IV surface. Finally, the performance of the SV(1) model deteriorates between July and October of 2021, when IVRMSE ratios frequently exceed 1.5.

Figure 6 plots the daily variance estimates of the stochastic volatility process in equation (11) for various option pricing models. The time series for the stochastic volatility models that do not include jumps are similar and close to Figure 2, as expected. The models with jumps (SVJR and SVDE) result in slightly smaller variance estimates because the jump component captures some of the large movements in the option surfaces.

Table 5 reports the average computing time (in seconds) to fit one daily IV surface with the various models we consider. The SHAR computing time is not incorporated in this Table, because the model parameters are estimated sequentially in closed form, and therefore the estimation time is identical for all surface fit models. The non-parametric OptionMetrics and random forest models are very fast compared to the neural network and especially compared to the dynamic option pricing models. For example, it takes about 3 seconds to fit the IV surface for the random forest, compared to 20 seconds and 4 minutes for the ANN(3) and SVDE models respectively. We conclude that when implementing high frequency, e.g. five minutes, streaming IV surface forecasting, dynamic option pricing models are prohibitively slow.

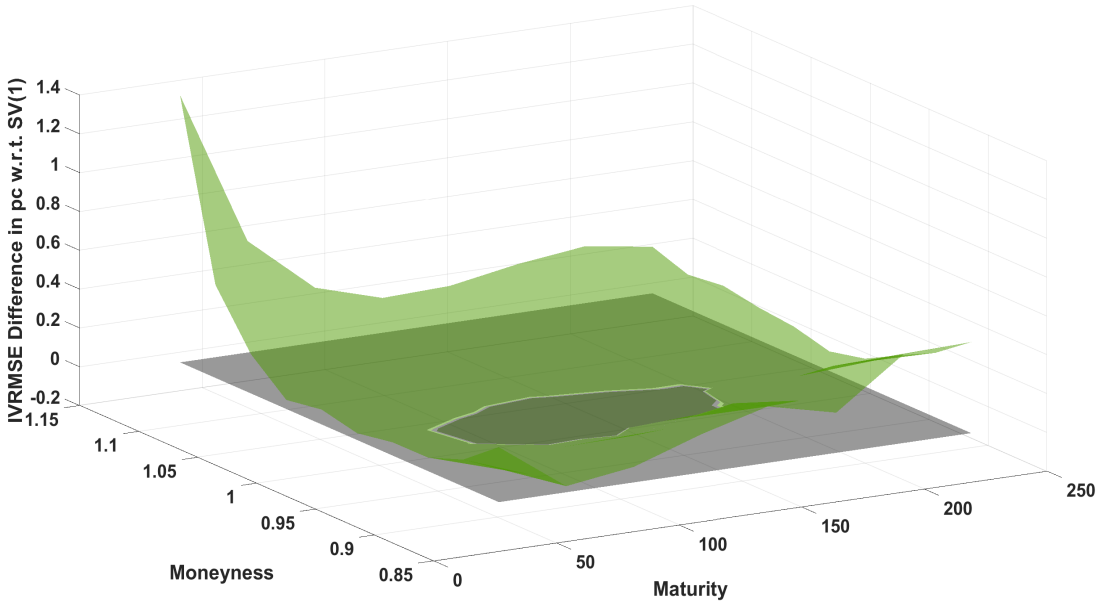
Table 5: Computing Time for Various Option Fitting Models.

Parametric Models		**Non-Parametric Models**		**Dynamic 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

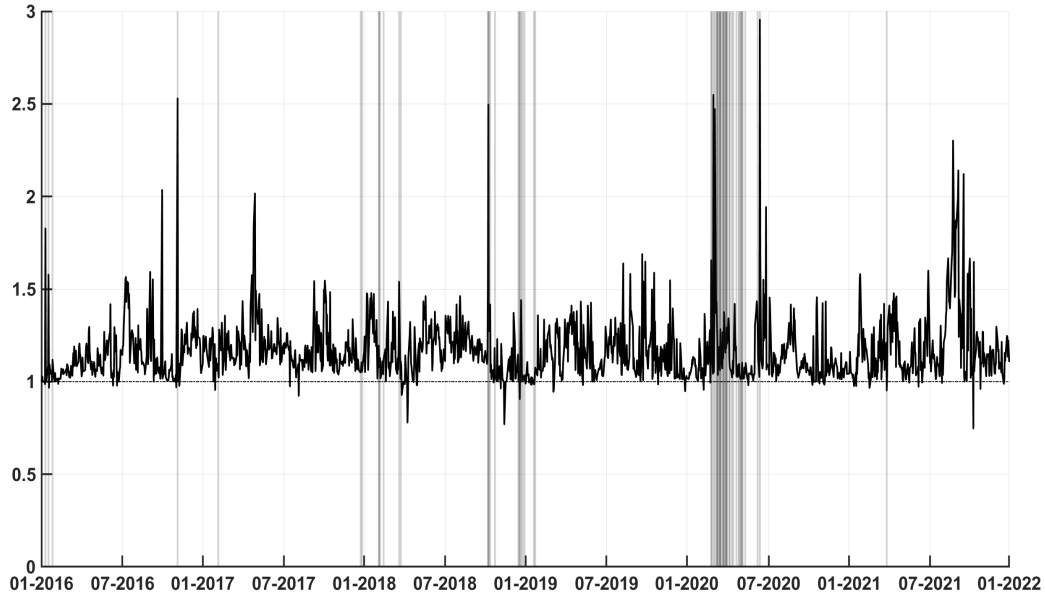
This table reports the average time (expressed in seconds) for estimating the models on one day of options. We compute the average over daily options using 2015-2021 data. Opt-Met and RF stand for OptionMetrics and random forest, respectively.

Figure 5: Daily IVRMSE Fit. SV(1) versus SVDE

(a) Panel A: SV(1) - SVDE IVRMSE



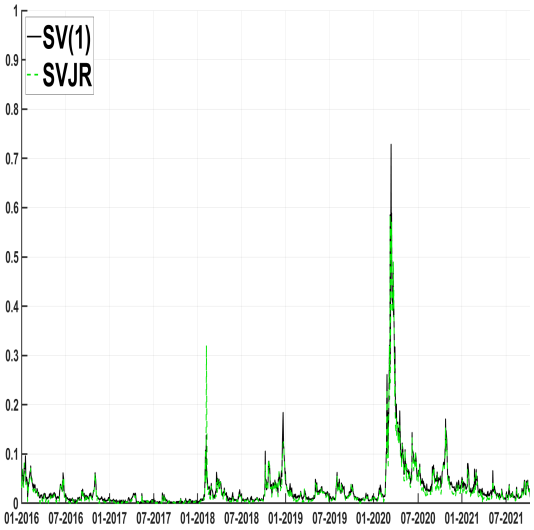
(b) Panel B: Daily Time Series of SV(1)/SVDE IVRMSE Ratio



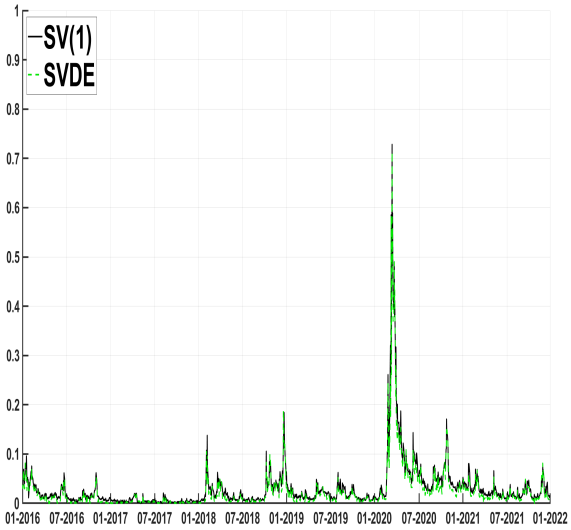
Notes: Panel A plots the difference between the IVRMSEs for the SV(1) and SVDE models on average over the sample. A positive value means that the SVDE model outperforms the SV(1) model. Panel B shows the ratio of daily IVRMSEs between the SV(1) and the SVDE models. Values higher than one indicate that SVDE improves over SV(1).

Figure 6: The Time Series of Spot Variances in Dynamic Option Pricing Models

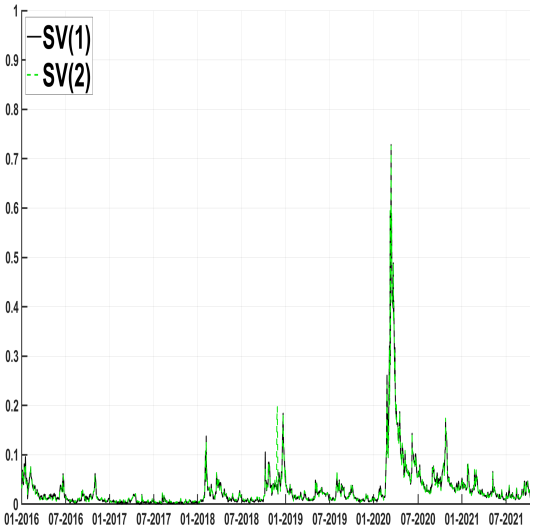
(a) Panel A: SV(1) - SVJR



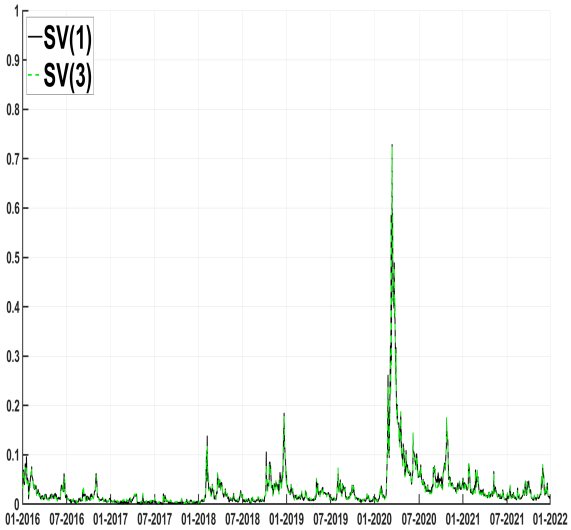
(b) Panel B: SV(1) - SVDE



(c) Panel C: SV(1) - SV(2)



(d) Panel D: SV(1) - SV(3)



Notes: We plot the spot variance for various option pricing models at the daily frequency. For multi-factor models, the spot variance is given by the sum of the latent factors.

6 Forecasting IV Surfaces: Empirical Results

This section contains our main empirical results on out-of-sample forecasting of IV surfaces. Section 6.1 discusses the implementation of the forecasting exercise. Section 6.2 reports the forecasting performance of the proposed SHAR approach for multiple forecast horizons. Section 6.3 shows that refinements to the SHAR approach can further improve forecast performance.

6.1 Implementation

We start the forecasting exercise in 2016 and implement it sequentially until the end of 2021. Every day, when a new IV surface becomes available, we fit it using all the models we consider, we update the sequential estimator for the dynamic surface SHAR model for three forecast horizons (1 day, 5 days, and 20 days), and we produce the forecasts. We compare the random walk (RW), SHAR, and SHAR-Robust forecasts. Table 6 reports the results for the one-day horizon, Table 7 reports the results for the 5-day (one-week) ahead IV surface forecasts, and Table 8 reports on the 20-day (one-month) ahead IV surface forecasts.

In all three tables, Panel A evaluates forecast performance using the out-of-sample implied volatility root-mean-square-error (IVRMSE) for day 1 to t, which is defined as $100\sqrt{\sum_{l=1}^{t}\sum_{i=1}^{N_l}\left(IV(O_{i,l})-\widehat{IV}(O_{i,l})\right)^2/\sum_{l=1}^{t}N_l}$. Panel B reports on forecast performance in dollar RMSE ($RMSE) defined as $100\sqrt{\sum_{l=1}^{t}\sum_{i=1}^{N_l}\left(\left(IV(O_{i,l})-\widehat{IV}(O_{i,l})\right)\text{vega}_{i,l}\right)^2/\sum_{l=1}^{t}N_l}$, where $\text{vega}_{i,l}$ is the option's vega. The latter loss function provides additional economic content compared to the IVRMSE loss in Panel A. We report the IVRMSEs and dollar RMSEs for every year in the sample separately to understand the performance of our framework over time.

6.2 SHAR Forecasting Performance

The key question we seek to answer is if dynamic modeling of the IV surface is superior to the RW assumption. The results in Tables 6, 7, and 8 indicate that the answer is a resounding yes. To facilitate the interpretation of the results, we report the IV and dollar loss for the RW forecast, but the row labeled SHAR reports the ratio of the SHAR and RW forecasts, and the row labeled SHAR-R. reports the ratio of the SHAR-Robust and RW forecasts. Our key finding is reflected in the fact that the large majority of these ratios are below one. We test for equal predictive ability between the RW and SHAR approaches using Clark and West (2007) and find that all differences in favor of SHAR are statistically significant at the five percent level. Comparing the ratios for the SHAR and SHAR-Robust approaches, the latter typically further reduces the SHAR IVRMSEs and dollar RMSEs. We conclude that updating the sequential estimator exclusively with normal IV surfaces provides modest additional benefits.

In addition to the main finding, another important conclusion is that there is quite a bit of variation in these results in several dimensions. The most important observation is that the ratios in Table 7 and especially Table 8 are smaller than the corresponding ratios in Table 6. The benefits of the SHAR approach are thus more important at longer forecast horizons. For example, in the case of the SVDE option pricing model, the $RMSE ratios are 0.98, 0.95 and 0.90 for the next day, week, and month forecast horizons respectively. This finding is robust across all models. Importantly, this conclusion is entirely consistent with our priors, because given the properties of the time series of the variance, we would expect the benefits of dynamic modeling of the surface to be more important for longer forecast horizons.

A second, perhaps more surprising, source of variation in the results is across models. The benefits of dynamic modeling are very pronounced for the ANN(3) and Random Forest models, but smaller for the other models. Note that the improvements due to dynamic mod-

eling for the ANN(3) and Random Forest models are interesting given the RW benchmarks. The RW forecast associated with the Random Forest model is a consistently good performer compared to other models.[12] In contrast, the RW forecast associated with the ANN(3) model consistently underperforms the other models, especially for the dollar loss function in Panel B of all three tables.

The differences in performance between the dynamic option pricing models are small but relatively robust. For the RW benchmarks, the SV(2) model leads to systematic improvements over SV(1), but the benefits from including a third factor are minor. Moreover, the addition of Normal or double exponential jumps provides minor improvements over the SV(1) model but not over the multi-factor models. The improvements resulting from the SHAR and SHAR-Robust forecasts are relatively similar across all these models for a given forecast horizon, both in Panel A and Panel B.

Perhaps unsurprisingly, both the quality of the benchmark RW forecast as well as the improvement provided by the SHAR and SHAR-Robust forecasts also greatly change over time. Both the IV and the dollar loss substantially increase in 2020 compared to 2016-2019, regardless of the forecast horizon. However, in 2021 the losses revert back to pre-Covid levels. Consistent with intuition, in the volatile and chaotic Covid year 2020, the RW forecast sometimes outperforms the SHAR and SHAR-Robust forecasts. For example, RW forecasts generally dominate the SHAR approach for the 5-day forecast horizon.

The results for the more economically intuitive dollar loss function in Panel B of the three tables are overall very consistent with those for the IVRMSE in Panel A. If anything, the results are strengthened. This is most obvious when considering the ANN(3) and Random Forest models, where the improvements from SHAR and SHAR-Robust are largest. For instance, for the 5-day horizon, the ratio between the SHAR-R. and RW dollar loss in Panel B is 0.66 (0.78) for the ANN(3) and Random Forest models respectively in 2016, compared

[12]However, note that the improvements in out-of-sample performance provided by the Random Forest approach are small compared to the in-sample fit improvement in Table 2.

to 0.81 (0.83) in Panel A for the IVRMSE ratio.

Finally, the SHAR-Robust approach only updates the streaming estimator when IV surfaces are classified as normal, but Tables 6-8 report on the forecast performance based on all available days, including the detected abnormal surface days. Panels B of Tables A.1-A.3 in the Appendix provide additional insight by assessing the performance of all methods for normal periods only. As expected given the evidence on outliers in Table 3, this lowers the IVRMSEs in 2018 and 2020. In those two years, the jump models (SVDE and SVJR) are now more strongly outperformed by the multi-factor stochastic volatility models, highlighting that the jump models are particularly well-suited for capturing abnormal surfaces.

Table 6: Out-of-sample RMSEs in IV and Dollars (1-day horizon)

	Panel A: IV loss						**Panel B: Dollar loss**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	4.75	4.45	4.84	4.47	6.53	5.29	7.81	8.10	11.02	10.88	22.88	21.70
				Model - AHBS						*Model - AHBS*		
RW	1.34	1.11	1.94	1.35	3.57	1.67	**2.61**	**2.41**	5.36	**3.96**	12.55	7.89
SHAR	**1.00**	1.00	**0.99**	**0.99**	1.00	0.99	1.00	1.01	**1.00**	1.00	1.01	1.00
SHAR-R.	1.01	**0.99**	1.00	1.00	**0.94**	**0.99**	1.02	1.00	1.00	1.01	**0.94**	**0.99**
				Model - ANN(3)						*Model - ANN(3)*		
RW	1.25	0.83	1.85	1.21	3.51	1.50	2.89	2.17	6.17	4.10	12.58	7.63
SHAR	0.90	0.96	0.85	**0.82**	0.91	0.83	0.84	0.86	0.70	0.76	0.89	0.83
SHAR-R.	**0.90**	**0.95**	**0.85**	0.82	**0.90**	**0.83**	**0.83**	**0.86**	**0.70**	**0.76**	**0.88**	**0.82**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	1.16	0.76	1.76	1.13	3.44	1.43	2.36	1.70	4.93	3.48	12.15	6.98
SHAR	**1.00**	0.99	0.99	**0.99**	1.00	**0.99**	**1.00**	0.99	0.99	**1.00**	1.01	**0.99**
SHAR-R.	1.00	**0.99**	**0.98**	1.00	**0.96**	0.99	1.00	**0.99**	**0.98**	1.00	**0.97**	0.99
				Model - Random Forest						*Model - Random Forest*		
RW	1.18	0.78	1.77	1.15	3.49	1.44	2.43	1.76	4.99	3.54	12.33	7.03
SHAR	**0.89**	**0.97**	**0.87**	0.82	0.92	0.83	**0.84**	0.91	**0.83**	0.77	0.90	**0.80**
SHAR-R.	0.89	0.97	0.87	**0.82**	**0.90**	**0.83**	0.84	**0.91**	0.83	**0.77**	**0.88**	0.80
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	1.28	0.91	1.89	1.19	3.60	1.48	2.54	1.99	5.19	3.66	12.75	7.29
SHAR	**0.99**	0.99	**0.98**	**0.99**	1.00	0.99	**0.99**	0.96	**0.98**	**1.00**	1.00	0.99
SHAR-R.	1.00	**0.98**	0.99	1.00	**0.96**	**0.98**	1.00	**0.96**	1.00	1.01	**0.97**	**0.98**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	1.21	0.84	1.82	1.16	3.55	1.48	2.45	1.84	5.05	3.53	12.54	7.16
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.97	**0.99**	**1.00**	1.01	0.99
SHAR-R.	1.01	**0.97**	1.00	1.00	**0.98**	**0.98**	1.01	**0.95**	1.00	1.01	**0.99**	**0.98**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	1.21	0.82	1.81	1.16	3.52	1.47	2.43	1.80	5.04	3.53	12.46	7.15
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.99	**0.99**	**1.00**	1.01	0.99
SHAR-R.	1.00	**0.98**	1.00	1.00	**0.98**	**0.98**	1.00	**0.97**	1.00	1.01	**0.99**	**0.98**
				Model - SVJR						*Model - SVJR*		
RW	1.27	0.87	1.86	1.18	3.52	1.48	2.53	1.96	5.14	3.65	12.65	7.18
SHAR	**0.99**	0.99	**0.98**	**0.99**	1.00	0.98	**0.99**	0.96	**0.98**	**1.00**	1.00	0.99
SHAR-R.	0.99	**0.98**	1.00	1.00	**0.99**	**0.98**	0.99	**0.95**	1.00	1.01	**0.99**	**0.98**
				Model - SVDE						*Model - SVDE*		
RW	1.24	0.85	1.83	1.17	3.51	1.46	2.51	1.91	5.11	3.62	12.56	7.17
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.97	**0.99**	**1.00**	1.01	0.99
SHAR-R.	1.00	**0.98**	1.00	1.00	**0.93**	**0.98**	0.99	**0.96**	1.00	1.00	**0.94**	**0.98**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the next-day implied volatility surface. The option pricing models are trained using the vega loss function. Differences in IV loss between RW and SHAR are statistically significant at the five percent level according to the Clark and West (2007) test.

Table 7: Out-of-sample RMSEs in IV and Dollars (5-day horizon).

	Panel A: IV loss						**Panel B: Dollar Loss**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	4.97	4.52	5.24	4.70	7.90	5.51	8.60	8.35	12.40	12.06	28.49	23.30
				Model - AHBS						*Model - AHBS*		
RW	2.34	1.42	3.20	2.26	**6.25**	2.59	4.66	3.19	8.71	6.73	**22.48**	12.50
SHAR	0.98	1.01	0.98	0.94	1.07	0.97	1.00	1.05	0.99	0.97	1.07	0.98
SHAR-R.	**0.97**	**0.98**	**0.98**	**0.94**	1.04	**0.95**	**0.98**	**1.00**	**0.99**	**0.97**	1.05	**0.96**
				Model - ANN(3)						*Model - ANN(3)*		
RW	2.32	1.40	3.15	2.21	6.16	2.50	5.65	4.23	9.86	7.35	22.61	12.69
SHAR	0.81	0.85	**0.83**	0.80	**0.93**	0.80	0.67	0.62	**0.72**	0.72	**0.92**	0.77
SHAR-R.	**0.81**	**0.85**	0.84	**0.79**	0.94	**0.79**	**0.66**	**0.62**	0.72	**0.72**	0.93	**0.76**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	2.22	1.18	3.07	2.12	**6.08**	2.46	4.56	2.69	8.55	6.50	**22.18**	12.05
SHAR	0.98	**0.98**	0.97	0.95	1.08	0.97	0.98	**0.99**	0.99	0.97	1.08	0.97
SHAR-R.	**0.98**	0.98	**0.97**	**0.95**	1.05	**0.95**	**0.98**	0.99	**0.98**	**0.97**	1.05	**0.95**
				Model - Random Forest						*Model - Random Forest*		
RW	2.25	1.19	3.07	2.13	6.12	2.47	4.63	2.71	8.60	6.55	22.35	12.12
SHAR	**0.83**	**0.95**	**0.84**	0.81	0.93	0.81	**0.78**	**0.90**	**0.80**	0.78	0.91	0.78
SHAR-R.	0.83	0.95	0.84	**0.81**	**0.92**	**0.80**	0.78	0.90	0.81	**0.78**	**0.90**	**0.77**
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	2.30	1.26	3.09	2.15	**6.12**	2.48	4.64	2.80	8.64	6.57	**22.43**	12.26
SHAR	0.97	1.00	0.97	0.95	1.07	0.97	0.99	0.99	0.98	0.97	1.07	0.97
SHAR-R.	**0.97**	**0.98**	**0.97**	**0.94**	1.04	**0.94**	**0.98**	**0.97**	**0.98**	**0.96**	1.04	**0.94**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	2.25	1.21	3.07	2.12	**6.09**	2.48	4.60	2.72	8.58	6.51	**22.31**	12.18
SHAR	0.97	0.99	**0.98**	0.95	1.07	0.97	0.98	0.99	**0.98**	0.97	1.07	0.97
SHAR-R.	**0.96**	**0.97**	0.98	**0.94**	1.06	**0.94**	**0.97**	**0.96**	0.99	**0.97**	1.06	**0.94**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	2.25	1.20	3.06	2.12	**6.09**	2.48	4.58	2.70	8.57	6.51	**22.26**	12.19
SHAR	0.98	0.99	0.98	0.95	1.07	0.97	0.99	1.01	**0.98**	0.97	1.07	0.97
SHAR-R.	**0.97**	**0.97**	**0.98**	**0.94**	1.06	**0.94**	**0.97**	**0.97**	0.98	**0.97**	1.06	**0.94**
				Model - SVJR						*Model - SVJR*		
RW	2.29	1.23	3.09	2.15	**6.11**	2.48	4.62	2.79	8.60	6.57	**22.47**	12.19
SHAR	0.97	1.00	0.97	0.95	1.07	0.97	0.99	0.99	0.98	0.97	1.07	0.97
SHAR-R.	**0.96**	**0.99**	**0.97**	**0.94**	1.05	**0.94**	**0.98**	**0.97**	**0.98**	**0.96**	1.05	**0.95**
				Model - SVDE						*Model - SVDE*		
RW	2.26	1.22	3.07	2.13	**6.08**	2.47	4.62	2.76	8.57	6.56	**22.35**	12.19
SHAR	0.98	0.99	0.97	0.95	1.07	0.97	0.98	0.99	**0.98**	0.97	1.07	0.97
SHAR-R.	**0.97**	**0.97**	**0.97**	**0.94**	1.03	**0.94**	**0.97**	**0.97**	0.98	**0.96**	1.04	**0.95**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the five day ahead implied volatility surface. The option pricing models are trained using the vega loss function. Differences in IV loss between RW and SHAR are statistically significant at the five percent level according to the Clark and West (2007) test.

Table 8: Out-of-sample RMSEs in IV and Dollars (20-day horizon).

	IV Loss						**Dollar Loss**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	5.35	4.59	5.96	5.16	12.11	5.80	9.87	8.73	14.92	14.36	44.01	25.72
				Model - AHBS						*Model - AHBS*		
RW	3.25	**1.69**	4.65	3.46	12.13	3.40	**6.63**	**4.03**	12.95	10.77	42.41	17.11
SHAR	1.01	1.11	**0.93**	0.88	**0.92**	0.95	1.04	1.14	**0.93**	0.90	**0.92**	0.94
SHAR-R.	**0.99**	1.07	0.93	**0.88**	0.94	**0.92**	1.01	1.08	0.94	**0.90**	0.94	**0.92**
				Model - ANN(3)						*Model - ANN(3)*		
RW	3.26	1.61	4.48	3.42	11.91	3.34	7.31	4.88	13.72	11.43	42.23	17.44
SHAR	0.87	0.89	**0.81**	0.77	**0.86**	0.80	0.78	0.70	**0.74**	0.73	**0.86**	0.76
SHAR-R.	**0.86**	**0.89**	0.82	**0.76**	0.86	**0.78**	**0.77**	**0.70**	0.74	**0.72**	0.86	**0.74**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	**3.16**	**1.48**	4.43	3.35	11.83	3.29	**6.60**	**3.62**	12.82	10.68	41.84	16.89
SHAR	1.01	1.08	0.93	0.89	**0.93**	0.94	1.02	1.08	0.93	0.90	**0.93**	0.91
SHAR-R.	1.01	1.08	**0.92**	**0.89**	0.94	**0.92**	1.02	1.08	**0.93**	**0.90**	0.94	**0.90**
				Model - Random Forest						*Model - Random Forest*		
RW	3.18	1.49	4.44	3.37	11.84	3.30	6.66	3.63	12.87	10.74	41.87	16.93
SHAR	0.86	0.96	**0.82**	0.77	0.87	0.81	0.84	0.90	**0.79**	0.76	0.86	0.77
SHAR-R.	**0.86**	**0.96**	0.82	**0.77**	**0.87**	**0.80**	**0.84**	**0.90**	0.79	**0.75**	**0.86**	**0.76**
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	**3.20**	**1.55**	4.44	3.35	11.71	3.29	**6.59**	**3.69**	12.87	10.67	41.58	17.02
SHAR	1.01	1.11	**0.92**	0.90	**0.94**	0.94	1.04	1.09	**0.93**	0.91	**0.94**	0.91
SHAR-R.	1.00	1.08	0.93	**0.88**	0.95	**0.92**	1.02	1.05	0.93	**0.90**	0.95	**0.90**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	3.17	**1.50**	4.42	3.34	11.70	3.29	**6.57**	**3.64**	12.84	10.66	41.57	16.92
SHAR	1.01	1.10	**0.92**	0.90	**0.94**	0.94	1.03	1.09	**0.93**	0.91	**0.94**	0.91
SHAR-R.	**1.00**	1.06	0.93	**0.88**	0.95	**0.92**	1.00	1.04	0.94	**0.89**	0.95	**0.90**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	3.17	**1.50**	4.42	3.34	11.69	3.29	**6.57**	**3.62**	12.82	10.65	41.53	16.95
SHAR	1.01	1.10	**0.93**	0.90	**0.94**	0.94	1.03	1.10	**0.93**	0.91	**0.94**	0.91
SHAR-R.	**1.00**	1.07	0.93	**0.88**	0.95	**0.92**	1.01	1.05	0.94	**0.89**	0.95	**0.90**
				Model - SVJR						*Model - SVJR*		
RW	3.20	**1.52**	4.47	3.37	11.75	3.30	**6.55**	**3.67**	12.85	10.67	41.81	16.93
SHAR	1.01	1.11	**0.92**	0.89	**0.93**	0.94	1.04	1.09	**0.93**	0.91	**0.93**	0.91
SHAR-R.	**1.00**	1.09	0.92	**0.88**	0.95	**0.91**	1.02	1.06	0.93	**0.90**	0.95	**0.90**
				Model - SVDE						*Model - SVDE*		
RW	**3.17**	**1.51**	4.42	3.36	11.71	3.29	**6.56**	**3.65**	12.81	10.69	41.63	16.94
SHAR	1.01	1.10	**0.92**	0.90	**0.93**	0.94	1.03	1.09	**0.93**	0.91	**0.93**	0.91
SHAR-R.	1.00	1.08	0.93	**0.88**	0.95	**0.91**	1.01	1.06	0.94	**0.90**	0.95	**0.90**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the twenty day ahead implied volatility surface. The option pricing models are trained using the vega loss function. Differences in IV loss between RW and SHAR are statistically significant at the five percent level according to the Clark and West (2007) test.

6.3 Refining the SHAR Forecast

The SHAR results are based on a parsimonious model with four parameters for fitting the entire surface. While this configuration already improves over the RW approach, it is natural to ask whether these improvements are uniform across the IV surface and whether allowing for local SHAR specifications could further improve predictive performance.

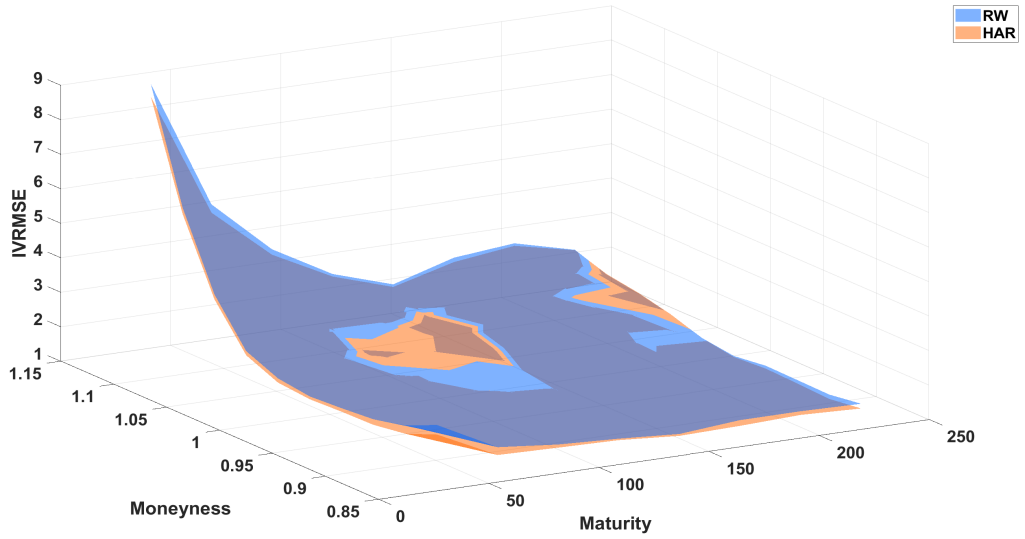
Figure 7 addresses the first question by plotting the IVRMSE surface using 2016–2021 data for the RW (blue) and SHAR (orange) forecasting approaches, based on the SV(1) and SVDE option pricing models. For the SV(1) model in Panel A, the surface is predominantly blue away from the center of the IV surface, indicating lower IVRMSEs for SHAR at moneyness levels far from one across all maturities. For at-the-money options with mid-range maturities (30-120 days), SHAR and RW perform similarly, with a few localized regions where RW slightly outperforms.

The results for the SVDE model in Panel B convey a similar message. SHAR dominates RW over large portions of the surface, while differences are smaller for moneyness levels close to one and medium maturities, as well as for some out of the money options with maturities exceeding six months. Overall, the figure shows that the gains from SHAR are substantial but not entirely uniform across the IV surface. In addition, irrespective of the surface fit model or forecasting method, a sizable contribution to total IVRMSE originates from options with short maturities and high moneyness, which are particularly difficult to forecast.

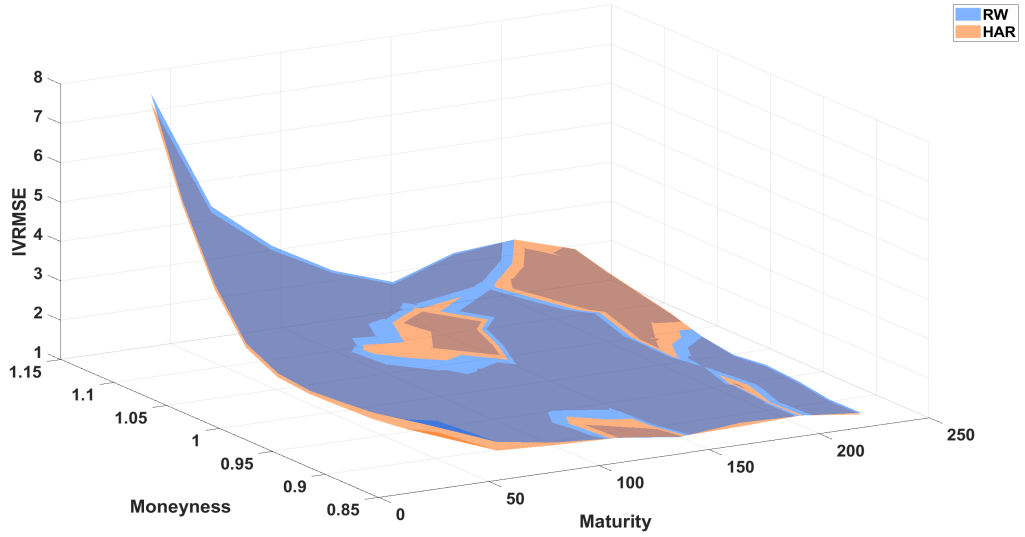
These observations motivate the consideration of more flexible SHAR specifications that allow parameters to vary across regions of the IV surface. By increasing model flexibility, such refinements may better capture local surface dynamics in areas where the gains from the global SHAR specification are weaker. To illustrate this idea, we consider two refinements based on local surface splits. The first refinement allows SHAR parameters to differ according to option maturity, more or less than 30 days, and moneyness, based on the thresholds [0.95; 1; 1.05]. This specification, denoted SHAR-REF1, yields eight local models with four

Figure 7: The IVRMSE surface. RW and SHAR Next Day Surface Forecasts

(a) Panel A: SV(1)



(b) Panel B: SVDE



Notes: We plot the IVRMSE surface, on average for 2016-2021, implied by the RW (blue) and SHAR (orange) forecasting approaches for the SV(1) and SVDE models.

parameters each, estimated in the same way as the baseline SHAR model. Table 9 reports the results. To save space, we focus on a selected subset of surface fit models. The AHBS model benefits the most from this refinement. SHAR-REF1 outperforms SHAR in all years for the one day forecast horizon, in five out of six cases for the five day forecast horizon and in four out of six cases for the twenty day forecast horizon. For example, in 2016 the AHBS one day ahead IVRMSE declines from 1.34 to 1.24, an 8% reduction. The dynamic option pricing models also benefit, although not in all years.

Another possible refinement of the SHAR approach adapts the model to each option characteristic. That is, the model differs for each point on the surface. We denote this approach as SHAR-REF2. Table 9 shows that this approach yields inferior results compared to SHAR for all years and forecast horizons. While this model is highly flexible, this flexibility does not translate into improved out-of-sample fit. Note that the computational cost of this approach is very high, since it requires generating the full history of IVs for each option using the surface model.

We conclude that the benefits offered by refinements to the SHAR approach depend on the particular implementation. We find that while moderate increases in flexibility can improve performance in some cases, more flexible refinements are not necessarily preferable.

7 Comparison with Existing Approaches

In this section, we compare the SHAR approach with some existing approaches in the literature. Section 7.1 investigates if the error correction method proposed by Almeida et al. (2023) provides additional improvements when implemented in conjunction with our approach. Section 7.2 compares our forecasting approach with the methodology of Goncalves and Guidolin (2006), who model the dynamics of the IV surfaces through the parameters of the surface fit models rather than through the surfaces themselves.

Table 9: Out-of-sample IVRMSEs for multiple SHAR specifications.

	Panel A: 1-day horizon						**Panel B: 5-day horizon**						**Panel C: 20-day horizon**					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
	Model - AHBS						*Model - AHBS*						*Model - AHBS*					
SHAR	1.34	1.11	1.93	1.34	3.58	1.65	2.29	1.43	3.14	2.14	**6.66**	2.52	3.28	1.87	4.30	3.06	**11.12**	3.24
SHAR-REF1	**1.24**	**0.98**	**1.82**	**1.22**	**3.50**	**1.54**	**2.09**	**1.34**	**2.98**	**1.89**	6.93	**2.40**	**2.54**	2.01	**3.88**	**2.40**	14.17	**3.19**
SHAR-REF2	1.62	1.25	2.35	1.60	3.91	1.99	2.46	1.46	3.35	2.31	6.87	2.62	3.25	**1.69**	4.58	3.41	12.26	3.29
	Model - OptionMetrics						*Model - OptionMetrics*						*Model - OptionMetrics*					
SHAR	1.15	0.75	1.74	1.13	**3.45**	1.42	2.17	**1.15**	2.98	2.02	**6.54**	2.38	3.19	1.60	4.10	3.00	**10.98**	**3.10**
SHAR-REF1	**1.14**	**0.75**	**1.73**	**1.11**	3.49	**1.40**	**2.03**	1.18	**2.91**	**1.89**	7.04	**2.33**	**2.51**	1.85	**3.80**	**2.54**	11.70	3.15
SHAR-REF2	1.46	0.94	2.17	1.42	3.86	1.80	2.33	1.23	3.19	2.16	6.76	2.49	3.15	**1.48**	4.33	3.29	11.94	3.16
	Option pricing model - SV(1)						*Option pricing model - SV(1)*						*Option pricing model - SV(1)*					
SHAR	1.27	0.90	1.86	**1.19**	**3.60**	**1.46**	2.24	**1.26**	3.01	2.04	**6.54**	2.40	3.24	1.71	4.10	3.01	10.96	**3.10**
SHAR-REF1	**1.24**	**0.89**	**1.83**	1.19	3.64	1.47	**2.10**	1.29	**2.94**	**1.95**	7.24	**2.38**	**2.57**	1.95	**3.80**	**2.65**	**10.73**	3.15
SHAR-REF2	1.56	1.05	2.26	1.46	3.96	1.84	2.40	1.30	3.23	2.18	6.80	2.50	3.18	**1.53**	4.35	3.29	11.85	3.16
	Option pricing model - SV(2)						*Option pricing model - SV(2)*						*Option pricing model - SV(2)*					
SHAR	1.21	0.83	1.80	**1.15**	**3.55**	1.45	2.20	**1.20**	2.99	2.02	**6.53**	2.40	3.20	1.65	4.09	3.00	10.94	**3.10**
SHAR-REF1	**1.19**	**0.82**	**1.78**	1.15	3.63	**1.45**	**2.06**	1.22	**2.92**	**1.93**	7.21	**2.37**	**2.55**	1.88	**3.79**	**2.70**	**10.64**	3.15
SHAR-REF2	1.51	0.99	2.23	1.43	3.92	1.83	2.36	1.25	3.21	2.16	6.76	2.50	3.15	**1.50**	4.34	3.28	11.83	3.16

Each day, we estimate the various models using the available options for that day, and we forecast the next-day implied volatility surface. The option pricing models are trained using the vega loss function.

7.1 Enhancing the Surface Fit with Deep Learning

Our proposed methodology consists of a two-step procedure. The first step involves fitting an IV model on a daily basis. The second step adds dynamics that are used to forecast IVs using the SHAR forecasting setup. Almeida et al. (2023) propose a very different two-step approach for enhancing the fit of the IV surface. It proceeds as follows:

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

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

Conceptually this two-step procedure is complementary to our SHAR forecasting setup, which combines daily surface fits with time-series dynamics. We can therefore apply our approach in conjunction with the extra step in Almeida et al. (2023), effectively yielding a three-step procedure. Next we investigate if the IVRMSE improvements provided by the SHAR forecasting approach differ from the non-linearities captured by the ANN applied to the residuals. We implement the neural network step using an ANN(3) consisting of 3 hidden layers with 32, 16 and 8 neurons, each activated by a sigmoid function as in Almeida et al. (2023).

Table 10 shows the IVRMSEs for various models for one-step ahead forecasts. We either rely on the RW approach, the two-step neural network error correction procedure, or the SHAR approach applied in conjunction with this two-step procedure, which is effectively a three-step procedure. We report results on a year-by-year basis. There are two important findings. First, we confirm the findings of Almeida et al. (2023) since their two-step approach improves the predictive performance of any model we consider in each sample year. Even the advanced option pricing models such as the SV(2) and SVDE models exhibit significant

improvements in predictive accuracy when combined with ANN(3). Second, the three-step procedure which dynamically models the surfaces with the SHAR approach further lowers the IVRMSEs. For each year and for all the models except the AHBS model in 2016, the SHAR approach provides additional gains. We conclude that our approach is complementary to the one proposed by Almeida et al. (2023), in the sense that both approaches provide IVRMSE improvements, but they do so by capturing different stylized facts. The ANN model captures additional non-linearities, while the SHAR forecasting approach captures the temporal persistence. The five-day ahead IV surface forecast results in Table 11 confirm this conclusion. In fact, the gains from dynamic surface modeling when combined with deep learning and a surface fit model are larger compared to the next-day forecast results in every year, except for 2020.

7.2 Forecasting Surfaces or Forecasting Parameters?

We find that we can significantly outperform the RW approach when we fit IV surfaces on a daily basis and use them as predictors in the time-series SHAR model. Recall that the RW forecasting approach is based on the following specification:

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

Note from equation (15) that the optimal model prediction is $IV^M(O_{i,t+h}, \boldsymbol{\Theta}_{t+h})$. However, since we do not know $\boldsymbol{\Theta}_{t+h}$ on day t, the RW approach defaults to using the most recent parameter estimates, $\hat{\boldsymbol{\Theta}}_t$. Since the information set at time t includes all the parameter estimates $\hat{\boldsymbol{\Theta}}_{1:t} = \{\hat{\boldsymbol{\Theta}}_1, \ldots, \hat{\boldsymbol{\Theta}}_t\}$, the RW approach can therefore also be understood as a process with a martingale property, i.e. $\mathbb{E}(\hat{\Theta}_{t+h}|\hat{\boldsymbol{\Theta}}_{1:t}) = \hat{\boldsymbol{\Theta}}_t$. Goncalves and Guidolin (2006) relax this RW assumption on the model parameters and instead suggest forecasting the model parameters using a VAR(p) model: $\hat{\boldsymbol{\Theta}}_{t+h} = \boldsymbol{\gamma}_0 + \sum_{i=1}^{p} \boldsymbol{\Gamma}_i \hat{\boldsymbol{\Theta}}_{t+1-i} + \boldsymbol{\eta}_{t+h}$, where the number

Table 10: IVRMSEs for the Two- and Three-Step Procedures (1-Day Horizon).

Model	2016	2017	2018	2019	2020	2021
			Model - AHBS			
RW	1.34	1.11	1.94	1.35	3.57	1.67
ANN(3)+RW	1.34	1.10	1.92	1.28	3.55	1.57
ANN(3)+SHAR	**1.34**	1.09	**1.90**	**1.27**	3.55	1.54
ANN(3)+SHAR-R.	1.36	**1.08**	1.90	1.28	**3.33**	**1.54**
			Model - SV(1)			
RW	1.28	0.91	1.89	1.19	3.60	1.48
ANN(3)+RW	1.26	0.87	1.83	1.19	3.53	1.46
ANN(3)+SHAR	**1.25**	0.87	**1.80**	**1.17**	3.52	1.44
ANN(3)+SHAR-R.	1.26	**0.86**	1.84	1.18	**3.42**	**1.43**
			Model - SV(2)			
RW	1.21	0.84	1.82	1.16	3.55	1.48
ANN(3)+RW	1.19	0.82	1.78	1.15	3.52	1.48
ANN(3)+SHAR	**1.19**	0.81	**1.76**	**1.14**	3.51	1.45
ANN(3)+SHAR-R.	1.22	**0.80**	1.79	1.15	**3.45**	**1.44**
			Model - SVJR			
RW	1.27	0.87	1.86	1.18	3.52	1.48
ANN(3)+RW	1.26	0.84	1.80	1.19	3.50	1.45
ANN(3)+SHAR	**1.25**	0.84	**1.78**	**1.17**	3.50	1.43
ANN(3)+SHAR-R.	1.26	**0.83**	1.82	1.18	**3.43**	**1.42**

We report IVRMSEs on a year-by-year basis. The RW approach fits the model daily and uses it as the next-day forecast. ANN(3)+RW refers to fitting a model daily, training a neural network on the implied errors, and constructing the next-day forecast as model fit plus error fit. ANN(3)+SHAR (or SHAR-R) refers to fitting a model daily, training a neural network on the implied errors, and estimating the SHAR (or SHAR-R) time-series model to produce the next-day forecast.

Table 11: IVRMSEs for the Two- and Three-Step Procedures (5-Day Horizon).

Model	2016	2017	2018	2019	2020	2021
			Model - AHBS			
RW	2.34	1.42	3.20	2.26	6.25	2.59
ANN(3)+RW	2.35	1.38	3.14	2.25	**6.21**	2.57
ANN(3)+SHAR	2.29	1.39	3.09	2.12	6.63	2.48
ANN(3)+SHAR-R.	**2.27**	**1.35**	**3.08**	**2.11**	6.48	**2.43**
			Model - SV(1)			
RW	2.30	1.26	3.09	2.15	6.12	2.48
ANN(3)+RW	2.28	1.23	3.05	2.16	**6.10**	2.50
ANN(3)+SHAR	2.22	1.22	2.98	2.03	6.52	2.41
ANN(3)+SHAR-R.	**2.20**	**1.20**	**2.97**	**2.01**	6.31	**2.34**
			Model - SV(2)			
RW	2.25	1.21	3.07	2.12	6.09	2.48
ANN(3)+RW	2.24	1.19	3.05	2.15	**6.08**	2.50
ANN(3)+SHAR	2.18	1.17	**2.98**	2.02	6.50	2.42
ANN(3)+SHAR-R.	**2.16**	**1.14**	2.99	**2.00**	6.45	**2.34**
			Model - SVJR			
RW	2.29	1.23	3.09	2.15	6.11	2.48
ANN(3)+RW	2.27	1.22	3.06	2.16	**6.10**	2.49
ANN(3)+SHAR	2.21	1.21	2.99	2.03	6.52	2.41
ANN(3)+SHAR-R.	**2.20**	**1.19**	**2.99**	**2.01**	6.41	**2.34**

We report IVRMSEs on a year-by-year basis. The RW approach fits the model daily and uses it as the 5-day ahead forecast. ANN(3)+RW refers to fitting a model daily, training a neural network on the implied errors, and constructing the 5-day ahead forecast as model fit plus error fit. ANN(3)+SHAR (or SHAR-R) refers to fitting a model daily, training a neural network on the implied errors, and estimating the SHAR (or SHAR-R) time-series model to produce the forecast.

of lags p is determined using the Bayesian Information Criterion (BIC). Consequently, their forecast of the IV surface at day t for horizon h is given by:

$$\widehat{IV}(O_{i,t+h}) = IV^M\big(O_{i,t+h}, \mathbb{E}(\hat{\Theta}_{t+h}|\hat{\mathbf{\Theta}}_{1:t})\big). \tag{16}$$

The SHAR approach offers two advantages compared to fitting a time series process on the model parameters. First, it is more versatile, as it is applicable across a wide range of IV surface models. Specifically, the Goncalves and Guidolin (2006) method requires predicting the model parameters, which is not feasible for highly parameterized 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 in those models 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 error, since $\mathbb{E}(IV^M\big(O_{i,t+h}, \hat{\Theta}_{t+h}\big)|\hat{\mathbf{\Theta}}_{1:t}) \neq IV^M\big(O_{i,t+h}, \mathbb{E}(\hat{\Theta}_{t+h}|\hat{\mathbf{\Theta}}_{1:t})\big)$. The first limitation makes the VAR method difficult to use for many IV models. However, it is of course suitable for the AHBS model because it is linear and not highly parameterized. We refer to this implementation as AHBS-VAR(p).

Table 12 presents IVRMSEs for the AHBS-VAR(p) approach for various forecast horizons. For ease of comparison, the table also repeats the RMSE performance of the RW and SHAR methods. The results across sample years and forecast horizons indicate that the AHBS-VAR(p) method is the best performer in five out of 18 instances. It performs particularly well in 2019, when the differences in IVRMSEs are substantial for the 5- and 20-day ahead horizons. The SHAR approach is the best performer in nine cases. In some years, notably 2016 and 2018, the difference in favor of the SHAR model is large, especially at the 5-day and 20-day horizons. Note that the SHAR approach for the more sophisticated surface fitting models substantially outperforms the AHBS model, see Tables 6 to 8.

The underlying intuition for these findings is that the use of a VAR to predict the model

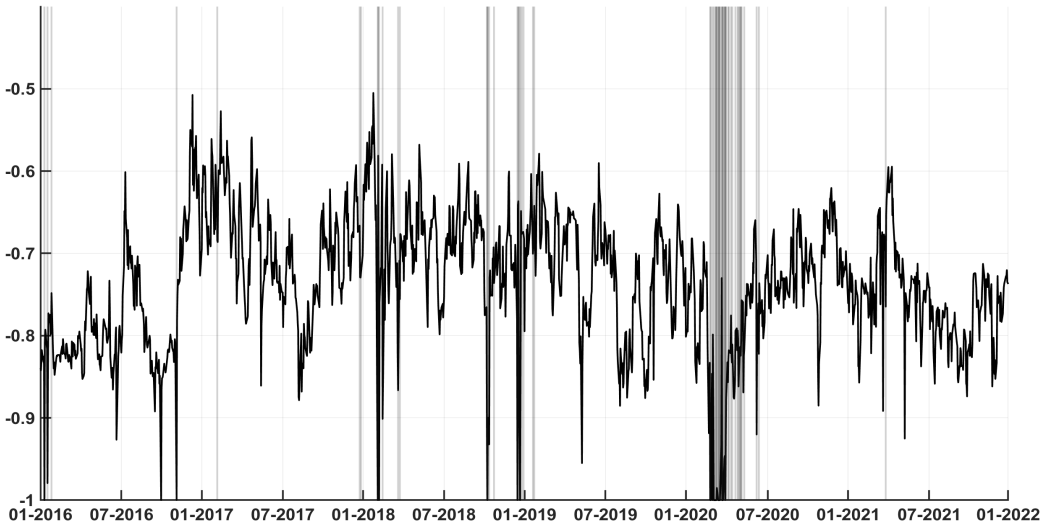
Table 12: IVRMSEs for the AHBS-VAR(p) Approach.

Model	2016	2017	2018	2019	2020	2021
			Horizon - 1			
RW	1.34	1.11	1.94	1.35	3.57	1.67
SHAR	**1.34**	**1.11**	**1.93**	1.34	3.58	**1.65**
VAR(p)	1.38	1.12	1.94	**1.33**	**3.54**	1.68
			Horizon - 5			
RW	2.34	**1.42**	3.20	2.26	**6.25**	2.59
SHAR	**2.29**	1.43	**3.14**	2.14	6.66	**2.52**
VAR(p)	2.52	1.56	3.15	**2.10**	7.17	2.53
			Horizon - 20			
RW	**3.25**	**1.69**	4.65	3.46	12.13	3.40
SHAR	3.28	1.87	**4.30**	3.06	**11.12**	3.24
VAR(p)	3.39	2.61	4.50	**2.80**	15.73	**3.03**

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

parameters is effective when the parameter estimates exhibit sufficient persistence. Figure 8 provides additional insight and illustrates some of the limitations of this approach. We plot daily estimates of the leverage correlation coefficient ρ in the SV(1) model. First, the parameter is confined within the bounded support $(-1, 1)$, which creates obvious pitfalls when forecasting with a VAR model. Second, daily estimation does not necessarily guarantee persistence over time. The parameter estimates change substantially in certain periods, especially when abnormal option surfaces occur or when limited information is available. For instance, the estimate for ρ is close to -1 on certain days, which induces a lack of persistence in the time series for this parameter.

Figure 8: Daily ρ Estimates in the SV(1) Model.



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

8 Conclusion

Because of the explosive growth in option trading volume, the study of option markets and option prices in particular is the subject of a large and growing body of research. Within this literature, the option IV surface has received a lot of attention, and it is modeled by a variety of approaches, including dynamic option pricing models, nonparametric methods and, more recently, machine learning techniques. While the ultimate objective of this literature is (implicitly) to generate forecasts of future IV surfaces, many studies focus on fitting the option surface, and the relative forecasting performance of different models and forecasting approaches remains relatively unexplored. This paper follows Goncalves and Guidolin (2006) and Almeida et al. (2023) in taking a more formal forecasting perspective and proposes a novel framework for forecasting the option IV surface.

The IV surfaces are impacted in real time by fluctuations in economic conditions. Moreover, the availability of option contracts with different moneyness and maturity of available option contracts is highly time-varying, and in recent years has trended upwards. Models of the IV surface therefore need to be fitted at high frequencies, at least on a daily basis. The fitted daily surfaces are typically used as forecasts for surfaces at future horizons, which amounts to a random walk assumption. While this is a very convenient and practical approach, we argue that it can be improved by including an additional step that exploits the time-series properties of the fitted (model-implied) daily IV surfaces. We show how to implement this model sequentially in closed form using ordinary least squares.

We test our approach using a HAR-type dynamic using daily data on the S&P 500 index IV surface for 2016-2021. We find strong evidence that the random walk forecast, which exclusively relies on the IV surface for the most recent day, typically yields inferior forecasts. Because the evolution of the daily IV surface often seems to be characterized by structural breaks, we introduce a robust version of our approach that mitigates the impact of such

abnormal surfaces on model estimates and subsequent surface forecasts. We show that this robust approach also provides further improvements.

Our proposed time-series approach is based on the HAR process, which is popular in the realized volatility literature. We believe that recent contributions in this literature may allow us to exploit the historical information in the IV surface even better, and therefore may result in improved forecasting performance. For example, our approach could potentially be extended using 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). Alternatively, forecasts can be further improved by considering refinements of the SHAR approach, as in Section 6.3. Another interesting extension is to investigate how IV forecasts change when including short and ultrashort option maturities, which have recently emerged as a very important component of the option market.

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.

Almeida, Caio, Gustavo Freire, and Rodrigo Hizmeri, 2024, 0DTE Asset Pricing, SSRN Working Paper.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015a, Parametric Inference and Dynamic State Recovery From Option Panels, *Econometrica* 83, 1081–1145.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015b, The Risk Premia Embedded in Index Options, *Journal of Financial Economics* 117, 558–584.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015c, The Risk Premia Embedded in Index Options, *Journal of Financial Economics* 117, 558–584.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2020, The Pricing of Tail Risk and the Equity Premium: Evidence From International Option Markets, *Journal of Business & Economic Statistics* 38, 662–678.

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 Renò, 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.

Bates, David S., 2012, U.S. Stock Market Crash Risk, 1926-2010, *Journal of Financial Economics* 105, 229–259.

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.

Chen, Ying, Maria Grith, and Hannah Lai, 2025, Neural Tangent Kernel in Implied Volatility Forecasting: A Nonlinear Functional Autoregression Approach Forthcoming in *Journal of Business & Economic Statistics*.

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.

Clark, Todd E, and Kenneth D West, 2007, Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138, 291–311.

Corsi, Fulvio, 2009, A Simple Approximate Long-memory Model of Realized Volatility, *Journal of Financial Econometrics* 7, 174–196.

Dim, Chukwuma, Bjorn Eraker, and Grigory Vilkov, 2024, 0DTEs: Trading, Gamma Risk and Volatility Propagation, Working Paper.

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. Härdle, 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, Álvaro 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.

SpiderRock, 2024, Platform Documentation - Live Vol Surfaces, https://docs.spiderrockconnect.com/docs/next/Documentation/PlatformFeatures/Analytics/LiveVolSurfaces/, Accessed: 2025-05-08.

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 RMSEs in IV (1-day horizon).

	All periods						Normal Periods					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
Model - BS												
RW	4.75	4.45	4.84	4.47	6.53	5.29	4.74	4.45	4.70	4.46	5.85	5.29
Model - AHBS												
RW	1.34	1.11	1.94	1.35	3.57	1.67	1.31	1.11	1.62	1.34	2.84	1.67
SHAR	**1.00**	1.00	**0.99**	**0.99**	1.00	0.99	**1.00**	1.00	**0.98**	**0.99**	1.02	0.99
SHAR-R.	1.01	**0.99**	1.00	1.00	**0.94**	**0.99**	1.01	**0.99**	0.99	1.00	**0.95**	**0.99**
Model - ANN(3)												
RW	1.25	0.83	1.85	1.21	3.51	1.50	1.22	0.83	1.57	1.20	3.03	1.50
SHAR	0.90	0.96	0.85	**0.82**	0.91	0.83	0.90	0.96	0.80	**0.82**	**0.91**	0.83
SHAR-R.	**0.90**	**0.95**	**0.85**	0.82	**0.90**	**0.83**	**0.90**	**0.95**	**0.80**	0.82	0.92	**0.83**
Model - OptionMetrics												
RW	1.16	0.76	1.76	1.13	3.44	1.43	1.15	0.76	1.51	1.13	2.97	1.43
SHAR	**1.00**	0.99	0.99	**0.99**	1.00	**0.99**	**1.00**	0.99	0.98	**0.99**	1.03	**0.99**
SHAR-R.	1.00	**0.99**	**0.98**	1.00	**0.96**	0.99	1.00	**0.99**	**0.96**	1.00	**0.98**	0.99
Model - Random Forest												
RW	1.18	0.78	1.77	1.15	3.49	1.44	1.17	0.78	1.52	1.15	3.00	1.44
SHAR	**0.89**	**0.97**	**0.87**	0.82	0.92	0.83	**0.89**	**0.97**	**0.83**	0.82	0.92	0.83
SHAR-R.	0.89	0.97	0.87	**0.82**	**0.90**	**0.83**	0.89	0.97	0.83	**0.82**	**0.91**	**0.83**
Option pricing model - SV(1)												
RW	1.28	0.91	1.89	1.19	3.60	1.48	1.24	0.91	1.40	1.18	2.28	1.49
SHAR	**0.99**	0.99	**0.98**	**0.99**	1.00	0.99	**0.99**	0.99	0.96	**1.00**	1.01	0.99
SHAR-R.	1.00	**0.98**	0.99	1.00	**0.96**	**0.98**	1.00	**0.98**	**0.93**	1.01	**0.98**	**0.98**
Option pricing model - SV(2)												
RW	1.21	0.84	1.82	1.16	3.55	1.48	1.17	0.83	1.35	1.14	1.98	1.48
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.99	0.97	**1.00**	1.02	0.99
SHAR-R.	1.01	**0.97**	1.00	1.00	**0.98**	**0.98**	1.01	**0.97**	**0.95**	1.01	**0.94**	**0.98**
Option pricing model - SV(3)												
RW	1.21	0.82	1.81	1.16	3.52	1.47	1.17	0.82	1.35	1.14	1.97	1.47
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.99	0.97	**1.00**	1.02	0.99
SHAR-R.	1.00	**0.98**	1.00	1.00	**0.98**	**0.98**	1.00	**0.98**	**0.96**	1.01	**0.94**	**0.98**
Model - SVJR												
RW	1.27	0.87	1.86	1.18	3.52	1.48	1.23	0.86	1.41	1.17	2.13	1.48
SHAR	**0.99**	0.99	**0.98**	**0.99**	1.00	0.98	0.99	0.99	0.95	**0.99**	1.02	0.98
SHAR-R.	0.99	**0.98**	1.00	1.00	**0.99**	**0.98**	**0.99**	**0.98**	**0.92**	1.00	**0.96**	**0.98**
Model - SVDE												
RW	1.24	0.85	1.83	1.17	3.51	1.46	1.21	0.85	1.38	1.15	2.34	1.46
SHAR	**0.99**	0.99	**0.99**	**0.99**	1.00	0.99	**0.99**	0.99	0.96	**1.00**	1.02	0.99
SHAR-R.	1.00	**0.98**	1.00	1.00	**0.93**	**0.98**	1.00	**0.98**	**0.96**	1.01	**0.98**	**0.98**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the one day ahead implied volatility surface. The option pricing models are trained using the vega loss function.

Table A.2: Out-of-sample RMSEs in IV (5-day horizon).

	All periods						Normal Periods					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	4.97	4.52	5.24	4.70	7.90	5.51	4.96	4.52	5.00	4.69	6.68	5.51
				Model - AHBS						*Model - AHBS*		
RW	2.34	1.42	3.20	2.26	**6.25**	2.59	2.34	1.42	2.78	2.26	**4.71**	2.59
SHAR	0.98	1.01	0.98	0.94	1.07	0.97	0.97	1.01	0.96	0.94	1.03	0.97
SHAR-R.	**0.97**	**0.98**	**0.98**	**0.94**	1.04	**0.95**	**0.97**	**0.98**	**0.95**	**0.94**	1.02	**0.95**
				Model - ANN(3)						*Model - ANN(3)*		
RW	2.32	1.40	3.15	2.21	6.16	2.50	2.32	1.40	2.84	2.21	5.03	2.49
SHAR	0.81	0.85	**0.83**	0.80	**0.93**	0.80	0.81	0.85	**0.78**	0.80	0.92	0.80
SHAR-R.	**0.81**	**0.85**	0.84	**0.79**	0.94	**0.79**	**0.81**	**0.85**	0.79	**0.79**	**0.91**	**0.79**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	2.22	1.18	3.07	2.12	**6.08**	2.46	2.22	1.18	2.84	2.12	**5.12**	2.46
SHAR	0.98	**0.98**	0.97	0.95	1.08	0.97	0.98	**0.98**	0.96	0.95	1.04	0.97
SHAR-R.	**0.98**	0.98	**0.97**	**0.95**	1.05	**0.95**	**0.98**	0.98	**0.95**	**0.95**	1.04	**0.95**
				Model - Random Forest						*Model - Random Forest*		
RW	2.25	1.19	3.07	2.13	6.12	2.47	2.25	1.19	2.85	2.13	5.24	2.47
SHAR	**0.83**	**0.95**	**0.84**	0.81	0.93	0.81	**0.83**	**0.95**	**0.81**	0.81	0.92	0.81
SHAR-R.	0.83	0.95	0.84	**0.81**	**0.92**	**0.80**	0.83	0.95	0.81	**0.81**	**0.89**	**0.80**
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	2.30	1.26	3.09	2.15	**6.12**	2.48	2.28	1.26	2.36	2.15	3.90	2.48
SHAR	0.97	1.00	0.97	0.95	1.07	0.97	0.97	1.00	0.93	0.95	1.00	0.97
SHAR-R.	**0.97**	**0.98**	**0.97**	**0.94**	1.04	**0.94**	**0.96**	**0.98**	**0.91**	**0.94**	**0.96**	**0.94**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	2.25	1.21	3.07	2.12	**6.09**	2.48	2.23	1.21	2.34	2.12	3.61	2.48
SHAR	0.97	0.99	**0.98**	0.95	1.07	0.97	0.97	0.99	0.92	0.95	0.98	0.97
SHAR-R.	**0.96**	**0.97**	0.98	**0.94**	1.06	**0.94**	**0.96**	**0.97**	**0.91**	**0.94**	**0.94**	**0.94**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	2.25	1.20	3.06	2.12	**6.09**	2.48	2.23	1.20	2.38	2.12	3.59	2.48
SHAR	0.98	0.99	0.98	0.95	1.07	0.97	0.97	0.99	0.92	0.95	0.99	0.97
SHAR-R.	**0.97**	**0.97**	**0.98**	**0.94**	1.06	**0.94**	**0.96**	**0.97**	**0.91**	**0.94**	**0.94**	**0.94**
				Model - SVJR						*Model - SVJR*		
RW	2.29	1.23	3.09	2.15	**6.11**	2.48	2.27	1.23	2.36	2.15	3.76	2.49
SHAR	0.97	1.00	0.97	0.95	1.07	0.97	0.97	1.00	0.93	0.95	0.98	0.97
SHAR-R.	**0.96**	**0.99**	**0.97**	**0.94**	1.05	**0.94**	**0.96**	**0.99**	**0.92**	**0.94**	**0.95**	**0.94**
				Model - SVDE						*Model - SVDE*		
RW	2.26	1.22	3.07	2.13	**6.08**	2.47	2.26	1.22	2.38	2.13	4.18	2.47
SHAR	0.98	0.99	0.97	0.95	1.07	0.97	0.97	0.99	0.93	0.95	1.01	0.97
SHAR-R.	**0.97**	**0.97**	**0.97**	**0.94**	1.03	**0.94**	**0.96**	**0.97**	**0.91**	**0.94**	**0.99**	**0.94**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the five day ahead implied volatility surface. The option pricing models are trained using the vega loss function.

Table A.3: Out-of-sample RMSEs in IV (20-day horizon).

	All periods						Normal Periods					
Model	2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020	2021
				Model - BS						*Model - BS*		
RW	5.35	4.59	5.96	5.16	12.11	5.80	5.33	4.59	5.63	5.16	9.69	5.80
				Model - AHBS						*Model - AHBS*		
RW	3.25	**1.69**	4.65	3.46	12.13	3.40	3.21	**1.69**	4.13	3.44	9.42	3.40
SHAR	1.01	1.11	**0.93**	0.88	**0.92**	0.95	1.01	1.11	**0.91**	0.89	**0.84**	0.95
SHAR-R.	**0.99**	1.07	0.93	**0.88**	0.94	**0.92**	**0.99**	1.07	0.92	**0.88**	0.85	**0.92**
				Model - ANN(3)						*Model - ANN(3)*		
RW	3.26	1.61	4.48	3.42	11.91	3.34	3.23	1.61	4.12	3.40	8.56	3.35
SHAR	0.87	0.89	**0.81**	0.77	**0.86**	0.80	0.86	0.89	**0.79**	0.77	0.82	0.80
SHAR-R.	**0.86**	**0.89**	0.82	**0.76**	0.86	**0.78**	**0.86**	**0.89**	0.79	**0.77**	**0.82**	**0.78**
				Model - OptionMetrics						*Model - OptionMetrics*		
RW	**3.16**	**1.48**	4.43	3.35	11.83	3.29	**3.15**	**1.48**	4.17	3.35	9.73	3.29
SHAR	1.01	1.08	0.93	0.89	**0.93**	0.94	1.01	1.08	0.91	0.89	**0.87**	0.94
SHAR-R.	1.01	1.08	**0.92**	**0.89**	0.94	**0.92**	1.01	1.08	**0.91**	**0.89**	0.88	**0.92**
				Model - Random Forest						*Model - Random Forest*		
RW	3.18	1.49	4.44	3.37	11.84	3.30	3.18	1.49	4.19	3.37	9.90	3.30
SHAR	0.86	0.96	**0.82**	0.77	0.87	0.81	0.86	0.96	**0.80**	0.77	0.81	0.81
SHAR-R.	**0.86**	**0.96**	0.82	**0.77**	**0.87**	**0.80**	**0.86**	**0.96**	0.80	**0.77**	**0.79**	**0.80**
				Option pricing model - SV(1)						*Option pricing model - SV(1)*		
RW	**3.20**	**1.55**	4.44	3.35	11.71	3.29	3.18	**1.55**	3.75	3.29	7.05	3.29
SHAR	1.01	1.11	**0.92**	0.90	**0.94**	0.94	1.01	1.11	**0.90**	0.91	**0.84**	0.94
SHAR-R.	1.00	1.08	0.93	**0.88**	0.95	**0.92**	**1.00**	1.08	0.90	**0.90**	0.86	**0.92**
				Option pricing model - SV(2)						*Option pricing model - SV(2)*		
RW	3.17	**1.50**	4.42	3.34	11.70	3.29	3.11	**1.51**	3.64	3.32	6.63	3.30
SHAR	1.01	1.10	**0.92**	0.90	**0.94**	0.94	1.01	1.10	**0.89**	0.90	**0.84**	0.94
SHAR-R.	**1.00**	1.06	0.93	**0.88**	0.95	**0.92**	**0.99**	1.06	0.89	**0.89**	0.86	**0.92**
				Option pricing model - SV(3)						*Option pricing model - SV(3)*		
RW	3.17	**1.50**	4.42	3.34	11.69	3.29	3.11	**1.50**	3.67	3.32	6.68	3.29
SHAR	1.01	1.10	**0.93**	0.90	**0.94**	0.94	1.01	1.10	**0.89**	0.90	**0.83**	0.94
SHAR-R.	**1.00**	1.07	0.93	**0.88**	0.95	**0.92**	**0.99**	1.07	0.90	**0.89**	0.85	**0.92**
				Model - SVJR						*Model - SVJR*		
RW	3.20	**1.52**	4.47	3.37	11.75	3.30	3.14	**1.52**	3.79	3.30	6.97	3.31
SHAR	1.01	1.11	**0.92**	0.89	**0.93**	0.94	1.00	1.11	0.90	0.91	**0.79**	0.94
SHAR-R.	**1.00**	1.09	0.92	**0.88**	0.95	**0.91**	**0.99**	1.09	**0.90**	**0.89**	0.82	**0.91**
				Model - SVDE						*Model - SVDE*		
RW	**3.17**	**1.51**	4.42	3.36	11.71	3.29	**3.13**	**1.51**	3.70	3.34	7.41	3.29
SHAR	1.01	1.10	**0.92**	0.90	**0.93**	0.94	1.01	1.10	**0.89**	0.90	**0.86**	0.94
SHAR-R.	1.00	1.08	0.93	**0.88**	0.95	**0.91**	1.00	1.08	0.89	**0.89**	0.88	**0.91**

For the SHAR and the SHAR-R models, we provide the RMSE ratio with respect to the RW approach. A value smaller than 1 indicates better performance than the RW model. Each day, we estimate the various models using the available options for that day, and we forecast the twenty day ahead implied volatility surface. The option pricing models are trained using the vega loss function.