Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Capturing Heterogeneity: Machine Learning Approaches to Implied Volatility Forecasting

Hyung Joo Kim and Dong Hwan Oh

2026-049

Please cite this paper as:
Kim, Hyung Joo, and Dong Hwan Oh (2026). "Capturing Heterogeneity: Machine Learning Approaches to Implied Volatility Forecasting," Finance and Economics Discussion Series 2026-049. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2026.049.

Capturing Heterogeneity: Machine Learning Approaches to Implied Volatility Forecasting*

Hyung Joo Kim† Dong Hwan Oh‡

June 24, 2026

Abstract

Despite documented heterogeneity in volatility dynamics across the option surface, standard implied volatility forecasting models apply homogeneous parameters throughout. We introduce a machine-learning framework that uses regression trees to partition the surface along both moneyness and maturity dimensions, identifying data-driven regions where distinct forecasting models perform best. Extending the Surface Heterogeneous Autoregressive (SHAR) framework of Dufays, Jacobs, and Rombouts (2025), we develop tree-based SHAR specifications that preserve interpretable structure while allowing model parameters to vary across the surface. Empirical analysis using S&P 500 options demonstrates that the boosted tree-based specification achieves the lowest out-of-sample forecast errors across all horizons, reducing one-month-ahead RMSE by 13% versus the benchmark SHAR model. The improvements are statistically significant and particularly pronounced during stress periods. The estimated tree presents economically interpretable segmentation: short-dated options exhibit higher daily persistence but lower monthly persistence than long-dated options, while deep out-of-the-money calls or puts display distinct dynamics from near-the-money contracts.

Keywords: implied volatility forecasting, option surface, machine learning, regression trees, ensemble methods, heterogeneous autoregressive models

JEL codes: C14, C22, C32, C51, C53, C58, G12

*We thank Andrew Patton for a detailed discussion of the paper. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.

†Federal Reserve Board; hyungjoo.kim@frb.gov

‡Federal Reserve Board; donghwan.oh@frb.gov

1 Introduction

Accurate forecasting of the implied volatility (IV) surface is crucial for practitioners, as it directly informs option pricing, risk management decisions, and trading strategies. Recent work by Dufays, Jacobs, and Rombouts (2025) proposes the Surface Heterogeneous Autoregressive (SHAR) model, which effectively and efficiently captures volatility surface dynamics by exploiting persistence over daily, weekly, and monthly horizons. However, SHAR and other existing models typically apply one set of forecasting model parameters uniformly across all strikes and maturities. This homogeneous specification may be restrictive given that option markets exhibit substantial heterogeneity: short-dated, long-dated, out-of-the-money (OTM), and near-the-money (NTM) options behave very differently in levels and in how they respond to news (see, e.g., Pan, 2002; Christoffersen, Heston, and Jacobs, 2009). Indeed, due to violations of Black and Scholes (1973) assumptions, implied volatilities vary systematically across strikes and maturities, forming the so-called implied volatility surface. In the time series, S&P 500 OTM put implied volatilities tend to be much higher and more volatile than at-the-money (ATM) or OTM call implied volatilities. Short-dated option implied volatilities are, on average, significantly more volatile and less persistent than long-dated option implied volatilities. When dynamics systematically vary across moneyness and maturity dimensions, imposing homogeneous model parameters across the entire surface may limit the potential for performance gains.

In this paper, we propose a machine learning approach to optimally partition the cross section of S&P 500 index options in implied volatility forecasting by extending the SHAR model. Specifically, we identify regions of the moneyness-maturity space where the SHAR model requires different parameter sets, acknowledging that distinct option segments exhibit distinct SHAR dynamics. We allow SHAR parameters to vary across the moneyness-maturity space, using regression trees to identify data-driven partitions where distinct forecasting models perform best. This approach substantially improves forecasting accuracy by adapting the model to regional heterogeneity across the surface and leveraging these localized dynamics.

Market segmentation and variation in investor clienteles across the option surface are well-documented in the literature. Studies have shown that option markets are segmented across strikes and maturities, with different investor clienteles focusing on specific regions of the option

surface (Bollen and Whaley, 2004; Gârleanu, Pedersen, and Poteshman, 2009; Bakshi, Madan, and Panayotov, 2010). These clienteles—including volatility sellers focusing on short-dated options, directional speculators trading out-of-the-money options, and institutional hedgers using near-the-money contracts—create distinct supply and demand patterns across the surface. Buraschi and Jiltsov (2006) show that considering different beliefs/information from heterogeneous agents improves explaining option market dynamics. Such documented market segmentation provides compelling motivation for our approach: allowing for localized dynamics substantially improves the performance of implied volatility forecasting across the surface.

Our empirical approach employs the partitioning logic of regression trees of Breiman, Friedman, Olshen, and Stone (1984, 2017) and lets the data decide where the option surface should be split, using option characteristics as splitting variables. Concretely, we use moneyness and maturity as splitting variables and build a tree that recursively partitions the option panel into subregions that are more homogeneous in terms of forecasting performance. At each terminal node, we thereby obtain a separate model estimated using only the options that fall into that region. This tree-based local estimation scheme preserves the structure and economic interpretation of the underlying models, while allowing their parameters to adapt flexibly across the surface. We particularly assess whether the tree structure can improve out-of-sample forecasts of the implied volatility surface.

We apply this tree-based partitioning approach to the SHAR framework for implied volatility surfaces, and we call this approach a tree-augmented Surface HAR model (Tree-SHAR).[1] Dufays, Jacobs, and Rombouts (2025) propose a two-step approach in which each daily surface is first fitted by a chosen option model or nonparametric method, and then the SHAR model linearly combines past fitted surfaces to forecast future implied volatilities on the unbalanced option panel, exploiting strong persistence over daily, weekly, and monthly horizons. We extend their setup by allowing the SHAR coefficients to vary across the tree's moneyness–maturity nodes. In the resulting Tree-SHAR model, each node has its own set of HAR-type parameters, so that short-dated, long-dated, OTM call, or OTM put options can have different implied volatility dynamics among themselves. We repeat the analysis by further adapting advanced machine learning approaches, a

[1] Although we choose the SHAR forecasting model, our tree-based methodology can be applied to other forecasting models to capture regional differences that are not directly implied by the models.

bootstrap-aggregated ensemble of tree models following Breiman (1996) and a boosted tree model following Friedman (2001).

Our empirical analysis using a large cross section of daily S&P 500 index option data shows that tree-based SHAR models deliver statistically significant improvements in out-of-sample forecast accuracy relative to the Random Walk and the baseline global SHAR model, while the Boosted Tree-SHAR model achieves the lowest forecast errors overall. The improvements are particularly pronounced in regions with greater cross-sectional heterogeneity: OTM calls, deep OTM puts, and short-maturity options. The estimated tree structures reveal systematic partitioning of the implied volatility surface by moneyness and maturity, capturing distinct dynamics across different option segments. For example, region-specific HAR coefficient estimates show that short-term option implied volatilities exhibit very high daily persistence, while long-term option implied volatilities display relatively higher weekly and monthly persistence.

The tree structure remains relatively stable under similar market conditions but adjusts substantially following market shocks. For example, (deep) OTM call IVs depend more on the weekly moving averages and are negatively associated with the monthly moving averages during bad times, whereas they depend more on the previous-day information during good times. These different dynamics of implied volatilities may be due to different market expectations toward upside potentials during good and bad periods. For long-term OTM puts, implied volatilities show sizable dependence on both past weekly and monthly information during good times, but the dependence on monthly information disappears during bad times. Moreover, the tree algorithm tends to divide the volatility surface into more heterogeneous dynamics in OTM calls and NTM options but more homogeneous dynamics in a wide range of short-term OTM puts during bad times. These types of variations are not captured by the global SHAR specification.

Our Tree-SHAR results convey the following messages: appropriately incorporating cross-sectional option characteristics—here, via tree partitions in moneyness and maturity—helps identify the segmentation of option markets and their heterogeneous features and systematically improves time-series implied-volatility forecasting. Our approach is deliberately conservative: it builds on a simple but well-understood benchmark model, preserves their economic structure, and uses the tree only as a data-driven device to choose where separate local models should be estimated. At the same time, the method is flexible and modular; it can be combined with alternative surface fitters,

forecasting models, or other splitting variables, and it can be implemented in real time using only modest computational resources.

A large body of empirical literature documents heterogeneity in option markets. Dumas, Fleming, and Whaley (1998), Constantinides, Jackwerth, and Perrakis (2009), and Fan and Mancini (2009) show that standard option pricing models produce unstable parameters over time and systematic pricing errors that vary with moneyness and maturity, underscoring the limitations of a single functional form. Additional examples of inferring heterogeneous implied model parameters include Bakshi, Cao, and Chen (1997), Hurn, Lindsay, and McClelland (2015), and Dufays, Jacobs, Liu, and Rombouts (2023). Our paper differs from these studies in that we focus on forecasting the time series evolution of the volatility surface rather than fitting option prices.

Our paper is most closely related to the literature on forecasting implied volatility surfaces, particularly Dufays, Jacobs, and Rombouts (2025), who propose the SHAR forecasting model. Recent works by Almeida, Fan, Freire, and Tang (2023) and Chen, Grith, and Lai (2025) apply neural networks to forecast implied volatility surfaces. The key distinction of our approach is that we allow forecasting model parameters to vary across the moneyness-maturity dimension through data-driven surface partitions. This not only improves forecasting accuracy compared to benchmark models but also reveals economically meaningful surface segmentation. In contrast to purely machine learning methods, our approach leverages machine learning to capture local heterogeneity while preserving the structural tractability of standard models.

Methodologically, our framework aligns with the literature on "tree-structured" parametric models. Audrino and Bühlmann (2001) pioneered the use of regression trees to estimate local GARCH parameters, and Patton and Simsek (2026) extend this approach to GAS models. Medeiros, Vasconcelos, Veiga, and Zilberman (2021) and Goulet Coulombe (2024) demonstrate that tree-based methods successfully identify regime-dependent dynamics in macroeconomic forecasting. While these studies primarily exploit tree-based partitioning in the time-series dimension, Brownlees and Souza (2025) and Oh and Patton (2026) apply similar methods to the cross-section of stocks and mutual funds, respectively. We extend this logic to the cross section of options, using trees to partition the moneyness-maturity space and estimate local SHAR parameters. To the best of our knowledge, this is the first application of tree-structured parameter heterogeneity to option-implied

volatilities. Our work also relates to the broader "local estimation" literature, which allows model parameters to vary with state variables. This literature has developed along two paths: kernel-based localization for (quasi-) maximum likelihood estimation (Tibshirani and Hastie, 1987; Fan, Farmen, and Gijbels, 1998; Fan, Wu, and Feng, 2009; Oh and Patton, 2024; Kim and Oh, 2025), and tree-based localization for regression, parametric distributions, and GMM models (Breiman, 2001; Athey, Tibshirani, and Wager, 2019; Schlosser, Hothorn, Stauffer, and Zeileis, 2019).

The remainder of the paper is organized as follows. Section 2 reviews the SHAR framework for dynamic implied volatility surfaces and outlines the Tree-SHAR extension, together with its bagged and boosted enhancements, emphasizing improvements in time-series forecasts. Section 3 describes the option data and filtering rules. Section 4 presents the empirical results for forecast performance, including out-of-sample forecasts of implied volatility for benchmarks, Tree-SHAR and its extensions. Section 5 concludes.

2 Tree-Based Surface HAR Models

This section develops tree-based extensions of the Surface HAR (SHAR) model for implied-volatility (IV) surface forecasting. Section 2.1 introduces the baseline SHAR model, which extends the HAR structure of Corsi (2009) to the option panel by modeling IV as a function of lagged fitted-surface components. This specification provides an interpretable decomposition of IV dynamics into short-, medium-, and long-run components. Section 2.2 provides empirical evidence of divergent IV dynamics across the moneyness and maturity dimensions. This empirical observation motivates considering heterogeneous dynamics over the volatility surface for forecasting. Section 2.3 generalizes the baseline model by allowing the HAR dynamics to vary across the moneyness–maturity space. Specifically, we partition the implied-volatility surface using a regression tree and estimate leaf-specific HAR coefficients, thereby capturing cross-sectional heterogeneity while preserving the HAR interpretation within each region. Section 2.4 introduces a bagged Tree-SHAR model that improves stability by averaging forecasts across bootstrap trees. Finally, Section 2.5 develops a boosted Tree-SHAR model that constructs an additive expansion of local HAR corrections, sequentially reducing remaining forecast errors and capturing finer nonlinearities and interactions over the option surface.

2.1 Baseline Surface HAR model

Our benchmark time-series model for the implied volatility surface is the Surface Heterogeneous Autoregressive (SHAR) framework of Dufays, Jacobs, and Rombouts (2025). Their forecasting setup is explicitly two-step.

Step 1: daily cross-sectional surface fit. From an initial day onward, we observe a daily stream of implied volatility (IV) surfaces. On each day ℓ, we observe N_ℓ option-implied volatilities with option i's characteristics $\mathbf{z}_{i,\ell} = (m_{i,\ell}, \tau_{i,\ell})'$, where $m_{i,\ell}$ denotes moneyness and $\tau_{i,\ell}$ denotes time-to-maturity. Let $\mathrm{IV}(\mathbf{z}_{i,\ell})$ denote the Black–Scholes implied volatility of option i on day ℓ.

In the first step, each day's IV surface is fitted cross-sectionally using an arbitrary model M for implied volatilities:

$$\mathrm{IV}(\mathbf{z}_{i,\ell}) = \mathrm{IV}^M\big(\mathbf{z}_{i,\ell}; \Theta_\ell\big) + \varepsilon_{i,\ell}, \qquad i = 1, \ldots, N_\ell, \quad \ell = 1, \ldots, T, \tag{1}$$

where $\mathrm{IV}^M\big(\cdot; \Theta_\ell\big)$ is the model-implied volatility surface generated by parameter vector Θ_ℓ (estimated on day ℓ), and $\varepsilon_{i,\ell}$ is an error. The model M can be a structural option pricing model (for example, Heston's (1993) stochastic volatility model), a parametric IV surface, or a flexible nonparametric or machine-learning method. Equation (1) therefore delivers, for each day ℓ, a fitted surface $\mathrm{IV}^M\big(\mathbf{z}_{i,\ell}; \widehat{\Theta}_\ell\big)$ evaluated on the unbalanced set of observed $\mathbf{z}_{i,\ell} = (m_{i,\ell}, \tau_{i,\ell})'$ points. In this paper, we employ the ad-hoc Black-Scholes (AHBS) specification, which provides an easy-to-implement but sufficiently flexible approach used by Dumas, Fleming, and Whaley (1998) and Dufays, Jacobs, and Rombouts (2025), among others. The AHBS model approximates the IV surface using a second-degree polynomial in moneyness and maturity:

$$\mathrm{IV}(\mathbf{z}_{i,\ell}) = \beta_0^{\mathrm{M}} + \beta_1^{\mathrm{M}} m_{i,\ell} + \beta_2^{\mathrm{M}} m_{i,\ell}^2 + \beta_3^{\mathrm{M}} \tau_{i,\ell} + \beta_4^{\mathrm{M}} \tau_{i,\ell}^2 + \beta_5^{\mathrm{M}} m_{i,\ell} \tau_{i,\ell} + \varepsilon_{i,\ell}.$$

A widely used benchmark forecast treats the fitted surface as a random walk in shape: on day t, estimate $\widehat{\Theta}_t$ using only the day-t cross-section, and then forecast the IV surface at day $t+h$ by evaluating the *same* fitted model at the future option characteristics:

$$\widehat{\mathrm{IV}}^{\mathrm{RW}}_{t+h\,|\,t}(\mathbf{z}_{i,t+h}) = \mathrm{IV}^M\big(\mathbf{z}_{i,t+h}; \widehat{\Theta}_t\big), \qquad i = 1, \ldots, N_{t+h}. \tag{2}$$

That is, $\mathrm{IV}^M(\mathbf{z}_{i,t+h}; \widehat{\Theta}_t)$ denotes the model-implied volatility obtained by evaluating the day-t fitted volatility surface, parameterized by $\widehat{\Theta}_t$, at the target moneyness and time-to-maturity point $\mathbf{z}_{i,t+h} = (m_{i,t+h}, \tau_{i,t+h})'$ for date $t+h$. The random-walk benchmark therefore assumes that the shape of the fitted volatility surface remains unchanged between dates t and $t+h$. This "RW" model is computationally convenient because it only requires re-estimating the cross-sectional surface model on a single day's option panel.

As an alternative approach, we also consider generating the volatility surface using the parsimonious Stochastic Volatility Inspired (SVI) method of Gatheral (2004). This method uses a different functional form of IV than the AHBS method to span the implied volatility curve given maturity. We specifically follow Beason and Schreindorfer (2022) to extend it to the volatility surface, *i.e.*, as a function of both moneyness and maturity. We discuss more details of the method in Appendix A. Overall, we obtain consistent forecasting performance regardless of whether we generate the surface following the AHBS or the SVI method.

The surface-fitting procedure in Step 1 is common to all forecasting specifications considered in this paper. The baseline SHAR model, Tree-SHAR, Bagged Tree-SHAR, and Boosted Tree-SHAR differ only in Step 2, which maps the history of fitted IV surfaces into forecasts of future implied volatility. Thus, the improvements from the tree-based methods come from modifying the second-step forecasting rule—in particular, by allowing the SHAR coefficients to vary across the moneyness–maturity surface—while holding the first-step surface construction fixed. We now describe this second step.

Step 2: dynamic modeling of fitted surfaces. The SHAR approach replaces the RW benchmark (2) with an autoregressive forecasting rule that exploits persistence in the sequence of fitted IV surfaces. In our empirical implementation, we follow Dufays, Jacobs, and Rombouts (2025) and adopt a heterogeneous autoregressive (HAR) specification in the spirit of Corsi (2009). For a given forecast horizon h, define the HAR regressor vector

$$\mathbf{x}_{h,i,\ell} = \left(1,\ \mathrm{IV}^M\big(\mathbf{z}_{i,\ell}; \widehat{\Theta}_{\ell-h}\big),\ \frac{1}{5}\sum_{r=0}^{4} \mathrm{IV}^M\big(\mathbf{z}_{i,\ell}; \widehat{\Theta}_{\ell-h-r}\big),\ \frac{1}{22}\sum_{r=0}^{21} \mathrm{IV}^M\big(\mathbf{z}_{i,\ell}; \widehat{\Theta}_{\ell-h-r}\big)\right)'. \tag{3}$$

The first element is an intercept, while the remaining three elements are the fitted IV surface from the previous day, the average of fitted IV surfaces over the previous week, and the average of fitted IV surfaces over the previous month, all evaluated at the same option characteristics $\mathbf{z}_{i,\ell}$.

Using this notation, the baseline SHAR model is

$$\mathrm{IV}(\mathbf{z}_{i,\ell}) = \mathbf{x}'_{h,i,\ell}\boldsymbol{\beta}_h + \varepsilon_{h,i,\ell}, \qquad i = 1, \ldots, N_\ell, \quad \ell = h + 22, \ldots, T, \tag{4}$$

where $\boldsymbol{\beta}_h = (\beta_h^{(0)}, \beta_h^{(d)}, \beta_h^{(w)}, \beta_h^{(m)})'$ depends on the forecast horizon h but is constant across options and days.

To estimate the SHAR coefficients at horizon h using information available at date t, we pool observations across days $\ell = h + 22, \ldots, t$ and across all options within each day. The pooled ordinary least squares estimator is

$$\widehat{\boldsymbol{\beta}}_{h|t} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^4} \sum_{\ell=h+22}^{t} \sum_{i=1}^{N_\ell} \left(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\boldsymbol{\beta}\right)^2. \tag{5}$$

Given $\widehat{\boldsymbol{\beta}}_{h|t}$ and the sequence of fitted surfaces up to date t, the h-step-ahead SHAR forecast for an option with characteristics $\mathbf{z}_{i,t+h}$ is[2]

$$\widehat{\mathrm{IV}}^{\mathrm{SHAR}}_{t+h\,|\,t}(\mathbf{z}_{i,t+h}) = \mathbf{x}'_{h,i,t+h}\widehat{\boldsymbol{\beta}}_{h|t}, \tag{6}$$

where

$$\mathbf{x}_{h,i,t+h} = \left(1,\ \mathrm{IV}^M\big(\mathbf{z}_{i,t+h}; \widehat{\Theta}_t\big),\ \frac{1}{5}\sum_{r=0}^{4} \mathrm{IV}^M\big(\mathbf{z}_{i,t+h}; \widehat{\Theta}_{t-r}\big),\ \frac{1}{22}\sum_{r=0}^{21} \mathrm{IV}^M\big(\mathbf{z}_{i,t+h}; \widehat{\Theta}_{t-r}\big)\right)'.$$

Thus, the SHAR forecast is a linear combination of the daily, weekly, and monthly fitted-surface components evaluated at the future option characteristics $\mathbf{z}_{i,t+h}$.

For later use in the tree-based extensions, it is convenient to collect the observations used in the second-step SHAR estimation into a single notation. Fix the forecast horizon h. At forecast origin

[2]The option characteristics and implied volatility at date $t + h$ are not observed at forecast origin t. We interpret the forecast as conditional on a target moneyness–maturity point $\mathbf{z}_{i,t+h} = (m_{i,t+h}, \tau_{i,t+h})'$. Given this point, $\mathbf{x}_{h,i,t+h}$ is computable from fitted IV surfaces estimated using information available up to date t. Hence, the only unknown quantity in the forecasting equation is the future implied volatility $\mathrm{IV}(\mathbf{z}_{i,t+h})$, which is the forecast target.

t, define the available SHAR sample as

$$\mathcal{D}_t = \{(\mathrm{IV}(\mathbf{z}_{i,\ell}), \mathbf{z}_{i,\ell}, \mathbf{x}_{h,i,\ell}) : i = 1, \ldots, N_\ell,\ \ell = h + 22, \ldots, t\}\,. \tag{7}$$

Each element of $\mathcal{D}_t$ contains the implied volatility $\mathrm{IV}(\mathbf{z}_{i,\ell})$, the option characteristics $\mathbf{z}_{i,\ell} = (m_{i,\ell}, \tau_{i,\ell})'$, and the SHAR regressor vector $\mathbf{x}_{h,i,\ell}$ defined in (3). This notation will be used below when constructing Tree-SHAR, Bagged Tree-SHAR, and Boosted Tree-SHAR.

In our empirical analysis, this SHAR specification serves as the baseline time-series model for the implied volatility surface, alongside the RW benchmark in (2). The tree-augmented version introduced in Section 2.3 retains the SHAR structure in (4)–(6) but allows the coefficients to differ across regions of the moneyness–maturity space.

2.2 Heterogeneous characteristics in data

Before we describe details of the tree-augmented SHAR model, we briefly discuss heterogeneous characteristics of the implied volatility surface observed in data.

Figure 1 shows average implied volatilities by moneyness or maturity groups over time. We use three groups where the moneyness or maturity is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile, on each date. Panel A shows the IV by moneyness, and Panel B shows the IV by maturity. In Panel A, the blue solid line intrinsically represents the average IV for OTM calls, and the black dotted line represents the average IV for deep OTM puts. We observe that the average IV increases and is more volatile as moneyness increases (or moves toward deeper OTM puts). In Panel B, the level of IVs across maturity dimension remains similar, but the volatility of IVs differs significantly. Shorter-maturity IVs (blue solid line) show a much more volatile pattern than relatively longer-maturity IVs (red dashed and black dotted lines).

Figure 2 depicts autocorrelation functions (ACF) of implied volatilities and their first differences up to 22 lags. Panels A and B plot the ACF of IVs by moneyness and maturity groups, where the groups are identically defined as in Figure 1. Although the ACFs of IVs look similar across different moneyness levels, they increase as maturity increases. This pattern is more pronounced as the number of lags increases. On the other hand, daily changes in IVs in Panels C and D show a

different pattern. Although their ACFs appear similar across different maturity levels, the ACFs vary across different moneyness levels. Specifically, OTM call IVs (blue bars) show relatively more negative first-order and positive second-order autocorrelation coefficients, respectively, implying that they are more likely to depend on lagged IVs.[3]

All of the above observations clearly indicate that the time series dynamics of IVs are heterogeneous across different moneyness and maturity. We now discuss how we incorporate such heterogeneity of the volatility surface into the forecasting framework through a regression tree.

2.3 Tree-augmented SHAR model

We next extend the benchmark SHAR specification in Section 2.1 by allowing the HAR dynamics to vary across regions of the moneyness–maturity space. The key idea is as follows: we use a regression tree $\mathcal{T}$ to partition the option observations into terminal nodes (leaves), and we estimate a separate SHAR model within each leaf. This yields a localized forecasting rule that can differ across economically meaningful regions of the option surface, such as short-dated, long-dated, OTM, and near-the-money options.

Let $m_{i,\ell}$ and $\tau_{i,\ell}$ denote the moneyness (underlying price divided by strike price) and time to maturity of option i on day ℓ. We collect these option characteristics in[4]

$$\mathbf{z}_{i,\ell} \equiv (z_{i,\ell}^{(1)}, z_{i,\ell}^{(2)})' \;=\; (m_{i,\ell}, \tau_{i,\ell})' \in \mathcal{Z}.$$

A tree $\mathcal{T}$ on $\mathcal{Z}$ induces a recursive partition

$$\mathcal{Z} \;=\; R_1 \cup \cdots \cup R_{J(\mathcal{T})},$$

[3]For example, consider a variable following an AR(2) process. The higher the coefficient of the 2nd lagged term, the bigger the negative first-order ACF of its first difference.

[4]Our tree-based method can accommodate many potential splitting variables, including other option characteristics (e.g., liquidity, option Greeks) and time-series state variables (e.g., VIX, past returns). However, we focus exclusively on moneyness and maturity because our primary objective is to characterize heterogeneity in forecasting dynamics across the option surface itself. Extending the framework to include additional option characteristics or time-varying state variables represents interesting directions for future research but would shift the focus away from identifying cross-sectional surface segmentation.

where terminal nodes (leaves) R_j are mutually exclusive, and $J(\mathcal{T})$ denotes the number of leaves. Each region R_j corresponds to a terminal node (leaf) and is characterized by a sequence of binary splits of the form $m \leq c$ or $\tau \leq c$ applied to the coordinates of $\mathbf{z}_{i,\ell}$.

Using the HAR regressor vector $\mathbf{x}_{h,i,\ell}$ defined in (3), the Tree-SHAR model replaces the global coefficient vector in (4) with a leaf-specific coefficient vector:

$$\mathrm{IV}(\mathbf{z}_{i,\ell}) \;=\; \mathbf{x}'_{h,i,\ell}\,\boldsymbol{\beta}_{h,j(i,\ell)} \;+\; \varepsilon_{h,i,\ell}, \tag{8}$$

where $\boldsymbol{\beta}_{h,j}$ collects the intercept and HAR coefficients for leaf j, and $j(i,\ell)$ denotes the index of the terminal node (leaf) to which option (i,ℓ) is assigned under the tree partition, i.e., the unique j such that $\mathbf{z}_{i,\ell} \in R_{j(i,\ell)}$.

For a fixed tree $\mathcal{T}$, the leaf-specific coefficients are estimated by ordinary least squares within each leaf. That is, for each leaf j,

$$\widehat{\boldsymbol{\beta}}_{h,j}(\mathcal{T}) \;=\; \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^4} \sum_{(i,\ell):\,\mathbf{z}_{i,\ell}\in R_j} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\boldsymbol{\beta}\big)^2, \tag{9}$$

and the Tree-SHAR objective is the sum of these squared residuals across leaves.

Directly solving equation (9) over all possible trees is infeasible, so we construct the tree greedily in the spirit of Classification and Regression Trees (CART; Breiman, Friedman, Olshen, and Stone, 1984, 2017). We start from the stump tree with a single leaf, denoted by $R_1^{(0)}$, which contains all option-day observations in the sample $\mathcal{D}_t$ defined in (7). The coefficient vector associated with this initial leaf coincides with the SHAR coefficient vector from Section 2.1. At a generic splitting step $s \geq 1$, suppose we have a tree $\mathcal{T}^{(s-1)}$ with $J^{(s-1)}$ leaves and associated parameter vectors $\{\widehat{\boldsymbol{\beta}}_{h,j}^{(s-1)}\}_{j=1}^{J^{(s-1)}}$.[5] For each current leaf $R_j^{(s-1)}$ and each splitting variable $k \in \{1,2\}$ (moneyness or maturity), we consider splitting $R_j^{(s-1)}$ into left and right child nodes at a candidate threshold c taken from a grid of $z_{i,\ell}^{(k)}$ within that leaf.[6] For each candidate split (j,k,c) we:

[5]The stump tree is denoted by $\mathcal{T}^{(0)}$ and contains a single leaf. Thus, $J^{(0)} = 1$, and the unique leaf $R_1^{(0)}$ contains all observations in $\mathcal{D}_t$. The coefficient vector associated with this leaf coincides with the SHAR estimator, $\widehat{\boldsymbol{\beta}}_{h,1}^{(0)} = \widehat{\boldsymbol{\beta}}_{h|t}$, obtained from the pooled OLS estimator in (5) in Section 2.1. Since each subsequent step adds exactly one split, the number of terminal leaves increases by one at each step; hence, before step s, the tree $\mathcal{T}^{(s-1)}$ has $J^{(s-1)} = s$ leaves.

[6]For each splitting variable $z_{i,\ell}^{(k)}$, $k = 1, 2$, candidate thresholds c are chosen from the 5%, 10%, ..., 95% sample quantiles of $z_{i,\ell}^{(k)}$ within the current leaf $R_j^{(s-1)}$. This produces 19 candidate split points for each splitting variable.

1. Partition the observations in $R_j^{(s-1)}$ into left and right subsets according to $z_{i,\ell}^{(k)} \leq c$ or $> c$, discarding the split if either child contains fewer than a pre-specified minimum number of options:

$$R_{j,k,c,L}^{(s)} = \{\, \mathbf{z}_{i,\ell} \in R_j^{(s-1)} : z_{i,\ell}^{(k)} \leq c \,\},$$
$$R_{j,k,c,R}^{(s)} = \{\, \mathbf{z}_{i,\ell} \in R_j^{(s-1)} : z_{i,\ell}^{(k)} > c \,\}.$$

2. Re-estimate the SHAR parameters separately in the left and right child nodes by minimizing the leaf-specific sum of squared errors (SSE) within each child node, while leaving the parameter estimates in all other leaves unchanged at their previous values $\widehat{\boldsymbol{\beta}}_{h,j'}^{(s-1)}$:

$$\widehat{\boldsymbol{\beta}}_{h,j,k,c,L}^{(s)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^4} \sum_{(i,\ell):\, \mathbf{z}_{i,\ell}\in R_{j,k,c,L}^{(s)}} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\boldsymbol{\beta}\big)^2,$$
$$\widehat{\boldsymbol{\beta}}_{h,j,k,c,R}^{(s)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^4} \sum_{(i,\ell):\, \mathbf{z}_{i,\ell}\in R_{j,k,c,R}^{(s)}} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\boldsymbol{\beta}\big)^2.$$

3. Compute the resulting total loss $\mathcal{L}(\mathcal{T}_{j,k,c}^{(s)})$ implied by the candidate tree $\mathcal{T}_{j,k,c}^{(s)}$, obtained by replacing $R_j^{(s-1)}$ with its two children and updating the corresponding parameter vectors:

$$\mathcal{L}(\mathcal{T}_{j,k,c}^{(s)}) = \sum_{(i,\ell):\, \mathbf{z}_{i,\ell}\in R_{j,k,c,L}^{(s)}} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\widehat{\boldsymbol{\beta}}_{h,j,k,c,L}^{(s)}\big)^2 + \sum_{(i,\ell):\, \mathbf{z}_{i,\ell}\in R_{j,k,c,R}^{(s)}} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\widehat{\boldsymbol{\beta}}_{h,j,k,c,R}^{(s)}\big)^2$$
$$+ \sum_{j'\neq j} \sum_{(i,\ell):\, \mathbf{z}_{i,\ell}\in R_{j'}^{(s-1)}} \big(\mathrm{IV}(\mathbf{z}_{i,\ell}) - \mathbf{x}'_{h,i,\ell}\widehat{\boldsymbol{\beta}}_{h,j'}^{(s-1)}\big)^2.$$

Among all candidate splits (j,k,c) we select the one that produces the largest reduction in total loss, i.e.

$$(j_s, k_s, c_s) \;=\; \arg\min_{(j,k,c)} \mathcal{L}\big(\mathcal{T}_{j,k,c}^{(s)}\big),$$

and set $\mathcal{T}^{(s)} = \mathcal{T}_{j_s,k_s,c_s}^{(s)}$, where the two child leaves use $\widehat{\boldsymbol{\beta}}_{h,j_s,k_s,c_s,L}^{(s)}$ and $\widehat{\boldsymbol{\beta}}_{h,j_s,k_s,c_s,R}^{(s)}$, and all other leaves keep $\widehat{\boldsymbol{\beta}}_{h,j'}^{(s-1)}$. This procedure is repeated until the number of leaves reaches a pre-specified optimal tree size.

To select the optimal tree size, denoted by J_h^*, we use the first half of the full sample and further split it into training and validation subsamples. We consider candidate trees with at most 15 terminal leaves and impose a minimum leaf size of 10,000 option observations. Among these candidates, we choose $J_h^* \in \{2, \ldots, 15\}$ using the CART cost-complexity criterion based on IV forecasting errors in the validation subsample, as described in Appendix B.

Given the selected tree $\widehat{\mathcal{T}}$ and coefficients $\{\widehat{\boldsymbol{\beta}}_{h,j}(\widehat{\mathcal{T}})\}$, the h-step-ahead Tree-SHAR IV forecast for any option with characteristics $\mathbf{z}_{i,t+h}$ is obtained by assigning it to its leaf under $\widehat{\mathcal{T}}$ and applying the corresponding local HAR coefficients,

$$\widehat{\mathrm{IV}}^{\mathrm{Tree}}_{t+h\,|\,t}(\mathbf{z}_{i,t+h}) \;=\; \mathbf{x}'_{h,i,t+h}\,\widehat{\boldsymbol{\beta}}_{h,j(i,t+h)}(\widehat{\mathcal{T}}), \tag{10}$$

which is the direct analogue of the SHAR forecast in (6), but with coefficients that adapt across the moneyness–maturity surface.

2.4 Bagged Tree-SHAR model

The Tree-SHAR model in Section 2.3 introduces cross-sectional heterogeneity by estimating a single tree partition and a set of leaf-specific SHAR coefficients. While this approach improves forecasting performance and remains interpretable, a single estimated tree can be sensitive to sampling variation in the data. To reduce this estimation noise, we construct a bootstrap-aggregated ("bagged") ensemble of Tree-SHAR models following Breiman (1996). The procedure is closely related to random forests (Breiman, 2001), with one important distinction: we do not randomly subsample the set of splitting variables at each split. Instead, every tree uses the same two splitting variables, moneyness and maturity.

We fix the target number of terminal leaves, denoted by J_h^*, at the value selected for Tree-SHAR using the CART cost-complexity criterion described in Appendix B. We then construct an ensemble of $B = 100$ trees as follows.

Step 1: Bootstrap resampling. For each $b = 1, \ldots, B$, we draw a bootstrap sample $\mathcal{D}_t^{[b]}$ from $\mathcal{D}_t$ in (7) by sampling observations independently with replacement.[7] Each bootstrap observation

[7]We use the standard nonparametric bootstrap rather than a block bootstrap because our objective is to account for sampling variability in tree estimation rather than serial dependence in the data.

therefore retains the full triplet consisting of the observed implied volatility, its associated option characteristics, and its SHAR regressor vector.

Step 2: Estimate a fixed-size Tree-SHAR on each bootstrap sample. For each bootstrap sample $\mathcal{D}_t^{[b]}$, we estimate a fixed-size tree $\mathcal{T}^{[b]}$ with J_h^* terminal leaves and the corresponding leaf-specific SHAR coefficients $\{\widehat{\boldsymbol{\beta}}_{h,j}^{[b]}\}_{j=1}^{J_h^*}$ using the same greedy splitting algorithm as in Section 2.3. For all bootstrap trees, we restrict the splitting variables to moneyness and maturity and continue splitting until the tree reaches J_h^* leaves, subject to the same minimum leaf-size constraint imposed in Tree-SHAR: each terminal leaf must contain at least 10,000 option observations. When evaluating a candidate split, we re-estimate the SHAR coefficients only in the two candidate child nodes, while holding the coefficient estimates in all unaffected leaves fixed at their values in the current tree before the split.

Step 3: Bagged forecasts by averaging across trees. To forecast the implied volatility at date $t+h$ for option i, conditional on its target moneyness–maturity point $\mathbf{z}_{i,t+h}$, each bootstrap tree $\mathcal{T}^{[b]}$ assigns the option to a terminal leaf $j^{[b]}(i,t+h)$. Given this leaf assignment, the corresponding leaf-specific HAR coefficients yield the h-step-ahead forecast

$$\widehat{\mathrm{IV}}_{t+h\,|\,t}^{[b]}(\mathbf{z}_{i,t+h}) = \mathbf{x}'_{h,i,t+h}\,\widehat{\boldsymbol{\beta}}_{h,j^{[b]}(i,t+h)}^{[b]}.$$

The Bagged Tree-SHAR forecast is obtained by averaging these tree-specific forecasts:

$$\widehat{\mathrm{IV}}_{t+h\,|\,t}^{\mathrm{Bag}}(\mathbf{z}_{i,t+h}) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mathrm{IV}}_{t+h\,|\,t}^{[b]}(\mathbf{z}_{i,t+h}). \tag{11}$$

This aggregation preserves the interpretability of Tree-SHAR: each component tree partitions the moneyness–maturity surface and estimates a local HAR specification within each terminal leaf. At the same time, averaging across bootstrap trees reduces the sensitivity of the forecast to any single sample or tree partition.

2.5 Boosted Tree-SHAR model

This subsection introduces a boosted version of Tree-SHAR. The construction follows the gradient boosting framework of Friedman (2001) (see also Hastie, Tibshirani, and Friedman, 2009), which builds an additive predictor by repeatedly fitting a tree-based model to the current fitting error and updating the forecast in a shrunken step in that direction. In our setting, each boosting step fits a local SHAR regression to the remaining errors, allowing the HAR coefficients to be refined sequentially across the moneyness–maturity surface.

Boosting is useful in this setting because a single Tree-SHAR model must capture all heterogeneity using one partition of the option surface. The IV surface, however, may contain several distinct nonlinear patterns across moneyness and maturity. Rather than relying on one partition to capture all such patterns, the Boosted Tree-SHAR model adds trees sequentially, with each new tree targeting the errors left by the previous additive fit. Under squared-error loss, this procedure corresponds to a regularized gradient-descent improvement of the forecasting function (see, e.g., Chapter 10 in Hastie, Tibshirani, and Friedman, 2009).

Fix the forecast horizon h. Let $\mathcal{D}_t$ be the sample available at forecast origin t, as defined in (7). The initial forecast function is the baseline SHAR fit,

$$F_0(\mathbf{z}_{i,\ell}) = \mathbf{x}'_{h,i,\ell}\,\widehat{\boldsymbol{\beta}}_{h|t},$$

where $\widehat{\boldsymbol{\beta}}_{h|t}$ is the pooled OLS estimate from (4) using $\mathcal{D}_t$.

For boosting iteration $m = 1, \ldots, M$, we first compute the current fitting error,

$$e_{i,\ell}^{(m)} = \mathrm{IV}(\mathbf{z}_{i,\ell}) - F_{m-1}(\mathbf{z}_{i,\ell})\,, \qquad (i,\ell) \in \mathcal{D}_t.$$

We then fit a Tree-SHAR model to these errors using the same splitting variables $\mathbf{z}_{i,\ell}$ as in Section 2.3. Let $\mathcal{T}^{(m)}$ denote the tree fitted at iteration m, and let $\{R_j^{(m)}\}_{j=1}^{J_h^*}$ denote its terminal leaves. The leaf assignment for observation (i,ℓ) is denoted by $j^{(m)}(i,\ell)$, so that $\mathbf{z}_{i,\ell} \in R^{(m)}_{j^{(m)}(i,\ell)}$.

Within each terminal leaf j, we estimate a local SHAR correction by regressing the current fitting errors on the SHAR regressor vector:

$$\widehat{\boldsymbol{\delta}}_{h,j}^{(m)} = \arg\min_{\boldsymbol{\delta}} \sum_{(i,\ell):\, \mathbf{z}_{i,\ell} \in R_j^{(m)}} \left(e_{i,\ell}^{(m)} - \mathbf{x}_{h,i,\ell}' \boldsymbol{\delta} \right)^2 . \tag{12}$$

This produces the tree-based correction

$$g_m(\mathbf{z}_{i,\ell}) = \mathbf{x}_{h,i,\ell}' \, \widehat{\boldsymbol{\delta}}_{h,j^{(m)}(i,\ell)}^{(m)}.$$

The boosted forecast function is updated using a learning rate $\nu \in (0, 1]$:

$$F_m(\mathbf{z}_{i,\ell}) = F_{m-1}(\mathbf{z}_{i,\ell}) + \nu \, g_m(\mathbf{z}_{i,\ell}) \, . \tag{13}$$

After M boosting iterations, the Boosted Tree-SHAR forecast of implied volatility at date $t+h$ for option i, conditional on its target moneyness–maturity point $\mathbf{z}_{i,t+h}$, is given by

$$\widehat{\mathrm{IV}}_{t+h\,|\,t}^{\mathrm{Boost}}(\mathbf{z}_{i,t+h}) = F_M(\mathbf{z}_{i,t+h}) = \mathbf{x}_{h,i,t+h}' \, \widehat{\boldsymbol{\beta}}_{h|t} + \nu \sum_{m=1}^{M} \mathbf{x}_{h,i,t+h}' \, \widehat{\boldsymbol{\delta}}_{h,j^{(m)}(i,t+h)}^{(m)}. \tag{14}$$

Tree fitting within each boosting iteration follows the same greedy splitting logic as in Tree-SHAR, except that candidate splits are evaluated using the residual-regression criterion implied by (12). In the empirical implementation, we keep each tree component deliberately simple. We set the number of terminal leaves equal to J_h^*, the value selected for Tree-SHAR using the CART cost-complexity criterion described in Appendix B, and impose the same minimum leaf-size constraint as in Tree-SHAR: each terminal leaf must contain at least 10,000 option observations. For the boosting regularization, we set the learning rate to $\nu = 0.1$ and the number of boosting iterations to $M = 100$, which are standard benchmark choices under squared-error loss and provide a stable bias–variance trade-off in practice.

This boosted specification differs from a standard boosted regression tree in one important respect. In a usual boosted tree, each terminal leaf contributes a constant fitted value. Here, each terminal leaf contributes a local HAR regression, summarized by a vector of coefficient adjustments $\widehat{\boldsymbol{\delta}}_{h,j}^{(m)}$. As a result, the contribution of each boosting tree depends not only on the moneyness–maturity

region assigned by $\mathbf{z}_{i,\ell}$, but also on the SHAR regressor vector $\mathbf{x}_{h,i,\ell}$. The model therefore preserves the HAR interpretation within each region while allowing the HAR coefficients to vary flexibly over the implied-volatility surface through repeated, region-specific coefficient adjustments.

3 Data

We use daily S&P 500 out-of-the-money (OTM) call and put options between January 2011 and August 2023 obtained from OptionMetrics. Due to the large cross-sectional size of the options data, which include less reliable data points, we use a filtered dataset, mainly following Dufays, Jacobs, and Rombouts (2025). We keep (i) options with maturities between 20 and 240 days, (ii) options with implied volatility above 5% and below 150%, (iii) options with moneyness (defined as underlying price/strike) between 0.8 and 1.6, (iv) options with mid price and bid price at least \$0.50 and \$0.375, respectively, and (v) options not subject to fundamental data errors, where the bid price exceeds the offer price or where a negative price is implied by put-call parity. We further exclude options whose volume and open interest are less than 100 contracts to ensure we include only sufficiently liquid contracts. The final filtered dataset comprises 786,867 option contracts.

We obtain the time series of S&P 500 index prices and returns from CRSP. The risk-free rate is proxied by the one-month Treasury bill rate obtained from CRSP as well. We download the VIX data from Bloomberg. Following the standard implementation in the literature, options are valued using a maturity-specific risk-free rate obtained from OptionMetrics.

4 Empirical Results

In this section, we discuss the forecasting performance of tree-based SHAR models. Section 4.1 reports the overall forecasting results and compare different models, especially between models with global homogeneous specifications and tree-based localized models. Section 4.2 examines cross-sectional performance variations across moneyness and maturity dimensions. Lastly, Section 4.3 illustrates the estimated tree structures on two dates representing good and bad periods. We further discuss the leaf-specific model coefficient estimates and their economic interpretations.

4.1 IV forecasting performance

This subsection evaluates the out-of-sample implied-volatility (IV) forecasting performance of the baseline Surface HAR (SHAR) model and its tree-based extensions proposed in Section 2. We assess the models' ability to predict the future implied volatilities over horizons of $h \in \{1, 5, 22\}$ forecasting days.

We evaluate forecasting performance over the OOS period 2018–2023 using an expanding-window scheme. For each forecast origin t in the OOS period, every model is estimated using only data available up to date t, and the resulting estimates are used to forecast implied volatility at date $t + h$. For the SHAR benchmark, this daily re-estimation updates the global HAR coefficients. For the tree-based specifications, this daily re-estimation updates both the tree partition structure—including splitting thresholds and leaf assignments—and the leaf-specific HAR coefficients. Thus, the Tree-SHAR, Bagged-SHAR, and Boosted-SHAR forecasts are generated from tree structures and coefficient estimates that are re-optimized at each OOS forecast origin.

We compare the proposed machine-learning specifications against three benchmarks:

1. **Random Walk (RW):** Assumes the future surface equals the current fitted surface.

2. **SHAR:** The baseline model of Dufays, Jacobs, and Rombouts (2025) with a single set of HAR coefficients for the entire surface.

3. **SHAR with 3×3 Grids:** An ad-hoc local model that partitions the moneyness–maturity space into nine fixed rectangular bins and estimates separate HAR coefficients within each bin. The bins are constructed by dividing each dimension into three empirical terciles: the moneyness cutoffs are the 1/3 and 2/3 sample quantiles of all observed moneyness values, and the maturity cutoffs are the 1/3 and 2/3 sample quantiles of all observed maturities. This produces $3 \times 3 = 9$ fixed regions and serves as a control to distinguish the benefits of *any* localization from the benefits of *data-driven* partitioning.

The complexity of the tree models is determined via the cost-complexity pruning described in Appendix B. The algorithm selects parsimonious structures, $J_1^* = 6$ leaves for the one-day horizon, $J_5^* = 8$ leaves for the five-day horizon, and increasing to $J_{22}^* = 11$ leaves for the one-month horizon ($h = 22$).

Table 1 reports the OOS Root Mean Squared Error (RMSE) in percentage points, revealing a strict hierarchy of model performance that holds across all forecast horizons. First, incorporating local heterogeneity is uniformly superior to global modeling. At the one-day horizon ($h = 1$), SHAR and Random Walk models produce aggregate RMSEs of roughly 2.17. Introducing the fixed 3×3 partition reduces this to 2.119. However, the data-driven Tree-SHAR significantly outperforms the fixed grid, lowering the RMSE to 1.985. This suggests that the optimal boundaries for volatility dynamics do not align with arbitrary moneyness–maturity cutoffs but are better identified algorithmically. Second, ensemble learning delivers substantial incremental gains. Bagging the Tree-SHAR (Bagged-SHAR) reduces the RMSE to 1.965, while gradient boosting (Boosted-SHAR) achieves the lowest error of 1.852. The superiority of boosting is consistent with its ability to sequentially correct biases in the surface fit, rather than simply reducing variance through averaging. The economic magnitude of these improvements increases with the forecast horizon. At the one-month horizon ($h = 22$), the SHAR model yields an RMSE of 5.910. The Tree-SHAR and Boosted-SHAR specifications reduce this value to 5.323 and 5.124, respectively. The Boosted-SHAR result represents a reduction in forecast error of approximately 13% relative to the global SHAR benchmark and over 18% relative to the Random Walk (RMSE 6.248).

The year-by-year decomposition in Table 1 further highlights the robustness of the tree-based methods during the extreme volatility period of 2020 (the onset of the COVID-19 pandemic). In 2020, all models experience elevated errors. However, the adaptability of the tree partitions proves crucial. For the 22-day horizon, the SHAR model suffers an RMSE of 12.177 in 2020. The Tree-SHAR and Boosted-SHAR models mitigate this error to 11.091 and 10.745, respectively. These outcomes indicate that when the implied volatility surface becomes steep and distorted during crises, the ability to isolate specific surface regions (such as deep OTM or short-term options) allows for more accurate dynamic modeling than a rigid global specification. The same argument holds for shorter-horizon forecasts ($h = 1$ or 5).[8]

[8]For $h = 1$ or 5, we observe that the RMSEs of the SHAR and ad-hoc 3×3 SHAR models in 2020 are slightly larger than those of the Random Walk. In contrast, tree-based SHAR models all outperform the Random Walk in 2020 regardless of the forecasting horizon. Furthermore, the performance difference of the ad-hoc 3×3 SHAR model suggests that our tree-based models' improvements are not simply due to partitioning the surface but due to assigning an optimal partition. Lastly, Dufays, Jacobs, and Rombouts (2025) also examine the SHAR model's behavior in 2020 and demonstrate that its performance can be enhanced through their Robust-SHAR extension, which effectively addresses the influence of volatility surface outliers.

Table 2 formalizes these comparisons using Diebold-Mariano (DM; Diebold and Mariano, 1995) tests and the Model Confidence Set (MCS; Hansen, Lunde, and Nason, 2011) confirming that the observed reductions in error are statistically significant. The DM statistics (comparing row vs. column) are positive and large across the board, rejecting the null hypothesis of equal predictive accuracy in favor of the more sophisticated models. For example, at $h = 22$, the test statistic comparing SHAR against Boosted-SHAR is 92.48, providing overwhelming evidence against the global model. Furthermore, the tests confirm that the gains from machine learning are statistically distinct: Boosting outperforms Bagging (statistic 45.98), and Bagging outperforms the single Tree (statistic 30.13). Finally, the Model Confidence Set results (last column of Table 2) are definitive. For every forecast horizon ($h = 1, 5, 22$), the Boosted-SHAR model is the *only* model retained in the confidence set at the 95% level. All other specifications—including the single Tree-SHAR and the Bagged-SHAR—are eliminated. This confirms that the stagewise additive expansion provided by gradient boosting captures complex nonlinearities in the evolution of the implied volatility surface that other methods miss.

In summary, the empirical evidence demonstrates that allowing HAR dynamics to vary across algorithmically selected regions of the moneyness–maturity space systematically improves forecasting accuracy.[9] The data-driven partitions outperform ad-hoc grids, and gradient boosting provides the most effective method for aggregating these local dynamics.

Panel A of Figure 3 provides a time series perspective on these forecast gains by plotting the daily out-of-sample RMSEs of the SHAR and Tree-SHAR models for the one-day horizon. The figure shows that the Tree-SHAR error series is almost uniformly below the SHAR error series throughout the evaluation period. This pattern indicates that the improvement from tree-based localization is not driven by a small number of isolated dates, but instead reflects a persistent reduction in forecast errors over time. The gains are visible both during relatively calm periods and during episodes of market stress, suggesting that the tree partition improves forecast accuracy systematically rather than only in crisis periods.

[9]When using the SVI method to generate the volatility surface as an alternative, we obtain highly consistent results and implications, as shown in Tables A.2 and A.3. The forecasting error sizes and patterns remain comparable to those obtained with the AHBS method. One exception is the five-day horizon forecast in 2020, where the Random Walk performs better than the Boosted-SHAR.

The advantage of Tree-SHAR is especially pronounced on days with large forecasting errors. Whenever daily RMSEs spike, the Tree-SHAR model typically produces a smaller error than the baseline SHAR model. This suggests that the tree partition is particularly valuable when the implied-volatility surface is harder to forecast, because it allows different regions of the surface to respond through region-specific HAR coefficients. In other words, during periods when a global SHAR specification generates large errors, the localized SHAR coefficients help mitigate those errors by adapting the forecasting dynamics across the moneyness–maturity surface.

Panels B and C of Figure 3 show that the same pattern holds across longer forecast horizons. For $h = 5$ and $h = 22$, the Tree-SHAR RMSE series remains generally below the SHAR series, and the gap is again most visible around dates with large forecast errors. Thus, the benefits of tree-based localization are not confined to one-day-ahead forecasting. They also appear at weekly and monthly horizons, where the implied-volatility surface is more difficult to predict and where regional differences in HAR dynamics become especially important.

Although Figure 3 reports only SHAR and Tree-SHAR for visual clarity, the aggregate results in Table 1 show that the ensemble extensions further improve on the single-tree specification. Bagged Tree-SHAR reduces the sensitivity of forecasts to any single tree partition, while Boosted Tree-SHAR sequentially corrects remaining forecast errors. Therefore, the figures should be interpreted as conservative illustrations of the gains from machine-learning-based localization: even the single Tree-SHAR model already dominates the global SHAR benchmark over time, and the bagged and boosted versions deliver additional reductions in forecast error.

4.2 Cross-sectional variations in IV forecasting performance

To further investigate the sources of forecasting improvements, we examine how the tree-based models perform across different regions of the moneyness-maturity space. Tables 3 and 4 decompose out-of-sample forecasting errors for the one-day horizon across moneyness and maturity buckets, respectively. At each point in time, we partition the sample into three groups based on the 20th and 80th percentiles of the respective dimension. This cross-sectional analysis reveals economically meaningful patterns that help explain why localized SHAR dynamics deliver superior forecasting performance.

In the moneyness dimension (Table 3), the benefits of tree-based localization are most pronounced at the extremes of the moneyness distribution. For OTM calls (Moneyness $\leq q_{20}$) and deep OTM puts (Moneyness $\geq q_{80}$), Boosted-SHAR consistently outperforms all competing specifications across all years. Tree-SHAR also dominates the Random Walk, SHAR, and SHAR with 3×3 grids benchmarks in these regions. These OTM options are particularly challenging to forecast using globally homogeneous models, as they exhibit distinct volatility dynamics driven by tail-risk concerns, crash premia, or jackpot expectations that differ fundamentally from NTM options. The localized HAR coefficients estimated by tree-based methods successfully capture these region-specific patterns, leading to substantial forecast improvements.

For NTM options and slightly OTM puts ($q_{20} <$ Moneyness $< q_{80}$), Boosted-SHAR continues to perform best in most years, with exceptions in 2020 and 2022. During the 2020 market stress period, the Random Walk assumption actually produces better short-horizon forecasting accuracy for NTM IVs compared to dynamic forecasting models. This pattern suggests that when the volatility surface undergoes rapid and dramatic shifts, high persistence dominates over dynamic patterns for NTM options. Importantly, the differences among model specifications in this moneyness bucket are substantially smaller than in the extreme moneyness regions, indicating that the Random Walk already performs reasonably well for NTM options. Nevertheless, the overall performance still improves when we apply tree-based SHAR approaches, particularly at longer horizons (see Tables A.4 and A.6 in Appendix). The outperformance of the Random Walk specification in the NTM bucket during 2020 remains confined to 5-day horizon forecasting, while it underperforms tree-based models in 22-day horizon forecasting. This result implies that during highly volatile periods, the very high persistence of NTM IVs is well-captured by the Random Walk over short horizons; the economic implication embedded in the HAR structure becomes valuable as the forecast horizon extends.

The largest improvements occur at the tails of the moneyness distribution when we employ Boosted-SHAR. Under the Random Walk or global SHAR specifications, OTM call and deep OTM put IVs exhibit substantially larger forecasting errors than NTM options. However, this pattern reverses when we use Boosted-SHAR. Specifically, the improvements of Boosted-SHAR relative to the global SHAR benchmark are approximately 30% and 26% for OTM calls and deep OTM puts, respectively, while the improvement in the NTM region is only about 2%. This stark contrast confirms that allowing for heterogeneous volatility dynamics across the moneyness dimension is

particularly valuable in regions where global models struggle and where the tree algorithm identifies the need for separate local specifications.

In the maturity dimension (Table 4), we observe a different pattern. Overall, IV forecast errors are larger for shorter-maturity options regardless of the forecasting approach. This baseline observation reflects the higher volatility of volatility for near-term contracts. Nevertheless, Boosted-SHAR performs best across all maturity buckets and all years for one-day horizon forecasts. The same pattern persists at longer horizons ($h = 5$ and 22), with only a few isolated cases where Bagged-SHAR achieves marginally better performance than Boosted-SHAR (see Tables A.5 and A.7).

Taken together, the cross-sectional results demonstrate that the tree-based partitioning successfully identifies economically meaningful regions of the implied volatility surface where distinct forecasting dynamics prevail. The localized SHAR parameters adapt to the unique persistence and mean-reversion characteristics of different option segments, thereby improving forecast accuracy most substantially in deep OTM and short-maturity regions.

4.3 Heterogeneity of the volatility surface

To understand how the tree-based approaches achieve their forecasting improvements, we examine Tree-SHAR's estimated surface partitions and the corresponding local HAR dynamics by focusing on one-day forecast horizon.[10] Figures 4 and 5 illustrate representative Tree-SHAR partitions estimated for two contrasting dates: January 4, 2018 (a calm market) and March 17, 2020 (a stressed market). Table 5 reports the estimated HAR coefficients for both the global SHAR model and the leaf-specific parameters from Tree-SHAR on these dates. These data-driven partitions, by and large, align with the regions where we observed the noticeable forecast improvements in Tables 3 and 4. This alignment confirms that the tree algorithm identifies economically meaningful surface segments where a localized model outperforms global specifications.

In Figure 4, Panel A shows the tree structure containing six terminal leaves, and Panel B maps these leaves onto the moneyness-maturity space, during a stable market state. The tree algorithm first isolates deep OTM calls (Leaf 1: Moneyness < 0.96), which constitute 10% of observations.

[10]We analyze Tree-SHAR rather than Boosted-SHAR because it is difficult to specify a single tree structure and a single set of parameters in Boosted-SHAR.

This region experiences distinct volatility dynamics, as evidenced by the substantially different HAR coefficients in Table 5. While the global SHAR model estimates a daily persistence coefficient of 0.909, Leaf 1 exhibits a notably lower daily coefficient of 0.744 and a negative monthly component of -0.096. In other words, deep OTM call IVs respond relatively less to yesterday's level and exhibit faster mean reversion than the surface average implied in the SHAR model. Deep OTM call options during good times are likely to be held by speculators seeking leveraged directional exposure. Although our tree structure does not particularly divide between short- and long-term OTM calls, Eaton, Green, Roseman, and Wu (2026) show that retail investors exhibit strong lottery-like preferences (e.g., OTM calls) and concentrate on trading short-term options. If IVs in Leaf 1 are primarily driven by such trading activities, we conjecture that option prices rapidly incorporate new information that is not captured by historical IVs and do not simply inherit yesterday's level, hence relatively lower daily coefficient, $\hat{\beta}^{(d)}_{h=1}$.

The tree then partitions the remaining surface along maturity and moneyness dimensions, creating economically interpretable regions that correspond to different investor clienteles and hedging demands. Leaf 6 (Moneyness ≥ 0.96, Maturity ≥ 140 days) captures the broad region of long-maturity options spanning NTM to deep OTM puts, representing the segment where institutional investors typically seek longer-term portfolio protection and hedging exposure. This region displays distinctive dynamics in Table 5. The daily persistence coefficient is only 0.412, but weekly and monthly coefficients are 0.442 and 0.279. In comparison, other leaves containing shorter-maturity options show much higher daily persistence and lower weekly or monthly persistence. This structure indicates that while these long-dated options do not simply inherit yesterday's volatility level, they exhibit substantial persistence at weekly and monthly horizons, consistent with the observation in Figure 2. A potential interpretation is that long-term hedging demand responds to persistent shifts in risks and macroeconomic uncertainty that evolve gradually over time, rather than reacting primarily to daily market fluctuations.

The shorter-maturity segment (Maturity < 39 days) is further partitioned by moneyness into Leaves 2, 3, and 4. Leaves 2 and 3, which cover NTM and slightly OTM puts, exhibit very high daily persistence with coefficients of 0.907 and 0.996, respectively. Leaf 4 captures short-dated deep OTM puts and shows an even higher daily persistence coefficient of 1.010. These short-maturity regions account for substantial trading volume from both hedgers and market makers, and their

IVs track relatively more recent market conditions closely. The uniformly high daily persistence of IV across short-term options suggests that these IVs are strongly anchored to yesterday's level, consistent with short-horizon market expectations that evolve smoothly day-to-day.[11]

Figure 5 shows how the tree partition adapts dramatically during the 2020 COVID-19 crisis period. While the tree still produces six leaves, the splitting thresholds and structure differ markedly from the good period. The tree algorithm now isolates deep OTM calls at a higher moneyness threshold (Moneyness < 0.92 for Leaf 1, compared to 0.96 in Figure 4), creating a more extreme partition. The moneyness cutoffs for other leaves shift leftward overall, implying more heterogeneous dynamics in OTM calls and NTM options but more homogeneous dynamics in a wide range of short-term OTM puts. In fact, the finer partitions in OTM put regions during good times align with empirical observations that investors are more fearful of negative market shocks during good periods, differentiating among various tail-risk scenarios and embedding more heterogeneous information in OTM puts (see, e.g., Kim, 2025; Schreindorfer and Sichert, 2025). Conversely, during bad times, the coarser short-term OTM put partitions suggest that investors respond more uniformly to crash-risk concerns.

The estimated HAR coefficients in Table 5 for March 17, 2020 present even more pronounced heterogeneity than the good period. The global SHAR model now estimates substantially lower daily persistence (0.843 versus 0.909 in 2018) and a sizable positive weekly coefficient (0.191 versus -0.012 in 2018), reflecting the general increase in the importance of weekly dynamics during the crisis. However, this global pattern masks extreme variation across the surface.

Leaf 1, which captures deep OTM calls, now displays radically different dynamics compared to the deep OTM calls segment during good times (Leaf 1 in Figure 4). The daily persistence coefficient is only 0.277, but a weekly coefficient is 0.892, and a monthly coefficient is largely negative, -0.831. This structure implies that during extreme stress, such speculative call options' IVs depend far more on the weekly moving average of recent surfaces than on yesterday's observation alone. The highly negative monthly coefficient suggests sharp mean reversion from any sustained elevation of volatilities. We conjecture that this pattern aligns with a market where speculative demand for lottery-like upside exposure persists for some days but not a long period of time, along with market expectations that a crisis may persist for some period.

[11] Implied volatilities represent forward-looking short-horizon market expectations of volatility.

Similarly, Leaf 6 (long-maturity OTM puts and NTM options) shows a daily coefficient of only 0.532 but a weekly coefficient of 0.635 during the crisis, again emphasizing the shift away from daily toward weekly persistence. However, its monthly persistence disappears in contrast to the sizable monthly persistence observed for long-term OTM put IVs during good times (-0.008 versus 0.279 in 2018). On the other hand, Leaf 4 (short-maturity OTM puts) maintains very high daily persistence of 0.939 even during the crisis, implying high persistence of short-horizon expectation of crash risk. All in all, this heterogeneity explains why the tree-based approach substantially outperforms global models during stressed periods: no single set of HAR coefficients can adequately describe the divergent dynamics across the surface when market conditions fracture the usual relationships.

Across our OOS evaluation period, we find that the tree structure remains relatively stable under similar market conditions but adjusts noticeably following major market shocks or volatility regime shifts. This combination of persistence in tree structure (reducing estimation noise over time) and adaptability when needed (capturing evolution of market segmentation) contributes to the tree-based models' robust OOS forecasting performance. The localized HAR specifications effectively allow short-term options, long-term options, and deep OTM calls and puts to exhibit distinct dynamics reflecting their specialized clienteles. A single global set of coefficients cannot accommodate such adjustments, but the data consistently identify these adjustments as improving forecast accuracy across the surface.

5 Conclusion

Standard implied volatility forecasting models typically impose homogeneous dynamics across the entire option surface. While this assumption is convenient, it overlooks the substantial heterogeneity in volatility dynamics across moneyness and maturity. In this paper, we propose a machine-learning approach that uses regression trees to identify data-driven regions of the moneyness–maturity space and estimate local forecasting dynamics within each region. Building on the SHAR framework, our Tree-SHAR model preserves the interpretability of HAR-type implied volatility dynamics while allowing the coefficients to vary across economically meaningful regions of the option surface.

Using daily S&P 500 index options, we show that incorporating cross-sectional heterogeneity substantially improves out-of-sample implied volatility forecasts. Tree-SHAR outperforms the global

SHAR benchmark and the Random Walk benchmark across forecast horizons. It also improves on an ad-hoc 3×3 grid specification, indicating that the gains are not simply due to local estimation, but rather to the data-driven selection of moneyness–maturity partitions. The ensemble extensions further strengthen these results. Bagged Tree-SHAR reduces the sensitivity of forecasts to any single estimated partition, while Boosted Tree-SHAR delivers the lowest forecast errors overall by sequentially correcting remaining prediction errors. These gains are statistically significant and are observed under a daily expanding-window forecasting exercise in which both the tree structures and the local coefficients are re-estimated at each out-of-sample forecast origin.

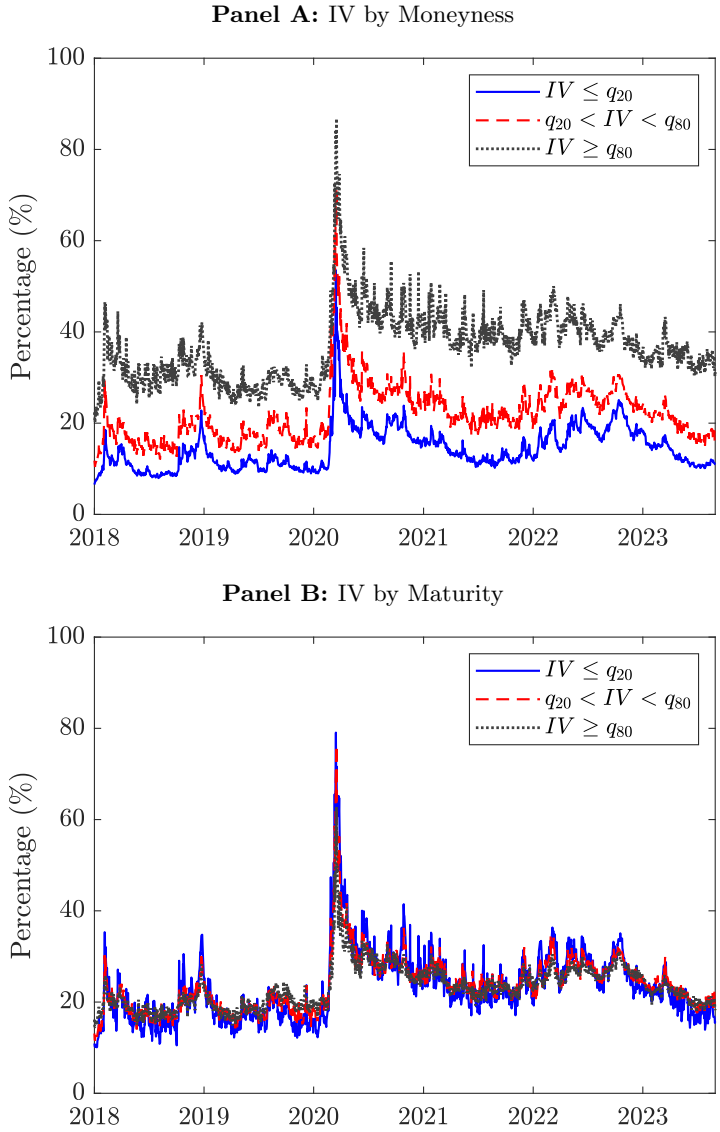
The forecasting improvements are economically informative. The gains from tree-based localization are especially large in regions where global models perform poorly, such as OTM calls, deep OTM puts, and short-maturity options. They are also pronounced during periods with large forecast errors, including the COVID-19 stress period, but remain visible during calmer market conditions. The estimated tree partitions and leaf-specific coefficient estimates show that different regions of the option surface display distinct persistence patterns. Short-dated options tend to place more weight on recent fitted surfaces, whereas longer-dated options exhibit stronger weekly and monthly components. These findings suggest that the implied-volatility surface is governed by heterogeneous dynamic behavior that cannot be adequately captured by a single global set of coefficients.

Overall, the results highlight the value of combining interpretable financial forecasting models with machine-learning-based localization. Regression trees provide a flexible but transparent way to determine where local models should be estimated, while the SHAR structure maintains a clear economic interpretation of daily, weekly, and monthly volatility persistence. One natural extension is to enrich the cross-sectional splitting variables used to partition the option surface. Beyond moneyness and maturity, additional option-level characteristics such as option Greeks, trading volume, open interest, or bid–ask spreads may help identify more refined forms of heterogeneity across the option panel. A second direction is to extend the tree-based approach from the cross-sectional dimension to the time-series dimension. For example, one could use market-wide state variables such as the VIX, past returns, realized volatility, or other measures of financial conditions to determine when forecasting dynamics change over time. This would connect our framework to the broader literature on tree-structured time-series models, which uses observed state variables to identify regimes with distinct dynamic parameters (Audrino and Bühlmann, 2001; Patton and

Simsek, 2026). Combining cross-sectional option characteristics with time-series state variables may provide a richer framework for modeling how implied-volatility dynamics vary both across the option surface and across market conditions.

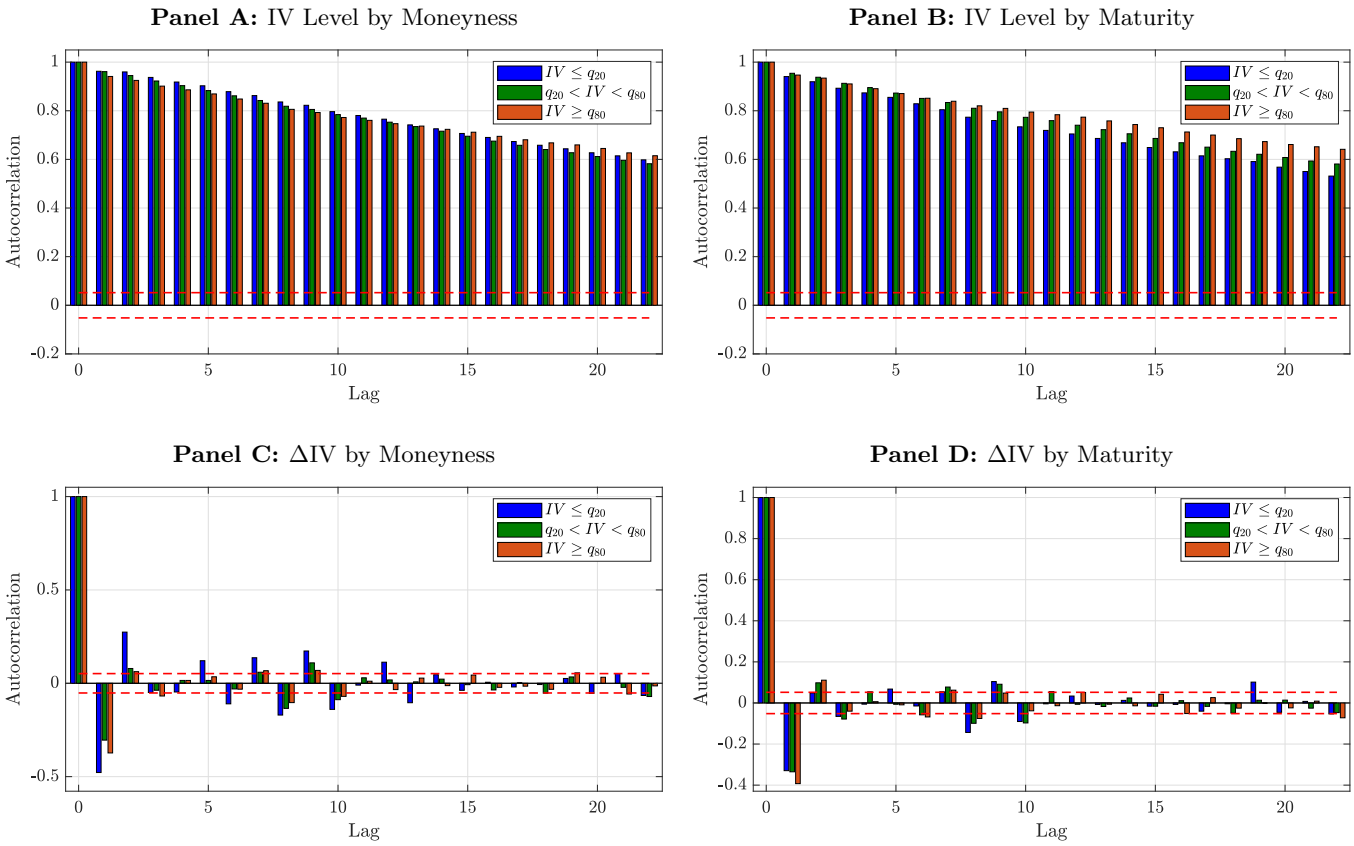
6 Figures and Tables

Figure 1: Implied Volatilities by Moneyness and Maturity in Data



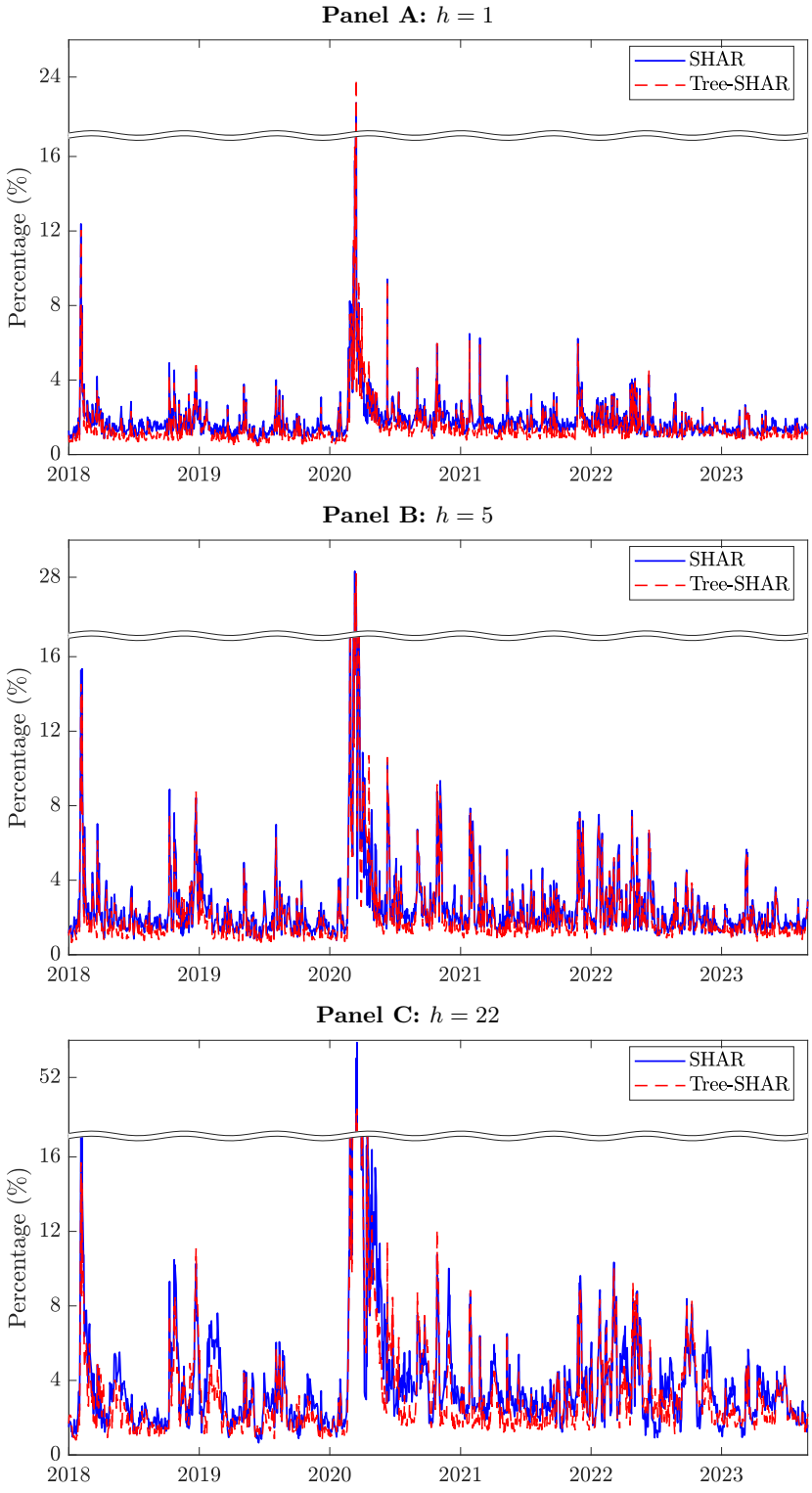
Notes: This figure depicts the average implied volatilities (IV) by moneyness and maturity groups over time from January 2018 to August 2023. On each date, three groups are classified such that the moneyness or maturity is less than or equal to its 20th percentile (blue solid line), between the 20th and 80th percentiles (red dashed line), and greater than or equal to the 80th percentile (black dotted line). Panel A shows IV by moneyness and Panel B shows IV by maturity. All IVs are expressed in percentage points.

Figure 2: Autocorrelation Function of Implied Volatilities



Notes: This figure depicts the autocorrelation functions (ACF) of implied volatilities (IVs) and their first differences up to 22 lags. Sample ACFs are computed using the average IVs within three groups classified such that the moneyness or maturity is less than or equal to its 20th percentile (blue bar), between the 20th and 80th percentiles (green bar), and greater than or equal to the 80th percentile (orange bar). Panels A and B plot the ACF of IVs by moneyness and maturity groups. Panels C and D plot the ACF of daily changes (first differences) in IVs. The red dashed lines represent the 95% confidence intervals.

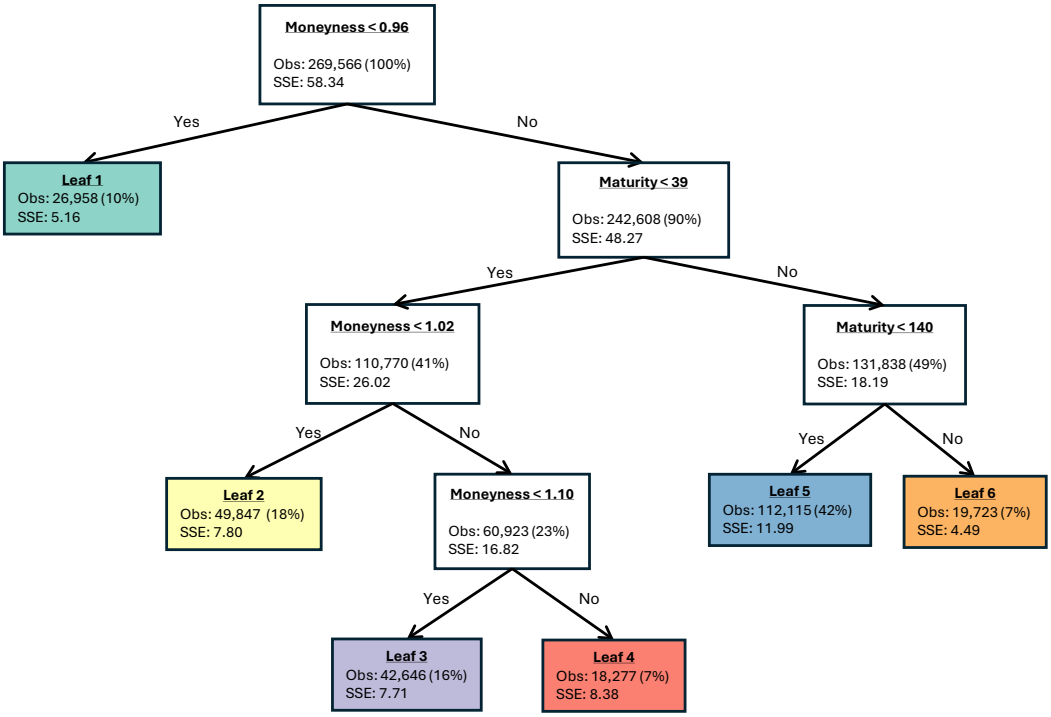
Figure 3: Daily Out-of-Sample RMSEs for SHAR and Tree-SHAR Across Forecast Horizons



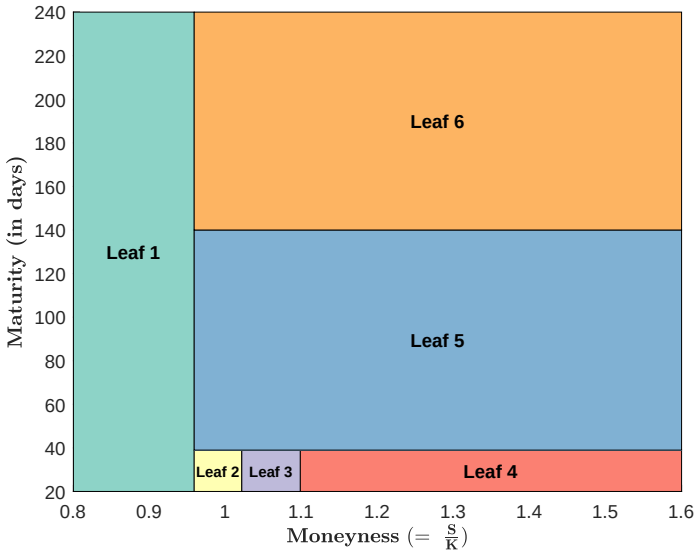
Notes: This figure plots the daily out-of-sample RMSEs of implied-volatility forecasts for the SHAR and Tree-SHAR models across forecast horizons $h = 1$, 5, and 22. Panel A reports the one-day-ahead horizon, Panel B reports the five-day-ahead horizon, and Panel C reports the 22-day-ahead horizon. The blue solid line reports the daily RMSE of the baseline SHAR model, while the red dashed line reports the daily RMSE of the Tree-SHAR model. Daily RMSEs are computed across all options observed on each forecast target date and are expressed in percentage points of implied volatility.

Figure 4: Estimated Tree-SHAR Partition on January 4, 2018: $h = 1$

Panel A: Estimated Tree Structure



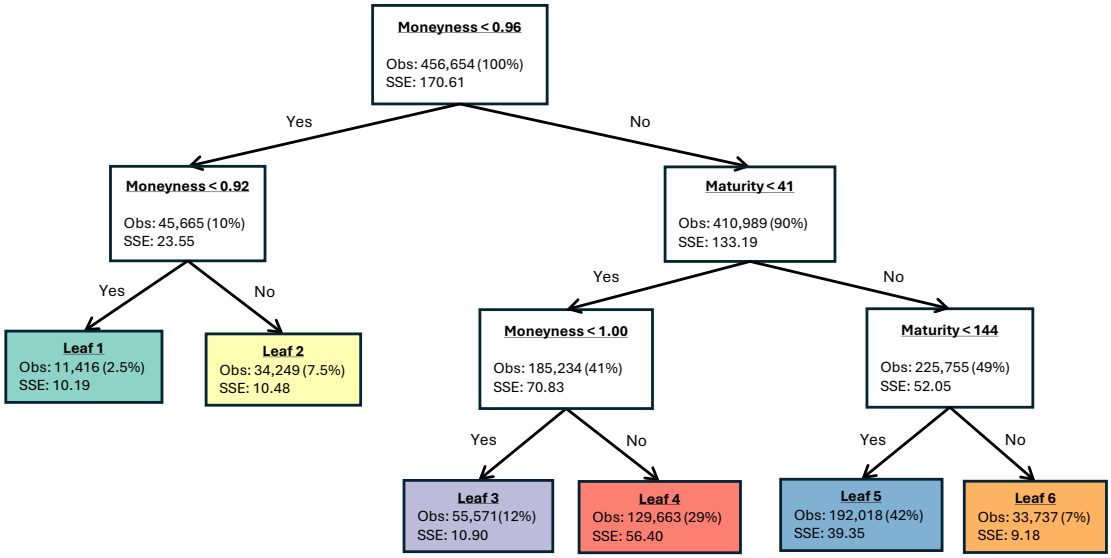
Panel B: Induced Moneyness–Maturity Partition



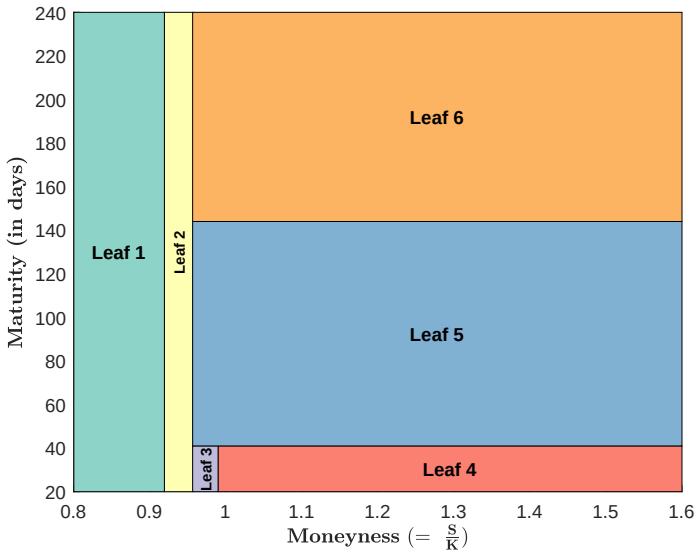
Notes: This figure illustrates the estimated Tree-SHAR partition for the one-day-ahead forecast horizon on January 4, 2018, using AHBS-fitted implied-volatility surfaces. Panel A reports the estimated regression tree. Each internal node shows the selected splitting variable and threshold, while each terminal leaf reports the number of option observations assigned to that leaf and the corresponding sum of squared errors (SSE). Panel B maps the same terminal leaves into the moneyness–maturity space. The figure shows how the Tree-SHAR algorithm partitions the option surface into economically interpretable regions and estimates separate local HAR coefficients within each region.

Figure 5: Estimated Tree-SHAR Partition on March 17, 2020: $h = 1$

Panel A: Estimated Tree Structure



Panel B: Induced Moneyness–Maturity Partition



Notes: This figure illustrates the estimated Tree-SHAR partition for the one-day-ahead forecast horizon on March 17, 2020, using AHBS-fitted implied-volatility surfaces. Panel A reports the estimated regression tree. Each internal node shows the selected splitting variable and threshold, while each terminal leaf reports the number of option observations assigned to that leaf and the corresponding sum of squared errors (SSE). Panel B maps the same terminal leaves into the moneyness–maturity space. Compared with Figure 4, the partition highlights how the tree structure adapts during the COVID-19 stress period by isolating more extreme regions (Leaf 1: deep OTM calls) of the moneyness–maturity surface.

Table 1: Out-of-Sample Implied Volatility Forecasting Performance (RMSE in %)

	2018	2019	2020	2021	2022	2023	Sum
			$\underline{h=1}$				
Random Walk	2.206	1.521	3.530	2.045	1.792	1.442	2.173
SHAR	2.186	1.514	3.563	2.029	1.786	1.443	2.173
SHAR with 3×3 grids	2.108	1.404	3.565	1.912	1.758	1.358	2.119
Tree-SHAR (6 leaves)	1.909	1.260	3.480	1.691	1.650	1.220	1.985
Bagged-SHAR (6 leaves)	1.897	1.245	3.457	1.673	1.617	1.193	1.965
Boosted-SHAR (6 leaves)	**1.767**	**1.112**	**3.393**	**1.511**	**1.490**	**0.985**	**1.852**
			$\underline{h=5}$				
Random Walk	3.522	2.293	6.380	2.855	2.936	1.987	3.571
SHAR	3.448	2.214	6.534	2.803	2.874	2.000	3.577
SHAR with 3×3 grids	3.337	2.035	6.459	2.670	2.881	1.885	3.496
Tree-SHAR (8 leaves)	3.071	1.884	6.316	2.485	2.729	1.774	3.343
Bagged-SHAR (8 leaves)	3.061	1.866	6.294	2.470	2.706	1.740	3.325
Boosted-SHAR (8 leaves)	**2.912**	**1.716**	**6.283**	**2.346**	**2.617**	**1.656**	**3.250**
			$\underline{h=22}$				
Random Walk	5.013	3.500	12.781	3.512	4.986	2.763	6.248
SHAR	4.646	3.174	12.177	3.512	4.455	2.922	5.910
SHAR with 3×3 grids	4.372	2.781	11.555	3.215	4.573	2.570	5.615
Tree-SHAR (11 leaves)	4.006	2.443	11.091	2.956	4.303	2.496	5.323
Bagged-SHAR (11 leaves)	3.987	2.404	11.036	2.899	4.253	2.441	5.284
Boosted-SHAR (11 leaves)	**3.792**	**2.253**	**10.745**	**2.813**	**4.135**	**2.362**	**5.124**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models. The forecasting performance is evaluated across three different horizons (1 day, 5 days, and 22 days) and decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table 2: Pairwise Diebold-Mariano Test Statistics and Model Confidence Set (MCS) for SHAR

	RW	SHAR	SHAR (3×3)	Tree-SHAR	Bagged-SHAR	Boosting-SHAR	MCS
			$\underline{h = 1}$				
Random Walk		-0.65	36.85	60.41	68.82	87.56	
SHAR	0.65		43.67	63.32	72.20	91.80	
SHAR with 3×3 grids	-36.85	-43.67		45.52	53.99	80.76	
Tree-SHAR (6 leaves)	-60.41	-63.32	-45.52		33.42	61.81	
Bagged-SHAR (6 leaves)	-68.82	-72.20	-53.99	-33.42		53.37	
Boosted-SHAR (6 leaves)	-87.56	-91.80	-80.76	-61.81	-53.37		✓
			$\underline{h = 5}$				
Random Walk		-3.82	24.24	49.37	56.30	53.43	
SHAR	3.82		30.01	53.15	60.69	56.14	
SHAR with 3×3 grids	-24.24	-30.01		39.08	45.81	48.44	
Tree-SHAR (8 leaves)	-49.37	-53.15	-39.08		21.02	22.82	
Bagged-SHAR (8 leaves)	-56.30	-60.69	-45.81	-21.02		19.13	
Boosted-SHAR (8 leaves)	-53.43	-56.14	-48.44	-22.82	-19.13		✓
			$\underline{h = 22}$				
Random Walk		62.70	90.90	97.86	101.12	104.12	
SHAR	-62.70		78.67	80.27	86.84	92.48	
SHAR with 3×3 grids	-90.90	-78.67		53.37	62.52	74.83	
Tree-SHAR (11 leaves)	-97.86	-80.27	-53.37		30.13	50.84	
Bagged-SHAR (11 leaves)	-101.12	-86.84	-62.52	-30.13		45.98	
Boosted-SHAR (11 leaves)	-104.12	-92.48	-74.83	-50.84	-45.98		✓

Notes: This table reports the pairwise Diebold-Mariano (DM) test statistics comparing the out-of-sample forecasting errors of the competing models across 1, 5, and 22-day horizons. A negative t-statistic indicates that the model in the row outperforms (has lower error than) the model in the column, while a positive value indicates that the column model outperforms the row model. Statistical significance at the 1% level corresponds to an absolute t-statistic greater than 2.33 ($|t| > 2.33$). The table also indicates with a checkmark the models that are included in the Model Confidence Set (MCS) at the 95% confidence level.

Table 3: Out-of-Sample Implied Volatility Forecasting Performance by Moneyness (RMSE in %, $h = 1$)

	2018	2019	2020	2021	2022	2023	Total
			Moneyness $\leq q_{20}$				
Random Walk	2.330	1.852	3.781	2.316	1.959	1.720	2.395
SHAR	2.305	1.817	3.721	2.276	1.940	1.678	2.358
SHAR with 3x3 grids	2.165	1.637	3.616	2.104	1.821	1.515	2.224
Tree-SHAR	1.667	1.272	3.357	1.604	1.591	1.348	1.910
Bagged-SHAR	1.660	1.274	3.373	1.605	1.559	1.287	1.901
Boosted-SHAR	**1.368**	**1.073**	**3.075**	**1.319**	**1.281**	**0.991**	**1.643**
			$q_{20} <$ Moneyness $< q_{80}$				
Random Walk	1.950	1.295	**3.387**	1.663	1.506	1.088	1.930
SHAR	1.929	1.314	3.454	1.659	**1.502**	1.114	1.947
SHAR with 3x3 grids	1.905	1.277	3.474	1.607	1.541	1.086	1.941
Tree-SHAR	1.902	1.237	3.494	1.558	1.599	1.037	1.941
Bagged-SHAR	1.888	1.223	3.479	1.547	1.575	1.026	1.927
Boosted-SHAR	**1.833**	**1.142**	3.509	**1.511**	1.541	**0.953**	**1.902**
			Moneyness $\geq q_{80}$				
Random Walk	2.713	1.738	3.667	2.663	2.302	1.948	2.565
SHAR	2.698	1.703	3.704	2.641	2.304	1.947	2.563
SHAR with 3x3 grids	2.561	1.497	3.766	2.444	2.222	1.823	2.471
Tree-SHAR	2.141	1.312	3.560	2.095	1.843	1.535	2.176
Bagged-SHAR	2.134	1.277	3.477	2.047	1.784	1.504	2.130
Boosted-SHAR	**1.922**	**1.061**	**3.357**	**1.679**	**1.535**	**1.067**	**1.903**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of one-day horizon ($h = 1$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three moneyness groups. On each date, three groups are classified such that the moneyness is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table 4: Out-of-Sample Implied Volatility Forecasting Performance by Maturity (RMSE in %, $h = 1$)

	2018	2019	2020	2021	2022	2023	Total
			Maturity $\leq q_{20}$				
Random Walk	2.520	1.793	4.313	2.441	2.223	1.755	2.634
SHAR	2.550	1.810	4.380	2.484	2.260	1.797	2.675
SHAR with 3x3 grids	2.488	1.710	4.313	2.379	2.199	1.704	2.602
Tree-SHAR	2.213	1.464	4.114	2.004	2.081	1.468	2.383
Bagged-SHAR	2.193	1.446	4.121	1.982	2.044	1.444	2.368
Boosted-SHAR	**2.068**	**1.313**	**4.044**	**1.828**	**1.867**	**1.166**	**2.240**
			$q_{20} <$ Maturity $< q_{80}$				
Random Walk	2.126	1.317	3.342	1.840	1.673	1.310	2.014
SHAR	2.074	1.302	3.357	1.779	1.633	1.282	1.988
SHAR with 3x3 grids	2.067	1.280	3.375	1.771	1.685	1.263	1.995
Tree-SHAR	1.865	1.146	3.353	1.525	1.511	1.082	1.861
Bagged-SHAR	1.862	1.136	3.329	1.511	1.482	1.063	1.844
Boosted-SHAR	**1.750**	**1.033**	**3.289**	**1.396**	**1.407**	**0.949**	**1.768**
			Maturity $\geq q_{80}$				
Random Walk	1.936	1.533	2.694	1.882	1.393	1.275	1.827
SHAR	1.905	1.503	2.701	1.848	1.383	1.269	1.812
SHAR with 3x3 grids	1.644	1.211	2.805	1.497	1.229	1.049	1.653
Tree-SHAR	1.571	1.201	2.786	1.568	1.287	1.157	1.666
Bagged-SHAR	1.556	1.182	2.708	1.550	1.245	1.110	1.627
Boosted-SHAR	**1.367**	**0.987**	**2.611**	**1.281**	**1.087**	**0.809**	**1.456**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of one-day horizon ($h = 1$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three maturity groups. On each date, three groups are classified such that the maturity is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table 5: SHAR and Tree-SHAR Coefficient Estimates on Representative Forecast Dates

	SHAR	Tree–SHAR					
		Leaf 1	Leaf 2	Leaf 3	Leaf 4	Leaf 5	Leaf 6
				January 4th, 2018			
Const.	0.001	0.047	-0.005	0.001	0.016	0.006	-0.023
	(6.32)	(52.97)	(-6.65)	(0.85)	(8.12)	(29.67)	(-27.92)
$\beta_{h=1}^{(d)}$	0.909	0.744	0.907	0.996	1.010	0.846	0.412
	(52.90)	(19.82)	(28.53)	(23.44)	(18.90)	(37.79)	(8.55)
$\beta_{h=1}^{(w)}$	-0.012	-0.029	-0.020	-0.085	-0.162	0.034	0.442
	(-0.68)	(-0.67)	(-0.63)	(-2.05)	(-2.62)	(1.41)	(7.50)
$\beta_{h=1}^{(m)}$	0.097	-0.096	0.098	0.100	0.141	0.077	0.279
	(11.83)	(-4.39)	(6.25)	(5.62)	(4.98)	(6.96)	(7.84)
				March 17th, 2020			
Const.	0.002	0.085	0.043	0.007	-0.013	0.007	-0.028
	(5.96)	(24.31)	(22.41)	(8.20)	(-20.56)	(28.06)	(-34.42)
$\beta_{h=1}^{(d)}$	0.843	0.277	0.653	0.800	0.939	0.783	0.532
	(49.75)	(3.00)	(16.25)	(23.84)	(32.08)	(36.08)	(13.38)
$\beta_{h=1}^{(w)}$	0.191	0.892	0.178	0.052	0.054	0.267	0.635
	(7.98)	(5.42)	(2.40)	(1.52)	(1.52)	(7.46)	(9.95)
$\beta_{h=1}^{(m)}$	-0.037	-0.831	-0.162	0.018	0.100	-0.097	-0.008
	(-2.32)	(-10.61)	(-4.27)	(0.97)	(4.27)	(-3.88)	(-0.18)

Notes: This table reports coefficient estimates from the baseline SHAR model and the Tree-SHAR model for two representative one-day-ahead forecasting dates. The first panel reports estimates for January 4, 2018, a relatively normal market day, while the second panel reports estimates for March 17, 2020, during the COVID-19 stress period. The SHAR column reports the global coefficient estimates from the baseline SHAR model. The Tree-SHAR columns report leaf-specific coefficient estimates for the six terminal leaves selected by the Tree-SHAR partition. These leaves correspond to the moneyness–maturity partitions displayed in Figures 4 and 5. Newey–West t-statistics computed with 10 lags are reported in parentheses. The results illustrate how Tree-SHAR allows the HAR dynamics to vary across regions of the moneyness–maturity surface, while the baseline SHAR model imposes a single set of coefficients across all options.

Appendix

A SVI Method

Although we use the AHBS method to interpolate and extrapolate IVs to generate the entire volatility surface in our main analyses, we consider the SVI method of Gatheral (2004) as an alternative approach. We specifically use its extended version suggested by Beason and Schreindorfer (2022) to span IVs along both the moneyness and maturity dimensions.

The SVI method differs from the AHBS method in two main ways. First, the functional form of IVs is given as

$$\mathrm{IV}^2(\mathbf{z}) = a + b\left(\rho(\kappa-\mu) + \sqrt{(\kappa-\mu)^2+\sigma^2}\right), \tag{A.1}$$

where $\mathbf{z} = (m, \tau)$ represents the option characteristics (moneyness and time to maturity), $\kappa = \log(K/F_\tau)/(VIX/100 \times \sqrt{\tau})$ is the standardized moneyness, which is a function of the strike K, forward price F_τ, and VIX, and the remaining five coefficients ($a,\ b,\ \rho,\ \mu,\ \sigma$) are linear functions of τ. For example, $a = a_0 + a_1\tau$. Second, the SVI method incorporates no-arbitrage constraints, which are absent in the AHBS method. Specifically, we impose the following restrictions: IVs should be nonnegative, monotonic in K, and convex in K; additionally, τ-maturity model-implied volatility $\mathrm{IV}^M(\mathbf{z}) \times \sqrt{\tau}$ should be monotonic in τ.

For implementation, on each day, we find the optimal set of parameters $\{a_0,\ a_1,\ b_0,\ b_1,\ \rho_0,\ \rho_1,\ \mu_0,\ \mu_1,\ \sigma_0,\ \sigma_1\}$ by minimizing the IV errors calculated through equation (A.1). Then, we replace the model-implied volatility $\mathrm{IV}^M(\mathbf{z})$ with the fitted value from equation (A.1) for any option characteristics $\mathbf{z}$ of interest.

B Selecting the Number of Leaves for Tree–SHAR

We select the number of terminal leaves in Tree–SHAR by combining a validation-risk measure in implied-volatility space with the cost-complexity principle of CART (Breiman, Friedman, Olshen, and Stone, 1984, 2017). The key object is a validation-risk curve indexed by tree size. Let d denote the number of accepted splits, so that a tree with size d has $L(d) = d + 1$ terminal leaves. For each candidate size d, we compute a validation forecasting loss and then add a linear penalty in the number of splits. The selected tree size is the one that minimizes this penalized validation criterion.

Training–validation split with expanding-window evaluation. We choose the Tree–SHAR tree size based on validation forecasting performance. We take the first half of the full sample as the in-sample period and split this in-sample period chronologically at its midpoint. The first half of the in-sample period is used as the initial training subsample, while the second half of the in-sample period is used as the validation subsample. For each forecast horizon h and each candidate tree size corresponding to $J \in \{2, \ldots, 15\}$ terminal leaves, equivalently $d \in \{1, \ldots, 14\}$ accepted splits, we impose a minimum leaf size of 10,000 option observations.

For each candidate size d, we conduct an expanding-window forecasting exercise within the in-sample period. Let $\mathcal{I}_{\text{val},h}$ denote the set of validation forecast origins t such that t lies in the validation subsample and the target date $t+h$ also lies in the validation subsample. At each forecast origin $t \in \mathcal{I}_{\text{val},h}$, we estimate the Tree–SHAR partition and the leaf-specific SHAR coefficients using only observations available up to date t. This expanding-window estimation sample starts with the initial training subsample and then expands through the validation period by adding all validation observations dated no later than t. We then form h-step-ahead forecasts of implied volatility at date $t+h$, conditional on information available at date t. Thus, the forecast is produced at time t, while its forecast error is evaluated ex post using the implied volatilities at the validation target date $t+h$.

Let $\widehat{\text{IV}}^{\text{Tree}}_{t+h\,|\,t}(\mathbf{z}_{i,t+h}; d)$ denote the Tree–SHAR forecast for option i at date $t+h$, produced at forecast origin t using a tree with $L(d) = d+1$ terminal leaves. Let $n_{\text{val},h} = \sum_{t \in \mathcal{I}_{\text{val},h}} N_{t+h}$ be the total number of options in the validation samples. We summarize validation performance by the average of squared forecast errors in implied-volatility space:

$$\mu_h^{\text{Tree}}(d) = \frac{1}{n_{\text{val},h}} \sum_{t \in \mathcal{I}_{\text{val},h}} \sum_{i=1}^{N_{t+h}} \left(\widehat{\text{IV}}^{\text{Tree}}_{t+h\,|\,t}(\mathbf{z}_{i,t+h}; d) - \text{IV}(\mathbf{z}_{i,t+h}) \right)^2 .$$

The curve $\mu_h^{\text{Tree}}(d)$ therefore measures the validation forecasting loss associated with each candidate tree size at horizon h. Because the validation exercise is conducted entirely within the first half of the sample, the second-half out-of-sample evaluation period is not used when selecting the tree size.

Cost-complexity criterion. Following the cost-complexity logic of CART (Breiman, Friedman, Olshen, and Stone, 1984, 2017), we regularize tree size by penalizing additional splits. For each

forecast horizon h, we define the penalized validation criterion

$$K_h(d) = \mu_h^{\text{Tree}}(d) + \lambda \cdot \frac{L(d) - 1}{n_{\text{val},h}},$$

where $\lambda \geq 0$ governs the strength of the complexity penalty. Since $L(d) - 1 = d$, the penalty increases linearly with the number of accepted splits. Intuitively, an additional split is selected only if the reduction in validation forecasting loss is large enough to offset the added complexity. In practice, we set $\lambda = 0.25$ as a rule-of-thumb complexity penalty. Our results are not sensitive to small variations around this value.

We then select the optimal number of splits as

$$\widehat{d}_h = \arg \min_{d \in \{1,\ldots,14\}} K_h(d),$$

and set the selected number of terminal leaves to

$$J_h^* = L(\widehat{d}_h) = \widehat{d}_h + 1.$$

We use this selected value J_h^* for the Tree–SHAR model at horizon h, and we use the same value when constructing the bagged and boosted Tree–SHAR specifications.

Implementation and selected J_h^*. Table A.1 reports the penalized criterion values $K_h(d)$, equivalently as a function of the number of leaves $J = L(d)$, for each candidate Tree–SHAR size and each forecast horizon h when we use the AHBS method to generate the volatility surface. We set J_h^* equal to the number of leaves associated with the smallest reported criterion value. Accordingly, Table A.1 implies $J_1^* = 6$, $J_5^* = 8$, and $J_{22}^* = 11$ for forecast horizons $h = 1$, 5, and 22, respectively. Using the SVI method to generate the volatility surface, we obtain $J_1^* = 6$, $J_5^* = 4$, and $J_{22}^* = 9$ for forecast horizons $h = 1$, 5, and 22, respectively.

Table A.1: CART $K_h(d)$ (10^{-4}) with $\lambda = 0.25$ to select optimal number of leaves

d	Leaves	Tree–SHAR ($h = 1$)	Tree–SHAR ($h = 5$)	Tree–SHAR ($h = 22$)
1	2	2.311	6.067	10.716
2	3	2.185	5.915	10.540
3	4	2.124	5.787	10.255
4	5	2.085	5.736	9.997
5	6	**2.083**	5.679	9.956
6	7	2.108	5.660	9.880
7	8	2.102	**5.651**	9.816
8	9	2.097	5.658	9.800
9	10	2.093	5.663	9.799
10	11	2.096	5.667	**9.791**
11	12	2.103	5.661	9.814
12	13	2.114	5.674	9.822
13	14	2.124	5.691	9.833
14	15	2.138	5.707	9.846

C Additional Tables

Table A.2: Out-of-Sample Implied Volatility Forecasting Performance (RMSE in %) (SVI)

	2018	2019	2020	2021	2022	2023	Sum
			$\underline{h=1}$				
Random Walk	2.035	1.248	3.394	1.762	1.695	1.405	2.020
SHAR	1.989	1.219	3.416	1.722	1.674	1.363	2.001
SHAR with 3×3 grids	1.903	1.155	3.427	1.637	1.591	1.218	1.943
Tree-SHAR (6 leaves)	1.875	1.134	3.399	1.632	1.592	1.189	1.925
Bagged-SHAR (6 leaves)	1.866	1.125	3.388	1.623	1.584	1.183	1.917
Boosted-SHAR (6 leaves)	**1.796**	**1.083**	**3.359**	**1.562**	**1.517**	**1.077**	**1.864**
			$\underline{h=5}$				
Random Walk	3.272	2.044	**6.105**	2.623	2.790	1.931	3.371
SHAR	3.178	1.958	6.305	2.539	2.731	1.893	3.380
SHAR with 3×3 grids	3.088	1.863	6.268	2.446	2.696	1.761	3.319
Tree-SHAR (4 leaves)	3.085	1.850	6.285	2.501	2.743	1.797	3.339
Bagged-SHAR (4 leaves)	3.082	1.846	6.284	2.500	2.736	1.795	3.337
Boosted-SHAR (4 leaves)	**2.971**	**1.715**	6.171	**2.378**	**2.632**	**1.694**	**3.236**
			$\underline{h=22}$				
Random Walk	4.643	3.219	12.143	3.286	4.733	2.678	5.909
SHAR	4.284	2.925	11.703	3.183	4.227	2.731	5.614
SHAR with 3×3 grids	4.122	2.699	11.225	2.990	4.296	2.461	5.399
Tree-SHAR (9 leaves)	4.027	2.493	11.041	3.031	4.305	2.532	5.324
Bagged-SHAR (9 leaves)	4.006	2.459	10.992	3.003	4.270	2.480	5.291
Boosted-SHAR (9 leaves)	**3.844**	**2.270**	**10.730**	**2.824**	**4.150**	**2.397**	**5.133**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models. The daily volatility surface is fitted using the SVI method as the first step of the forecasting procedure described in Section 2.1. The forecasting performance is evaluated across three different horizons (1 day, 5 days, and 22 days) and decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table A.3: Pairwise Diebold-Mariano Test Statistics and Model Confidence Set (MCS) for SHAR (SVI)

	RW	SHAR	SHAR (3×3)	Tree-SHAR	Bagged-SHAR	Boosting-SHAR	MCS
			$\underline{h = 1}$				
Random Walk		15.50	37.55	40.03	44.31	52.61	
SHAR	-15.50		38.88	39.42	45.27	52.12	
SHAR with 3x3 grids	-37.55	-38.88		10.65	16.27	36.23	
Tree-SHAR (6 leaves)	-40.03	-39.42	-10.65		24.83	31.52	
Bagged-SHAR (6 leaves)	-44.31	-45.27	-16.27	-24.83		28.47	
Boosted-SHAR (6 leaves)	-52.61	-52.12	-36.23	-31.52	-28.47		✓
			$\underline{h = 5}$				
Random Walk		-5.35	17.71	10.58	11.90	34.53	
SHAR	5.35		24.03	16.21	18.19	40.48	
SHAR with 3x3 grids	-17.71	-24.03		-8.80	-8.02	29.51	
Tree-SHAR (4 leaves)	-10.58	-16.21	8.80		4.97	36.63	
Bagged-SHAR (4 leaves)	-11.90	-18.19	8.02	-4.97		37.06	
Boosted-SHAR (4 leaves)	-34.53	-40.48	-29.51	-36.63	-37.06		✓
			$\underline{h = 22}$				
Random Walk		54.71	82.75	80.65	85.61	89.36	
SHAR	-54.71		65.48	63.13	72.95	79.78	
SHAR with 3x3 grids	-82.75	-65.48		22.04	34.41	56.44	
Tree-SHAR (9 leaves)	-80.65	-63.13	-22.04		33.46	52.06	
Bagged-SHAR (9 leaves)	-85.61	-72.95	-34.41	-33.46		47.23	
Boosted-SHAR (9 leaves)	-89.36	-79.78	-56.44	-52.06	-47.23		✓

Notes: This table reports the pairwise Diebold-Mariano (DM) test statistics comparing the out-of-sample forecasting errors of the competing models across 1, 5, and 22-day horizons. The daily volatility surface is fitted using the SVI method as the first step of the forecasting procedure described in Section 2.1. A negative t-statistic indicates that the model in the row outperforms (has lower error than) the model in the column, while a positive value indicates that the column model outperforms the row model. Statistical significance at the 1% level corresponds to an absolute t-statistic greater than 2.33 ($|t| > 2.33$). The table also indicates with a checkmark the models that are included in the Model Confidence Set (MCS) at the 95% confidence level.

Table A.4: Out-of-Sample Implied Volatility Forecasting Performance by Moneyness (RMSE in %, $h = 5$)

	2018	2019	2020	2021	2022	2023	Total
			Moneyness $\leq q_{20}$				
Random Walk	3.524	2.456	5.714	2.937	3.114	2.042	3.468
SHAR	3.448	2.387	5.717	2.825	2.991	2.002	3.407
SHAR with 3x3 grid	3.122	2.048	5.309	2.551	2.811	1.733	3.117
Tree-SHAR	2.212	1.531	4.784	2.033	2.184	1.735	2.603
Bagged-SHAR	2.199	1.501	4.782	2.025	2.165	1.685	2.588
Boosted-SHAR	**1.889**	**1.364**	**4.414**	**1.825**	**1.969**	**1.604**	**2.362**
			$q_{20} <$ Moneyness $< q_{80}$				
Random Walk	3.294	2.254	**6.569**	2.611	2.737	1.709	3.493
SHAR	3.218	2.207	6.754	2.556	**2.664**	1.761	3.515
SHAR with 3x3 grid	3.198	2.087	6.734	2.501	2.791	1.699	3.504
Tree-SHAR	3.180	2.024	6.713	2.516	2.838	1.661	3.496
Bagged-SHAR	3.169	2.010	6.693	2.499	2.817	1.630	**3.479**
Boosted-SHAR	**3.114**	**1.900**	6.818	**2.440**	2.796	**1.585**	3.484
			Moneyness $\geq q_{80}$				
Random Walk	4.106	2.232	6.463	3.393	3.290	2.591	3.883
SHAR	4.039	2.048	6.661	3.401	3.306	2.566	3.907
SHAR with 3x3 grid	3.893	1.863	6.700	3.205	3.190	2.454	3.819
Tree-SHAR	3.469	1.781	6.494	2.786	2.891	2.099	3.544
Bagged-SHAR	3.463	1.765	6.449	2.774	2.863	2.070	3.521
Boosted-SHAR	**3.149**	**1.455**	**6.283**	**2.528**	**2.652**	**1.891**	**3.318**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of five-day horizon ($h = 5$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three moneyness groups. On each date, three groups are classified such that the moneyness is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table A.5: Out-of-Sample Implied Volatility Forecasting Performance by Maturity (RMSE in %, $h = 5$)

	2018	2019	2020	2021	2022	2023	Total
			Maturity $\leq q_{20}$				
Random Walk	4.202	2.757	7.530	3.417	3.621	2.403	4.302
SHAR	4.148	2.695	7.766	3.530	3.643	2.520	4.381
SHAR with 3x3 grid	4.046	2.489	7.645	3.324	3.574	2.327	4.257
Tree-SHAR	3.712	2.218	7.364	2.972	3.351	2.088	3.998
Bagged-SHAR	3.693	2.192	**7.330**	2.948	3.322	2.039	3.972
Boosted-SHAR	**3.454**	**2.086**	7.456	**2.798**	**3.215**	**1.961**	**3.921**
			$q_{20} <$ Maturity $< q_{80}$				
Random Walk	3.440	2.149	6.280	2.714	2.876	1.882	3.453
SHAR	3.358	2.056	6.409	2.534	2.736	1.815	3.413
SHAR with 3x3 grid	3.285	1.972	6.281	2.541	2.835	1.796	3.374
Tree-SHAR	3.004	1.872	6.243	2.379	2.643	1.722	3.252
Bagged-SHAR	2.998	1.855	6.224	2.369	2.624	1.684	3.236
Boosted-SHAR	**2.909**	**1.682**	**6.166**	**2.292**	**2.574**	**1.631**	**3.169**
			Maturity $\geq q_{80}$				
Random Walk	2.690	1.913	4.830	2.331	1.995	1.618	2.731
SHAR	2.599	1.834	4.899	2.250	1.947	1.626	2.709
SHAR with 3x3 grid	2.365	1.462	5.027	1.939	1.874	1.417	2.603
Tree-SHAR	2.250	1.403	4.893	1.998	1.958	1.445	2.563
Bagged-SHAR	2.243	1.400	4.884	1.988	1.937	1.442	2.555
Boosted-SHAR	**2.105**	**1.211**	**4.721**	**1.801**	**1.772**	**1.267**	**2.408**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of five-day horizon ($h = 5$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three maturity groups. On each date, three groups are classified such that the maturity is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table A.6: Out-of-Sample Implied Volatility Forecasting Performance by Moneyness (RMSE in %, $h = 22$)

	2018	2019	2020	2021	2022	2023	Total
			Moneyness $\leq q_{20}$				
Random Walk	5.339	3.809	11.801	3.541	5.519	3.082	6.158
SHAR	4.846	3.526	11.009	3.223	4.797	3.354	5.693
SHAR with 3x3 grid	4.009	2.791	9.133	2.547	4.245	**2.724**	4.740
Tree-SHAR	2.725	1.878	7.569	2.278	3.167	2.857	3.826
Bagged-SHAR	2.712	1.852	7.537	2.258	3.088	2.805	3.792
Boosted-SHAR	**2.340**	**1.752**	**6.543**	**2.232**	**2.734**	2.775	**3.379**
			$q_{20} <$ Moneyness $< q_{80}$				
Random Walk	4.799	3.592	13.196	3.299	4.844	2.579	6.295
SHAR	4.386	3.259	12.512	3.086	**4.242**	2.709	5.895
SHAR with 3x3 grid	4.327	2.907	12.164	3.082	4.750	2.422	5.810
Tree-SHAR	4.255	2.629	11.921	2.955	4.718	2.298	5.680
Bagged-SHAR	4.244	2.609	11.868	2.907	4.683	2.255	5.647
Boosted-SHAR	**4.158**	**2.512**	**11.796**	**2.894**	4.659	**2.186**	**5.597**
			Moneyness $\geq q_{80}$				
Random Walk	5.273	2.845	12.519	4.039	4.826	2.938	6.197
SHAR	5.144	2.478	12.322	4.733	4.694	3.054	6.162
SHAR with 3x3 grid	4.828	2.368	11.935	4.076	4.364	2.814	5.845
Tree-SHAR	4.333	2.387	11.591	3.509	4.022	2.654	5.547
Bagged-SHAR	4.286	2.276	11.510	3.404	3.946	2.567	5.479
Boosted-SHAR	**3.869**	**1.880**	**10.982**	**3.087**	**3.664**	**2.407**	**5.147**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of 22-day horizon ($h = 22$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three moneyness groups. On each date, three groups are classified such that the moneyness is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

Table A.7: Out-of-Sample Implied Volatility Forecasting Performance by Maturity (RMSE in %, $h = 22$)

	2018	2019	2020	2021	2022	2023	Total
			Maturity $\leq q_{20}$				
Random Walk	5.878	4.058	15.524	4.073	6.084	3.195	7.613
SHAR	5.588	3.797	14.966	4.515	5.667	3.589	7.378
SHAR with 3x3 grid	5.291	3.262	13.950	4.014	5.590	3.021	6.875
Tree-SHAR	4.803	2.832	13.255	3.453	5.261	2.745	6.435
Bagged-SHAR	4.785	2.823	13.124	3.366	5.188	**2.678**	6.366
Boosted-SHAR	**4.455**	**2.663**	**12.508**	**3.175**	**5.008**	2.721	**6.070**
			$q_{20} <$ Maturity $< q_{80}$				
Random Walk	5.032	3.464	12.247	3.397	4.988	2.660	6.011
SHAR	4.571	3.048	11.542	3.080	4.249	2.677	5.548
SHAR with 3x3 grid	4.315	2.807	11.060	3.031	4.561	2.486	5.381
Tree-SHAR	3.977	2.472	10.637	2.890	4.229	2.447	5.109
Bagged-SHAR	3.951	2.407	10.622	**2.837**	4.181	2.397	5.079
Boosted-SHAR	**3.792**	**2.266**	**10.497**	2.838	**4.107**	**2.313**	**4.991**
			Maturity $\geq q_{80}$				
Random Walk	3.695	2.760	9.601	2.966	3.231	2.391	4.687
SHAR	3.388	2.516	9.085	2.838	2.954	2.488	4.420
SHAR with 3x3 grid	3.086	1.980	8.873	2.379	2.980	2.116	4.198
Tree-SHAR	2.856	1.789	8.726	2.392	2.974	2.275	4.123
Bagged-SHAR	2.851	1.758	8.720	2.372	2.966	2.223	4.111
Boosted-SHAR	**2.815**	**1.586**	**8.638**	**2.266**	**2.844**	**1.970**	**4.022**

Notes: This table reports the out-of-sample Root Mean Squared Error (RMSE) in % of 22-day horizon ($h = 22$) implied volatility forecasts for the Random Walk, SHAR, SHAR with 3×3 grids, Tree–SHAR, Bagged–SHAR, and Boosted–SHAR models by three maturity groups. On each date, three groups are classified such that the maturity is less than or equal to its 20th percentile, between the 20th and 80th percentiles, and greater than or equal to the 80th percentile. The forecasting performance is also decomposed by year to identify the specific periods where improvements occurred. Boldface values represent the lowest RMSEs among the six forecasting models.

References

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.

Athey, Susan, Julie Tibshirani, and Stefan Wager, 2019, Generalized random forests, *Annals of Statistics* 47, 1148–1178.

Audrino, Francesco, and Peter Bühlmann, 2001, Tree-structured GARCH models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 727–744.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical Performance of Alternative Option Pricing Models, *Journal of Finance* 52, 2003–2049.

Bakshi, Gurdip, Dilip Madan, and George Panayotov, 2010, Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels, *Journal of Financial Economics* 97, 130–154.

Beason, Tyler, and David Schreindorfer, 2022, Dissecting the Equity Premium, *Journal of Political Economy* 130, 2203–2222.

Black, Fischer, and Myron Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637–654.

Bollen, Nicolas PB, and Robert E Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.

Breiman, Leo, 1996, Bagging Predictors, *Machine Learning* 24, 123–140.

Breiman, Leo, 2001, Random Forests, *Machine Learning* 45, 5–32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone, 1984, *Classification and Regression Trees*. (Wadsworth International Group).

Breiman, Leo, Jerome H Friedman, Richard A Olshen, and Charles J Stone, 2017, *Classification and regression trees*. (Routledge).

Brownlees, Christian, and Gustavo Souza, 2025, How to bet on winners and losers, *Working Paper*.

Buraschi, Andrea, and Alexei Jiltsov, 2006, Model Uncertainty and Option Markets with Heterogeneous Beliefs, *Journal of Finance* 61, 2841–2897.

Chen, Ying, Maria Grith, and Hannah LH Lai, 2025, Neural tangent kernel in implied volatility forecasting: a nonlinear functional autoregression approach, *Journal of Business & Economic Statistics* pp. 1–15.

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.

Constantinides, George M., Jens Carsten Jackwerth, and Stylianos Perrakis, 2009, Mispricing of S&P 500 Index Options, *The Review of Financial Studies* 22, 1247–1277.

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

Diebold, Francis X, and Roberto S Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Dufays, Arnaud, Kris Jacobs, Yuguo Liu, and Jeroen Rombouts, 2023, Fast Filtering with Large Option Panels: Implications for Asset Pricing, *Journal of Financial and Quantitative Analysis* pp. 1–56.

Dufays, Arnaud, Kris Jacobs, and Jeroen Rombouts, 2025, A Framework for Real-Time Modeling and Forecasting of Large Unbalanced Option Implied Volatility Surfaces, Working paper, University of Houston.

Dumas, Bernard, Jeff Fleming, and Robert E. Whaley, 1998, Implied Volatility Functions: Empirical Tests, *Journal of Finance* 53, 2059–2106.

Eaton, Gregory W, T Clifton Green, Brian S Roseman, and Yanbin Wu, 2026, Retail option traders and the implied volatility surface, *Journal of Financial Economics* 177, 104238.

Fan, Jianqing, Marianne Farmen, and Irene Gijbels, 1998, Local maximum likelihood estimation and inference, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60, 591–608.

Fan, Jianqing, and Loriano Mancini, 2009, Option Pricing With Model-Guided Nonparametric Methods, *Journal of the American Statistical Association* 104, 1351–1372.

Fan, Jianqing, Yichao Wu, and Yang Feng, 2009, Local quasi-likelihood with a parametric guide, *Annals of Statistics* 37, 4153–4183.

Friedman, Jerome H., 2001, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics* 29, 1189–1232.

Gârleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Poteshman, 2009, Demand-Based Option Pricing, *Review of Financial Studies* 22, 4259–4299.

Gatheral, Jim, 2004, A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives, *Presentation at Global Derivatives & Risk Management, Madrid* p. 0.

Goulet Coulombe, Philippe, 2024, The macroeconomy as a random forest, *Journal of Applied Econometrics* 39, 452–476.

Hansen, Peter R., Asger Lunde, and James M. Nason, 2011, The Model Confidence Set, *Econometrica* 79, 453–497.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer) 2 edn.

Heston, Steven L., 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–343.

Hurn, A. Stan, Kenneth A. Lindsay, and Andrew J. McClelland, 2015, Estimating the Parameters of Stochastic Volatility Models Using Option Price Data, *Journal of Business & Economic Statistics* 33, 579–594.

Kim, Hyung Joo, 2025, Characterizing the conditional pricing kernel: A new approach, Working paper.

Kim, Hyung Joo, and Dong Hwan Oh, 2025, Local Estimation for Option Pricing: Improving Forecasts with Market State Information, Working 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.

Oh, Dong Hwan, and Andrew J. Patton, 2024, Better the devil you know: Improved forecasts from imperfect models, *Journal of Econometrics* 242, 105767.

Oh, Dong Hwan, and Andrew J. Patton, 2026, Skill and Efficiency in the U.S. Mutual Fund Industry, Working paper.

Pan, Jun, 2002, The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, *Journal of Financial Economics* 63, 3–50.

Patton, Andrew J., and Yasin Simsek, 2026, Generalized Autoregressive Score Trees and Forests, *Journal of Business & Economic Statistics* Forthcoming.

Schlosser, Lisa, Torsten Hothorn, Reto Stauffer, and Achim Zeileis, 2019, Distributional regression forests for probabilistic precipitation forecasting in complex terrain, *Annals of Applied Statistics* 13, 1564–1589.

Schreindorfer, David, and Tobias Sichert, 2025, Conditional risk and the pricing kernel, *Journal of Financial Economics, Forthcoming.*

Tibshirani, Robert, and Trevor Hastie, 1987, Local likelihood estimation, *Journal of the American Statistical Association* 82, 559–567.