Université du Québec en Outaouais

Département des sciences administratives

Working Paper No. 7

The Options-Implied Information Content for Cross-Asset Return and Volatility Prediction: Evidence from 3.8 Billion Option Contracts

Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

October 2026

Version 1.2

Abstract

This paper measures the information content of the implied volatility surface for returns and realized variance in a panel of 69 liquid US underlyings—49 large stocks, 17 ETFs and three equity indices—over 2010–2025, built from 3.83 billion end-of-day option records and 5-minute prices. Option-implied features predict the cross-section of weekly returns in a statistically detectable but economically marginal way: they explain 0.3% of the variation in 5-day stock returns, 0.6% for ETFs and 1.0% for indices; under two-way clustered standard errors only the 25-delta skew remains significant, with a positive sign consistent with compensation for skewness risk; and of fourteen long-short strategies implemented in calendar time with measured turnover, only a sort on the implied volatility level is profitable after a 10 basis-point cost, with a t-statistic of 2.2 that does not survive a multiple-testing correction. The surface is more informative about second moments. On a common sample it raises the in-sample R^2 of the HAR-RV model from 0.415 to 0.457 for next-day realized variance and from 0.421 to 0.515 for weekly realized variance, in every one of nine subperiods; out of sample the gain is reliable under mean squared error but not under QLIKE, and at the weekly horizon the level of implied volatility alone is the best forecaster under both losses. The ratio of implied to realized correlation among S&P 500 constituents predicts stress episodes at 5- to 20-day horizons beyond the VIX (t of 2.3–4.1 with overlap-robust standard errors). At-the-money implied volatility Granger-causes realized variance in every ticker, but implied-volatility shocks account for 20–29% of its 20-day forecast-error variance. Open-interest concentration does not attract the underlying price, and tree-based learners trained on the SPX surface do not beat HAR-RV out of sample. The revision corrects a target-alignment error in the previous version, whose weekly return results are withdrawn.

Keywords: option-implied information, implied volatility surface, realized variance, HAR-RV, implied correlation, portfolio sorts, transaction costs, Granger causality

JEL Classification: G12, G13, G14, G17, C53, C58

Contents

1 Introduction **3**

2 Related Literature **5**
2.1 Where Informed Traders Trade 5
2.2 Option-Implied Moments and the Cross-Section of Returns 5
2.3 Implied Volatility and Realized-Variance Forecasting 6
2.4 Implied Correlation and Market Stress 7
2.5 Greeks, Open Interest and the Underlying Price 7

3 Data and Variable Construction **7**
3.1 Data Sources 7
3.2 Analysis Sample 8
3.3 Option-Implied Features 9
3.4 Realized Measures and Forward Targets 9
3.5 Descriptive Statistics 10

4 Methodology **12**
4.1 Return Predictability (RQ1) 12
4.2 Realized-Variance Forecasting (RQ2) 13
4.3 Implied Correlation and Market Stress (RQ3) 14
4.4 Greeks by Time to Expiration and Price Magnets (RQ4) 15
4.5 Machine Learning on the SPX Surface (RQ5) 15
4.6 Granger Causality and Vector Autoregressions 15

5 Results **16**
5.1 RQ1: Return Predictability 16
5.1.1 Panel Regressions 16
5.1.2 Portfolio Sorts 18
5.1.3 Economic Significance After Trading Costs 20
5.2 RQ2: The Implied Volatility Surface and Realized-Variance Forecasting 22
5.3 RQ3: Implied Correlation and Market Stress 23
5.4 RQ4: Greeks by Time to Expiration and Price Magnets 25
5.5 RQ5: Machine Learning on the SPX Surface 26
5.6 Granger Causality and Vector Autoregressions 27

6 Robustness **29**
6.1 Standard-Error Specification . . . 29
6.2 Subperiod Stability . . . 30
6.3 VIX Regimes . . . 31
6.4 Additional Controls, Quantile Regressions and Ticker-Level Fit . . . 32
6.5 Winsorization and HAC-Lag Sensitivity . . . 33
6.6 Rank-Based Information Coefficients . . . 34
6.7 Decile Sorts . . . 35
6.8 Temporal Stability and a Placebo Test . . . 36

7 Discussion **37**
7.1 Economic Interpretation . . . 37
7.2 Relation to the Literature on Option-Implied Predictability . . . 39
7.3 Limitations . . . 40

8 Conclusion **40**

A Analysis Universe **49**

B Implied Correlation Around Market Events **49**

C Per-Ticker Out-of-Sample Comparisons (RQ2) **50**

D SPX Options Surface Features (RQ5) **50**

E Random Forest Feature Importance (RQ5) **51**

1. Introduction

Option prices aggregate the beliefs of heterogeneous investors about the distribution of future returns. Since Black and Scholes (1973), the implied volatility surface—the map from moneyness and time to expiration to Black–Scholes implied volatility—has been read as a forward-looking summary of that distribution: its level measures expected variance, its slope across strikes measures the asymmetry of perceived risks, and its curvature measures the weight of the tails (Bakshi et al., 2003). Whether this information predicts returns and realized volatility, and to what extent it is already embedded in the underlying asset's price, remain central questions in empirical finance; Christoffersen et al. (2013) survey a large literature with mixed answers.

This paper revisits those questions for the most liquid segment of the US market. From a database of 3.83 billion end-of-day option records on 11,077 underlyings and 11.5 billion intraday price bars, I build a daily panel of option-implied features and 5-minute realized variances for 69 underlyings—49 large-capitalization stocks, 17 exchange-traded funds and three equity indices—over January 2010 to December 2025 (264,383 ticker-days). The panel is used to answer five questions.

RQ1. Do option-implied moments extracted from the strike–expiry surface predict next-day and next-week returns in the cross-section of stocks, ETFs and indices, and is the predictability exploitable after trading costs?

RQ2. Does the implied volatility term structure and smile improve realized-variance forecasts relative to the heterogeneous autoregressive model of Corsi (2009), in sample and out of sample?

RQ3. Does the gap between option-implied and realized correlation among S&P 500 constituents predict episodes of market stress?

RQ4. How does the information content of aggregated option Greeks vary with time to expiration, and does open-interest concentration at a strike attract the underlying price?

RQ5. Can machine-learning models trained on the full daily SPX surface outperform the VIX and HAR-RV in forecasting the next-day realized variance of the S&P 500?

Three features distinguish the study. First, the analysis spans three asset groups with a common methodology, so that the cross-asset gradient in predictability is measured rather than assumed. Second, economic significance is evaluated with a calendar-time implementation of the long-short portfolios that measures realized turnover and reports performance net of proportional trading costs, following Jegadeesh and Titman (1993) and Novy-Marx and Velikov (2016). Third, inference is deliberately conservative: overlapping multi-day targets are treated with Newey–West standard errors (Newey and West, 1987; Hansen and Hodrick, 1980), panel regressions are also estimated with two-way clustering (Cameron et al., 2011; Petersen, 2009) and by the method of Fama and MacBeth (1973), and the predictors are confronted with the multiple-testing hurdle of Harvey et

al. (2016). Lead–lag relations between implied and realized measures are examined with Granger causality tests and a bivariate vector autoregression whose variance decomposition is reported under both recursive orderings.

The results are more sober than the ones usually associated with option-implied predictors. Across the 69 underlyings the ten option-implied predictors explain less than one percent of the variation in 5-day returns for stocks and about one percent for indices; only the 25-delta skew retains a t-statistic above two under two-way clustering, and no predictor clears the multiple-testing threshold under every inference method. Long-short quintile portfolios earn gross Sharpe ratios between -0.5 and 0.6; the single strategy that remains profitable after a 10 basis-point one-way cost, a sort on the level of at-the-money implied volatility, has a t-statistic of 2.2 and would not survive a correction for the fourteen sorts examined. The results for volatility are stronger. Adding the surface to HAR-RV raises the in-sample R^2 for next-day realized variance from 0.415 to 0.457 and for 5-day realized variance from 0.421 to 0.515 on a common sample, with a positive improvement in each of nine subperiods; out of sample, the augmented model lowers mean squared error for 14 of 30 tickers at the 5% level and never raises it significantly, but it is worse under the QLIKE loss for half of the tickers, so the forecasting gain is loss-function dependent. The ratio of implied to realized correlation predicts stress episodes at horizons of 5 to 20 days (t-statistics of 2.3 to 4.1 with overlap-robust standard errors). At-the-money implied volatility Granger-causes realized variance in every ticker, but the orthogonalized variance decomposition attributes 20 to 29 percent of the 20-day forecast-error variance of realized variance to implied-volatility shocks, depending on the ordering, so realized variance remains driven mostly by its own innovations. The price-magnet hypothesis is rejected, and tree-based learners trained on the SPX surface do not beat HAR-RV out of sample.

Relation to the previous version. Version 1.1 of this paper reported weekly return R^2 values of 5–19%, a long-short Sharpe ratio of 2.33 for a sort on implied kurtosis, and an R^2_{OOS} of 0.99 for 22-day SPX realized variance. A review of the code established that the multi-day targets had been constructed with a trailing instead of a forward window, so that the "5-day forward return" at date t was in fact $r_{t-3} + \cdots + r_{t+1}$ and overlapped the contemporaneous option features. All multi-day results have been recomputed with correctly aligned targets; the headline weekly findings of version 1.1 were artefacts of that error and are withdrawn. Section 3 documents the construction and the companion change log lists every affected number.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 describes the data and the construction of the variables. Section 4 presents the econometric methods. Section 5 reports the results for the five questions, Section 6 the robustness analyses, and Section 7 discusses the findings and their limitations. Section 8 concludes.

2. Related Literature

2.1. Where Informed Traders Trade

The idea that options markets may lead the stock market goes back to Black (1975) and Manaster and Rendleman (1982). Easley et al. (1998) formalize it in a sequential-trade model in which informed investors choose between stocks and options according to leverage and liquidity, and show that signed option volume predicts stock returns. Pan and Poteshman (2006) find that the put–call volume ratio of non-market-maker trades predicts returns for up to two weeks, with the information concentrated in out-of-the-money puts, and Cao et al. (2005) document that option volume is informative before takeover announcements. Johnson and So (2012) show that the option-to-stock volume ratio predicts returns most strongly when short-sale constraints bind, and Ge et al. (2016) trace the effect to the direction of option order flow. Ni et al. (2008) show that option trades also carry information about future volatility, and Hu (2014) documents that the price impact of option order flow has increased over time.

The evidence on price discovery is less favourable to options. Chakravarty et al. (2004) attribute roughly one sixth of price discovery to the option market, and Muravyev et al. (2013) find that option quotes do not contain economically significant information about future stock prices once the stock's own quotes are accounted for. Muravyev et al. (2025) argue that a large part of the return predictability attributed to option-implied measures reflects temporary price pressure in the stock that is impounded in option quotes, and that the predictability is concentrated in less liquid stocks. These findings matter for the present study, whose universe consists of the most liquid names in the market: if option-implied return predictability is a limits-to-arbitrage phenomenon, it should be weakest precisely in this segment.

2.2. Option-Implied Moments and the Cross-Section of Returns

Bakshi and Madan (2000) and Bakshi et al. (2003) show how risk-neutral variance, skewness and kurtosis can be recovered from option prices without a parametric model, and Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) develop the model-free implied variance that underlies the VIX. Dennis and Mayhew (2002) document that risk-neutral skewness of individual stocks relates to firm characteristics, and a large literature asks whether implied moments predict returns. Xing et al. (2010) find that a steeper volatility smirk predicts lower future returns; Cremers and Weinbaum (2010) show that deviations from put–call parity predict returns; Bali and Hovakimian (2009) relate call–put implied volatility spreads to expected returns; Rehman and Vilkov (2012) and Stilger et al. (2017) find that more negative risk-neutral skewness predicts lower returns, the latter attributing the effect to overpricing that is costly to arbitrage; Conrad et al. (2013) report the opposite sign for ex ante skewness over longer horizons; An et al. (2014) analyse the joint

cross-section of stocks and options and find that increases in call (put) implied volatility predict higher (lower) returns; Bali and Murray (2013) show that risk-neutral skewness also prices option portfolios; and Bali et al. (2019) assemble a comprehensive set of implied moments. The term structure of the surface has received separate attention: Mixon (2007) and Vasquez (2017) study the slope of the implied volatility term structure, and Borochin et al. (2020) the term structure of risk-neutral skewness. Martin (2017) and Martin and Wagner (2019) derive lower bounds on expected returns from option prices under weak assumptions, offering a theoretical rationale for a positive relation between implied variance and expected returns.

Taken together these studies show statistically reliable but modest predictability, typically at the weekly to monthly horizon and typically strongest among small, illiquid or hard-to-short stocks (Johnson and So, 2012; Stilger et al., 2017). Two general lessons from the anomalies literature also apply. First, predictability measured with overlapping multi-period returns requires overlap-robust inference (Hansen and Hodrick, 1980), and predictability measured across many candidate predictors requires a multiple-testing adjustment (Harvey et al., 2016; Chordia et al., 2020). Second, anomalies shrink or disappear once trading costs are accounted for (Novy-Marx and Velikov, 2016) or once they are replicated on liquid stocks with careful procedures (Hou et al., 2020; McLean and Pontiff, 2016).

2.3. Implied Volatility and Realized-Variance Forecasting

Christensen and Prabhala (1998) establish that implied volatility is a biased but informationally efficient forecast of subsequent realized volatility, a result Blair et al. (2001) confirm for the VIX using high-frequency realized measures and Taylor et al. (2010) extend to options on individual stocks. The realized-variance literature initiated by Andersen and Bollerslev (1998) and Andersen et al. (2001, 2003) provides the forecasting target; Liu et al. (2015) show that the 5-minute realized variance used here is hard to beat among realized measures. The heterogeneous autoregressive model of Corsi (2009) is the standard benchmark; Busch et al. (2011) find that implied volatility adds to HAR-RV for equity, bond and foreign-exchange markets, Prokopczuk and Wese Simen (2014) that correcting implied volatility for the variance risk premium matters for its forecasting value, and Kourtis et al. (2016) that implied volatility forecasts compare favourably with GARCH forecasts across international index markets. Bekaert and Hoerova (2014) decompose the squared VIX into expected variance and a variance risk premium; Bollerslev et al. (2009) and Carr and Wu (2009) study the variance risk premium itself. Refinements of HAR include the semivariance-based specifications of Patton and Sheppard (2015), the measurement-error-adjusted HARQ of Bollerslev et al. (2016) and the Realized GARCH of Hansen et al. (2012). On the evaluation side, Hansen and Lunde (2005) show how difficult it is to beat simple volatility models, Patton (2011) establishes that MSE and QLIKE are the loss functions robust to noise in the realized proxy, Diebold and Mariano

(1995) and Giacomini and White (2006) provide tests of equal predictive accuracy, Clark and West (2007) the adjustment needed for nested models, and Hansen et al. (2011) the model confidence set. Poon and Granger (2003) survey the earlier literature. Machine-learning approaches are evaluated by Bucci (2020) and Christensen et al. (2023), who find that nonlinear learners offer gains that are real but modest and that depend on careful tuning; Gu et al. (2020) report larger gains for cross-sectional return prediction.

2.4. Implied Correlation and Market Stress

Skintzi and Refenes (2005) propose an implied correlation index as a measure of diversification, Driessen et al. (2009) show that the correlation risk premium is large and positive, and Buss and Vilkov (2012) use option-implied correlations to improve covariance estimates. Pollet and Wilson (2010) find that average realized correlation predicts market returns, Buraschi et al. (2014) provide an equilibrium model in which disagreement drives comovement and the variance risk premium, and Faria et al. (2022) document the correlation risk premium internationally. Related measures of tail and systemic risk are developed by Kelly and Jiang (2014), Bollerslev and Todorov (2011) and Acharya et al. (2017).

2.5. Greeks, Open Interest and the Underlying Price

Ni et al. (2005) document that stock prices cluster at option strike prices on expiration dates and attribute the effect to delta hedging and manipulation, Avellaneda and Lipkin (2003) provide a hedging-based mechanism for pinning, and Golez and Jackwerth (2012) find pinning in S&P 500 futures. Ni et al. (2021) show that option-trading-induced hedging has a pervasive effect on stock price dynamics, and Baltussen et al. (2021) and Barbon and Buraschi (2021) relate aggregate dealer gamma exposure to intraday momentum and to the fragility of the market. These studies motivate the question of whether open-interest concentration attracts the underlying on ordinary trading days rather than only at expiration.

3. Data and Variable Construction

3.1. Data Sources

Options. The option data are end-of-day records for 11,077 US underlyings from January 4, 2010 to December 31, 2025 (4,025 trading days), 3,831,907,488 contract-days in total. Each record contains the trade date, strike, expiration date, call/put flag, bid and ask prices, bid and ask implied volatilities, open interest, volume and the Black–Scholes sensitivities δ, Γ, $\mathcal{V}$, Θ and ρ supplied by the vendor. Between 92 and 95 percent of records in a given year carry valid Greeks.

Intraday prices. Realized measures are computed from 5-minute OHLCV bars for 7,789 US equities, 4,302 ETFs, 125 indices, 131 futures, 78 currency pairs and 74 cryptocurrencies—about 11.5 billion bars (Table 1). Only the equity, ETF and index stores enter the analysis below; the futures, currency and cryptocurrency series are part of the same data lake but no option chains are available for them, so the "cross-asset" dimension of this paper is limited to the three groups with listed options.

Table 1. Data coverage.

Asset class	Symbols	Rows (millions)	Freq.	Period
Panel A: Options				
Option contract-days	11,077	3,832	Daily	2010–2025
Panel B: Intraday OHLCV				
US equities	7,789	1,225	5-min	2000–2026
ETFs	4,302	301	5-min	2000–2026
Equity indices	125	40	5-min	2008–2026
Futures	131	73	5-min	2008–2026
FX	78	83	5-min	2010–2026
Crypto	74	39	5-min	2013–2026
Total		**≈11,500**		

Notes. Row counts of the raw stores (frozen inventory of the extraction step). Only the option store and the equity, ETF and index bars are used in the analysis.

3.2. Analysis Sample

The analysis panel consists of 69 underlyings with both liquid option chains and complete intraday coverage: 49 large-capitalization stocks (the constituents of a fixed list drawn from the largest US companies by market capitalization, listed in Appendix A), 17 ETFs (broad market, sector and bond or gold funds) and the three cash indices with listed options (SPX, NDX, RUT). The list is fixed over the sample and was chosen for liquidity rather than by a point-in-time screen, so it is subject to a mild survivorship bias: all 49 stocks are large companies in 2025, and several joined the option universe late (META in July 2021 after its ticker change, GOOGL in 2014, AVGO in 2016, RTX in 2018). The merge of the daily option features with the realized measures yields 264,383 ticker-days (188,093 for stocks, 64,256 for ETFs and 12,034 for indices). Because the moneyness-specific features require quotes at several deltas and maturities, the regressions that use all ten option-implied predictors are estimated on the 119,081 ticker-days (79,943 for stocks) on which every predictor is observed; the N reported with each table is the relevant estimation sample.

3.3. Option-Implied Features

Table 2 defines the ten daily features computed from each underlying's option chain, using contracts with 7 to 180 days to expiration (DTE) and a positive ask price. Implied volatilities are the mid of the bid and ask implied volatilities. The 30-day at-the-money implied volatility $IV_{ATM,30d}$ is the average mid-implied volatility of calls with $|\delta - 0.5| < 0.10$ and 20 to 40 DTE; the 90-day analogue uses 80 to 100 DTE. The skew measures compare out-of-the-money puts and calls at the 25-delta and 10-delta points. The "implied skewness" and "implied kurtosis" labels follow version 1.1 of the paper but should be read as smile-based proxies: the former is the 10-delta put–call implied volatility difference scaled by the at-the-money level, and the latter the ratio of the average wing implied volatility ($0.03 < |\delta| < 0.15$) to the at-the-money level. They are monotone in, but not equal to, the model-free risk-neutral moments of Bakshi et al. (2003), which require integrating option prices over a dense strike grid at a fixed maturity; the proxies have the advantage of being computable for every ticker-day with a handful of quotes. Put–call ratios and the net gamma exposure are aggregated over the whole chain. The gamma exposure has a heavy-tailed distribution (a few ticker-days with corrupted vendor Greeks produce values of order 10^{40}), which is why every predictor is winsorized before estimation.

Table 2. Option-implied variables.

Variable	Definition
$IV_{ATM,30d}$	Average mid-IV of calls with $\|\delta - 0.5\| < 0.10$, 20–40 DTE
$IV_{ATM,90d}$	Same, 80–100 DTE
IV term slope	$IV_{ATM,90d} - IV_{ATM,30d}$
$\text{Skew}_{25\delta}$	$IV_{25\delta P,30d} - IV_{25\delta C,30d}$, puts with $\|\delta + 0.25\| < 0.07$, calls with $\|\delta - 0.25\| < 0.07$
Implied skewness (proxy)	$(IV_{10\delta P,30d} - IV_{10\delta C,30d})/IV_{ATM,30d}$
Implied kurtosis (proxy)	$\overline{IV}_{wings,30d}/IV_{ATM,30d}$, wings defined by $0.03 < \|\delta\| < 0.15$
PC volume ratio	$\sum V^{put}/\sum V^{call}$ over the chain
PC OI ratio	$\sum OI^{put}/\sum OI^{call}$ over the chain
Net gamma exposure	$\sum \Gamma^{call} \cdot OI^{call} - \sum \Gamma^{put} \cdot OI^{put}$

Notes. All features are computed at the close of day t from contracts with 7–180 days to expiration and a positive ask. The "30d" maturity bucket is 20–40 DTE. Return and variance regressions add RV_t and $RV_t^{(w)}$ as the ninth and tenth predictors.

3.4. Realized Measures and Forward Targets

Following Andersen et al. (2003), the daily realized variance of underlying i is the sum of squared 5-minute log returns, the first return of the day being measured from the previous day's last bar so

that the overnight return is included:

$$RV_{i,t} = \sum_{k=1}^{N_{i,t}} r_{i,t,k}^2, \qquad r_{i,t,k} = \ln\!\left(P_{i,t,k} \,/\, P_{i,t,k-1}\right), \tag{1}$$

where $N_{i,t}$ is the number of bars (days with fewer than 20 bars are dropped). The 5-minute frequency is the one Liu et al. (2015) find hardest to beat across asset classes. The daily return $r_{i,t} = \sum_k r_{i,t,k}$ is the sum of the intraday log returns. The HAR components are the daily variance $RV_{i,t}$ and the trailing weekly and monthly averages

$$RV_{i,t}^{(w)} = \frac{1}{5}\sum_{j=0}^{4} RV_{i,t-j}, \qquad RV_{i,t}^{(m)} = \frac{1}{22}\sum_{j=0}^{21} RV_{i,t-j}. \tag{2}$$

The forward targets are the next-day return and realized variance, $r_{i,t+1}$ and $RV_{i,t+1}$, and their 5-day forward sums,

$$r_{i,t+1:t+5}^{(5)} = \sum_{j=1}^{5} r_{i,t+j}, \qquad RV_{i,t+1:t+5}^{(5)} = \sum_{j=1}^{5} RV_{i,t+j}, \tag{3}$$

which require all five future observations and are therefore undefined on the last five trading days of each series. Every predictor is measured at the close of day t, before the first return that enters the target.[1]

3.5. Descriptive Statistics

Table 3 summarizes the panel by asset group. Stocks have higher at-the-money implied volatility (26.6%) and daily return volatility (1.90%) than ETFs (17.8%, 1.25%) and indices (18.2%, 1.29%). The put–call volume ratio is higher for ETFs (2.35) and indices (1.75) than for stocks (0.89), consistent with the use of index and ETF puts for portfolio hedging. The mean 5-day forward return is 25 basis points with a standard deviation of 3.8%, so that the predictors face a target whose cross-sectional variation is dominated by noise.

[1] Version 1.1 computed the 5-day targets as the trailing five-day sum of the series shifted one day forward, that is $\sum_{j=-3}^{1} r_{i,t+j}$, which contains the contemporaneous return $r_{i,t}$ and three lagged returns. Because option features such as the put–call volume ratio and the smile react strongly to the same day's price move, this construction produced spurious "weekly" predictability (R^2 of 5–19%, long-short Sharpe ratios above two) and spuriously high fit for the 5-day variance target, whose trailing terms overlap the HAR regressors. The correct construction in equation (3) is implemented in `wp7.data_io.fix_forward_targets`; the shipped processed data are unchanged and the legacy columns are retained for audit.

Table 3. Descriptive statistics by asset group.

	Stocks	ETFs	Indices	All
N (ticker-days)	188,093	64,256	12,034	264,383
N tickers	49	17	3	69
Mean $IV_{ATM,30d}$	0.2664	0.1784	0.1823	0.2406
Mean $\text{skew}_{25\delta}$	0.0361	0.0425	0.0562	0.0387
Mean PC volume ratio	0.890	2.345	1.754	1.283
Mean daily RV ($\times 10^4$)	4.86	2.00	1.35	4.01
Mean daily return (%)	0.054	0.035	0.047	0.049
Std. daily return (%)	1.90	1.25	1.29	1.74

Notes. Sample period: January 2010–December 2025. $IV_{ATM,30d}$ is the 30-day at-the-money call implied volatility. Skew is the 25-delta put minus 25-delta call IV. PC volume ratio is total put volume over total call volume. RV is the realized variance from 5-minute returns; the daily return is the sum of 5-minute log returns.

Table 4 reports the average within-ticker autocorrelations. At-the-money implied volatility is highly persistent (0.959 at one lag, 0.538 at 22 lags); realized variance is far less persistent at the daily level (0.320 at one lag), which is the usual motivation for the HAR aggregation; daily returns show the mild negative first-order autocorrelation typical of large stocks. The persistence of the implied series matters for the time-series analyses of Section 5: a bivariate VAR in levels of implied volatility and realized variance is estimated on highly persistent but stationary series, and the Granger tests use five lags of each.

Table 4. Average autocorrelation structure.

Variable	Lag 1	Lag 5	Lag 10	Lag 22
ATM IV (30d)	0.959	0.860	0.753	0.538
Skew (25δ)	0.832	0.711	0.593	0.416
Realized variance	0.320	0.211	0.139	0.057
Daily return	−0.062	−0.010	−0.006	−0.030
PC volume ratio	0.299	0.218	0.172	0.118

Notes. Cross-sectional average of within-ticker autocorrelations.

4. Methodology

4.1. Return Predictability (RQ1)

Panel regressions. For horizon $h \in \{1, 5\}$ trading days, the pooled regression is

$$r^{(h)}_{i,t+1:t+h} = \alpha + \boldsymbol{\beta}' \, \mathbf{x}_{i,t} + \varepsilon_{i,t+h}, \tag{4}$$

where $\mathbf{x}_{i,t}$ stacks the eight option-implied features of Table 2 together with $RV_{i,t}$ and $RV^{(w)}_{i,t}$, each winsorized at the 1st and 99th percentiles of the estimation sample and standardized to unit variance, so that β_k is the change in the h-day return, in basis points, per one-standard-deviation change in predictor k. The target is winsorized at the same percentiles. Baseline t-statistics use the heteroskedasticity-robust HC1 estimator of White (1980). Because observations are correlated across tickers on the same day and, for $h = 5$, across the four overlapping days of the target, the baseline inference overstates precision; Section 6 re-estimates (4) with Newey–West standard errors with 5 and 22 lags (Newey and West, 1987), with two-way clustering by ticker and date (Cameron et al., 2011; Petersen, 2009), and with additional controls, quantile regressions (Koenker and Bassett, 1978) and ticker-by-ticker estimation. The regression is estimated separately for stocks, ETFs, indices and the full panel.

Fama–MacBeth regressions. For stocks, equation (4) is also estimated cross-sectionally on each day t with at least ten stocks, giving a time series of slopes $\hat{\boldsymbol{\beta}}_t$ (Fama and MacBeth, 1973). The reported coefficient is the time-series mean, and its standard error is the Newey–West long-run standard deviation of the slope series with $h - 1$ lags, the minimum required for an h-day overlapping target (Hansen and Hodrick, 1980).

Multiple testing. Ten predictors are examined at two horizons. A predictor is said to clear the Bonferroni threshold when $|t| > 3.02$, the 5% critical value for 20 tests, which is close to the $t > 3$ hurdle that Harvey et al. (2016) recommend for new cross-sectional predictors.

Portfolio sorts. On each day t the stocks with a non-missing value of a sorting variable are ranked into equal-weighted quintiles, and the long-short return is the Q5 minus Q1 average forward return $r^{(h)}_{t+1:t+h}$. For $h = 5$ the daily series of long-short returns overlaps, so its mean is tested with a Newey–West standard error with four lags and the annualized Sharpe ratio uses 52 periods; for $h = 1$ the series does not overlap and 252 periods are used. Both quintile and decile sorts are reported, and a 3×3 independent double sort on $IV_{ATM,30d}$ and the implied skewness proxy examines whether skew predictability depends on the level of uncertainty; cell means are time-series means of daily cell portfolio returns and the high-minus-low differences are tested on the daily difference series.

Calendar-time implementation and trading costs. Sorting every day on a target held for h days produces h overlapping cohorts. Following Jegadeesh and Titman (1993), the calendar-time strategy holds the h live cohorts with weight $1/h$ each. Let w_s be the vector of long-short weights formed at the close of s (positive $1/n_{Q5}$ for stocks in the top quintile, negative $1/n_{Q1}$ for those in the bottom quintile, zero otherwise). The position held at the close of day t, the strategy's daily return, its two-sided turnover and its net return under a one-way proportional cost c are

$$W_t = \frac{1}{h}\sum_{k=0}^{h-1} w_{t-k}, \qquad R^{LS}_{t+1} = W_t' \, \mathbf{r}_{t+1}, \qquad TO_t = \sum_i |W_{t,i} - W_{t-1,i}|, \qquad R^{net}_{t+1} = R^{LS}_{t+1} - c\, TO_t. \quad (5)$$

The daily series R^{LS}_{t+1} does not overlap, so its Sharpe ratio is annualized with 252 days and its mean tested with Newey–West standard errors (five lags). The break-even cost is $c^* = \overline{R^{LS}}/\overline{TO}$. Costs of 5 to 25 basis points of traded value bracket the effective half-spreads of large-capitalization US stocks over the sample (Novy-Marx and Velikov, 2016); Lo (2002) discusses the sampling error of Sharpe ratios, which for a 16-year daily series is about 0.25 in annualized units.

4.2. Realized-Variance Forecasting (RQ2)

The benchmark is the heterogeneous autoregressive model of Corsi (2009),

$$RV_{i,t+1} = \alpha + \beta_d \, RV_{i,t} + \beta_w \, RV^{(w)}_{i,t} + \beta_m \, RV^{(m)}_{i,t} + \varepsilon_{i,t+1}, \quad (6)$$

with the components of equation (2); for the 5-day target the left-hand side is $RV^{(5)}_{i,t+1:t+5}$. Four alternatives are compared with it: an ARCH-type benchmark with $RV_{i,t}$ and the squared daily return $r^2_{i,t}$ as regressors, in the spirit of the GARCH(1,1) of Bollerslev (1986) but estimated by least squares; an "IV only" model with $IV_{ATM,30d}$ and $IV_{ATM,90d}$; an "IV surface" model that adds $\text{Skew}_{25\delta}$, the implied skewness proxy and the implied kurtosis proxy; and the HAR-RV model augmented with the IV-surface block.[2] All regressors and targets are winsorized at the 1st and 99th percentiles within ticker, since realized variance levels differ by an order of magnitude between indices and single stocks, and all five models are estimated on the common sample of ticker-days on which every regressor of the largest model is observed, so that differences in R^2 reflect the information in the regressors rather than the composition of the sample.

Out-of-sample performance is evaluated for the 30 tickers with the longest histories with a rolling scheme: each model is estimated by least squares on a window of 500 trading days and used to forecast the next 250 days, the window then advancing by 250 days; forecasts are floored at the

[2] Version 1.1 included $IV_{ATM,30d}$, $IV_{ATM,90d}$ and their difference in the same regression, an exact linear dependency that is harmless in sample but made the rolling out-of-sample forecasts of the IV models numerically unstable. The term slope is therefore excluded from the forecasting models; it remains a return predictor in RQ1. Version 1.1 also used RV_{t-1} rather than RV_t as the daily HAR component.

first percentile of the training-window target to avoid non-positive variance forecasts. The stacked forecasts of ticker i are evaluated with the two losses that Patton (2011) shows to be robust to noise in the realized proxy,

$$\mathrm{MSE}_i = \frac{1}{T_i}\sum_t \big(RV_{i,t} - \widehat{RV}_{i,t}\big)^2, \qquad \mathrm{QLIKE}_i = \frac{1}{T_i}\sum_t \Big[\frac{RV_{i,t}}{\widehat{RV}_{i,t}} - \ln\frac{RV_{i,t}}{\widehat{RV}_{i,t}} - 1\Big], \tag{7}$$

together with $R^2_{OOS} = 1 - \sum_t (RV_{i,t} - \widehat{RV}_{i,t})^2 / \sum_t (RV_{i,t} - \overline{RV}_i)^2$, where $\overline{RV}_i$ is the mean of the realized target over the forecast sample (Campbell and Thompson, 2008). The IV-augmented model is compared with HAR-RV ticker by ticker with the Diebold and Mariano (1995) statistic, the mean of the loss differential $d_t = L(\mathrm{HAR})_t - L(\mathrm{HAR+IV})_t$ divided by its Newey–West standard error with h lags; positive values favour the augmented model. Because HAR-RV is nested in the augmented model the DM statistic is conservative under the null (Clark and West, 2007); the direction of the bias works against finding an improvement.

4.3. Implied Correlation and Market Stress (RQ3)

Let $\sigma_{SPX,t}$ be the 30-day at-the-money implied volatility of SPX and $\sigma_{i,t}$ that of constituent i, $i = 1, \dots, n_t$, for a fixed list of 30 large S&P 500 constituents (Appendix A). Writing the index variance as that of an equally weighted portfolio with a common pairwise correlation ρ, $\sigma^2_{SPX} = n^{-2}\sum_i \sigma_i^2 + \rho\, n^{-2}\sum_{i\neq j}\sigma_i\sigma_j$, and approximating $n^{-2}\sum_{i\neq j}\sigma_i\sigma_j$ by $(1 - n^{-1})\bar{\sigma}^2$ with $\bar{\sigma}$ the mean constituent volatility, the implied correlation is

$$\hat{\rho}_{impl,t} = \frac{\sigma^2_{SPX,t} - n_t^{-1}\,\overline{\sigma^2_{i,t}}}{\big(1 - n_t^{-1}\big)\,\bar{\sigma}^2_t}, \tag{8}$$

clipped to $[0, 1]$. This is the equal-weight analogue of the index of Skintzi and Refenes (2005) and of the CBOE implied correlation index, and it treats the 30 constituents as a proxy for the index basket. The realized correlation $\hat{\rho}_{real,t}$ is the average pairwise correlation of the constituents' daily returns over the trailing 22 trading days; the divergence is $\hat{\rho}_{impl,t} - \hat{\rho}_{real,t}$ and the ratio $\hat{\rho}_{impl,t}/\max(\hat{\rho}_{real,t}, 0.01)$. A stress event is a day on which the VIX closes above 25 or has risen by more than 20% over five days, and the target $S_{t,h}$ is the indicator that at least one stress day occurs in $t+1, \dots, t+h$ for $h \in \{5, 10, 20\}$. Linear probability models of $S_{t,h}$ on standardized predictors, with the VIX level and SPX implied volatility as controls, are estimated separately for the ratio, the divergence and the two components, with Newey–West standard errors with h lags. Including the ratio, the divergence and both components in one regression, as in version 1.1, produces an exactly collinear design.

4.4. Greeks by Time to Expiration and Price Magnets (RQ4)

For ten liquid underlyings the option chain is partitioned into six DTE buckets (1–7, 8–14, 15–30, 31–60, 61–90 and 91–180 days), and within each bucket the next-day return and realized variance are regressed on the bucket's average gamma, vega, $|\Theta|$ and implied volatility and on its call and put gamma-weighted open interest. For the price-magnet test, let $K^*_{i,t}$ be the strike with the largest open interest among options expiring within 30 days, and define

$$\text{Magnet}_{i,t} = \mathbf{1}\Big[\frac{|P^{close}_{i,t} - K^*_{i,t}|}{P^{close}_{i,t}} < \frac{|P^{open}_{i,t} - K^*_{i,t}|}{P^{open}_{i,t}}\Big] \tag{9}$$

on days when the open is between 0.1% and 10% from $K^*_{i,t}$. Under the null of no attraction, $\Pr(\text{Magnet}_{i,t} = 1) = 0.5$; the share is tabulated by quintile of the concentration of open interest at K^*.

4.5. Machine Learning on the SPX Surface (RQ5)

Twenty-one daily features of the SPX surface (Appendix D) and the three HAR components are used to forecast next-day SPX realized variance with a random forest (200 trees, maximum depth 10, minimum leaf size 20) and gradient boosting (200 trees, depth 5, learning rate 0.05, subsampling 0.8), trained on 2010–2019 and evaluated on 2020–2025 against HAR-RV and the VIX benchmark $\widehat{RV}^{VIX}_{t+1} = (VIX_t/100)^2/252$. This step requires the raw stores and its results are frozen outputs of the original run (Section 5).

4.6. Granger Causality and Vector Autoregressions

For each ticker and each ordered pair (x, y) of daily series, the null that x does not Granger-cause y is tested with the F-statistic for the joint exclusion of five lags of x from a regression of y_t on five lags of y and five lags of x (Granger, 1969). A bivariate VAR(5) in standardized $IV_{ATM,30d}$ and RV is estimated for the 20 tickers with the longest histories,

$$\mathbf{y}_t = \mathbf{c} + \sum_{\ell=1}^{5} \mathbf{A}_\ell\, \mathbf{y}_{t-\ell} + \mathbf{u}_t, \qquad \mathbf{y}_t = (IV_t, RV_t)', \quad \mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \mathbf{\Sigma}. \tag{10}$$

Structural shocks are identified recursively: with $\mathbf{P}$ the lower-triangular Cholesky factor of $\mathbf{\Sigma}$ under a given ordering, the orthogonalized responses are $\mathbf{\Theta}_s = \mathbf{\Phi}_s \mathbf{P}$, where $\mathbf{\Phi}_s$ are the reduced-form moving-average coefficients, and the share of the H-step forecast-error variance of variable j attributable to shock k is

$$\omega_{jk}(H) = \frac{\sum_{s=0}^{H-1} \Theta^2_{s,jk}}{\sum_{s=0}^{H-1} \sum_m \Theta^2_{s,jm}} \tag{11}$$

(Sims, 1980; Kilian and Lütkepohl, 2017). Because the residuals of the two equations are contemporaneously correlated, the decomposition depends on the ordering; both orderings are reported.[3]

5. Results

5.1. RQ1: Return Predictability

5.1.1. Panel Regressions

Table 5 reports the explanatory power of the ten predictors by asset group and horizon. The option-implied features explain 0.08% of the variation in next-day stock returns and 0.29% of the variation in 5-day returns; the corresponding figures are 0.14% and 0.61% for ETFs and 0.43% and 1.01% for the three indices. The ordering stocks < ETFs < indices is the cross-asset gradient the paper set out to measure, and the increase from the daily to the weekly horizon is consistent with information that is incorporated over several days rather than overnight. The magnitudes, however, are small: even for the indices, where the "cross-section" consists of three series and the R^2 is mostly time-series predictability of index returns by their own implied volatility, barely one percent of the weekly variation is explained. Four stock predictors are individually significant at the 5% level at the 5-day horizon under HC1 standard errors and three clear the Bonferroni threshold; at the daily horizon two are significant at 5% (the 25-delta skew and the implied volatility level) and none clears it.

Table 5. Return predictability: pooled panel regressions by asset group.

			1-day return				5-day return			
Group	N	Tickers	R^2 (%)	Adj. R^2 (%)	#sig$_{5\%}$	#sig$_B$	R^2 (%)	Adj. R^2 (%)	#sig$_{5\%}$	#sig$_B$
Stocks	79,943	49	0.08	0.07	2	0	0.29	0.27	4	3
ETFs	30,152	17	0.14	0.11	2	1	0.61	0.58	3	2
Indices	8,986	3	0.43	0.32	0	0	1.01	0.90	2	0
All	119,081	69	0.10	0.09	4	1	0.31	0.30	6	3

Notes. Pooled OLS of the h-day forward log return on the ten standardized predictors of Table 2 (plus RV_t and $RV_t^{(w)}$), HC1 standard errors; all variables winsorized at the 1st and 99th percentiles. The 5-day return is $\sum_{j=1}^{5} r_{t+j}$. #sig$_{5\%}$ counts predictors with $|t| > 1.96$; #sig$_B$ counts those clearing the Bonferroni threshold for the 20 tests of each horizon pair ($|t| > 3.02$). N is the number of ticker-days with all ten predictors observed.

Table 6 shows which predictors carry the signal for stocks. In the pooled 5-day regression a one-standard-deviation steeper 25-delta skew is associated with a 16.9 basis-point higher return over

[3] Version 1.1 computed "shares" from the reduced-form unit-shock responses, ignoring $\boldsymbol{\Sigma}$; the resulting 73.8% figure was not a variance decomposition.

the following week ($t = 4.98$), a higher put–call volume ratio with 6.1 basis points ($t = 4.49$), a higher implied kurtosis proxy with 6.1 basis points ($t = 3.14$) and a higher at-the-money implied volatility with 8.2 basis points ($t = 2.51$); the implied skewness proxy enters negatively (−5.8 basis points, $t = -1.93$). The Fama–MacBeth estimates, which weight each day equally and whose standard errors account for the overlap of the weekly target, tell a more cautious story: only the 25-delta skew remains significant (18.3 basis points, $t = 1.99$), and the kurtosis proxy, whose quintile sort was the headline result of version 1.1, has a Fama–MacBeth coefficient of 0.2 basis points ($t = 0.03$). At the daily horizon the pooled regression attributes 4.7 basis points to the skew ($t = 2.89$) and 3.5 basis points to the implied volatility level ($t = 2.40$), but no Fama–MacBeth coefficient reaches $|t| = 2$.

Table 6. Return predictability for individual stocks: coefficients and t-statistics.

	1-day return				5-day return			
	Pooled OLS		Fama–MacBeth		Pooled OLS		Fama–MacBeth	
Predictor	b	(t)	b	(t)	b	(t)	b	(t)
$IV_{ATM,30d}$	3.5**	(2.40)	3.3	(0.85)	8.2**	(2.51)	14.6	(1.30)
IV term slope	1.3	(1.48)	1.0	(0.51)	1.7	(0.81)	4.2	(0.69)
$\text{Skew}_{25\delta}$	4.7***	(2.89)	4.2	(1.12)	16.9***	(4.98)	18.3**	(1.99)
Implied skewness	−1.6	(−1.12)	−1.4	(−0.35)	−5.8*	(−1.93)	−13.6	(−1.54)
Implied kurtosis	0.4	(0.49)	−2.3	(−0.95)	6.1***	(3.14)	0.2	(0.03)
PC volume ratio	1.0*	(1.66)	1.4	(0.79)	6.1***	(4.49)	5.1	(1.46)
PC OI ratio	−0.4	(−0.57)	2.0	(0.76)	0.5	(0.30)	0.5	(0.11)
Net gamma exposure	0.2	(0.32)	3.6*	(1.79)	−1.0	(−0.78)	3.9	(0.77)
RV_t	−2.8*	(−1.89)	5.8	(0.93)	−1.3	(−0.45)	−1.6	(−0.10)
$RV_t^{(w)}$	−1.0	(−0.71)	−13.3*	(−1.79)	1.6	(0.52)	−21.8	(−1.42)

Notes. Individual stocks (49 tickers, 79,943 ticker-days). Coefficients are in basis points of h-day return per one-standard-deviation change in the predictor. Pooled OLS: HC1 t-statistics. Fama–MacBeth: daily cross-sectional regressions (2,415 days at 1 day, 2,415 at 5 days; at least 10 stocks per day); t-statistics from the time series of slopes with Newey–West standard errors (0 lags at 1 day, 4 lags at 5 days). ***, **, *: $|t| > 2.576, 1.96, 1.645$.

The positive sign on the 25-delta skew deserves comment. Xing et al. (2010) find that stocks with steeper smirks earn *lower* subsequent returns, which they attribute to informed trading in puts ahead of bad news, and Stilger et al. (2017) and Rehman and Vilkov (2012) find that more negative risk-neutral skewness predicts lower returns. Conrad et al. (2013), by contrast, find that more negative ex ante skewness is associated with higher returns, which they interpret as compensation for skewness risk. The present estimate is of the latter sign. It is obtained in a universe of 49 very large stocks for which the informed-trading channel is least plausible—information about these

companies is quickly impounded in the stock itself (Muravyev et al., 2013)—and it is the only coefficient whose sign and significance survive every inference method of Section 6. Its magnitude is nevertheless modest, and the sort on the same variable (Table 7) earns an insignificant long-short return, so the skew effect is better described as a weak tilt than as a tradable premium.

5.1.2. Portfolio Sorts

Table 7 translates the regressions into long-short quintile portfolios. With 5-day forward returns, three sorts produce long-short returns that are significant at the 5% level once the overlap of the daily series is accounted for: the sort on at-the-money implied volatility (14.5% per year, Sharpe ratio 0.57, $t_{NW} = 2.71$), the sort on the put–call volume ratio (5.1%, Sharpe 0.42, $t_{NW} = 2.64$) and the sort on the put–call open-interest ratio (5.5%, Sharpe 0.42, $t_{NW} = 2.04$). The sort on the implied skewness proxy earns −8.8% per year ($t_{NW} = -2.41$): stocks with the most negatively skewed smiles earn lower returns, in line with the regression coefficient. The sort on the implied kurtosis proxy, which version 1.1 reported with a Sharpe ratio of 2.33, earns −6.1% per year with a Sharpe ratio of −0.33 ($t_{NW} = -1.66$). The comparison of t_{NW} with t_{iid} shows that ignoring the overlap inflates the t-statistics by a factor of 1.4 to 1.8. At the daily horizon only the implied-volatility and put–call volume sorts are marginally significant ($t \approx 2.1$).

Table 7. Long-short quintile portfolios sorted on option-implied variables (individual stocks).

Sorting variable	Ann. ret. (%)	Ann. vol. (%)	Sharpe	t_{NW}	t_{iid}	N_{days}
Panel A: 5-day forward returns						
$IV_{ATM,30d}$	14.51	25.46	0.570	2.71^{***}	4.86	3,777
PC volume ratio	5.06	12.15	0.416	2.64^{***}	3.66	4,024
PC OI ratio	5.48	13.17	0.416	2.03^{**}	3.66	4,024
$\text{Skew}_{25\delta}$	5.86	19.24	0.305	1.54	2.60	3,778
IV term slope	−2.86	19.79	−0.145	−0.64	−1.04	2,676
Implied kurtosis	−6.12	18.32	−0.334	-1.66^{*}	−2.85	3,777
Implied skewness	−8.80	18.02	−0.488	-2.41^{**}	−4.16	3,777
Panel B: 1-day forward returns						
$IV_{ATM,30d}$	13.88	25.55	0.543	2.10^{**}	2.10	3,777
PC volume ratio	6.24	12.04	0.518	2.07^{**}	2.07	4,024
PC OI ratio	5.13	13.28	0.386	1.54	1.54	4,024
$\text{Skew}_{25\delta}$	3.08	19.70	0.157	0.61	0.61	3,778
IV term slope	−4.83	20.46	−0.236	−0.77	−0.77	2,676
Implied kurtosis	−7.85	18.30	−0.429	-1.66^{*}	−1.66	3,777
Implied skewness	−8.33	17.91	−0.465	-1.80^{*}	−1.80	3,777

Notes. Equal-weighted quintile portfolios of individual stocks formed on each trading day; the long-short portfolio is Q5−Q1. Panel A: 5-day forward returns $\sum_{j=1}^{5} r_{t+j}$, annualized with 52 periods; the daily series of 5-day returns overlaps, so t_{NW} uses Newey–West standard errors with 4 lags while t_{iid} ignores the overlap (the statistic reported in version 1.1). Panel B: next-day returns, annualized with 252 periods. ***, **, *: $|t_{NW}| > 2.576, 1.96, 1.645$.

The double sort in Table 8 refines the picture. Every one of the nine cells earns a positive average weekly return, and the dominant pattern is in the rows rather than the columns: high-implied-volatility stocks earn 16 to 23 basis points per week more than low-implied-volatility stocks within each skewness tercile (t between 1.4 and 2.5), whereas the skewness differences within implied-volatility terciles are small and insignificant (t between −1.5 and 1.0). The interaction that version 1.1 emphasized—a large negative return for high-volatility, high-skewness stocks—does not exist once the target is correctly aligned.

Table 8. Double sort: ATM IV × implied skewness → 5-day returns (bps/week).

	Low skew	Medium	High skew	*High−Low skew*
Low IV	12.2*	13.7**	18.8***	5.3
	(1.70)	(2.52)	(3.56)	(1.00)
Medium IV	22.3***	30.3***	25.7***	4.0
	(3.02)	(4.11)	(3.37)	(0.72)
High IV	40.1***	33.0***	34.0***	−17.9
	(3.55)	(3.31)	(2.59)	(−1.52)
High−Low IV	23.1**	20.1**	15.7	
	(2.00)	(2.45)	(1.42)	

Notes. Independent daily tercile sorts of individual stocks on $IV_{ATM,30d}$ and implied skewness; cells report the time-series mean of the daily equal-weighted cell portfolio's 5-day forward return in basis points per week, with Newey–West (4 lags) t-statistics in parentheses. The last row and column are differences between the daily High and Low portfolio returns. ***, **, *: $|t| > 2.576$, 1.96, 1.645.

5.1.3. Economic Significance After Trading Costs

Table 9 reports the calendar-time strategies of equation (5). With a five-day holding period the gross Sharpe ratios range from −0.52 (implied skewness) to 0.57 (put–call volume ratio), and only three strategies have $|t_{NW}| > 2$. Turnover is the decisive quantity. The put–call volume ratio changes from day to day, so its strategy turns over 51% of the long leg every day; its break-even one-way cost is 3.8 basis points and its Sharpe ratio is negative at a cost of 10 basis points. The put–call open-interest ratio is more persistent (turnover 14% per day, break-even 14.6 basis points) but its gross Sharpe ratio of 0.41 falls to 0.13 net of 10 basis points. The implied-volatility sort is the only strategy that remains profitable at realistic costs: implied volatility is highly persistent (Table 4), the strategy turns over 20% of the long leg per day, its break-even cost is 27 basis points and its Sharpe ratio is 0.34 net of 10 basis points and 0.04 net of 25 basis points. With daily rebalancing ($h = 1$) turnover roughly doubles and no strategy retains a meaningful Sharpe ratio after a 10 basis-point cost (the largest is 0.06).

Table 9. Calendar-time long-short strategies: gross performance, turnover and net-of-cost Sharpe ratios.

	Gross						Net Sharpe	
Sorting variable	Ret. (%)	Vol. (%)	Sharpe	t_{NW}	Turnover	Break-even (bps)	10 bps	25 bps
Panel A: 5-day holding period								
PC volume ratio	4.91	8.62	0.57	2.29**	0.51	3.8	−0.93	−3.18
$IV_{ATM,30d}$	13.59	25.06	0.54	2.19**	0.20	27.0	0.34	0.04
PC OI ratio	5.22	12.66	0.41	1.66*	0.14	14.6	0.13	−0.29
$\text{Skew}_{25\delta}$	4.92	18.01	0.27	1.12	0.39	5.0	−0.27	−1.09
IV term slope	−3.31	18.54	−0.18	−0.69	0.45	−2.9	−0.79	−1.70
Implied kurtosis	−6.01	17.03	−0.35	−1.41	0.39	−6.1	−0.93	−1.79
Implied skewness	−8.72	16.86	−0.52	−2.09**	0.35	−10.0	−1.04	−1.81
Panel B: 1-day holding period								
$IV_{ATM,30d}$	13.88	25.55	0.54	2.16**	0.49	11.1	0.06	−0.68
PC volume ratio	6.24	12.04	0.52	2.09**	2.32	1.1	−4.33	−11.50
PC OI ratio	5.13	13.28	0.39	1.56	0.29	7.1	−0.16	−0.98
$\text{Skew}_{25\delta}$	3.08	19.70	0.16	0.63	1.30	0.9	−1.51	−3.99
IV term slope	−4.83	20.46	−0.24	−0.79	1.04	−1.8	−1.52	−3.43
Implied kurtosis	−7.85	18.30	−0.43	−1.65*	1.27	−2.5	−2.17	−4.77
Implied skewness	−8.33	17.91	−0.46	−1.82*	1.18	−2.8	−2.13	−4.61

Notes. Calendar-time implementation of the long-short quintile strategies (49 stocks): each day a new Q5−Q1 cohort is formed and held for h trading days, so the strategy holds h overlapping cohorts with weight $1/h$ each and its daily return series does not overlap (Jegadeesh and Titman, 1993). Returns and volatilities are annualized with 252 days; t_{NW} uses 5 Newey–West lags. Turnover is the average daily two-sided traded value $\sum_i |W_{t,i} - W_{t-1,i}|$ per unit of long leg; the break-even cost is the one-way proportional cost (basis points of traded value) at which the mean net return is zero; net Sharpe ratios deduct one-way costs of 10 and 25 bps (Novy-Marx and Velikov, 2016). ***, **, *: $|t_{NW}| > 2.576, 1.96, 1.645$.

Two qualifications apply to the implied-volatility strategy. First, a long position in the highest-implied-volatility quintile of large stocks and a short position in the lowest is a bet on high-beta, high-volatility names whose return over 2010–2025 reflects, at least in part, exposure to market and sector factors rather than information in option prices; the strategy's annualized volatility of 25% is consistent with a substantial net beta, and no factor adjustment is performed here. Second, its t-statistic of 2.2 is one of fourteen computed in Table 9 and would not survive the $t > 3$ hurdle of Harvey et al. (2016). The appropriate conclusion is that, in this universe, option-implied cross-sectional return predictability is statistically detectable but economically marginal, which is what the limits-to-arbitrage interpretation of the literature predicts for the most liquid stocks (Johnson and So, 2012; Stilger et al., 2017; Muravyev et al., 2025).

5.2. RQ2: The Implied Volatility Surface and Realized-Variance Forecasting

Table 10 compares the five models in sample on the common set of 119,093 ticker-days. For the next-day target, HAR-RV explains 41.5% of the variation in realized variance; the two option-only models explain 35.5–36.1%, less than HAR-RV but more than the ARCH-type benchmark; and adding the surface to HAR-RV raises the R^2 to 45.7%, an improvement of 9.9% relative to HAR-RV. For the 5-day target the ranking changes: the implied volatility level alone ($R^2 = 0.463$) now beats HAR-RV (0.421), the full surface adds little to the level (0.466), and the augmented model reaches 0.515, 22.3% above HAR-RV. The pattern is the one the implied-volatility literature leads one to expect (Christensen and Prabhala, 1998; Busch et al., 2011): implied volatility is a forecast of average variance over the life of the option, so its relative value increases with the forecast horizon, while the daily HAR components track the short-run dynamics that matter most for the next day.

Table 10. In-sample fit of realized-variance forecasting models (common sample).

		RV_{t+1}		$\sum_{j=1}^{5} RV_{t+j}$	
Model	k	R^2	ΔR^2 (%)	R^2	ΔR^2 (%)
$RV_t + r_t^2$ (ARCH-type)	2	0.360	−13.3	0.317	−24.7
HAR-RV	3	0.415	+0.0	0.421	+0.0
IV only (30d, 90d)	2	0.355	−14.6	0.463	+9.8
IV surface	5	0.361	−13.1	0.466	+10.5
HAR-RV + IV surface	8	0.457	+9.9	0.515	+22.3

Notes. Pooled OLS on the common sample of 119,093 ticker-days (69 tickers) for which every regressor of the largest model is observed; regressors and targets winsorized at the 1st and 99th percentiles within ticker and standardized. k is the number of regressors. ΔR^2 is the change relative to HAR-RV in percent of the HAR-RV R^2. Model definitions in Section 4.

Table 11 evaluates the models out of sample for the 30 tickers with the longest histories. For the next-day target the median R^2_{OOS} rises from 0.241 for HAR-RV to 0.309 for the augmented model, the median ratio of mean squared errors is 0.967, and the Diebold–Mariano test rejects equal MSE in favour of the augmented model for 14 of the 30 tickers while never rejecting in favour of HAR-RV (Table A2). Under the QLIKE loss the conclusion reverses: the median QLIKE ratio is 1.17, and the augmented model is significantly worse for 15 tickers and significantly better for none. QLIKE penalizes under-prediction of variance much more heavily than over-prediction and is the loss function most robust to noise in the realized proxy (Patton, 2011); the IV-augmented model, with eight estimated coefficients in a 500-day window, produces occasional very low forecasts in calm periods that MSE barely notices but QLIKE punishes. For the 5-day target the two losses agree

more closely: the augmented model is better under MSE for 8 tickers and worse for 1, and under QLIKE better for 6 and worse for 5, while the parsimonious IV-only model has the best median R^2_{OOS} (0.276) and the lowest median QLIKE (0.220) of all five models. The out-of-sample evidence therefore supports a qualified version of the in-sample result: the surface adds information about future variance, but the additional parameters cost precision, and at the weekly horizon the level of implied volatility alone is the most reliable forecaster.

Table 11. Out-of-sample realized-variance forecasting: HAR-RV against IV-surface models.

	R^2_{OOS}		QLIKE		
Model	Median	Mean	Median	Mean	$R^2_{OOS} > 0$
Panel A: target RV_{t+1}					
$RV_t + r_t^2$ (ARCH-type)	0.212	0.194	0.308	0.297	28/30
HAR-RV	0.241	0.248	0.296	0.285	28/30
IV only (30d, 90d)	0.227	0.227	0.354	0.355	30/30
IV surface	0.204	0.202	0.408	0.453	28/30
HAR-RV + IV surface	0.309	0.296	0.341	0.350	29/30
HAR-RV + IV surface vs. HAR-RV: median MSE ratio 0.967, $\mathrm{DM}_{MSE} > 1.96$ for 14/30 tickers, < -1.96 for 0; median QLIKE ratio 1.171, $\mathrm{DM}_{QLIKE} > 1.96$ for 0/30, < -1.96 for 15.					
Panel B: target $\sum_{j=1}^{5} RV_{t+j}$					
$RV_t + r_t^2$ (ARCH-type)	0.128	0.112	0.262	0.254	26/30
HAR-RV	0.204	0.184	0.247	0.240	26/30
IV only (30d, 90d)	0.276	0.260	0.220	0.235	30/30
IV surface	0.235	0.209	0.248	0.302	28/30
HAR-RV + IV surface	0.264	0.241	0.246	0.264	29/30
HAR-RV + IV surface vs. HAR-RV: median MSE ratio 0.941, $\mathrm{DM}_{MSE} > 1.96$ for 8/30 tickers, < -1.96 for 1; median QLIKE ratio 1.001, $\mathrm{DM}_{QLIKE} > 1.96$ for 6/30, < -1.96 for 5.					

Notes. Rolling-window forecasts for the 30 tickers with the longest histories: each model is re-estimated by OLS on 500 trading days and used for the next 250 days, the window then advancing by 250 days; the stacked forecasts yield per-ticker $R^2_{OOS} = 1 - \sum(RV - \widehat{RV})^2 / \sum(RV - \overline{RV})^2$ and QLIKE $= \frac{1}{T}\sum[RV/\widehat{RV} - \ln(RV/\widehat{RV}) - 1]$ (Patton, 2011). Columns report the median and mean across tickers and the number of tickers with positive R^2_{OOS}. Diebold–Mariano statistics (Diebold and Mariano, 1995) are computed per ticker on the loss differential with Newey–West variance (h lags); positive values favour the IV-augmented model. Per-ticker statistics are in Table A2.

5.3. RQ3: Implied Correlation and Market Stress

Table 12 reports the stress regressions for February 2016 to December 2025, the period over which the realized-correlation series is available. Controlling for the VIX level and SPX implied volatility, a one-standard-deviation increase in the ratio of implied to realized correlation raises the probability of a stress episode in the next 5, 10 and 20 trading days by 5.2, 7.4 and 4.7 percentage points, with overlap-robust t-statistics of 3.27, 4.13 and 2.33 (Panel A). The arithmetic divergence carries the

same sign but is weaker (3.4, 4.1 and 3.7 percentage points; t of 2.14, 1.89 and 1.38; Panel B). Entering the two components separately (Panel C) shows why the ratio works better: the probability of stress rises with implied correlation (t of 2.78 at five days) and falls with realized correlation (t of −1.71), so that the informative configuration is a high implied correlation relative to a low realized one, which the ratio captures more sharply than the difference. The incremental explanatory power is modest—the R^2 values of the three specifications differ by at most 1.4 percentage points at any horizon—and the VIX level itself is the strongest predictor, as it must be when the stress indicator is defined from the VIX. Appendix B shows that implied correlation rose after each of nine major market events of the sample, from +0.03 after the June 2022 rate shock to +0.33 after the February 2018 volatility spike.

Table 12. Implied versus realized correlation as a predictor of market stress.

	5-day		10-day		20-day	
Predictor	b	(t)	b	(t)	b	(t)
Panel A: $\hat{\rho}_{impl}/\hat{\rho}_{real}$						
Corr. ratio	5.22***	(3.27)	7.39***	(4.13)	4.68**	(2.33)
VIX level	23.81**	(2.57)	31.69***	(2.71)	44.17***	(3.27)
SPX ATM IV	−2.57	(−0.27)	−15.23	(−1.29)	−34.08**	(−2.52)
R^2	0.222		0.151		0.088	
Panel B: $\hat{\rho}_{impl} - \hat{\rho}_{real}$						
Corr. divergence	3.37**	(2.14)	4.07*	(1.89)	3.73	(1.38)
VIX level	26.08***	(2.78)	35.83***	(3.01)	45.27***	(3.33)
SPX ATM IV	−4.17	(−0.43)	−18.44	(−1.53)	−34.56**	(−2.54)
R^2	0.215		0.136		0.085	
Panel C: implied and realized correlation entered separately						
$\hat{\rho}_{impl}$	10.64***	(2.78)	9.26**	(2.02)	5.44	(1.16)
$\hat{\rho}_{real}$	−3.70*	(−1.71)	−4.84*	(−1.66)	−4.76	(−1.30)
VIX level	29.70***	(3.21)	38.49***	(3.29)	46.25***	(3.46)
SPX ATM IV	−14.92	(−1.41)	−26.33**	(−2.06)	−37.48***	(−2.69)
R^2	0.222		0.140		0.085	
N (days)	2,486		2,481		2,471	
Base rate	0.356		0.493		0.693	

Notes. Linear probability models of the indicator that a stress event (VIX > 25 or a 5-day VIX increase above 20%) occurs within the next h trading days, on standardized predictors observed at t; February 2016–December 2025 (the period over which the realized-correlation series is available). Coefficients are in percentage points of probability per standard deviation; t-statistics in parentheses use Newey–West standard errors with h lags because the forward indicator overlaps. The base rate is the unconditional frequency of the indicator. ***, **, *: $|t| > 2.576, 1.96, 1.645$.

5.4. RQ4: Greeks by Time to Expiration and Price Magnets

Table 13 reports, for ten liquid underlyings, the fit of bucket-level Greeks for next-day realized variance and returns; it is a frozen output of the original run because it requires the raw option store. The R^2 for realized variance rises from 3.8% in the one-week bucket to 9.5% in the three-month bucket and 8.2% in the six-month bucket, and in every bucket the average implied volatility is the regressor that carries the fit (t between 31 and 36), the other Greeks being nearly collinear within a bucket. Longer-dated implied volatility is a less noisy measure of expected variance than the implied volatility of options about to expire, whose prices are dominated by the time value of a few days. Next-day returns are unpredictable from bucket-level Greeks at every maturity ($R^2 \leq 0.17\%$).

Table 13. Predictive content of bucket-level Greeks by days to expiry.

		RV_{t+1}		r_{t+1}	
DTE bucket (days)	N	R^2 (%)	t_{IV}	R^2 (%)	t_{IV}
1 week (1–7)	34,660	3.77	33.1	0.00	−0.13
2 weeks (8–14)	31,912	6.70	31.8	0.03	1.98
1 month (15–30)	33,715	6.85	32.0	0.02	1.87
2 months (31–60)	36,449	7.52	34.0	0.05	2.55
3 months (61–90)	31,354	9.47	35.1	0.17	2.28
6 months (91–180)	36,719	8.18	35.9	0.04	2.04

Notes. Ten liquid underlyings (AAPL, MSFT, AMZN, GOOG, TSLA, NVDA, META, JPM, SPY, QQQ). Within each days-to-expiry bucket the next-day RV or return is regressed on the bucket's average gamma, vega, $|\Theta|$, IV and call/put gamma-weighted open interest (HC1). Only the R^2 and the IV t-statistic are reported: the Greeks are nearly collinear within a bucket and their individual t-statistics are not interpretable. Frozen output of `scripts/05` part A (raw-dependent).

Table 14 tests the price-magnet hypothesis on 10,480 ticker-days. The underlying moved toward the maximum-open-interest strike on 47.0% of days, below rather than above the 50% null, and the share does not increase with the concentration of open interest at that strike (45.8% to 48.0% across quintiles). The binomial standard error of the pooled share is about 0.5 percentage points, so the deviation from one half is in the wrong direction for the hypothesis and is not large in economic terms. The "max pain" narrative receives no support on ordinary trading days; this does not contradict the expiration-day clustering documented by Ni et al. (2005), which concerns the last hours before expiration and is a different phenomenon.

Table 14. Price-magnet test: does the underlying drift toward the maximum-open-interest strike?

OI-concentration quintile	*N*	Moved toward strike (%)	Open distance (%)	Close distance (%)
Q1 (lowest)	2,096	47.3	4.07	4.24
Q2	2,096	48.0	3.98	4.16
Q3	2,096	45.8	4.13	4.34
Q4	2,096	46.1	3.72	3.87
Q5 (highest)	2,096	47.8	3.33	3.41
All	10,480	47.0		

Notes. Ticker-days (ten liquid underlyings, 2010–2025) on which the open was within 0.1%–10% of the maximum-open-interest strike among options expiring within 30 days. "Moved toward strike" is the share of days on which the close was relatively closer to that strike than the open; under the null of no attraction the expected share is 50% (binomial standard error 0.49 percentage points for the pooled sample, ignoring dependence). OI concentration is the share of the maximum strike in total open interest. Built from the shipped `price_magnet_data.parquet`.

5.5. RQ5: Machine Learning on the SPX Surface

Table 15 reports the out-of-sample comparison for next-day SPX realized variance over 2020–2025, a frozen output of the original run. HAR-RV attains $R^2_{OOS} = 0.437$ and dominates the VIX benchmark (0.336), the random forest (0.143) and gradient boosting (0.095). The weak performance of the tree-based learners is not surprising: they are trained on 2010–2019 and tested on a period that begins with the COVID-19 variance burst, which lies outside the range of the training targets, and trees cannot extrapolate. Christensen et al. (2023) find that machine-learning methods improve on HAR only when carefully tuned and that the gains are concentrated in periods of market stress with sufficiently rich training data. The random-forest importances (Appendix E) are informative nonetheless: the two shortest at-the-money tenors account for two thirds of the total importance (50.8% for the two-week tenor, 15.1% for the one-week tenor), echoing the finding of RQ2 that the level of short-dated implied volatility is the surface's most valuable element. The original run also estimated OLS models on the 21 surface features; because that feature set contained exact linear dependencies, the forecasts were numerically degenerate and are not reported, and the 5- and 22-day targets of the original run were affected by the alignment error of Section 3.

Table 15. Out-of-sample SPX realized-variance forecasting, 2020–2025 (next-day target).

Model	MSE ($\times 10^7$)	MAE ($\times 10^4$)	R^2_{OOS}
VIX benchmark, $(VIX_t/100)^2/252$	1.835	1.371	0.335
OLS: HAR-RV	1.556	1.018	0.437
Random forest (surface + HAR)	2.368	1.148	0.143
Gradient boosting (surface + HAR)	2.501	1.282	0.095

Notes. Target: next-day 5-minute realized variance of the S&P 500. Training 2010–2019 (2,515 days), test 2020–2025 (1,507 days). Frozen output of `scripts/06` (raw-dependent; not re-run in the 2026-10-17 revision). The OLS surface models of the original run are omitted because their design matrix contained exact linear dependencies (term-structure slopes defined as differences of included tenors) and their forecasts were numerically degenerate; the 5- and 22-day targets of the original run were mis-aligned (Section 3) and are not reported.

5.6. Granger Causality and Vector Autoregressions

Table 16 reports the Granger causality tests. At-the-money implied volatility Granger-causes realized variance in every one of the 69 tickers (mean $F = 62.4$), whereas realized variance Granger-causes implied volatility in 61% of them (mean $F = 17.9$). The 25-delta skew also Granger-causes realized variance (94% of tickers), and returns Granger-cause the skew (96%)—the asymmetric volatility feedback by which negative returns steepen the smirk—while the skew Granger-causes returns in only 13% of tickers. No option-implied series Granger-causes returns in more than a quarter of the tickers, and the put–call ratio does so in 6%, which is the rejection rate expected under the null. These results are consistent with the regression evidence: option prices lead second moments, not first moments.

Table 16. Granger causality tests (5 lags).

Null hypothesis rejected: $X \to Y$	Mean F	Mean p	% tickers sig. (5%)	Tickers
ATM IV → RV	62.38	0.000	100.0	69
RV → ATM IV	17.93	0.150	60.9	69
$\text{Skew}_{25\delta}$ → RV	20.79	0.018	94.2	69
RV → $\text{Skew}_{25\delta}$	6.59	0.063	75.4	69
ATM IV → Return	1.80	0.269	24.6	69
Return → ATM IV	3.98	0.174	58.0	69
$\text{Skew}_{25\delta}$ → Return	1.41	0.382	13.0	69
Return → $\text{Skew}_{25\delta}$	13.58	0.013	95.7	69
PC volume ratio → Return	1.01	0.500	5.8	69
Return → PC volume ratio	2.82	0.164	46.4	69
Implied skewness → RV	1.34	0.386	17.4	69
Implied kurtosis → RV	5.62	0.086	75.4	69

Notes. Per-ticker bivariate Granger causality F-tests with 5 lags (Granger, 1969) on winsorized daily series; the table reports the cross-ticker mean of the F-statistic and of the p-value, and the share of tickers rejecting the null of no Granger causality at the 5% level.

The strength of the Granger rejections should not be read as a statement about the share of variance that implied volatility explains. Table 17 reports the orthogonalized variance decomposition of the bivariate VAR(5). With realized variance ordered first, so that its innovations may move implied volatility within the day but not the reverse, implied-volatility shocks account for 10.7% of the forecast-error variance of realized variance at five days and 20.2% at twenty days; with implied volatility ordered first the shares are 18.8% and 28.5%. In both orderings realized-variance shocks explain less than 6% of the variance of implied volatility at any horizon. Realized variance is thus driven mainly by its own innovations, and implied volatility is an informative but secondary source of its predictable variation. The gap between the two orderings reflects a mean contemporaneous residual correlation of 0.20; version 1.1's figure of 73.8% at twenty days was obtained from unit reduced-form shocks that ignore this correlation and is not comparable.

Table 17. Forecast-error variance decomposition (%), bivariate VAR(5) in ATM IV and RV.

	RV variance due to IV shock		IV variance due to RV shock	
Horizon (days)	RV first	IV first	RV first	IV first
1	0.0	4.7	4.7	0.0
2	3.7	10.0	4.5	2.2
5	10.7	18.8	4.5	4.1
10	15.2	23.9	4.5	4.8
15	18.2	26.7	4.6	5.2
20	20.2	28.5	4.6	5.4

Notes. Cholesky-orthogonalized forecast-error variance decomposition of the bivariate VAR(5) in standardized ATM IV and RV, averaged over the 20 tickers with the longest histories (mean $N = 3{,}622$ days per ticker). "RV first" orders RV before IV (RV shocks may move IV contemporaneously, not the reverse); "IV first" is the opposite ordering (Sims, 1980; Kilian and Lütkepohl, 2017). The mean contemporaneous residual correlation is 0.20; mean equation R^2: 0.94 (IV) and 0.40 (RV).

6. Robustness

6.1. Standard-Error Specification

Table 18 re-estimates the pooled return regressions on the full 69-ticker panel and reports t-statistics under four covariance estimators. Moving from HC1 to Newey–West standard errors with 5 or 22 lags reduces the 5-day t-statistics by roughly a quarter to 40%; two-way clustering by ticker and date (Cameron et al., 2011; Petersen, 2009), which Petersen (2009) shows to be the appropriate choice when both a firm and a time effect are present in the residuals, reduces them by a further half. Under two-way clustering only the 25-delta skew remains significant at the 5% level for the 5-day return ($t = 2.91$); the put–call open-interest ratio ($t = -1.93$) and the implied skewness proxy ($t = -1.80$) are marginal, and the kurtosis proxy, the put–call volume ratio and the implied volatility level, all significant under HC1, are not (t of 0.97, 1.66 and 0.80). For the next-day return no predictor is significant under two-way clustering (the skew has $t = 1.80$). The cross-sectional correlation of returns on a given day is the dominant source of dependence in this panel: with 69 tickers and some 2,700 days, the effective number of independent observations is far closer to the number of days than to the number of ticker-days.

Table 18. Robustness of return predictability to the standard-error specification.

		t-statistic			
Predictor	b (bps)	HC1	NW(5)	NW(22)	DC
Panel A: 5-day return (R^2 = 0.31%, N = 119,081)					
$IV_{ATM,30d}$	5.02	1.99	1.17	1.08	0.80
IV term slope	−0.30	−0.19	−0.11	−0.10	−0.07
$\text{Skew}_{25\delta}$	19.14	7.59	4.78	4.48	2.91***
Implied skewness	−8.23	−3.60	−2.46	−2.32	−1.80*
Implied kurtosis	3.05	2.08	1.37	1.32	0.97
PC volume ratio	3.04	2.86	2.20	2.08	1.66*
PC OI ratio	−4.54	−3.68	−2.19	−1.96	−1.93*
Net gamma exposure	−1.21	−1.18	−0.69	−0.65	−0.90
RV_t	−0.69	−0.30	−0.27	−0.28	−0.12
$RV_t^{(w)}$	1.02	0.42	0.29	0.28	0.14
Panel B: 1-day return (R^2 = 0.10%, N = 119,081)					
$IV_{ATM,30d}$	2.80	2.47	2.62	2.74	0.92
IV term slope	0.81	1.17	1.24	1.27	0.43
$\text{Skew}_{25\delta}$	5.59	4.66	5.08	5.41	1.80*
Implied skewness	−2.74	−2.51	−2.72	−2.81	−1.27
Implied kurtosis	0.32	0.45	0.49	0.50	0.20
PC volume ratio	−0.01	−0.03	−0.03	−0.03	−0.02
PC OI ratio	−0.64	−1.14	−1.18	−1.18	−0.84
Net gamma exposure	0.05	0.11	0.11	0.12	0.11
RV_t	−3.21	−2.75	−2.79	−2.79	−0.97
$RV_t^{(w)}$	−0.47	−0.40	−0.41	−0.41	−0.12

Notes. Pooled regressions on the full 69-ticker panel. b is in basis points per standard deviation. HC1: heteroskedasticity-robust (White, 1980); NW(L): Newey–West with L lags (Newey and West, 1987); DC: two-way clustering by ticker and date (Cameron et al., 2011; Petersen, 2009). Stars refer to the DC statistic: ***, **, *: $|t| > 2.576, 1.96, 1.645$.

6.2. Subperiod Stability

Table 19 splits the sample into nine subperiods. Return predictability is concentrated in volatile markets: the 5-day R^2 is 0.13% in the low-volatility years 2017–18 and 0.8% in 2010–12 and 2019, but 3.7% during the 2022 rate-hiking cycle and 4.2% in 2020. The realized-variance result is stable in sign across all nine subperiods: the IV surface raises the R^2 of HAR-RV in each of them, from +4.3% in 2020, when the variance burst made the recent realized history unusually informative, to more than +100% in 2017–18, when realized variance was so low that the HAR components explained only 7% of its variation and option prices carried most of the predictable signal. The pre-

and post-March-2020 split gives the same picture: the HAR-RV R^2 is 0.268 before and 0.565 after, and the augmented model raises it to 0.358 and 0.591.

Table 19. Subperiod stability.

	Return predictability			RV_{t+1} forecasting			
Subperiod	N	1-day R^2 (%)	5-day R^2 (%)	N	HAR-RV	HAR + IV	ΔR^2 (%)
Post-GFC (2010–12)	8,565	0.46	0.79	8,566	0.205	0.250	+22.0
Bull (2013–16)	25,950	0.46	2.14	25,950	0.236	0.335	+41.7
Low vol (2017–18)	16,271	0.15	0.13	16,273	0.070	0.149	+114.2
Pre-COVID (2019)	8,374	0.32	0.78	8,376	0.204	0.281	+37.8
COVID (2020)	8,892	1.79	4.18	8,892	0.709	0.739	+4.3
Post-COVID (2021)	9,734	0.75	1.06	9,737	0.398	0.436	+9.6
Rate hikes (2022)	10,068	1.00	3.73	10,070	0.336	0.415	+23.7
Recovery (2023–24)	20,396	0.57	1.78	20,398	0.268	0.342	+27.5
Recent (2025)	10,831	0.70	1.46	10,831	0.340	0.383	+12.7

Notes. Pooled regressions re-estimated within each subperiod (full panel). Return regressions: ten predictors, HC1, pooled winsorization. RV regressions: HAR-RV and HAR-RV + IV surface estimated on the same observations (complete IV surface), per-ticker winsorization; ΔR^2 is relative to HAR-RV.

6.3. VIX Regimes

Table 20 conditions on the VIX level at the time of the forecast. Weekly return predictability rises monotonically from 0.11% when the VIX is below 15 to 3.0% when it exceeds 35, and the fit of the augmented variance model from 0.35 to 0.62. Option-implied information is most valuable when uncertainty is high, but the crisis regime covers 2.4% of the panel, and the predictability measured there is a mixture of genuine conditional information and of the mechanical increase in the dispersion of both returns and predictors in turbulent markets.

Table 20. Predictability by VIX regime.

VIX regime	% of panel	N	1-day R^2 (%)	5-day R^2 (%)	HAR + IV R^2
Very low (<15)	37.6	43,913	0.07	0.11	0.346
Low (15–20)	34.0	40,018	0.06	0.19	0.298
Medium (20–25)	15.4	19,029	0.21	0.55	0.353
High (25–35)	10.6	13,069	1.01	1.27	0.373
Crisis (>35)	2.4	2,529	2.32	2.99	0.623

Notes. Observations are assigned to regimes by the VIX close on day t. Return regressions as in Table 5 (full panel); the last column is the in-sample R^2 of the HAR-RV + IV surface model for RV_{t+1} within the regime.

6.4. Additional Controls, Quantile Regressions and Ticker-Level Fit

Table 21 adds the log of option volume and open interest, the absolute daily return, the lagged daily return and the lagged weekly return to the 5-day regression. The R^2 rises from 0.31% to 0.40%, mostly because the lagged daily return ($t = -4.17$) and the absolute return ($t = -6.10$) capture the short-term reversal of large stocks; the 25-delta skew ($t = 8.08$), the implied skewness proxy ($t = -4.09$) and the put–call open-interest ratio ($t = -3.73$) retain their signs and HC1 significance, while the implied volatility level and the kurtosis proxy fall to $t \approx 2.5$.

Table 21. Return predictability with additional controls (5-day returns).

Regressor	b (bps)	t_{HC1}
$IV_{ATM,30d}$	6.37	2.51**
IV term slope	−0.21	−0.13
$\text{Skew}_{25\delta}$	20.45	8.08***
Implied skewness	−9.42	−4.08***
Implied kurtosis	3.74	2.51**
PC volume ratio	2.48	2.33**
PC OI ratio	−4.87	−3.73***
Net gamma exposure	−1.01	−0.95
RV_t	2.56	1.06
$RV_t^{(w)}$	1.49	0.61
log option volume	8.34	3.40***
log open interest	−5.94	−2.45**
$\lvert r_t \rvert$	−10.34	−6.10***
r_t (lagged return)	−6.16	−4.17***
$\sum_{j=1}^{5} r_{t-j}$	1.31	0.88

Notes. Pooled 5-day return regression on the full panel with five additional controls; R^2 = 0.40% against 0.31% without controls. ***, **, *: $|t| > 2.576, 1.96, 1.645$.

The quantile regressions in Table 22 (Koenker and Bassett, 1978) report coefficients only, since the iteratively reweighted estimator used here does not produce standard errors; they document the shape rather than the significance of the relations. The implied volatility level has by far the largest coefficients and they change sign across the distribution, from −173 basis points at the 10th percentile to +179 at the 90th: high implied volatility predicts a wider distribution of 5-day returns, not a shift in its location, which is exactly what a volatility forecast should do and explains why its effect on the mean is small. The 25-delta skew is positive at every quantile and the put–call

open-interest ratio negative at every quantile, so these two predictors shift the whole distribution; the coefficients of the kurtosis and net gamma proxies change sign across quantiles.

Table 22. Quantile regression coefficients (5-day returns, bps).

Predictor	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$IV_{ATM,30d}$	−173.4	−81.4	3.6	92.7	178.9
IV term slope	−1.4	−2.3	−2.2	0.3	1.7
$\text{Skew}_{25\delta}$	5.3	14.3	26.1	20.4	6.8
Implied skewness	−5.4	−1.6	−5.7	−3.6	−4.2
Implied kurtosis	12.0	5.2	1.6	−3.2	−6.9
PC volume ratio	3.0	2.1	2.7	1.4	1.6
PC OI ratio	−4.5	−4.4	−3.6	−3.3	−5.2
Net gamma exposure	5.2	2.1	−2.2	−4.4	−3.6
RV_t	−8.5	−6.2	−1.3	5.7	2.2
$RV_t^{(w)}$	0.8	−7.8	3.5	15.3	0.3

Notes. Quantile regression coefficients (Koenker and Bassett, 1978) of the 5-day return on the standardized predictors (basis points per standard deviation), full panel, estimated by iteratively reweighted least squares. No standard errors are computed; the table documents the sign and magnitude of the coefficients across the conditional distribution only.

Estimating the 5-day regression ticker by ticker gives R^2 values with mean 2.3%, median 2.0% and interquartile range [1.4%, 2.7%] across the 69 tickers. These figures are larger than the pooled 0.31% because each ticker regression has its own ten coefficients and about 1,700 observations, for which the expected R^2 under the null of no predictability is about 0.6%; the pooled estimate is the one that constrains the coefficients to be common. A 252-day rolling-window regression on the full panel, re-estimated every 63 days, gives a 5-day return R^2 with mean 2.2% (standard deviation 1.4%, range 0.5–5.8%) and a next-day HAR + IV variance R^2 with mean 0.32 (range 0.03–0.77); the return R^2 never approaches the levels reported in version 1.1.

6.5. Winsorization and HAC-Lag Sensitivity

Table 23 examines the two implementation choices most likely to drive the regression results. Panel A shows that they are sensitive to the treatment of the tails. Without winsorization the kurtosis proxy ($t = -0.47$), the implied skewness proxy ($t = 0.29$) and the put–call volume ratio ($t = 0.28$) are all insignificant and four predictors in total are significant at the 5% level; at the baseline 1% cutoff the three have t-statistics of 2.08, −3.60 and 2.86 and six predictors are significant; the kurtosis proxy loses significance again at the 5% cutoff. The option-implied proxies have heavy-tailed distributions

(Section 3), so some trimming is unavoidable, but a result that appears only after trimming cannot be given much weight. Panel B shows that the lag length of the Newey–West estimator is immaterial: moving from 5 to 22 lags changes the t-statistics by less than 0.3 and the number of significant predictors from four to three.

Table 23. Winsorization and HAC-lag sensitivity (5-day returns).

	R^2 (%)	$t_{\text{impl. skew}}$	$t_{\text{impl. kurt}}$	$t_{\text{PC vol}}$	#sig. (5%)
Panel A: winsorization cutoff (5-day return, HC1 t-statistics)					
None	0.30	0.29	−0.47	0.28	4/10
0.5%	0.28	−1.91	1.48	2.44	4/10
1% (baseline)	0.31	−3.60	2.08	2.86	6/10
2.5%	0.34	−4.66	1.89	3.41	5/10
5%	0.38	−3.50	0.81	4.06	6/10
Panel B: Newey–West lag length (5-day return, baseline winsorization)					
NW(5)		−2.46	1.37	2.20	4/10
NW(10)		−2.37	1.32	2.11	4/10
NW(22)		−2.32	1.32	2.08	3/10

Notes. Pooled 5-day return regression on the full panel (ten predictors). Panel A varies the symmetric winsorization cutoff applied to predictors and target. Panel B varies the Newey–West lag length. The last column counts predictors with $|t| > 1.96$.

6.6. Rank-Based Information Coefficients

Table 24 reports daily cross-sectional Spearman correlations between each predictor and the forward stock return, a nonparametric check that is insensitive to the tails. At the 5-day horizon the largest mean information coefficients are +0.036 for the implied volatility level ($t = 4.78$; positive on 56% of days), +0.025 for the 25-delta skew and −0.020 for net gamma exposure; the kurtosis proxy has a mean IC of −0.019 and the implied skewness proxy −0.018. Two features of these results matter for interpretation. First, no information coefficient exceeds 0.04 in absolute value, which is small by the standards of the equity anomalies literature. Second, the rank correlation of the kurtosis proxy is negative while its pooled regression coefficient is positive, and the put–call open-interest ratio has a positive rank correlation but a negative coefficient on the full panel: for these predictors the parametric estimates are dominated by observations in the tails of the predictor distribution and their sign is not robust. The t-statistics in the table use i.i.d. standard errors, as in version 1.1; with four Newey–West lags the 5-day statistics would be roughly halved. At the daily horizon no mean IC exceeds 0.02 in absolute value.

Table 24. Daily cross-sectional Spearman information coefficients (individual stocks).

	5-day return			1-day return		
Predictor	Mean IC	t	% days > 0	Mean IC	t	% days > 0
$IV_{ATM,30d}$	0.0364	4.78***	56.0	0.0140	1.84*	52.2
$\text{Skew}_{25\delta}$	0.0246	4.41***	53.0	0.0200	3.51***	53.1
PC volume ratio	0.0164	3.85***	53.4	0.0133	3.10***	54.2
RV_t	0.0160	2.55**	51.1	−0.0004	−0.06	50.0
$RV_t^{(w)}$	0.0157	2.36**	52.0	0.0019	0.28	50.1
PC OI ratio	0.0135	2.98***	51.9	0.0071	1.56	51.0
IV term slope	−0.0042	−0.82	48.8	−0.0013	−0.25	49.2
Implied skewness	−0.0178	−3.14***	46.5	−0.0018	−0.32	50.3
Implied kurtosis	−0.0188	−3.21***	46.7	−0.0083	−1.44	49.0
Net gamma exposure	−0.0203	−4.24***	46.9	−0.0107	−2.17**	48.1

Notes. For each trading day with at least 20 stocks, the Spearman rank correlation between the predictor and the forward return is computed across stocks (1,965 days); the table reports the time-series mean, its t-statistic (i.i.d. standard error, as in version 1.1; with 4 Newey–West lags the 5-day statistics would be roughly halved) and the share of days with a positive IC. ***, **, *: $|t| > 2.576, 1.96, 1.645$.

6.7. Decile Sorts

Table 25 repeats the three sorts emphasized in version 1.1 with decile portfolios (Fama and French, 2008). Sharpening the sort does not strengthen the results: the kurtosis-proxy decile spread is −7.0% per year (Sharpe ratio −0.30, $t_{NW} = -1.51$), the put–call volume spread is 5.6% (Sharpe ratio 0.34, $t_{NW} = 2.29$) and the 25-delta skew spread 3.6% (Sharpe ratio 0.14, $t_{NW} = 0.70$). None of the premia is monotone enough in the sorting variable for the extreme deciles to add to the quintile spread.

Table 25. Quintile versus decile long-short sorts (5-day returns).

Sorting variable	Scheme	Ann. ret. (%)	Sharpe	t_{NW}
Implied kurtosis	Q5−Q1	−6.12	−0.334	−1.66*
	D10−D1	−7.02	−0.298	−1.51
PC volume ratio	Q5−Q1	5.06	0.416	2.64***
	D10−D1	5.55	0.339	2.29**
$\text{Skew}_{25\delta}$	Q5−Q1	5.86	0.305	1.54
	D10−D1	3.59	0.136	0.70

Notes. Equal-weighted daily-rebalanced portfolios of individual stocks, 5-day forward returns, annualization with 52 periods; long-short is the top minus the bottom portfolio. t_{NW}: Newey–West with 4 lags. ***, **, *: $|t| > 2.576, 1.96, 1.645$.

6.8. Temporal Stability and a Placebo Test

In-sample predictability is often concentrated in a few episodes (Welch and Goyal, 2008; Campbell and Thompson, 2008). Table 26 re-estimates the two headline regressions sixteen times, excluding one calendar year at a time. The 5-day return R^2 stays between 0.26% and 0.53%; the largest value is obtained when 2022 is excluded, the smallest when 2023 is excluded, so no single year creates the (small) predictability. The augmented variance model's R^2 stays between 0.458 and 0.465 for every exclusion except 2020, whose removal lowers both the HAR-RV fit (to 0.330) and the augmented fit (to 0.394) while raising the relative improvement to 19.5%; the improvement from the surface lies between +9.2% and +19.5% in all sixteen estimations.

Table 26. Leave-one-year-out panel R^2.

Excluded year	5-day return R^2 (%)	HAR-RV R^2	HAR + IV R^2	ΔR^2 (%)
2010	0.33	0.418	0.461	+10.4
2011	0.37	0.417	0.461	+10.4
2012	0.30	0.419	0.462	+10.3
2013	0.32	0.415	0.458	+10.4
2014	0.29	0.419	0.461	+9.8
2015	0.28	0.422	0.465	+10.2
2016	0.27	0.418	0.460	+10.0
2017	0.36	0.417	0.459	+9.9
2018	0.32	0.425	0.464	+9.2
2019	0.32	0.420	0.464	+10.3
2020	0.27	0.330	0.394	+19.5
2021	0.30	0.416	0.460	+10.5
2022	0.53	0.420	0.460	+9.6
2023	0.26	0.423	0.465	+9.9
2024	0.32	0.420	0.462	+10.1
2025	0.28	0.418	0.462	+10.4

Notes. Each row re-estimates the pooled 5-day return regression (ten predictors, HC1) and the two realized-variance models for RV_{t+1} (common sample, per-ticker winsorization) on the full panel excluding the indicated calendar year.

Finally, a placebo experiment permutes the block of option-implied features within each ticker, preserving the cross-feature correlation but destroying the alignment with dates, and re-estimates the 5-day regression. Across ten seeded draws the placebo R^2 averages 0.04% (maximum 0.05%) against an actual R^2 of 0.31%. The actual fit is thus eight times the fit obtainable from the panel structure alone and is statistically different from zero—with 119,081 observations an R^2 of 0.31% corresponds to an F-statistic of about 37—but the placebo also calibrates what "small" means here: the predictors move the explained share of weekly return variation from one twentieth of a percent to one third of a percent.

7. Discussion

7.1. Economic Interpretation

Option prices forecast second moments, not first moments. The contrast between the two halves of the paper is the main finding. The same ten features that explain a third of a percent of weekly return variation raise the explained share of next-day realized variance by a tenth and of weekly

realized variance by a fifth, Granger-cause realized variance in every ticker and are themselves barely affected by realized-variance shocks. This asymmetry is what theory predicts. An option price is a risk-neutral expectation of a payoff that depends on the path of the underlying, so implied volatility is a forecast of variance plus a risk premium (Bollerslev et al., 2009; Carr and Wu, 2009); it is informative about returns only through the risk premium, which is small relative to the noise in weekly returns, or through informed trading, which is a limits-to-arbitrage phenomenon that should be weakest in the most liquid stocks (Johnson and So, 2012; Muravyev et al., 2013, 2025). The universe studied here is that segment, and the results are consistent with the view that for large, heavily traded companies the options market and the stock market price the same information at nearly the same time.

The implied volatility level and the cross-section of large-stock returns. The one portfolio strategy that survives transaction costs is long high-implied-volatility and short low-implied-volatility stocks. In the full cross-section, Ang et al. (2006) document the opposite pattern for idiosyncratic and total volatility. The two findings are not inconsistent: the low-volatility anomaly is concentrated among small, illiquid and hard-to-short stocks, whereas among 49 of the largest US companies over 2010–2025 the high-volatility names were predominantly large technology firms whose returns were high over this particular period. The quantile regressions show that the implied volatility level mainly predicts the dispersion of future returns, and its effect on the mean has a t-statistic of 2.2 after costs and no factor adjustment. It is best regarded as a period-specific pattern rather than as evidence of information in option prices, and the lower bounds on expected returns derived by Martin (2017) and Martin and Wagner (2019), which imply a positive relation between risk-neutral variance and expected returns, provide at most a partial rationale.

The sign of the skew effect. The 25-delta skew is the only return predictor that is significant under two-way clustering and in the Fama–MacBeth regressions, with a positive sign: stocks whose out-of-the-money puts are expensive relative to calls earn higher returns over the following week. This is the sign found by Conrad et al. (2013) for ex ante skewness and the opposite of the one found by Xing et al. (2010), Rehman and Vilkov (2012) and Stilger et al. (2017). The informed-trading mechanism of Xing et al. (2010) requires that put buying by informed traders steepens the smirk before bad news reaches the stock; among the largest companies, with heavy analyst coverage and deep stock liquidity, that channel is weak, and what remains is the compensation for bearing the skewness risk that the smirk measures. The effect is small, and the corresponding sort is not profitable, so the paper does not claim to resolve the disagreement in the literature; it documents that in the liquid segment the risk-compensation sign prevails.

Information in the surface for volatility: in sample versus out of sample. In sample the surface adds to HAR-RV in every subperiod and every leave-one-year-out estimation, and in the weekly horizon the level of implied volatility alone outperforms HAR-RV, as in Christensen and Prabhala (1998), Blair et al. (2001) and Taylor et al. (2010). Out of sample the gain is loss-function dependent: it is reliable under MSE and absent, or negative, under QLIKE. Patton (2011) shows that QLIKE is the more robust loss when the realized proxy is noisy, so a reader who follows that recommendation should conclude that the augmented eight-parameter model does not improve on HAR-RV for next-day variance, while the two-parameter IV-only model does improve on it at the weekly horizon under both losses. The variance risk premium embedded in implied volatility is a plausible reason: implied variance exceeds expected variance on average, and a model that does not correct for the premium (Prokopczuk and Wese Simen, 2014; Bekaert and Hoerova, 2014) inherits a time-varying bias that QLIKE penalizes. Richer HAR specifications with measurement-error corrections (Bollerslev et al., 2016) or semivariances (Patton and Sheppard, 2015) might absorb part of the gain attributed here to the surface; they were not estimated.

Implied correlation as a stress indicator. A high implied correlation relative to realized correlation signals that option markets price a common shock that the recent return history does not yet show. The ratio predicts stress episodes beyond the information in the VIX level, with t-statistics between 2.3 and 4.1 after accounting for the overlap of the forward indicator, and the decomposition into components shows that both a high implied and a low realized correlation contribute. The incremental explanatory power is small because the stress indicator is itself a function of the VIX, and a cleaner test would define stress from realized market returns or drawdowns; the finding is nonetheless in line with the evidence that correlation risk is priced (Driessen et al., 2009; Faria et al., 2022) and that implied correlation rises in stress (Buss and Vilkov, 2012).

Price magnets. The absence of attraction toward the maximum-open-interest strike on ordinary days is consistent with the mechanism of Avellaneda and Lipkin (2003) and the evidence of Ni et al. (2005) and Golez and Jackwerth (2012), in which delta-hedging flows become strong enough to move the underlying only in the last hours before expiration, when gamma is concentrated at the money. The dealer-gamma effects documented by Baltussen et al. (2021) and Barbon and Buraschi (2021) operate on market-wide intraday momentum and fragility, which the present test does not address.

7.2. Relation to the Literature on Option-Implied Predictability

The results agree with the two most robust regularities of the literature surveyed in Section 2: implied volatility is an efficient but biased forecast of future variance, and its forecasting value

relative to realized-variance models rises with the horizon. They are more cautious than most of the cross-sectional return literature, for reasons the design makes explicit. The universe is restricted to the most liquid names, the inference accounts for overlap, cross-sectional dependence and multiple testing, and economic significance is measured net of realized turnover. Under these conditions the weekly return predictability documented by Xing et al. (2010), Cremers and Weinbaum (2010) or An et al. (2014) for the full cross-section shrinks to a few basis points per standard deviation, in the same way that many anomalies shrink in the replications of Hou et al. (2020) and after the trading costs of Novy-Marx and Velikov (2016). The results are also a reminder, illustrated by the history of this paper itself, that overlapping multi-day targets must be aligned and tested with care (Hansen and Hodrick, 1980): an error of four days in the window of a weekly return turned a third of a percent of explained variation into five percent and a Sharpe ratio of -0.3 into one of 2.3.

7.3. Limitations

Several limitations bound the conclusions. The universe is a fixed list of 69 liquid underlyings chosen ex post, so the stock results carry a survivorship bias and do not speak to the small-capitalization segment where the literature locates most option-implied return predictability. The "implied skewness" and "implied kurtosis" variables are smile-based proxies, not the model-free risk-neutral moments of Bakshi et al. (2003); the sensitivity of the results to winsorization suggests that the proxies are noisy in the wings. The long-short portfolios are equal-weighted and not adjusted for factor exposures, and transaction costs are proportional assumptions rather than measured spreads; the strategies' net performance is indicative. The realized-variance forecasts use least squares on levels and do not separate the variance risk premium from the expected-variance component of implied volatility. The stress indicator of RQ3 is defined from the VIX, which is also a regressor. The RQ4 Greeks-decay results and the RQ5 machine-learning results are frozen outputs of the original run that could not be re-executed because the raw option and intraday stores are no longer available; their 1-day components are unaffected by the alignment error, but they could not be extended or re-tuned. Finally, the sample covers a single market and sixteen years dominated by a long equity bull market, and the implied correlation series starts only in 2016.

8. Conclusion

This paper measures the information content of the implied volatility surface for returns and realized variance in a panel of 69 liquid US underlyings over 2010–2025, built from 3.83 billion option records and 5-minute prices. Five conclusions emerge.

First, option-implied features predict the cross-section of weekly returns in a statistically detectable but economically marginal way. They explain 0.3% of the variation in 5-day stock returns,

0.6% for ETFs and 1.0% for indices; under two-way clustered standard errors only the 25-delta skew remains significant, with a positive sign that points to compensation for skewness risk rather than to informed trading; and of fourteen long-short strategies only the sort on the implied volatility level is profitable after a 10 basis-point cost, with a t-statistic that does not survive a multiple-testing correction. Daily predictability is absent.

Second, the surface improves realized-variance forecasts. On a common sample the IV surface raises the in-sample R^2 of HAR-RV from 0.415 to 0.457 for next-day variance and from 0.421 to 0.515 for weekly variance, in every one of nine subperiods and every leave-one-year-out estimation. Out of sample the gain is reliable under mean squared error but not under QLIKE, and at the weekly horizon the level of implied volatility alone is the best forecaster under both losses.

Third, the ratio of implied to realized correlation among S&P 500 constituents predicts stress episodes at 5- to 20-day horizons beyond the VIX level, with overlap-robust t-statistics of 2.3 to 4.1.

Fourth, at-the-money implied volatility Granger-causes realized variance in every ticker, but implied-volatility shocks account for 20 to 29 percent of the 20-day forecast-error variance of realized variance, depending on the recursive ordering; realized variance is driven mainly by its own innovations. Information in bucket-level Greeks for next-day variance rises with time to expiration, and open-interest concentration does not attract the underlying price on ordinary trading days.

Fifth, tree-based learners trained on the SPX surface do not outperform HAR-RV out of sample over 2020–2025, and the surface features they rank highest are the shortest at-the-money tenors.

The revision of this paper corrected an alignment error in the construction of multi-day targets that had produced weekly return R^2 values of 5–19% and a long-short Sharpe ratio of 2.33 in the previous version. Those results are withdrawn. Future work should extend the analysis to a point-in-time universe including smaller stocks, where the limits-to-arbitrage literature predicts stronger option-implied return predictability; replace the smile proxies with model-free risk-neutral moments; separate the variance risk premium from expected variance in the forecasting models and evaluate them with the model confidence set of Hansen et al. (2011); and adjust the implied-volatility strategy for factor exposures before interpreting its net return.

References

Acharya, Viral V., Lasse H. Pedersen, Thomas Philippon, and Matthew Richardson, "Measuring systemic risk," *Review of Financial Studies*, 2017, *30* (1), 2–47.

An, Byeong-Je, Andrew Ang, Turan G. Bali, and Nusret Cakici, "The joint cross section of stocks and options," *Journal of Finance*, 2014, *69* (5), 2279–2337.

Andersen, Torben G. and Tim Bollerslev, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 1998, *39* (4), 885–905.

_ , _ , Francis X. Diebold, and Heiko Ebens, "The distribution of realized stock return volatility," *Journal of Financial Economics*, 2001, *61* (1), 43–76.

_ , _ , _ , and Paul Labys, "Modeling and forecasting realized volatility," *Econometrica*, 2003, *71* (2), 579–625.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, "The cross-section of volatility and expected returns," *Journal of Finance*, 2006, *61* (1), 259–299.

Avellaneda, Marco and Michael D. Lipkin, "A market-induced mechanism for stock pinning," *Quantitative Finance*, 2003, *3* (6), 417–425.

Bakshi, Gurdip and Dilip Madan, "Spanning and derivative-security valuation," *Journal of Financial Economics*, 2000, *55* (2), 205–238.

_ , Nikunj Kapadia, and Dilip Madan, "Stock return characteristics, skew laws, and the differential pricing of individual equity options," *Review of Financial Studies*, 2003, *16* (1), 101–143.

Bali, Turan G. and Armen Hovakimian, "Volatility spreads and expected stock returns," *Management Science*, 2009, *55* (11), 1797–1812.

_ and Scott Murray, "Does risk-neutral skewness predict the cross-section of equity option portfolio returns?," *Journal of Financial and Quantitative Analysis*, 2013, *48* (4), 1145–1171.

_ , Jianfeng Hu, and Scott Murray, "Option implied volatility, skewness, and kurtosis and the cross-section of expected stock returns," 2019. Working paper, Georgetown University. SSRN 2322945.

Baltussen, Guido, Zhi Da, Sten Lammers, and Martin Martens, "Hedging demand and market intraday momentum," *Journal of Financial Economics*, 2021, *142* (1), 377–403.

Barbon, Andrea and Andrea Buraschi, "Gamma fragility," 2021. Working paper, University of St. Gallen and Imperial College London. SSRN 3725454.

Bekaert, Geert and Marie Hoerova, "The VIX, the variance premium and stock market volatility," *Journal of Econometrics*, 2014, *183* (2), 181–192.

Black, Fischer, "Fact and fantasy in the use of options," *Financial Analysts Journal*, 1975, *31* (4), 36–41.

_ **and Myron Scholes**, "The pricing of options and corporate liabilities," *Journal of Political Economy*, 1973, *81* (3), 637–654.

Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor, "Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns," *Journal of Econometrics*, 2001, *105* (1), 5–26.

Bollerslev, Tim, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 1986, *31* (3), 307–327.

_ **and Viktor Todorov**, "Tails, fears, and risk premia," *Journal of Finance*, 2011, *66* (6), 2165–2211.

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

_ **, George Tauchen, and Hao Zhou**, "Expected stock returns and variance risk premia," *Review of Financial Studies*, 2009, *22* (11), 4463–4492.

Borochin, Paul, Hao Chang, and Yangru Wu, "The information content of the term structure of risk-neutral skewness," *Journal of Empirical Finance*, 2020, *58*, 247–274.

Britten-Jones, Mark and Anthony Neuberger, "Option prices, implied price processes, and stochastic volatility," *Journal of Finance*, 2000, *55* (2), 839–866.

Bucci, Andrea, "Realized volatility forecasting with neural networks," *Journal of Financial Econometrics*, 2020, *18* (3), 502–531.

Buraschi, Andrea, Fabio Trojani, and Andrea Vedolin, "When uncertainty blows in the orchard: Comovement and equilibrium volatility risk premia," *Journal of Finance*, 2014, *69* (1), 101–137.

Busch, Thomas, Bent Jesper Christensen, and Morten Ørregaard Nielsen, "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," *Journal of Econometrics*, 2011, *160* (1), 48–57.

Buss, Adrian and Grigory Vilkov, "Measuring equity risk with option-implied correlations," *Review of Financial Studies*, 2012, *25* (10), 3113–3140.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2), 238–249.

Campbell, John Y. and Samuel B. Thompson, "Predicting excess stock returns out of sample: Can anything beat the historical average?," *Review of Financial Studies*, 2008, *21* (4), 1509–1531.

Cao, Charles, Zhiwu Chen, and John M. Griffin, "Informational content of option volume prior to takeovers," *Journal of Business*, 2005, *78* (3), 1073–1109.

Carr, Peter and Liuren Wu, "Variance risk premiums," *Review of Financial Studies*, 2009, *22* (3), 1311–1341.

Chakravarty, Sugato, Huseyin Gulen, and Stewart Mayhew, "Informed trading in stock and option markets," *Journal of Finance*, 2004, *59* (3), 1235–1257.

Chordia, Tarun, Amit Goyal, and Alessio Saretto, "Anomalies and false rejections," *Review of Financial Studies*, 2020, *33* (5), 2134–2179.

Christensen, Bent Jesper and Nagpurnanand R. Prabhala, "The relation between implied and realized volatility," *Journal of Financial Economics*, 1998, *50* (2), 125–150.

Christensen, Kim, Mathias Siggaard, and Bezirgen Veliyev, "A machine learning approach to volatility forecasting," *Journal of Financial Econometrics*, 2023, *21* (5), 1680–1727.

Christoffersen, Peter, Kris Jacobs, and Bo Young Chang, "Forecasting with option-implied information," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2A, Amsterdam: Elsevier, 2013, pp. 581–656.

Clark, Todd E. and Kenneth D. West, "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, 2007, *138* (1), 291–311.

Conrad, Jennifer, Robert F. Dittmar, and Eric Ghysels, "Ex ante skewness and expected stock returns," *Journal of Finance*, 2013, *68* (1), 85–124.

Corsi, Fulvio, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 2009, *7* (2), 174–196.

Cremers, Martijn and David Weinbaum, "Deviations from put-call parity and stock return predictability," *Journal of Financial and Quantitative Analysis*, 2010, *45* (2), 335–367.

Dennis, Patrick and Stewart Mayhew, "Risk-neutral skewness: Evidence from stock options," *Journal of Financial and Quantitative Analysis*, 2002, *37* (3), 471–493.

Diebold, Francis X. and Roberto S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, 1995, *13* (3), 253–263.

Driessen, Joost, Pascal J. Maenhout, and Grigory Vilkov, "The price of correlation risk: Evidence from equity options," *Journal of Finance*, 2009, *64* (3), 1377–1406.

Easley, David, Maureen O'Hara, and P. S. Srinivas, "Option volume and stock prices: Evidence on where informed traders trade," *Journal of Finance*, 1998, *53* (2), 431–465.

Fama, Eugene F. and James D. MacBeth, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 1973, *81* (3), 607–636.

_ **and Kenneth R. French**, "Dissecting anomalies," *Journal of Finance*, 2008, *63* (4), 1653–1678.

Faria, Gonçalo, Robert Kosowski, and Tianyu Wang, "The correlation risk premium: International evidence," *Journal of Banking & Finance*, 2022, *136*, 106399.

Ge, Li, Tse-Chun Lin, and Neil D. Pearson, "Why does the option to stock volume ratio predict stock returns?," *Journal of Financial Economics*, 2016, *120* (3), 601–622.

Giacomini, Raffaella and Halbert White, "Tests of conditional predictive ability," *Econometrica*, 2006, *74* (6), 1545–1578.

Golez, Benjamin and Jens Carsten Jackwerth, "Pinning in the S&P 500 futures," *Journal of Financial Economics*, 2012, *106* (3), 566–585.

Granger, Clive W. J., "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 1969, *37* (3), 424–438.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Empirical asset pricing via machine learning," *Review of Financial Studies*, 2020, *33* (5), 2223–2273.

Hansen, Lars Peter and Robert J. Hodrick, "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis," *Journal of Political Economy*, 1980, *88* (5), 829–853.

Hansen, Peter R. and Asger Lunde, "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?," *Journal of Applied Econometrics*, 2005, *20* (7), 873–889.

_ **,** _ **, and James M. Nason**, "The model confidence set," *Econometrica*, 2011, *79* (2), 453–497.

_ , Zhuo Huang, and Howard H. Shek, "Realized GARCH: A joint model for returns and realized measures of volatility," *Journal of Applied Econometrics*, 2012, *27* (6), 877–906.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, ". . . and the cross-section of expected returns," *Review of Financial Studies*, 2016, *29* (1), 5–68.

Hou, Kewei, Chen Xue, and Lu Zhang, "Replicating anomalies," *Review of Financial Studies*, 2020, *33* (5), 2019–2133.

Hu, Jianfeng, "Does option trading convey stock price information?," *Journal of Financial Economics*, 2014, *111* (3), 625–645.

Jegadeesh, Narasimhan and Sheridan Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 1993, *48* (1), 65–91.

Jiang, George J. and Yisong S. Tian, "The model-free implied volatility and its information content," *Review of Financial Studies*, 2005, *18* (4), 1305–1342.

Johnson, Travis L. and Eric C. So, "The option to stock volume ratio and future returns," *Journal of Financial Economics*, 2012, *106* (2), 262–286.

Kelly, Bryan and Hao Jiang, "Tail risk and asset prices," *Review of Financial Studies*, 2014, *27* (10), 2841–2871.

Kilian, Lutz and Helmut Lütkepohl, *Structural Vector Autoregressive Analysis*, Cambridge: Cambridge University Press, 2017.

Koenker, Roger and Gilbert Bassett, "Regression quantiles," *Econometrica*, 1978, *46* (1), 33–50.

Kourtis, Apostolos, Raphael N. Markellos, and Lazaros Symeonidis, "An international comparison of implied, realized, and GARCH volatility forecasts," *Journal of Futures Markets*, 2016, *36* (12), 1164–1193.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard, "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes," *Journal of Econometrics*, 2015, *187* (1), 293–311.

Lo, Andrew W., "The statistics of Sharpe ratios," *Financial Analysts Journal*, 2002, *58* (4), 36–52.

Manaster, Steven and Richard J. Rendleman, "Option prices as predictors of equilibrium stock prices," *Journal of Finance*, 1982, *37* (4), 1043–1057.

Martin, Ian, "What is the expected return on the market?," *Quarterly Journal of Economics*, 2017, *132* (1), 367–433.

Martin, Ian W. R. and Christian Wagner, "What is the expected return on a stock?," *Journal of Finance*, 2019, *74* (4), 1887–1929.

McLean, R. David and Jeffrey Pontiff, "Does academic research destroy stock return predictability?," *Journal of Finance*, 2016, *71* (1), 5–32.

Mixon, Scott, "The implied volatility term structure of stock index options," *Journal of Empirical Finance*, 2007, *14* (3), 333–354.

Muravyev, Dmitriy, Neil D. Pearson, and John Paul Broussard, "Is there price discovery in equity options?," *Journal of Financial Economics*, 2013, *107* (2), 259–283.

_ , _ , and Joshua M. Pollet, "Why does options market information predict stock returns?," *Journal of Financial Economics*, 2025, *172*, 104153.

Newey, Whitney K. and Kenneth D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 1987, *55* (3), 703–708.

Ni, Sophie X., Jun Pan, and Allen M. Poteshman, "Volatility information trading in the option market," *Journal of Finance*, 2008, *63* (3), 1059–1091.

Ni, Sophie Xiaoyan, Neil D. Pearson, Allen M. Poteshman, and Joshua White, "Does option trading have a pervasive impact on underlying stock prices?," *Review of Financial Studies*, 2021, *34* (4), 1952–1986.

_ , _ , and _ , "Stock price clustering on option expiration dates," *Journal of Financial Economics*, 2005, *78* (1), 49–87.

Novy-Marx, Robert and Mihail Velikov, "A taxonomy of anomalies and their trading costs," *Review of Financial Studies*, 2016, *29* (1), 104–147.

Pan, Jun and Allen M. Poteshman, "The information in option volume for future stock prices," *Review of Financial Studies*, 2006, *19* (3), 871–908.

Patton, Andrew J., "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 2011, *160* (1), 246–256.

_ and Kevin Sheppard, "Good volatility, bad volatility: Signed jumps and the persistence of volatility," *Review of Economics and Statistics*, 2015, *97* (3), 683–697.

Petersen, Mitchell A., "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies*, 2009, *22* (1), 435–480.

Pollet, Joshua M. and Mungo Wilson, "Average correlation and stock market returns," *Journal of Financial Economics*, 2010, *96* (3), 364–380.

Poon, Ser-Huang and Clive W. J. Granger, "Forecasting volatility in financial markets: A review," *Journal of Economic Literature*, 2003, *41* (2), 478–539.

Prokopczuk, Marcel and Chardin Wese Simen, "The importance of the volatility risk premium for volatility forecasting," *Journal of Banking & Finance*, 2014, *40*, 303–320.

Rehman, Zahid and Grigory Vilkov, "Risk-neutral skewness: Return predictability and its sources," 2012. Working paper, Goethe University Frankfurt. SSRN 1301648.

Sims, Christopher A., "Macroeconomics and reality," *Econometrica*, 1980, *48* (1), 1–48.

Skintzi, Vasiliki D. and Apostolos-Paul N. Refenes, "Implied correlation index: A new measure of diversification," *Journal of Futures Markets*, 2005, *25* (2), 171–197.

Stilger, Przemysław S., Alexandros Kostakis, and Ser-Huang Poon, "What does risk-neutral skewness tell us about future stock returns?," *Management Science*, 2017, *63* (6), 1814–1834.

Taylor, Stephen J., Pradeep K. Yadav, and Yuanyuan Zhang, "The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks," *Journal of Banking & Finance*, 2010, *34* (4), 871–881.

Vasquez, Aurelio, "Equity volatility term structures and the cross section of option returns," *Journal of Financial and Quantitative Analysis*, 2017, *52* (6), 2727–2754.

Welch, Ivo and Amit Goyal, "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 2008, *21* (4), 1455–1508.

White, Halbert, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 1980, *48* (4), 817–838.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao, "What does the individual option volatility smirk tell us about future equity returns?," *Journal of Financial and Quantitative Analysis*, 2010, *45* (3), 641–662.

A. Analysis Universe

The 49 stocks are AAPL, MSFT, AMZN, GOOG, GOOGL, TSLA, NVDA, META, JPM, BAC, WFC, GS, JNJ, UNH, PFE, XOM, CVX, HD, PG, KO, DIS, NFLX, INTC, AMD, CRM, COST, ABBV, MRK, TMO, ABT, V, MA, BRK_B, LLY, AVGO, ADBE, CSCO, PEP, WMT, MCD, TXN, QCOM, LOW, UPS, CAT, GE, BA, RTX, HON, LMT. The 17 ETFs are SPY, QQQ, IWM, DIA, TLT, GLD, XLF, XLE, XLK, XLV, XLI, XLP, XLU, XLB, XLC, XLRE, XLY. The three indices are SPX, NDX, RUT. The 30 constituents used for the implied-correlation index (RQ3) are AAPL, MSFT, AMZN, GOOG, GOOGL, TSLA, NVDA, META, JPM, BAC, WFC, GS, JNJ, UNH, PFE, XOM, CVX, HD, PG, KO, DIS, NFLX, INTC, AMD, CRM, V, MA, LLY, AVGO, ADBE; the ten underlyings of the Greeks and price-magnet analyses (RQ4) are AAPL, MSFT, AMZN, GOOG, TSLA, NVDA, META, JPM, SPY, QQQ. Lists are taken from `src/wp7/config.py`; BRK_B denotes Berkshire Hathaway class B.

B. Implied Correlation Around Market Events

Table A1 reports the implied correlation of the 30 constituents, the realized correlation where available, and the VIX before and after the nine market events used in Section 5. Implied correlation rises after every event; the largest increases follow the February 2018 volatility spike (+0.328) and the August 2015 devaluation of the renminbi (+0.258), and the smallest the June 2022 rate shock (+0.030) and the March 2023 regional-bank episode (+0.045), two events that were anticipated by a VIX already above 20.

Table A1. Implied and realized correlation around major market events.

	Implied correlation			Realized corr.		VIX	
Event	Pre	Post	Δ	Pre	Post	Pre	Post
Flash Crash (May 2010)	0.223	0.404	+0.181	–	–	17.7	28.1
Euro crisis (Aug. 2011)	0.313	0.468	+0.156	–	–	20.3	40.3
China devaluation (Aug. 2015)	0.231	0.489	+0.258	–	–	14.3	30.6
Volmageddon (Feb. 2018)	0.108	0.436	+0.327	0.257	0.638	11.9	28.4
COVID-19 (Mar. 2020)	0.473	0.648	+0.175	0.611	0.788	36.5	70.0
Meme stocks (Jan. 2021)	0.227	0.326	+0.099	0.160	0.349	23.2	27.7
Rate shock (June 2022)	0.417	0.447	+0.030	0.553	0.610	26.9	31.4
SVB (Mar. 2023)	0.328	0.373	+0.045	0.327	0.371	20.3	24.9
VIX spike (Aug. 2024)	0.138	0.316	+0.178	0.093	0.260	15.6	24.2

Notes. Pre: mean over the 30 calendar days before the event date; Post: mean over the 10 calendar days from the event date. The realized-correlation series starts in February 2016, hence the missing entries for earlier events.

C. Per-Ticker Out-of-Sample Comparisons (RQ2)

Table A2. Per-ticker Diebold–Mariano tests: HAR-RV + IV surface versus HAR-RV.

		RV_{t+1}				$\sum_{j=1}^{5} RV_{t+j}$			
Ticker	T	MSE ratio	DM	QLIKE ratio	DM	MSE ratio	DM	QLIKE ratio	DM
AAPL	2,052	0.985	0.39	1.326	−2.34	0.920	1.35	0.866	2.10
ADBE	885	0.980	0.68	1.738	−5.50	0.997	0.04	1.330	−2.40
AMD	865	0.943	2.56	1.110	−1.40	0.843	3.35	0.916	0.79
AMZN	1,667	0.946	2.02	1.022	−0.23	0.836	2.66	0.926	0.45
BA	1,500	0.824	4.77	1.082	−1.32	0.969	0.26	0.902	1.03
BAC	1,623	1.037	−0.73	1.340	−3.89	1.117	−1.00	1.421	−2.90
CAT	1,419	0.877	2.96	1.043	−0.57	0.957	0.82	0.928	1.10
COST	714	0.882	2.17	1.035	−0.67	0.976	0.40	1.001	−0.01
CRM	1,500	1.004	−0.15	1.159	−2.30	0.900	1.60	0.842	2.01
CSCO	1,410	0.998	0.09	1.250	−3.71	0.989	0.30	1.018	−0.21
CVX	1,175	0.977	0.77	1.273	−1.43	0.968	0.90	1.144	−0.90
DIA	1,922	1.018	−0.32	1.405	−4.66	1.220	−1.69	1.773	−2.76
DIS	1,433	0.949	2.55	1.277	−1.55	0.870	3.85	0.841	2.53
GLD	2,425	0.860	5.26	0.970	0.59	0.859	3.29	1.001	−0.00
GOOG	1,618	0.937	3.20	1.183	−2.13	0.799	3.29	1.097	−0.43
GS	1,120	1.061	−1.06	0.956	0.95	1.006	−0.10	0.918	1.27
HD	1,408	1.046	−1.02	1.107	−2.09	1.037	−0.38	0.947	0.68
HON	371	0.833	2.91	1.347	−2.80	0.754	2.19	1.126	−0.89
INTC	1,673	1.034	−1.22	1.033	−0.59	0.882	1.69	0.721	4.23
IWM	2,482	0.962	1.37	1.406	−2.73	1.313	−2.99	2.176	−4.29
JPM	1,680	0.792	4.25	1.195	−1.29	0.791	2.79	1.002	−0.02
KO	1,222	0.826	3.84	0.985	0.23	0.981	0.47	1.248	−1.31
LLY	1,216	0.986	0.50	1.092	−1.48	0.874	1.52	0.754	2.22
LMT	424	0.909	2.37	2.145	−2.23	0.917	1.43	1.917	−1.83
LOW	626	0.826	2.67	1.089	−0.96	1.046	−0.37	1.581	−4.20
MA	975	1.028	−0.60	1.305	−2.23	0.889	1.73	1.273	−0.96
MCD	1,128	0.850	3.35	1.242	−2.56	0.847	2.49	0.883	1.02
MRK	1,398	1.045	−1.55	1.277	−2.82	0.977	0.26	1.024	−0.32
MSFT	1,867	0.974	1.10	1.075	−0.98	0.925	1.75	0.873	2.26
NVDA	1,421	0.971	0.98	1.151	−2.75	0.960	0.81	0.955	0.58

Notes. Ratios are losses of HAR-RV + IV surface over HAR-RV (values below one favour the IV-augmented model); DM statistics are positive when the IV-augmented model has the lower loss. T is the number of out-of-sample forecasts. Rolling scheme as in Table 11.

D. SPX Options Surface Features (RQ5)

The 21 SPX surface features used in the machine-learning comparison are: at-the-money implied volatility at six tenors (1 week, 2 weeks, 1, 2, 3 and 6 months, defined by DTE windows of 5–10, 13–17, 25–35, 55–65, 85–95 and 170–200 days); 25-delta skew at 1 and 3 months; 10-delta skew at 1 month; the 1-month butterfly ratio (average wing IV over ATM IV); two term-structure slopes (3-month minus 1-month and 6-month minus 1-month ATM IV); total gamma-weighted open

interest; net gamma exposure; total vega-weighted open interest; average Θ; the put–call volume and open-interest ratios; total option volume; total open interest; and the average relative bid–ask spread. The two term-structure slopes are exact linear combinations of the tenor-level implied volatilities, which is why the OLS surface models of the original run were numerically degenerate (Section 5); the tree-based learners are unaffected by the collinearity.

E. Random Forest Feature Importance (RQ5)

Table A3. Random forest feature importance (next-day SPX RV, 2010–2019 training).

Feature	Importance
$IV_{ATM,2w}$	0.508
$IV_{ATM,1w}$	0.151
Total option volume	0.083
PC volume ratio	0.067
Net gamma exposure	0.037
$IV_{ATM,1m}$	0.023
$RV^{(w)}$	0.022
Skew (25δ, 3m)	0.013
Average Θ	0.012
$IV_{ATM,2m}$	0.011
All other features (14)	*0.073*

Notes. Mean-decrease-in-impurity importances of the random forest fitted on 2010–2019 (frozen output of `scripts/06`).