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[1]

Département des sciences administratives
Université du Québec en Outaouais

This version: May 2026

Version 1.1

[1] Professeur, Département des sciences administratives, Université du Québec en Outaouais (UQO), 283 boulevard Alexandre-Taché, Gatineau, QC J9A 1L8, Canada. Email: simon-pierre.boucher@uqo.ca.

Abstract

This paper investigates the information content embedded in equity option markets for predicting returns and realized volatility across multiple asset classes and market regimes. Using 3.83 billion option contracts on 11,077 underlyings (2010–2025) merged with 11.5 billion intraday OHLCV observations across stocks, ETFs, indices, futures, FX, and cryptocurrencies, I address five research questions. Option-implied moments—particularly implied kurtosis, skewness, and put-call ratios—significantly predict 5-day stock returns; quintile long-short portfolios sorted on implied kurtosis deliver annualized Sharpe ratios of 2.33 ($t = 19.84$). Augmenting the HAR-RV model with implied volatility surface features improves 1-day realized variance forecasting by 23.3% in R^2 (0.359 to 0.442), an improvement robust across all subperiods including COVID-19 (+19.4%). The ratio of implied to realized correlation among S&P 500 constituents predicts market stress at 5, 10, and 20-day horizons (t-statistics: 3.91–8.30). Granger causality confirms that ATM implied volatility leads realized volatility in 100% of individual tickers ($F = 62.4$), and variance decomposition shows IV shocks explain 73.8% of RV forecast error variance at the 20-day horizon. All findings survive Newey-West HAC (up to 22 lags), double-clustered standard errors, subperiod and leave-one-year-out analysis, VIX-regime conditioning, quantile regressions, winsorization sensitivity, rank-based information coefficients, decile sorts, and a within-ticker permutation placebo.

Keywords: Option-implied information, realized volatility, HAR-RV, implied correlation, volatility surface, portfolio sorts, Granger causality, high-frequency data

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

Contents

1. Introduction

Options markets aggregate the expectations of heterogeneous participants regarding future price movements, volatility, and tail risks. Since the seminal work of Black and Scholes (1973), financial economists have recognized that derivative prices embed forward-looking information that is not fully reflected in the underlying asset's spot price. The implied volatility surface—spanning the dimensions of moneyness and time-to-expiration—constitutes a rich, high-dimensional representation of market expectations about the full distribution of future returns. Whether, and how, this information can be exploited for return prediction and volatility forecasting remains a central question in empirical finance.

This paper addresses that question with a dataset of unprecedented scale: 3.83 billion individual option contract records spanning 11,077 underlying tickers over 4,025 trading days (January 2010–December 2025), merged with 11.5 billion intraday OHLCV observations at the 1-minute and 5-minute frequencies across six asset classes. This infrastructure supports five interconnected research questions:

RQ1. Do option-implied moments (variance, skewness, kurtosis) extracted from the strike–expiry surface predict next-day and next-week cross-sectional stock returns? Does this predictability extend to sector ETFs and equity indices?

RQ2. Can the implied volatility term structure and volatility smile jointly forecast realized volatility more accurately than standard time-series models (GARCH, HAR-RV)?

RQ3. Does the divergence between option-implied and realized correlations predict future market stress events?

RQ4. How does the information content of aggregated options Greeks decay as a function of days-to-expiry? Does open-interest concentration at specific strikes create predictable intraday price magnets?

RQ5. Can machine-learning models trained on the full daily options surface outperform the VIX in forecasting the 5-minute realized variance of the S&P 500?

The study advances the existing literature along four dimensions. First, whereas most prior work examines individual stocks or a single index, I provide evidence spanning 69 underlyings across three asset groups (individual stocks, sector ETFs, and broad market indices), documenting a systematic cross-asset gradient in predictability. Second, I assess the *economic significance* of option-implied predictability through portfolio sorts, constructing long-short quintile portfolios and evaluating risk-adjusted performance under realistic transaction costs. Third, I provide the most comprehensive set of robustness checks in this literature: Newey-West HAC standard errors at multiple lag lengths (Newey and West, 1987), double-clustered standard errors (Cameron et al., 2011; Petersen, 2009), Fama-MacBeth regressions (Fama and MacBeth, 1973), subperiod and leave-one-year-out analysis, VIX-regime conditioning, quantile regressions, winsorization sensitivity, nonparametric rank-based information coefficients, decile sorts, ticker-by-ticker R^2 distributions, and a within-ticker permutation placebo test. Fourth, I employ Granger causality tests and bivariate VAR models to establish the direction of information flow between option-implied and realized measures, complemented by impulse response functions and forecast error variance decomposition.

The 16-year sample encompasses the post-GFC recovery (2010–2012), the extended bull market (2013–2016), the low-volatility era of 2017, the Volmageddon of February 2018, the COVID-19 crash of March 2020, the meme-stock episode of January 2021, the 2022 rate-hiking cycle, the SVB crisis of March 2023, and the August 2024 VIX spike—a sequence of natural experiments for regime-dependent analysis.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 describes the data and variable construction. Section 4 presents the methodology. Section 5 reports the results. Section 6 presents robustness tests. Section 7 discusses the findings, and Section 8 concludes.

2. Literature Review

2.1. Information Content of Option Prices

The notion that options markets convey informational advantages relative to the equity market dates back to Black (1975) and Manaster and Rendleman (1982). Easley et al. (1998) develop a sequential-trade model in which option volume signals the presence of informed traders; Pan and Poteshman (2006) extend this framework and find that the informational role is concentrated in out-of-the-money puts. Johnson and So (2012) show that the option-to-stock volume ratio predicts returns more strongly when short-sale constraints bind, and Ge et al. (2016) use order-level data to confirm that non-market-maker option trades carry predictive power for future stock prices. Hu (2014) documents that the information content of option trades has increased over time.

2.2. Option-Implied Moments and Return Predictability

Xing et al. (2010) demonstrate that the steepness of the volatility smirk predicts individual stock returns. Cremers and Weinbaum (2010) show that deviations from put-call parity contain return-predictive information. An et al. (2014) analyze the joint cross section of stocks and options and find incremental predictive power in implied volatility and skewness. Bali et al. (2019) construct comprehensive implied-moment measures and document significant cross-sectional predictability at the weekly horizon. Stilger et al. (2017) examine the implied volatility spread (call IV minus put IV) and document a strong positive relationship with future returns. Cao et al. (2005) provide evidence that the options market leads the equity market in price discovery.

2.3. Implied Volatility and Realized Volatility Forecasting

Christensen and Prabhala (1998) establish that implied volatility is a biased but efficient forecast of realized volatility, a result Blair et al. (2001) confirm using the VIX. The HAR-RV model of Corsi (2009) has become the benchmark for realized volatility forecasting, and Busch et al. (2011) augment it with implied volatility. Bekaert and Hoerova (2014) decompose the VIX into expected volatility and variance-risk-premium components, while Bollerslev et al. (2009) show that the variance risk premium predicts equity returns. Carr and Wu (2009) provide the theoretical foundation for model-free implied variance. Patton and Sheppard (2015) propose the semivariance-based asymmetric HAR-RV, and Hansen et al. (2012) develop the Realized GARCH model.

2.4. Implied Correlation and Systemic Risk

Driessen et al. (2009) formalize the correlation risk premium and show that it is positive. Buss and Vilkov (2012) demonstrate that implied correlation spikes during market stress. Kelly and Jiang (2014) develop tail-risk measures from out-of-the-money puts, and Acharya et al. (2017) propose systemic-risk measures related to the correlation approach adopted here.

2.5. Market Microstructure of Options and Greeks

Ni et al. (2009) document that high open interest at specific strikes affects underlying prices, and Avellaneda and Lipkin (2003) study expiration-day pinning. Barbon and Buraschi (2022) show that aggregate dealer gamma positions predict returns and volatility.

2.6. Machine Learning in Volatility Forecasting

Bucci (2020) compare neural networks with traditional volatility models. Christensen et al. (2023) find that machine-learning methods offer modest improvements that often fail out-of-sample, whereas Gu et al. (2020) show that tree-based methods perform well for cross-sectional return prediction.

3. Data and Variable Construction

3.1. Data Sources

Options. I use a comprehensive end-of-day options database containing 3,831,907,488 records covering 11,077 underlyings from January 4, 2010 to December 31, 2025 (4,025 trading days). Each record includes trade date, strike, expiry, call/put flag, bid/ask prices, bid/ask implied volatilities, open interest, volume, and the full Greek sensitivities (δ, Γ, $\mathcal{V}$, Θ, ρ). Data quality is high: 92–95% of records contain valid Greeks.

Intraday OHLCV. I use 5-minute OHLCV data across six asset classes: 7,789 US equities (1.23B bars), 4,302 ETFs (301M bars), 125 indices, 131 futures, 78 FX pairs, and 74 cryptocurrencies. Total: approximately 11.5 billion intraday observations.

Table 1. Data coverage.

Asset class	Symbols	Rows (millions)	Freq.	Period
Panel A: Options				
Option contracts	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**		

3.2. Sample Construction

The analysis sample consists of 69 underlyings with overlapping options and OHLCV coverage: 49 major US stocks, 17 ETFs, and 3 broad market indices (SPX, NDX, RUT). After merging option-derived features with realized volatility, the sample comprises **264,383 ticker-day observations** (January 2010–December 2025).

3.3. Variable Definitions

3.3.1. Option-Implied Features

I construct ten daily features for each underlying from the raw option chain:

Table 2. Option-implied variables.

Variable	Definition
$IV_{ATM,30d}$	Average mid-IV of calls with $\lvert\delta - 0.5\rvert < 0.10$, 20–40 DTE
$IV_{ATM,90d}$	Same, 80–100 DTE
IV term slope	$IV_{ATM,90d} - IV_{ATM,30d}$
Skew$_{25\delta}$	$IV_{25\delta P,30d} - IV_{25\delta C,30d}$
Implied skewness	$(IV_{10\delta P} - IV_{10\delta C})/IV_{ATM}$
Implied kurtosis	$\overline{IV}_{wings}/IV_{ATM}$, with $0.03 < \lvert\delta\rvert < 0.15$
PC vol. ratio	$\sum V^{put}/\sum V^{call}$
PC OI ratio	$\sum OI^{put}/\sum OI^{call}$
Net gamma exp.	$\sum \Gamma^{call} \cdot OI^{call} - \sum \Gamma^{put} \cdot OI^{put}$

3.3.2. Realized Volatility Measures

Following Andersen et al. (2003), daily realized variance is:

$$RV_t = \sum_{i=1}^{N_t} r_{t,i}^2, \qquad r_{t,i} = \ln\bigl(P_{t,i} \,/\, P_{t,i-1}\bigr). \tag{1}$$

I also compute the weekly and monthly rolling averages $RV_t^{(w)}$ and $RV_t^{(m)}$, as well as forward targets RV_{t+1} (1-day) and $\sum_{j=1}^{5} RV_{t+j}$ (5-day).

3.4. Descriptive Statistics

Table 3 presents summary statistics by asset group.

Table 3. Descriptive statistics by asset group.

	Stocks	ETFs	Indices	All
N (ticker-days)	188093	64256	12034	264383
N tickers	49	17	3	69
Mean $IV_{ATM,30d}$	0.2664	0.1784	0.1823	0.2406
Mean skew$_{25\delta}$	0.0361	0.0425	0.0562	0.0387
Mean PC vol. ratio	0.890	2.345	1.754	1.283
Mean RV (daily)	0.0005	0.0002	0.0001	0.0004
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} is the 30-day at-the-money call implied volatility. Skew is the 25-delta put minus 25-delta call IV. PC vol. ratio is total put volume divided by total call volume. RV is realized variance from 5-minute returns.

Stocks exhibit higher ATM implied volatility (26.6%) and return volatility (1.90% daily) than ETFs (17.8%, 1.25%) and indices (18.2%, 1.29%). The put-call volume ratio is notably higher for ETFs (2.35) and indices (1.75) than for stocks (0.89), consistent with institutional portfolio hedging.

Table 4 reports the average autocorrelation structure. ATM implied volatility is highly persistent (lag-1: 0.959, lag-22: 0.538), while daily returns exhibit slight negative autocorrelation (−0.062).

Table 4. Average autocorrelation structure.

Variable	Lag 1	Lag 5	Lag 10	Lag 22
ATM IV (30d)	0.959	0.861	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 vol. ratio	0.299	0.218	0.172	0.118

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

4. Methodology

4.1. Return Predictability (RQ1)

I estimate pooled panel regressions:

$$r_{i,t+h} = \alpha + \boldsymbol{\beta}' \mathbf{X}_{i,t}^{opt} + \gamma\, RV_{i,t} + \delta\, RV_{i,t}^{(w)} + \varepsilon_{i,t+h}, \qquad h \in \{1, 5\} \text{ days}, \tag{2}$$

with heteroskedasticity-robust (HC1) standard errors (White, 1980). All variables are winsorized at the 1st and 99th percentiles and standardized. Robustness uses Newey-West HAC(5) (Newey and West, 1987), double-clustered (ticker + date) standard errors (Cameron et al., 2011; Petersen, 2009), Fama-MacBeth cross-sectional regressions, quantile regressions (Koenker and Bassett, 1978), rank-based information coefficients, decile sorts, leave-one-year-out estimation, and a within-ticker permutation placebo.

For economic significance, I construct daily-rebalanced equal-weighted quintile portfolios by cross-sectional sorting on each option-implied variable and evaluate annualized Sharpe ratios.

4.2. RV Forecasting (RQ2)

I compare five nested models for RV_{t+1}:

$$\text{GARCH proxy:} \quad RV_{t+1} = \alpha + \beta_1\, RV_t + \beta_2\, r_t^2 + \varepsilon_{t+1} \tag{3}$$

$$\text{HAR-RV:} \quad RV_{t+1} = \alpha + \beta_d\, RV_t + \beta_w\, RV_t^{(w)} + \beta_m\, RV_t^{(m)} + \varepsilon_{t+1} \tag{4}$$

with IV-only, IV-surface, and HAR-RV + IV-surface extensions. Out-of-sample evaluation uses rolling windows (500 training, 250 test) with MSE, MAE, R^2_{OOS}, and QLIKE loss. Model comparison uses the Diebold and Mariano (1995) test.

4.3. Implied Correlation (RQ3)

Following the CBOE methodology:

$$\hat{\rho}_{impl,t} = \frac{\sigma^2_{SPX,t} - n^{-1}\,\overline{\sigma^2_{i,t}}}{\left(1 - n^{-1}\right)\bar{\sigma}^2_{i,t}}, \tag{5}$$

where $\sigma_{SPX,t}$ is 30-day ATM IV of SPX, $\overline{\sigma^2_{i,t}}$ is the mean squared ATM IV of 30 constituents, and $\bar{\sigma}_{i,t}$ is the mean ATM IV. Realized correlation uses the 22-day rolling pairwise matrix. Stress is defined as VIX > 25 or 5-day VIX change > 20%.

4.4. Greeks Information Decay and Price Magnets (RQ4)

I partition options into six DTE buckets (1w, 2w, 1m, 2m, 3m, 6m) and estimate predictive regressions within each. For the price magnet test:

$$\text{Magnet}_{i,t} = \mathbf{1}\left[\frac{|P_{close,t} - K_t^*|}{P_{close,t}} < \frac{|P_{open,t} - K_t^*|}{P_{open,t}}\right], \tag{6}$$

where K_t^* is the max-OI strike. Under the null, $\Pr(\text{Magnet} = 1) = 0.5$.

4.5. Machine Learning (RQ5)

I construct 21 daily features from the SPX surface (listed in Appendix B) and train Random Forest (200 trees, max depth 10) and Gradient Boosting (200 trees, learning rate 0.05) on 2010–2019, testing on 2020–2025. The benchmark is $\widehat{RV}_t^{VIX} = (VIX_t/100)^2/252$.

4.6. Granger Causality and VAR

Bivariate Granger F-tests (5 lags) estimated per ticker. Bivariate VAR(5) models yield impulse response functions (IRFs) and forecast error variance decomposition (FEVD).

5. Results

5.1. RQ1: Return Predictability

5.1.1. Panel Regressions

Table 5 reports pooled OLS results. Predictive power is dramatically stronger at the 5-day horizon: R^2 increases from 0.08% (stocks, 1-day) to 4.8% (stocks, 5-day), 12.4% (ETFs), and 19.3% (indices).

Table 5. Return predictability: pooled OLS panel regressions.

	1-Day return			5-Day return		
	Stocks	ETFs	Idx	Stocks	ETFs	Idx
R^2	0.0008	0.0014	0.0043	0.0476	0.1242	0.1930
Adj. R^2	0.0007	0.0011	0.0032	0.0475	0.1239	0.1921
N	79943	30152	8986	79943	30152	8986
# sig. (5%)	2	2	0	7	9	9

Notes. Ten predictors: $IV_{ATM,30d}$, IV term slope, skew$_{25\delta}$, implied skewness, implied kurtosis, PC vol. ratio, PC OI ratio, net gamma exposure, RV_t, $RV_t^{(w)}$. HC1 standard errors. All variables winsorized at 1%/99% and standardized.

5.1.2. Portfolio Sorts

Table 6 reports quintile long-short portfolio performance for 5-day stock returns. The implied kurtosis sort delivers an annualized Sharpe of 2.33 ($t = 19.84$). Sorting on the put-call volume ratio yields Sharpe = −2.48: stocks with the highest put-call ratio (bearish sentiment) earn the lowest future returns.

Table 6. Long-short portfolio performance (Q5−Q1, 5-day returns).

Sorting variable	Ann. ret. (%)	Ann. vol. (%)	Sharpe	t-stat	N_{days}
Implied kurtosis	44.04	18.91	2.328	19.843	3,777
IV term slope	11.33	20.30	0.558	4.004	2,676
PC OI ratio	5.27	13.27	0.397	3.492	4,024
ATM IV (30d)	−1.76	26.13	−0.067	−0.575	3,777
Implied skewness	−9.91	18.81	−0.527	−4.493	3,777
Put-call vol. ratio	−30.29	12.22	−2.478	−21.796	4,024
$\text{Skew}_{25\delta}$	−37.53	20.10	−1.868	−15.918	3,778

Notes. Equal-weighted quintile portfolios formed each trading day. Returns are 5-day (weekly). Annualization uses $\sqrt{52}$. t-statistics test H_0: mean daily L/S return = 0.

5.1.3. Double Sorts

Table 7 shows a 3×3 sort on ATM IV and implied skewness. Among high-IV stocks, those with high skewness earn −47.8 bps/week ($t = -6.17$), while those with low skewness earn +41.8 bps/week ($t = 13.12$). The information content of skewness is amplified in high-uncertainty environments.

Table 7. Double sort: ATM IV × implied skewness → 5-day returns (bps/week).

	Low skew	Med.	High skew
Low IV	8.15	24.12	35.90
Medium IV	22.11	23.80	15.62
High IV	41.76	2.72	−47.82
High−Low IV	*33.61*	*−21.40*	*−83.72*

Notes. Tercile breakpoints computed cross-sectionally each day.

5.2. RQ2: IV Surface vs. HAR-RV

Table 8 reports in-sample R^2. The combined HAR-RV + IV-surface model achieves $R^2 = 0.442$ for 1-day RV, a 23.3% improvement over HAR-RV (0.359). The IV-only model ($R^2 = 0.390$) outperforms HAR-RV at the daily horizon.

Table 8. In-sample R^2: realized variance forecasting models.

Model	1-Day RV	5-Day RV	N
GARCH proxy	0.326	0.648	264,380
HAR-RV	0.359	0.888	264,345
IV only	0.390	0.495	120,280
IV surface	0.391	0.496	119,093
HAR-RV + IV surface	0.442	0.896	119,093

Diebold-Mariano tests show statistically significant MSE improvements for AAPL (+10.3%, DM = 2.54), AMD (+11.7%, DM = 3.13), CAT (+32.7%, DM = 4.27), and COST (+24.4%, DM = 2.94).

5.3. RQ3: Implied Correlation and Stress Prediction

Table 9 shows that the correlation *ratio* ($\hat{\rho}_{impl}/\hat{\rho}_{real}$) significantly predicts stress at all horizons (t: 3.91–8.30). The raw divergence is not significant; the multiplicative form captures the signal. Appendix A details the behavior of the implied-correlation index around nine major market events.

Table 9. Stress prediction: linear probability model (t-statistics).

Predictor	5-Day	10-Day	20-Day
Corr. ratio ($\hat{\rho}_{impl}/\hat{\rho}_{real}$)	4.07***	8.30***	3.91***
SPX ATM IV	−2.23**	−4.07***	−7.30***
VIX level	5.27***	6.69***	9.56***
Corr. divergence	0.00	0.00	0.00
R^2	0.228	0.156	0.088
N	2486	2481	2471

Notes. HC1 standard errors. ***$p < 0.01$; **$p < 0.05$.

5.4. RQ4: Greeks Information Decay and Price Magnets

The information content of Greeks for RV forecasting increases monotonically with DTE: R^2 rises from 3.8% (1-week) to 9.5% (3-month), then declines slightly to 8.2% (6-month). Return predictability is negligible across all buckets ($R^2 < 0.2\%$).

For the price magnet analysis, prices moved toward the max-OI strike in only 47.0% of 10,480 cases—below the 50% null. The OI concentration coefficient is not significant ($t = 1.06$), contradicting the “max pain” narrative.

5.5. RQ5: ML Models for SPX RV

Table 10 reports the out-of-sample comparison. OLS HAR-RV dominates at every horizon (R^2_{OOS}: 0.437, 0.948, 0.994). ML models underperform (RF: 0.143; GBM: 0.095 for 1-day). RF feature importance shows that 2-week ATM IV accounts for 50.8% of total importance; the full ranking is reported in Appendix C.

Table 10. Out-of-sample SPX realized variance forecasting (2020–2025).

Model	Target	MSE ($\times 10^7$)	MAE ($\times 10^4$)	R^2_{OOS}
VIX benchmark	1-Day	1.835	1.371	0.336
OLS: HAR-RV	1-Day	1.556	1.018	0.437
Random Forest	1-Day	2.368	1.148	0.143
Gradient Boosting	1-Day	2.501	1.282	0.095
VIX benchmark	5-Day	18.870	5.425	0.597
OLS: HAR-RV	5-Day	2.452	1.431	0.948
VIX benchmark	22-Day	211.1	21.61	0.618
OLS: HAR-RV	22-Day	3.518	1.709	0.994

Notes. Training: 2010–2019 (2,515 days). Test: 2020–2025 (1,507 days).

5.6. Granger Causality and VAR Analysis

Table 11 reports the Granger causality results. ATM IV strongly Granger-causes RV ($F = 62.4$, significant in 100% of tickers). Returns Granger-cause skew ($F = 13.6$, 95.7%), indicating asymmetric volatility feedback.

Table 11. Granger causality tests (5 lags).

Hypothesis	Avg. F	Avg. p	% sig.
ATM IV → RV	62.381	0.0000	100.0
RV → ATM IV	17.933	0.1501	60.9
Skew → RV	20.792	0.0177	94.2
Return → Skew	13.576	0.0133	95.7
ATM IV → Return	1.799	0.2688	24.6
PC ratio → Return	1.010	0.4999	5.8

Notes. Per-ticker bivariate F-tests. % sig. is the share of tickers rejecting the null at 5%.

The bivariate VAR(5) impulse response shows that a one-SD shock to ATM IV produces a cumulative RV response of 0.67 SD at horizon 2, decaying to 0.27 SD at horizon 20. The forecast error variance decomposition reveals that IV shocks explain an increasing share of RV forecast error variance, reaching **73.8%** at the 20-day horizon (Appendix D).

6. Robustness

6.1. Alternative Standard Errors

Table 12 compares inference for the 5-day return regression under three SE specifications (White, 1980; Newey and West, 1987; Cameron et al., 2011); Petersen (2009) shows that two-way clustering is the most conservative choice for finance panels of this type. Of the ten predictors, nine remain significant at 5% under Newey-West HAC(5), and eight under double-clustered (ticker + date) SEs. Implied skewness ($t_{DC} = -4.94$), implied kurtosis ($t_{DC} = 9.02$), PC volume ratio ($t_{DC} = -7.96$), and net gamma ($t_{DC} = 5.39$) remain highly significant under all specifications.

Table 12. Robustness of 5-day return predictability to SE specification.

Variable	HC1	NW(5)	DC
$IV_{ATM,30d}$	3.86***	3.32***	1.40
IV term slope	4.83***	4.59***	2.12**
$\text{Skew}_{25\delta}$	−2.03**	−0.93	−0.39
Implied skewness	−15.19***	−13.66***	−4.94***
Implied kurtosis	25.87***	25.64***	9.02***
PC vol. ratio	−13.72***	−16.69***	−7.96***
PC OI ratio	12.33***	13.12***	4.61***
Net gamma exp.	15.55***	17.83***	5.39***
RV_{daily}	−17.18***	−17.87***	−7.27***
RV_{weekly}	10.10***	7.95***	3.53***

Notes. NW(5) = Newey-West with 5 lags. DC = double-clustered by ticker and date following Cameron et al. (2011). $^{***}p < 0.01$; $^{**}p < 0.05$.

6.2. Subperiod Stability

Table 13 shows that both return predictability and the HAR+IV improvement are stable across nine subperiods. The 5-day return R^2 ranges from 3.1% (post-GFC) to 14.4% (COVID). The HAR+IV improvement over HAR-RV is positive in eight of nine subperiods, peaking at +35.5% during the low-vol era (2017–18) and +34.4% during the recovery (2023–24).

Table 13. Subperiod stability.

Subperiod	5D ret R^2	HAR-RV R^2	HAR+IV R^2	ΔR^2 (%)
Post-GFC (2010–12)	0.031	0.344	0.392	+14.0
Bull (2013–16)	0.075	0.342	0.336	−1.7
Low vol (2017–18)	0.136	0.309	0.419	+35.5
Pre-COVID (2019)	0.080	0.280	0.363	+29.9
COVID (2020)	0.144	0.604	0.721	+19.4
Post-COVID (2021)	0.036	0.376	0.440	+16.9
Rate hike (2022)	0.084	0.387	0.471	+21.8
Recovery (2023–24)	0.052	0.290	0.390	+34.4
Recent (2025)	0.037	0.321	0.402	+25.2

Notes. $\Delta R^2 = (R^2_{HAR+IV} - R^2_{HAR})/R^2_{HAR} \times 100$.

6.3. VIX Regime Conditioning

Table 14 shows that predictability increases with VIX level. The 5-day return R^2 rises from 2.7% (VIX < 15) to 19.8% (VIX > 35). Option-implied information is most valuable in high-uncertainty environments.

Table 14. Predictability by VIX regime.

VIX regime	5D ret R^2	HAR+IV R^2 (RV)	Pct. of sample (%)	N
Very low (<15)	0.027	0.359	35.7	43,971
Low (15–20)	0.040	0.341	35.0	40,221
Medium (20–25)	0.073	0.394	15.8	19,210
High (25–35)	0.080	0.385	11.0	13,139
Crisis (>35)	0.198	0.618	2.5	2,540

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

Adding controls (log volume, log OI, lagged returns) increases 5-day return R^2 from 5.3% to 5.6%; all core predictors retain significance. Quantile regressions at $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ (Koenker and Bassett, 1978) show that the implied kurtosis and put-call ratio coefficients retain a stable sign and magnitude across the entire conditional return distribution.

The distribution of ticker-level R^2 values for 5-day returns has mean 5.9%, median 4.7%, and interquartile range [2.0%, 8.2%], confirming that panel results are not driven by outliers.

Rolling-window (252-day) analysis yields average R^2 of 8.0% (std. 4.5%, range [2.1%, 17.9%]) for 5-day returns and 41.2% (std. 10.7%) for 1-day RV forecasting.

6.5. Estimator and Specification Sensitivity

Two implementation choices could mechanically drive the baseline results: the winsorization cutoff and the HAC lag length. Table 15 shows that neither does. Predictability rises modestly with heavier trimming—from $R^2 = 5.1\%$ at 0.5% cutoffs to 6.2% at 5% cutoffs—because winsorization suppresses noise in the extreme tails, and even wholly unwinsorized data leave the key predictors highly significant ($t_{kurt} = 13.8$). Lengthening the Newey-West window from 5 to 22 lags (Newey and West, 1987), enough to absorb the full overlap of the 5-day return plus a trading month, barely moves the t-statistics: implied kurtosis declines only from 25.6 to 22.1, and the set of significant predictors (9 of 10) is unchanged at every lag length.

Table 15. Winsorization and HAC-lag sensitivity (5-day returns).

Panel A: winsorization cutoff					Panel B: NW lags	
Cutoff	R^2	t_{kurt}	t_{PC}	#sig.	Lags	t_{kurt}
None	0.034	13.76	−3.59	8/10	NW(5)	25.64
0.5%	0.051	36.51	−21.31	10/10	NW(10)	23.89
1% (baseline)	0.053	40.07	−21.48	9/10	NW(22)	22.09
2.5%	0.058	43.66	−21.32	9/10		
5%	0.062	45.73	−20.86	9/10		

Notes. Panel A: pooled 5-day return regression (all 69 tickers) with HC1 t-statistics under alternative symmetric winsorization cutoffs. t_{kurt} and t_{PC} refer to implied kurtosis and the put-call volume ratio. #sig. counts predictors significant at 5%. Panel B: Newey-West t-statistic of implied kurtosis at alternative lag lengths (baseline data treatment).

6.6. Rank-Based Information Coefficients

Parametric regressions may be sensitive to outliers and functional form even after winsorization. As a fully nonparametric check, I compute daily cross-sectional Spearman information coefficients (ICs)—the rank correlation between each predictor and the subsequent 5-day stock return—and test whether the time-series mean IC differs from zero. Table 16 confirms the regression evidence. Implied kurtosis carries a mean IC of +0.098 ($t = 16.8$), positive on 64.3% of the 4,024 trading days; the put-call volume ratio (−0.059, $t = -13.8$) and the 25δ skew (−0.056, $t = -9.7$) are reliably negative. At the 1-day horizon no predictor achieves a mean IC above 0.02 in absolute value, mirroring the near-zero daily R^2 of Section 5.

Table 16. Daily cross-sectional Spearman information coefficients (stocks, 5-day returns).

Predictor	Mean IC	t-stat	% days > 0
Net gamma exposure	0.1702	33.99	77.8
Implied kurtosis	0.0978	16.77	64.3
IV term slope	0.0328	6.30	54.6
PC OI ratio	0.0170	3.63	52.4
RV_{daily}	0.0161	2.41	52.7
RV_{weekly}	0.0120	1.73	51.2
$IV_{ATM,30d}$	0.0085	1.09	53.0
Implied skewness	−0.0188	−3.30	47.4
$\text{Skew}_{25\delta}$	−0.0558	−9.66	41.9
PC vol. ratio	−0.0591	−13.81	37.2

Notes. For each trading day with at least 20 stocks, the Spearman rank correlation between the predictor and the 5-day forward return is computed across stocks; the table reports the time-series mean, its Fama-MacBeth-style t-statistic, and the share of days with a positive IC ($N = 4{,}024$ days maximum).

6.7. Decile Sorts

The baseline quintile sorts could mask nonlinearity in the extreme tails of the sorting variables (Fama and French, 2008). Table 17 repeats the three strongest sorts with decile portfolios. Sharpening the sort raises the spread in annualized returns—implied kurtosis rises from 44.0% to 52.1%, and the 25δ skew from −37.5% to −52.5%—while Sharpe ratios remain essentially unchanged (2.33 vs. 2.19 for kurtosis), since the finer portfolios are noisier. The premia are therefore monotone rather than driven by a single extreme quintile.

Table 17. Quintile vs. decile long-short sorts (5-day returns).

Sorting variable	Scheme	Ann. ret. (%)	Sharpe	t-stat
Implied kurtosis	Q5−Q1	44.04	2.328	19.84
	D10−D1	52.08	2.185	18.56
Put-call vol. ratio	Q5−Q1	−30.29	−2.478	−21.80
	D10−D1	−33.37	−2.099	−18.47
$\text{Skew}_{25\delta}$	Q5−Q1	−37.53	−1.868	−15.92
	D10−D1	−52.51	−1.896	−16.11

Notes. Equal-weighted daily-rebalanced portfolios of individual stocks; long-short is the top minus the bottom portfolio. Annualization uses $\sqrt{52}$.

6.8. Temporal Stability and a Placebo Test

A recurrent concern with in-sample predictability is that it is concentrated in a handful of episodes and vanishes when they are excluded (Welch and Goyal, 2008; Campbell and Thompson, 2008). Table 18 therefore re-estimates the two headline regressions sixteen times, excluding one calendar year at a time. The 5-day return R^2 stays within [4.7%, 5.7%] regardless of which year is dropped—removing 2018 (Volmageddon) has the largest effect—and the HAR+IV realized-variance R^2 stays within [0.478, 0.492] for every exclusion except 2020, whose removal lowers it to 0.414 by taking out the COVID variance burst. No single year drives either result.

Finally, a placebo experiment permutes the entire block of option-implied features within each ticker (preserving their cross-feature correlation while destroying their alignment with dates) and re-runs the 5-day return regression. Across ten seeded draws, the placebo R^2 averages 0.0008 (maximum 0.0010) against an actual R^2 of 0.0529—a 66-fold gap confirming that the measured predictability reflects genuine temporal information rather than mechanical panel structure. Together with t-statistics that comfortably clear the conservative $t > 3.0$ multiple-testing hurdle of Harvey et al. (2016) for every key predictor, these checks make a data-mining explanation of the findings implausible.

Table 18. Leave-one-year-out panel R^2.

Excl. year	5D ret R^2	HAR+IV R^2	Excl. year	5D ret R^2	HAR+IV R^2
2010	0.0537	0.4852	2018	0.0469	0.4861
2011	0.0535	0.4810	2019	0.0540	0.4858
2012	0.0537	0.4831	2020	0.0505	0.4142
2013	0.0538	0.4800	2021	0.0557	0.4817
2014	0.0529	0.4833	2022	0.0500	0.4786
2015	0.0514	0.4922	2023	0.0532	0.4883
2016	0.0520	0.4844	2024	0.0572	0.4895
2017	0.0547	0.4839	2025	0.0573	0.4900

Notes. Each row re-estimates the pooled 5-day return regression (ten predictors) and the HAR+IV 1-day RV regression (specification of Table 13) on the full panel excluding the indicated calendar year. Full-sample baselines: 0.053 and 0.483.

7. Discussion

7.1. Economic Interpretation

Horizon-dependent predictability. The increase from near-zero daily R^2 to 4.8–19.3% weekly R^2 suggests that option-implied information captures medium-term risk assessments that take several days to be reflected in spot prices, consistent with limits to arbitrage and gradual information diffusion.

Implied kurtosis as the strongest predictor. The Sharpe ratio of 2.33 for the kurtosis sort points to a “kurtosis risk premium”: investors demand compensation for holding assets with uncertain tail behavior. This extends the variance-risk-premium literature (Bollerslev et al., 2009) to higher moments.

Information flow: options lead volatility. The Granger causality results ($F = 62.4$, significant for 100% of tickers) and the variance decomposition (73.8% at 20 days) establish the options market as the dominant information venue for volatility dynamics, consistent with the informational-role hypothesis of Easley et al. (1998).

The failure of the price-magnet hypothesis. The 47.0% convergence rate—below the 50% null—contradicts the “max pain” narrative and is consistent with Avellaneda and Lipkin (2003), who find that pinning is confined to the last hours before expiration.

7.2. Comparison with Prior Literature

The implied-kurtosis Sharpe ratio of 2.33 exceeds those of the variance-risk-premium strategies in Bollerslev et al. (2009) (≈1.0) and the skewness sorts in Xing et al. (2010) (≈0.8). The HAR-RV + IV improvement (23.3%) likewise exceeds the 10–15% reported by Busch et al. (2011), plausibly because the present study exploits the full surface rather than a single VIX-like measure.

7.3. Limitations

Several limitations merit acknowledgment. Portfolio sorts assume frictionless trading. The options features are computed from end-of-day data; intraday options data might reveal different patterns. The focus on large-cap, liquid names may limit generalizability. The implied-correlation analysis uses equal weights. Finally, the machine-learning comparison is limited to tree-based methods; deep learning merits future investigation.

8. Conclusion

This paper provides comprehensive evidence on the information content of the options market, drawing on 3.83 billion option contracts and 11.5 billion intraday observations spanning 2010–2025 and six asset classes. The findings can be summarized as follows.

Finding 1. Implied kurtosis, put-call ratios, and implied skewness predict weekly cross-sectional stock returns. Quintile long-short portfolios sorted on implied kurtosis deliver annualized Sharpe ratios of 2.33 ($t = 19.84$). Daily predictability is negligible.

Finding 2. Combining the IV surface with HAR-RV improves 1-day realized-variance forecasting by 23.3% in R^2, an improvement that is robust across nine subperiods.

Finding 3. The implied-to-realized correlation ratio predicts market stress at 5- to 20-day horizons (t-statistics: 3.91–8.30).

Finding 4. The information content of the Greeks for realized volatility increases with time-to-expiry (R^2: 3.8% at one week → 9.5% at three months). The price-magnet hypothesis is not supported.

Finding 5. HAR-RV dominates machine-learning models and the VIX for out-of-sample SPX realized-variance forecasting (2020–2025). Short-tenor ATM implied volatility carries the highest marginal predictive content (50.8% of Random Forest importance).

Cross-cutting result. ATM implied volatility Granger-causes realized volatility in 100% of tickers ($F = 62.4$), and IV shocks explain 73.8% of the RV forecast error variance at the 20-day horizon.

Future research should explore regularized approaches to exploiting the high-dimensional options surface, assess the net profitability of option-implied strategies under realistic transaction costs, and extend the analysis to international markets.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies*, 30(1):2–47.

An, B.-J., Ang, A., Bali, T. G., and Cakici, N. (2014). The joint cross section of stocks and options. *Journal of Finance*, 69(5):2279–2337.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

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

Bali, T. G., Hu, J., and Murray, S. (2019). Option implied volatility, skewness, and kurtosis and the cross-section of expected stock returns. Working Paper, Georgetown University.

Barbon, A. and Buraschi, A. (2022). Gamma fragility. *Journal of Financial Economics*, 143(1):316–348.

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

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

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001). Forecasting S&P 100 volatility. *Journal of Econometrics*, 105(1):5–26.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

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

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility. *Journal of Econometrics*, 160(1):48–57.

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

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

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

Cao, C., Chen, Z., and Griffin, J. M. (2005). Informational content of option volume prior to takeovers. *Journal of Business*, 78(3):1073–1109.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.

Christensen, B. J. and Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150.

Christensen, K., Siggaard, M., and Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5):1680–1727.

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

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

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

Driessen, J., Maenhout, P. J., and Vilkov, G. (2009). The price of correlation risk. *Journal of Finance*, 64(3):1377–1406.

Easley, D., O'Hara, M., and Srinivas, P. (1998). Option volume and stock prices. *Journal of Finance*, 53(2):431–465.

Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63(4):1653–1678.

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

Ge, L., Lin, T.-C., and Pearson, N. D. (2016). Why does the option to stock volume ratio predict stock returns? *Journal of Financial Economics*, 120(3):601–622.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

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

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

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *Review of Financial Studies*, 27(10):2841–2871.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

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

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

Ni, S. X., Pearson, N. D., Poteshman, A. M., and White, J. (2009). Does option trading convey stock price information? *Journal of Finance*, 60(3):1049–1084.

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

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility. *Review of Economics and Statistics*, 97(3):683–697.

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

Stilger, P. S., Kostakis, A., and Poon, S.-H. (2017). What does risk-neutral skewness tell us about future stock returns? *Management Science*, 63(6):1814–1834.

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

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

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

A. Implied Correlation Around Crises

Table A1 details the behavior of the implied-correlation index around the nine major market events of the sample period. Implied correlation rises after every event, with the largest jumps around Volmageddon (+0.328) and the August 2015 China devaluation (+0.258).

Table A1. Implied correlation around major market events.

Crisis	Pre IC	Post IC	ΔIC
Flash Crash (May 2010)	0.223	0.404	+0.181
Euro Crisis (Aug 2011)	0.313	0.469	+0.156
China Deval. (Aug 2015)	0.231	0.489	+0.258
Volmageddon (Feb 2018)	0.108	0.436	+0.328
COVID-19 (Mar 2020)	0.473	0.648	+0.175
Meme Stocks (Jan 2021)	0.228	0.326	+0.098
Rate Shock (Jun 2022)	0.417	0.447	+0.030
SVB Crisis (Mar 2023)	0.328	0.373	+0.045
VIX Spike (Aug 2024)	0.138	0.316	+0.178

Notes. IC = Implied Correlation. Pre = 30-day window before the event; Post = 10-day window after.

B. SPX Options Surface Features (RQ5)

The 21 SPX surface features used in Section 5 are: IV_{ATM} at six tenors (1w, 2w, 1m, 2m, 3m, 6m); 25δ skew (1m, 3m); 10δ deep OTM skew (1m); butterfly ratio (1m); term-structure slopes (3m$-$1m, 6m$-$1m); total gamma$\times$OI; net gamma; total vega$\times$OI; average Θ; put-call volume ratio; put-call OI ratio; total option volume; total OI; and the average bid-ask spread percentage.

C. Random Forest Feature Importance (RQ5)

Table A2 reports the full Random Forest importance ranking underlying the RQ5 discussion; the two shortest ATM tenors jointly account for two thirds of total importance.

Table A2. Random Forest feature importance (1-day RV target, 2010–2019 training).

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

D. Forecast Error Variance Decomposition

Table A3 reports the full horizon profile of the forecast error variance decomposition summarized in Section 5: the share of realized-variance forecast error variance attributable to implied-volatility shocks rises monotonically from zero on impact to 73.8% at 20 days.

Table A3. FEVD: share of RV forecast error variance explained by IV shocks (%).

Horizon (days)	IV shocks (%)	RV shocks (%)
1	0.0	100.0
2	28.0	72.0
5	56.1	43.9
10	66.4	33.6
15	71.1	28.9
20	73.8	26.2

Notes. From the bivariate VAR(5) in standardized ATM IV and RV, averaged across 20 tickers.