Implied Moment Indices and Volatility-Managed Portfolios

Michael Gao*

Department of Statistics & Data Science

Yale University

m.gao@yale.edu

*I am grateful to my research advisor, David Shimko (Department of Finance and Risk Engineering, NYU Tandon School of Engineering), for his guidance throughout this project. All remaining errors are my own.

Abstract

Moreira and Muir (2017) show that scaling equity market exposure inversely with realized variance generates economically large risk-adjusted returns relative to a buy-and-hold benchmark. We re-examine this finding from the perspective of a long-term investor who evaluates strategies by their Sharpe and information ratios, and ask whether publicly available implied moment indices (VIX, CBOE SKEW, and VVIX) add incremental timing value over realized variance within the same framework. Using a 100-year sample of U.S. market returns (1926–2025, 1,194 months), four findings emerge. First, the realized variance (RV) managed portfolio closely replicates Moreira and Muir (2017): $\alpha = 4.56\%$ ($t = 3.10$) versus their 4.86% ($t = 3.11$). Second, the VIX-managed portfolio underperforms RV on both alpha (2.91% versus 4.16%) and information ratio (0.317 versus 0.354) despite being a substantially better variance forecast (OOS $R^2 = 0.32$ versus -0.07); raw Sharpe rankings disagree only because VIX inherits more equity-premium exposure through a higher portfolio beta. Third, all eight hand-crafted higher-moment adjustments reduce risk-adjusted returns relative to the variance-only baseline. Fourth, Gradient Boosting attains the highest information ratio among machine learning specifications (0.357), but a feature ablation shows the gain is driven primarily by nonlinear variance modeling: variance-only GB already reaches IR = 0.332. The most attractive practical strategy in our tests is RV-managed with a 1.5x leverage cap, which delivers the highest Sharpe ratio (0.701) and the highest information ratio (0.374) of any VIX-era strategy considered.

Keywords: volatility-managed portfolios, realized variance, implied moments, machine learning, market timing, VIX, CBOE SKEW

JEL Classification: G11, G12, G13, C58

1 Introduction

A long-term investor who already holds the market portfolio faces a basic question: can the information embedded in equity options markets be used to do better, in risk-adjusted terms, than buy-and-hold? Options continuously price the full return distribution. CBOE publishes scalar indices that approximate implied volatility (VIX), implied skewness (SKEW), and the volatility of volatility (VVIX). If these publicly available indices contain incremental information about future market risk, they should improve on a rule that conditions only on past realized variance. This paper asks whether they do, evaluating each candidate strategy not by its raw spanning regression alpha alone but by the risk-adjusted return measures that matter to a long-term investor: the Sharpe ratio of the portfolio, and the information ratio (alpha per unit of idiosyncratic risk) that strips out mechanical differences in equity-premium exposure across strategies.

Moreira and Muir (2017) establish the benchmark: the volatility-managed portfolio scales monthly market exposure by $c/\hat{\sigma}_t^2$, where $\hat{\sigma}_t^2$ is within-month realized variance. We replicate this result on a 100-year sample (July 1926 to December 2025), obtaining $\alpha = 4.56\%$ ($t = 3.10$) and Sharpe ratio of 0.518 versus 0.450 for buy-and-hold. The replication is essentially exact, both in alpha and in significance: Moreira and Muir (2017) report $\alpha = 4.86\%$ ($t = 3.11$) on a slightly shorter sample. Cederburg et al. (2020) caution that spanning regression alphas do not always translate to real-time investment gains, attributing the gap to structural instability across market regimes. We take this seriously and conduct all machine learning and out-of-sample evaluations using strict expanding-window protocols.

VIX is the market's 30-day model-free implied volatility and, by construction, a forward-looking variance estimate. Bollerslev et al. (2009) show that the variance risk premium predicts aggregate returns at the quarterly horizon, and Kostakis et al. (2011) demonstrate that the full option-implied distribution improves portfolio allocation relative to historical moments. Yet we find that VIX-managed portfolios generate *lower* alpha and a *lower* information ratio than RV-managed portfolios over 1990–2025 (alpha 2.91% versus 4.16%; IR

0.317 versus 0.354), despite VIX being a substantially better variance forecast (out-of-sample $R^2 = 0.32$ versus -0.07). The raw Sharpe ratio reverses the ranking (VIX 0.670 versus RV 0.653), but only because VIX-managed weights stay closer to full market exposure (portfolio beta 0.795 versus 0.629) and so inherit more of the equity premium itself. Once mechanical beta differences are stripped out, the information ratio agrees with alpha. Forecasting accuracy and risk-adjusted timing value are not the same objective.

Beyond variance, options markets encode information about skewness and kurtosis. CBOE SKEW approximates 30-day implied skewness; VVIX (the volatility of VIX) serves as a kurtosis proxy. Harvey and Siddique (2000) and Conrad et al. (2013) establish that skewness is priced in the cross-section of expected returns, motivating its use as a timing signal. We find that hand-crafted z-score adjustments using realized and implied skewness and kurtosis uniformly reduce risk-adjusted returns: every moment extension hurts the Sharpe ratio and the information ratio relative to the variance-only baseline. Gradient Boosting, given access to the same signals plus interactions, attains the highest information ratio among machine learning strategies (0.357), but a feature ablation reveals that most of this gain reflects nonlinear variance modeling rather than higher-moment information: a variance-features-only specification already reaches IR = 0.332.

The most practically attractive strategy in our tests is in fact the simplest: the RV-managed portfolio with a 1.5x leverage cap. The unconstrained RV portfolio occasionally takes positions of five times market exposure during low-volatility months, weights that no real trading desk would carry. Imposing the cap eliminates these noisy positions and produces the highest Sharpe ratio (0.701) and the highest information ratio (0.374) of any VIX-era strategy in our analysis. The improvement matters because real-world risk limits are binding constraints, not optional refinements.

Our results speak to three literatures. On volatility-managed portfolios, we document that implied variance (VIX) does not improve on realized variance for risk-adjusted timing, consistent with Božović (2024) and complementary to Kostakis et al. (2011), who find

that the forward-looking mean rather than higher moments drives outperformance in their utility-maximization setting. On options-implied information, we test the three CBOE distributional indices (VIX, SKEW, VVIX) jointly as moment proxies in a unified machine learning framework, an approach not previously studied in this context. On machine learning in asset pricing, we find that tree-based methods produce the strongest risk-adjusted timing among ML specifications, primarily through nonlinear variance modeling rather than higher-moment information, while linear regularization (Ridge, Lasso) destroys value through variance extrapolation that no risk-limited fund could implement. This is consistent with Gu et al. (2020), who find that nonlinear methods dominate across a broad set of equity return prediction tasks.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes our data. Section 4 outlines the methodology. Section 5 presents the main results. Section 6 provides robustness checks. Section 7 concludes.

2 Related Literature

2.1 Volatility-Managed Portfolios

The idea that volatility timing creates economic value has a long pedigree. Fleming et al. (2001) provide early evidence that dynamically adjusting portfolio weights in response to conditional volatility improves certainty-equivalent returns for a mean-variance investor. Moreira and Muir (2017) formalize this insight into the volatility-managed portfolio framework. Scaling monthly factor exposure by c/RV_t (the inverse of trailing realized variance) generates spanning regression alphas of approximately 4.86% for the market factor and significant alphas across nine equity factors including value, momentum, profitability, and investment. The economic intuition is that variance is strongly mean-reverting, so periods of high current variance tend to be followed by poor risk-return trade-offs; reducing exposure during these periods expands the mean-variance frontier.

Barroso and Santa-Clara (2015) extend the framework to momentum, where the gains are particularly striking. Momentum carries the highest unconditional Sharpe ratio among standard factors (0.53) but also the worst crash risk, with excess kurtosis of 18.24 and skewness of -2.47. Momentum risk is highly time-varying and predictable: an AR(1) model on realized variance achieves an out-of-sample R^2 of 57.82%. RV-scaling nearly doubles the Sharpe ratio (0.53 to 0.97) and dramatically reduces crash exposure (excess kurtosis falls from 18.24 to 2.68; maximum drawdown from -96.69% to -45.20%). Their results confirm that volatility management generalizes well beyond the market factor.

The most comprehensive critique comes from Cederburg et al. (2020), who evaluate volatility management across 103 equity trading strategies. While they confirm Moreira and Muir's in-sample spanning regression alphas (77 of 103 are positive, 23 statistically significant), they argue that these regressions are not implementable in real time because they require ex post optimal portfolio weights. Out-of-sample combination strategies earn lower certainty-equivalent returns in 72 of 103 cases. The root cause is structural instability: the

regression coefficients that map variance signals into portfolio weights shift across subsamples. We address this by training all models only on data available at portfolio formation time.

Božović (2024) tests VIX-managed portfolios specifically, using the cumulative previous-month VIX to gauge leverage. VIX-managed strategies yield more stable portfolio weights and require less rebalancing than RV-based strategies, which Božović argues improves net-of-cost performance. Our paper corroborates the smoother weight path but finds that VIX produces lower gross alpha and a lower information ratio than realized variance, with the raw Sharpe ranking favoring VIX only because of the higher portfolio beta its smoother weights induce (Section 5.2). We extend this work by incorporating higher-moment signals (SKEW, VVIX), machine learning methods for nonlinear signal combination, and explicit risk-limit considerations through leverage constraints.

2.2 Options-Implied Information and Return Predictability

Bakshi et al. (2003) provide the theoretical foundation for extracting distributional information from option prices without parametric assumptions. They derive model-free expressions linking a continuum of option prices to risk-neutral variance, skewness, and kurtosis, and establish "skew laws" decomposing individual stock risk-neutral skewness into systematic and idiosyncratic components. Crucially, they show that risk aversion introduces negative skewness into the risk-neutral density even when the physical distribution is symmetric. This theoretical apparatus underpins the interpretation of CBOE SKEW as a measure of implied tail risk: when SKEW rises above its neutral value of 100, the option-implied distribution is more negatively skewed, reflecting greater demand for downside protection.

Bollerslev et al. (2009) demonstrate that the variance risk premium (the difference between model-free implied variance and realized variance) explains a nontrivial fraction of aggregate stock return variation. The predictive power is strongest at the quarterly horizon, where the variance risk premium dominates the price-earnings ratio, the default spread, and

the consumption-wealth ratio over the 1990–2007 sample. This result explains why VIX contains information beyond realized variance: it reflects the premium investors pay to hedge variance uncertainty, which varies with time-varying risk aversion.

Kostakis et al. (2011) is the closest antecedent to our work. They extract constant-maturity one-month S&P 500 implied distributions from CME futures options (January 1986 to December 2009) and use them to construct optimal portfolios via expected utility maximization. The forward-looking approach beats backward-looking (historical) distributions over 1992–2007 under multiple utility specifications. However, their conclusion directly foreshadows ours: "higher-order moments are not the source of outperformance within our setting," with the forward-looking mean driving their gains rather than skewness or kurtosis. Our paper differs in three ways: we use scalar CBOE proxies rather than the full implied PDF, we rely on publicly available indices implementable in real time rather than full option chains, and we add machine learning as a method for nonlinear signal combination. The practical upside is that our approach requires no proprietary data: VIX, SKEW, and VVIX are freely available daily from CBOE, making the strategies directly replicable by any practitioner without access to option chains or specialized data vendors.

Harvey and Siddique (2000) establish that conditional systematic skewness is priced in the cross-section of expected returns, commanding a risk premium of approximately 3.60% per year. They show that momentum portfolio returns are related to systematic skewness, providing a partial risk-based explanation for the momentum anomaly. Their analysis uses backward-looking realized skewness; our paper tests whether forward-looking (implied) skewness adds value in a timing context. Conrad et al. (2013) complement this by using option prices to estimate ex ante higher moments of individual securities' risk-neutral distributions. They find that more negatively skewed securities have higher subsequent returns, consistent with risk compensation. Their analysis is cross-sectional rather than time-series, but it motivates our use of CBOE SKEW as a market-level timing signal at the aggregate level.

2.3 Risk-Neutral Density Estimation

Breeden and Litzenberger (1978) establish the foundational result that state-contingent claim prices are implicit in option prices. The risk-neutral density satisfies $q(K) = e^{rT}\,\partial^2 C/\partial K^2$, linking the second derivative of the call price function with respect to strike to the Arrow–Debreu state price density. Shimko (1993) provides the first practical implementation: smooth the implied volatility smile with a quadratic fit $\sigma(K) = a + bK + cK^2$, reconstruct call prices via Black-Scholes, and apply the Breeden-Litzenberger formula numerically. This approach remains widely employed due to its simplicity and robustness.

zur Linden (2023) introduces the Analytic Recovery Method (ARM), a parametric approach that recovers the full risk-neutral density via a polynomial transformation of the standard normal, yielding closed-form call prices and analytical implied moments. We use ARM as a proof-of-concept variance signal in Section 6.4 and describe the method in Appendix A. Jackwerth and Rubinstein (1996) provide an alternative nonparametric approach; together, these methods situate our distributional analysis in the broader risk-neutral density estimation literature.

3 Data

3.1 Equity Returns

We obtain daily and monthly equity market returns from the Kenneth French Data Library (Fama and French, 1993). Our primary variable is the market excess return (Mkt-RF): the value-weighted return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate. The full sample spans July 1926 to December 2025, comprising 1,194 monthly observations and 26,151 daily observations. Daily returns are used to construct within-month realized moments (Section 3.3); monthly returns serve as the portfolio formation and evaluation frequency. The average monthly market excess return is 0.69% (approximately 8.3% annualized) with a standard deviation of 5.31%.

3.2 CBOE Volatility Indices

We use four CBOE volatility indices to capture different dimensions of the option-implied distribution.

VIX. The CBOE Volatility Index measures 30-day model-free implied volatility for S&P 500 options. We convert daily VIX levels to a monthly implied variance estimate: $\hat{\sigma}^2_{\text{impl}} = (\text{VIX}_t/100)^2/12$. The VIX subsample runs from January 1990 to December 2025 (432 months). Over this period, VIX averages 19.49 with a standard deviation of 7.41, ranging from a low of 9.51 to a crisis peak of 59.89.

CBOE SKEW. The SKEW index measures the perceived cost of tail protection via out-of-the-money puts. It equals 100 when the implied distribution is symmetric and rises above 100 when the distribution is left-skewed. We normalize it as $\text{SKEW}_{\text{norm}} = \text{SKEW} - 100$. Data are available from January 2010 (192 months). The average level is 131.10 (SD = 12.44), confirming persistent negative implied skewness in S&P 500 options. Because moment-managed strategies require a 24-month minimum estimation window, the effective analysis period for SKEW-based strategies begins in early 2012.

VVIX. The volatility-of-VIX index measures uncertainty about future volatility levels. It serves as a proxy for implied kurtosis and tail risk of the variance process itself. Available from January 2007 (228 months), VVIX averages 93.63 (SD = 14.29).

VIX3M. The three-month implied volatility index captures the term structure of implied volatility. We use the slope VIX3M − VIX as a feature in our machine learning models. A negative slope indicates backwardation, where near-term fear exceeds longer-term expectations.

3.3 Realized Moments

Following Moreira and Muir (2017), we compute realized variance for each calendar month as the sum of squared daily excess return deviations: $\text{RV}_t = \sum_{d \in t}(r_d - \bar{r}_t)^2$, where r_d is the daily market excess return and $\bar{r}_t$ is the within-month mean. We also compute realized skewness using the unbiased estimator and realized excess kurtosis (Fisher definition) from daily excess returns within each month. Over the full sample, average monthly realized variance is 2.43×10^{-3} (corresponding to an annualized volatility of approximately 17%), with an AR(1) coefficient of 0.534. Realized skewness averages −0.11 and realized excess kurtosis averages 0.65. Notably, realized kurtosis is essentially unpredictable (AR(1) = 0.015), which helps explain why hand-crafted kurtosis adjustments add little in Section 5.3.

Table 1: Summary Statistics

Series	Mean	Std Dev	AR(1)	N	Sample
Market excess return (%, monthly)	0.69	5.31	0.09	1,194	1926–2025
Realized variance ($\times 10^{-3}$)	2.43	5.01	0.53	1,194	1926–2025
Realized skewness	−0.11	0.69	0.11	1,194	1926–2025
Realized excess kurtosis	0.65	1.66	0.02	1,194	1926–2025
VIX (end-of-month)	19.49	7.41	0.81	432	1990–2025
VIX3M (end-of-month)	21.20	7.38	0.85	234	2006–2025
CBOE SKEW	131.10	12.44	0.77	192	2010–2025
VVIX (monthly avg)	93.63	14.29	0.74	228	2007–2025

Notes: Market excess returns and realized moments are computed from the Kenneth French Data Library (daily and monthly files). VIX, VIX3M, SKEW, and VVIX are from CBOE. AR(1) denotes the first-order autocorrelation coefficient. Realized variance is reported in units of 10^{-3} (multiply by 12 and take the square root for approximate annualized volatility). VIX3M is the 90-day implied volatility index; we use the slope VIX3M−VIX as a term structure feature in machine learning models.

4 Methodology

4.1 The Volatility-Managed Portfolio Framework

We adopt the framework of Moreira and Muir (2017). Let f_{t+1} denote the excess return on the market portfolio in month $t+1$. The volatility-managed portfolio return is

$$f^{\sigma}_{t+1} = \frac{c}{\hat{\sigma}^2_t} f_{t+1}, \tag{1}$$

where $\hat{\sigma}^2_t$ is a variance estimate formed using information available at the end of month t, and c is a scaling constant defined as

$$c = \frac{\text{std}(f_{t+1})}{\text{std}(f_{t+1}/\hat{\sigma}^2_t)}, \tag{2}$$

computed over the full evaluation sample so that the managed portfolio matches the buy-and-hold standard deviation. Since α and its standard error scale equally with c, this normalization does not affect t-statistics. Realized variance for month t is computed from daily

simple excess returns:

$$\mathrm{RV}_t = \sum_{d \in t} \left(r_d - \bar{r}_t \right)^2, \tag{3}$$

where r_d is the daily market excess return and $\bar{r}_t$ is the within-month mean. Performance is evaluated via the spanning regression

$$f^{\sigma}_{t+1} = \alpha + \beta \, f_{t+1} + \varepsilon_{t+1}, \tag{4}$$

with HC0-robust standard errors. Because managed portfolio returns inherit serial correlation through the persistence of variance weights, we also verify all main results using Newey-West standard errors with 6 lags (Section 5.1); all findings survive. A positive, statistically significant α indicates that the managed strategy expands the mean-variance frontier relative to buy-and-hold. We annualize α by multiplying by 12.

Three variance signals plug into $\hat{\sigma}^2_t$. Realized variance from Equation (3) is the backward-looking benchmark of Moreira and Muir (2017). VIX implied variance converts the end-of-month VIX reading to monthly units:

$$\hat{\sigma}^2_{\mathrm{impl},t} = \frac{(\mathrm{VIX}_t/100)^2}{12}. \tag{5}$$

Machine learning predicted variance, described in Section 4.3, combines realized and implied signals nonlinearly.

4.2 Moment-Managed Portfolios

A natural extension of Equation (1) is to condition portfolio weights on higher moments as well. A risk-averse investor should reduce exposure when the distribution is unusually left-skewed (higher crash probability) or fat-tailed (greater tail uncertainty). We implement this by augmenting the variance-scaled weight with multiplicative skewness and kurtosis adjustment factors, using both realized and implied (CBOE SKEW) signals as inputs. Ad-

justment magnitudes are set to $\lambda = 0.2$: a one-standard-deviation skewness shock shifts the weight by only $\pm 20\%$, ensuring that moment signals remain strictly secondary to the variance signal even under extreme distributional readings. Weights are clipped to $[0.5, 1.5]$ so that no single signal can more than double or halve the variance-implied position. These choices bound the moment contribution without eliminating it. The qualitative finding that all eight extensions reduce performance is unlikely to depend on the exact parameterization, because the root cause is the near-zero persistence of realized higher moments (AR(1) of 0.11 for skewness, 0.02 for kurtosis): last month's moments carry almost no information about next month's distribution, so no reasonable scaling can extract a reliable timing signal from them. Expanding-window z-scores with a 24-month minimum are used to avoid look-ahead bias.

The moment-managed portfolio augments Equation (1) as follows:

$$w_t = \frac{c}{\hat{\sigma}_t^2} \cdot g_s(\text{skew}_t) \cdot g_k(\text{kurt}_t), \tag{6}$$

where the skewness and kurtosis adjustment factors are

$$g_s = \text{clip}(1 + \lambda_s \cdot z_{\text{skew},t},\ 0.5,\ 1.5)\,, \tag{7}$$

$$g_k = \text{clip}(1 - \lambda_k \cdot z_{\text{kurt},t},\ 0.5,\ 1.5)\,, \tag{8}$$

and $z_{\cdot,t}$ is the expanding-window z-score:

$$z_{\text{skew},t} = \frac{\text{skew}_{t-1} - \hat{\mu}_{t-1}^{\text{skew}}}{\hat{\sigma}_{t-1}^{\text{skew}}}, \tag{9}$$

where $\hat{\mu}_{t-1}^{\text{skew}}$ and $\hat{\sigma}_{t-1}^{\text{skew}}$ are the sample mean and standard deviation computed over all months through $t-1$. The one-period lag ensures that the weight at time t uses only information available at the end of month $t-1$; an analogous formula applies for $z_{\text{kurt},t}$. We set $\lambda_s = \lambda_k = 0.2$ and require a minimum of 24 months before z-scores are computed.

4.3 Machine Learning Variance Forecasting

We train four machine learning models to predict next-month realized variance RV_{t+1} using seven features known at the end of month t: current realized variance, VIX implied variance, realized skewness, realized excess kurtosis, lagged realized variance, month-over-month VIX change, and an interaction term ($\text{rv}_t \times \text{skew}_t$). The four models are Ridge regression, Lasso, Random Forest, and Gradient Boosting.[1]

All models use an expanding-window out-of-sample protocol: train on months $1, \ldots, t$ and predict RV_{t+1}. Features are normalized to mean zero and unit variance using statistics from the training window only, refitting at each OOS iteration so no future distributional information contaminates the standardization. With a minimum training window of 60 months, this yields 369 out-of-sample predictions. Predictions are floored at 10^{-6} before inversion to prevent extreme portfolio weights from near-zero forecasts. The predicted variance replaces $\hat{\sigma}_t^2$ in Equation (1) to form the machine learning managed portfolios.

The inverse-variance transformation creates an important asymmetry between linear and tree-based models. Portfolio weights are proportional to $1/\hat{\sigma}_t^2$, so small forecast errors near zero are mechanically amplified. Linear models (Ridge, Lasso) can extrapolate beyond the training range, occasionally predicting near-zero or negative variances that, when inverted, generate extreme portfolio weights. Tree-based models (Random Forest, Gradient Boosting) are bounded by construction: their predictions are averages over training-sample leaves and cannot fall below the minimum observed training variance. This distinction is specific to inverse-variance weighting and does not imply linear forecasters are universally inferior.

[1]Hyperparameters: Ridge ($\alpha = 1.0$), Lasso ($\alpha = 0.0001$), Random Forest (100 trees, max depth 3, minimum 10 samples per leaf), Gradient Boosting (100 trees, max depth 2, learning rate 0.05, minimum 10 samples per leaf). No hyperparameter tuning is performed within the expanding window.

5 Results

5.1 Replication: Realized Variance Management, 1926–2025

Table 2 reports results for the RV-managed portfolio over the full 1926–2025 sample. The spanning regression yields an annualized alpha of 4.56% ($t = 3.10$) with $\beta = 0.601$. The information ratio is 0.310, and the Sharpe ratio rises from 0.450 (buy-and-hold) to 0.518 (managed), an improvement of 15%. These results are extremely close to Moreira and Muir (2017), who report $\alpha = 4.86\%$, $t = 3.11$, and $\beta = 0.61$ on a slightly shorter sample. The replication is essentially exact, both in economic magnitude and in statistical significance, and adding the post-Moreira and Muir decade of data neither sharpens nor weakens the original finding meaningfully. We use HC0-robust standard errors throughout; Moreira and Muir (2017) use Newey-West.[2]

Table 2: Replication: RV-Managed Portfolio, Full Sample (1926–2025)

Strategy	N	Alpha (%)	t-stat	β	Sharpe	IR
RV-Managed (ours)	1,193	4.56	3.10	0.601	0.518	0.310
Moreira & Muir (2017)	—	4.86	3.11	0.61	—	—

Notes: Alpha is the intercept from the spanning regression $f^{\sigma}_{t+1} = \alpha + \beta\, f_{t+1} + \varepsilon_{t+1}$, annualized by multiplying by 12. Both alpha and SE(alpha) are scaled by 12, so the t-statistic is unitless and invariant to the choice of horizon. The information ratio (IR) is annualized as $(\alpha_m/\sigma(\varepsilon_m))\sqrt{12}$, expressing alpha per unit of idiosyncratic risk after stripping out beta-driven equity-premium exposure. Standard errors are HC0-robust. Moreira and Muir (2017) report their benchmark using Newey-West standard errors over a shorter sample ($\approx$1,050 months); their paper does not report N, Sharpe, or IR for this specification. The buy-and-hold benchmark achieves a Sharpe ratio of 0.450 over our 1,193-month sample.

5.2 Implied Variance: Does VIX Improve on Realized Variance?

Table 3 compares RV-managed and VIX-managed portfolios over the common VIX-era sample (1990–2025, 431 months). The economically relevant message is contained in the alpha

[2]Managed portfolio returns inherit serial correlation through the persistence of variance weights, so Newey-West standard errors with 6 lags may be more appropriate than HC0. NW reduces t-statistics by roughly 5–10% for most strategies (full-sample RV: 3.10 $\rightarrow$ 2.84; VIX-era RV: 1.99 $\rightarrow$ 1.85; VIX: 1.80 $\rightarrow$ 1.70; RV Cap=1.5: 2.10 $\rightarrow$ 2.18; GB All features: 1.90 $\rightarrow$ 1.83). The full-sample RV result remains highly significant under both inference methods; VIX-era strategies hover near the conventional 5% threshold under either.

and information ratio columns. The RV-managed portfolio generates an annualized alpha of 4.16% ($t = 1.99$) with an information ratio of 0.354. The VIX-managed portfolio produces a lower alpha of 2.91% ($t = 1.80$) and a lower information ratio of 0.317. Both metrics agree that RV is the better timing signal, even though VIX is by construction a much better variance forecast: Table 5 shows that VIX implied variance achieves an out-of-sample R^2 of 0.32, compared to -0.07 for RV persistence (a negative R^2 indicates forecasts worse than the unconditional mean).

The raw Sharpe ratio appears to send a different signal. VIX-managed delivers Sharpe 0.670 versus 0.653 for RV, a narrow win for VIX, and a slightly lower maximum drawdown (-40.8% versus -42.5%). This disagreement is mechanical, not economic. Both managed portfolios are normalized by the constant c to match the buy-and-hold standard deviation, so total volatility cannot drive the Sharpe gap. The driver is beta: VIX-managed weights move less aggressively (the portfolio carries $\beta = 0.795$), so a larger share of the strategy's return is simply inherited equity premium. RV-managed weights respond more sharply to volatility shocks, pulling the portfolio further from full market exposure ($\beta = 0.629$) and generating more market-orthogonal timing value but less mechanical equity-premium pickup. The information ratio strips out this beta-driven channel and measures only the alpha component per unit of idiosyncratic risk. By that measure, which is the relevant one for a long-term investor who already holds market exposure and is asking how much incremental risk-adjusted return a timing signal delivers, RV wins. VIX's narrow Sharpe advantage reflects more inherited market beta, not better timing. The interpretation is consistent with Božović (2024), who documents that VIX-managed strategies produce smoother weight paths and lower rebalancing costs, with the net-of-cost advantage potentially offsetting the gross alpha and IR gap we document here.

Table 3: RV-Managed vs. VIX-Managed Portfolios (1990–2025)

Strategy	N	Alpha (%)	t-stat	β	Sharpe	IR	Max DD (%)
RV-Managed	431	4.16	1.99	0.629	0.653	0.354	−42.5
VIX-Managed	431	2.91	1.80	0.795	0.670	0.317	−40.8

Notes: Same spanning regression specification as Table 2. The buy-and-hold benchmark over the same 431-month sample achieves a Sharpe ratio of 0.601. VIX is a substantially better variance forecast than RV persistence (OOS $R^2 = 0.32$ versus -0.07; see Table 5), yet produces lower alpha and a lower information ratio. Raw Sharpe rankings reverse only because VIX inherits more equity-premium exposure through a higher portfolio beta; the information ratio strips this channel out.

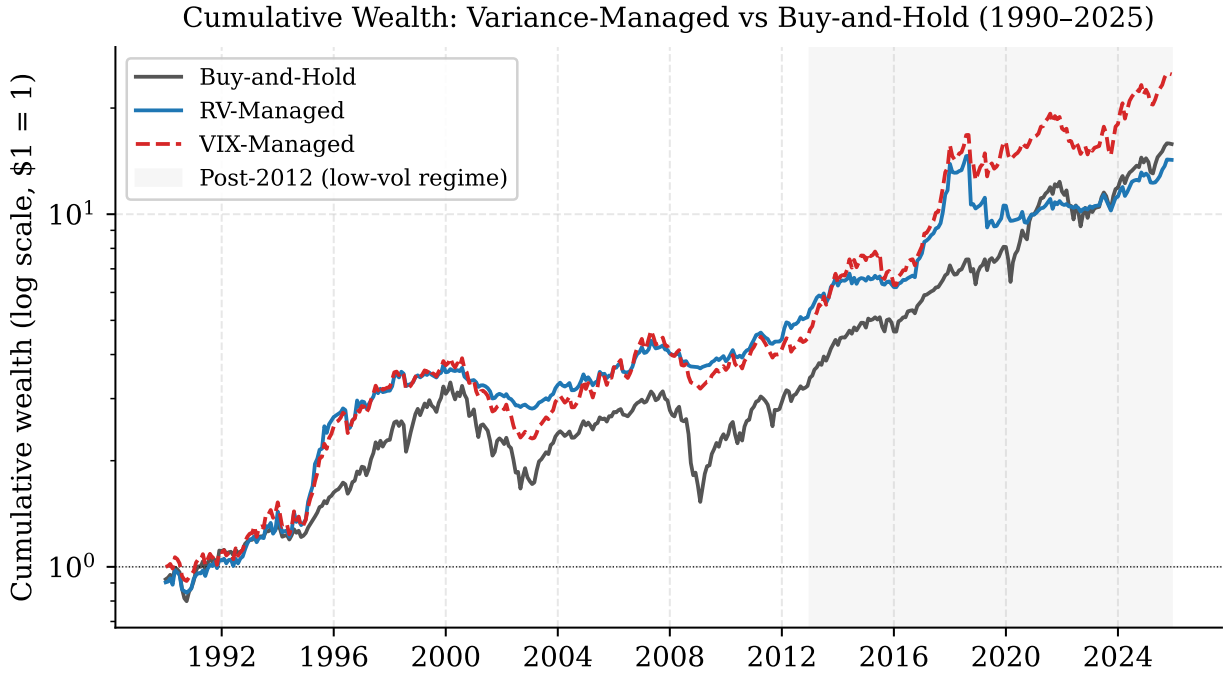

Figure 1: Cumulative wealth of the RV-managed, VIX-managed, and buy-and-hold portfolios from January 1990 to December 2025 (log scale). The shaded region marks the post-2012 low-volatility regime, during which both managed strategies offer limited improvement over buy-and-hold. RV-managed generates higher alpha (4.16%, $t = 1.99$) and a higher information ratio (0.354 vs. 0.317), while VIX-managed achieves a marginally higher raw Sharpe ratio (0.670 vs. 0.653) due to its higher beta (0.795 vs. 0.629).

5.3 Higher-Moment Adjustments

Table 4 reports results for moment-managed portfolios across three sample periods. The pattern is uniform: every moment extension reduces alpha, Sharpe ratio, and information ratio relative to the corresponding variance-only baseline. In the full sample (1926–2025),

adding realized skewness to the RV-managed strategy lowers alpha from 4.56% to 4.34% (t falls from 3.10 to 2.90; IR from 0.310 to 0.291); adding kurtosis lowers it further to 4.25% ($t = 2.79$, IR $= 0.283$); and combining both yields 4.29% ($t = 2.82$, IR $= 0.285$). The Sharpe ratio deteriorates monotonically from 0.518 to 0.484.

The VIX-era results (1990–2025) show the same pattern with larger magnitudes. Adding realized skewness and kurtosis to the VIX-managed strategy reduces alpha from 2.91% to 2.36% (t from 1.80 to 1.33), the information ratio from 0.317 to 0.243, and Sharpe from 0.670 to 0.613. Using CBOE SKEW as an implied skewness signal (2011–2025, 168 months) yields an alpha of 1.53% ($t = 0.54$, IR $= 0.161$). The short sample and high buy-and-hold Sharpe (0.808) make this result difficult to interpret. Adding realized kurtosis on top of CBOE SKEW drives alpha to 0.47% ($t = 0.16$, IR $= 0.049$).

The failure does not stem from going in the wrong direction. The adjustments are economically sensible: reduce exposure when skewness is unusually negative (more crash risk) or kurtosis is unusually high (fatter tails). The problem lies in functional form. Linear z-score adjustments reduce exposure during months of extreme distributional shape, but these are often precisely the months preceding strong recoveries. The low persistence of realized higher moments exacerbates this problem: realized skewness has an AR(1) coefficient of only 0.11, and realized kurtosis just 0.02 (Table 1), so last month's moments carry almost no information about next month's distribution. This finding echoes Kostakis et al. (2011), who conclude that higher-order moments are not the source of outperformance in their option-implied utility-maximization framework. Harvey and Siddique (2000) show that skewness is priced in the cross-section, but our results suggest that cross-sectional pricing does not translate into time-series timing power through linear adjustments.

5.4 Machine Learning: Forecasting and Portfolio Performance

Tables 5 and 6 report out-of-sample variance forecasting and the corresponding portfolio performance for the four ML models over the 369 out-of-sample months.

Table 4: Moment-Managed Portfolio Comparison

Era	Strategy	N	Alpha (%)	t-stat	Sharpe	IR
Full (1926–)	RV only	1,193	4.56	3.10	0.518	0.310
	RV + Skew	1,170	4.34	2.90	0.493	0.291
	RV + Kurt	1,170	4.25	2.79	0.483	0.283
	RV + Skew + Kurt	1,170	4.29	2.82	0.484	0.285
VIX (1990–)	VIX only	431	2.91	1.80	0.670	0.317
	VIX + R.Skew	408	2.79	1.59	0.643	0.289
	VIX + R.Kurt	408	2.57	1.51	0.638	0.275
	VIX + R.Skew+Kurt	408	2.36	1.33	0.613	0.243
SKEW (2011–)	VIX + CBOE SKEW	168	1.53	0.54	0.808	0.161
	VIX + CBOE SKEW + R.Kurt	168	0.47	0.16	0.730	0.049

Notes: Moment adjustments use expanding-window z-scores with $\lambda_s = \lambda_k = 0.2$, clipped to $[0.5, 1.5]$, and a 24-month minimum window. The information ratio (IR) decays monotonically with each moment extension, mirroring the alpha and Sharpe declines. The different sample start dates across panels reflect data availability for each signal. The SKEW-era panel (2011–) has only 168 months, limiting statistical power.

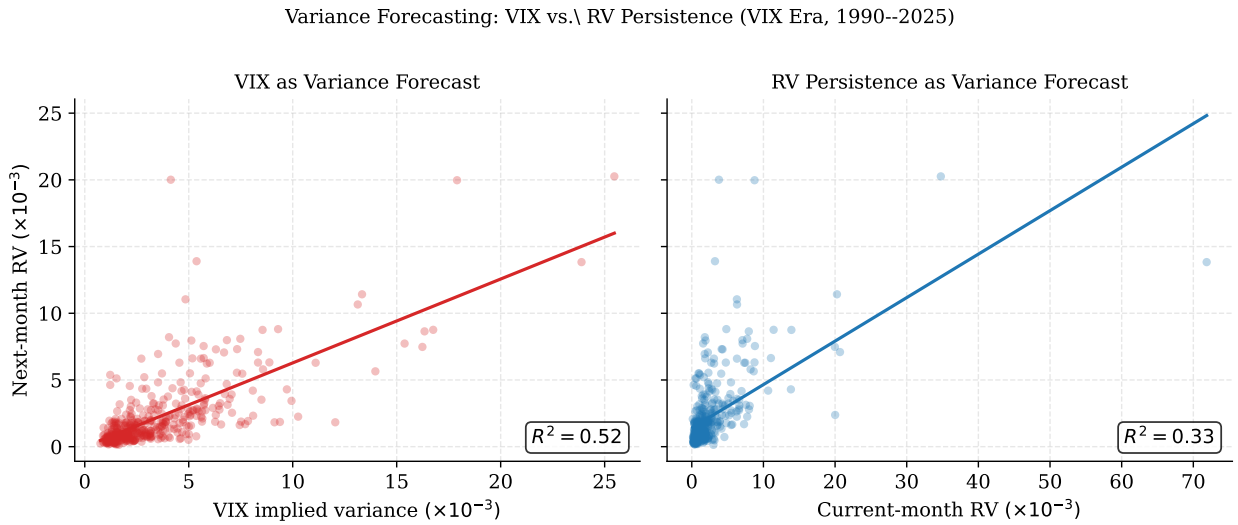

Figure 2: Scatter plots of variance forecasts against next-month realized variance over the VIX era (1990–2025), winsorised at the 99th percentile for readability. Left: VIX implied variance ($R^2 = 0.32$, visible positive relationship). Right: RV persistence ($R^2 = -0.07$, essentially no relationship relative to the OOS mean benchmark). Despite being a substantially better variance forecaster, VIX produces lower portfolio alpha than RV, illustrating that forecasting accuracy and timing value are not equivalent objectives.

Variance forecasting. VIX implied variance is the best pure forecaster, with an out-of-sample R^2 of 0.32 and a correlation of 0.59 with next-month realized variance. RV persistence (using this month's variance as next month's forecast) yields a negative out-of-sample R^2 of -0.07, indicating forecasts worse than the unconditional mean of realized variance over the OOS period.[3] The four ML models fall between these benchmarks: Ridge ($R^2 = 0.14$), Lasso ($R^2 = 0.21$), Random Forest ($R^2 = 0.20$), and Gradient Boosting ($R^2 = 0.21$). No ML model matches VIX as a pure variance forecaster.

Portfolio performance. The portfolio results in Table 6 reveal a sharp divide between linear and tree-based models. Linear models destroy value: Ridge produces an alpha of -1.91% ($t = -0.59$) with a near-zero Sharpe ratio (-0.004) and a negative information ratio (-0.125); Lasso fares similarly ($\alpha = -0.87\%$, $t = -0.29$, IR $= -0.056$). Although these negative point estimates fall short of conventional statistical significance, the economically relevant fact is the magnitude and the structural reason behind it: any real trading desk that attempted to implement these signals would be blocked by its risk management before the alpha had a chance to materialize. The Ridge weights in our sample exceed an order of magnitude on multiple occasions (Figure 3); no fund operating under realistic risk limits could carry such positions, regardless of their statistical properties. Tree-based models, by contrast, perform well: Random Forest generates $\alpha = 3.07\%$ ($t = 1.71$, IR $= 0.322$), and Gradient Boosting attains the highest information ratio among ML strategies ($\alpha = 3.72\%$, $t = 1.90$, Sharpe $= 0.687$, IR $= 0.357$).

The sharp linear-versus-tree divide traces directly to the inverse-variance weighting rule. Portfolio performance depends on the joint distribution of forecast errors and subsequent returns, not on mean-squared error alone. Forecast errors near zero are mechanically amplified: a small underestimate of $\hat{\sigma}_t^2$ produces a disproportionately large portfolio weight. Ridge and

[3]The OOS R^2 benchmark is the full OOS period mean, not an expanding historical mean. The negative value for RV persistence reflects the highly non-stationary nature of monthly variance: structural breaks (the 2008 crisis, COVID-19) cause the persistence forecast to overshoot during regime transitions. Variance is strongly persistent at the daily frequency but considerably less so at the monthly frequency when large shocks are present.

Lasso extrapolate beyond the training range, occasionally producing near-zero or negative variance forecasts that, when floored and inverted, generate extreme weights. Tree-based models avoid this by construction: their predictions are averages over training-sample leaves and stay within the range of observed training variances, so their weights remain within bounds a real risk manager would tolerate. That models with lower R^2 than VIX produce higher information ratios follows directly from this asymmetry between forecasting accuracy and timing utility under inverse-variance weighting. Gu et al. (2020) document a similar dominance of nonlinear methods over linear regularization across asset pricing applications.

To isolate the incremental contribution of higher moments, we run an ablation: Gradient Boosting with variance-related features only (rv, iv_monthly, rv_lag1, vix_change; 4 features) versus the full 7-feature specification. The variance-only model achieves $\alpha = 3.33\%$ ($t = 1.76$, IR $= 0.332$); adding moment features raises this to $\alpha = 3.72\%$ ($t = 1.90$, IR $= 0.357$). The increment is modest: 0.39 percentage points of alpha, 0.14 in t-statistic, and 0.025 in information ratio. Feature importances confirm the picture: variance-related features account for 85.7% of total importance, with realized skewness and kurtosis contributing only 3.9% directly (the interaction term rv $\times$ skewness adds another 10.4%). This evidence indicates that Gradient Boosting's outperformance is *primarily* driven by nonlinear variance modeling rather than higher-moment information. Moment features provide a marginal improvement, but the bulk of the gains comes from the tree model's ability to capture nonlinear dynamics in the variance signal that linear forecasters miss.

Table 5: Out-of-Sample Variance Forecasting Performance (369 months)

Model	OOS R^2	Correlation
RV (persistence)	-0.07	0.47
VIX (implied var)	0.32	0.59
Ridge	0.14	0.49
Lasso	0.21	0.51
Random Forest	0.20	0.46
Gradient Boosting	0.21	0.47

Notes: OOS $R^2 = 1 - \sum(\mathrm{RV}_{t+1} - \hat{\mathrm{RV}}_{t+1})^2 / \sum(\mathrm{RV}_{t+1} - \overline{\mathrm{RV}}_{\mathrm{OOS}})^2$, where the benchmark is the mean of realized variance over the OOS period. A negative value indicates forecasts worse than this unconditional mean. All ML models use an expanding-window protocol with a 60-month minimum training period.

Table 6: Machine Learning Managed Portfolios (369 OOS months, 1990–2025)

Strategy	Alpha (%)	t-stat	β	Sharpe	IR
RV only	4.20	1.75	0.615	0.641	0.341
VIX only	3.36	1.84	0.790	0.692	0.351
ML: Ridge	-1.91	-0.59	0.196	-0.004	-0.125
ML: Lasso	-0.87	-0.29	0.156	0.039	-0.056
ML: Random Forest	3.07	1.71	0.791	0.675	0.322
ML: Gradient Boosting	**3.72**	**1.90**	0.743	0.687	**0.357**

Notes: RV and VIX baselines are evaluated over the same 369-month OOS period for comparability. Bold formatting highlights the strategy with the highest information ratio. Linear models (Ridge, Lasso) produce extreme portfolio weights due to near-zero variance predictions from extrapolation; their negative IR reflects both the alpha sign and the extreme idiosyncratic volatility their weights induce. No real fund could carry these positions under typical risk limits.

6 Robustness

6.1 Subsample Stability

Table 7 divides the VIX era (1990–2025) into three subsamples to assess the temporal stability of variance-managed strategies. The risk-adjusted return measures reveal substantial regime dependence. During 1990–2002, a period encompassing the dot-com bubble and burst, RV-managed portfolios dominate: alpha of 6.44% ($t = 2.05$), Sharpe of 0.692, information ratio of 0.560. VIX-management generates a smaller alpha of 2.77% ($t = 1.18$) with a substantially lower information ratio of 0.328. The pattern is similar in 2003–2012 (RV: $\alpha = 6.95\%$, $t = 1.83$, IR = 0.608; VIX: $\alpha = 3.31\%$, $t = 0.98$, IR = 0.330). From 2013 on-

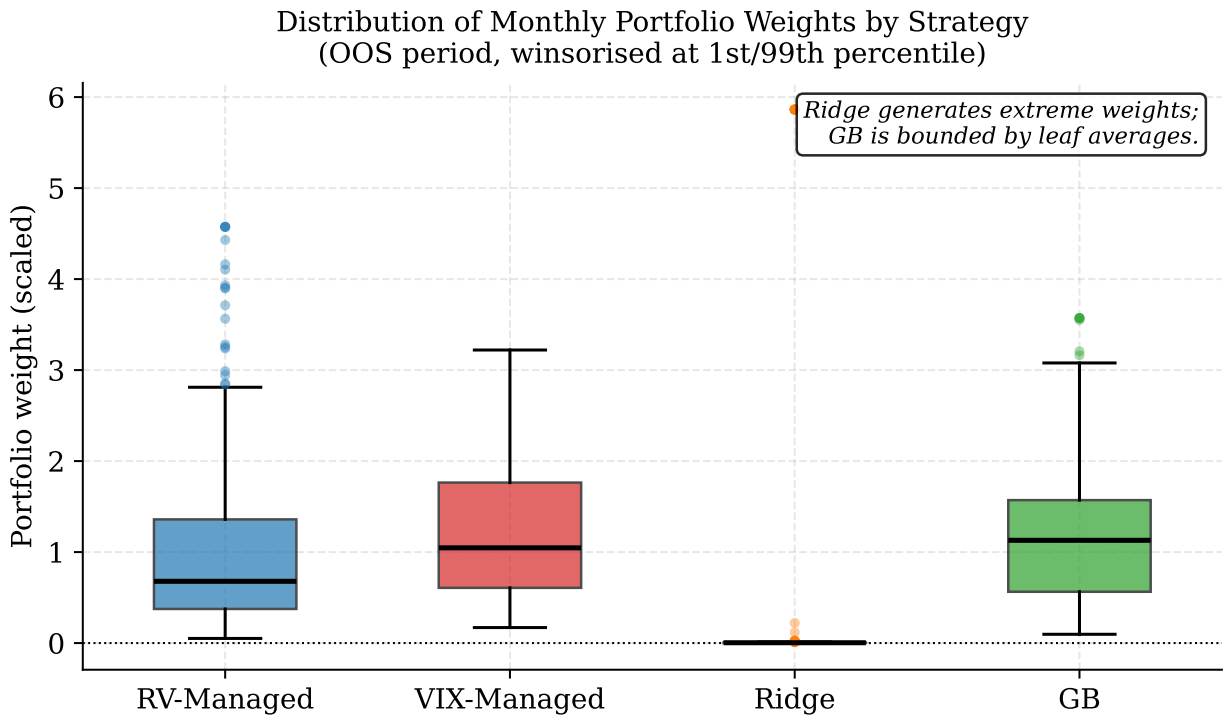

Figure 3: Distribution of monthly portfolio weights by strategy (OOS period, winsorised at the 1st/99th percentile). Ridge weights span a much wider range than tree-based models, reflecting inverse-variance amplification of near-zero variance predictions. Gradient Boosting weights are compact and stable by construction.

ward, however, the pattern reverses on every metric. The post-2012 period is characterized by low realized volatility, quantitative easing, and compressed risk premia, with a buy-and-hold Sharpe of 0.882. RV-managed alpha collapses to 0.65% ($t = 0.17$, IR $= 0.056$) and its Sharpe ratio (0.588) falls below buy-and-hold. VIX retains a small positive alpha of 2.01% ($t = 0.75$, IR $= 0.227$) and is the only managed strategy whose Sharpe (0.842) approaches the buy-and-hold benchmark in this regime. This subsample instability is consistent with Cederburg et al. (2020)'s finding that spanning regression parameters shift across market regimes, and it implies that the full-sample alpha and information ratio estimates are driven primarily by the pre-2013 period. Our main results are therefore best interpreted as conditional on a volatility regime that includes meaningful variance fluctuations, rather than as unconditional structural features of equity markets.

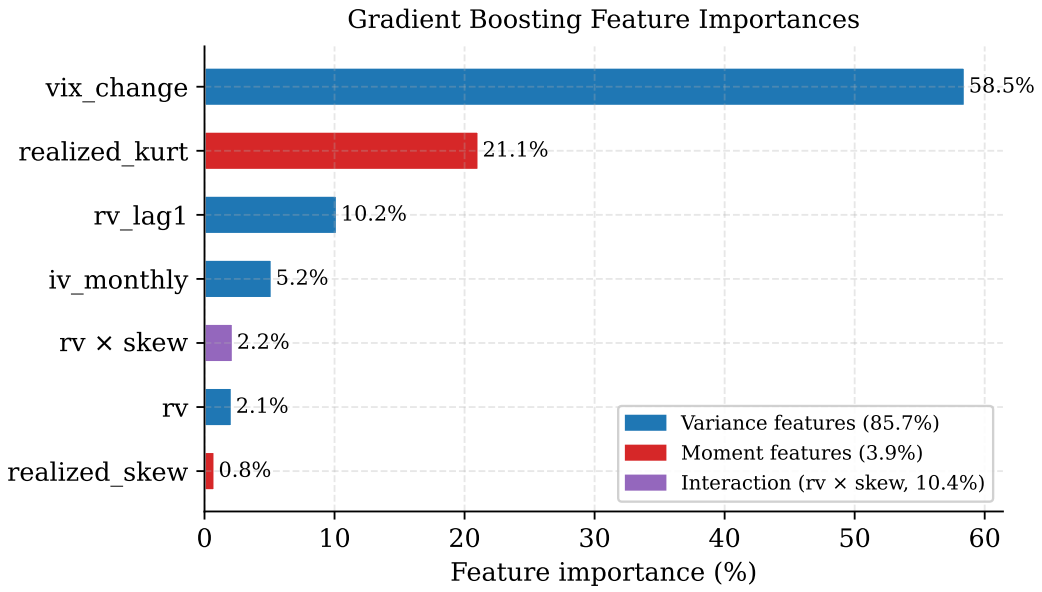

Figure 4: Gradient Boosting feature importances from the full 7-feature specification. Variance-related features (blue) account for 85.7% of total importance, with `iv_monthly` alone contributing 60.4%. Realized skewness and kurtosis (red) contribute only 3.9% directly. The interaction term `rv` × `skew` (purple) accounts for 10.4%, but its predictive content is primarily driven by the variance component.

Table 7: Subsample Stability: RV-Managed vs. VIX-Managed Portfolios

Subsample	N	RV-Managed				VIX-Managed			
		Alpha (%)	t-stat	Sharpe	IR	Alpha (%)	t-stat	Sharpe	IR
1990–2002	155	6.44	2.05	0.692	0.560	2.77	1.18	0.524	0.328
2003–2012	119	6.95	1.83	0.782	0.608	3.31	0.98	0.587	0.330
2013–2025	155	0.65	0.17	0.588	0.056	2.01	0.75	0.842	0.227
Full VIX era	431	4.16	1.99	0.653	0.354	2.91	1.80	0.670	0.317

Notes: Same spanning regression specification as Table 2. Buy-and-hold Sharpe ratios are 0.412 (1990–2002), 0.492 (2003–2012), and 0.882 (2013–2025). The post-2012 row is the most striking: the RV information ratio falls from roughly 0.6 in the earlier subsamples to 0.056, and its Sharpe drops below the buy-and-hold benchmark.

6.2 Leverage Constraints

Unconstrained variance-managed strategies can take extreme long positions during low-volatility months: the 99th percentile weight for the RV-managed portfolio is 5.16, meaning that on roughly 1% of months the strategy is nominally five times leveraged into the equity market. Real trading desks operate under hard risk limits and cannot carry such positions,

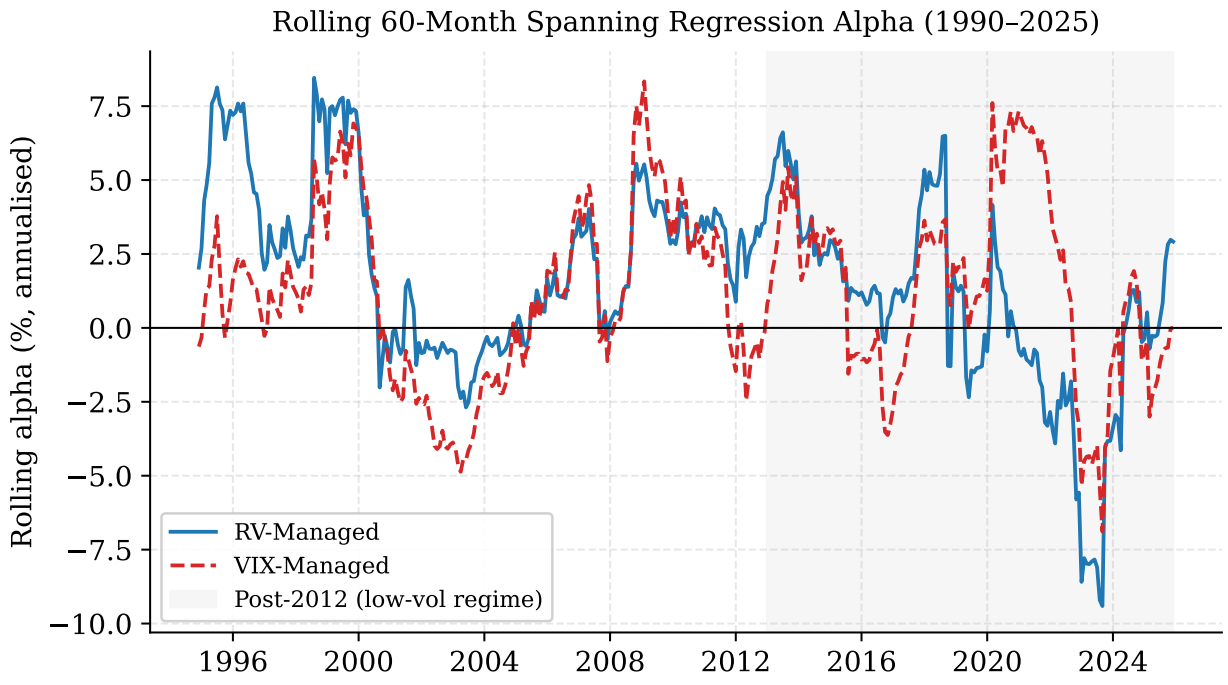

Figure 5: Rolling 60-month spanning regression alpha (annualised %) for RV-managed and VIX-managed portfolios, 1990–2025. The shaded region marks the post-2012 period. RV alpha collapses toward zero after 2013, mirroring the collapse in the RV information ratio reported in Table 7. VIX alpha is more stable across regimes, though both strategies benefit primarily from the high-volatility environment before 2013.

so any practical implementation has to cap leverage. We re-run the analysis with weight caps of 1.5 (the standard 150% Regulation T limit) and 1.0 (no leverage at all). Table 8 reports the results.

Table 8: Leverage-Constrained Strategies (1990–2025, 431 months)

Strategy	Alpha (%)	t-stat	Sharpe	IR	99th pct. weight
RV (No cap)	4.16	1.99	0.653	0.354	5.16
RV (Cap=1.5)	**3.67**	**2.10**	**0.701**	**0.374**	1.50
RV (Cap=1.0)	2.99	1.87	0.688	0.342	1.00
VIX (No cap)	2.91	1.80	0.670	0.317	3.19
VIX (Cap=1.5)	2.09	1.52	0.656	0.272	1.50
VIX (Cap=1.0)	1.34	1.16	0.633	0.208	1.00

Notes: Bold formatting highlights the strategy with the highest Sharpe and information ratio of any VIX-era specification in this paper. Caps are applied after the initial volatility scaling and the weight is renormalized so that the managed portfolio matches buy-and-hold volatility. The 99th percentile weight summarizes the upper tail of the position distribution.

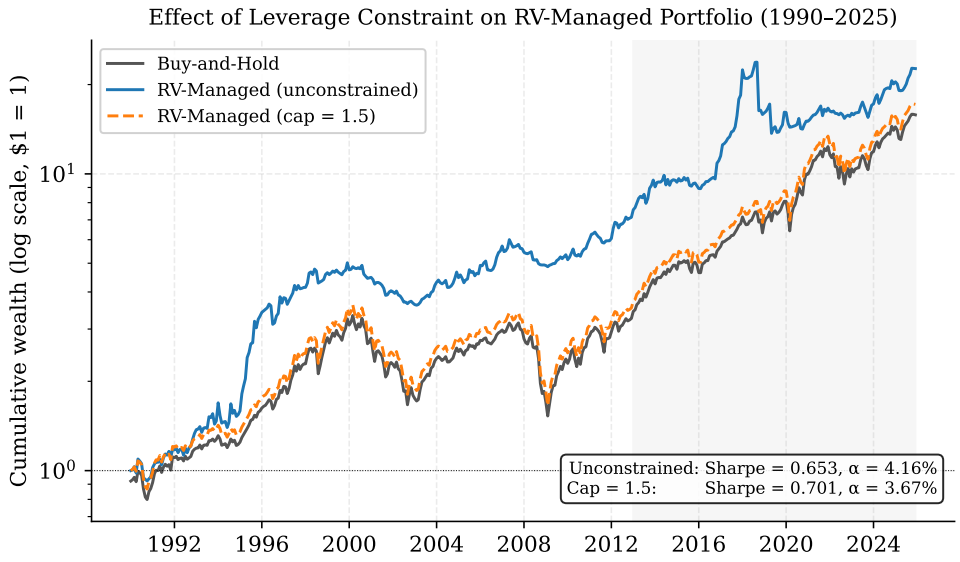

Figure 6: Cumulative wealth of the unconstrained and leverage-capped (cap = 1.5) RV-managed portfolios versus buy-and-hold, 1990–2025 (log scale). The capped strategy achieves a higher Sharpe ratio (0.701 vs. 0.653) and a higher information ratio (0.374 vs. 0.354) at the cost of a modest alpha reduction (3.67% vs. 4.16%).

Constraining leverage *improves* risk-adjusted performance for RV-managed strategies. The 1.5x cap raises the Sharpe ratio from 0.653 to **0.701** and the information ratio from 0.354 to **0.374**, both the highest of any VIX-era strategy considered in the paper. Alpha falls only modestly, from 4.16% to 3.67%, but the t-statistic actually rises (from 1.99 to 2.10) because the residual variance contracts when the noisiest weights are truncated. The 1.0x cap (zero leverage) produces a more conservative result, with Sharpe 0.688 and information ratio 0.342, both still above the unconstrained version. For VIX-managed strategies, by contrast, leverage caps reduce performance across the board, because VIX-managed weights rarely reach the cap and the binding effect is to suppress legitimate timing variation rather than truncate noisy outliers.

The economics are intuitive. When realized variance falls very low, the inverse-variance weight grows very large, but expected returns do not grow proportionally. The extreme-leverage months are noise rather than signal. Truncating these positions at 150% removes the worst of the noise without sacrificing the core timing benefit. Barroso and Santa-Clara

(2015) document the same pattern for momentum. The practical implication is that the most attractive volatility-managed strategy in our tests is also the simplest and the most implementable: scale by inverse realized variance, and cap leverage at 150%.

6.3 Distributional Timing: VIX, SKEW, and VVIX Jointly

Our core hypothesis is that VIX, SKEW, and VVIX jointly approximate the first three moments of the implied distribution (variance, skewness, kurtosis proxy), and that combining these signals via ML should outperform VIX alone. Table 9 tests this hypothesis in the two available subsamples: the full VVIX sample (2007–2025, 228 months) and the SKEW sample (2011–2025, with CBOE SKEW available). Features include VIX implied variance, VVIX level, the volatility term structure slope (VIX3M − VIX), normalized SKEW, realized variance, lagged realized variance, and VIX change.

The results are uniformly insignificant on every metric. In the full sample the best performer is Gradient Boosting ($\alpha = 4.5\%$, $t = 1.42$, IR ≈ 0.38); in the SKEW sample, GB achieves $\alpha = 4.2\%$ ($t = 0.96$, IR ≈ 0.30). No strategy reaches $t > 1.5$ in either subsample. Even the baseline RV and VIX strategies are insignificant in these short samples ($t < 1.1$), indicating a broader power problem rather than a specific failure of distributional signals. GB feature importances show that realized variance dominates (66% importance), with the distributional indices (VVIX, slope, normalized SKEW) contributing modestly. The information ratios echo the alpha story: the joint distributional signal does not deliver materially better risk-adjusted timing than VIX alone in the data we can access.

Two explanations cannot be distinguished empirically. First, the samples are short: 228 months total (168 OOS) for the VVIX sample and 180 months total (120 OOS) for the SKEW sample, providing limited statistical power. Second, the 2007–2025 period includes the post-2012 bull market in which all variance-managed strategies structurally underperform (Section 6.1). If historical SPX option chains were available (e.g., via OptionMetrics), ARM-implied monthly moments could be computed back to 1996, enabling a proper long-sample

test of the distributional timing hypothesis.

Table 9: Distribution-Managed Portfolios: VIX, SKEW, and VVIX

Sample	Strategy	N	Alpha (%)	t-stat	Sharpe	IR
Full (2007+)	RV baseline	228	3.0	0.90	0.564	0.21
	VIX baseline	228	2.7	1.08	0.664	0.25
	ML: Ridge	168	< 0	< 0	—	< 0
	ML: Gradient Boosting	168	4.5	1.42	0.380	0.38
SKEW (2011+)	VIX baseline	179	1.8	0.73	0.812	0.19
	VIX+SKEW+VVIX (moment)	179	2.0	0.79	0.823	0.20
	ML: Gradient Boosting	120	4.2	0.96	0.314	0.30
No strategy achieves $t > 1.5$. Buy-and-hold Sharpe: 0.637 (2007+), 0.865 (2011+).						

Notes: The full sample (2007+) is limited by VVIX availability; the SKEW sample (2011+) additionally requires CBOE SKEW. Features include VIX implied variance, VVIX, term structure slope (VIX3M−VIX), normalized SKEW, realized variance, lagged RV, and VIX change. ML models use expanding-window OOS evaluation with a 60-month minimum training period. The moment strategy (VIX+SKEW+VVIX) uses an additive weight adjustment $w_t = (c/\hat{\sigma}_t^2)(1 + \alpha_s z_{\text{skew},t} + \alpha_k z_{\text{kurt},t})$ with $\alpha_s = \alpha_k = -0.05$, a 12-month minimum window, and a zero lower bound on weights. This specification differs from the main moment-managed portfolios in Section 5.3 ($\lambda = 0.2$, clip $[0.5, 1.5]$, 24-month minimum) due to the shorter available sample.

6.4 ARM-Implied Volatility as an Alternative Signal

As a proof of concept, we calibrate ARM (zur Linden, 2023) to 114 trading days of SPX options (October 2024 to March 2025) and compare its implied variance to VIX as a predictor of next-day squared excess returns. ARM achieves an out-of-sample R^2 of 0.073 versus 0.059 for VIX over the same period. This comparison operates at the daily frequency and is not a portfolio timing test; we include it only to illustrate that ARM's full-distribution implied variance contains incremental information beyond VIX. A proper long-sample evaluation at the monthly portfolio frequency would require historical SPX option chains (e.g., from OptionMetrics, covering 1996 onward) and remains an important direction for future work.

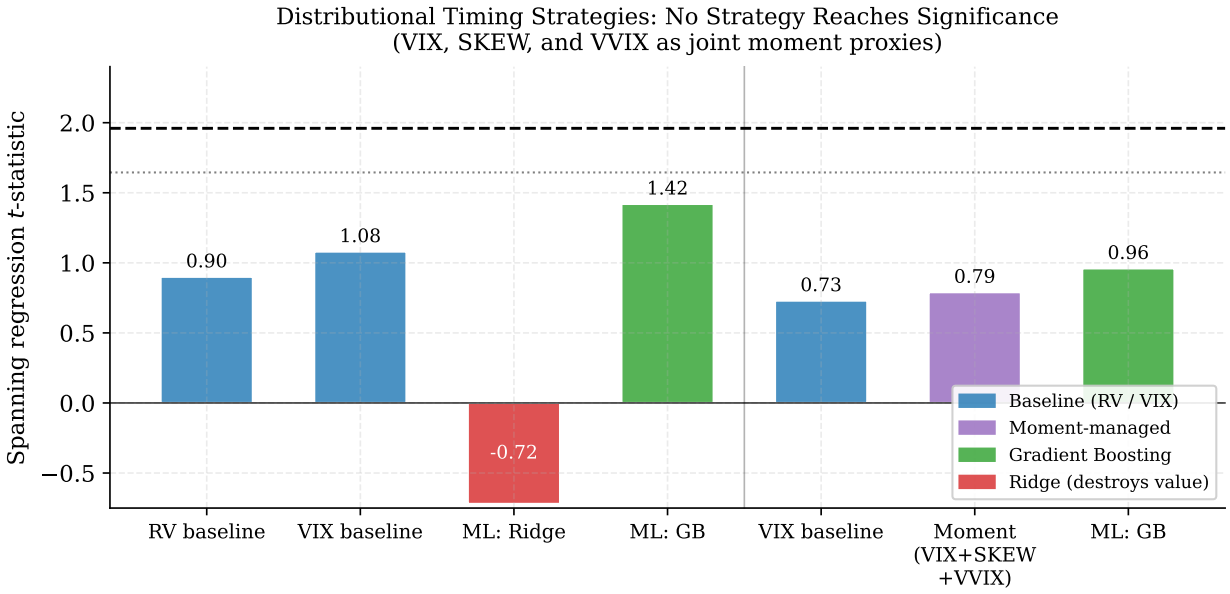

Figure 7: Spanning regression t-statistics for distributional timing strategies. The left group covers the full VVIX sample (2007–2025); the right group covers the shorter SKEW sample (2011–2025). Dashed and dotted lines mark the 5% ($t = 1.96$) and 10% ($t = 1.645$) significance thresholds. No strategy reaches conventional significance in either sample, and even the RV and VIX baselines are insignificant, indicating a power and regime problem rather than a specific shortcoming of the distributional signals.

7 Conclusion

Options markets price the full return distribution, yet for a long-term investor evaluating strategies by risk-adjusted return, the additional information they encode does not reliably improve on a simple backward-looking rule. On a 100-year sample, realized variance timing generates a spanning regression alpha of 4.56% ($t = 3.10$, IR = 0.310), matching Moreira and Muir (2017)'s reported 4.86% ($t = 3.11$) almost exactly. Replacing realized variance with VIX, a substantially better variance forecast (OOS $R^2 = 0.32$ versus -0.07), *lowers* alpha to 2.91% and the information ratio to 0.317. The raw Sharpe ratio narrowly favors VIX, but only because VIX-managed weights inherit more equity-premium beta; once that channel is stripped out, the information ratio agrees with alpha. All eight hand-crafted higher-moment adjustments reduce both alpha and the information ratio. Better forecasting and better risk-adjusted timing are not the same objective, and the additional moments captured by

SKEW and VVIX do not bridge the gap.

Gradient Boosting is the most successful machine learning specification, achieving $\alpha = 3.72\%$ ($t = 1.90$, IR $= 0.357$), the highest information ratio of any ML strategy in our analysis. An ablation shows why. With variance features alone, GB already reaches IR $= 0.332$; adding skewness and kurtosis raises this only to 0.357. The tree structure captures nonlinear variance dynamics that linear forecasters miss, with higher moments contributing a small secondary increment. This is consistent with Kostakis et al. (2011), who find that higher moments are not the source of outperformance in their utility-maximization framework. There is also a practical lesson tied to risk management: because portfolio weights are proportional to $1/\hat{\sigma}_t^2$, any forecast method that can produce near-zero variance predictions will generate extreme portfolio positions. Tree-based models avoid this by construction; linear methods do not, and their negative alphas in our tests are largely academic, since no risk-limited fund could carry the resulting weights.

The most attractive practical strategy in our tests is in fact the simplest. RV-managed with a 150% leverage cap delivers Sharpe 0.701 and information ratio 0.374, both higher than any other VIX-era strategy considered in this paper, including Gradient Boosting. The unconstrained version reaches the 99th-percentile weight of 5.16, a position no real desk would carry. Capping leverage at 1.5x removes the noisiest tail of the weight distribution without sacrificing the core timing benefit, and the resulting strategy has the additional virtue of being directly implementable under standard Reg-T constraints.

One important caveat applies specifically to the ML results. We do not report certainty-equivalent gains, portfolio turnover, or transaction costs, which Cederburg et al. (2020) identify as essential for assessing real-time implementability. ML-managed portfolios likely turn over more than simple RV-scaling, and the alpha gap between Gradient Boosting (3.72%) and simple RV (4.16%) is already narrow enough that transaction costs could close or reverse it. The same caveat is much weaker for the leverage-capped RV strategy, which differs from the unconstrained M&M baseline only by an additive position constraint and should have

similar turnover.

Two directions for future research follow naturally. First, the distributional timing tests in Section 6.3 are limited to 180–228 months by CBOE data availability; accessing historical SPX option chains via OptionMetrics would allow ARM-implied moments to extend the test back to 1996, enabling a proper long-sample evaluation of whether the full implied distribution improves on VIX alone. Second, neural networks, which Gu et al. (2020) find dominate tree-based methods for equity prediction, remain untested here and may capture richer nonlinear variance dynamics. Whether the gap between what options markets encode and what risk-adjusted timing strategies can exploit will close with longer samples or richer methods is the central open question this paper raises.

References

Gurdip Bakshi, Nikunj Kapadia, and Dilip Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16 (1):101–143, 2003. doi: 10.1093/rfs/16.1.101.

Pedro Barroso and Pedro Santa-Clara. Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120, 2015. doi: 10.1016/j.jfineco.2014.11.010.

Tim Bollerslev, George Tauchen, and Hao Zhou. Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492, 2009. doi: 10.1093/rfs/hhp008.

Miloš Božović. VIX-managed portfolios. *International Review of Financial Analysis*, 95: 103353, 2024. doi: 10.1016/j.irfa.2024.103353.

Douglas T. Breeden and Robert H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978. doi: 10.1086/296025.

Scott Cederburg, Michael S. O'Doherty, Feifei Wang, and Xuemin (Sterling) Yan. On the performance of volatility-managed portfolios. *Journal of Financial Economics*, 138(1): 95–117, 2020. doi: 10.1016/j.jfineco.2020.04.015.

Jennifer Conrad, Robert F. Dittmar, and Eric Ghysels. Ex ante skewness and expected stock returns. *Journal of Finance*, 68(1):85–124, 2013. doi: 10.1111/j.1540-6261.2012.01795.x.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. doi: 10.1016/0304-405X(93) 90023-5.

Jeff Fleming, Chris Kirby, and Barbara Ostdiek. The economic value of volatility timing. *Journal of Finance*, 56(1):329–352, 2001. doi: 10.1111/0022-1082.00327.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Campbell R. Harvey and Akhtar Siddique. Conditional skewness in asset pricing tests. *Journal of Finance*, 55(3):1263–1295, 2000. doi: 10.1111/0022-1082.00247.

Jens Carsten Jackwerth and Mark Rubinstein. Recovering probability distributions from option prices. *Journal of Finance*, 51(5):1611–1631, 1996. doi: 10.1111/j.1540-6261.1996.tb05219.x.

Alexandros Kostakis, Nikolaos Panigirtzoglou, and George Skiadopoulos. Market timing with option-implied distributions: A forward-looking approach. *Management Science*, 57 (7):1231–1249, 2011. doi: 10.1287/mnsc.1110.1346.

Alan Moreira and Tyler Muir. Volatility-managed portfolios. *Journal of Finance*, 72(4): 1611–1644, 2017. doi: 10.1111/jofi.12513.

David Shimko. Bounds of probability. *Risk*, 6(4):33–37, 1993.

David Shimko and Masumeh (Nadia) Babaei. Dynamic implied probability (DIP): An alternative option pricing framework. Working paper, Wall Street Scholars LLC, 2025.

Ernst zur Linden. ARM: The analytic recovery method. *Journal of Derivatives*, 30(4):8–27, 2023. doi: 10.3905/jod.2023.30.4.008.

A The Analytic Recovery Method

The Analytic Recovery Method (ARM), introduced by zur Linden (2023), recovers arbitrage-free risk-neutral densities from option prices via a parametric transformation of the standard normal distribution. The $N = 1$ (cubic) specification defines the asset price as a polynomial function of a standard normal variate y:

$$S(y) = a_0 + a_1 y + a_2 y^2 + a_3 y^3, \quad y \sim \mathcal{N}(0,1). \tag{10}$$

The risk-neutral density of S follows from the change-of-variables formula:

$$q\big(S(y)\big) = \frac{n(y)}{S'(y)}, \tag{11}$$

where $n(y)$ is the standard normal density and $S'(y) = a_1 + 2a_2 y + 3a_3 y^2$. Monotonicity of $S(\cdot)$ requires $a_2^2 < 3a_1 a_3$ with $a_1, a_3 > 0$, ensuring a valid density. The forward price constraint sets the risk-neutral mean: $\mu = a_0 + a_2$, reducing the free parameters to three (a_1, a_2, a_3).

ARM yields closed-form call prices under this specification. Following zur Linden (2023), the call price is

$$C(K) = D(T)\left[\Phi(-y_K)(\mu - K) + n(y_K)\big(a_1 + a_2 y_K + a_3(y_K^2 + 2)\big)\right], \tag{12}$$

where $y_K = S^{-1}(K)$, $D(T) = e^{-rT}$ is the discount factor, and $\Phi(\cdot)$ denotes the standard normal CDF. Calibration proceeds by minimizing the sum of squared pricing errors between model and observed option mid-prices via Nelder-Mead optimization. All implied moments (mean, variance, skewness, and kurtosis) are expressible in terms of $a_0, \ldots, a_3$ using standard normal moment formulas.

We calibrate ARM to SPX options over 114 trading days (October 2024 to March 2025), using calls expiring March 21, 2025, with strikes in the range 85–115% of the forward price

($\mu = S_0 e^{rT}$, $r = 4.5\%$). The average root-mean-square pricing error across all 114 days is \$1.43. Implied volatility is near a random walk (AR(1) = 0.994, $R^2 = 97.1\%$), while skewness and kurtosis are moderately persistent ($R^2 \approx 48\%$). Changes in implied skewness and kurtosis are highly negatively correlated ($\rho = -0.876$), consistent with the -0.77 reported by Shimko and Babaei (2025) for SPX options using a related dynamic implied probability framework.

B Additional Results

B.1 Full Strategy Summary

Table 10 reports results for all 15 primary strategies evaluated in the paper, organized by signal type. The RV-managed portfolio over the full 1926–2025 sample achieves the highest t-statistic (3.10), reflecting its 100-year sample length, while Gradient Boosting attains the highest information ratio (0.357) among the 369-month out-of-sample VIX-era strategies. The RV-managed portfolio with a 150% leverage cap (reported in Section 6.2 rather than this table) reaches Sharpe 0.701 and IR 0.374, the highest of any VIX-era specification considered. Buy-and-hold Sharpe ratios range from 0.450 (full sample) to 0.601 (VIX era), reflecting the different sample compositions.

B.2 Realized Moments Time Series

Realized variance computed from daily market excess returns is strongly clustered, with extreme spikes during the Great Depression (1929–1932), the 1987 crash, the 2008 financial crisis, and the COVID-19 shock of March 2020. Realized skewness fluctuates around its mean of -0.11 with modest persistence (AR(1) = 0.11), while realized excess kurtosis is essentially white noise (AR(1) = 0.02), spiking only during crash months. The near-zero persistence of realized kurtosis explains why hand-crafted kurtosis adjustments fail to improve portfolio performance: last month's kurtosis carries almost no information about next

Table 10: Full Strategy Summary

Type	Strategy	N	Alpha (%)	t-stat	β	Sharpe	IR
Baseline	RV (full sample)	1,193	4.56	3.10	0.601	0.518	0.310
	RV (VIX era)	431	4.16	1.99	0.629	0.653	0.354
	VIX	431	2.91	1.80	0.795	0.670	0.317
Moments (full)	RV + Skew	1,170	4.34	2.90	0.593	0.493	0.291
	RV + Kurt	1,170	4.25	2.79	0.582	0.483	0.283
	RV + Skew + Kurt	1,170	4.29	2.82	0.578	0.484	0.285
Moments (VIX)	VIX + R.Skew	408	2.79	1.59	0.766	0.643	0.289
	VIX + R.Kurt	408	2.57	1.51	0.783	0.638	0.275
	VIX + R.Skew+Kurt	408	2.36	1.33	0.764	0.613	0.243
	VIX + CBOE SKEW	168	1.53	0.54	0.751	0.808	0.161
	VIX + CBOE SKEW + R.Kurt	168	0.47	0.16	0.747	0.730	0.049
ML	Ridge	369	−1.91	−0.59	0.196	−0.004	−0.125
	Lasso	369	−0.87	−0.29	0.156	0.039	−0.056
	Random Forest	369	3.07	1.71	0.791	0.675	0.322
	Gradient Boosting	369	**3.72**	**1.90**	0.743	0.687	**0.357**

month's distribution.

B.3 CBOE Indices Time Series

VIX, CBOE SKEW, and VVIX each capture a different dimension of the option-implied distribution. VIX is the most persistent (AR(1) = 0.81), consistent with the slow-moving nature of variance expectations. CBOE SKEW averages 131.10 (well above its neutral value of 100), indicating persistent negative implied skewness in S&P 500 options.

B.4 ML Feature Importance

In the distributional timing model of Section 6.3, realized variance accounts for approximately 66% of Gradient Boosting's total feature importance in the SKEW-sample specification. VIX implied variance and the monthly VIX change contribute moderately, while CBOE SKEW, VVIX, and the term structure slope add only marginal incremental value. These importances confirm that the distributional indices provide limited information beyond what is already captured by variance signals in this sample.