VolEdge: Measuring and Trading the Variance Risk Premium with Model-Free Risk-Neutral Moments

Tanishk Yadav
MS Financial Engineering
NYU Tandon School of Engineering
New York, NY, USA

July 2026

Abstract

VolEdge is a full-stack research platform built around a single thesis: the gap between what options *price* and what the underlying *realizes* is the risk premium, and that gap can be measured and traded. The platform extracts the risk-neutral distribution model-free via Bakshi-Kapadia-Madan (BKM) moments and compares it, horizon-matched, against the physical distribution estimated from realized returns; the spread $Q - P$ is surfaced for variance, skew, and kurtosis. A no-look-ahead walk-forward engine then backtests premium-driven strategies. We report two honest results on free data (2019-2024, net of modeled costs): a GARCH(1,1) volatility-targeted strategy beat SPY buy-and-hold on a risk-adjusted basis (Sharpe 0.85 vs. 0.68) with 41% lower maximum drawdown, while a naive long-only variance-premium harvest *lost* money (Sharpe -0.16), empirically confirming the premium as compensation for crash risk rather than harvestable alpha. The BKM extractor is validated against Black-Scholes chains of known volatility, and the backtest engine reproduces buy-and-hold to ± 0.000 pp at any rebalance frequency.

Keywords: variance risk premium, risk-neutral moments, BKM, volatility targeting, walk-forward backtesting, options.

1 Introduction

In a frictionless world the option-implied (risk-neutral, Q) distribution would equal the physical (real-world, P) distribution. It does not: Q embeds a risk premium, so an option's price is *not* a probability. The wedge between the two measures is the economic object of interest. VolEdge is organized entirely around keeping Q and P as distinct, separately estimated quantities and studying their difference.

The platform measures the premium live for any ticker, and - because a historical per-name option surface is not available on free data - backtests the idea at the index level using the VIX as the risk-neutral volatility proxy. The two measurements are deliberately distinct; the backtest's role is to test whether the premium the live tool surfaces is economically real and harvestable, not to reproduce BKM historically.

2 Methods

2.1 Risk-neutral moments (BKM)

Following Bakshi, Kapadia, and Madan (2003), any twice-differentiable payoff is spanned by a continuum of out-of-the-money (OTM) options. The variance, cubic, and quartic contracts therefore have known strike weights, and their prices V, W, X are obtained by integrating OTM call and put prices across strikes. With log-moneyness $\ell = \ln(K/S)$ the call and put integrands unify:

$$V = \int \frac{2(1-\ell)}{K^2} O(K)\, dK, \quad W = \int \frac{6\ell - 3\ell^2}{K^2} O(K)\, dK, \quad X = \int \frac{12\ell^2 - 4\ell^3}{K^2} O(K)\, dK,$$

where $O(K)$ is the OTM option price. Risk-neutral variance, skewness, and (excess) kurtosis follow in closed form, with a single e^{rT} entering the moment formulas (not the integrands). Integration uses composite Simpson's rule over the irregular strike grid with a trapezoidal fallback. No distributional assumption is made.

2.2 Physical moments and the premium

The physical distribution is estimated horizon-matched: because skew and kurtosis are horizon-dependent, they are computed on *overlapping T-day log returns* rather than daily returns, where T matches the BKM tenor. The premium is the difference at each tenor,

$$\text{VRP} = \sigma_Q^2 - \sigma_P^2, \qquad \text{skew premium} = s_Q - s_P, \qquad \text{kurt premium} = k_Q - k_P.$$

A persistently positive variance premium means options price more variance than realizes; a more negative Q-skew than P-skew means the market pays up for crash protection beyond what history delivered.

2.3 Mean-reversion diagnostics

As a descriptive layer, the platform reports three independent reversion estimators on price, log-price, and realized volatility: the Ornstein-Uhlenbeck half-life from the AR(1) speed ($-\ln 2 / \ln \phi$), the Hurst exponent, and a Dickey-Fuller-style t-statistic, plus a rolling half-life. A series is labelled mean-reverting only when a finite in-sample half-life corroborates a Hurst below 0.5.

2.4 Walk-forward engine

Strategies are evaluated day-by-day; at each decision point the strategy sees only data up to the prior close. Strategy code is AST-validated and sandboxed (whitelisted numerical imports, no I/O or network). Positions are marked to market *before* any rebalance, so returns accrue correctly at every rebalance frequency; turnover cost is charged on the actual drift-to-target trade with a per-leg model (spread, impact, slippage, commission).

3 Validation

BKM. On a Black-Scholes chain of known 20% volatility the extractor recovers $\sigma_Q \approx 0.20$ with skewness and excess kurtosis near zero, and converges toward truth as strike coverage widens; the residual bias is strike truncation, an inherent limitation of BKM on any finite chain. A synthetic put-skew chain correctly yields negative risk-neutral skew. On live SPY chains (via free Yahoo

Finance data) the moments show a sensible term structure (e.g. 30d skew ≈ -2.2, kurtosis ≈ 9; 60d ≈ -1.6, 3.4), consistent with published equity-index estimates.

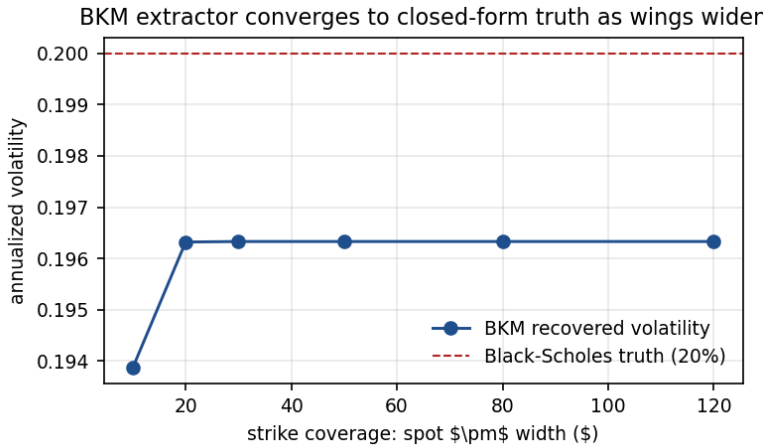


Figure 1: BKM extractor versus closed-form truth. On a Black-Scholes chain of known 20% volatility, the recovered risk-neutral volatility converges to the pricing volatility as strike coverage widens; the small residual at narrow wings is strike truncation, inherent to BKM on any finite chain. This is the regression test that guards the engine (an earlier version carried a spurious $1/T$ and doubled e^{rT}, inflating moments $\sim$2.7x).

Engine. A 100%-SPY strategy reproduces SPY buy-and-hold to ± 0.000 percentage points at rebalance frequencies of 1, 5, 21, and 63 days, confirming the mark-to-market is correct and frequency-independent.

4 Results

All figures are out-of-sample, on free Yahoo Finance data, net of modeled transaction costs.

Strategy	Return	Ann.	Vol	Sharpe	Max DD
GARCH vol-managed SPY (2020-24)	+75.1%	11.9%	13.9%	**0.85**	**−19.7%**
SPY buy-and-hold (same)	+94.6%	14.3%	21.0%	0.68	−33.7%
VIX variance-premium harvest (2019-24)	−15.4%	−2.9%	17.8%	−0.16	−44.0%
SPY (2019-24)	+123.9%	15.1%	20.1%	0.75	−33.7%

Table 1: Backtested premium-driven strategies vs. SPY.

GARCH volatility targeting (the positive result). A GARCH(1,1) conditional-variance forecast scales SPY exposure to a constant 15% volatility target (Moreira-Muir), the remainder in T-bills. On an apples-to-apples basis it delivered a higher Sharpe (0.85 vs. 0.68, +25%) at 41% lower maximum drawdown and a realized 13.9% volatility. Because volatility is forecastable while expected returns are not, scaling down in high-forecast-vol states improves the realized risk/return trade-off - and the variance premium makes those states expensive to be long, so the de-risking is compensated.

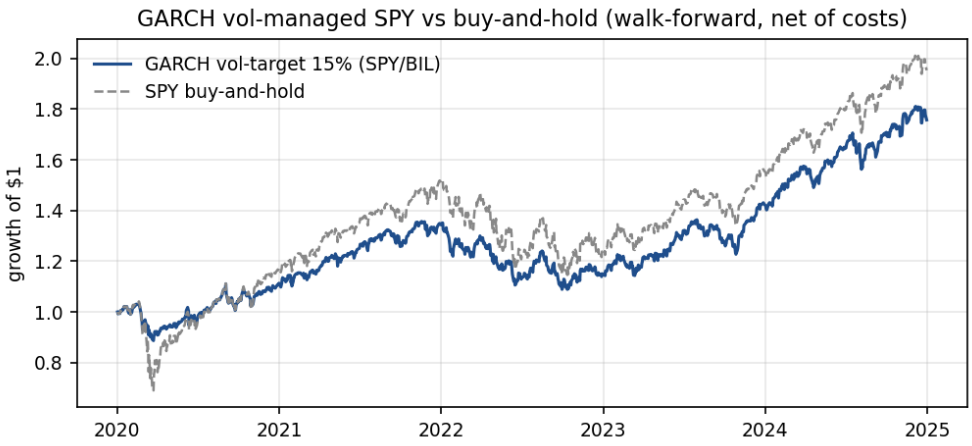


Figure 2: GARCH vol-managed SPY versus buy-and-hold, walk-forward with a 2019 training burn-in so the evaluation window (2020-2024) includes the COVID crash. The strategy gives up total return for a markedly better path: Sharpe 0.86 versus 0.68, maximum drawdown −19.7% versus −33.7%. Reproduce with `scripts/make_paper_figures.py`.

Variance-premium harvest (the reality check). A long-only short-volatility harvester (short via SVXY when VIX^2−realized is positive, scaling down as VIX rises and stepping out above VIX 30) *lost* money over 2019-2024 net of costs. This is the intended finding, not a hidden failure: across the COVID-2020 and 2022 volatility spikes the premium behaves as compensation for crash risk, not harvestable alpha with simple long-only instruments - which is precisely why the premium exists. The platform thus both measures the premium and demonstrates that harvesting it naively is dangerous.

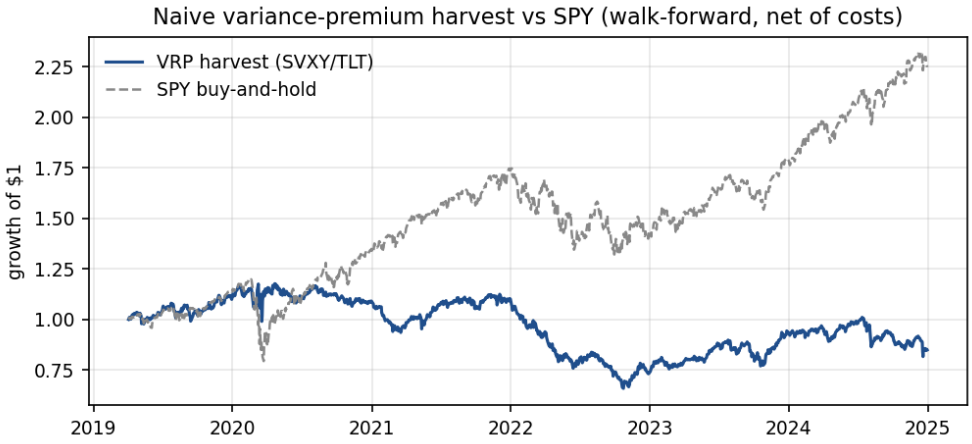


Figure 3: The reality check: the naive variance-premium harvest versus SPY, walk-forward, net of costs (2019-2024). The COVID-2020 and 2022 volatility spikes are exactly where the short-vol position pays out its accumulated premium as crash losses.

5 Limitations

Yahoo Finance quotes are ∼15-minute-delayed snapshots with no SLA; deep-OTM strikes can be stale (mitigated by a liquidity filter and a ±20% strike window). Kurtosis is the least reliable BKM

moment on free data. The live premium (per-name BKM snapshot) and the backtest (index VIX proxy) measure the idea two different ways. The walk-forward accounting is weight-based rather than a full position ledger. None of the reported numbers are tuned via parameter search behind the headline.

6 Conclusion

VolEdge closes the loop from measurement to execution for the variance risk premium: model-free risk-neutral moments versus horizon-matched realized moments, validated extractors, and a frequency-correct walk-forward engine. The empirical takeaway is that the premium is easy to measure and hard to keep - disciplined volatility targeting is compensated, while naive premium harvesting is not.