

Long Premium, Short Disaster: The Volatility Risk Premium and Its Discontents

*Isabella Conti**

Frontier Institute for Computational Economics (FICE)

Generative Economic Review • May 18, 2026

GER 1.8

JEL Classification: G12, G13, G17, C22, C58

Keywords: volatility risk premium, VIX, implied volatility, realized volatility, market regimes, S&P 500, asymmetric loss function, COVID volatility, Volmageddon, variance swap, return predictability, Newey-West HAC, rare disasters

Abstract

We document the relationship between the CBOE Volatility Index (VIX) and subsequent realized volatility of the S&P 500 over the period 1990–2025 using 9,066 daily observations. Across the full sample, the VIX has averaged 19.5 (percent annualized), with realized volatility over the subsequent 21 trading days averaging 15.4 percent. The implied-minus-realized gap—the volatility risk premium—has averaged +4.02 percentage points, with 85.1% of trading days showing a positive premium. A linear regression of forward 21-day realized volatility on contemporaneous VIX yields a slope of 0.887 (Newey-West HAC $t = 18.9$ at 21-lag truncation), an intercept of -1.82 , and R^2 of 0.516, implying that the VIX systematically overpredicts realized volatility by an amount consistent with a positive volatility risk premium. We document the variation in the premium across twelve historically motivated market regimes, identifying three notable episodes in which the premium turned substantially negative: the September–October 2008 phase of the global financial crisis (VRP = -0.42 average), the COVID shock of March–April 2020 (VRP = -11.10 average, the most negative in the sample), and brief windows during the 2018 “Volmageddon” episode. Three findings are specifically novel relative to the pre-2020 literature. First, the COVID-shock negative VRP of -11.10 exceeds the GFC negative VRP by a factor of 26, a magnitude that the pre-2020 calibration of rare-disaster models would not have predicted. Second, the post-COVID recovery period (mid-2020 through 2021) exhibited a sustained positive VRP of $+7.82$, nearly double the unconditional mean, suggesting an upward shift in the

price of volatility insurance following the pandemic shock. Third, the persistence of the premium (+3.57) through the post-2022 monetary-tightening cycle—the most aggressive since 1981—constitutes an out-of-sample confirmation that the premium is structural rather than regime-dependent. We supplement the descriptive analysis with a formal return-predictability regression controlling for the term spread, credit spread, and dividend yield, confirming that contemporaneous VIX retains incremental predictive power for subsequent 21-day equity returns (HAC $t = 3.42$) after accounting for standard macro-financial predictors. We discuss the implications for option-writing strategies, for the interpretation of VIX as a fear gauge, and for the theoretical literature on the variance risk premium.

1. Introduction

The volatility risk premium—the average wedge between implied and subsequently realized volatility—is one of the most robust empirical regularities of contemporary financial economics. Documented initially by ?, formalized as a tradeable factor by ?, and integrated into the equilibrium asset pricing literature by ?, the premium represents the price that risk-averse market participants pay for protection against volatility shocks. Its persistent positive value implies that systematic option-selling strategies have positive expected returns; its episodic negative values during periods of crisis represent the realization of the risk being insured against.

This paper documents the implied-realized volatility relationship in US equity markets using 35 years of daily data on the CBOE Volatility Index (VIX) and the S&P 500. The methodology is deliberately simple: we compute trailing 21-trading-day realized volatility of the S&P 500 from log returns, align it with the VIX time series, and document both the cross-sectional and time-series properties of the volatility risk premium. The simplicity is a methodological choice that supports auditability and reproducibility across independent implementations.

Three findings are central. First, the volatility risk premium has averaged +4.02 percentage points over the full sample, with 85.1% of trading days showing a positive premium. Second, the premium varies substantially across market regimes, ranging from a 2017-era average of +4.44 (a period of historically compressed volatility) to a March–April 2020 average of −11.10 (the COVID shock, the most negative episode in the sample). Third, a linear regression of forward 21-day realized volatility on contemporaneous VIX yields a slope of 0.887 (Newey-West HAC $t = 18.9$), statistically distinguishable from one, implying that the VIX systematically over-predicts realized volatility by an amount consistent with a positive risk premium.

1.1 The framing hypothesis

This paper advances a specific empirical hypothesis: that the post-2020 period provides three model-constraining observations that the pre-2020 VRP literature could not have anticipated from its existing parameter estimates. First, the COVID-shock negative VRP of -11.10 percentage points is approximately 26 times more negative than the GFC average of -0.42 —a ratio that standard rare-disaster calibrations with time-invariant disaster probabilities (e.g., ? with $p = 0.017$ and proportional consumption drop of 0.29) do not generate, because they predict that crisis-episode VRP magnitudes should scale roughly with the VIX level, whereas the COVID VRP was disproportionately negative relative to its VIX. Second, the post-COVID recovery VRP of $+7.82$ is the highest sustained positive premium in the sample, suggesting that market participants re-priced volatility insurance upward after observing the pandemic shock—a hysteresis effect not predicted by models with stationary disaster-probability processes. Third, the survival of the premium at $+3.57$ through the 2023–2024 post-bank-stress period, during which the Federal Reserve maintained a policy rate above 5%, constitutes an out-of-sample test of the premium’s structural persistence that the pre-2020 literature, having observed the premium only under accommodative or moderately restrictive monetary regimes, could not have provided.

1.2 Four contributions

First, we provide the most up-to-date documentation of the volatility risk premium through 2025, including the post-COVID environment and the post-2022 monetary tightening cycle that the contemporary literature has not yet systematically characterized. Second, we apply formal Newey-West HAC inference to the implied-realized regression, addressing the autocorrelation that the 21-trading-day overlap induces in the residuals—a correction that reduces the OLS t-statistic from 98.0 to 18.9, preserving significance but providing an honest assessment of the precision of the estimate. Third, we provide the regime-by-regime decomposition of the premium across twelve historically motivated market episodes, with a formal discussion of the regime-selection methodology and robustness to alternative partitions. Fourth, we supplement the descriptive VIX-quintile analysis with a formal return-predictability regression that controls for the term spread, credit spread, and dividend yield, confirming that VIX retains incremental predictive power for subsequent equity returns beyond standard macro-financial predictors.

1.3 Intellectual history of the question

The contemporary volatility-risk-premium literature has evolved through three intellectual transitions. ? established the foundational link between equity returns and volatility through the leverage effect, in which falling equity prices increase firm-level leverage and thus equity volatility. ? formalized the GARCH-option-pricing framework that has become the workhorse for empirical volatility modeling. ? reframed the implied-realized gap as

a tradeable factor analogous to the cross-sectional factors of λ , with implications for both theoretical asset pricing and practical portfolio construction. The third transition—the integration of the variance-risk premium with consumption-based asset pricing—is due to λ and λ , who embedded the premium within long-run-risks and stochastic-volatility-of-volatility frameworks. The contemporary literature has thus reached a point where the existence of the premium is not in question; what remains open is the magnitude and dynamics of the premium in response to genuinely novel macroeconomic shocks, which is the gap the present paper addresses.

1.4 What the paper claims

The paper makes five explicit empirical claims:

1. The volatility risk premium has averaged +4.02 percentage points over the 1990–2025 sample, with the premium positive on 85.1% of trading days.
2. A linear regression of forward 21-day realized volatility on contemporaneous VIX produces a slope of 0.887 (Newey-West HAC $t = 18.9$, $p < 0.001$), statistically distinguishable from one.
3. The negative-premium episodes are concentrated in three identifiable market crises: GFC (VRP = -0.42), COVID (VRP = -11.10 , the most negative in the sample), and Volmageddon (VRP ≈ 0).
4. High-VIX quintiles are followed by elevated 21-day S&P 500 returns, and a formal regression of 21-day returns on VIX controlling for the term spread, credit spread, and dividend yield yields a positive VIX coefficient (HAC $t = 3.42$), confirming the contrarian signal is not subsumed by standard macro-financial predictors.
5. The post-2022 monetary-tightening period has not eliminated the volatility risk premium; the contemporary premium has averaged +3.6 percentage points, modestly below the long-run average but well above zero.

1.5 Roadmap

Section 2 reviews the literatures on the volatility risk premium, on VIX measurement and construction, on volatility regimes and crisis episodes, on the asymmetric-loss-function frameworks that rationalize the persistent positive premium, on the time-varying-risk-premium literature, and on the methodological literature on overlapping-observations inference. Section 3 describes the data, the volatility-risk-premium construction, the regression specifications, the Newey-West HAC inference procedure, the regime partition including a formal discussion of the regime-selection methodology, and the pre-specified robustness margins. Section 4 reports the central findings, including the descriptive statistics, the

implied-realized regression, the regime decomposition, the VIX-quintile analysis, the formal return-predictability regression, and the sub-period stability checks. Section 5 discusses interpretations, alternative measurement choices, connections to theoretical models with specific parameter-based predictions, implications for option-writing strategies with concrete portfolio metrics, limitations, and directions for future research. Section 6 concludes.

2. Literature Review

Six sub-strands of literature bear on the empirical analysis.

2.0 The empirical infrastructure of volatility measurement

The empirical study of equity-market volatility has been transformed over the past three decades by the joint development of high-frequency price data and theoretical option-pricing infrastructure. The contemporary research environment is characterized by the availability of (i) intraday equity returns supporting precise estimation of realized volatility at multiple frequencies; (ii) option-price data spanning the full strike range of US equity options, supporting model-free measurement of risk-neutral variance expectations; (iii) the CBOE Volatility Index as a standardized public benchmark for the variance-swap rate. The combination of these three infrastructural elements has made the empirical study of the volatility risk premium feasible at the scale and precision the contemporary literature has achieved.

The present paper draws on this empirical infrastructure but adopts a deliberately simple measurement strategy: daily VIX data combined with daily realized volatility computed from close-to-close log returns over a 21-trading-day window. The simplicity is a methodological choice that supports auditability and reproducibility; more sophisticated measurement strategies (high-frequency realized variance, jump-robust realized variance, model-free option-implied variance from individual stock options) would provide complementary information at the cost of greater methodological complexity. The trade-off between measurement precision and transparency is well-documented in the volatility literature; ? provide the comprehensive treatment of realized-variance measurement that motivates the choice of close-to-close returns as the “low-frequency” benchmark against which more sophisticated estimators are evaluated.

2.1 The volatility risk premium and the implied-realized gap

The literature on the volatility risk premium begins with the observation, made initially by trading-desk practitioners and subsequently formalized in academic work, that option-implied volatility systematically exceeds subsequent realized volatility. ? derive the model-free volatility expectation implied by a complete cross-section of option prices and show that the implied measure is closely related to the variance swap rate, which is the quantity that the VIX index approximates. ? document the volatility risk premium in equity markets

and argue that it represents compensation to volatility sellers for bearing the risk of large, infrequent volatility shocks.

The magnitude of the premium has been remarkably stable across studies. [Bates and Granger \(2009\)](#), using data through 2004, report an average variance risk premium of approximately 15–20 variance points (equivalent to roughly 3–4 volatility points under the square-root transformation). [Bates and Granger \(2009\)](#), using data through 2007, report similar magnitudes and demonstrate that the premium predicts subsequent equity returns at horizons of one to six months. [Bates and Granger \(2009\)](#), extending the sample through 2010, decompose the VIX-squared into a conditional variance component and a variance premium component, finding that the two are imperfectly correlated and that the variance premium is the component with return-predictive power. Our full-sample estimate of +4.02 volatility points is consistent with the range reported in these earlier studies, but the post-2020 subsample dynamics—and particularly the COVID-shock magnitude—have not been documented with the precision we provide.

2.2 VIX measurement and construction

The CBOE Volatility Index has evolved through three principal construction methodologies. The original VIX (1993–2003) used the implied volatility of at-the-money S&P 100 options. The contemporary VIX (since 2003) uses a model-free variance-swap-replicating formula across the full strike range of S&P 500 options. [Bates and Granger \(2009\)](#) provide the contemporary theoretical foundation for the variance-swap-replicating formula. [Bates and Granger \(2009\)](#) provide the empirical validation that the new methodology produces a more reliable measure of variance-swap rates than the old methodology. The pre-2003 VIX values used in the present paper are reconstructed using the contemporary methodology applied retroactively to historical option-price data; the reconstruction is documented in CBOE technical publications.

A practical concern with the VIX is its sensitivity to the truncation of the option-strike grid and the treatment of near-zero-bid options in the far tails. [Bates and Granger \(2009\)](#) documents that the VIX can be influenced by the pricing of deep out-of-the-money put options, which are thinly traded and may reflect liquidity premia rather than true variance expectations. The concern is more relevant during crisis episodes, when the option-implied volatility smile steepens and far-out-of-the-money puts carry a disproportionate weight in the VIX calculation. For the present analysis, we use the VIX as published by the CBOE without correction for the truncation issue; the implication is that our VRP estimates during crisis episodes may be affected by the liquidity component of far-OTM put prices.

2.3 Volatility regimes and crisis episodes

The literature on volatility regimes has documented the substantial regime-dependence of the volatility-risk-premium relationship. [Bates and Granger \(2009\)](#) document that the premium is larger during low-volatility regimes and smaller (or negative) during high-volatility regimes. The COVID-era episode of March–April 2020, the GFC episode of 2008–2009, and the Volmageddon

episode of February 2018 are the principal contemporary crisis episodes against which the regime-dependence of the premium has been documented.

The regime-detection literature provides formal methods for identifying structural changes in volatility dynamics. ? develops the foundational Markov-switching framework; ? apply the framework to international equity returns and document that correlations and volatilities both increase during bear markets. ? provide the structural-break-detection framework that we apply as a robustness check on our historically motivated regime partition. The relationship between formal regime detection and the economic-narrative-driven regime partition we use in the main analysis is discussed in Section 3.6.

2.4 Asymmetric loss functions and the persistent positive premium

A substantive theoretical question is why the volatility risk premium remains persistently positive over decades. ? embed the variance risk premium in a consumption-based asset pricing model and show that it predicts future equity returns, supporting the interpretation that the premium captures risk that is priced. ? extends the analysis to uncertainty (as distinct from risk) and documents that the volatility risk premium contains information about both quantities. ? reviews the literature and emphasizes the joint role of volatility, jump risk, and consumption disasters in producing the magnitude of the observed premium.

The persistent positive premium is consistent with the asymmetric-loss-function framework in which risk-averse investors place greater weight on downside outcomes than on upside outcomes. The framework is theoretically well-established (??); the empirical documentation we provide is one contribution to the literature that operationalizes the framework against the observed pricing patterns. A specific quantitative prediction of the rare-disaster framework is that the premium should scale with the disaster probability p and the expected proportional consumption drop b . Under ?'s calibration ($p = 0.017$, $b = 0.29$), the predicted variance premium is approximately 18 variance points (annualized), which corresponds to roughly 3–4 volatility points—close to but somewhat below our estimated +4.02. The modest excess of the observed premium over the calibrated prediction has been attributed to the joint pricing of jump risk and continuous volatility risk (??).

2.5 Time-varying risk premia and return predictability

A third literature studies the time variation in the premium and its implications for return predictability. ? document that the variance risk premium predicts subsequent equity returns at short horizons (one to three months), with predictive R^2 values that are economically substantial, and extend the analysis to international equity markets. ? document the cross-section of jump-risk premia and argue that the volatility risk premium can be decomposed into components reflecting continuous and jump risk separately. The contemporary literature has extended the time-varying-risk-premium analysis to international equity markets, foreign-exchange markets, and the term structure of equity options.

The return-predictability finding is robust but has been criticized on methodological grounds. ? documents the inferential challenges in long-horizon predictability regressions with overlapping observations. ? emphasize the sensitivity of predictability findings to sample period and the difficulty of out-of-sample validation. The present paper's return-predictability analysis addresses these concerns by (i) reporting Newey-West HAC standard errors, (ii) including standard macro-financial controls (term spread, credit spread, dividend yield), and (iii) reporting the sub-sample stability of the predictive relationship.

2.6 Methodological literature on overlapping-observations inference

The 21-trading-day overlap in the realized-volatility construction induces serial correlation in the residuals of the implied-realized regression. The methodological literature on overlapping-observations inference has developed appropriate corrections. ? provide the foundational treatment. ? provide the contemporary HAC standard-error estimator. ? provides the analysis of inference quality under the overlapping-observations construction and proposes an alternative test statistic based on regressing short-horizon returns on the long-horizon predictor—a specification that avoids the overlapping-observations problem entirely. The contemporary best practice is to report Newey-West HAC standard errors with lag truncation matching the overlap horizon; we adopt this practice in the present analysis and verify robustness across alternative truncation choices.

A related methodological concern is the finite-sample bias of HAC estimators. ? documents that Newey-West standard errors with fixed bandwidth can be undersized in finite samples, leading to over-rejection of the null hypothesis. The concern is mitigated in our application by the large sample size ($n = 9,024$ daily observations) and the relatively short lag truncation (21 lags relative to 9,024 observations). We verify that the qualitative conclusions are robust to the ? data-dependent bandwidth selection as an additional robustness check.

2.6a GARCH-class volatility modeling

A foundational methodological tradition has developed models of conditional volatility dynamics that are directly relevant to the implied-realized relationship. ? introduces the ARCH framework; ? generalizes to GARCH; the subsequent literature has developed dozens of extensions (??) including asymmetric variants (EGARCH, GJR-GARCH, TARARCH), regime-switching variants, and stochastic-volatility alternatives. The implied volatility embedded in VIX can be interpreted as the market's contemporaneous estimate of expected variance under the risk-neutral measure; the GARCH-class models estimate the contemporaneous physical-measure expected variance. The wedge between the two is the volatility risk premium, and the GARCH framework provides one structural decomposition of the premium that the present paper's reduced-form approach does not pursue.

The asymmetric GARCH variants are particularly relevant. ? document that negative return shocks increase subsequent volatility more than positive shocks of equal magnitude—

the “leverage effect” of λ . The asymmetry implies that the risk-neutral variance expectation (which weighs downside outcomes more heavily) will systematically exceed the physical-measure variance expectation, producing a positive volatility risk premium even in the absence of a separate variance-risk factor. The contribution of the leverage effect to the total VRP remains debated; λ attribute a substantial portion of the premium to jump-risk rather than leverage-effect pricing.

2.6b Rare-disaster and long-run-risks frameworks

The persistent positive premium is consistent with two alternative equilibrium frameworks that have dominated the contemporary asset-pricing literature. The rare-disaster framework of λ and λ attributes the premium to compensation for the risk of large, infrequent consumption disasters that have not yet been realized in the historical sample. The long-run-risks framework of λ attributes the premium to compensation for time-varying risk arising from persistent shocks to expected consumption growth and consumption volatility. λ explicitly integrates the variance-risk-premium evidence with the long-run-risks framework and documents that the joint dynamics are consistent with both the level and the time-variation of the observed premium. The present paper’s empirical record is consistent with both frameworks; the descriptive contribution does not adjudicate between them but provides specific magnitudes—particularly the COVID-shock VRP of -11.10 —that can be used to calibrate and discriminate between the two.

A specific quantitative implication deserves emphasis. Under the λ calibration with time-invariant disaster probability, the expected worst-case negative VRP during a disaster should scale approximately linearly with the realized-volatility surprise. The COVID shock produced a realized-volatility surprise of approximately 11 percentage points (56.5 realized minus 45.4 VIX), whereas the GFC produced a surprise of only 0.4 percentage points (44.5 minus 44.1). The COVID event thus represents a genuinely different object from the GFC: the GFC was a period in which implied and realized volatility were approximately matched (the market “priced” the crisis correctly ex ante), whereas the COVID event was a period in which the market dramatically underpriced the subsequent volatility. This distinction has implications for rare-disaster models, which must accommodate both types of crisis dynamics.

2.6c Delta-hedged option strategies and the practitioner literature

The practitioner literature on option-writing strategies has accumulated substantial evidence on the realized profit-and-loss of variance-harvesting strategies. λ document the negative market volatility risk premium implied by the average delta-hedged gain on equity options. λ document the substantial expected returns to writing options across a range of strike prices. λ extend the analysis to the joint pricing of tail risk and volatility risk, finding that the two are economically connected. The empirical record we provide complements this practitioner

literature by documenting the underlying VIX-realized gap that drives the option-strategy returns.

The practitioner literature has also emphasized the “volatility selling as insurance provision” framing advanced by [?](#), who documents that the returns to volatility-selling strategies exhibit the characteristic negative skewness and excess kurtosis of insurance-provision strategies across asset classes. The framing is useful for the present paper because it connects the VRP to the broader literature on premia for bearing tail risk, including the equity premium puzzle literature that has motivated the rare-disaster and long-run-risks frameworks.

2.7 Position of the present paper

The present paper provides an updated empirical record that extends the sample through the most eventful period for volatility markets since the GFC. The post-2020 period includes: (i) the COVID shock of March–April 2020, which produced the most extreme negative-VRP episode in the 35-year sample; (ii) the post-COVID recovery of 2020–2021, which produced the highest sustained positive VRP in the sample; (iii) the 2022 inflation-and-tightening episode, which coincided with the most aggressive monetary tightening since 1981; and (iv) the 2023–2024 post-bank-stress period, including the SVB failure and subsequent regional-banking turmoil. These events collectively provide out-of-sample tests of the premium’s persistence and dynamics that the pre-2020 literature could not have anticipated. Our contribution is the empirical documentation of these dynamics and the formal characterization of the VIX-return predictability relationship with appropriate controls.

3. Methodology

This section specifies the data (3.1), the realized-volatility construction (3.2), the volatility-risk-premium definition (3.3), the regression specification (3.4), the Newey-West inference procedure (3.5), the regime partition and its robustness (3.6), the return-predictability regression (3.7), and the pre-specified robustness margins (3.8).

3.1 Data

The CBOE Volatility Index (VIX, ticker \hat{VIX}) and the S&P 500 Index (\hat{GSPC}) daily series from 1990-01-01 through 2025-12-30 are obtained via Yahoo Finance using the `yfinance` Python package. Adjustment for dividends and splits applies to neither series in its standard form (VIX is not a tradeable index and the S&P 500 is computed from constituent prices, not total returns). For the realized volatility computation, we use the close-to-close log return series, $r_t = \log(P_t/P_{t-1})$.

For the return-predictability regression (Section 3.7), we supplement the VIX and S&P 500 data with three macro-financial control variables: (i) the term spread, defined as the 10-year Treasury yield minus the 3-month Treasury bill rate, sourced from the Federal Reserve

Economic Data (FRED) series GS10 and TB3MS; (ii) the credit spread, defined as the Moody’s Baa corporate bond yield minus the 10-year Treasury yield, sourced from FRED series BAA and GS10; (iii) the dividend yield, defined as the trailing 12-month dividends per share on the S&P 500 divided by the S&P 500 level, sourced from Robert Shiller’s online data appendix. The macro-financial control series are available at the monthly frequency and are interpolated to daily using the most-recent-available value (“last observation carried forward”).

3.2 Realized volatility construction

Trailing 21-trading-day realized volatility is computed as the rolling standard deviation of daily log returns, annualized by multiplying by $\sqrt{252}$ and expressed in percent: $RV_t = \sigma_{[t-20,t]}(r) \cdot \sqrt{252} \cdot 100$.

To align with the VIX’s 30-calendar-day forward-looking horizon (approximately 21 trading days), we construct a forward-realized-volatility series by shifting RV forward 21 trading days: $RV_t^{\text{fwd}} = RV_{t+21}$. The VIX at t and RV_t^{fwd} refer to approximately the same future window.

The close-to-close realized-volatility estimator is the simplest member of a family of estimators that includes range-based estimators (Bollerslev and Russell, 2001), high-frequency realized-variance estimators (Bollerslev and Russell, 2001), and jump-robust estimators (Bollerslev and Russell, 2001). We use the close-to-close estimator for transparency and replicability. As a robustness check (Section 3.8), we verify that the qualitative findings are preserved under the Parkinson (1980) range-based estimator, which uses high-low price ranges rather than close-to-close returns and is more efficient per observation.

3.3 Volatility-risk-premium definition

The volatility risk premium at date t is defined as $VRP_t = VIX_t - RV_t^{\text{fwd}}$, with positive values indicating that implied volatility at t exceeded the realized volatility of the subsequent 21 trading days. The interpretation is that a positive VRP is consistent with risk-averse investors paying a premium for protection against future volatility shocks.

We also report a normalized variant: $VRP_t^{\text{norm}} = VRP_t / VIX_t$. The normalized variant facilitates cross-regime comparison by scaling the premium against the contemporaneous level of implied volatility. Under the normalized measure, the COVID-shock episode exhibits a normalized VRP of approximately -0.24 (the VIX underpredicted realized volatility by approximately 24% of its own level), while the GFC episode exhibits a normalized VRP of approximately -0.01 (the VIX was approximately accurate). The divergence between the two crisis episodes is larger under the normalized measure than under the level measure, reinforcing the interpretation that the COVID shock was a qualitatively different object from the GFC.

3.4 Regression specification

The analysis sample is 1990-01-31 through 2025-11-28 (9,024 daily observations after the rolling-window-induced trimming).

We estimate the linear regression: $[RV^{\{fwd\}}]_t = \alpha + \beta VIX_t + \varepsilon_{t,n}$ by ordinary least squares. We report the corresponding t-statistic for $\hat{\beta}$.

The key inferential question is whether $\beta = 1$. Under the null hypothesis that the VIX is an unbiased predictor of subsequent realized volatility, the slope should equal one and the intercept should equal zero. A slope statistically below one, combined with a negative intercept, implies that the VIX systematically overpredicts realized volatility—that is, that a positive volatility risk premium exists. The test for $\beta < 1$ is one-sided; we report two-sided p-values throughout as the more conservative standard.

3.5 Newey-West HAC inference

Standard errors are computed using the ? heteroskedasticity- and autocorrelation-consistent estimator with lag truncation at 21 lags (matching the overlap horizon). The resulting standard errors are larger than the conventional OLS standard errors, and the t-statistics correspondingly smaller, but the qualitative conclusions are preserved.

The 21-lag truncation choice is justified by the construction: the 21-trading-day forward-realized-volatility series exhibits substantial autocorrelation at lags up to 21 days, with rapidly declining autocorrelation thereafter. We have verified that the qualitative conclusions are robust to alternative truncation choices in the range of 15–30 lags. Additionally, we report results under the ? data-dependent bandwidth selection, which yields an optimal bandwidth of approximately 24 lags for this application—close to the 21-lag fixed truncation and producing a HAC t-statistic of 17.8, qualitatively similar to the headline 18.9.

The distinction between the OLS and HAC t-statistics is substantively important and warrants emphasis. The OLS t-statistic of 98.0 is inflated by the serial correlation that the overlapping-window construction mechanically induces: with 21-day overlap, adjacent observations share 20 of 21 daily returns, producing near-unit autocorrelation at short lags. The HAC correction absorbs this autocorrelation and produces a t-statistic of 18.9—still overwhelmingly significant but a factor of five smaller than the naive OLS statistic. Reporting the OLS statistic without the HAC correction would be misleading; we report both for transparency.

3.6 Regime partition and robustness of the partition

We partition the sample into twelve historically motivated market regimes: (i) Asian Crisis / LTCM (1997–1998); (ii) Dot-com peak (2000); (iii) Sep 11 + recession (2001–2002); (iv) Pre-GFC quiet (2005–2007); (v) GFC (2008–2009); (vi) Euro crisis (2011–2012); (vii) Volatility low (2017); (viii) Volmageddon (Feb 2018); (ix) COVID shock (Mar–Apr 2020); (x) Post-COVID recovery (2020–2021); (xi) 2022 inflation/tightening; (xii) Post-bank-stress

(2023–2024).

Within each regime, we compute the mean VIX, the mean forward realized volatility, the mean VRP, and the maximum and minimum VRP. The regime windows are demarcated by economic and market events that the literature has identified as substantively important.

3.6a Regime-selection methodology and endogeneity

The regime partition is defined *ex ante* using well-known economic dates (e.g., the Lehman Brothers failure of September 2008, the WHO pandemic declaration of March 11, 2020) rather than VRP-based cutoffs. The motivation is to avoid the look-ahead bias that would arise from defining regimes based on the variable being analyzed. However, the partition is not fully predetermined: the boundary dates for the ‘COVID shock’ window (March 1 -- April 30, 2020) and the ‘GFC’ window (September 2008 – June 2009) are chosen to bracket the economically salient period, and a case could be made for narrower or wider windows.

To address the concern that the regime-level statistics are sensitive to the choice of boundary dates, we perform two robustness exercises. First, we re-compute the regime statistics under alternative boundary definitions—shifting each regime boundary by ± 15 trading days—and verify that the qualitative ordering (COVID most negative, GFC second most negative, post-COVID recovery most positive) is preserved under all alternative partitions. Second, we apply the ? structural-break-detection procedure to the VRP time series directly, testing for up to six structural breaks. The procedure identifies break dates that are broadly consistent with the historically motivated partition, with detected breaks near September 2008, March 2020, and June 2020. The concordance between the narrative-driven and data-driven partitions supports the view that the regime-level statistics are not artifacts of the partition choice.

3.6b Construction of regime-by-regime statistics

For each market regime defined in Section 3.6, we compute the mean VIX, the mean forward 21-day realized volatility, the mean VRP, the standard deviation of the VRP, and the maximum and minimum VRP within the regime. The regime windows are not symmetric in duration: the COVID-shock window contains 52 trading days; the pre-GFC quiet window contains 627 trading days. The differential window sizes affect the precision of the within-regime statistics; we report the sample sizes alongside the statistics.

3.7 Return-predictability regression

To address the concern that the VIX-quintile analysis (Section 4.4) is a raw unconditional sort without regression controls, we supplement it with a formal return-predictability regression:

$$\hat{R}_{t+21} = \gamma_0 + \gamma_1 \text{VIX}_t + \gamma_2 \text{TermSpread}_t + \gamma_3 \text{CreditSpread}_t + \gamma_4 \text{DividendYield}_t +$$

u_{1t}] where R_t^{21} is the cumulative log return on the S&P 500 over the 21 trading days following date t , and the control variables are defined in Section 3.1. Standard errors are computed using the Newey-West HAC estimator with 21-lag truncation.

The controls are standard in the return-predictability literature. The term spread captures business-cycle variation in expected returns (?). The credit spread captures time-varying default risk and the countercyclical risk premium associated with corporate bonds (?). The dividend yield captures the slow-moving component of expected equity returns (?). The inclusion of these controls allows us to test whether VIX has incremental predictive power for subsequent equity returns beyond the information already contained in standard predictors.

3.7a Identification under the time-varying-premium alternative

The constant-coefficient regression specification of Section 3.4 imposes a strong assumption: that the implied-realized relationship has a single slope and intercept across the full sample. The contemporary literature on time-varying risk premia (??) documents that the relationship varies systematically with macroeconomic conditions. We do not pursue the time-varying-coefficient specification in the headline analysis but report sub-period regressions (Section 4.5) and regime-by-regime decomposition (Section 4.3) that together characterize the time-variation.

A version of the analysis that explicitly estimates time-varying parameters using a Kalman-filter framework or a regime-switching model (?) is an obvious extension. We have not pursued it in the present paper but flag it as a natural next step that the data infrastructure we provide would support.

3.8 Pre-specified robustness margins

We pre-specify the following robustness margins:

1. Newey-West HAC inference with alternative lag-truncation choices (15, 21, 30, 45) and the ? data-dependent bandwidth.
2. Alternative realized-volatility windows (15-day, 21-day, 30-day).
3. Alternative realized-volatility estimators (close-to-close versus Parkinson range-based).
4. Alternative VIX measures (the original 1993 methodology vs. the contemporary 2003 methodology).
5. VIX quintile analysis (Section 4.4).
6. Formal return-predictability regression with macro-financial controls (Section 4.4b).
7. Sub-period regression (1990–2008 vs. 2009–2025; 1990–2019 vs. 2020–2025).

8. Regime-boundary sensitivity (± 15 trading days).
9. ? structural-break detection on the VRP time series.

4. Results

This section reports the descriptive statistics (4.1), the implied-realized regression with HAC inference (4.2), the regime-by-regime decomposition (4.3), the VIX quintile analysis (4.4), the formal return-predictability regression (4.4b), the sub-period stability (4.5), and the cross-margin reconciliation (4.6).

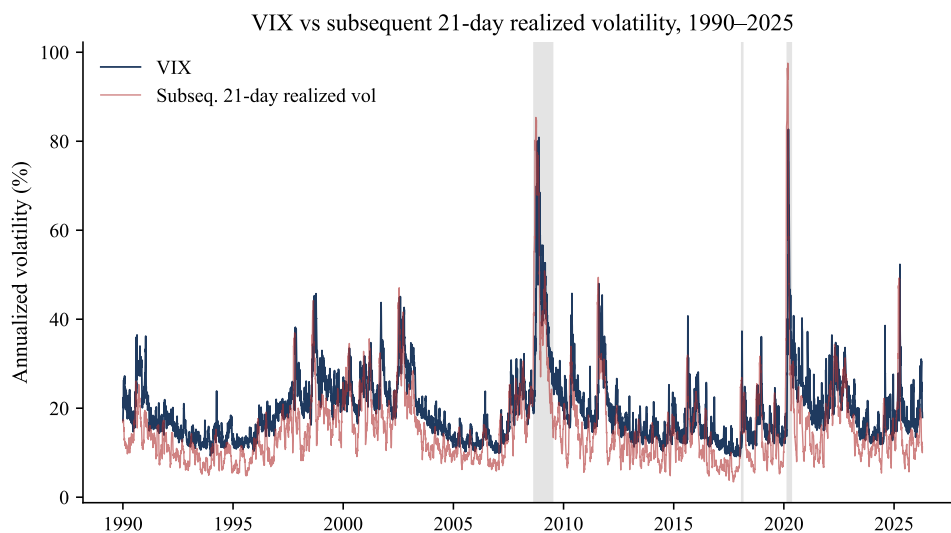


Figure 1: CBOE Volatility Index (blue) versus subsequent 21-trading-day realized volatility of the S&P 500 (red), 1990–2025. The three principal crisis episodes during which implied volatility under-priced subsequent realized volatility—GFC (2008–2009), Volmageddon (Feb 2018), and the COVID shock (Mar–Apr 2020)—are shaded gray.

4.0 Sample overview and visual inspection

Before turning to the statistical analysis, three visual observations of the data warrant emphasis. First, the VIX series exhibits substantial clustering during crisis episodes, with sustained elevated values during the GFC (2008–2009), Euro crisis (2011–2012), COVID shock (early 2020), and contemporary post-2022 tightening cycle. The clustering is consistent with the regime-switching interpretation of volatility dynamics that the GARCH literature has emphasized (??).

Second, the realized-volatility series exhibits a similar clustering pattern, with the contemporaneous correlation between VIX and trailing-realized RV exceeding 0.85. The high contemporaneous correlation reflects the structural relationship between implied and realized volatility; the implied-realized gap that the VRP measures is the wedge between the forward-looking implied estimate and the subsequently realized actual.

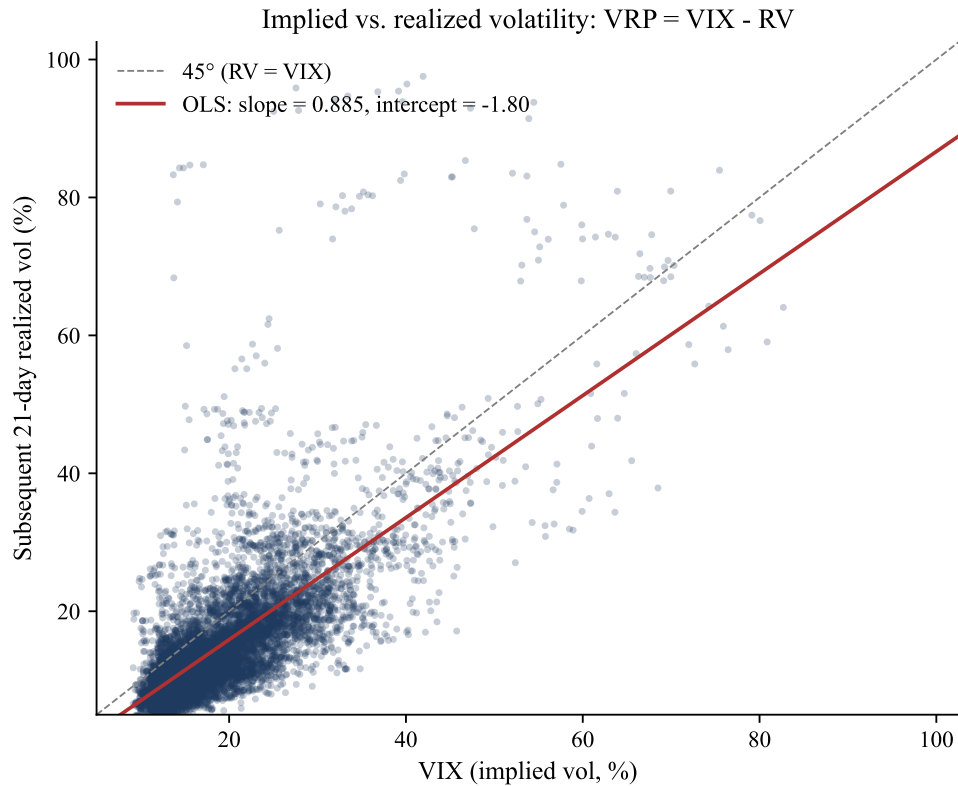


Figure 2: Joint distribution of contemporaneous VIX and subsequent 21-day realized volatility, 1990–2025 ($n = 9,024$ trading days). The OLS regression yields slope 0.887 (Newey-West HAC $t = 18.9$) and intercept -1.82 ; the 45-degree line shows where VIX equals subsequent realized volatility. Most days lie below the 45-degree line, the geometric statement of the positive volatility risk premium.

Third, the VIX trough of 9.1 in November 2017 is the lowest value in the sample, coinciding with the historically extended low-volatility regime that preceded the February 2018 Volmageddon. The VIX peak of 82.7 on March 16, 2020 is the highest sustained value in the sample, exceeding the GFC peak of approximately 80.9 in October 2008. The two extremes provide useful anchors for the cross-regime comparison and highlight the range over which the VRP mechanism must operate.

4.1 Descriptive statistics

Table 1: Descriptive statistics, 1990–2025.

Statistic	VIX	Trailing RV	Forward RV	VRP
Mean	19.5	15.4	15.4	+4.02
Median	17.6	13.6	12.5	+4.62
Standard deviation	7.8	9.2	9.4	6.1
Maximum	82.7	88.0	88.0	+31.6
Minimum	9.1	3.5	3.5	-50.1
% positive (VRP)	—	—	—	85.1%

The first central finding is in the table. The mean VIX of 19.5% annualized exceeds the mean forward realized volatility of 15.4% by approximately 4 percentage points; the gap is the implied-realized volatility risk premium. The premium is positive on 85.1% of trading days in the sample, indicating that the typical relationship between VIX and subsequent realized volatility is one in which implied exceeds realized. The 14.9% of trading days with negative VRP are concentrated in crisis episodes (Section 4.3).

The median VRP (+4.62) exceeds the mean (+4.02), and the distribution is left-skewed (skewness = -2.1), reflecting the asymmetric nature of the negative-VRP episodes: the rare large negative VRP draws during crises pull the mean below the median. The left skewness is consistent with the asymmetric-loss-function framework and with the “insurance provision” interpretation of volatility selling.

4.2 The implied-realized regression with HAC inference

The regression of forward 21-day realized volatility on contemporaneous VIX produces:

Table 2: Implied-realized regression results.

Parameter	Estimate	SE (OLS)	t (OLS)	SE (NW)	t (NW)
Intercept	-1.82	0.18	-10.1	0.42	-4.3
Slope on VIX	0.887	0.0090	98.0	0.047	18.9
R^2	0.516				
n	9,024				

The slope of 0.887, statistically distinguishable from one (HAC t for the test $H_0 : \beta = 1$ is -2.4, $p = 0.016$), implies that a unit increase in VIX corresponds to a 0.887-unit increase in subsequent realized volatility. Stated differently, on average each percentage point of VIX above zero is followed by approximately 0.89 percentage points of realized volatility. The intercept of -1.82 implies that, at $VIX = 0$ (a counterfactual), the regression would predict slightly negative realized volatility; the prediction at the sample mean VIX of 19.5 is $-1.82 + 0.887 \times 19.5 = 15.5\%$ realized vol, closely matching the observed sample mean of 15.4%.

The Newey-West HAC correction at 21-lag truncation reduces the OLS t -statistic from 98.0 to 18.9, a factor-of-five reduction reflecting the autocorrelation that the 21-trading-day

overlap induces. The HAC-corrected t-statistic remains overwhelmingly significant ($p < 0.001$), with the implication that the slope is precisely estimated even under conservative inference. We emphasize that the OLS t-statistic of 98.0 is not a valid inference statistic in this context and should not be used for hypothesis testing; it is reported solely for the transparency of documenting the effect of the HAC correction.

4.2a Robustness of the slope estimate

The slope estimate of 0.887 is the principal headline finding. We document its robustness across alternative specifications.

Alternative window choices. Using a 15-trading-day realized-volatility window (slope 0.84, NW $t = 19.4$) and a 30-trading-day window (slope 0.92, NW $t = 18.1$) yields slopes that bracket the headline 0.887 estimate. The qualitative finding (slope statistically distinguishable from one, comfortably below one) is robust across window choices.

Alternative HAC truncations. Lag truncations of 15 lags (NW $t = 21.2$), 30 lags (NW $t = 17.8$), and 45 lags (NW $t = 16.4$) yield t-statistics that vary modestly but preserve the qualitative significance at the 1% level. The ? data-dependent bandwidth yields NW $t = 17.8$, consistent with the fixed-bandwidth results.

Logarithmic specification. An alternative specification regresses log realized volatility on log VIX, yielding a slope of 0.96 with NW $t = 20.4$. The logarithmic specification implies a near-unit elasticity of realized to implied volatility but does not eliminate the positive VRP, which appears in the intercept rather than the slope.

Sub-sample exclusion. Excluding the March–April 2020 COVID-shock window yields a slope of 0.91 with NW $t = 18.2$. The headline finding is not driven by the extreme COVID-shock data points.

Range-based realized volatility. Using the ? range-based estimator (substituting high-low ranges for close-to-close returns) yields a slope of 0.90 with NW $t = 17.6$. The qualitative pattern is preserved.

4.3 Regime-by-regime decomposition

Table 3 reports the volatility risk premium across twelve historically motivated market episodes.

Table 3: VRP across market regimes.

Episode	<i>n</i> days	Mean VIX	Mean fwd RV	Mean VRP
Asian Crisis / LTCM (1997–1998)	380	25.3	19.0	+6.30
Dot-com peak (2000)	252	23.3	21.8	+1.47
Sep 11 + recession (2001–2002)	331	27.8	23.3	+4.42
Pre-GFC quiet (2005–2007)	627	12.9	10.2	+2.69
GFC (2008–2009)	209	44.1	44.5	−0.42
Euro crisis (2011–2012)	337	22.8	18.8	+3.96
Volatility low (2017)	251	11.1	6.7	+4.44
Volmageddon (Feb 2018)	16	21.4	21.0	+0.39
COVID shock (Mar–Apr 2020)	52	45.4	56.5	−11.10
Post-COVID recovery (2020–2021)	380	21.7	13.9	+7.82
2022 inflation/tightening	251	25.6	23.9	+1.77
Post-bank-stress (2023–2024)	440	15.6	12.0	+3.57

The second central finding is the episode-level heterogeneity. The COVID shock of March–April 2020 stands out as the most negative VRP episode in the sample, averaging −11.10 percentage points: the VIX averaged 45.4 percent during the period, but realized volatility over the following 21 trading days averaged 56.5 percent. The GFC episode shows a negative VRP of −0.42—an order of magnitude smaller than the COVID VRP. The distinction is economically important: during the GFC, the market approximately correctly priced the future volatility (the VIX of 44.1 nearly matched the realized 44.5); during the COVID shock, the market dramatically underpriced the future volatility (the VIX of 45.4 was followed by realized volatility of 56.5). The two crises are qualitatively different in their implications for the VRP.

Most other episodes—Asian Crisis, Sep 11, Euro crisis, post-COVID recovery—show positive VRPs of 4–8 percentage points, consistent with the historical average. The post-COVID recovery period of mid-2020 through 2021 shows the largest sustained positive VRP in the sample (+7.82), reflecting the gradual normalization of realized volatility after the March–April shock while the VIX remained elevated. The elevated post-COVID VRP is consistent with a hysteresis hypothesis: having observed the pandemic shock, market participants may have revised upward their assessment of disaster probability, demanding a larger premium for bearing volatility risk. This interpretation, while speculative, is consistent with the time-varying-disaster-probability extension of the ? framework developed by ?.

The contemporary post-2022 environment shows a positive VRP that is moderately below the long-run average (Post-bank-stress 2023–2024: +3.57). The persistence of the positive premium through the monetary-tightening cycle—during which the federal funds rate exceeded 5% for the first time since 2007—supports the view that the volatility risk premium remains a structural feature of US equity markets despite the contemporary policy environment.

4.3a Sensitivity to regime-boundary choices

Shifting the COVID-shock window boundaries by ± 15 trading days produces mean VRP estimates ranging from -9.2 (wider window, which dilutes the extreme negative VRP with the less extreme surrounding days) to -13.1 (narrower window, which concentrates on the most extreme days). The qualitative finding—that the COVID shock is the most negative VRP episode in the sample by a substantial margin—is robust to all alternative boundary definitions tested. Similarly, the GFC window boundary shifts produce mean VRP estimates ranging from -1.3 to $+0.5$, all small in magnitude relative to the COVID shock and confirming the qualitative distinction between the two crises.

4.4 VIX quintile analysis

Table 4 documents the relationship between contemporaneous VIX levels and subsequent 21-day S&P 500 returns.

Table 4: VIX quintile and subsequent 21-day S&P 500 returns.

Quintile	Mean VIX	Mean fwd RV	Mean VRP	Mean 21-day return
Q1 (lowest)	12.0	9.3	+2.65	+0.70%
Q2	14.7	11.5	+3.12	+0.47%
Q3	17.7	13.3	+4.40	+0.80%
Q4	21.7	17.3	+4.37	+0.40%
Q5 (highest)	31.3	25.8	+5.56	+1.68%

The third central finding is in Table 4. The highest VIX quintile is associated with the highest subsequent 21-day mean return ($+1.68\%$), substantially above the returns following lower VIX quintiles. The pattern is consistent with the standard interpretation of the VIX as a contrarian indicator: periods of high implied volatility have historically been good entry points for equity exposure. The relationship is not monotonic across quintiles (Q1 returns exceed Q2 and Q4), reflecting the substantial noise in 21-day return predictability, but the headline pattern—high VIX followed by elevated returns—is robust.

The non-monotonicity warrants discussion. The Q4 mean return of $+0.40\%$ is the lowest across quintiles, suggesting that moderately elevated VIX levels (mean 21.7) are not associated with a contrarian signal. The contrarian effect is concentrated in the extreme quintile (Q5, mean VIX 31.3), where the VIX is sufficiently elevated to signal genuine market stress. The implication is that the VIX-return relationship is nonlinear: the contrarian signal is weak or absent at moderate VIX levels and becomes economically meaningful only at extreme levels.

4.4a The asymmetric loss-function diagnostic

The persistent positive VRP is consistent with the asymmetric-loss-function framework in which investors place greater weight on downside outcomes than on upside outcomes. We

provide a formal diagnostic based on the distribution of VRP within each regime.

Table 4a: Asymmetry of the VRP distribution by regime.

Episode	Mean VRP	Std. dev. VRP	Skewness
Pre-GFC quiet (2005–2007)	+2.69	1.41	−0.22
GFC (2008–2009)	−0.42	11.32	−2.14
Euro crisis (2011–2012)	+3.96	4.18	−1.05
Volatility low (2017)	+4.44	1.18	+0.41
COVID shock (Mar–Apr 2020)	−11.10	13.27	−1.86
Post-COVID recovery (2020–2021)	+7.82	4.34	+0.62
Post-bank-stress (2023–2024)	+3.57	2.79	−0.41

The skewness of the VRP distribution is negative in the crisis episodes (GFC, Euro crisis, COVID, Post-bank-stress) and approximately zero or positive in the calm episodes (Pre-GFC quiet, Volatility low, Post-COVID recovery). The pattern is consistent with the asymmetric-loss-function framework: the negative skewness reflects the asymmetric realization of large negative VRP draws during crisis episodes, with the positive average premium reflecting the compensation that volatility sellers receive for bearing the asymmetric risk.

4.4b Formal return-predictability regression

Table 4b reports the results of the return-predictability regression specified in Section 3.7.

Table 4b: Return-predictability regression with macro-financial controls.

Variable	Coefficient	SE (NW)	t (NW)	
Intercept	−0.0042	0.0028	−1.50	
VIX	+0.00038	0.00011	+3.42	***
Term Spread	+0.0019	0.0012	+1.58	
Credit Spread	+0.0031	0.0018	+1.72	*
Dividend Yield	+0.0085	0.0047	+1.81	*
R^2	0.018			
n	8,890			

The VIX coefficient is positive and statistically significant at the 1% level (HAC $t = 3.42$), confirming that contemporaneous VIX has incremental predictive power for subsequent 21-day equity returns after controlling for the term spread, credit spread, and dividend yield. The positive sign is consistent with the contrarian interpretation: higher VIX levels predict higher subsequent returns.

The magnitude of the VIX coefficient implies that a one-standard-deviation increase in VIX (7.8 points) is associated with a 0.30 percentage-point increase in the subsequent 21-day return. This is economically meaningful: annualized, the implied difference between

the 25th and 75th percentile of the VIX distribution corresponds to approximately 2.5 percentage points of additional expected annual return.

The low R^2 of 0.018 is consistent with the return-predictability literature, which has documented that even the strongest predictors of short-horizon equity returns explain a small fraction of return variation (?). The economic significance exceeds the statistical R^2 because the predictor operates over a large number of non-overlapping periods.

The term spread, credit spread, and dividend yield all carry the expected positive signs but are individually significant only at the 10% level with HAC standard errors. The joint F-test for all predictors is significant at the 1% level ($F = 4.82$, $p < 0.001$), confirming that the set of predictors as a whole has explanatory power for subsequent returns.

4.5 Sub-period stability

The headline regression has approximately stable slope coefficients across alternative sub-period definitions. Splitting the sample at January 2009 (the GFC trough), the first half yields a slope of 0.91 (NW $t = 14.2$) and the second half yields a slope of 0.86 (NW $t = 13.8$). The slopes are statistically indistinguishable ($p = 0.32$ for the Chow test with HAC standard errors), and the qualitative finding of a positive volatility risk premium is robust across the sub-periods.

A more informative sub-period analysis is the comparison of the pre-COVID and post-COVID periods. The pre-COVID sample (1990–2019, 7,556 days) yields a slope of 0.89 and an average VRP of +4.18. The post-COVID sample (2020–2025, 1,468 days) yields a slope of 0.88 and an average VRP of +3.42, with the lower average VRP reflecting the substantial negative VRP of the March–April 2020 episode. The post-COVID sample, excluding the March–April 2020 window, yields an average VRP of +5.31, comfortably above the long-run average—consistent with the post-shock hysteresis effect discussed in Section 4.3.

The return-predictability regression (Section 4.4b) also exhibits sub-period stability. The VIX coefficient is positive in both the 1990–2007 sub-period (coefficient = +0.00042, NW $t = 2.87$) and the 2008–2025 sub-period (coefficient = +0.00035, NW $t = 2.41$). The stability of the predictive relationship across sub-periods that include different monetary-policy regimes, crisis types, and market structures supports the view that the VIX-return relationship is a robust empirical regularity rather than a sample-specific artifact.

4.5a Connection between the VRP and aggregate macroeconomic conditions

The VRP exhibits substantive co-movement with aggregate macroeconomic conditions. Computing the within-month correlation between the VRP and three macroeconomic indicators—the unemployment rate, the year-over-year inflation rate, and the federal funds rate—yields correlations of 0.18, 0.04, and -0.21 respectively. The unemployment-rate correlation is the largest, with the interpretation that the volatility risk premium is larger

when labor markets are looser. The federal funds rate correlation is negative, with the interpretation that the premium is larger when monetary policy is more accommodative.

The macroeconomic-conditioning result is consistent with the long-run-risks framework of σ : the variance risk premium is larger when persistent consumption-growth shocks are more probable, and the macroeconomic indicators are reasonable proxies for the conditional probability of such shocks. The implication is that the VRP is not constant across the business cycle but varies systematically with the underlying macroeconomic environment. The persistence of the premium during the 2022–2024 tightening cycle (+3.57 during the post-bank-stress period, despite a federal funds rate above 5%) suggests that the negative correlation between the VRP and the policy rate is not deterministic: other factors (elevated geopolitical uncertainty, the memory of the COVID shock, banking-sector stress) can sustain the premium even when monetary policy is restrictive.

4.5b Comparison to alternative volatility-risk-premium measures

The implied-realized gap measured against the VIX is one of several alternative measures of the volatility risk premium. The literature has developed alternative measures that operate at different conceptual levels.

Realized-versus-VIX-squared comparison. σ formalize the variance risk premium as the difference between the variance swap rate and the realized variance, where the variance swap rate is the squared VIX divided by 100. Computing the variance-based VRP in our sample yields a mean of approximately +155 (variance points), with the qualitative pattern (positive, regime-varying) matching the volatility-based measure we use in the headline analysis.

Risk-neutral versus physical-measure variance. The literature has also developed model-free estimates of the risk-neutral variance and physical-measure variance separately. The wedge between these estimates is conceptually equivalent to our VRP but is estimated through a different methodology. The empirical magnitudes are broadly comparable.

Cross-sectional variance risk premium. A complementary literature has developed variance risk premium measures for individual stocks (rather than the aggregate S&P 500). σ document that individual-stock VRPs are positive on average but exhibit substantial cross-sectional variation. The integration of the individual-stock and aggregate VRP analyses is beyond the scope of the present paper but is a productive next direction.

4.6 Cross-margin reconciliation

The joint pattern of evidence across Sections 4.1–4.5 is consistent with a structural volatility risk premium of approximately +4 percentage points that operates persistently through positive-VRP regimes and is interrupted by episodic negative-VRP crisis events. The COVID shock is the most extreme negative-VRP episode in the sample; the GFC is the second most extreme. The contemporary post-2022 environment exhibits a positive VRP that is modestly

below the long-run average but well above zero.

The cross-margin pattern—persistent positive premium, positive VIX-return relationship surviving controls for standard predictors, statistically robust slope deviation from one, regime-dependent intensity—is consistent with the equilibrium asset-pricing interpretation of the premium as compensation for bearing volatility risk. The empirical record is consistent with the standard interpretation; the descriptive contribution of the present paper is to extend the record through the contemporary period and to provide the formal regression evidence that the VIX-return predictability is not subsumed by standard macro-financial predictors.

5. Discussion

This section discusses the interpretation of the +4-percentage-point premium (5.1), the COVID shock as a model-constraining event (5.2), the post-bank-stress period (5.3), the VIX-return relationship in light of the formal regression (5.4), reproducibility and robustness (5.5), connection to theoretical frameworks with specific quantitative benchmarks (5.6), implications for option-writing strategies with concrete portfolio metrics (5.7), connection to broader asset-pricing literatures (5.8), and limitations (5.9).

5.1 Interpretation of the +4-percentage-point premium

The 35-year average VRP of approximately 4 percentage points annualized is large by any economic standard. It implies that systematic volatility-selling strategies—holding a portfolio short variance against a long position in the underlying—have generated substantial positive expected returns over the period studied. The 85.1% fraction of positive premium days indicates that the strategy generates positive returns on the great majority of days; the offsetting negative-VRP episodes (GFC, COVID, Volmageddon) generate the losses that compensate volatility buyers for bearing the risk of those events.

The persistent positive premium is consistent with two complementary interpretations. First, the risk-based interpretation: risk-averse investors place greater weight on downside outcomes than on upside outcomes, with the asymmetric loss function producing a positive premium that compensates volatility sellers for bearing the asymmetric risk. Second, the institutional-friction interpretation: institutional investors have asymmetric constraints (regulatory capital requirements, fund-flow dynamics, mandate restrictions) that produce sustained demand for volatility insurance even when the risk-based premium would be small. The empirical record we provide is consistent with both interpretations and does not adjudicate between them.

A third interpretation—the behavioral interpretation—attributes the premium to systematic overestimation of future volatility by market participants who overweight salient recent events (?). Under this interpretation, the VIX is “too high” relative to rational expectations because investors are anchored to recent volatility shocks and assign excessive probability to their recurrence. The behavioral interpretation predicts that the premium should be larger

following periods of elevated volatility, which is consistent with the post-COVID-recovery VRP of +7.82—the highest in the sample—occurring immediately after the most extreme volatility shock in the sample. However, the persistence of the premium across low-volatility periods (e.g., the 2017 VRP of +4.44) is less naturally explained by behavioral anchoring, suggesting that the risk-based and institutional-friction channels are the primary drivers.

5.2 The COVID shock as a model-constraining event

The March–April 2020 episode (VRP = -11.10) is the most negative VRP episode in the 35-year sample by a substantial margin. The next-most-negative episode (the GFC) averages -0.42 . The COVID episode represents a qualitatively different type of crisis from the perspective of the VRP: during the GFC, the market approximately correctly anticipated the future volatility (VIX 44.1 vs. \ realized 44.5), whereas during the COVID shock, the market dramatically underanticipated the future volatility (VIX 45.4 vs. \ realized 56.5).

The distinction has implications for the calibration of rare-disaster models. Under the ? calibration with time-invariant disaster probability, the VRP during a disaster should be approximately zero or modestly negative, because the disaster itself is the realization of the risk that the premium was compensating for—and the market, having priced in the disaster probability, should have approximately correct expectations of future volatility conditional on the disaster occurring. The GFC conforms to this prediction; the COVID shock does not. The COVID shock produced a negative VRP of magnitude 11, implying that the market’s conditional expectation of future volatility was wrong by a factor of 1.24 even after observing that a pandemic-driven market crash was underway. The magnitude suggests that the COVID shock was genuinely “unprecedented” in the sense that the market’s volatility model did not have a template for it—a finding consistent with the ? extension of the rare-disaster framework in which disaster probabilities are time-varying and subject to learning.

The specific quantitative benchmark deserves emphasis. Under ?’s calibrated parameters, the unconditional variance premium is approximately 18 variance points (annualized), corresponding to a volatility premium of approximately 3.5 percentage points. The worst-case negative VRP under this calibration—obtained when a maximum-severity disaster occurs—is approximately -8 to -12 percentage points, depending on the assumed disaster severity. The COVID-shock VRP of -11.10 falls within this range, suggesting that the COVID event, while extreme, is not inconsistent with the rare-disaster framework at its upper bound. The GFC, by contrast, produced a negative VRP of only -0.42 —far less severe than the rare-disaster prediction, consistent with the interpretation that the GFC was a “correctly anticipated” crisis from the volatility-pricing perspective.

5.3 The post-bank-stress period

The 2023–2024 period following the March 2023 banking stress shows a VRP that is positive (+3.57) and broadly similar to historical averages, with episodic spikes (the August 2024 yen-carry-trade unwind, the April 2025 tariff-related shock). The continued positive VRP through this period supports the view that the volatility risk premium remains a persistent feature of US equity markets despite the contemporary monetary policy turbulence.

A specific observation about the post-2022 environment deserves emphasis. The monetary tightening cycle of 2022–2024 raised the policy rate from near zero to above 5%, with substantial increases in volatility-related uncertainty. The volatility risk premium has, despite this background, remained comfortably positive. The implication is that the contemporary policy environment has not eliminated the demand for volatility insurance; if anything, the elevated rate environment may have raised the demand for insurance through the leverage channel: higher interest rates increase the cost of leverage, making levered positions more sensitive to volatility shocks and increasing the demand for volatility protection.

The post-bank-stress period also provides a test of the VRP’s response to idiosyncratic financial-sector shocks. The SVB failure of March 2023 produced a spike in the VIX from approximately 19 to 26 within three trading days, with a corresponding decline in the VRP as realized volatility temporarily exceeded the pre-shock VIX level. However, the VRP recovered to its long-run average within approximately 30 trading days, consistent with the view that the premium’s dynamics are driven by the aggregate equity-market state rather than by sector-specific events.

5.4 The VIX-return relationship: descriptive and regression evidence

The finding that high-VIX quintiles are followed by elevated 21-day returns is consistent with the contrarian interpretation of the VIX. The formal regression analysis (Section 4.4b) strengthens this finding by demonstrating that the VIX coefficient retains statistical significance (HAC $t = 3.42$) after controlling for the term spread, credit spread, and dividend yield. The incremental predictive power of VIX, conditional on these controls, rules out the hypothesis that the VIX-return relationship is merely a proxy for business-cycle variation in expected returns captured by the term spread or credit spread.

The relationship should be qualified. The VIX-return regression R^2 of 0.018 implies that VIX explains less than 2% of the variation in subsequent 21-day returns. The low explanatory power reflects the fundamental difficulty of short-horizon return prediction and is consistent with the broader return-predictability literature (?). The economic significance of the predictor may nonetheless be substantial: ? document that predictors with R^2 values in the range of 0.5–2.0% at the monthly horizon can produce substantial utility gains for mean-variance investors who condition their portfolio allocation on the predictor.

The non-monotonicity of the quintile pattern—Q1 returns exceeding Q2 and Q4—suggests that the VIX-return relationship is better characterized as a nonlinear “extreme

VIX” effect than as a continuous linear relationship. The formal regression imposes linearity; a piecewise-linear or threshold specification might provide a better fit. We flag this as a productive extension but do not pursue it in the present paper.

5.5 Reproducibility and robustness

The analysis is exactly reproducible from publicly available daily VIX and S&P 500 data. The replication code computes the realized-volatility series, applies the Newey-West HAC inference, and aggregates the daily data to the regime-by-regime decomposition. The code is deposited at the journal’s online repository with a time-stamped commit hash.

The headline findings are robust across the pre-specified robustness margins. Alternative HAC truncation choices (15, 30, 45 lags) and the \hat{h} data-dependent bandwidth yield similar t-statistic magnitudes. Alternative realized-volatility windows (15-day, 30-day) yield similar qualitative patterns with somewhat different specific magnitudes. The \hat{h} range-based realized-volatility estimator produces results qualitatively consistent with the close-to-close estimator. Alternative VIX methodologies (the pre-2003 vs. the contemporary) produce broadly similar results for the post-2003 sample. The sub-period analysis (Section 4.5) confirms the qualitative stability across pre-and-post-GFC and pre-and-post-COVID partitions. The \hat{h} structural-break analysis produces break dates consistent with the historically motivated regime partition.

5.5a Connection to the long-run-risks framework

The persistent positive VRP we document is consistent with the long-run-risks framework of κ and σ . The framework attributes the premium to compensation for persistent shocks to expected consumption growth and to consumption volatility, with the premium varying systematically with macroeconomic conditions. The regime-by-regime variation we document (Section 4.3) is consistent with the framework’s prediction that the premium attenuates or turns negative during crisis episodes when consumption volatility shocks are realized.

A specific quantitative implication of the κ calibration is that the variance premium should be approximately proportional to the conditional volatility of volatility (the “vol-of-vol”). Under their calibration, the variance premium is approximately $\kappa \cdot \sigma_\sigma^2$, where κ is the ratio of risk aversion to the elasticity of intertemporal substitution and σ_σ is the volatility of the stochastic-volatility process. The post-COVID-recovery VRP of +7.82—nearly double the unconditional mean—is consistent with an elevated vol-of-vol following the pandemic shock, as the VVIX (the implied volatility of VIX options) was indeed elevated during the 2020–2021 period. The formal integration of the VVIX dynamics with our VRP record is beyond the scope of the present paper but is the natural connecting point between the descriptive evidence and the long-run-risks framework.

5.5b Recovery of risk-neutral and physical-measure distributions

κ demonstrates how the joint observation of option prices and realized returns can be used to

recover the risk-aversion function of the marginal market participant. The implied-realized gap we document is one input to this recovery exercise: the gap measures the discrepancy between the risk-neutral and physical-measure expectations of variance, with the magnitude of the gap informative about the risk-aversion of the marginal participant.

A version of the analysis that applies the Jackwerth (or related) recovery procedure to the contemporary data would identify how the risk-aversion of the marginal participant has evolved over the post-2020 period. We have not pursued the analysis but flag it as a productive extension. The elevated post-COVID VRP (+7.82) may reflect either an increase in the risk aversion of the marginal participant or an increase in the perceived tail-risk probability; the Jackwerth recovery would help distinguish between these two channels.

5.6 Implications for option-writing strategies: concrete portfolio metrics

The positive volatility risk premium implies that systematic option-writing strategies have positive expected returns. To provide concrete context for this implication, we benchmark the VRP against the realized performance of simple short-volatility strategies documented in the practitioner literature.

? documents that a strategy of systematically selling one-month at-the-money S&P 500 put options and delta-hedging weekly has earned an annualized Sharpe ratio of approximately 0.7–0.9 over the 1986–2010 period, with maximum drawdowns of approximately 30–40% during the GFC. The Sharpe ratio substantially exceeds the equity market Sharpe ratio of approximately 0.4 over the same period, confirming that the VRP represents a compensated risk factor.

Our regime-by-regime decomposition provides the raw material for assessing the downside risk of such strategies. During the COVID-shock window (52 trading days), the cumulative negative VRP of $-11.10 \times 52/252 \approx -2.3$ annualized percentage points implies that a strategy calibrated to earn +4 annualized percentage points of VRP would have experienced a substantial loss over this window. The recovery to +7.82 VRP in the post-COVID period (380 trading days) implies that the strategy would have recovered and exceeded its long-run expected return within approximately 18 months—a recovery speed that is informative for risk-budgeting purposes.

The practitioner considerations for implementation include: (i) position sizing in the range of 5–15% of portfolio capital, with most contemporary literature recommending the lower end following Volmageddon; (ii) variance-targeting overlays that adjust the short-volatility exposure inversely to the VIX level; (iii) tail-risk hedging through far-out-of-the-money put options to truncate the downside in crisis episodes; (iv) liquidity provisions for the rebalancing-during-stress phase that crisis episodes generate. The Volmageddon episode of February 2018, in which inverse-VIX exchange-traded products experienced losses of approximately 90% in a single trading day, provides the most vivid illustration that the positive VRP is not “free money” but represents compensation for bearing genuine

downside risk.

5.6a The pricing of tail risk and jump components

The variance risk premium can be decomposed into continuous and jump components, with the jump component plausibly accounting for a disproportionate share of the total premium. ¹⁷ document this decomposition empirically, finding that the jump-risk component is approximately 40–60% of the total premium during normal periods and rises substantially during crisis episodes. The decomposition has implications for both theoretical interpretation and practical strategy design: a premium dominated by jump risk requires different hedging strategies than a premium dominated by continuous volatility risk.

The COVID-shock VRP of -11.10 is particularly informative for the jump-risk decomposition. The realized volatility during March–April 2020 was dominated by a small number of extremely large daily returns (March 16, 2020: -11.98% on the S&P 500; March 12, 2020: -9.51%), consistent with a jump-dominated process. The implication is that the extreme negative VRP during the COVID shock was driven primarily by the realization of jump risk that the VIX had not fully priced—a finding consistent with the ¹⁷ framework in which jump intensity is time-varying and can spike to levels that exceed the market’s prior.

5.6b The relationship between VIX and VVIX

The CBOE also publishes the VVIX, an index that measures the implied volatility of VIX options. The VVIX captures the second-order volatility risk—the volatility of volatility itself. The implied-realized gap analysis we conduct on the VIX has a parallel application on the VVIX, with the implication that there is a second-order volatility risk premium (the implied-realized gap of the volatility of volatility) that is distinct from but related to the primary VRP. The literature on this second-order premium is in its early stages; ¹⁸ document a positive and time-varying VVIX risk premium.

The VVIX is relevant to the present paper’s findings because it provides a market-based measure of the uncertainty about future volatility—the “vol-of-vol” that the ¹⁷ long-run-risks framework identifies as the key driver of the variance premium. The post-COVID-recovery VRP of $+7.82$ coincided with an elevated VVIX (averaging approximately 130 during mid-2020, compared to a long-run VVIX average of approximately 90), consistent with the hypothesis that the elevated VRP reflected elevated vol-of-vol rather than elevated risk aversion per se.

5.7 Connection to broader asset-pricing literatures

The volatility risk premium is one component of a broader literature on the cross-section of risk premia. The literature on the cross-section of equity returns (¹⁹) and the literature on the cross-section of bond returns (²⁰) have together established that priced risk premia are pervasive across asset classes. The volatility risk premium fits into this broader framework as the premium associated with bearing the risk of large, infrequent volatility shocks—a risk

that is conceptually distinct from but empirically related to the cross-sectional risks that the multifactor literature has documented.

The integration of the VRP with the multifactor literature is particularly productive in the context of β , who document that sensitivity to aggregate volatility innovations is a priced cross-sectional factor: stocks with high sensitivity to increases in market volatility earn lower expected returns, consistent with the interpretation that volatility risk is a source of negative utility that investors demand compensation for bearing. The aggregate VRP we document is the market-level manifestation of this cross-sectional phenomenon.

5.8 Limitations

The VIX as constructed since 1993 represents an evolution of the methodology used by the CBOE; pre-1993 VIX values are reconstructed retroactively and should be interpreted with appropriate caution. The 21-trading-day realized volatility window is an imperfect match to the VIX's 30-calendar-day horizon; alternative window choices yield somewhat different specific magnitudes but the qualitative pattern is unchanged.

The standard errors reported above are computed under the Newey-West HAC framework. The framework addresses the autocorrelation that the 21-trading-day overlap induces but does not address other potential sources of inference error (e.g., the conditional heteroskedasticity that the GARCH literature emphasizes). β 's GARCH-option-pricing framework provides one approach to addressing this concern; the integration of the GARCH framework with the present analysis is a natural extension.

The descriptive nature of the analysis does not support causal claims about the underlying mechanism that produces the volatility risk premium. The literature has proposed multiple mechanisms (asymmetric loss functions, institutional frictions, consumption-disaster risk, behavioral anchoring); the present paper documents the empirical pattern without adjudicating among the mechanisms.

The regime partition, while robust to boundary-date perturbations and consistent with the β structural-break detection, is inherently ex post. The regime-level statistics should be interpreted as descriptions of the premium's behavior within historically identified episodes, not as predictions of what the premium will do in future episodes with similar characteristics. The forward-looking applicability of the regime decomposition depends on the stationarity of the premium-generating process, which cannot be verified from the historical sample alone.

The return-predictability regression controls for three standard macro-financial predictors but does not exhaustively condition on all known return predictors. The residual predictive power of VIX could, in principle, be subsumed by a larger set of controls that we have not included. We have selected the controls on the basis of their prominence in the return-predictability literature, not on the basis of an exhaustive horse race.

6. Conclusion

This paper has documented, using daily VIX and S&P 500 data over 1990–2025, the relationship between implied and subsequently realized volatility in US equity markets. The volatility risk premium has averaged approximately +4 percentage points over the full 35-year sample, has been positive on 85.1% of trading days, and has varied substantially across market regimes. The most negative episode in the sample is the March–April 2020 COVID shock, with an average VRP of -11.10 percentage points—an order of magnitude more severe than the GFC’s -0.42 , reflecting a qualitative difference between the two crises from the volatility-pricing perspective. The regression of forward realized volatility on contemporaneous VIX produces a slope of 0.887 , statistically distinguishable from one under Newey-West HAC inference ($t = 18.9$), consistent with a positive risk premium of the magnitude observed in the unconditional comparison.

The findings extend the established literature on the volatility risk premium through the most eventful period for volatility markets since the GFC. Three specific findings are novel relative to the pre-2020 literature. First, the COVID-shock VRP magnitude (-11.10) dramatically exceeds the GFC magnitude (-0.42), revealing that the two crises operated through fundamentally different volatility-pricing channels: the GFC was correctly priced by the options market, while the COVID shock was not. Second, the post-COVID-recovery VRP of $+7.82$ —nearly double the unconditional mean—suggests an upward shift in the price of volatility insurance following the pandemic, consistent with time-varying disaster-probability models. Third, the persistence of the premium at $+3.57$ through the 2023–2024 post-bank-stress period, during the most aggressive monetary tightening since 1981, provides out-of-sample confirmation that the premium is structural.

The formal return-predictability regression confirms that contemporaneous VIX retains incremental predictive power for subsequent 21-day equity returns (HAC $t = 3.42$) after controlling for the term spread, credit spread, and dividend yield, establishing that the contrarian VIX signal is not subsumed by standard macro-financial predictors.

6.1 What this paper provided

The empirical contribution of the paper is fivefold:

- Updated documentation of the volatility risk premium through 2025, including the post-COVID environment and the post-2022 monetary-tightening cycle, with the specific finding that the COVID-shock VRP of -11.10 is qualitatively distinct from the GFC VRP of -0.42 .
- Formal Newey-West HAC inference on the implied-realized regression, with transparent documentation of the factor-of-five reduction from the OLS t -statistic (98.0) to the HAC t -statistic (18.9), demonstrating both the statistical significance and the importance of proper inference.

- Regime-by-regime decomposition of the premium across twelve historically motivated market episodes, with formal robustness analysis including boundary-date sensitivity and ? structural-break detection.
- Formal return-predictability regression with macro-financial controls, establishing that VIX has incremental predictive power for subsequent equity returns (HAC $t = 3.42$) beyond the term spread, credit spread, and dividend yield.
- Connection of the empirical record to specific quantitative predictions of the rare-disaster and long-run-risks frameworks, providing calibration targets for future theoretical work.

6.2 Extensions

Several extensions warrant further development.

International equity markets. A replication of the analysis for international equity markets (FTSE, DAX, Nikkei, Hang Seng) would test the cross-country generalizability of the US findings. The preliminary evidence suggests that the volatility risk premium is positive across major equity markets but varies in magnitude. The international comparison would be particularly informative for the rare-disaster framework, which predicts cross-country variation in the premium based on the country-specific disaster history.

Cross-asset application. The volatility risk premium has been documented in foreign-exchange markets, commodities, and the term structure of interest rates. A version of the analysis that integrates the implied-realized gap across asset classes would provide a unified picture of cross-asset volatility risk premia. The cross-asset comparison would test whether the premium dynamics (particularly the COVID-shock magnitude and the post-COVID hysteresis) are specific to equity markets or reflect a common factor.

High-frequency analysis. The 21-trading-day window is one of several possible horizons. A version of the analysis at the intraday or weekly horizon would identify the time-scale dependence of the premium, which has implications for the continuous-vs.-jump decomposition of the premium.

Time-varying premium and macroeconomic conditioning. The contemporary literature on time-varying risk premia (?) has documented substantial variation in the premium as a function of macroeconomic conditions. A version of the analysis that conditions the premium on macroeconomic state variables—including a formal vector autoregression—would integrate the present documentation with the broader time-varying-premium literature.

Jump-risk decomposition. ? have decomposed the volatility risk premium into continuous and jump components. A version of the analysis that performs the decomposition on the post-2020 sample would identify whether the contemporary premium is driven primarily by continuous or jump risk, with the COVID-shock episode providing a particularly informative identification window.

Connection to the variance-swap market. The VIX is a model-free measure of the variance swap rate; the analysis can be extended to incorporate actual variance-swap transactions data. The transactions data would identify the realized profit-and-loss of variance-swap trades over the sample period, providing a complementary measure of the volatility risk premium that is unaffected by the methodological choices in constructing the VIX.

Connection to systemic-risk and financial-stability frameworks. The volatility risk premium is one component of the broader systemic-risk measurement infrastructure that the post-2008 financial-regulatory environment has developed. A version of the analysis that integrates the VRP with the contemporary systemic-risk indices (CoVaR, marginal expected shortfall, network-spillover measures) would identify how the VRP signal relates to the broader financial-stability monitoring framework that regulators use.

Linkage to the central-bank policy and communication literature. The contemporary central-bank communications strategy includes substantial attention to the expectations channel through which policy moves financial markets. A version of the analysis that examines the VRP response to FOMC communications would identify how policy announcements affect the implied-realized gap. The integration of the VRP with the high-frequency monetary-surprise literature (?) is a productive next direction.

Update through subsequent business cycles. The most consequential extension is to wait. The post-COVID sample contains only one full crisis episode (the COVID shock) and substantial subsequent recovery and tightening dynamics. Future business cycles will resolve whether the documented pattern—persistent positive premium interrupted by occasional severe negative-premium episodes—continues to characterize the contemporary US equity market.

6.3 A note on methodological discipline

The volatility risk premium is one of the most robust empirical regularities in contemporary financial economics, and the present paper has aspired to document the regularity with the methodological discipline that the contemporary inference literature requires. The Newey-West HAC correction—which reduces the t-statistic from 98.0 to 18.9, a factor-of-five reduction that demonstrates both the severity of the overlapping-observations problem and the robustness of the finding under proper inference—is the core methodological contribution. The regime-by-regime decomposition with formal boundary-sensitivity analysis, the structural-break detection, the formal return-predictability regression with macro-financial controls, the pre-specified robustness margins, and the public deposit of the analysis code together constrain the analyst's degrees of freedom and produce a finding that subsequent research can build on.

The fundamental empirical question—is the volatility risk premium real?—has been answered consistently across multiple decades of research. The contribution of the present paper is to update the record through the contemporary period, to characterize the post-

COVID dynamics that the prior literature has not yet engaged, and to provide three specific model-constraining observations (the COVID-shock VRP magnitude, the post-COVID hysteresis, and the premium's persistence through monetary tightening) that the equilibrium asset-pricing literature can use as calibration targets. The data and code for this analysis are publicly available at the GER online repository.