Identifying a Robust Risk-Return Relation

Benjamin Knox*

Federal Reserve Board

August 26, 2026

Abstract

Is there a conditional risk-return tradeoff in aggregate stock markets? This paper argues that the difficulty of answering this question reflects not only low power, but also finite-sample fragility: when return variance is the regressor, conventional OLS inference depends disproportionately on a small number of especially noisy observations. These statistical challenges motivate the use of more efficient and robust estimators, direct forward-looking measurement of market risk from option-prices, and an expanded testable dataset covering global markets. Combining these methods and data, I document a robust, significant and economically meaningful relation between risk-neutral variance and subsequent realized market returns. Local risk-neutral variance not only predicts returns within countries over time, but also explains differences in market returns across countries. A forecast of risk-neutral variance that is based on realized variance extends U.S. evidence back to 1885. Overall, the results reveal a stronger, more systematic relationship between forward-looking market risk and stock market returns over time and across countries than previously recognized.

*The views expressed in the paper are those of the author and do not necessarily represent the views of the Federal Reserve Board or the Federal Reserve System. All errors are my own. Contact: ben.knox@frb.gov. I thank Andrew Chen, Niels Joachim Gormsen, Theis Ingerslev Jensen, Eben Lazaras, Ian Martin, Tyler Muir, Lasse Heje Pedersen, Ivan Shaliastovich, Yannick Timmer, Annette Vissing-Jorgensen and Paul Whelan for comments, as well as seminar participants at the Federal Reserve Board.

1 Introduction

The risk–return tradeoff is a foundational pillar of modern financial theory. It is therefore striking that, even for the U.S. stock market— arguably the most studied asset in finance— the empirical strength of the time-series relation between expected returns and conditional market risk remains unresolved. While average returns on the U.S. stock market have far exceeded risk-free returns historically, indicating substantial compensation for bearing equity market risk, evidence on whether expected market returns rise and fall with the variance of market returns is mixed (see Merton (1980); French et al. (1987); Campbell (1987); Glosten et al. (1993); Ghysels et al. (2005); Lundblad (2007); Moreira and Muir (2017), among others).

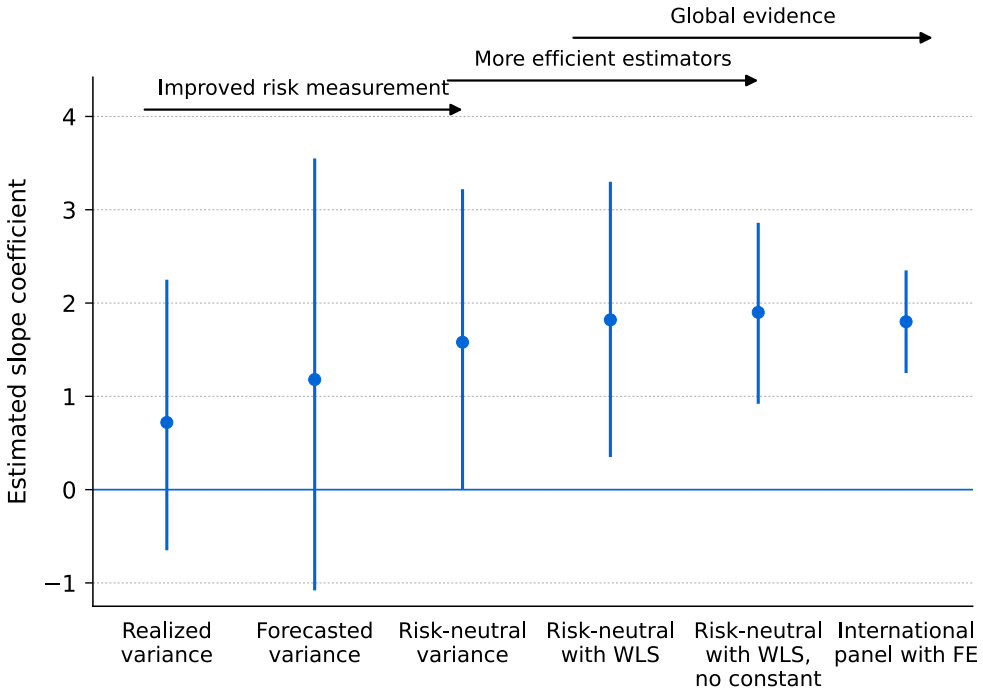
A central reason that decisive empirical evidence has remained elusive is that risk–return regressions are unusually difficult to estimate. Not only is the empirical setting inherently low in power, as is well-established in the literature, but I show an additional challenge that inference is particularly fragile in finite samples. This fragility arises because the predictor of returns is return risk itself and, importantly, is also heavily right-skewed. In practice, these features of the risk-return data generating process, which I refer to as *regressor-variance alignment*, result in OLS estimators loading almost entirely on a small set of particularly noisy observations.

In light of these statistical issues, this paper reassesses the empirical evidence on risk-return relations using new methods and data. The empirical analysis has three key dimensions. First, because low power magnifies the cost of measurement error, I emphasize the importance of the precise measurement of investors' assessment of market risk. Second, because OLS is inefficient and fragile in this setting, I focus on estimators that improve efficiency under heteroskedasticity and reduce sensitivity to influential yet noisy observations. Third, I expand the analysis across international markets. This both substantially increases the testable dataset and enables cross-sectional tests that exploit differences in risk across markets at the same point in time. Importantly, the cross-sectional variation in market risk is much less right-skewed than the time-series variation within countries, and thus provides a complementary test of conditional risk-return relations that is less exposed to the regressor-variance alignment problem even under OLS.

Figure 1 summarizes the central empirical results of the paper and illustrates how these ingredients progressively strengthen risk–return inference. The figure starts with the slope estimate from an OLS regression using backward-looking realized variance as the predictor. Even in this baseline specification, the estimated risk–return slope is positive and economically meaningful, but it is imprecisely estimated and statistically insignificant. Improving

Figure 1: Risk-return slope estimates with improved measurement, more efficient estimators, and the introduction of global data.

The figure reports slope estimates from predictive regressions of monthly excess market returns on alternative measures of the monthly variance of market returns over the sample 1996-2025. Whiskers denote 90% confidence intervals. The first three specifications are estimated with OLS and move from realized variance to forecasted variance and risk-neutral variance. The next two specifications estimate the risk-neutral variance specification using WLS, with and without a constant. The final specification uses the international panel (sample beginning 2003) using WLS and with index and date fixed effects.



measurement of the regressor strengthens the evidence, as moving to forecasted variance and then to option-implied risk-neutral variance increases both the estimated slope and precision. Estimators designed to down-weight high-variance observations and increase power then further improve precision. Finally, the introduction of international data produces still more precise and statistically significant estimates across major equity market indices.

A consistent theme throughout the results is the substantially stronger return predictability obtained using risk-neutral variance rather than using simple proxies for physical conditional variance. While the conventional risk–return question concerns physical conditional variance, physical conditional variance is latent and must be estimated, and in this sense risk-neutral variance provides the desirable properties of being a directly observable, forward-looking measure of market risk that is theoretically tied to physical expected returns (Martin, 2017). An important distinction, however, is that risk-neutral variance also reflects the pricing of variance risk. The stronger predictive power may therefore reflect better information

about investors' perception of underlying physical risk, economically relevant risk-neutral pricing information, or both. Indeed, a straight read of the out-performance of risk-neutral variance relative to the simple physical variance proxies used in this paper suggests a role for variance risk premium in return predictability (Bollerslev et al., 2009; Drechsler and Yaron, 2011; Bekaert and Hoerova, 2014). However, contrary to prior findings, I emphasize that risk-neutral variance subsumes the predictive information in the variance risk premium.[1]

An interesting question is the extent to which the optimal statistical forecast of the unobserved physical variance can improve return predictability (either directly, or via the implied variance risk premium). However, this is not the focus of this paper, which instead focuses on the directly observable risk-neutral measure. One additional benefit of this approach is that it avoids the need for a first-stage forecasting model in a setting that is already characterized by low power and finite-sample fragility. That said, an important trade-off is the short history of option-price data, which also raises a natural question: what can be learned about the risk-return relation over a much longer historical sample? To answer this question, I use the modern relation between realized and risk-neutral variance to construct a forecast of risk-neutral variance from realized variance, extending the U.S. evidence back to 1885. Using insights from the main sample, this analysis provide evidence on risk–return relations that have existed over the past 140 years.

Empirical setting and statistical challenges. The empirical analysis in this paper centers around the estimation of the simple regression:

$$r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1} \tag{1}$$

where r^{ex}_{t+1} is the excess return realized on a market index from time t to $t+1$, the regressor x_t is a measure of return variance and is the time t return signal, and ϵ_{t+1} is the unexplained return with time varying variance σ^2_{t+1}.

A well-known challenge in estimating the risk–return relation is that the economic signal is relatively small. Even economically large values of the slope coefficient β imply modest explanatory power because the average variance of returns is large relative to the time-series variation in this return variance itself (Merton, 1980; Lundblad, 2007). A second, less appreciated challenge is that this low-power problem is amplified by the distribution of return variance and its mechanical correlation with Equation 1 regression errors. Variance-based regressors are highly right-skewed and, by definition, increase precisely in periods when return innovations are most volatile. As a result, the observations with the greatest leverage

[1] These new results are driven by the extended sample and more efficient estimators uses in this paper.

for estimating the slope are small in number and occur precisely when the regression error is particularly noisy. I use calibrated Monte Carlo simulations to show how OLS remains centered on the correct slope under *regressor-variance alignment*, but produces highly fragile inference in finite samples. The simulated distribution of sample R^2 and estimated β exhibit a far larger dispersion than under a counterfactual case where x_t has the same distributional properties (and hence the same population R^2) but the regressor is orthogonal to the return variance.

An initial empirical application of Equation 1 to S&P 500 returns over 1996–2025 also illustrates these econometric issues in practice. I show how model estimates and inference under OLS depend almost entirely on two brief episodes — the Global Financial Crisis (GFC) and the early onset of COVID-19 crisis — during which realized (risk-neutral) return variance rose to more than 3.4 (2.5) times its maximum level at any other point in recent history. In the data, the relationship between return variance and realized returns is weak during the GFC but strong during COVID, so that the two episodes largely offset each other. As a result, full-sample OLS estimation yields a positive but statistically weak relationship between variance and one-month-ahead returns.[2]

However, conclusions about the fundamental relationship between variance and realized returns over a 30-year sample based on OLS estimation alone should be drawn with caution. These two episodes of extreme volatility account for over 60 percent of estimator leverage—defined as the fraction of demeaned variation in the regressor—despite representing less than 3 percent of the total observations. At the same time, these periods are exactly when unexpected returns were particularly dominant drivers of realized returns. These concerns motivate the deeper analysis of the U.S. time series and the introduction of international evidence developed in the remainder of the paper.

Efficient and robust evidence in the U.S. timeseries. Weighted least squares (WLS) estimation provides a natural improvement over OLS under time-varying volatility (Westerlund and Narayan, 2015; Johnson, 2019). By weighting each observation inversely to a measure of its return variance, the approach improves estimation efficiency since observations in high conditional return-variance periods are down-weighted. The efficiency benefits hold for any *any* forecaster of market returns, but are particularly valuable in the regressor-variance alignment setting since they also down-weight observations that are the source of the estimator fragility. I estimate conditional volatility using the Corsi (2009) Heterogeneous Autoregressive (HAR) model, which reproduces key features of realized volatility using past

[2]The relation is found to be stronger when extending the to 6-month and 1-year excess returns (Knox and Vissing-Jorgensen, 2024; Martin, 2025). However, given concerns about low power, this paper focuses on the one-month horizon to increase the number of unique observations relative to these longer horizons.

volatilities in a parsimonious and tractable specification. Applying this model to a WLS estimator yields economically and statistically stronger results: the estimated beta coefficient of 1.82 has a p-value of 0.035, indicating a substantially more precise relationship between risk-neutral variance and realized returns relative to the OLS estimate of 1.60 with a p-value of 0.097.

Although the WLS estimator choice improves precision relative to OLS, the approach does not fully eliminate fragility in the estimation of Equation 1. Forecasting conditional volatility can be particularly challenging in crisis periods. Even the most sophisticated econometric models of volatility exhibited signs of miss specification in the GFC (Bekaert and Hoerova, 2014), and the volatility implied by VIX futures also appeared miss-priced during Covid (Cheng, 2020). Since these crisis periods exert disproportionate leverage on the regression, and given errors in volatility expectations are systematically correlated with subsequent realized returns (Lochstoer and Muir, 2022), volatility estimation error could potentially affect WLS estimates and inference in a meaningful way.

However, evidence with a broader set of estimators confirms the robustness of the positive relationship between risk-neutral variance and returns in the U.S. stock market. I implement various alternative estimation approaches—trimming, quantile regression, and robust Huber-biweight M-estimation—that also mechanically reduce the influence of outliers. Across all these models, the link between risk-neutral variance and future returns is consistently positive and highly statistically significant with p-values ranging from 0.023 to 0.047. Results are also particularly strong when the two pre-specified crisis periods are excluded, with an OLS R^2 of 2.8 percent and the estimated beta above 3 with a p-value of less than 0.01. Although average realized returns at extreme levels of return variance are economically important, the fact that the observed positive risk-return relation is not driven by these brief but extreme episodes and, if anything, is stronger in other periods provides further evidence of a robust underlying relationship.

The baseline analysis of monthly returns also exploits overlapping daily observations to reduce estimation fragility relative to estimation based on sparse end-of-month data. Estimating Equation 1 separately for each trading day-of-the-month highlights the benefit: under OLS, slope coefficients for non-overlapping data range from 0.32 to 3.11 (s.d. = 0.77) depending on the day-of-month used, with model R^2 varying from 0.1 percent to 5.1 percent. Inference therefore depends wildly on which day of the month is used, highlighting the estimation fragility in the regressor-variance alignment setting. WLS substantially stabilizes estimates, narrowing the beta range to 0.75–2.75 (s.d. = 0.50), which is further evidence of the benefits of this estimator relative to OLS in the risk-return setting. Nevertheless, with the relatively short sample and extreme right-skew on the data generating process,

WLS remains somewhat sensitive to day-of-month effects. These concerns reinforce the importance of maximizing available where data, and help motivate our international results.

A separate issue concerns the intercept. Economically, expected excess returns should be zero when return variance is zero and, econometrically, the intercept is identified only by extrapolation since return variance is strictly positive. Moreover, unrestricted estimates produce constants that are small and statistically insignificant. I therefore also report zero-intercept specifications. This restriction allows the slope to be estimated from the total variation in the regressor rather than its demeaned variation, with this more powerful test reducing the standard error of the slope meaningfully (see Figure 1, for example).[3]

Consistent and stable evidence across international markets. Extending the analysis across countries provides a powerful way to strengthen inference on the conditional risk-return relation. Assuming a common slope parameter, international data identify the risk-return coefficient using both changes in market risk over time and differences in risk across countries at the same point in time. The latter variation is particularly useful in light of the econometric challenge documented above. While time-series variation in return variance is heavily right-skewed, much of this skewness reflects common movements in global market risk. Removing date-specific means reduces the skewness of risk-neutral variance in the international sample from 4.08 to 1.57. Cross-country variation therefore provides identification of the same risk-return coefficient that is substantially less exposed to the regressor-variance alignment fragility affecting time-series inference.

Using a panel of ten equity indices spanning four continents, I first show that the time-series relationship holds consistently across markets. Pooling estimation across countries improves precision relative to standalone index tests, such that even under OLS the relationship is statistically significant at the five percent level. These gains in power arise even with the high correlation in risk-neutral variance across countries, which to some extent limits the independent time series variation that is added in the global data, due to the economically meaningful variation across countries that the cross sectional tests illuminate.

Indeed, the international panel enables novel tests that examine the cross-sectional relationship between risk-neutral risk and market returns. Results show that local option prices not only predict domestic index returns, but also explain cross-country return differences with markets with higher ex-ante risk earning higher subsequent excess returns. The inter-

[3]The theoretical argument for dropping the intercept is much weaker than assuming the estimated model is true. Under the assumption of no arbitrage, any model has zero expected return on an asset with zero variance. That said, if the true model is non-linear, it might be that a linear function with an intercept is a better approximation of the true model than a linear function through the origin. For this reason, I always report estimates with and without an intercept.

national cross-sectional results remain stable across time periods and hold when controlling for both time and country fixed effects, as well as under efficient WLS estimation. This robustness across periods, specifications, and weighting schemes provides strong evidence of a systematic relationship between option-implied risk and stock-market returns.

Additional results and interpretation. The final section of the paper asks how the main risk-return evidence should be interpreted and how far it extends beyond the option sample. First, I decompose risk-neutral variance into proxies for expected physical variance and the variance risk premium (VRP). Consistent with Bekaert and Hoerova (2014), I show that the VRP estimates are the more informative component of risk-neutral variance for predicting realized returns. However, in contrast to the prior literature, I find the explanatory power of VRP is largely spanned by risk-neutral variance itself.[4] Second, I construct a synthetic risk-neutral variance measure from realized-variance data, using the modern option sample to discipline the mapping from realized volatility to option-implied risk. This generated regressor should be viewed as a conservative proxy for out-of-sample return predictability: even in the option sample, replacing observed risk-neutral variance with the synthetic measure reduces explanatory power. Despite this limitation, the long historical data strongly supports the paper's main conclusions. Using the same econometric lessons from the option sample—through WLS, crisis-robust specifications, and zero-intercept estimates—the estimated risk-return relation is consistently positive, economically meaningful, and statistically strong over nearly 140 years of U.S. equity market history.

1.1 Related literature

This paper builds first on the large empirical literature studying the intertemporal relation between market returns and conditional market risk that followed the theoretical foundations of the Merton (1973) intertemporal capital asset pricing model. Early evidence is mixed: French et al. (1987) find a positive relation between expected returns and predictable volatility, while Campbell (1987), Baillie and DeGennaro (1990), and Glosten et al. (1993) report weak or negative relations under alternative conditioning information and GARCH-in-mean specifications. Subsequent work finds stronger evidence using improved risk measurement, richer conditioning information, and longer samples, including mixed-frequency forecasts (Ghysels et al., 2005), high-frequency measures (Bali and Peng, 2006), factor-augmented conditioning information (Ludvigson and Ng, 2007), and much longer historical samples

[4]The difference in conclusion reflects new methods but also, to a large degree, an extended sample that is now less dominated by the GFC.

(Lundblad, 2007). I contribute to this literature by identifying a novel econometric mechanism that generates inference fragility and by providing new and robust empirical evidence in support of a positive risk-return relation.

The econometric contribution relates to the literature on finite-sample inference in return-predictability regressions. Stambaugh (1999) shows that persistence in predictive variables, together with correlation between predictor and return innovations, can generate substantial small-sample bias, motivating subsequent bias corrections and alternative inference procedures (Amihud and Hurvich, 2004; Lewellen, 2004; Campbell and Yogo, 2006; Jansson and Moreira, 2006). The mechanism emphasized here is distinct: regressor–variance alignment does not bias OLS under the maintained model, but reduces precision and makes inference unusually fragile because observations with the greatest leverage for estimating the slope are also those with the highest conditional return variance.

In the empirical analysis, the paper connects to work using WLS and related methods in return-predictability regressions. WLS has been applied with a broad range of predictors (Johannes et al., 2014; Westerlund and Narayan, 2015; Johnson, 2019) as well as in specifications in which the predictor is itself a measure of return volatility (French et al., 1987; Johannes et al., 2014; Campbell et al., 2018). I show why such methods are especially valuable under regressor–variance alignment and complement WLS with trimming, quantile regression, robust M-estimation, and international panel evidence.[5]

The empirical focus on risk-neutral variance relates to a broader literature using option-implied risk to predict returns (Bollerslev et al., 2015; Martin, 2017; Chabi-Yo and Loudis, 2020; Tetlock et al., 2024; Chabi-Yo et al., 2025). However, rather than constructing the most informative option-implied expected-return measure, this paper focuses on risk-neutral variance as a canonical measure of conditional risk.[6] The introduction of international data relates to work using global option-market data to study variance risk premia and stock returns (Bollerslev et al., 2014; Londono, 2015),[7] but differs by focusing directly on the relation between risk-neutral market risk and realized market returns across both time and

[5] A related and growing body of recent survey evidence finds positive relations between subjective expected returns and perceived risk across a range of asset classes and markets (Couts et al., 2023; Kremens et al., 2025; Pesch et al., 2024; Leombroni et al., 2026; Dahlquist and Ibert, 2024; Jensen, 2022). By using expected rather than realized returns, this literature avoids realized-return variance and thus the regressor–variance alignment problem emphasized here.

[6] The empirical evidence for the U.S. is stronger than previously presented in earlier studies (Martin, 2017; Bekaert and Hoerova, 2014; Eraker, 2025) for two reasons. First, the predictive relation with an OLS estimator has strengthened in the post-publication samples (Knox and Vissing-Jorgensen, 2024; Knox et al., 2024; Martin, 2025). Second, the robust and efficient estimation methods developed in this paper further sharpen inference.

[7] Londono and Xu (2024) also use U.S. variance risk premium decomposed into upside and downside risk to predict international stock returns.

countries.

The paper also contributes to the broader literature on equity market return predictability (Welch and Goyal (2008); Goyal et al. (2024)). The objective here, however, is not to estimate the true or optimal forecast model for the market return, but instead to establish whether conditional market risk is statistically and economically related to subsequent realized returns. For this reason, my analysis emphasizes in-sample estimation to maximize statistical power (Inoue and Kilian, 2005; Campbell and Thompson, 2008).

Finally, the paper is complementary to research showing that the compensation for market risk can vary across time and states. Moreira and Muir (2017) show that scaling market exposure down when volatility is high can improve Sharpe ratios, suggesting that expected returns do not rise sufficiently with volatility to eliminate the benefits of volatility timing, with this result holding even when considering the perspective of long-term investors who can buy and hold for longer periods Moreira and Muir (2019). One interpretation of the large returns to volatility-timed strategies is that the price of risk falls in high-volatility states (Schreindorfer and Sichert, 2025; Jensen and Lazarus, 2026) while another interpretation is that expectations adjust slowly after bad news and therefore realized returns in high volatility periods are lower than investors' expected (Moskowitz et al., 2012; Gómez-Cram, 2022; Lochstoer and Muir, 2022; Yang, 2026). Against this background, this paper asks a simpler empirical question: do realized market returns nevertheless rise reliably with conditional market risk on average? I find that they do, with the relation strongest when risk is measured using risk-neutral variance, inference accounts for the econometric challenges inherent in the setting, and international markets expand the testable dataset.

2 Empirical measurement

This section defines the measures of market return risk used throughout the paper and describes the data used to construct them. The section then summarizes the U.S. and international data that form the basis for the analyses.

2.1 Measurement of stock market return risk

Realized variance of market returns. The realized variance of index returns from time $t-1$ to t can be measured from the sum of squared returns within the period:

$$RV_t = \sum_{j=1}^{n} \left(r_{t-1+(j-1)/n} \right)^2 \tag{2}$$

where the use of higher frequency intervals has been shown to lead to more accurate approximations of the true return variation (Andersen et al., 2001, 2003; Barndorff-Nielsen and Shephard, 2002a,b). As is standard in the literature, this paper uses realized variance estimates based on 5-minute return intervals when intra-day data is available. In all other cases, monthly return variance is estimated from daily return realizations.

Forecasted variance of market returns. To take a stance on the conditional expectation of next month's return variance, as a baseline measure we use the parsimonious HAR (Heterogeneous Autoregressive) forecasting model of Corsi (2009). This model is well-known for reproducing key features of return volatility in a tractable way by using past volatility over multiple look-back horizons. Specifically, the model estimates:

$$RV_{t+1}^{month} = \alpha + \beta_d \, RV_t^{day} + \beta_w \, RV_t^{week} + \beta_m \, RV_t^{month} + \epsilon_{t+1}^{RV} \tag{3}$$

where RV_t^{day} is the realized variance on the last day of month t, RV_d^{week} is the realized variance in the last week of month t, and RV_d^{month} is is the realized variance over the month of t.

Risk-neutral variance of market returns. The conditional risk-neutral variance of the gross equity market return R_{t+1} can be computed from index option prices:

$$Var_t^* \left(R_{t+1} \right) = \frac{2R_{t+1}^f}{P_t^2} \left[\int_0^{F_t} \text{put}_t(K) dK + \int_{F_t}^{\infty} \text{call}_t(K) dK \right] \tag{4}$$

where R_{t+1}^f is the gross risk-free rate, P_t is the price of the market index, $F_t = P_t R_{t+1}^f$ is the forward price of the index, and $\text{put}_t(K)$ ($\text{call}_t(K)$) are prices of out-of-the-money put (call) options with strike price K expiring at date $t+1$. Implementing Equation 4 requires discretizing the integrals over the available range of option strike prices (see Martin (2017), among others).

2.2 Data Summary

End-of-day option price data is obtained from OptionsMetrics. Analysis for the U.S time series uses the S&P 500 index, with the sample beginning January 4, 1996 and running through to July 31, 2025. The international analysis includes the following 10 index (countries/regions): AEX (Netherlands), ASX 200 (Australia), CAC 40 (France), DAX (Germany), EURO STOXX 50 (Euro area), FTSE 100 (United Kingdom), FTSE MIB (Italy), Nikkei 225 (Japan), S&P 500 (United States), TAIEX (Taiwan). Aside from the US, the

Table 1: U.S. Summary Statistics (1996-2025).

This table reports summary statistics for monthly U.S. equity returns and variance measures from 1996-2025. Panel A presents distributional properties including count, mean, standard deviation, and selected percentiles for excess returns, alternative variance measures based on both risk-neutral and realized volatility, and various variance risk premium estimates. All variables are scaled by a factor 100, which means the realized returns in row 1 are expressed in percent. Panel B displays the correlation matrix across these return and variance series.

Panel A: Monthly S&P500 index returns and measures of the variance of the market return

	count	mean	sd	p5	p25	p50	p75	p95
Realized excess return (in %)	7,422	0.58	4.72	-7.39	-1.79	1.12	3.39	6.98
Realized variance (high-frequency measurement)	7,422	0.25	0.43	0.04	0.08	0.14	0.26	0.71
Realized variance (daily measurement)	7,422	0.31	0.59	0.04	0.09	0.16	0.32	0.85
Forecasted variance (HAR model)	7,422	0.24	0.28	0.11	0.13	0.17	0.24	0.51
Risk-neutral variance ($SVIX^2$)	7,422	0.34	0.34	0.10	0.16	0.25	0.41	0.84
Risk-neutral entropy (VIX^2)	7,422	0.40	0.41	0.11	0.17	0.29	0.47	0.98
VRP (forecasted variance)	7,422	0.12	0.17	-0.02	0.02	0.08	0.18	0.41
VRP (realized variance)	7,422	0.09	0.24	-0.07	0.05	0.09	0.16	0.32

Panel B: Correlation matrix of U.S. returns and variance measures

	Ex. return	RVar, hf	RVar, d	FVar	RN Var	RN Entropy	VRP, FV	VRP, RV
Ex. return	1.000							
RVar, hf	0.076	1.000						
RVar, d	0.066	0.984	1.000					
FVar	0.072	0.858	0.857	1.000				
RN Var	0.115	0.838	0.834	0.868	1.000			
RN Entropy	0.113	0.843	0.839	0.876	0.998	1.000		
VRP, FV	0.118	0.346	0.341	0.186	0.648	0.631	1.000	
VRP, RV	0.025	-0.634	-0.610	-0.333	-0.109	-0.121	0.288	1.000

data availability start dates range from January 2003 (for EURO STOXX 50, DAX and FTSE 100) to January 2007 (for AEX and FTSE MIB) and was collected through to June 2022.[8] The steps used to clean the option data and compute risk-neutral moments of returns is outlined in Appendix E.

Daily price levels for the international panel of stock market index are also obtained from OptionsMetrics, along with term structures of zero-coupon risk-free rates. High frequency price-level data for the S&P500 index is obtained from the London Stock Exchange Group Tick History database (formerly known as Refinitiv Tick History) and Bloomberg and is available from 1990. The 30-day CBOE Volatility Index (VIX) and the 3-month market yield on Treasury Securities are both retrieved from FRED, Federal Reserve Bank of St. Louis. Theses daily time series are available from January 1990 and September 1981 respectively, and were collected through to September 2025. In the extended U.S. time series sample analysis in section 5, I use daily market return data back to 1926 using data from CRSP and further back to 1885 using data from Schwert (1990). I then use risk-free rate data taken from Jordà et al. (2019).

Table 1 Panel A presents summary statistics of monthly excess returns on the S&P500 index over the sample 1996-2025, along with option-implied, model-implied and physical measures of the variance of the return on the market, and measures of the variance risk premium. Panel B then presents a correlation matrix of these variables.

3 Empirical setting and econometric considerations

This section formalizes the empirical setting underlying the paper's predictive regressions and explains why this setting is unusually challenging. The central issue is regressor-variance alignment: variance-based predictors are highly right-skewed and receive their greatest slope leverage precisely when return innovations are most volatile. These observations motivate the weighted, robust, and restricted-intercept estimators used in the empirical analysis, and their finite-sample implications are illustrated in a calibrated Monte Carlo exercise.

3.1 Empirical framework and key features of the data

As outlined in the introduction, the empirical analysis in the paper is based on estimating

$$r_{t+1}^{ex} = \alpha + \beta x_t + \epsilon_{t+1}, \tag{5}$$

[8]The index included for each index are based on the best data quality in terms of number of unique strike prices, total open interest and total volume. Future iterations of the paper will expand the sample of countries and dates where data quality allows.

where r^{ex}_{t+1} denotes the one-month excess return on the market index, x_t is a measure of return risk observed at time t, and ϵ_{t+1} is the return innovation. We allow for conditional heteroskedasticity, so that

$$\text{Var}_t(\epsilon_{t+1}) = \sigma^2_{t+1}, \tag{6}$$

where σ^2_{t+1} denotes the conditional variance of next month's excess return.

In fact, this heteroskedasticity is the underlying source of the return predictability that we are studying. All the regressors x_t considered in the paper are forward-looking measures of next month's return variance. In some specifications, x_t is a forecast of future variance under the physical measure, while in others it is the conditional *risk-neutral* variance inferred from option prices. The common feature is that each x_t is observed at time t and can be interpreted as proxy for variation in σ^2_{t+1}.

To illustrate the properties of these regressors in the data, Figure 2 plots the realized monthly variance and the risk-neutral return variance for the S&P 500 from 1996 to 2025. Both measures remain low and relatively stable for most of the sample, but are punctuated by two short and extreme volatility episodes: the global financial crisis in late 2008 and early 2009, and the onset of Covid-19 in March and April 2020. In these two episodes, monthly return variance briefly rises above 0.06. In comparison, over the rest of the sample monthly return variance does not cross the 0.02 threshold. As a result, realized variance is strongly right-skewed, and the same is true of the forward-looking risk measures, such as risk-neutral variance, that are the main predictors in our Equation 5 estimations.

3.2 Econometric Challenges

The distributional features of market returns illustrated in Figure 2 create two distinct econometric challenges for the estimation of Equation 5. We explain these issues in this subsection.

Challenge 1: Limited Statistical Power. The population R^2 from the OLS estimation of Equation (5) is:

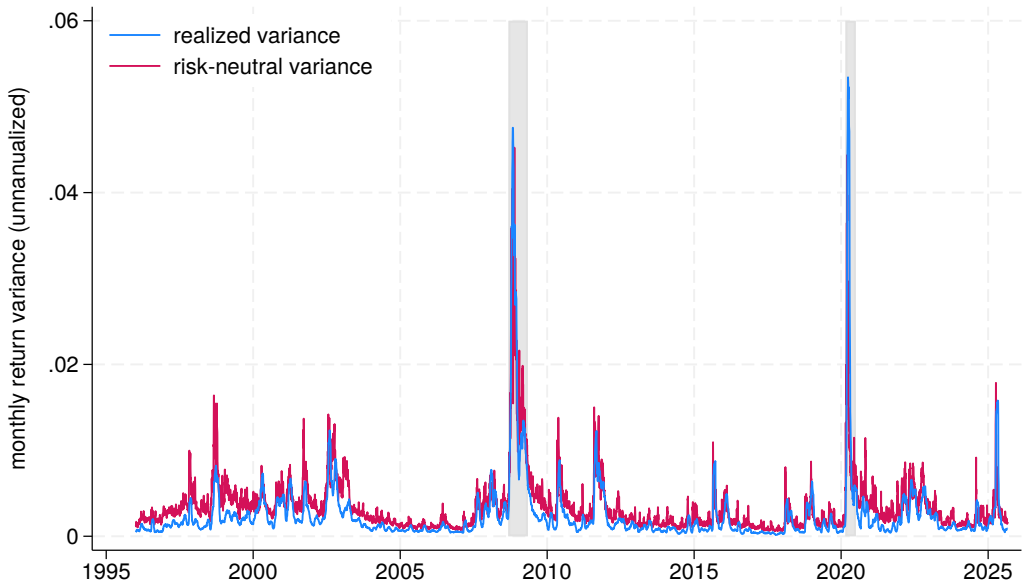
$$R^2_\infty = \frac{\beta^2 \text{Var}(x_t)}{\beta^2 \text{Var}(x_t) + \mathbb{E}[\sigma^2_{t+1}]} = \frac{1}{1 + \frac{\mathbb{E}[\sigma^2_{t+1}]}{\beta^2 \cdot \text{Var}(x_t)}}, \tag{7}$$

which shows the statistical power depends on strength of the relationship, β, and the time-series variation in the regressor, $\text{Var}(x_t)$, relative to the mean level of return variance, $\mathbb{E}[\sigma^2_{t+1}]$.

Table 2 presents the implied population R^2 across a range of plausible β values for alternative variance measures. Each entry is calculated from Equation 7 using the observed average monthly return variance, $\mathbb{E}[\sigma^2_{t+1}] = 0.0025$, together with the observed variance of

Figure 2: The time-variation and skew of realized S&P 500 return variance.

This figure shows the realized variance and risk-neutral variance of one-month returns on the S&P 500 index. The sample runs from January 1996 to July 2025 and data is plotted at a daily frequency. The two shaded areas indicate the global financial crisis and Covid-19 episodes, which are the only two periods in the sample during which realized variance exceeds 0.06 and risk-neutral variance exceeds 0.04.



the corresponding regressor, $\text{Var}(x_t)$, over the 1996–2025 sample. The first row uses the realized variance over the next month, $x_t = RV_{t+1}$, as the regressor. Under this benchmark, in which next month's variance is assumed known at time t, the implied R^2 is already quite low. As emphasized by Lundblad (2007), economically meaningful values of β yield low R^2 in standard risk-return regressions because the mean level of return variance $\mathbb{E}[\sigma_t^2]$ is large relative to its time-series variation $\text{Var}(\sigma_t^2)$ that was illustrated in Figure 2. For example, assuming $\beta = 2$, the population R^2 is only 2.87 percent.

In practice, investors do not observe ex post realized variance at time t, but instead form forecasts of future variance:

$$RV_{t+1} = x_t + \eta_{t+1} \tag{8}$$

where $x_t = \overline{RV}_{t+1}$ is the ex-ante forecast and η_{t+1} is the forecast error. By definition, a forecast will have lower variance than the underlying variable, with $\text{Var}\left(RV_{t+1}\right) > \text{Var}\left(x_t\right)$, if the forecast is optimal since the forecast and forecast error are uncorrelated. Consistent with this, the second row shows that the Corsi (2009) forecast of return variance has a time series variance substantially lower than the true realized variance. This reduction in regressor variance causes the population R^2 to fall dramatically—from 2.87 percent to 1.24 percent when $\beta = 2$. This shows that, while return variance is fairly well forecastable, the

Table 2: Implied population R^2 under alternative β values

This table reports implied population R^2 of Equation 5 under different slope coefficients, calibrated to the observed distributions of realized, forecasted and risk-neutral monthly return variance over 1996-2025. The population R^2 is calculated from Equation 7 using the observed sample variance of each measure, x_t, and assuming $E\left[\sigma_t^2\right] = 0.0025$, which is the sample mean of realized monthly return variance.

		Population R^2 (%)			
Variance Measure	$\text{Var}(x_t)$	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 5$
Realized variance	0.00185	0.74	2.87	6.25	15.62
Forecasted variance	0.00078	0.31	1.24	2.74	7.27
Risk-neutral variance	0.00116	0.46	1.81	4.00	10.38

unpredictable component of realized return variance mechanically reduces the achievable R^2 in risk-return regressions. The third row then presents analogous analysis calibrated to the observed risk-neutral variance. The time-series variance of risk-neutral variance exceeds that of the forecasted variance, increasing the population R^2 for fixed betas.

Challenge 2: Fragility from regressor–variance alignment. A distinct challenge arises from the combination of two features of the regressors that we study. First, they by definition increase with return risk and, second, they are highly right-skewed. Because slope leverage depends on the squared distance of x_t from its mean, right-skewness concentrates high-leverage observations in the upper tail of x_t, precisely where return variance also tends to be highest. The observations that matter most for estimating β are therefore also especially noisy. In finite samples, this alignment causes the slope to be disproportionately shaped by a handful of extreme observations, making OLS estimation highly fragile.

This mechanism can be viewed through the asymptotic variance of the OLS estimator. Let $\tilde{x}_t = x_t - E[x_t]$, so that $E[\tilde{x}_t^2] = \text{Var}(x_t)$, and define

$$H_t = \frac{\tilde{x}_t^2}{\text{Var}(x_t)}, \qquad E[H_t] = 1. \tag{9}$$

Under standard regularity conditions,

$$\sqrt{T}\left(\hat{\beta}_{OLS} - \beta\right) \xrightarrow{d} N(0, \Omega_{OLS}), \tag{10}$$

with

$$\Omega_{OLS} = \frac{E[\tilde{x}_t^2 \sigma_{t+1}^2]}{\left(\text{Var}(x_t)\right)^2}. \tag{11}$$

Under homoskedasticity, with $\sigma^2_{t+1} = \sigma^2$ for all t, this reduces to

$$\Omega_{OLS} = \frac{\sigma^2}{\text{Var}(x_t)}. \tag{12}$$

More generally, however, equation (11) can be written as

$$\Omega_{OLS} = \frac{E[\sigma^2_{t+1}]}{\text{Var}(x_t)} + \frac{\text{Cov}(H_t, \sigma^2_{t+1})}{\text{Var}(x_t)}. \tag{13}$$

This decomposition isolates the effect of regressor–variance alignment on OLS precision. The first term is the benchmark variance that allows for heteroskedasticity and arises absent any systematic relation between slope leverage and conditional volatility. The second term captures the additional effect of alignment. If high-leverage realizations of x_t occur when return innovation variance is also high, $\text{Cov}(H_t, \sigma^2_{t+1}) > 0$, and the asymptotic variance of OLS increases.

Importantly, a positive relation between x_t and σ^2_t alone need not imply positive regressor–variance alignment. Because H_t depends on squared deviations from $E(x_t)$, observations in either tail of the x_t distribution can have high slope leverage. If the distribution of x_t were symmetric, the low-x_t, low-variance observations balance the high-x_t, high-variance observations. However, skewness breaks this symmetry.[9] In our setting, slope leverage is concentrated overwhelmingly in the right tail of the x_t distribution, where return variance is also especially high. The combination of regressor-linked heteroskedasticity and right-skewness therefore generates positive alignment between slope leverage and return variance.

Table 3 shows the relevant distributional features of the three alternative return variance measures used in the paper. The concentration in the right tail is striking across all measures. For risk-neutral variance, observations in the top 5% of the x_t distribution account for 74.1% of total slope leverage,

$$h_t = \frac{(x_t - \bar{x})^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2}, \tag{14}$$

which is the finite sample equivalent of H_t in Equation 9.[10] In contrast, the bottom 5%

[9] Appendix B illustrates this mechanism in an affine benchmark in which the risk predictor is proportional to conditional return variance. In this case, the low-x_t, low-variance observations exactly offset the high-x_t, high-variance observations, and the alignment term is directly proportional to the third central moment of the regressor.

[10] In a simple regression with an intercept, the conventional hat-matrix leverage of observation t is

$$\frac{1}{T} + \frac{(x_t - \bar{x})^2}{\sum_{s=1}^{T}(x_s - \bar{x})^2}.$$

Thus, h_t is the component of conventional leverage attributable to the slope, normalized to sum to one across

Table 3: Leverage concentration in regressors.

This table reports the slope-leverage share (in percent) coming from observations in different parts of the x_t distribution. Leverage is defined in Equation 14 and measures the influence of observation t on the estimated slope coefficient $\hat{\beta}$. Columns (1)–(5) report leverage shares from the top 1%, top 5%, bottom 1%, bottom 5% of the x_t distribution, and crisis periods (including the GFC and Covid). The Gini coefficient in column (6) measures concentration in the leverage distribution, and skewness in column (7) is computed from the full x_t distribution. The sample period is January 1996–July 2025, using daily observations of monthly variance of returns on the S&P 500 index.

	Share of Slope Leverage (%)					Distribution	
	(1) $x_t > p_{99}$	(2) $x_t > p_{95}$	(3) $x_t < p_1$	(4) $x_t < p_5$	(5) $x_t =$ crisis	(6) Gini	(7) Skewness
Realized variance	73.56	89.12	0.23	1.06	78.41	0.926	8.73
Forecasted variance	71.99	87.34	0.27	1.22	80.47	0.914	7.03
Risk-neutral variance	50.20	74.08	0.65	2.92	63.26	0.845	4.68

account for only 2.9%—a ratio of approximately 25-to-1. This concentration is reflected in a Gini coefficient of 0.845, while the asymmetry of the underlying regressor distribution is captured by a sample skewness of 4.68.[11] The corresponding patterns are even more pronounced for the other variance measures shown in the table. These features imply that OLS inference depends heavily on a small set of extreme observations, which occur precisely the periods in which return innovations are also most volatile.

Taken together, the decomposition in Equation 13 and the distribution of regressors in Table 3 show that regressor–variance alignment is a central econometric issue in our setting. This motivates the use of estimators that either exploit known heteroskedasticity for efficiency or reduce sensitivity to extreme tail observations for robustness.

3.3 Estimation methods

We consider three aspects of estimation in Equation 5: efficiency under conditional heteroskedasticity, robustness to influential crisis observations, and the treatment of the intercept.

observations. We refer to h_t as the slope-leverage share.

[11] Gini coefficients approaching 1 indicate near-complete concentration.

3.3.1 More efficient estimators

The efficient linear estimator under heteroskedasticity is generalized least squares (GLS), which solves:

$$\arg\min_{\alpha,\beta} \sum_{t=1}^{T} \frac{(r_{t+1}^{ex} - \alpha - \beta x_t)^2}{\sigma_{t+1}^2} \tag{15}$$

and thus places less weight on observations with higher conditional error variance. The resulting sampling variance of the GLS slope coefficient is:

$$\text{Var}(\hat{\beta}_{\text{GLS}}) = \left(\sum_{t=1}^{T} \frac{(x_t - \bar{x}_w)^2}{\sigma_{t+1}^2} \right)^{-1}, \tag{16}$$

where $\bar{x}_w = \frac{\sum_{t=1}^{T} x_t/\sigma_{t+1}^2}{\sum_{t=1}^{T} 1/\sigma_{t+1}^2}$ is the variance-weighted mean. Whenever σ_{t+1}^2 varies across t, then $\text{Var}(\hat{\beta}_{\text{GLS}}) < \text{Var}(\hat{\beta}_{\text{OLS}})$ and thus GLS improves estimation efficiency relative to OLS.

However, since σ_{t+1}^2 is not directly observed, empirical analysis typically implements feasible GLS via weighted least squares (WLS). Let

$$(\hat{\alpha}_W, \hat{\beta}_W) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} w_t \left(r_{t+1}^{ex} - \alpha - \beta x_t \right)^2, \tag{17}$$

where $w_t > 0$ is measurable with respect to time-t information. Under the maintained condition $E_t(\epsilon_{t+1}) = 0$, WLS remains consistent for the same slope coefficient β for any admissible weight function w_t that depends only on time-t information. Its efficiency improvement, however, depends on how closely the weights track the inverse conditional variance of the regression error. Importantly, the relevant object for efficient weighting is therefore the physical conditional variance of the subsequent return, rather than its risk-neutral counterpart. Although risk-neutral variance is directly observable and forward-looking, it incorporates variance risk premia and therefore need not provide an efficient forecast of subsequent realized variance (Chernov, 2007).

In our main specifications, we implement WLS estimators using externally constructed variance forecasts as precision weights.[12] This approach can deliver meaningful reductions in estimator variance when next period's return volatility is predictable with current and past return volatility. While the efficiency gains hold in the estimation of Equation 5 with any regressor x_t, they are particularly relevant in our empirical setting with x_t themselves

[12] An alternative approach to feasible GLS is to use weights recovered from first-stage regression residuals. However, given that high-frequency return data allows precise measurement of realized return variance, and the fact that a large component of return variance is predictable, we prefer to use the externally constructed variance forecasts.

variance forecasts, since WLS down-weights the observations that are also the source of the inference fragility.

That said, return variance cannot be perfectly forecasted, and it is especially challenging to do so during crisis periods (Bekaert and Hoerova, 2014; Cheng, 2020). This means that, while WLS improves efficiency in principle, those gains are not automatic. If the weight function is misspecified, feasible WLS need not improve efficiency, and the resulting estimates may be more sensitive to modeling choices in precisely the states where alignment is most severe.

3.3.2 More robust estimators

The potential fragility of WLS under misspecification motivates a second class of estimators that are designed to be more robust and less sensitive to extreme observations. We consider three such approaches: trimming, quantile regression, and robust M-estimation.

Trimming. As leverage is concentrated almost entirely in the upper tail of the x_t distribution, one diagnostic is to re-estimate the predictive regression after excluding observations above a prespecified cutoff:

$$(\hat{\alpha}^{(c)}, \hat{\beta}^{(c)}) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} \mathbf{1}\{x_t \leq c\} \left(r_{t+1}^{ex} - \alpha - \beta x_t\right)^2, \tag{18}$$

where c may be chosen as, for example, the 99th or 95th percentile of the predictor distribution. In our setting, upper-tail trimming is more relevant than symmetric trimming because the predictor is strongly right-skewed. The purpose of this estimator is not to improve efficiency under a maintained model of heteroskedasticity, but rather to assess how much of the estimated mean relationship is driven by a very small number of crisis observations. A related exercise is to exclude crisis periods directly. Let D_t denote an indicator equal to one in crisis periods and zero otherwise. The corresponding estimator solves

$$(\hat{\alpha}^{(NC)}, \hat{\beta}^{(NC)}) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} (1 - D_t) \left(r_{t+1}^{ex} - \alpha - \beta x_t\right)^2, \tag{19}$$

where, in our setting, this exercise is closely related to upper-tail trimming because crisis periods typically coincide with unusually high values of x_t.

Quantile regression. For quantile index $\tau \in (0,1)$, the quantile regression estimator solves

$$(\hat{\alpha}_\tau, \hat{\beta}_\tau) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} \rho_\tau\left(r_{t+1}^{ex} - \alpha - \beta x_t\right), \tag{20}$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$ is the standard check-loss function. Quantile regression is less sensitive than least squares to large residuals because the objective function is linear rather than quadratic in the residual. The tradeoff is that quantile regression does not estimate the conditional mean slope β from the baseline model. Instead, it estimates quantile-specific slopes, and should therefore be interpreted as complementary evidence on robustness.

A third approach is robust M-estimation, which replaces the quadratic least-squares loss with a function that grows more slowly in large residuals. Specifically, let

$$(\hat{\alpha}_M, \hat{\beta}_M) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} \rho\left(\frac{r_{t+1}^{ex} - \alpha - \beta x_t}{\hat{s}}\right), \tag{21}$$

where $\rho(\cdot)$ is a robust loss function, such as the Huber or Tukey biweight loss, and $\hat{s}$ is a robust scale estimator. Unlike WLS, robust M-estimation does not require an ex ante model for the conditional variance. Instead, it limits the influence of observations with unusually large realized residuals. In the present application, this allows the data to downweight observations that behave like outliers in the regression itself, without mechanically assigning low weight to all high-volatility states. Robust M-estimation therefore provides a useful middle ground: unlike trimming, it does not discard extreme observations entirely, and unlike WLS, it does not rely on a correctly specified inverse-variance weighting scheme.

3.3.3 Intercept restrictions

A further consideration in the estimation of Equation 5 concerns the intercept. Assuming no arbitrage, the intercept should be zero with any asset since any asset exhibiting no variance should also have no excess return in theory. This observation is potentially powerful since estimating without a constant significantly increases slope precision, which can be a major benefit in this low power environment. However, there are also potential concerns with no constant estimation. Given the benefits and costs, which are elaborated on below, we therefore present estimation results in the main analysis both with and without an intercept.

Motivations for considering zero-intercept restrictions. In theory, expected excess returns on an asset should be approximately zero as the return variance approaches zero, which suggests that the economically relevant specification may satisfy $\alpha = 0$. In our setting, this restriction is also attractive for two practical reasons. First, by definition of it being a

variance, the support of the predictor is strictly positive. This means that the intercept is identified only by extrapolation to the boundary case $x_t = 0$. Second, and more importantly, estimating a free intercept reduces precision in the slope coefficient, which is already difficult to estimate in a low-R^2 environment.

Assuming the model is correctly specified, then the asymptotic variance Equation 11 with a zero intercept restriction becomes,

$$\Omega^{nc}_{OLS} = \frac{E[x_t^2 \sigma^2_{t+1}]}{(E[x_t^2])^2} \tag{22}$$

and, under homoskedasticity,

$$\Omega^{nc}_{OLS} = \frac{\sigma^2_{t+1}}{E[x_t^2]}, \tag{23}$$

which transparently shows why the precision improvement occurs. The numerator with the zero-intercept restriction is the total variation in the regressor, as compared to demeaned variation in the numerator of Equation 12 when the model is also forced to estimate the intercept. In other words, by dropping the intercept, the slope estimator harnesses total variation in the regressor rather than only demeaned variation. The simulation exercises in the next section show how powerful this force can be for more accurate inference on the slope.

Concerns with zero-intercept restrictions. Suppose the correctly specified population model is consistent with Equation 5 in that it is linear and does include a constant. When $x_t > 0$ and $E[r^{ex}_{t+1}] > 0$, a no-constant regression can load part of the "true" intercept mechanically onto the slope coefficient. To see this, suppose the correctly specified population model is consistent with Equation 5 and does include a constant. The no-constant slope estimate is then

$$\hat{\beta}^{nc} = \frac{E[x_t r^{ex}_{t+1}]}{E[x_t^2]} = \beta + \alpha \frac{E[x_t]}{E[x_t^2]}. \tag{24}$$

Equation (24) shows that omitting the intercept shifts the slope upward whenever $\alpha > 0$ and $E[x_t] > 0$. It also provides a useful benchmark to show what the spuriously estimated $\hat{\beta} > 0$ would be even if the true underlying $\beta = 0$. For example, given the statistics summarized in Table 1, we can estimate that $\hat{\beta}^{nc}$ is 1.02 and 0.85 using forecasted variance and risk-neutral variance respectively.

Now suppose the correctly specified population model has a zero intercept, consistent with no-arbitrage arguments outlined earlier, but that the true model is non-linear. Even if we know that the nonlinear curve has to pass through the origin, over the range of empirically relevant observations it is possible that an intercept and linear function is a better

approximation to the true nonlinear function than a linear function going through the origin.

For this reason, in the empirical analysis we report specifications with and without an intercept. These two estimations can be viewed as two extremes of a bias-variance trade off. The no constant specification maximizes precision of the β estimate if the population $\alpha = 0$ and true relation is linear. However, it introduces bias to the β estimate if $\alpha \neq 0$, and can also (potentially) increase bias if the true relationship is non-linear.[13]

3.4 Empirical implementation issues

Given the fragility of regression setting, our baseline analysis uses overlapping 30-day excess returns measured at the daily frequency rather than restricting attention to non-overlapping monthly observations. This approach exploits as much of the available time-series variation as possible and yields estimates that are less sensitive to sampling variation than regressions constructed from a single trading day per month. Results shown in Figure 7 will illustrate the benefit of this approach in practice. Moreover, the independent variation in the overlapping windows also helps to increase power in the low-power setting. That said, overlapping observations mechanically induce dependence across adjacent regression residuals and thus standard errors must be adjusted. We therefore conduct inference using a 30-day block bootstrap, with the blocks the same length as the realized returns to account for the serial dependence introduced by construction with the overlapping-returns.[14]

A separate implementation issue is measurement error when using proxies of physical variance measures as predictor variables. An important econometric distinction between risk-neutral return risk, which is directly recoverable from option prices and is therefore observed in real time, and forward-looking physical return risk is that the latter must instead be estimated. As a result, regressions that use physical risk as the regressor are naturally subject to measurement error, which biases the slope downward. Since physical variance is positive on average, attenuation in the slope also mechanically shifts more of the unconditional mean return into the intercept.

[13] To balance these competing considerations, in unreported results we have also considered intercept-shrinkage estimations that nest both the unrestricted and no-constant specifications. The degree to which one is concerned about bias vs. variance dictates how much shrinkage to enforce.

[14] Issues related to standard error bias with overlapping observations are less severe than previously documented in the literature focused on return predictability settings where valuation ratios are the regressor (Boudoukh et al., 2019, 2022). This is because the predictor variable in our setting is far less persistent than valuation ratios.

3.5 Monte Carlo Analysis

To quantify the finite-sample importance of the issues discussed above, we conduct stylized Monte Carlo simulations calibrated to S&P 500 return data.

Calibration. We consider two data-generating processes:

$$\text{(i) independent regressor:} \qquad r_t = \beta x_t + \sigma_t \varepsilon_t, \tag{25}$$

$$\text{(ii) aligned regressor:} \qquad r_t = \beta \sigma_t^2 + \sigma_t \varepsilon_t, \tag{26}$$

where $\varepsilon_t \sim N(0,1)$ and $\log \sigma_t^2 \sim N(\mu_\sigma, \tau_\sigma^2)$. The parameters μ_σ and τ_σ^2 are chosen to match the empirical mean and variance of monthly realized S&P 500 return variance over 1996–2025, so that $E(\sigma_t^2) = 0.0025$ and $sd(\sigma_t^2) = 0.0043$. This calibration also generates substantial right-skewness in σ_t^2, broadly consistent with the data.

In design (i), x_t is drawn independently from the same lognormal distribution, so $x_t \overset{d}{=} \sigma_t^2$ but $x_t \perp \sigma_t$. In design (ii), $x_t = \sigma_t^2$, so the regressor is fully aligned with conditional return variance by construction. Throughout, we set $\beta = 2$, which implies a population R^2_∞ of 2.8 percent in both designs, as in Table 2. The exercise is intended as a transparent benchmark for the role of alignment, rather than as a full structural model of return dynamics.[15]

Finite-sample fragility under alignment. Figure 3 reports OLS sampling distributions for R^2 and $\hat{\beta}$ across 5,000 Monte Carlo replications with $T = 600$ monthly observations, corresponding to 50 years of data. When $x_t \perp \sigma_t$, the distribution of R^2 is tightly centered around the population value $R^2_\infty = 2.8\%$. When $x_t = \sigma_t^2$, by contrast, the distribution becomes highly right-skewed, with many realizations near zero and or above 10 percent. The same contrast across the two model designs appears in the sampling distribution of $\hat{\beta}$s.

Since the two designs share the same marginal distribution of x_t and the same population R^2, the difference comes entirely from alignment between the regressor and conditional noise variance. Under alignment, the observations with the greatest leverage for identifying the slope are also those with the noisiest returns. As a result, convergence to the population relationship is much slower, and finite-sample OLS estimates become substantially more dispersed and fragile. In the calibrated exercise, only 0.06 percent of estimated slopes are negative in design (i), compared with 10.5 percent in design (ii).

[15] For simplicity, the simulation abstracts from intertemporal dynamics in volatility and focuses only on the cross-state relation between the regressor and conditional return variance.

Figure 3: Fragility of OLS estimation in simulated data.

The figure plots Monte Carlo sampling distributions of model fit (R^2) and the estimated slope coefficient ($\hat{\beta}$) from OLS regressions of $r_t = \beta x_t + \sigma_t u_t$, where $u_t \sim N(0,1)$ and $\log \sigma_t^2 \sim N(\mu_\sigma, \tau_\sigma^2)$ is calibrated to historical S&P 500 realized return volatility. Blue: x_t is drawn independently with $x_t \overset{d}{=} \sigma_t^2$ but $x_t \perp \sigma_t$. Red: $x_t = \sigma_t^2$. Both designs have identical population $R^2_\infty = 2.8\%$ (dashed line), but alignment generates substantially greater finite-sample dispersion in R^2 (left) and $\hat{\beta}$ (right). Sample size $T = 600$.

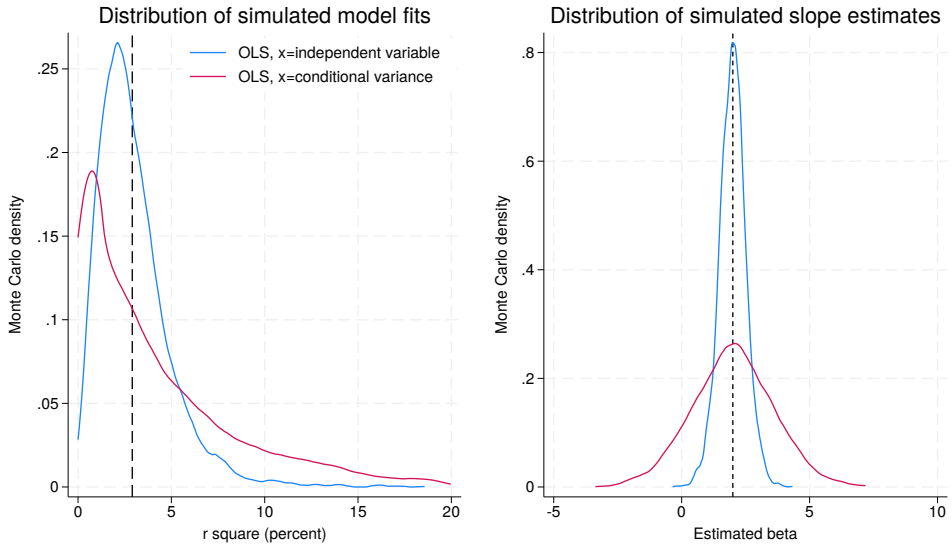
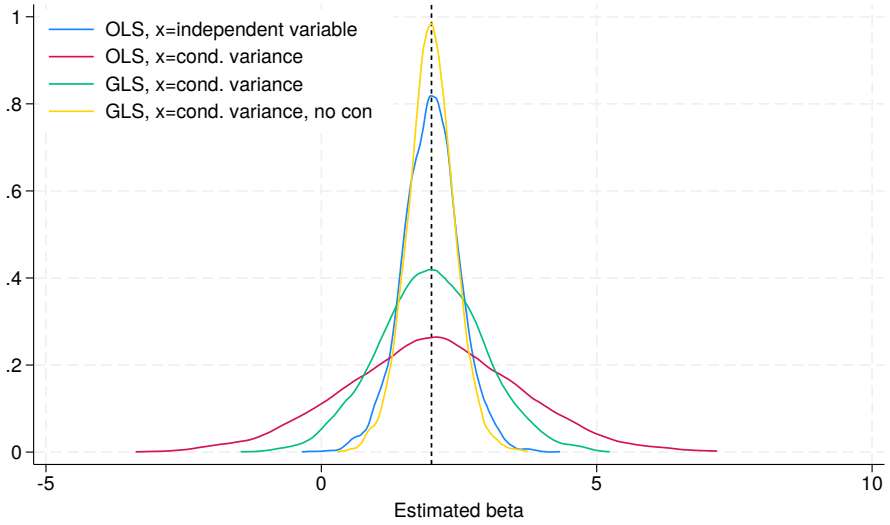


Figure 4: Efficiency gains under GLS in simulated data.

The figure plots Monte Carlo sampling distributions of the estimated slope coefficient from regressions of $r_t = \beta x_t + \sigma_t u_t$, where $u_t \sim N(0,1)$ and $\log \sigma_t^2 \sim N(\mu_\sigma, \tau_\sigma^2)$ is calibrated to historical S&P 500 realized return volatility. The blue curve corresponds to the case where x_t is drawn independently with the same marginal distribution as σ_t^2, while the red curve corresponds to the case where $x_t = \sigma_t^2$; both are estimated by OLS. The yellow and green curves report GLS estimates for the aligned case using oracle weights $w_t = 1/\sigma_t^2$, with and without an intercept, respectively.



Efficiency gains under GLS. Figure 4 shows how much of this dispersion can be reduced when the heteroskedastic structure is taken into account. Relative to OLS in the aligned design, GLS using weights $w_t = 1/\sigma_t^2$ substantially tightens the sampling distribution of $\hat{\beta}$ around the true parameter value. These gains should be interpreted as an upper bound, since the simulation assumes that conditional variance is known rather than estimated. Even so, the figure makes clear that a large part of the finite-sample fragility in OLS arises from the failure to downweight high-variance states.

The gains are larger still when GLS is estimated without a constant, which is the correct specification for the simulated data-generating process. Imposing this restriction avoids estimating a weakly identified intercept in a low-R^2 environment and concentrates precision on the slope coefficient. In fact, in the aligned design the resulting estimator is even more precise than OLS in the independent-regressor case.

4 Main Empirical Results

This section presents the main empirical evidence. The U.S. time series first illustrates the practical importance of regressor-variance alignment: equal-weighted OLS estimates are positive but fragile because a few high-volatility crisis observations dominate slope identification. More efficient and robust estimators reveal a stronger relation between option-implied variance and subsequent returns, and international panel evidence shows that this relation is systematic across both time and countries.

4.1 OLS fragility in practice

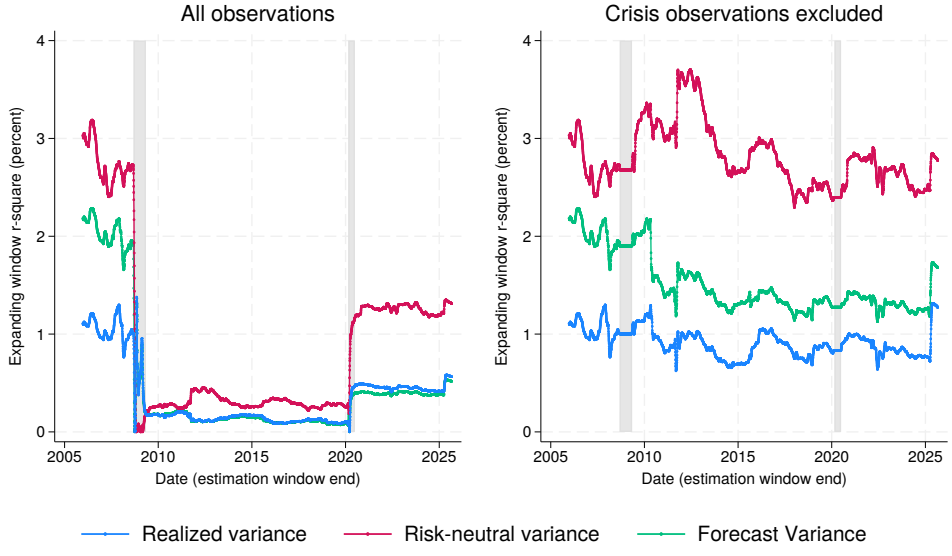
This subsection shows how the fragility of OLS discussed in Section 3 is quantitatively important in practice in S&P 500 data from 1996 to 2025. Figure 5 reports expanding-window OLS model fit for three variance-based predictors: lagged realized variance over the previous month, included as a backward-looking benchmark; risk-neutral variance; and the Corsi (2009) forecast of next month's variance. The shaded areas show the GFC and Covid-19 episodes.

The full-sample results in the left panel reveal substantial instability through crisis periods. For all three predictors, there is economically meaningful model fit prior to the GFC, but each R^2 falls dramatically as the GFC enters the sample, remains low for windows ending between the GFC and Covid, and then rises again when the Covid episode is included. The common pattern across the three predictors illustrates that full-sample OLS estimates are

Figure 5: Model fit from expanding-window OLS regressions on the S&P 500.
This figure plots the model fit (R^2) from expanding-window OLS estimations of

$$r_{t+1}^{ex} = \alpha + \beta x_t + \epsilon_{t+1},$$

where r_{t+1}^{ex} denotes the one-month excess return on the S&P 500 and x_t is the explanatory variable. The three predictors are: (1) realized variance over the previous month (blue), (2) risk-neutral variance (red), and (3) a Corsi (2009) forecast of next month's variance (green). Regressions use daily overlapping observations from January 1996 through the date shown on the x-axis; the last observation is July 30, 2025. The shaded areas indicate the GFC and Covid-19 episodes. The right panel excludes observations in these periods.



highly sensitive to the two short crisis episodes.[16] This sensitivity is somewhat unsurprising given the leverage distribution of the regressors that we discussed in Table 3. Depending on the measure of x_t, observations from the two crisis episodes account for 63 to 80 percent of total regressor leverage while representing less than 3 percent of the sample. Full-sample OLS estimates are therefore expected to be disproportionately shaped by this small set of observations.[17] The left panel of Figure 5 serves to illustrate just how powerful these forces are in practice.

A natural question is how strong is the predictability outside of these crisis periods. To explore this, the right panel of Figure 5 shows expanding window model fit with the extreme, high-leverage episodes removed. Model fit is now remarkably stable across subperiods for all three predictors. The risk-neutral variance measure, with an R^2 of roughly 3 percent over much of the sample delivers the highest fit. This exercise should be interpreted as a diagnostic rather than a preferred sample restriction: its purpose is to isolate the role of the crisis observations that dominate OLS in the full sample. The broader evidence on the risk–return relation, using the full sample and alternative estimators, is developed below.[18]

To illustrate the underlying mechanism more directly, Figure 6 focuses on the onset of the GFC using risk-neutral variance as the predictor. In the sample through early October 2008, the estimated relation between risk-neutral variance and subsequent excess returns is clearly positive and economically meaningful (left panel). As the October 2008 stock market crash observations enter the sample (middle panel, red diamonds), the fitted slope flattens. More importantly, following the crash, from November and early December 2008 the first extreme risk-neutral variance observations enter the sample (right panel, green triangles),. By December 10, the estimated slope is negative and the model R^2 has fallen from 2.6 percent with 3,191 observations to essentially zero with 3,237 observations.

The reason is visible in the figure. The newly added observations lie far to the right of the historical range of x_t and therefore carry enormous leverage in the OLS fit. Because realized returns were negative, the fitted line rotates sharply downward as these observations enter the sample. Without taking a stance on the degree to which these negative returns reflected low ex-ante return expectations or unexpectedly bad ex-post return realizations, the point we with to draw here is that the inclusion of only a few dozen observations, at a time when there was clearly great uncertainty about future realized returns, materially

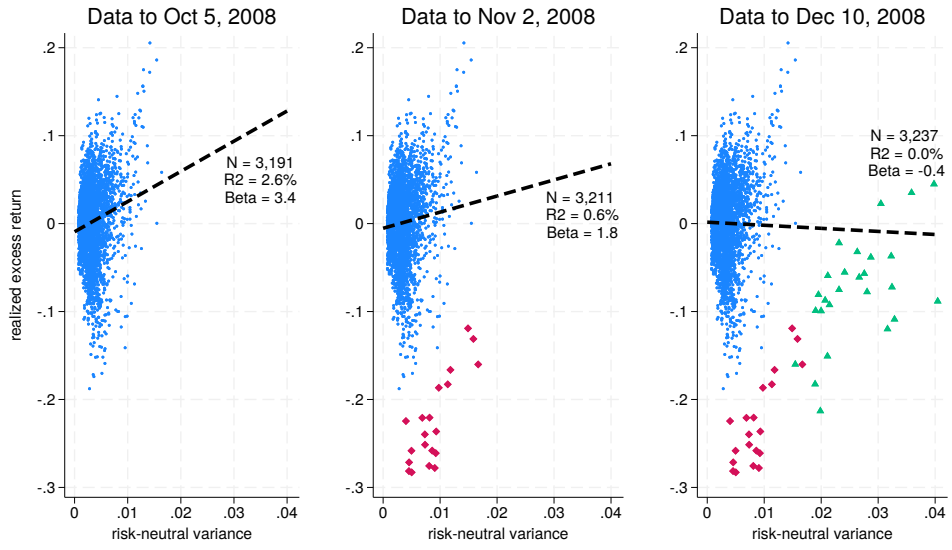
[16] For comparison, Figure A.2 shows model-fit for VRP measures as the regressor, which exhibit the same extreme sensitivity to crisis episodes. However, in contrast to variance measures, VRP measures tend to show an increase in the explanatory power through the GFC but a decline in power through covid.

[17] Because the regression is univariate, instability in model fit and instability in the estimated slope are closely related. Plots of expanding window beta estimates are provided in Figure A.1.

[18] In related work Yang (2026), identifies high and low volatility periods with a regime shifting model and shows that the risk-return relation holds stronger in low volatility periods.

Figure 6: OLS fragility at the onset of the GFC.

This figure plots one-month realized excess returns against one-month risk-neutral variance using daily overlapping observations from January 4, 1996 through return horizons ending October 5, 2008 (left panel), November 2, 2008 (middle panel), and December 10, 2008 (right panel). Each panel reports the OLS line of best fit for the sample available at that date, together with the number of observations, R^2, and the estimated slope coefficient. Observations from January 4, 1996 through October 5, 2008 are shown as blue dots, observations from October 5 to November 2, 2008 as red diamonds, and observations from November 2 to December 10, 2008 as green triangles.



alters full-sample OLS estimates to that point. This is precisely the fragility mechanism emphasized in subsection 3.2.

4.2 Robust (option-implied) risk-return evidence in the U.S.

This section provides the central U.S. evidence of a positive and economically meaningful relation between option-implied variance and subsequent excess returns. The benchmark OLS estimate is positive but only modestly precise. However, once estimation is made less sensitive to the high-variance observations, the estimated slope becomes uniformly larger and more precisely estimated. This pattern appears under inverse-variance weighting, trimming, quantile regression, robust M-estimation, and with estimator precision increased by the zero-intercept specification. Overall, the results are consistent with the view that the positive but relatively weak full-sample OLS result reflects finite-sample fragility rather than an absence of a risk-return relation.

Table 4 reports estimates of Equation 5 using risk-neutral variance as the regressor, based on the alternative estimation methods outlined in subsection 3.3. Column (1) presents the benchmark OLS results for the full 1996–2025 sample. The estimated slope coefficient is 1.60 (s.e.=0.96, p=0.096), with an R^2 of 1.31 percent.[19] Column (2) reports results from feasible GLS (weighted least squares), where the weights are the inverse of forecasted return variance obtained from the parsimonious Corsi (2009) HAR model. Relative to the equal-weighted OLS benchmark, the weighted specification yields a more efficient estimate: the slope rises to 1.82 ($p = 0.031$), indicating a materially stronger and more precisely estimated link between risk-neutral variance and subsequent returns. Not only does the slope estimate rise relative to OLS, but the standard error also shrinks which is consistent with the econometric objective of placing less weight on the high variance observations in order to increase precision.

Columns (3) and (4) report a simple trimming exercise that excludes periods in which the noise in realized returns is exceptionally high. This provides a diagnostic parallel to the WLS approach by effectively assigning a weight of zero to the highest-variance observations. In column (3), the GFC and covid periods are omitted from the sample. Without these episodes, OLS estimation yields a markedly stronger relationship between risk-neutral variance and subsequent returns: the slope coefficient increases to 3.46 and is significant at the 1 percent level, with the R^2 (shown in the right panel of Figure 5) exceeding 3 percent. Column (4) trims based on top 5 percent of x_t realizations as well as the bottom and top 1 percent of

[19]The OLS model R^2 can be seen in the last observation in Figure 5. We do not present R^2 across models in Table 4 because the estimators considered optimize different objective functions. Similarly, we present non-overlapping block bootstrap errors throughout since traditional robust standard errors that adjust for heteroskedasticity and autocorrelation are not easily implementable in some of the alternative econometric specifications used.

Table 4: Risk-neutral variance and returns under a broad range of models.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral variance at time t. Regressions use a sample of daily observations from January 1996 through July 2025. Panel A shows estimation results with a constant and panel B shows estimation results without a constant. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Estimations with a constant

	Dependent Variable: 1-month excess return					
	(1) OLS	(2) WLS	(3) ex.Crisis	(4) Trim	(5) Quantile	(6) Robust
Risk-neutral variance	1.602*	1.817**	3.456***	2.125**	2.055**	2.927***
	(0.964)	(0.863)	(0.904)	(0.936)	(0.981)	(1.036)
Constant	0.000	0.000	-0.004	-0.000	0.006**	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	7443	7443	7179	6998	7443	7443

Panel B: Estimations without a constant

	Dependent Variable: 1-month excess return					
	(1) OLS	(2) WLS	(3) ex.Crisis	(4) Trim	(5) Quantile	(6) Robust
Risk-neutral variance	1.663**	1.877***	2.584***	2.090***	3.293***	3.149***
	(0.648)	(0.577)	(0.580)	(0.577)	(0.671)	(0.658)
Observations	7443	7443	7179	6998	7443	7380

return realizations. The results remain similar, with the estimated slope highly significant at the 1 percent level and an unreported R^2 above 3 percent. Taken together, columns (3) and (4) indicate that the full-sample OLS evidence includes an attenuating effect of extreme realizations and crisis periods. Once these high-variance observations are excluded, the underlying relationship between risk-neutral variance and subsequent returns emerges as both strong and precisely estimated during more typical market conditions. The higher slope coefficient outside of crisis periods also helps to explains why the WLS approach, which naturally down weights these crisis periods, estimates a larger point estimate relative to the OLS approach.

Columns (5)-(6) then show estimation results using a quantile regression at the median (50th percentile) and robust M-estimation. Across these distinct estimation methods, the estimated slope coefficients are systematically larger than the OLS estimator and are significant at at-least the 5 percent level. The results are again consistent with a more robust relationship between risk-neutral variance and realized returns than suggested by the imprecise full sample OLS benchmark.[20]

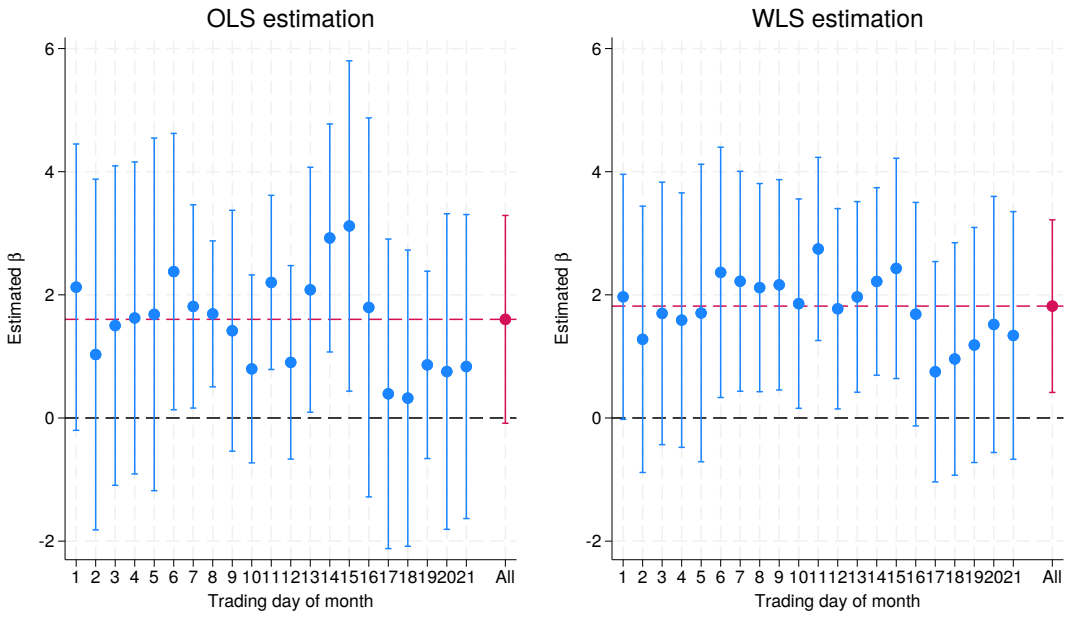
Table 4 Panel B reports estimations of Equation 5 imposing a zero–intercept restriction. As discussed in subsubsection 3.3.3, this restriction can be both theoretically and econometrically motivated. Moreover, in Panel A we observe that the estimated constants are economically negligible and statistically indistinguishable from zero across most specifications. Removing the unnecessary parameter therefore offers a direct means of improving estimation efficiency. Indeed, the estimates in Panel B show that standard errors shrink meaningfully once the intercept is omitted, while the slope coefficients remain similar to their Panel A counterparts. The resulting estimates therefore provide a cleaner and more efficient characterization of the positive relationship between risk–neutral variance and subsequent excess returns.

The results in Table 4 are based on monthly returns constructed from overlapping daily observations, with inference based on block bootstrap standard errors. As discussed in subsection 3.4, this approach uses the full information in the daily data, improving precision while accounting for the induced dependence in returns. A natural question, however, is how much the use of overlapping data matters in practice. Figure 7 addresses this by reporting estimates from predictive regressions constructed from non-overlapping monthly returns using different trading-days of the month. Each point corresponds to a separate regression, where the x-axis indicates the trading day used to form the monthly series.

[20] Applying standard Stambaugh (1999) bias corrections leaves the estimates essentially unchanged. This classic predictive-regression bias is less severe because the regressor is substantially less persistent than valuation ratios.

Figure 7: Non-overlapping monthly regressions, by trade-day of month.

This figure presents beta estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral variance at time t. Each dot shows the the estimated beta for a separate regression that is sampled at the monthly frequency and uses the n'th trading day of the month that is reported on x-axis. Confidence intervals based on heteroskedasticity-robust standard errors (Huber-White) at the 10% significance level are shown with the bars. The left figure is estimated using OLS and the right figure is estimated using WLS. The dotted final regression result and red horizontal line shows the beta estimate using all observations in a single regression with Newey West standard errors. The sample is from January 1996 through to July 2025.



The left panel reports OLS estimates and the right panel reports WLS estimates, with heteroskedasticity-robust confidence intervals.

The dispersion in estimates across these non-overlapping regressions is striking. Under OLS, estimated slope coefficients range from 0.32 to 3.11, with a standard deviation of 0.77 across trade-day estimates. There is similar variability in model-fit, with R^2 values ranging from 0.05 percent to 5.08 percent depending on the selected trading day. Notably, the conventional end-of-month specification yields one of the weakest estimates in the figure, with a slope of 0.84 and an R^2 of only 0.29 percent. Thus, in this setting, the apparent strength of the risk-return relation can depend materially on an essentially arbitrary calendar choice.

WLS partly stabilizes the estimates, narrowing the slope range to 0.75–2.75 and reducing the cross-day standard deviation of $\hat{\beta}$ by roughly 35 percent, to 0.50. This is consistent with the efficiency gains from down-weighting high-variance observations. Even so, substantial variation remains across trading-day alignments. Taken together, these results show that non-overlapping monthly regressions are even more fragile in finite samples, and that using overlapping observations is valuable not only for precision, but also for reducing sensitivity to arbitrary sampling noise.

4.3 New global evidence on risk-return relations

This section shows that the option-implied risk-return relation is not confined to the U.S. market. In a panel of ten major equity indices, higher risk-neutral variance predicts higher subsequent excess returns both when identification comes from pooled variation across countries over time and when it comes from cross-country differences within date after absorbing common global shocks through time fixed effects. The international evidence therefore strengthens the time-series case and, in addition, provides novel evidence of a cross-sectional risk-return relation across equity markets.

Figure 8 shows the evolution of risk-neutral variance across the international markets. Volatility spikes are strongly synchronized across countries—most notably during the GFC and the onset of Covid-19—but there is also useful cross-sectional dispersion across countries that will provide the basis for the empirical identification in this international analysis.

Table 5 reports panel-regression estimates of:

$$r^{ex}_{i,t+1} = \alpha + \beta x_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t+1}, \tag{27}$$

where $r^{ex}_{i,t+1}$ is the one-month excess return on index i, $x_{i,t}$ is the corresponding risk-neutral variance, and γ_i and γ_t denote index and time fixed effects. Standard errors are clustered by

Figure 8: Risk-neutral variance across international equity markets.

This figure shows estimates of the risk-neutral one-month variance of returns on a range of international stock markets indexes. Observations are daily and the sample is January 2, 2002 through to May 31, 2022.

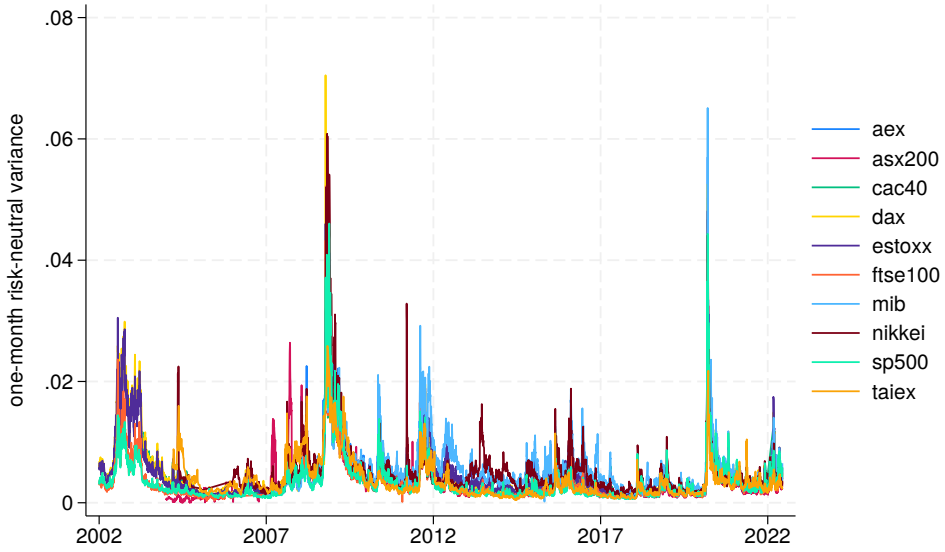


Table 5: Risk-neutral variance and returns across international stock markets.

This table presents estimates from the regression: $r^{ex}_{i,t+1} = \alpha + \beta x_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t+1}$, where $r^{ex}_{i,t+1}$ is the excess return of international stock market i over a one-month period from t to $t+1$, $x_{i,t}$ is the one-month risk-neutral variance for index i at time t, and γ_i and γ_t are index and time fixed effects. Regressions use a sample of daily observations from January 2002 to June 2022. Standard errors clustered by date and index are reported in parenthesis.

	Dependent Variable: 1-month realized excess return							
	OLS				WLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk-neutral variance	1.09***	1.16***	1.28***	1.82***	1.50***	1.79***	1.03*	1.90***
	(0.24)	(0.26)	(0.34)	(0.30)	(0.31)	(0.34)	(0.48)	(0.42)
Constant	-0.00				-0.00			
	(0.00)				(0.00)			
Index FE	no	yes	no	yes	no	yes	no	yes
Date FE	no	no	yes	yes	no	no	yes	yes
Adj R^2 (%)	0.73	0.92	77.92	78.18	0.46	0.99	74.05	70.49
Adj R^2 (Within, %)		0.80	0.78	1.25		0.62	0.26	0.68
Observations	45,207	45,207	43,717	43,717	45,180	45,180	43,690	43,690

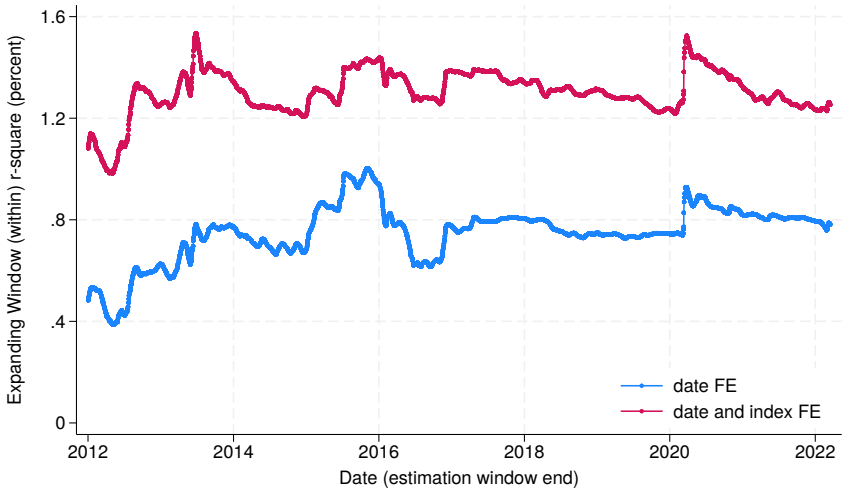
both index and date.

Across all specifications, the estimated slope coefficients are positive, economically large, and statistically significant. Under OLS, the estimated β ranges from 1.09 to 1.82 across mod-

Figure 9: Model fit from expanding window international regressions.
This figure plots the model fit beyond the variation absorbed through fixed effects (i.e., within R^2) from expanding window OLS estimations of:

$$r^{ex}_{i,t+1} = \alpha + \beta x_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t+1},$$

where $r^{ex}_{i,t+1}$ is the excess return of international stock market i over a one-month period from t to $t+1$, $x_{i,t}$ is the one-month risk-neutral variance for index i at time t, and γ_i and γ_t are index and time fixed effects. Regressions use a sample of daily observations from January 2002 through to the date shown on the x-axis. The last observation is May 31, 2022.



els (columns 1–4), with significance at the 1 percent level in each case. Controlling for both index and date fixed effects modestly reduces the point estimates but leaves inference unaffected. Weighted least squares (columns 5–8) produces slightly larger coefficients—ranging from 1.50 to 1.79—and comparable significance. These results indicate that the positive association between risk-neutral variance and subsequent returns holds systematically across markets, regardless of estimation method or inclusion of fixed effects. The statistical significance under OLS with index fixed effects (column 2) underscores how pooling information across countries substantially improves the precision and reliability of inference even in time series tests.

The date fixed-effects specifications in Table 5 also provide novel evidence of cross-country return differences related to risk-neutral variances. The results imply that, on average, markets with higher contemporaneous risk-neutral variance earn higher subsequent excess returns relative to other markets in the same period. Thus, the option-implied risk–return relationship is not confined to a single market's time series but is observable in the global cross-section of equity market returns as well.

Finally, Figure 9 evaluates the stability of the relationships over time using expanding-window regressions within the same international panel framework. Focusing on specifica-

tions with date fixed effects, and thus exploring the cross sectional relationship over time, the within-R^2 of remains relatively stable across the sample period. This stability is in marked contrast to the instability observed in the U.S. time series estimation under OLS (see Figure 5). Overall, the international evidence reveals a robust and temporally stable link between option-implied variance and realized returns for stock market indices. The consistency of results across estimation methods, time periods, and markets provides strong support for a systematic, globally pervasive risk–return relationship for stock market indices that is identifiable through the information in equity option prices.

5 Further results and discussion

This section interprets and extends the main risk-return evidence in three directions. I first show that, although the variance risk premium is informative, risk-neutral variance itself largely spans the return-predictive information contained in standard decompositions. I then extend the evidence back to 1885 using a synthetic risk-neutral variance proxy and use option-implied higher moments to connect the estimated slope magnitudes to the economic interpretation of the risk-return tradeoff.

5.1 The role of the variance risk premium

Risk-neutral variance can be decomposed into investors' expectation of the physical variance plus the variance risk premium (VRP),

$$\mathrm{Var}_t^* \left(R_{t+1}\right) = \mathrm{Var}_t \left(R_{t+1}\right) + \left(\underbrace{\mathrm{Var}_t^* \left(R_{t+1}\right) - \mathrm{Var}_t \left(R_{t+1}\right)}_{VRP_t} \right),$$

where the VRP is the difference between the risk-neutral and physical variance over the same horizon (Carr and Wu, 2009). Natural questions are therefore the extent to which each component of risk-neutral variance drives the variable's return predictability (Bekaert and Hoerova, 2014), and whether the separation of the risk-neutral variance into its components can increase the return predictability.

Table A.2 presents the same estimation results as in Table 4 but with last month's realized return variance used as the regressor. Across specifications, results are qualitatively similar to predicting returns with risk-neutral variance. However, quantitatively, the estimated slope coefficients are consistently lower. Taken at face value, the results suggest that the physical variance component of risk-neutral variance contributes to some, but not all, of the risk-

neutral variance predictability documented in section 4. However, a well-known challenge to this interpretation is that investors expectation of physical variance is unobserved to the econometrician (Zhou, 2018). In its place a proxy must be used, whether that be the backward looking past realized variance presented (eg. Bollerslev et al. (2009)) or a forward-looking forecast of variance (eg. Drechsler and Yaron (2011)). The former has the benefit of being directly observable, while the latter is likely more accurate but at the cost of being more subjective and also sample and model dependent.

To explore further, we re-estimate Equation 5 with a range of different regressors that are the components of risk-neutral variance under different proxies of physical variance. Most estimations are with simple OLS. However, since our proxies of investor's physical variance expectations are almost surely miss-measured, we also consider instrumental variables (IV) estimations with the lagged regressor used as an instrument. With measurement error, instrumented estimates should not only increase the estimated slope parameter, but also decrease the estimated intercept.[21]

Table 6 presents the results. Column (1) shows baseline estimates with risk-neutral variance with the remaining columns showing the predictability of its components, with columns (2)-(5) presenting results where physical variance is proxied using realized variance over the last month, and columns (6)-(9) presenting results where physical variance is proxied using the Corsi (2009) forecast of return variance over the next month. Panel A uses the full sample of data from 1996 and panel B excludes crisis episodes.

Panel A column (1) first shows the benchmark regression using risk-neutral variance delivers an estimated slope of 1.60 and an R^2 of 1.30 percent. By contrast, replacing risk-neutral variance with lagged realized variance in column (2) lowers the slope to 0.88 and reduces the R^2 to 0.56 percent. The intercept also becomes more positive, consistent with the measurement-error logic discussed above: if the physical variance proxy is a noisy measure of the true forward-looking variance, the slope is attenuated and more of the unconditional mean return is absorbed by the intercept. Column (3) provides direct evidence on this interpretation by instrumenting lagged realized variance with its own lag. The IV estimate raises the slope from 0.88 to 1.46 and lowers the intercept, consistent with attenuation bias in the OLS estimate. However, the standard error also increases and there is no improvement in model fit. This illustrates the basic tradeoff in using lagged instruments in this setting. The IV estimator reduces measurement-error bias by using only the persistent component of realized variance, but this also discards much of the high-frequency variation in x_t and therefore further reduces power in an already low-R^2 regression.

[21] Since physical variance is positive on average, attenuation in the slope will mechanically shift more of the unconditional mean return into the intercept in the OLS regressions.

Columns (4) and (5) then examine the variance risk premium constructed using lagged realized variance as the physical-variance proxy.[22] In isolation, this VRP proxy has little explanatory power: the slope in column (4) is positive but imprecisely estimated, and the R^2 is only 0.12 percent. When lagged realized variance and the corresponding VRP are included together in column (5), the VRP coefficient becomes significant and the R^2 rises to 1.42 percent. However, this increase relative to the simple risk-neutral variance benchmark in column (1) is small in economic terms. Decomposing risk-neutral variance into lagged realized variance and a residual option-price component raises the R^2 by only about 0.1 percentage points.

Columns (6)–(9) repeat the exercise using the Corsi (2009) forecast of future variance as the proxy for physical variance. The results are similar, although somewhat more favorable to the VRP component. The forecasted physical variance alone in column (6) has modest explanatory power, and the IV estimate in column (7) again raises the point estimate but at the cost of substantially lower precision. The corresponding VRP, by contrast, is statistically significant both on its own in column (8) and jointly with the physical-variance forecast in column (9). This is consistent with the variance risk premium containing economically meaningful forward-looking information on bad uncertainty and downside economic risk that requires risk compensation (Bekaert and Engstrom, 2017; Bekaert et al., 2023). Even so, the combined specification in column (9) raises the R^2 only to 1.6 percent, compared with 1.3 percent for risk-neutral variance alone. Thus, even with a more forward-looking proxy for physical variance, the decomposition produces only a modest improvement over the model-free risk-neutral variance benchmark.[23] In this sense, the conclusion is different to Bekaert and Hoerova (2014), as I do not find it necessary to isolate the variance risk premium component in order to generate return predictability from option prices.

Panel B shows that excluding the crisis episode observations produces substantially stronger and more stable estimates across the full range of predictors considered. The risk-neutral variance slope in column (1) rises to 3.46 and is significant at the 1 percent level, with an R^2 of 2.76 percent. The physical-variance proxies also perform better outside of crisis periods. Lagged realized variance in column (2) and forecasted variance in column (6) both have positive and significant coefficients, and their model fit increases relative to

[22] We do not report IV estimates for the variance risk premium because the variance risk premium is substantially less persistent than variance itself, which weakens the first stage and further reduces power.

[23] The exercise here is not exhaustive here as I do not review a full set of variance forecasting models (as in Bekaert and Hoerova (2014) and Bekaert et al. (2025). It is possible that more accurately capturing the true physical variance expectations of investors might provide enhanced return predictability from the variance risk premium component. However, this is not the focus of my analysis, which is instead focused on risk-neutral variance that is a model-free measure of market risk that is directly observable from option prices.

the full sample. This is consistent with the broader evidence that the risk-return relation is easier to estimate once the most noisy, high-leverage observations are removed.

The IV results in Panel B also suggest that measurement error in the physical-variance proxies is not the main issue outside the crisis episodes. For lagged realized variance, the IV slope in column (3) is almost identical to the OLS slope in column (2). For forecasted variance, the IV slope in column (7) is higher than the OLS estimate, but the standard error is much larger and the estimate is not statistically significant. A natural interpretation is that, in calmer periods, variance is more predictable and the physical-variance proxies are less severely contaminated by the large measurement errors that arise during crises and have high leverage on estimated parameters. Finally, the decompositions in Panel B show that, while the VRP component remains informative, the decomposition of risk-neutral variance does not improve meaningfully on using just risk-neutral variance itself outside of crisis periods.

5.2 The risk-return relation since 1885

The option-based evidence raises a natural question: what can be learned about the risk-return relation over a much longer historical sample? In this subsection, we use the lessons from the option-based sample to study the risk-return relation in U.S. equity data back to 1885.

Approach. Risk-neutral variance is not observable over the full historical period, so we construct a synthetic risk-neutral variance measure from realized-variance data. Specifically, over the option sample we estimate the relation between observed risk-neutral variance and realized variance measured over the past month, week, and daily horizons:

$$\text{Var}_t^* = \underbrace{\alpha}_{0.00192} + \underbrace{\beta_1}_{0.328} RV_t^{month} + \underbrace{\beta_2}_{0.578} RV_t^{week} + \underbrace{0.\beta_3}_{0.401} RV_t^{day} + \epsilon_t. \tag{28}$$

which replicates the Corsi (2009) model shown in Equation 3 but replaces next month's realized variance with the contemporaneous risk-neutral variance as the dependent variable. This simple and parsimonious projection of realized returns onto risk-neutral expectations has strong explanatory power, with an R^2 of 74 percent. We then apply the fitted relation in Equation 28 to the historical realized-variance data to obtain a synthetic risk-neutral variance series, $\text{Var}_t^{*,\text{syn}} = \widehat{\text{Var}}_t^{*,}$ back to 1885.

Limitations. This synthetic approach is useful because it maps historical realized-variance information into a proxy of the same object studied in the main analysis: the market's

Table 6: Decomposing risk-neutral variance and the role of variance risk premium.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral or physical variance at time t, or the estimated one-month variance risk premium at time t. Regressions use a sample of daily observations from January 1996 through July 2025. Panel A shows estimation results with the full sample and panel B shows estimation results with the sample excluding crisis periods. Columns (2) to (5) measure physical variance as the realized variance over the past month. Columns (6) to (9) measure physical variance as the Corsi (2009) forecast of variance for the next month. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Full sample

	Dependent Variable: 1-month realized excess return								
		Realized Variance				Forecasted Variance			
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) OLS	(7) IV	(8) OLS	(9) OLS
Risk-neutral variance	1.602*								
	(0.964)								
Physical variance		0.877	1.460		1.608	1.199	2.327		1.066
		(0.850)	(1.167)		(0.982)	(1.335)	(1.846)		(1.445)
Variance risk premium				0.768	2.404*			3.025**	2.898**
				(1.617)	(1.460)			(1.459)	(1.460)
Constant	0.000	0.004	0.002	0.005*	-0.000	0.003	0.001	0.002	0.000
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Observations	7443	7443	7422	7443	7443	7443	7422	7443	7443
R^2	1.297	0.555	0.336	0.117	1.423	0.483	0.071	1.208	1.585

Panel B: Excluding crisis episodes

	Dependent Variable: 1-month realized excess return								
		Realized Variance				Forecasted Variance			
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) OLS	(7) IV	(8) OLS	(9) OLS
Risk-neutral variance	3.456***								
	(0.904)								
Physical variance		2.831**	2.808		3.270***	4.402***	5.268		3.664***
		(1.111)	(2.012)		(1.139)	(1.280)	(3.703)		(1.301)
Variance risk premium				3.153**	3.825***			4.057***	3.253**
				(1.584)	(1.390)			(1.429)	(1.271)
Constant	-0.004	0.001	0.001	0.003	-0.004	-0.002	-0.004	0.002	-0.004
	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.003)	(0.007)	(0.002)	(0.003)
Observations	7179	7179	7158	7179	7179	7179	7158	7179	7179
R^2	2.760	1.385	1.447	0.958	2.768	1.795	1.789	1.575	2.753

forward-looking risk-neutral assessment. However, it also a lower-variance and incomplete proxy for observed risk-neutral variance, and its construction assumes that the modern relation between realized variance and risk-neutral variance is informative historically. Even in the option sample, the incomplete proxy limitation is quantitatively important. Despite the first-stage projection explaining a large fraction (74 percent) of the in-sample variation in observed risk-neutral variance, the synthetic measure delivers an R^2 of 0.4 percent in the second stage return predictability regression, compared with 1.3 percent for observed risk-neutral variance (see Table 6 column 1). This decline is larger than the mechanical loss implied by the first-stage fit: if the first-stage residual were not differentially priced, the expected R^2 would be approximately $0.7426 \times 1.3 = 1.0$ percent. The slope also falls from 1.59 to 1.03, while the intercept rises. These results suggest that the component of risk-neutral variance not spanned by realized variance contains disproportionate return-predictive information. The long-sample results should therefore be interpreted as a conservative test using an incomplete historical proxy for option-implied risk.[24]

Results. Figure 10 Panel A plots realized one-month return variance for U.S. equities from 1885 to 2025, with periods of extreme volatility highlighted in gray. The GFC and Covid-19 episodes, which played a central role in the option-based sample, are two of only a handful of episodes in the full historical record in which realized variance reaches comparable levels. The other major episodes are the Great Depression and the October 1987 crash. Among these, the Great Depression stands out as the only sustained period of exceptionally elevated return volatility, with Schwert (1989) emphasizing that this period of volatility is hard to explain with simple valuation or macroeconomic models.

Panel B reports expanding-window OLS estimates using synthetic risk-neutral variance as the predictor. The first point in the figure uses the first 50 years of data. In this initial sample, the estimated risk-return relation is positive, economically large, and statistically significant, with a slope coefficient of approximately 3.5 and R^2 of 1 percent. However, the figure also shows substantial sensitivity to extreme volatility episodes. In particular, the inclusion of the Great Depression causes a sharp decline in both model fit and the estimated slope when all observations are used (blue lines). This dramatic and out-sized influence on estimation of periods of elevated volatility mirror the role played by the GFC and Covid-19 in the option-based sample, but on the extended historical sample.

The red lines in Panel B repeat the expanding-window exercise after excluding crisis observations. The resulting estimates are much more stable. Model fit remains economically

[24]We have explored improving the risk-neutral variance forecast using lagged realized return variance and random fourier features machine learning techniques, but our findings to date suggest that the simple three linear factor model of Corsi (2009) is very hard to beat out-of-sample.

meaningful throughout the sample, and the estimated slope coefficient stays consistently positive and large. As before, the purpose of this exclusion is twofold. First, it shows the same fragility mechanism identified in the option sample is also present in the long historical data. Crisis observations contain important information about time-varying risk, but they also carry exceptionally noisy realized returns and therefore exert disproportionate influence on equal-weighted OLS estimates. Second, it shows that the strong and robust relationship between risk and returns outside of the extremely rare volatility spikes, first shown in the relatively short options-based sample, has been present for the past 140 years.

Table 7 reports formal estimates over two long samples. Panel A begins in 1941, after the Great Depression, while Panel B uses the full sample beginning in 1885. The table reports OLS, WLS, and crisis-exclusion specifications, both with and without an intercept.[25] The WLS specifications follow the logic of the main empirical analysis by placing less weight on observations with higher forecasted return variance.

The unrestricted estimates are positive across all specifications, although equal-weighted OLS is imprecise. In the post-1941 sample, the OLS slope is 1.06, while WLS raises the estimate to 1.99 and makes it statistically significant. In the full sample, the OLS slope is 1.28 and the WLS estimate is 1.36, significant at the 10 percent level. These estimates should be interpreted in light of the generated-regressor problem: synthetic risk-neutral variance almost surely misses some of the true risk-neutral variance predictability, and this measurement error should also bias the slope toward zero. Overall, the unrestricted specifications point to a positive relation, but also show that measurement error and crisis-driven noise make inference difficult.

The evidence strengthens substantially when the same crisis-fragility lesson from the option sample is applied. Excluding crisis observations raises the slope to 3.17 in the post-1941 sample and 3.55 in the full sample, with both estimates statistically significant at conventional levels. These magnitudes are close to those estimated in the option sample outside the GFC and Covid-19 episodes. The no-constant specifications provide additional support. As discussed in subsubsection 3.3.3, the zero-intercept restriction is economically natural if expected excess returns approach zero as return variance approaches zero, and it can substantially improve precision in low-R^2 settings. In both long samples, the no-constant estimates are positive and highly statistically significant across every specification. The full-sample estimates range from 1.54 under OLS to 2.43 when crisis observations are excluded, while the post-1941 estimates range from 1.62 to 2.45. The stability of these estimates is

[25] Instrumental-variable estimates, which instrument synthetic risk-neutral variance with its own lag, are reported in the appendix. They provide additional evidence consistent with measurement error in the historical proxy (with increases in beta estimates and decreases in alpha estimates) but also have larger standard errors because they rely only on the persistent component of the generated regressor.

notable given that the regressor is a synthetic proxy for risk-neutral variance rather than an option-implied measure observed directly from market prices.

Taken together, the long-sample evidence supports the main conclusions from the option-based analysis. Because synthetic risk-neutral variance is generated from historical realized-variance data, these estimates should be interpreted as conservative evidence on the risk-return relation. In the option sample, replacing observed risk-neutral variance with realized-variance proxies substantially reduces explanatory power. It is therefore unsurprising that the long-sample estimates are weaker under equal-weighted OLS. Nevertheless, once the same econometric issues are addressed—measurement error, heteroskedasticity, crisis leverage, and the weakly identified intercept—the historical data suggests a positive and economically meaningful relation between return risk and subsequent excess returns over nearly 140 years of U.S. equity market history.

5.3 Higher-order moments and economic interpretation

Analysis thus far has focused on the existence and robustness of the statistical relationship between option-implied risk and realized returns. However, as described below, the magnitude of the estimated slope coefficient, β, is also economically informative.

Theoretical Framework. Under no arbitrage, the expected return on the market is:

$$\mathbb{E}_t[R_{t+1}] - R^f_{t+1} = -R^f_{t+1} \operatorname{Cov}_t \left(M_{t+1}, R_{t+1}\right). \tag{29}$$

where M_{t+1} is a market pricing kernel. Define the risk-neutral measure $*$ in the usual way,

$$\mathbb{E}^*_t[X_{t+1}] \equiv R^f_{t+1}\mathbb{E}_t[M_{t+1}X_{t+1}], \tag{30}$$

then $\mathbb{E}^*_t[R_{t+1}] = R^f_{t+1}$, and the expected excess return on the market can also be expressed

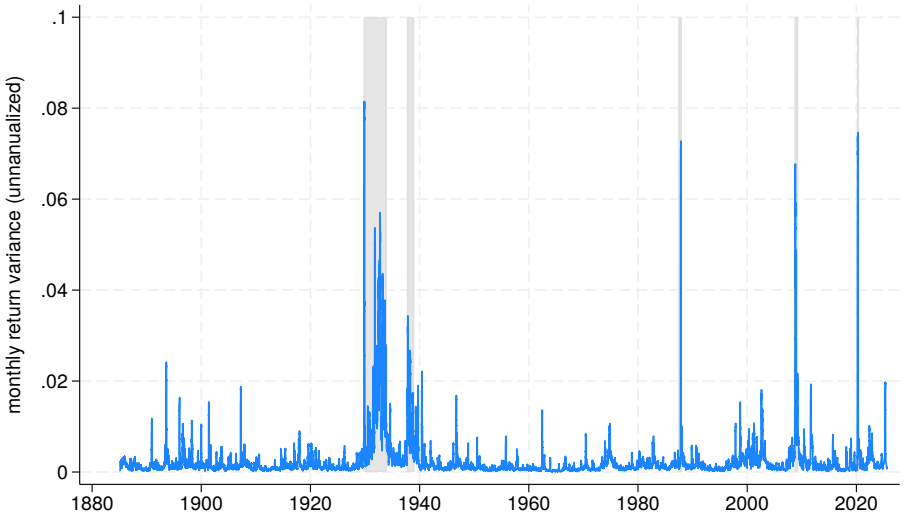
$$\mathbb{E}_t[R_{t+1}] - R^f_{t+1} = \frac{1}{R^f_{t+1}} \operatorname{Cov}^*_t \left(\frac{1}{M_{t+1}}, R_{t+1}\right). \tag{31}$$

The theoretical benchmark of Martin (2017) provides a clear point of reference: when a marginal investor is unconstrained, rational, has log utility and is fully invested in the market portfolio, their stochastic discount factor is $M_{t+1} = 1/R_{t+1}$. Substituting this into Equation 31, and the perceived expected excess return on the market is the risk-neutral

Figure 10: Extending the time series back to 1885.

This figure shows the realized variance of one-month returns on US equities from January 1885 to July 2025. The data is plotted at a daily frequency. The shaded areas indicate the the great depression 1929-1933 and 1937-1938, the flash crash of 1987, the global financial crisis and Covid-19, which are the only periods in the sample during which realized variance exceeds 0.06.

Panel A: Realized monthly variance of market returns



Panel B: Expanding window OLS estimation results

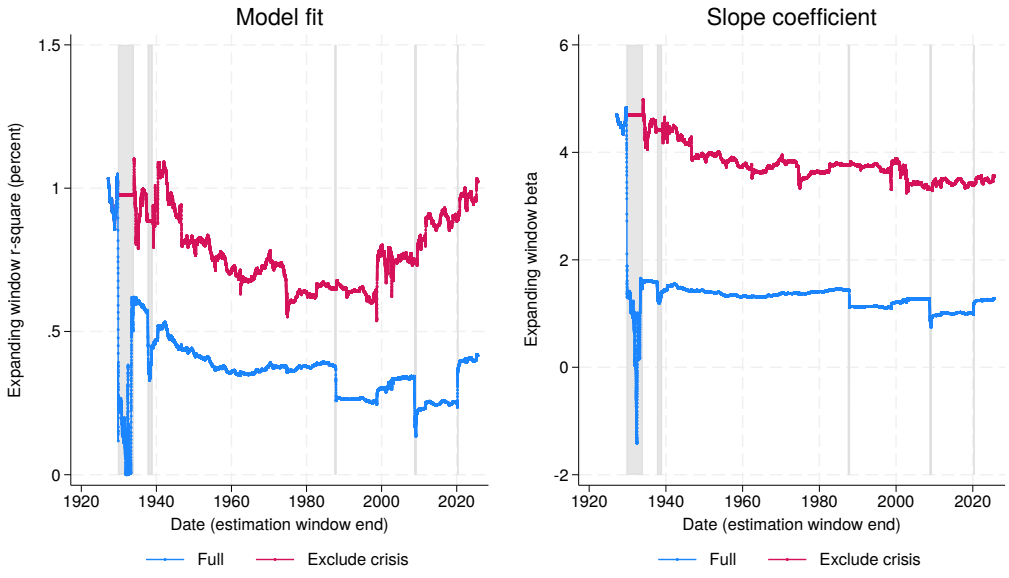


Table 7: Synthetic risk-neutral variance and realized returns since 1885.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral variance at time t. Regressions use a sample of daily observations from January 1941 (Panel A) or January 1885 (Panel B) through July 2025. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Risk-Return Relations since 1941

	Dependent Variable: 1-month realized excess return					
	Unrestricted estimation			No constant estimation		
	(1) OLS	(2) WLS	(3) ex-crisis	(4) OLS	(5) WLS	(6) ex-crisis
Syn. risk-neutral variance	1.064	1.993**	3.174***	1.618***	2.075***	2.448***
	(0.806)	(0.789)	(0.848)	(0.453)	(0.376)	(0.371)
Constant	0.003	0.000	-0.002			
	(0.002)	(0.002)	(0.002)			
Observations	21733	21733	21319	21733	21733	21319

Panel B: Risk-Return Relations since 1885

	Dependent Variable: 1-month realized excess return					
	Unrestricted estimation			No constant estimation		
	(1) OLS	(2) WLS	(3) ex-crisis	(4) OLS	(5) WLS	(6) ex-crisis
Syn. risk-neutral variance	1.275	1.358*	3.551***	1.537***	1.824***	2.427***
	(0.803)	(0.766)	(0.671)	(0.455)	(0.316)	(0.313)
Constant	0.001	0.001	-0.004**			
	(0.002)	(0.002)	(0.002)			
Observations	38345	38345	36278	38345	38345	36278

variance discounted by the risk-free rate,

$$E_t\left[R_{t+1} - R^f_{t+1}\right] = \frac{1}{R^f_{t+1}} Var^*_t\left(R_{t+1}\right), \tag{32}$$

which is also a lower bound of the equity premium under the much weaker assumption that the return on the market is negative correlated. In the case where the lower bound is a point estimate of the equity premium, then the prediction from the model is that the estimated coefficients from Equation 5 would be $\alpha = 0$ and $\beta = 1$.

In results presented to date using risk-neutral variance as the regressor, the estimated intercept has been consistently close to 0 but the estimated betas have are all greater than one. These findings suggest that observed returns are more sensitive to option-implied variance than predicted under log utility (as also argued by Back et al. (2022)). One simple and natural interpretation is that investors exhibit higher risk aversion than log utility, and thus higher-order risk-neutral moments also contribute to the expected returns on the market (Chabi-Yo and Loudis, 2020).

As shown in Martin (2025), if there is a rational investor fully invested in the market who has power utility with risk-aversion coefficient $\gamma > 1$, then there exists a market pricing kernel takes the form $M_{t+1} = \lambda_t R^{-\gamma}_{t+1}$. This leads to a perceived expected return on the market that depends on the higher-order risk-neutral moments of returns:

$$E_t\left[R_{t+1} - R^f_{t+1}\right] = E^*_t\left[R^{1+\gamma}_{t+1}\right] - R^f_{t+1} E^*_t[R^\gamma_{t+1}] \tag{33}$$

where higher order moments can be readily computed from option prices via the Breeden and Litzenberger (1978) approach using a generalization of Equation 4.

Results. Table 8 reports return predictability regression estimates incorporating these higher-moment measures in the regressor variable. Columns (1)–(3) show OLS results and columns (4)–(6) present the corresponding WLS estimates. The measures of risk-neutral risk use the definition in Equation 33 and set gamma as specified in column title. Slope coefficients decline monotonically with higher γ, and under the preferred WLS specification the beta is extremely close to 1 with $\gamma = 2$. The relationship between option prices and realized returns is therefore theoretically consistent with a representative investor with risk aversion $\gamma \approx 2$.[26]

[26]The average option-implied expected excess excess return assuming risk-aversion coefficients of 1 and 2 were 0.34 percent and 0.64 percent respectively (1996-2025). The average realized excess return on the market over the same period was 0.58 percent per month. The risk-aversion parameter of 2 therefore approximately matches the level of realized return. While Martin (2017) found that the lower bound option-implied equity premium (assuming log utility) was approximately close to average returns, that was to some extent driven

Table 8: Higher moment risk and U.S. stock market returns.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral risk at time t. Risk-neutral risk is defined in Equation 33 and uses gamma in the column title. Regressions use a sample of daily observations from January 1990 through to July 2025. Columns (1)-(3) are estimated with OLS and columns (4)-(6) are estimated with WLS. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	Dependent Variable: 1-month realized excess return					
	OLS			WLS		
	(Gamma=1)	(Gamma=2)	(Gamma=3)	(Gamma=1)	(Gamma=2)	(Gamma=3)
	(1)	(2)	(3)	(4)	(5)	(6)
risk-neutral risk	1.620	0.885*	0.632*	1.845**	0.975**	0.678**
	(0.994)	(0.538)	(0.384)	(0.835)	(0.450)	(0.320)
Constant	0.000	0.000	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Adjusted R^2 (%)	1.32	1.33	1.32	2.01	1.99	1.97
Observations	7422	7422	7422	7422	7422	7422

Beyond examining higher-order moments, an additional robustness check in Appendix Table D.5 considers the use of the VIX^2 as an alternative measure of option-implied risk. The VIX^2 represents the risk-neutral entropy of market returns and is therefore closely related to the risk-neutral variance measure used throughout the main body of the paper. In practice, it is computed from option prices with a formula close to Equation 4, but with relative more weight on left tail options. This increases the level and volatility, with the wedge between entropy and variance increasing in crisis periods. However, despite these differences, estimation of Equation 5 using VIX^2 as the regressor produces a model-fit that is almost identical as with risk-neutral variance, with a higher beta that reflects the more volatile VIX^2. So, while there is economically interesting differences between entropy and variance, it appears that the empirical wedge is not meaningful for the purpose of predicting future realized returns.

6 Conclusion

Is there an empirical relationship between risk and returns in aggregate stock markets? This paper revisits this fundamental question in asset pricing and argues that the challenges in

by a relatively short sample period that included GFC period. This period of elevated return variance therefore likely raised the in-sample implied lower bound of the expected return relative to historical norms.

answering it are larger than a simple problem of low statistical power. The key econometric message is that inference in risk–return regressions is fragile in finite samples because variance-based predictors receive their greatest leverage precisely when realized returns are most volatile. This regressor-variance alignment makes conventional OLS estimates highly sensitive to a small number of especially noisy observations.

This statistical challenge motivates the use of alternative methods and data. When risk is measured using option markets, estimation is designed to account for the heteroskedastic structure of returns, and the evidence is expanded beyond the U.S. time series to international equity markets, the risk–return relation becomes substantially clearer. Risk-neutral variance positively forecasts subsequent equity-market returns within countries over time and helps explain differences in realized returns across countries. A synthetic risk-neutral variance proxy further extends the U.S. evidence back to 1885, showing that this relation is not unique to the modern option-market period.

Overall, the results reveal a stronger, more stable, and more systematic relationship between market risk and realized stock market returns than suggested by conventional OLS time-series regressions using estimated proxies of investors' forward-looking perceptions of risk. The findings therefore support the view that expected compensation for bearing aggregate equity-market risk varies meaningfully over time and across markets, while also highlighting the importance of measurement, estimator choice, and sample construction for detecting this fundamental relation in a challenging statistical setting.

References

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(4):813–841.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76.

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

Back, K., Crotty, K., and Kazempour, S. M. (2022). Validity, tightness, and forecasting power of risk premium bounds. *Journal of Financial Economics*, 144(3):732–760.

Baillie, R. T. and DeGennaro, R. P. (1990). Stock returns and volatility. *Journal of financial and Quantitative Analysis*, 25(2):203–214.

Bali, T. G. and Peng, L. (2006). Is there a risk–return trade-off? evidence from high-frequency data. *Journal of Applied Econometrics*, 21(8):1169–1198.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied econometrics*, 17(5):457–477.

Bekaert, G. and Engstrom, E. (2017). Asset return dynamics under habits and bad environment–good environment fundamentals. *Journal of Political Economy*, 125(3):713–760.

Bekaert, G., Engstrom, E., and Ermolov, A. (2023). The variance risk premium in equilibrium models. *Review of Finance*, 27(6):1977–2014.

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

Bekaert, G., Xu, N. R., and Ye, T. (2025). Forecasting international stock market variances. *Available at SSRN 4831547*.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, 49(3):633–661.

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

Bollerslev, T., Todorov, V., and Xu, L. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134.

Boudoukh, J., Israel, R., and Richardson, M. (2019). Long-horizon predictability: A cautionary tale. *Financial Analysts Journal*, 75(1):17–30.

Boudoukh, J., Israel, R., and Richardson, M. (2022). Biases in long-horizon predictive regressions. *Journal of Financial Economics*, 145(3):937–969.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics*, 18(2):373–399.

Campbell, J. Y., Giglio, S., Polk, C., and Turley, R. (2018). An intertemporal capm with stochastic volatility. *Journal of Financial Economics*, 128(2):207–233.

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

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.

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

Chabi-Yo, F., Gourier, E., and Langlois, H. (2025). Option-implied risk premia with intertemporal hedging. *Available at SSRN 5404768*.

Chabi-Yo, F. and Loudis, J. (2020). The conditional expected market return. *Journal of Financial Economics*, 137(3):752–786.

Cheng, I.-H. (2020). Volatility markets underreacted to the early stages of the covid-19 pandemic. *The Review of Asset Pricing Studies*, 10(4):635–668.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business & Economic Statistics*, 25(4):411–426.

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

Couts, S. J., Gonçalves, A. S., and Loudis, J. (2023). The subjective risk and return expectations of institutional investors. *Fisher College of Business Working Paper*, 14.

Dahlquist, M. and Ibert, M. (2024). Equity return expectations and portfolios: Evidence from large asset managers. *The Review of Financial Studies*, 37(6):1887–1928.

Drechsler, I. and Yaron, A. (2011). What's vol got to do with it. *The Review of Financial Studies*, 24(1):1–45.

Eraker, B. (2025). Predictability puzzles. *Journal of Financial and Quantitative Analysis*, 60(1):482–523.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of financial economics*, 76(3):509–548.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

Gómez-Cram, R. (2022). Late to recessions: Stocks and the business cycle. *The Journal of Finance*, 77(2):923–966.

Goyal, A., Welch, I., and Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 37(11):3490–3557.

Inoue, A. and Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402.

Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.

Jensen, C. S. and Lazarus, E. (2026). Conditional excess volatility. *Working paper*.

Jensen, T. I. (2022). Subjective risk and return. *Working Paper*.

Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644.

Johnson, T. L. (2019). A fresh look at return predictability using a more efficient estimator. *The Review of Asset Pricing Studies*, 9(1):1–46.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *The quarterly journal of economics*, 134(3):1225–1298.

Knox, B., Londono, J. M., Samadi, M., and Vissing-Jorgensen, A. (2024). Equity premium events. *Available at SSRN 4773692*.

Knox, B. and Vissing-Jorgensen, A. (2024). A stock return decomposition using observables.

Kremens, L., Martin, I. W., and Varela, L. (2025). Long-horizon exchange rate expectations. *The Journal of Finance*, 80(6):3695–3724.

Leombroni, M., Pflueger, C. E., and Sunderam, A. (2026). Risk and return in government bonds. *Available at SSRN 7059018*.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of financial economics*, 74(2):209–235.

Lochstoer, L. A. and Muir, T. (2022). Volatility expectations and returns. *The Journal of Finance*, 77(2):1055–1096.

Londono, J. M. (2015). The variance risk premium around the world. *Available at SSRN 2517020*.

Londono, J. M. and Xu, N. R. (2024). The global determinants of international equity risk premiums. *Management Science*, 70(9):6374–6394.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of financial economics*, 83(1):171–222.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836–2003. *Journal of Financial Economics*, 85(1):123–150.

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

Martin, I. (2025). Information in derivatives markets: Forecasting prices with prices. *Working paper*.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Moreira, A. and Muir, T. (2019). Should long-term investors time volatility? *Journal of Financial Economics*, 131(3):507–527.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, 104(2):228–250.

Pesch, D., Piatti, I., and Whelan, P. (2024). Subjective risk premia in bond and fx markets. *Working paper*.

Schreindorfer, D. and Sichert, T. (2025). Conditional risk and the pricing kernel. *Journal of financial economics*, 171:104106.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153.

Schwert, G. W. (1990). Indexes of us stock prices from 1802 to 1987. *Journal of business*, pages 399–426.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of financial economics*, 54(3):375–421.

Tetlock, P. C., McCoy, J., and Shah, N. (2024). The implied equity premium. *Available at SSRN 4373579*.

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

Westerlund, J. and Narayan, P. (2015). Testing for predictability in conditionally heteroskedastic stock returns. *Journal of Financial Econometrics*, 13(2):342–375.

Yang, A. (2026). Understanding negative risk-return trade-offs. *Forthcoming, Management Science*.

Zhou, H. (2018). Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty. *Annual Review of Financial Economics*, 10(1):481–497.

A Supplementary empirical results

This appendix section provides supplementary empirical results to the main results in the paper.

Table A.1 provides summary statistics and a correlation matrix of variables that excludes crisis periods (1996-2025).

Table A.2 provides regression results of realized returns regressed on realized physical return variance.

Figure A.2 provides expanding window R^2 of regressing realized excess returns on two measures of the variance risk premium.

Table A.3 shows day-of-month results that summarized in Figure 7

Table A.4 presents international panel results using realized return variance and the associated implied variance risk premium as the regressor.

Table A.1: Summary Statistics (excluding crisis periods).

This table presents summary statistics exlcuding the GFC and Covid-19 periods.

Panel A: Monthly U.S. returns, measures of the variance of returns, and VRP (1996-2025)

	count	mean	sd	p5	p25	p50	p75	p95
Realized excess return (in %)	7,200	0.67	4.19	-6.89	-1.69	1.13	3.35	6.79
Realized variance (high-frequency measurement)	7,200	0.20	0.19	0.04	0.08	0.14	0.24	0.57
Realized variance (daily measurement)	7,200	0.24	0.25	0.04	0.08	0.16	0.30	0.71
Forecasted variance (HAR model)	7,200	0.21	0.13	0.11	0.13	0.17	0.23	0.42
Risk-neutral variance ($SVIX^2$)	7,200	0.30	0.20	0.10	0.15	0.25	0.39	0.70
Risk-neutral entropy (VIX^2)	7,200	0.35	0.24	0.11	0.17	0.28	0.44	0.82
VRP (forecasted variance)	7,200	0.11	0.13	-0.02	0.02	0.08	0.17	0.36
VRP (realized variance)	7,200	0.11	0.13	-0.05	0.05	0.09	0.16	0.31

Panel B: Correlation matrix of U.S. returns and variance measures

	Ex. return	RVar, hf	RVar, d	FVar	RN Var	RN Entropy	VRP, FV	VRP, RV
Ex. return	1.000							
RVar, hf	0.125	1.000						
RVar, d	0.098	0.946	1.000					
FVar	0.137	0.804	0.765	1.000				
RN Var	0.176	0.769	0.758	0.787	1.000			
RN Entropy	0.177	0.766	0.755	0.787	0.997	1.000		
VRP, FV	0.143	0.458	0.477	0.307	0.827	0.822	1.000	
VRP, RV	0.094	-0.228	-0.169	0.073	0.447	0.446	0.619	1.000

Figure A.1: Slope estimate from expanding-window OLS regressions.

This figure plots the beta estimate from expanding-window OLS estimations of

$$r_{t+1}^{ex} = \alpha + \beta x_t + \epsilon_{t+1},$$

where r_{t+1}^{ex} denotes the one-month excess return on the S&P 500 and x_t is the explanatory variable. The three predictors are: (1) realized variance over the previous month (blue), (2) risk-neutral variance (red), and (3) a Corsi (2009) forecast of next month's variance (green). Regressions use daily overlapping observations from January 1996 through the date shown on the x-axis; the last observation is July 30, 2025. The shaded areas indicate the GFC and Covid-19 episodes. The right panel excludes observations in these periods.

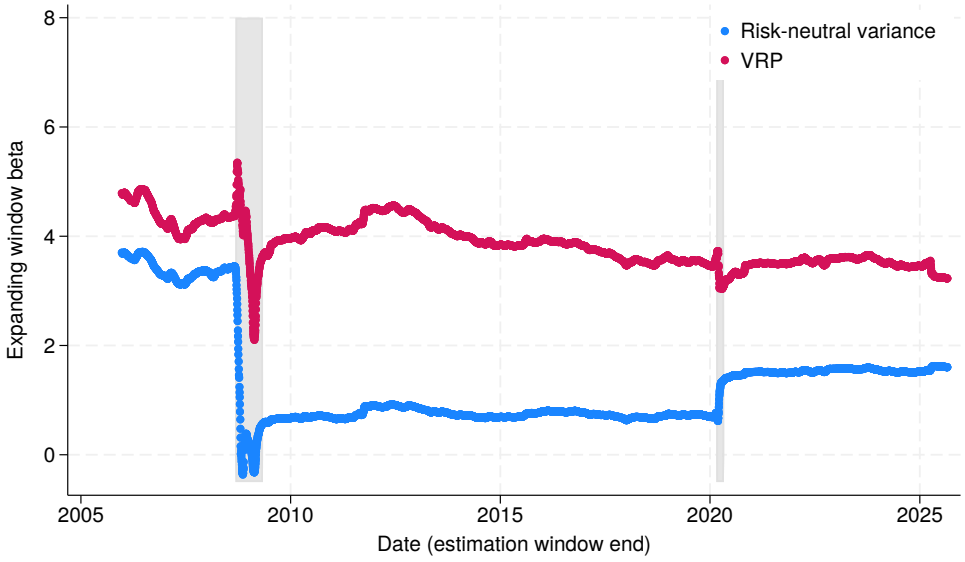


Figure A.2: Model fit from expanding window regressions on the S&P500.
This figure plots the model fit (R^2) from expanding window OLS estimations of:

$$r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1},$$

where r^{ex}_{t+1} denotes the one-month excess return on the S&P 500 and x_t is the explanatory variable. The results show model fit estimating with three different x_t: (1) risk-neutral variance (shown in blue), (2) a model-free VRP measured using lagged realized variance (shown in red), and (3) a VRP measured using forecasted variance (shown in green). Regressions use a sample of daily observations from January 1996 through to the date shown on the x-axis. The last observation is July 30, 2025. The shaded areas show the GFC and Covid-19 episodes, with observations in these periods excluded from the estimation in the right hand side figure.

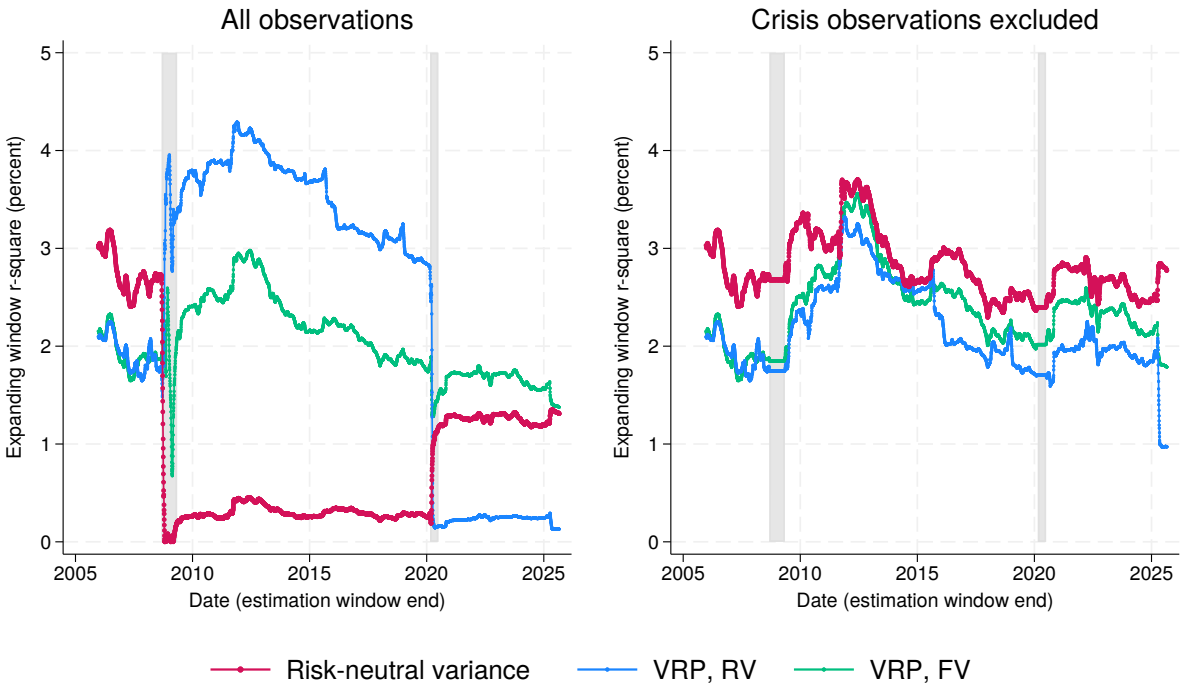


Table A.2: Physical variance and returns under a broad range of models.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month realized physical variance from time $t-1$ to time t. Regressions use a sample of daily observations from January 1996 through July 2025. Panel A shows estimation results with a constant and panel B shows estimation results without a constant. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Estimations with a constant

	Dependent Variable: 1-month excess return					
	(1) OLS	(2) WLS	(3) ex.Crisis	(4) Trim	(5) Quantile	(6) Robust
Realized physical variance	0.877	1.175	2.831**	2.533**	1.206	1.840*
	(0.850)	(0.836)	(1.111)	(1.262)	(0.916)	(1.067)
Constant	0.004	0.003	0.001	0.002	0.009***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	7443	7443	7179	6991	7443	7443

Panel B: Estimations without a constant

	Dependent Variable: 1-month excess return					
	(1) OLS	(2) WLS	(3) ex.Crisis	(4) Trim	(5) Quantile	(6) Robust
Realized physical variance	1.287*	2.060***	3.232***	3.309***	2.926***	2.991***
	(0.747)	(0.750)	(0.824)	(0.885)	(0.861)	(0.977)
Observations	7443	7443	7179	6991	7443	7380

Table A.3: Non-overlapping U.S. stock return regressions, by trade-day of month.

This table presents estimates from the regression: $r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral variance at time t. Regressions use monthly observations from January 1996 to July 2025, with separate estimations for each trade date of the month. Heteroskedasticity-robust standard errors (Huber-White) are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: OLS

	Dependent Variable: 1-month realized excess return using (n)th trading day of the month																				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(Last)
Risk-neutral variance	2.13	1.03	1.50	1.62	1.68	2.38*	1.81*	1.69**	1.42	0.80	2.20**	0.90	2.08*	2.92***	3.12*	1.80	0.39	0.32	0.86	0.75	0.84
	(1.41)	(1.73)	(1.57)	(1.54)	(1.74)	(1.36)	(1.00)	(0.72)	(1.19)	(0.92)	(0.86)	(0.95)	(1.21)	(1.12)	(1.63)	(1.87)	(1.52)	(1.46)	(0.92)	(1.55)	(1.50)
Constant	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2 (%)	2.13	0.41	0.82	1.20	1.10	2.58	1.57	1.69	1.10	0.32	3.50	0.50	2.83	5.08	4.60	1.62	0.07	0.05	0.39	0.26	0.29
Observations	356	356	356	356	356	356	356	356	356	356	356	356	356	356	356	355	355	354	351	312	356

Panel B: WLS

	Dependent Variable: 1-month realized excess return using (n)th trading day of the month																				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(Last)
Risk-neutral variance	1.97	1.28	1.70	1.59	1.70	2.36*	2.22**	2.12**	2.16**	1.86*	2.75***	1.77*	1.97**	2.22**	2.43**	1.69	0.75	0.96	1.19	1.52	1.34
	(1.21)	(1.31)	(1.29)	(1.25)	(1.46)	(1.23)	(1.08)	(1.02)	(1.04)	(1.03)	(0.90)	(0.99)	(0.94)	(0.92)	(1.08)	(1.10)	(1.08)	(1.14)	(1.16)	(1.26)	(1.22)
Constant	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2 (%)	0.91	0.36	0.60	0.58	0.61	1.18	1.02	0.98	0.99	0.71	1.46	0.55	0.74	0.99	1.26	0.62	0.11	0.19	0.29	0.45	0.39
Observations	356	356	356	356	356	356	356	356	356	356	356	356	356	356	356	355	355	354	351	312	356

B Skewness and regressor–variance alignment

This appendix section provides a simple benchmark that makes explicit the role of right-skewness in regressor–variance alignment. Suppose that the risk predictor is affine in conditional return variance,

$$x_t = a + b\sigma_{t+1}^2, \qquad b > 0, \tag{34}$$

where $\sigma_{t+1}^2 = \mathrm{Var}(\varepsilon_{t+1} \mid \mathcal{F}_t)$. This affine assumption is used here only to illustrate the mechanism and is not necessary for the main analysis.

Let

$$v_t \equiv \sigma_{t+1}^2, \qquad \mu_v \equiv E[v_t], \qquad s_v^2 \equiv \mathrm{Var}(v_t).$$

Equation (34) implies

$$\tilde{x}_t \equiv x_t - E[x_t] = b(v_t - \mu_v), \qquad \mathrm{Var}(x_t) = b^2 s_v^2.$$

Hence, slope leverage can be written as

$$H_t = \frac{\tilde{x}_t^2}{\mathrm{Var}(x_t)} = \frac{(v_t - \mu_v)^2}{s_v^2}, \tag{35}$$

and the covariance between slope leverage and conditional return variance is therefore

$$\begin{aligned} \mathrm{Cov}(H_t, v_t) &= E[H_t v_t] - E[H_t]E[v_t] \\ &= \frac{E[(v_t - \mu_v)^2 v_t]}{s_v^2} - \mu_v \\ &= \frac{E[(v_t - \mu_v)^3]}{s_v^2}. \end{aligned} \tag{36}$$

Let

$$\gamma_v \equiv \frac{E[(v_t - \mu_v)^3]}{s_v^3} \tag{37}$$

denote the skewness of conditional return variance. It follows immediately that

$$\mathrm{Cov}(H_t, \sigma_{t+1}^2) = \gamma_v s_v. \tag{38}$$

Thus, in this affine benchmark, regressor–variance alignment is governed directly by the skewness of conditional return variance. Recall from Equation 13 that the OLS asymptotic

variance can be written as

$$\Omega_{OLS} = \frac{E[\sigma^2_{t+1}]}{\text{Var}(x_t)} + \frac{\text{Cov}(H_t, \sigma^2_{t+1})}{\text{Var}(x_t)}. \tag{39}$$

Using (38), this becomes

$$\Omega_{OLS} = \frac{\mu_v}{\text{Var}(x_t)} + \frac{\gamma_v s_v}{\text{Var}(x_t)} = \frac{\mu_v + \gamma_v s_v}{b^2 s_v^2}. \tag{40}$$

Positive skewness therefore increases the asymptotic variance of OLS through the regressor–variance alignment term.

The result also clarifies why a positive relation between the regressor and conditional return variance is not, by itself, sufficient to generate positive alignment. Under (34),

$$\text{Cov}(x_t, \sigma^2_{t+1}) = b\text{Var}(\sigma^2_{t+1}) = b s_v^2 > 0, \tag{41}$$

yet if the distribution of conditional variance is symmetric, then $\gamma_v = 0$ and

$$\text{Cov}(H_t, \sigma^2_{t+1}) = 0.$$

The reason is that slope leverage depends on squared deviations of x_t from its mean. Under symmetry, large deviations below and above the mean generate comparable slope leverage: high leverage on the low-x_t, low-variance side offsets high leverage on the high-x_t, high-variance side. Positive skewness breaks this symmetry. Extreme deviations occur disproportionately in the upper tail, so observations with the greatest slope leverage also tend to have the highest conditional return variance.

C A simple economic interpretation of the international evidence

One direct interpretation of the international risk–return relation arises under market segmentation or when local investors are marginal in their domestic market. If compensation for bearing country-market risk is determined locally, standard asset-pricing logic implies that expected market returns increase with local conditional risk. With a common effective price of risk across countries, this gives $E_t[r^e_{i,t+1}] = \gamma\sigma^2_{it}$. Strong home bias or incomplete international risk sharing can generate a similar mechanism. Importantly, however, such segmentation is not required for the empirical prediction. We next show that the same cross-country relation can arise in fully integrated markets.

Let country i's excess market return satisfy

$$r^e_{i,t+1} = \mu_{it} + G_{t+1} + u_{i,t+1}, \qquad E_t[G_{t+1}] = E_t[u_{i,t+1}] = 0, \tag{42}$$

where G_{t+1} is a common global shock and $u_{i,t+1}$ is a country-level return innovation orthogonal to G_{t+1}. For expositional simplicity, assume that the $u_{i,t+1}$ are mutually conditionally orthogonal. Let

$$\mathrm{Var}_t(G_{t+1}) = q_t, \qquad \mathrm{Var}_t(u_{i,t+1}) = v_{it}. \tag{43}$$

Country i's conditional return variance is then

$$\sigma^2_{it} \equiv \mathrm{Var}_t(r^e_{i,t+1}) = q_t + v_{it}. \tag{44}$$

Thus, q_t generates common time-series movement in country variances, while v_{it} generates cross-country differences in risk.

Define a second global factor as the aggregate of the country-level innovations,

$$H_{t+1} = \sum_j u_{j,t+1}, \tag{45}$$

and consider the normalized linear stochastic discount factor

$$m_{t+1} = 1 - aG_{t+1} - \gamma H_{t+1}, \qquad a, \gamma > 0. \tag{46}$$

Under no arbitrage, the expected excess return for country i is

$$E_t[r^e_{i,t+1}] = -\operatorname{Cov}_t\left(m_{t+1}, r^e_{i,t+1}\right) \tag{47}$$

$$= aq_t + \gamma v_{it}, \tag{48}$$

because

$$\operatorname{Cov}_t(H_{t+1}, u_{i,t+1}) = \operatorname{Cov}_t\left(\sum_j u_{j,t+1}, u_{i,t+1}\right) = v_{it}. \tag{49}$$

Cross-country variation in v_{it} therefore captures variation in exposure to priced aggregate risk beyond the common global component G_{t+1}. More generally, H_{t+1} could be a weighted aggregate of country-level shocks; the equal-weight specification is the simplest case delivering a common cross-country slope.

At a given date, the first term in equation (48) is common across countries. Hence

$$E_t[r^e_{i,t+1}] - E_t[r^e_{j,t+1}] = \gamma(v_{it} - v_{jt}) = \gamma(\sigma^2_{it} - \sigma^2_{jt}). \tag{50}$$

Equivalently, substituting $v_{it} = \sigma^2_{it} - q_t$ into equation (48) yields

$$E_t[r^e_{i,t+1}] = \alpha_t + \gamma\sigma^2_{it}, \qquad \alpha_t \equiv (a - \gamma)q_t. \tag{51}$$

Thus, with unrestricted time effects, a common slope on local conditional variance follows without requiring the two sources of risk to have the same price. The stronger restriction

$$a = \gamma \tag{52}$$

eliminates α_t and gives the unified relation

$$E_t[r^e_{i,t+1}] = \gamma\sigma^2_{it}. \tag{53}$$

This stronger restriction also has a useful empirical implication. In the simple model, after removing any country fixed effects, a regression without time effects has population slope

$$\beta^{\text{pool}} = \gamma + (a - \gamma)\frac{\operatorname{Cov}(\sigma^2_{it}, q_t)}{\operatorname{Var}(\sigma^2_{it})}. \tag{54}$$

If common global risk accounts for a substantial share of variation in country variances, the pooled slope reflects both a and γ, whereas time effects identify γ from cross-country variation. Similar estimated slopes across the two specifications are therefore consistent with similar compensation for the two sources of risk.

The model provides a simple interpretation of the international evidence. The first global factor generates the strong comovement in country risk, while cross-country differences in local variance capture exposure to an additional priced aggregate factor. Market segmentation or local marginal investors provide an alternative route to the same prediction, but are not required. The priced factors need not coincide with observed equity-market portfolios; they may instead represent shocks, or factor-mimicking portfolios, that span relevant components of the stochastic discount factor.

D Entropy risk and the VIX

Beyond examining higher-order moments, an additional robustness check considers the use of the VIX^2 as an alternative measure of option-implied risk. The VIX^2 represents the risk-neutral entropy of market returns and is therefore closely related to the risk-neutral variance measure used throughout the main body of the paper. The two are tightly linked, with the VIX^2 computed from option prices in a similar fashion to risk-neutral variance:

$$\frac{VIX_t^2}{12} = 2R_{t+1}^f \left[\int_0^{F_t} \frac{1}{K^2} \text{put}_t(K) dK + \int_{F_t}^{\infty} \frac{1}{K^2} \text{call}_t(K) dK \right] \tag{55}$$

where the 1/12 on the left hand side reflects that the VIX (as reported by the CBOE) is annualized. Comparing the right hand side of Equation 55 with Equation 4, the key difference thus comes down to the weighting scheme placed on option prices within the integral. While the variance measure in Equation 4 is equal weighted with $\frac{1}{P_t^2}$ outside of the integral, the entropy measure in Equation 55 places more weight on left tail option prices through the $\frac{1}{K^2}$ weighting within the integral. The 30-day CBOE Volatility Index (VIX) is retrieved from FRED, Federal Reserve Bank of St. Louis, with the daily time series available from January 1990 through to present. Although the series is highly correlated with risk-neutral variance, the VIX^2 is typically larger and more volatile than its variance counterpart.[27]

Table D.5 reports regressions using VIX_t^2 as x_t. The model fit and coefficient magnitudes are nearly identical to those obtained with risk-neutral variance, indicating that the results are not sensitive to the precise option-based measure of implied risk. Under OLS, the estimated β for the baseline 1996–2025 sample is 1.29 (s.e. = 0.82), while under WLS the corresponding estimate is 1.54 (s.e. = 0.69) and significant at the 5 percent level. The OLS model-fit is marginally lower than for risk-neutral variance with an R^2 of 1.25 percent relative to the 1.31 percent reported in Table 4 column (1).

Since the VIX^2 series is availability back to 1990, it enables an out-of-sample validation of the main findings with risk-neutral variance. In the early 1990–1995 period, the estimated WLS slope coefficient is 1.44 (column 5), which is almost identical to the 1.54 estimate from the 1996–2025 sample (column 4). This provides further evidence that the option-implied risk–return relationship has remained stable over the 35 years from 1990 to 2025, and further supports the synthetic variance exercise used in the historical 1885 results.

[27] See Martin (2017) for further discussion of the theoretical and empirical distinctions between risk-neutral volatility (which he defines as the SVIX) and the VIX, as well as the economic implications of these distinctions.

E OptionMetrics data cleaning

To clean the OptionMetrics data and generate the risk-neutral moments of the return distribution I take the following steps. First I drop observations if the bid price or ask price is missing and if the best bid price is zero. I then calculate the mid price as the average of the best bid and best ask price and, for date-maturity-strike-type combinations where there are multiple mid-prices, we use the option with the highest open interest.[28] I next keep date-maturity-strike observations where there is both a call and put mid-price, and then select the option with the lowest mid-price. This step automatically deletes put options greater than $F_{t,T}$ and deletes call option prices less than $F_{t,T}$, as required in the estimation of Equation 4. Before implementing the discretizion of the integral, I then take two extra steps to ensure the accuracy of the discretizion approximation. First, I drop date-maturity observations where the number of unique strikes is less than 10. Second, I drop date-maturity observations where the difference between the maximum put strike and the minimum call strike is greater than 50. This step makes sure discretization is not too coarse in the most important range of the integral (where prices are at the highest levels).

Following these cleaning steps, Equation 4 is then estimated, providing date-maturity risk-neutral variance estimates at the maturities that correspond to the expiration dates of the options in issue. To generate constant maturity 30-day estimates across trade dates, I then linearly interpolate across maturities within dates.

[28] Type refers to call or put option. The existence of multiple maturity-strike-type observations is caused by the increase in the number of options issued. For example, if a newly issued weekly option has the same expiry as an existing annual option that is expiring in a weeks time, then we observe two mid-prices. Typically, shorter horizon options are the most liquid and have highest open interest (i.e. weekly options are more liquid than annual options). We drop all quarterly options (expiring end of quarter) from the sample as these are particularly illiquid throughout their issuance period.

Table A.4: Returns and other measures of risk, evidence from international stock markets.

This table presents estimates from the regression: $r_{i,t+1} = \alpha + \beta x_{i,t} + \gamma_i + \gamma_t + \epsilon_{i,t+1}$, where $r_{i,t+1}$ is the excess return of international stock market i over a one-month period from t to $t+1$, $x_{i,t}$ is the one-month realized variance for index i at time t (panel A) or the one-month VRP (panel B), and γ_i and γ_t are index and time fixed effects. Regressions use a sample of daily observations from January 1996 to June 2022. Standard errors clustered by date and index are reported in parenthesis.

Panel A: Realized return variance

	Dependent Variable: 1-month realized excess return							
	OLS				WLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Realized variance	0.30*	0.32*	0.29**	0.33**	0.60**	0.77***	0.26	0.66**
	(0.13)	(0.14)	(0.10)	(0.12)	(0.22)	(0.22)	(0.31)	(0.27)
Constant	0.00*				0.00			
	(0.00)				(0.00)			
Index FE	no	yes	no	yes	no	yes	no	yes
Date FE	no	no	yes	yes	no	no	yes	yes
Weighted Adj R^2	0.14	0.27	77.77	77.94	0.08	0.50	73.90	70.23
Weighted Adj R^2 (Within)		0.16	0.21	0.27		0.12	0.02	0.12
Observations	45,203	45,203	43,713	43,713	45,203	45,203	43,713	43,713

Panel B: Variance risk premium

	Dependent Variable: 1-month realized excess return							
	OLS				WLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variance risk premium	0.26	0.26	-0.06	-0.06	1.61***	1.66***	0.62**	0.67**
	(0.26)	(0.26)	(0.08)	(0.08)	(0.34)	(0.34)	(0.26)	(0.25)
Constant	0.00***				0.00			
	(0.00)				(0.00)			
Index FE	no	yes	no	yes	no	yes	no	yes
Date FE	no	no	yes	yes	no	no	yes	yes
Weighted Adj R^2	0.05	0.17	77.73	77.88	0.29	0.69	73.92	74.27
Weighted Adj R^2 (Within)		0.05	0.01	0.01		0.31	0.09	0.11
Observations	45,203	45,203	43,713	43,713	45,203	45,203	43,713	43,713

Table D.5: Risk-neutral entropy and U.S. stock market returns.

This table presents estimates from the regression: $r^{ex}_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$, where r^{ex}_{t+1} is the excess return of the S&P500 over a one-month period from t to $t+1$ and x_t is the one-month risk-neutral entropy at time t (i.e. the VIX^2_t). Regressions use a sample of daily observations from January 1990 through to July 2025. Columns (1)-(3) are estimated with OLS and columns (4)-(6) are estimated with WLS. Standard errors from 30 calendar day (non-overlapping) block bootstrap are presented in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	Dependent Variable: 1-month realized excess return					
	OLS			WLS		
	(1996–2025) (1)	(1990–1995) (2)	(1990–2025) (3)	(1996–2025) (4)	(1990–1995) (5)	(1990–2025) (6)
VIX^2	1.290	1.681	1.293	1.542**	1.441	1.516**
	(0.819)	(3.438)	(0.846)	(0.694)	(3.235)	(0.695)
Constant	0.001	0.000	0.001	0.000	0.001	0.000
	(0.003)	(0.006)	(0.003)	(0.002)	(0.006)	(0.002)
Adjusted R^2 (%)	1.25	0.61	1.22	2.01	2.13	2.04
Observations	7422	1516	8938	7422	1516	8938