ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

Fear, Policy Noise and Market Pain: Can Free Uncertainty Indicators Improve S&P 500 Risk Forecasts

Cailing Li

College of International Education, South China Agricultural University, Guangzhou, Guangdong, China

cl_0911@163.com

Abstract

This study asks whether freely available uncertainty and stress indicators improve weekly forecasts of S&P 500 risk once past returns and past volatility are already in the model. Using 1,898 weekly observations from January 1990 to May 2026, it models two targets, namely next-week realised volatility and a next-week downside event defined as the worst 5 per cent of weekly returns. Predictors are the VIX, economic policy uncertainty, a corporate credit spread, the yield curve slope, the short-term interest rate and oil market volatility, all lagged by one week. The VIX is the only added variable that improves the volatility forecast in a clear way, lifting the adjusted R-squared from 0.519 to 0.593 and reducing out-of-sample error, while policy uncertainty, the credit spread and the other controls add very little once it is present. For downside events the wider stress set raises in-sample fit but does not survive out of sample, where the models are statistically hard to separate and the leaner specification does at least as well. The findings support the value of market-based fear measures and offer a more cautious view of the broader uncertainty indicators.

Keywords

Realised volatility, downside risk, VIX, economic policy uncertainty, credit spreads, forecasting.

1. Introduction

Investors, risk managers and regulators all want an early warning of when equity markets are about to become more dangerous. Two forms of danger matter most for a typical equity portfolio. The first is a rise in volatility, because higher volatility widens the range of possible outcomes and raises the cost of holding or hedging a position. The second is a sharp fall in prices, because large negative returns cause the most painful losses and tend to arrive when liquidity is poor. A natural question for a practitioner with a modest budget is whether the free public indicators that are published every day can help to predict either form of danger, or whether the information they carry is already contained in the recent behaviour of the market itself.

The most familiar of these indicators is the VIX, the implied volatility index produced by the Chicago Board Options Exchange, often described as the market's fear gauge because it rises when investors pay more for downside protection [1, 2]. Alongside it sits a growing family of uncertainty and stress measures that are now available without charge. The economic policy uncertainty index built from newspaper coverage offers a text-based reading of policy noise [3]. Corporate credit spreads summarise the price of default risk and tend to widen when financing conditions tighten [4]. The slope of the yield curve has long been read as a signal about the future path of the economy [5]. Each of these series can be downloaded from the Federal Reserve Economic Data service at no cost, which makes them attractive inputs for a student, a small fund or anyone who cannot pay for a commercial data feed.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

The central question of this paper is straightforward. Do uncertainty and stress indicators improve the prediction of S&P 500 volatility and downside risk beyond what is already known from past market returns and past volatility? The question matters because a forecast that merely repeats the recent past is cheap and easy to build, so any additional indicator has to earn its place by improving on that baseline. It also matters because the academic literature has produced mixed evidence on whether macroeconomic and uncertainty variables add anything to volatility forecasts once the persistence of volatility itself is taken into account [6, 7].

This study takes a deliberately simple and transparent approach that suits a master's level analysis. It converts daily data into weekly observations, builds a continuous measure of realised volatility and a binary downside event, and then estimates a sequence of nested models that add the uncertainty indicators one block at a time. Ordinary least squares is used for the continuous volatility target and logistic regression is used for the binary downside target. The models are compared on in-sample fit and, more importantly, on out-of-sample performance through an expanding window exercise that mimics how a forecaster would actually have used the data through time. Robustness checks remove the two largest crises so that the results are not driven by a handful of extreme weeks.

The main message is honest and a little sobering for anyone hoping that more indicators always mean better forecasts. The VIX is genuinely useful and improves the volatility forecast by a clear margin, while the wider set of uncertainty and stress variables, including policy uncertainty, the credit spread, the yield curve, the interest rate and oil volatility, adds very little once the VIX is in the model. The contribution of the paper is therefore not a claim of strong new predictive power, but a careful and reproducible account of where free indicators help and where they do not.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature and sets out four hypotheses. Section 3 describes the data and the methodology, including the construction of the weekly variables and the lag structure that guards against look-ahead bias. Section 4 reports the descriptive statistics, the regression models and the out-of-sample comparison. Section 5 discusses what the results mean and how they relate to earlier work, and Section 6 concludes.

2. Literature Review And Hypotheses

This section reviews four strands of work, namely the forecasting of equity volatility, the information carried by the VIX, broader uncertainty measures, and credit and yield-curve signals, before stating the four hypotheses.

2.1. Forecasting Equity Volatility

The study of changing equity volatility has a long history. Schwert [8] documented that aggregate stock market volatility varies a great deal over time and is only loosely connected to measures of economic activity, which set an early puzzle about how far macroeconomic information can explain volatility. The autoregressive conditional heteroskedasticity framework of Engle [9] and its generalised form in Bollerslev [10] gave researchers a way to model the clustering of volatility, where calm periods and turbulent periods each tend to persist. A later strand used high-frequency data to measure volatility directly rather than to infer it from a model, and Andersen et al. [11] showed that realised volatility built from intraday returns can be modelled and forecast with comparatively simple time-series methods. Corsi [12] added a parsimonious heterogeneous autoregressive model, known as HAR, that captures the long memory of realised volatility by combining daily, weekly and monthly components, and it has become a common benchmark that the present study adopts.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

A practical lesson from this literature is that volatility is highly persistent, so a forecast that uses recent realised volatility is already quite accurate. This persistence sets a demanding baseline for any additional predictor. Paye [6] examined whether macroeconomic variables such as default spreads and other financial conditions help to predict aggregate volatility, and found that the gains are modest and uneven. Christiansen et al. [7] reached a similar conclusion in a broad study of economic predictors of financial volatility, reporting that many variables look useful in sample but offer limited out-of-sample improvement. These findings frame the present study, because they suggest that the bar for the uncertainty indicators is high and that honest out-of-sample testing is essential. Patton [13] further reminds us that an imperfect volatility proxy can distort comparisons between models, so the choice of evaluation metric needs care.

2.2. The VIX and Market Fear

The VIX draws on option prices to summarise the volatility that the market expects over the coming month. Whaley [1] introduced it as an investor fear gauge and later explained how it behaves and why it tends to spike during falling markets [2]. Because the index reflects the price that investors are willing to pay for protection, it should embed forward-looking information that backward-looking realised volatility cannot contain. Giot [14] found an asymmetric pattern in which very high implied volatility indices tend to be followed by positive stock index returns. A connected literature on the variance risk premium, the gap between option-implied and realised variance, links the VIX to expected returns and to economic uncertainty [15, 16, 17]. The common theme is that the VIX carries information beyond recent realised volatility, which motivates a direct test of whether it improves a weekly volatility forecast.

2.3. Economic Policy Uncertainty And Related Measures

A second family of indicators tries to capture uncertainty that originates outside the options market. Baker et al. [3] built an index of economic policy uncertainty from the frequency of newspaper articles that discuss policy and uncertainty together, and showed that it rises around elections, wars and major policy debates. Theoretical and empirical work suggests that policy uncertainty should command a risk premium and should raise volatility, because firms and investors delay decisions when the rules of the game are unclear [18, 19]. Bali et al. [20] reported that broad economic uncertainty is priced in the cross-section of stock returns, while Jurado et al. [21] constructed a statistical measure of macroeconomic uncertainty and argued that genuine uncertainty episodes are fewer and larger than newspaper-based measures imply. Related text and search-based indicators, such as news-implied volatility [22], investor attention from internet searches [23] and geopolitical risk from news coverage [24], all attempt to read sentiment and uncertainty from sources that are cheap to obtain. Whether such measures improve a weekly volatility or downside forecast after the VIX is included remains an open question that this paper addresses.

2.4. Credit Spreads, The Yield Curve And Downside Risk

A third group of predictors comes from fixed income and reflects financing conditions. Gilchrist and Zakrajšek [4] showed that corporate bond credit spreads contain useful information about future economic activity and that a component of the spread linked to investor sentiment is especially informative. Estrella and Hardouvelis [5] established that the slope of the yield curve predicts real activity, and Adrian et al. [25] extended the idea to the whole distribution of future growth, showing that financial conditions shape downside risk much more than upside potential. This asymmetry connects directly to the downside target in the present study. The asset pricing literature on downside risk argues that investors care more about losses than about gains and demand compensation for assets that fall hardest in bad times [26], while work

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

on tail risk shows that the risk of extreme outcomes varies over time and is priced [27]. Taken together, this strand suggests that credit and macro-financial stress indicators may be more useful for predicting downside events than for predicting the general level of volatility.

2.5. Out-of-Sample Evaluation

A recurring lesson from the return and volatility prediction literature is that in-sample fit can be misleading. Welch and Goyal [28] showed that many popular predictors of the equity premium fail to beat a simple historical average out of sample, and Campbell and Thompson [29] argued that sensible restrictions can restore a small but real degree of out-of-sample predictability. Formal tests of equal predictive accuracy provide a way to compare forecasts [30], and this study places weight on an expanding window exercise and on those tests rather than on in-sample statistics alone.

2.6. Hypotheses

The literature leads to four hypotheses that the empirical work will test directly.

H1. Higher implied volatility is positively associated with next week's realised volatility.

H2. Higher economic policy uncertainty is positively associated with next week's realised volatility.

H3. Credit stress indicators improve the prediction of downside events.

H4. A full uncertainty model improves forecasting performance compared with a baseline model that uses only past returns and past volatility.

3. Data and Methodology

This section sets out the data sources, the construction of the weekly variables and outcomes, the lag structure that prevents look-ahead bias, the models, and the out-of-sample design.

3.1. Data Sources

The study uses only free and publicly available secondary data. The S&P 500 daily closing level comes from Yahoo Finance, because the version of the index published on the Federal Reserve Economic Data service covers only the most recent ten years and would not reach the earlier crises. All other series come from the Federal Reserve Economic Data service through its free comma-separated download route, which does not require an API key. The VIX is the VIXCLS series, economic policy uncertainty is the daily USEPUINDXD series, the credit spread is the BAA10Y series measuring the gap between Moody's Baa corporate yield and the ten-year Treasury yield, the yield curve slope is the T10Y2Y series measuring the ten-year minus two-year Treasury spread, the short rate is the daily effective federal funds rate DFF, and oil is the WTI spot price DCOILWTICO. The high-yield option-adjusted spread was the first choice for credit stress, but the free route returned data only from 2023, so the long-history Baa spread was used instead and this decision is recorded in the data audit rather than made silently. Table 1 lists every variable, its construction and its source.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

Table 1. Variable definitions and sources

Variable	Description	Construction	Source
RET	Lagged weekly return	Sum of daily S&P 500 log returns in week t, in %	Yahoo Finance (^GSPC)
RV	Lagged realised volatility	sqrt(252 * mean squared daily log return in week t) * 100, annualised %	Yahoo Finance (^GSPC)
VIX	Lagged implied volatility	Weekly mean of the daily CBOE VIX in week t	FRED VIXCLS
EPU	Lagged policy uncertainty	Weekly mean of the daily US Economic Policy Uncertainty index in week t	FRED USEPUINDXD
Credit	Lagged credit spread	Weekly mean of Moody's Baa minus 10y Treasury spread in week t, in percentage points	FRED BAA10Y
Term	Lagged yield curve slope	Weekly mean of the 10y minus 2y Treasury spread in week t, in percentage points	FRED T10Y2Y
Rate	Lagged short rate	Weekly mean of the daily effective federal funds rate in week t, in %	FRED DFF
Oil vol	Lagged oil volatility	Sqrt (252 * mean squared daily WTI log return in week t) * 100, annualised %	FRED DCOILWTICO
RV (t+1)	Realised volatility (outcome)	RV in week t+1, annualised %, the continuous forecasting target	Yahoo Finance (^GSPC)
Downside (t+1)	Downside event (outcome)	1 if the week t+1 return is in the worst 5% of weekly returns, else 0	Yahoo Finance (^GSPC)

Note: All predictors are dated week t and the two outcomes are dated week t+1.

3.2. Sample and Weekly Construction

The raw daily data span January 1990 to May 2026. Daily observations are aggregated into calendar weeks that end on Friday. The weekly return is the sum of daily log returns within the week, expressed in per cent, which is equal to the log change from the previous Friday close to the current Friday close. Weekly realised volatility is the square root of 252 times the average squared daily log return within the week, multiplied by 100, so that it is an annualised figure in per cent that can be compared directly with the VIX. The mean-of-squares form is used because it is stable when a holiday shortens the trading week. A week is retained only if it contains at least three trading days. The uncertainty and stress predictors are weekly averages of their daily values. Oil market volatility is built with the same realised volatility formula applied to daily oil returns, and the single day in April 2020 on which the oil price settled below zero is treated as missing because a log return is undefined there. After these steps the final sample contains 1,898 complete weekly observations running from 5 January 1990 to 22 May 2026, and the multicollinearity diagnostics confirm that the predictors are not so closely related as to make the models unstable.

3.3. Outcome Variables

Two outcomes are studied. The first is next-week realised volatility, a continuous variable that captures the general level of market turbulence. The second is a next-week downside event, a binary variable equal to one when the weekly return falls in the worst 5 per cent of the weekly return distribution over the sample, and zero otherwise. The downside threshold is a weekly return of -3.47 per cent, and the rule produces 95 downside events, which is 5.01 per cent of the weeks by construction. The threshold is an unconditional cut used only to label the event and not as a predictor, so it does not give the models information about the future.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

3.4. Lag Structure And Look-Ahead Bias

The forecasting design is built to avoid look-ahead bias. Every predictor is measured in week t and every outcome is measured in week t+1. In practice this means that the realised volatility and the uncertainty indicators observed at the end of one week are used to predict the volatility and the downside event of the following week. No information from the outcome week enters the predictors, and the final week of the sample is dropped because it has no following week to predict. This timing rule is the most important safeguard in the study, because a model that used contemporaneous information would look more accurate than it could be in real use.

3.5. Models

The volatility target is modelled with ordinary least squares in a sequence of nested specifications. Model 1 is the baseline and uses only the lagged weekly return and the lagged realised volatility. Model 2 adds the lagged VIX. Model 3 instead adds lagged policy uncertainty. Model 4 is the full model and adds the VIX, policy uncertainty, the credit spread, the yield curve slope, the short rate and oil volatility together. Standard errors are computed with the Newey and West heteroskedasticity and autocorrelation consistent estimator, because overlapping volatility dynamics tend to produce serially correlated errors.

The downside target is modelled with logistic regression in three nested specifications. Model 5 is the baseline with the lagged return and lagged realised volatility. Model 6 adds the VIX. Model 7 is the full model with all of the uncertainty and stress variables. The volatility models are compared on adjusted R-squared and on root mean squared error, and the downside models are compared on the area under the receiver operating characteristic curve and on the fit of the model.

Two further points belong here. First, the VIX is itself the market's option-implied forecast of volatility over the coming month, so a positive link between the VIX and next-week realised volatility is close to definitional and the meaningful question is whether the VIX adds anything beyond past realised volatility, which the formal tests below address directly. Second, to give the volatility models a demanding benchmark from the volatility literature rather than only the simple baseline, a weekly heterogeneous autoregressive model is added, labelled HAR, which uses the past one week, the average of the past four weeks and the average of the past thirteen weeks of realised volatility, all lagged.

3.6. Out-of-Sample Validation And Robustness

In-sample fit is only part of the story, so the study uses an expanding window exercise to judge genuine forecasting performance. The first 520 weeks, roughly the first ten years, form the initial training window. For each later week the model is re-estimated on all weeks up to that point and then used to predict the next week, so the forecaster never sees the future. This produces 1,378 out-of-sample weekly forecasts running from December 1999 to May 2026. Volatility models are scored by out-of-sample root mean squared error and downside models by out-of-sample area under the curve. To avoid drawing conclusions from differences that could be noise, the out-of-sample squared errors of the volatility models are compared with the Diebold and Mariano [30] test of equal predictive accuracy, and the out-of-sample downside areas under the curve are given bootstrap confidence intervals and bootstrap tests of their differences, using two thousand paired resamples of the forecasts.

Two robustness checks re-estimate the full models after removing the global financial crisis from September 2008 to June 2009 and the COVID-19 shock from February to June 2020, so that the results can be judged on whether they depend on a small number of extreme weeks. A further check addresses the way the downside event is defined. In the main models the worst-5-per-cent threshold is fixed on the full sample, which is a reasonable way to define the event but does use information from across the whole period. As a real-time alternative, the threshold

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

is re-estimated on the training window at each step of the expanding window, so that the out-of-sample labels never use future information, and the downside area under the curve is recomputed on this strictly real-time basis.

4. Results

This section reports the descriptive statistics, the diagnostics, the volatility and downside models, the out-of-sample comparison with formal tests, and the robustness checks.

4.1. Descriptive Statistics

Table 2 reports the descriptive statistics, and Figures 1 to 3 show the data through time. Figure 1 plots the S&P 500 level from 1990 to 2026 with the two crisis windows shaded. The weekly return has a mean close to zero at 0.16 per cent and a standard deviation of 2.32 per cent, with a worst week of around minus 20 per cent. Realised volatility has a mean of 14.60 in annualised per cent and a maximum above 120. The VIX has a mean of 19.45 and ranges from roughly 9 in calm periods to around 75 at the peak of the crises, and Figure 2 shows that it moves with policy uncertainty while remaining distinct. Policy uncertainty has a mean of 121 and a long right tail that reaches above 700 during the pandemic. The credit spread averages 2.29 percentage points, the yield curve slope averages around 1 percentage point and turns negative during inversions, and the short rate averages 2.89 per cent over a period that includes both high-rate and near-zero-rate regimes. Figure 3 plots weekly realised volatility and confirms the volatility clustering described in the literature.

Table 2. Descriptive statistics, weekly sample 1990 to 2026 (1,898 weeks)

Variable	N	Mean	SD	Min	25%	Median	75%	Max
RET	1898.0	0.16	2.324	-20.084	-1.021	0.283	1.419	11.424
RV	1898.0	14.6	10.539	1.142	8.136	12.102	17.807	120.8
VIX	1898.0	19.445	7.619	9.34	14.002	17.625	22.605	74.618
EPU	1898.0	121.102	82.903	23.764	70.857	98.966	138.861	705.379
Credit	1898.0	2.286	0.712	1.262	1.745	2.148	2.688	6.118
Term	1898.0	1.004	0.911	-0.996	0.242	0.859	1.782	2.868
Rate	1898.0	2.888	2.333	0.046	0.318	2.894	5.192	8.324
Oil vol	1898.0	33.68	26.05	1.409	19.843	28.162	39.506	404.444
RV (t+1)	1898.0	14.597	10.54	1.142	8.134	12.097	17.807	120.8
Downside (t+1)	1898.0	0.05	0.218	0.0	0.0	0.0	0.0	1.0

Note: RV and VIX are annualised volatilities in per cent. RET is the weekly log return in per cent.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

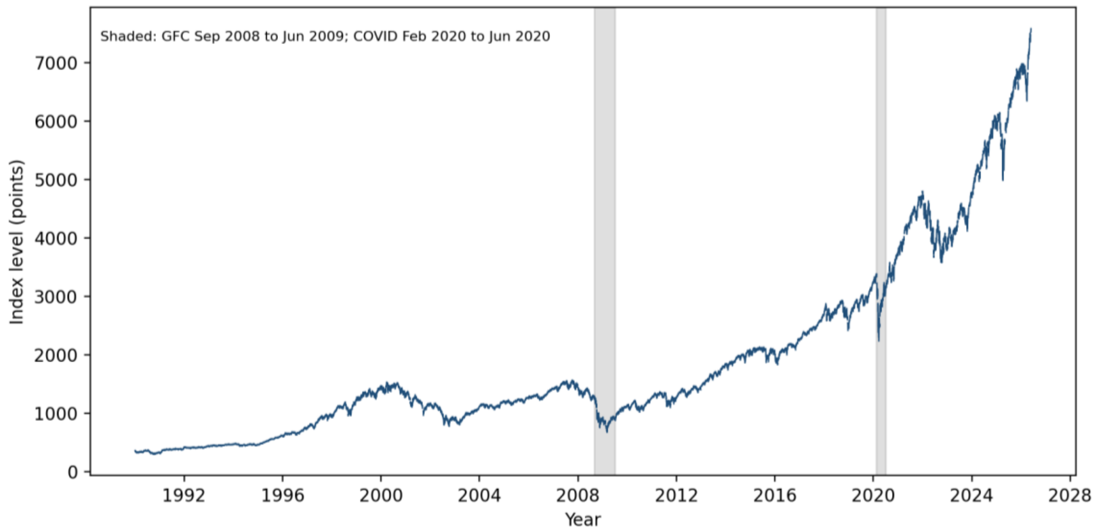

Figure 1. S&P 500 index level, 1990 to 2026, with crisis periods shaded
Note: Shaded bands mark the global financial crisis (September 2008 to June 2009) and the COVID-19 shock (February to June 2020).

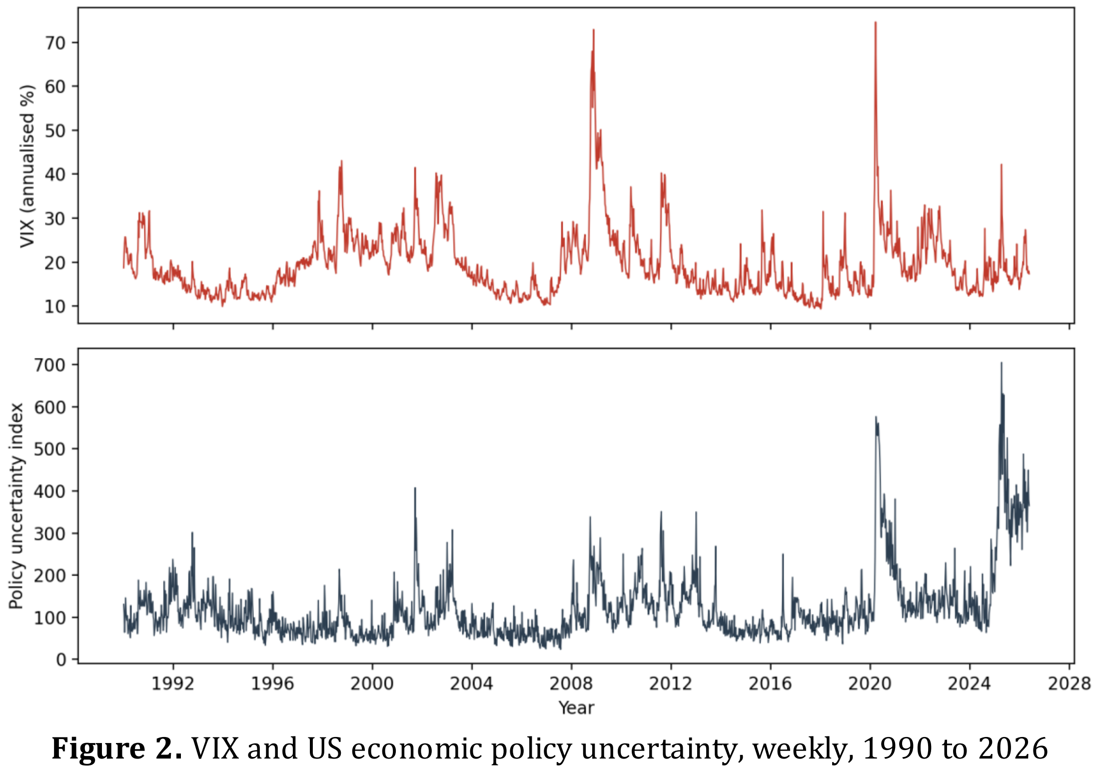

Figure 2. VIX and US economic policy uncertainty, weekly, 1990 to 2026
Note: Left axis: weekly mean VIX. Right axis: weekly mean economic policy uncertainty index.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

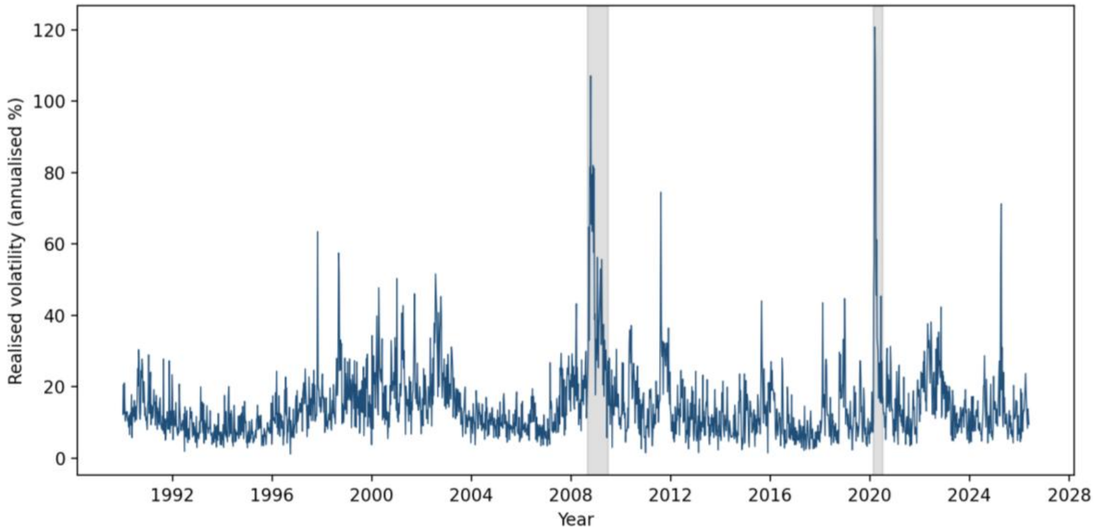

Figure 3. S&P 500 weekly realised volatility, annualised, 1990 to 2026

Note: Realised volatility is annualised from daily returns within each week.

4.2. Correlations and Multicollinearity

Table 3 presents the correlation matrix. Next-week realised volatility is most strongly correlated with the lagged VIX at 0.73 and with lagged realised volatility itself at 0.69, which already hints that these two variables will dominate the volatility models. The credit spread and oil volatility have moderate positive correlations with next-week volatility at 0.45 and 0.38, while policy uncertainty has a weaker correlation of 0.22. The variance inflation factors from the full predictor set stay below five, with the highest value of 4.56 attached to the VIX, so multicollinearity is present but not severe.

Table 3. Correlation matrix of predictors and outcomes

Variable	RET	RV	VIX	EPU	Credit	Term	Rate	Oil vol	RV (t+1)	Downside (t+1)
RET	1.0	-0.172	-0.153	0.028	-0.029	-0.012	0.005	-0.073	-0.316	-0.044
RV	-0.172	1.0	0.817	0.286	0.49	0.044	-0.083	0.439	0.693	0.206
VIX	-0.153	0.817	1.0	0.354	0.617	0.095	-0.089	0.478	0.725	0.212
EPU	0.028	0.286	0.354	1.0	0.18	0.046	-0.145	0.279	0.224	0.034
Credit	-0.029	0.49	0.617	0.18	1.0	0.497	-0.571	0.285	0.452	0.17
Term	-0.012	0.044	0.095	0.046	0.497	1.0	-0.63	-0.083	0.037	0.021
Rate	0.005	-0.083	-0.089	-0.145	-0.571	-0.63	1.0	-0.012	-0.078	-0.039
Oil vol	-0.073	0.439	0.478	0.279	0.285	-0.083	-0.012	1.0	0.377	0.085
RV (t+1)	-0.316	0.693	0.725	0.224	0.452	0.037	-0.078	0.377	1.0	0.392
Downside (t+1)	-0.044	0.206	0.212	0.034	0.17	0.021	-0.039	0.085	0.392	1.0

Note: Pearson correlations over the 1,898 week sample.

4.3. Volatility Models

Table 4 reports the four ordinary least squares models for next-week realised volatility. The baseline Model 1, which uses only the lagged return and lagged realised volatility, achieves an adjusted R-squared of 0.519, which confirms how much can be predicted from persistence alone. Adding the VIX in Model 2 raises the adjusted R-squared to 0.593 and the coefficient on the index is 0.651 with a very small standard error, which is a clear and economically meaningful improvement. Adding policy uncertainty on its own in Model 3 barely moves the fit,

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

lifting the adjusted R-squared only to 0.521, even though the policy uncertainty coefficient is statistically different from zero. The full Model 4 reaches an adjusted R-squared of 0.595, which is almost identical to Model 2, and within it the VIX keeps a strong coefficient of 0.612 while policy uncertainty, the credit spread, the short rate and oil volatility all become small and statistically indistinguishable from zero. The yield curve slope carries a significant negative coefficient of -0.689, which is consistent with a flatter or inverted curve preceding more turbulent weeks. The pattern across the four models is consistent and easy to summarise, because the VIX does almost all of the additional work and the other indicators add very little once it is present.

Table 4. OLS models for next-week realised volatility (RV_t1)

Term	Model 1	Model 2	Model 3	Model 4	HAR
Intercept	5.135*** (0.498)	-1.930*** (0.624)	4.644*** (0.523)	-1.988 (1.487)	2.325*** (0.471)
RET	-0.918*** (0.136)	-0.891*** (0.129)	-0.934*** (0.136)	-0.903*** (0.134)	
RV	0.658*** (0.040)	0.275*** (0.061)	0.645*** (0.040)	0.268*** (0.059)	0.408*** (0.088)
RV (4w)					0.333*** (0.093)
RV (13w)					0.100* (0.054)
VIX		0.651*** (0.056)		0.612*** (0.057)	
EPU			0.006*** (0.002)	-0.003 (0.003)	
Credit				0.860 (0.579)	
Term				-0.689*** (0.215)	
Rate				-0.100 (0.136)	
Oil vol				0.007 (0.009)	
Observations	1898	1898	1898	1898	1898
Adjusted R^2	0.5193	0.5929	0.5209	0.5953	0.5187
RMSE (in-sample)	7.3022	6.7186	7.2883	6.6898	7.3047

Note: Newey-West HAC standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

These results give clear support to the first hypothesis, since higher implied volatility is strongly and positively associated with next week's realised volatility. This association should be read with care, because the VIX is itself the market's forecast of volatility over the coming month, so a positive link is close to definitional. The result that matters for a forecaster is the incremental one, namely that adding the VIX to a model that already contains past realised volatility raises the adjusted R-squared by about seven points, and the out-of-sample tests in Section 4.5 confirm that this gain is genuine. The second hypothesis receives only weak support, because although policy uncertainty has the expected positive sign when it enters on its own, its contribution is tiny and it loses significance entirely in the full model once the VIX is included.

4.4. Downside Risk Models

Table 5 reports the three logistic regression models for the next-week downside event. In the baseline Model 5 the lagged realised volatility is a significant predictor of downside risk, which makes sense because turbulent weeks tend to cluster, and the model reaches an in-sample area under the curve of 0.709. Adding the VIX in Model 6 raises the area under the curve to 0.731,

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

and the index enters with a positive and significant coefficient while the coefficient on lagged realised volatility falls toward zero, which shows that the index absorbs much of the same information. The full Model 7 raises the in-sample area under the curve further to 0.752, and within it the VIX remains significant with a coefficient of 0.053 while the credit spread carries a positive coefficient of 0.305 that points in the expected direction but is not statistically significant. Figure 4 shows the receiver operating characteristic curves, which sit clearly above the diagonal of a non-informative classifier.

Table 5. Logistic regression models for the next-week downside event (DOWN_t1)

Term	Model 5	Model 6	Model 7
Intercept	-3.762*** (0.173)	-4.487*** (0.292)	-4.345*** (0.509)
RET	-0.003 (0.034)	-0.003 (0.034)	-0.006 (0.035)
RV	0.047*** (0.007)	0.011 (0.013)	0.017 (0.014)
VIX		0.062*** (0.019)	0.053** (0.026)
EPU			-0.002 (0.002)
Credit			0.305 (0.215)
Term			-0.232 (0.171)
Rate			-0.039 (0.070)
Oil vol			-0.005 (0.004)
Observations	1898	1898	1898
Pseudo R^2	0.0633	0.0768	0.0873
AUC (in-sample)	0.7092	0.731	0.7516

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10. AUC is in-sample.

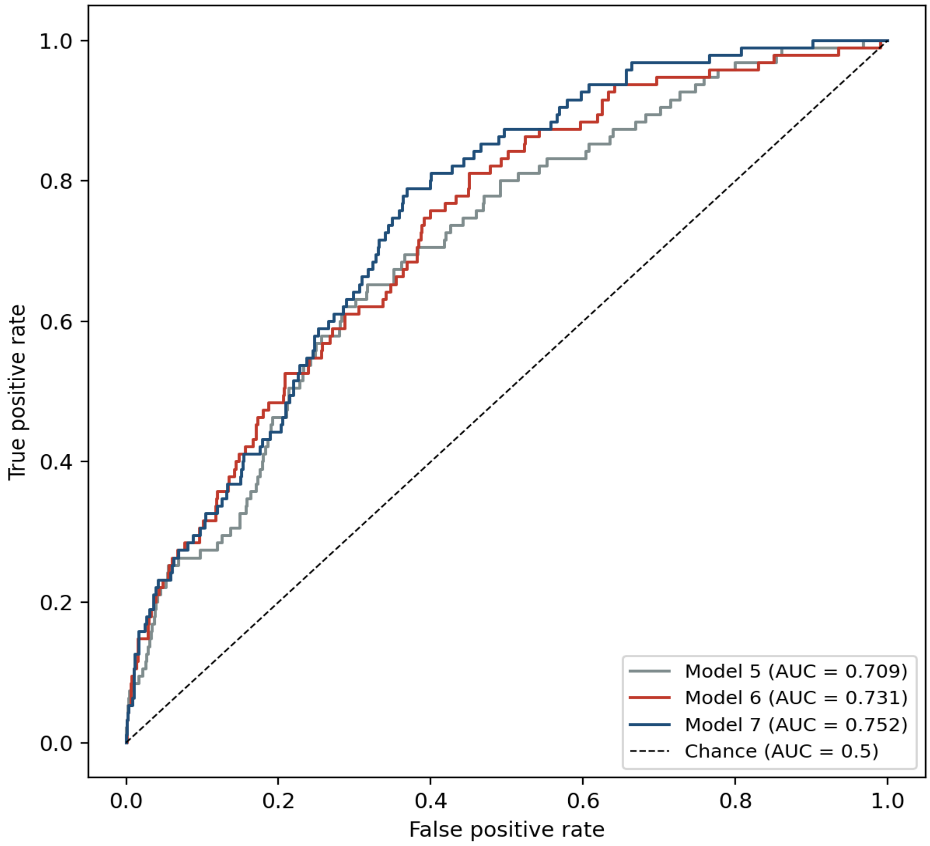

Figure 4. ROC curves for the downside risk models (in-sample)
Note: The diagonal marks a non-informative classifier with AUC of 0.5.

The third hypothesis, that credit stress indicators improve the prediction of downside events, receives only weak support from the in-sample results, because the credit spread has the right sign but does not reach significance once the other variables are present. The downside models therefore tell a similar story to the volatility models, in which the market-based fear gauge is the most reliable addition and the broader stress variables contribute less than their popularity might suggest.

4.5. Out-of-Sample Comparison

Table 6 brings the in-sample and out-of-sample results together and is the heart of the analysis, because it shows how the models would have performed in real forecasting use. For volatility the out-of-sample root mean squared errors are 7.85 for the baseline Model 1, 7.26 for Model 2 with the VIX, 7.85 for Model 3 with policy uncertainty, and 7.28 for the full Model 4. The ranking is informative, because the model with the VIX alone produces the lowest out-of-sample error and the full model does not improve on it and is in fact marginally worse. Policy uncertainty on its own delivers no out-of-sample gain over the baseline at all. The HAR benchmark from the volatility literature does not change this picture, because its out-of-sample root mean squared error of 7.93 is no better than the simple baseline and clearly worse than the model with the VIX. Figure 5 plots the out-of-sample predicted and actual volatility from the full model and shows that the model tracks the broad movements of volatility well but understates the sharpest spikes, which is a familiar feature of linear volatility forecasts.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

Table 6. Model comparison: in-sample fit and out-of-sample performance

Model	Target	Predictors (k)	In-sample fit	In-sample RMSE	Out-of-sample RMSE	In-sample AUC	Out-of-sample AUC
Model 1	Realised volatility	2	0.5193	7.3022	7.8467		
Model 2	Realised volatility	3	0.5929	6.7186	7.2603		
Model 3	Realised volatility	3	0.5209	7.2883	7.8461		
Model 4	Realised volatility	8	0.5953	6.6898	7.278		
HAR	Realised volatility	3	0.5187	7.3047	7.9285		
Model 5	Downside event	2	0.0633			0.7092	0.6485
Model 6	Downside event	3	0.0768			0.731	0.6743
Model 7	Downside event	8	0.0873			0.7516	0.6351

Note: OOS metrics use an expanding window starting after the first 520 weeks. Volatility models are scored by RMSE, downside models by AUC. HAR is the heterogeneous autoregressive benchmark using past 1, 4 and 13 week realised volatility.

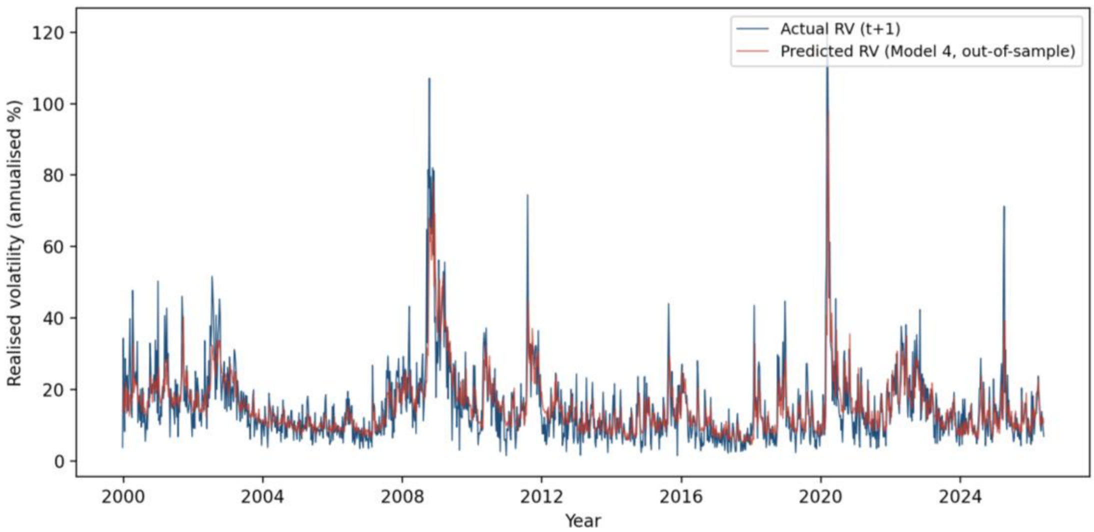

Figure 5. Out-of-sample predicted versus actual weekly realised volatility (full model)
Note: Predictions come from the expanding-window estimation of Model 4.

Table 7 reports formal tests of these differences rather than leaving them to the eye. The Diebold and Mariano test rejects equal predictive accuracy between the baseline and the VIX model, with a statistic of 4.16 and a p-value below 0.001, so the VIX genuinely improves the volatility forecast, and it also rejects equal accuracy between the HAR benchmark and the VIX model with a statistic of 4.63, so the VIX beats the standard HAR model as well. By contrast the test cannot separate the full model from the VIX-only model, with a statistic of -0.40 and a p-value of 0.69, which confirms that the wider uncertainty set adds nothing once the VIX is present.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

Table 7. Formal tests of out-of-sample forecast accuracy

Test	Comparison	Statistic	p-value
Diebold-Mariano (OOS squared error)	Baseline vs +VIX (Model 1 vs Model 2)	4.162	<0.001
Diebold-Mariano (OOS squared error)	+VIX vs Full (Model 2 vs Model 4)	-0.397	0.691
Diebold-Mariano (OOS squared error)	HAR vs +VIX (HAR vs Model 2)	4.634	<0.001
Diebold-Mariano (OOS squared error)	Baseline vs HAR (Model 1 vs HAR)	-0.374	0.708
Bootstrap AUC difference	+VIX vs baseline (Model 6 vs Model 5)	0.026	0.157
Bootstrap AUC difference	Full vs +VIX (Model 7 vs Model 6)	-0.039	0.080
Bootstrap AUC difference	Full vs baseline (Model 7 vs Model 5)	-0.012	0.677

Note: Diebold-Mariano tests compare squared-error loss for the volatility models (a positive statistic means the first model is worse). Bootstrap AUC tests use 2,000 paired resamples of the out-of-sample downside predictions.

The downside results out of sample are more cautionary still. The out-of-sample areas under the curve are 0.649 for the baseline Model 5, 0.674 for Model 6 with the VIX, and 0.635 for the full Model 7, so the leaner Model 6 has the highest point estimate while the full Model 7, which had the strongest in-sample fit, has the lowest. These differences are small relative to their sampling uncertainty. As Table 7 shows, the bootstrap 95 per cent confidence intervals are wide and overlapping, running from 0.585 to 0.712 for Model 5, from 0.611 to 0.736 for Model 6 and from 0.563 to 0.703 for Model 7, and the bootstrap test cannot reject equal performance between Model 6 and the baseline, with a p-value of 0.157. The one difference that approaches significance is that the full Model 7 is worse than the leaner Model 6, with a p-value of 0.080. The honest reading is therefore that no downside model is reliably better than the others out of sample, and that the gap between the strong in-sample fit of the full model and its weak out-of-sample standing is a sign of overfitting, in which the extra stress variables help the model to describe the past without helping it to predict the future. The fourth hypothesis, that a full uncertainty model improves forecasting performance, is therefore only partly supported, because for volatility the improvement over the baseline comes almost entirely from the VIX, and for downside risk the wider uncertainty set adds nothing reliable and, if anything, harms the forecast.

4.6. Robustness to Crisis Periods

Because volatility and downside events cluster in the two crises, the results might be driven by those weeks alone. The robustness checks remove the global financial crisis and the COVID-19 shock and re-estimate the full models. In the volatility model the VIX coefficient is essentially unchanged at 0.633 and remains significant at the one per cent level, while the adjusted R-squared falls from 0.595 to 0.485, which is expected because the crises contained much of the strongest signal. The credit spread, which was not significant in the full sample, becomes significant outside the crises with a coefficient of 0.945, which suggests that credit conditions may matter more in ordinary times than during the extreme weeks when the VIX dominates everything. In the downside model the VIX coefficient remains positive and significant at 0.078 when the crises are removed. The headline conclusion is that the value of the VIX does not depend on a few crisis weeks, which strengthens the case for treating it as a genuinely useful free indicator.

ISSN: 3080-9843 (Print), ISSN: 3080-9851 (Online)

A final check addresses the concern that the downside event is defined with a full-sample threshold. When the worst-5-per-cent cut is instead re-estimated on the training window at each step, so that the out-of-sample labels use no future information, the downside areas under the curve are 0.695 for Model 5, 0.714 for Model 6 and 0.659 for Model 7. The values are a little higher than in the main exercise, but the ranking is unchanged, with the VIX-based Model 6 ahead and the full Model 7 behind, so the earlier conclusion is not an artefact of how the threshold is set.

5. Discussion

The results offer a clear and consistent answer to the research question. Free public indicators can improve S&P 500 risk forecasts, but the improvement comes overwhelmingly from one indicator, the VIX, rather than from the broader collection of uncertainty and stress measures. This fits the literature on the information content of the VIX, which argues that option prices embed forward-looking expectations that backward-looking volatility cannot capture [2, 15, 17]. The strength and stability of the VIX coefficient across specifications and the crisis robustness checks indicate that it carries durable information about next-week risk.

The weaker performance of the other indicators deserves careful interpretation rather than dismissal. Policy uncertainty has the expected positive association with volatility when it enters on its own, which is consistent with the theory that policy noise raises risk premia and delays decisions [18, 19]. Its contribution fades once the VIX is present, which most likely reflects overlap, because periods of high policy uncertainty are also periods when option markets price in more volatility, so the two series move together and the market-based measure captures the information first. This interpretation is in keeping with the broad finding that macroeconomic and uncertainty variables often look promising in sample yet add little out of sample once the persistence of volatility is accounted for [6, 7, 28].

The credit spread results are more nuanced and arguably the most interesting secondary finding. In the full sample the credit spread does not significantly improve either target, yet once the two crises are removed it becomes a significant predictor of volatility. A plausible reading is that during the most extreme weeks the VIX moves so sharply that it absorbs almost all of the predictable variation, while in calmer times credit conditions carry independent information about building stress. This is consistent with the evidence that credit spreads track financing conditions and future activity [4] and with the idea that financial conditions shape the downside of the outcome distribution in particular [25]. The downside models echo this, because the credit spread always carries the expected positive sign even where it is not significant.

The gap between in-sample and out-of-sample performance for the downside target is the clearest practical warning in the study. The full model had the best in-sample classification yet the worst out-of-sample classification, which shows how easily a richer model can fit the past without forecasting the future. For a practitioner the message is that more indicators are not automatically better and that a parsimonious model anchored on the VIX is likely to be more reliable in real use. This caution mirrors the wider out-of-sample literature in which complexity is often punished rather than rewarded [28, 29].

It is also worth being clear about what the study does not claim. It does not claim that the uncertainty indicators are useless, because they may well be valuable at other horizons, for other assets or in combination with richer non-linear methods. Text and attention-based measures such as news-implied volatility, search-based attention and geopolitical risk have been shown to matter in other settings [22-24], and statistical uncertainty measures may capture episodes that newspaper indices miss [21]. The contribution here is narrower and practical, a transparent weekly test of whether these free series beat a simple baseline for two S&P 500 risk targets, and the answer is that the VIX does while the others mostly do not.

Several limitations should temper the conclusions. The sample begins in 1990, so it cannot speak to earlier decades, and the equity series is drawn from a free public source rather than from the licensed index data used in some studies. The realised volatility measure is built from daily rather than intraday returns, which is appropriate at this level but is noisier than a high-frequency estimator [11, 13]. The downside target is demanding for a binary model, because only 95 of the 1,898 weeks are downside events, which leaves about ten events for each predictor in the full logistic model and helps to explain why that model overfits. The models are linear and additive, so they cannot capture the interactions or threshold effects that a more advanced method might reveal, and although the forecast comparisons are backed by Diebold and Mariano [30] tests and bootstrap intervals, they still rest on a single expanding window with one chosen starting point, so the precise figures would shift a little under a different split. These choices keep the analysis clear and reproducible, which is the right balance for the question at hand, but they also mark out sensible directions for further work.

6. Conclusion

This study asked whether free uncertainty and stress indicators improve weekly forecasts of S&P 500 volatility and downside risk beyond what is already known from past returns and past volatility. Using 1,898 weekly observations from 1990 to 2026, a transparent set of nested ordinary least squares and logistic models, and an honest expanding window out-of-sample exercise, the answer is a qualified yes that rests almost entirely on a single indicator. The VIX improves the volatility forecast by a clear margin, raising the adjusted R-squared from 0.519 to 0.593 and lowering out-of-sample error in a way that the formal tests of predictive accuracy confirm, and it also beats a standard HAR benchmark. For downside risk no model is reliably better than the others out of sample, so the gains there are weaker than the in-sample fit suggests. The broader uncertainty and stress set, including policy uncertainty, the credit spread, the yield curve, the short rate and oil volatility, adds very little once the VIX is included, and for downside events the full model overfits. The value of the VIX survives the removal of the two largest crises, while the credit spread shows some independent value for volatility outside crisis periods.

For a student, a small fund or a risk manager working with free data, the practical lesson is encouraging in one respect and sobering in another. A single freely available series, the VIX, genuinely helps to anticipate next-week risk and is well worth using, yet adding more free indicators in the hope of a better forecast can backfire by encouraging overfitting. Future work could extend the analysis to non-linear and machine learning methods, to longer forecast horizons, to intraday realised volatility and to combinations of indicators. The present contribution is a careful and reproducible benchmark that shows where free uncertainty indicators help and where they do not, without overstating what they can do.

References

[1] Whaley, R. E. (2000). The investor fear gauge. The Journal of Portfolio Management, 26(3), 12–17. https://doi.org/10.3905/jpm.2000.319728

[2] Whaley, R. E. (2009). Understanding the VIX. The Journal of Portfolio Management, 35(3), 98–105. https://doi.org/10.3905/jpm.2009.319728

[3] Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

[4] Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. American Economic Review, 102(4), 1692–1720. https://doi.org/10.1257/aer.102.4.1692

[5] Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. The Journal of Finance, 46(2), 555–576. https://doi.org/10.1111/j.1540-6261.1991.tb02674.x

[6] Paye, B. S. (2012). 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. Journal of Financial Economics, 106(3), 527–546. https://doi.org/10.1016/j.jfineco.2012.06.006

[7] Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. Journal of Applied Econometrics, 27(6), 956–977. https://doi.org/10.1002/jae.2287

[8] Schwert, G. W. (1989). Why does stock market volatility change over time? The Journal of Finance, 44(5), 1115–1153. https://doi.org/10.1111/j.1540-6261.1989.tb02647.x

[9] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

[10] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

[11] Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579–625. https://doi.org/10.1111/1468-0262.00418

[12] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

[13] Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.003

[14] Giot, P. (2005). Relationships between implied volatility indexes and stock index returns. The Journal of Portfolio Management, 31(3), 92–100. https://doi.org/10.3905/jpm.2005.433714

[15] Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. The Review of Financial Studies, 22(11), 4463–4492. https://doi.org/10.1093/rfs/hhp080

[16] Drechsler, I., & Yaron, A. (2011). What's vol got to do with it. The Review of Financial Studies, 24(1), 1–45. https://doi.org/10.1093/rfs/hhq075

[17] Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. Journal of Econometrics, 183(2), 181–192. https://doi.org/10.1016/j.jeconom.2014.05.002

[18] Pástor, Ľ., & Veronesi, P. (2013). Political uncertainty and risk premia. Journal of Financial Economics, 110(3), 520–545. https://doi.org/10.1016/j.jfineco.2013.08.007

[19] Brogaard, J., & Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. Management Science, 61(1), 3–18. https://doi.org/10.1287/mnsc.2014.2001

[20] Bali, T. G., Brown, S. J., & Tang, Y. (2017). Is economic uncertainty priced in the cross-section of stock returns? Journal of Financial Economics, 126(3), 471–489. https://doi.org/10.1016/j.jfineco.2017.09.006

[21] Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. American Economic Review, 105(3), 1177–1216. https://doi.org/10.1257/aer.20131193

[22] Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. Journal of Financial Economics, 123(1), 137–162. https://doi.org/10.1016/j.jfineco.2016.09.013

[23] Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. The Review of Financial Studies, 28(1), 1–32. https://doi.org/10.1093/rfs/hhu060

[24] Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. American Economic Review, 112(4), 1194–1225. https://doi.org/10.1257/aer.20191823

[25] Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. American Economic Review, 109(4), 1263–1289. https://doi.org/10.1257/aer.20171929

[26] Ang, A., Chen, J., & Xing, Y. (2006). Downside risk. The Review of Financial Studies, 19(4), 1191–1239. https://doi.org/10.1093/rfs/hhj035

[27] Kelly, B., & Jiang, H. (2014). Tail risk and asset prices. The Review of Financial Studies, 27(10), 2841–2871. https://doi.org/10.1093/rfs/hhu039

[28] Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), 1455–1508. https://doi.org/10.1093/rfs/hhm002

[29] Campbell, J. Y., & 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. https://doi.org/10.1093/rfs/hhm003

[30] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599