News Implied Volatility and Disaster Concerns

Asaf Manela Alan Moreira*

First draft: September 2012

This draft: December 2015

Abstract

We construct a text-based measure of uncertainty starting in 1890 using front-page articles of the *Wall Street Journal*. News implied volatility (NVIX) peaks during stock market crashes, times of policy-related uncertainty, world wars and financial crises. In US post-war data, periods when NVIX is high are followed by periods of above average stock returns, even after controlling for contemporaneous and forward-looking measures of stock market volatility. News coverage related to wars and government policy explains most of the time variation in risk premia our measure identifies. Over the longer 1890–2009 sample that includes the Great Depression and two world wars, high NVIX predicts high future returns in normal times, and rises just before transitions into economic disasters. The evidence is consistent with recent theories emphasizing time variation in rare disaster risk as a source of aggregate asset prices fluctuations.

JEL Classification: G12, C82, E44

Keywords: Text-based analysis, implied volatility, rare disasters, equity premium, return predictability, machine learning

*Washington University in St. Louis, amanela@wustl.edu; and Yale University, alan.moreira@yale.edu. We thank anonymous reviewers, Fernando Alvarez, Jacob Boudoukh, Diego García (discussant), Armando Gomes, Gerard Hoberg, Bryan Kelly (discussant), Ralitsa Petkova (discussant), Jacob Sagi (discussant), Jesse Shapiro, Chester Spatt (discussant), Paul Tetlock (discussant), seminar participants at Ohio State and Wash U, and conference participants at the AFA meetings, NBER BE meetings, IDC Herzlia, SFS Cavalcade, Texas Finance Festival, and BFI Media and Communications for helpful comments.

1 Introduction

Looking back, people's concerns about the future more often than not seem misguided and overly pessimistic. Only when these concerns are borne out in some tangible data, do economists tip their hat to the wisdom of the crowds. This gap between measurement and the concerns of the average investor is particularly severe for rare events. In this case, concerns might change frequently, but real economic data often makes these concerns seem puzzling and unwarranted. This paper aims to quantify this "spirit of the times", which after the dust settles is forgotten, and only hard data remains to describe the period. Specifically, our goal is to measure people's perception of uncertainty about the future, and to use this measurement to investigate what types of uncertainty drive aggregate stock market risk premia.

We start from the idea that time-variation in the topics covered by the business press is a good proxy for the evolution of investors' concerns regarding these topics.[1] We estimate a news-based measure of uncertainty based on the co-movement between the front-page coverage of the *Wall Street Journal* and options-implied volatility (VIX). We call this measure News Implied Volatility, or NVIX for short. NVIX has two useful features that allow us to further our understanding of the relationship between uncertainty and expected returns: (i) it has a long time-series, extending back to the last decade of the nineteen century, covering periods of large economic turmoil, wars, government policy changes, and crises of various sorts; (ii) its variation is interpretable and provides insight into the origins of risk variation. The first feature enables us to study how compensation for risks reflected in newspaper coverage has fluctuated over time, and the second feature allows us to identify which kinds of risk were important to investors.

We rely on machine learning techniques to uncover information from this rich and unique text dataset. Specifically, we estimate the relationship between option prices and the frequency of words using Support Vector Regression. The key advantage of this method over Ordinary Least Squares is its ability to deal with a large feature space. We find that NVIX predicts VIX well out-of-sample, with a root mean squared error of 7.48 percentage points ($R^2 = 0.19$). When we replicate our methodology with realized volatility instead of VIX, we find that it works well even as we go decades back in time, suggesting newspaper word-choice is fairly stable over this period.[2]

[1]This idea is consistent with the Gentzkow and Shapiro (2006) empirically supported model of news firms.

[2]We analyze word-choice stability and measurement error in Section 2.3. One could potentially improve on

Asset pricing theory predicts that fluctuation in options implied volatility is a strong predictor of stock market returns as it measures fluctuation in expected stock market volatility (Merton, 1973), in the variance risk premium (Drechsler, 2008; Drechsler and Yaron, 2011), and in the probability of large economic disasters (Gabaix, 2012; Wachter, 2013; Gourio, 2008, 2012). Motivated by this work we study whether fluctuations in NVIX encode information about equity risk premia.

We begin by focusing on the post-war period commonly studied in the literature for which high-quality stock market data is available. We find strong evidence that times of greater investor uncertainty are followed by times of above average stock market returns. A one standard deviation increase in NVIX predicts annualized excess returns higher by 3.3 percentage points over the next year and 2.9 percentage points annually over the next two years.

We dig deeper into the nature of the uncertainty captured by NVIX and find three pieces of evidence that these return predictability results are driven by variation in investors' concerns regarding rare disasters as in Gabaix (2012), Wachter (2013) and Gourio (2008, 2012). First, we find that the predictive power of NVIX is orthogonal to risk measures based on contemporaneous or forward-looking measures of stock market volatility. Second, we use alternative option based measures, which are more focused on left tail risk, to estimate their news-based counterparts. Specifically, our news-based extensions of the variance premium (Bollerslev, Tauchen, and Zhou, 2009), the model free left tail risk measure of Bollerslev and Todorov (2011), and implied volatility slope give similar predictability results.

Interpretability, a key feature of the text-based approach, enables us to investigate what type of news drive the ability of NVIX to predict returns. We decompose the text into five categories plausibly related (to a varying degree) to disaster concerns: war, financial intermediation, government policy, stock markets, and natural disasters. We find that a large part of the variation in risk premia is related to wars (53%) and government policy (27%). A substantial part of the time-series variation in risk premia NVIX identifies is driven by concerns tightly related to the type of events discussed in the rare disasters literature (Rietz, 1988; Barro, 2006). We find that government-related concerns are related to redistribution risk, as our measure traces remarkably well tax-policy changes in the US. Interestingly, even though uncertainty regarding the stock mar-

this out-of-sample fit using financial variables (e.g. past volatility, default spreads, etc.) at the cost of losing the interpretability of the text-based index, which is central to our analysis.

ket itself—an NVIX component highly correlated with realized volatility—drive a substantial part of the variation in NVIX, this variation is not priced. By contrast, while concerns related to wars or government policy do not drive most of the variation in news implied volatility, they do drive most of its priced variation. These results suggest that time-varying disaster risk in particular is priced in the post-war US stock market.

Our paper is the first to extract information about aggregate uncertainty from news coverage using machine learning techniques. Other recent work uses a more human-centric approach to extract such information. Leading examples are the economic policy uncertainty index of Baker, Bloom, and Davis (2013) and the word-list-based measures of Loughran and McDonald (2011). We find that NVIX is unique in its ability to relate text with variation in aggregate risk premia.

We then extend our analysis to include the earlier and turbulent 1896–1944 period to directly test whether NVIX predicts economic disasters. According to the theory, a variable that measures disaster concerns should forecast not only returns but also disasters. We develop a Bayesian framework based on Nakamura, Steinsson, Barro, and Ursúa (2013) to estimate the exact timing of disasters. The estimated posterior probability goes to one during three clear and distinct disaster periods, all between the two major world wars. It also identifies several periods of near misses, when the posterior probability is below one, but increases sharply, such as the 2008 financial crisis. Consistent with the notion that NVIX encodes disaster concerns, NVIX predicts innovations in this posterior probability. A one standard deviation increase in NVIX predicts a 2.5 percentage points higher probability of a disaster over the next year. These results are robust to the inclusion of several controls for contemporaneous and forward-looking measures of stock market variance. Furthermore, the relationship between NVIX and future returns is strikingly similar to our post-war estimates, once we adjust the estimation for disasters that realized in the pre-war sample.

Our paper fits in a large literature that studies asset pricing consequences of large and rare economic disasters. At least since Rietz (1988), financial economists have been concerned about the pricing consequences of large events that happened not to occur in US data. Brown, Goetzmann, and Ross (1995) argues the fact we can measure the equity premium in the US stock market using such a long sample suggests that its history is special. Barro (2006) and subsequently Barro and Ursua (2008); Nakamura, Steinsson, Barro, and Ursúa (2013); Barro (2009) show that calibrations consistent with 20th century world history can make quantitative sense of equity premium point estimates in the empirical literature. Gabaix (2012), Wachter (2013), Gourio (2008), and Gourio (2012) further show that calibrations of a time-varying rare disaster risk model can also explain the amount of time-variation in the data.

A major challenge of this literature is whether those calibrations are reasonable. As Gourio (2008) puts it, "this crucial question is hard to answer, since the success of this calibration is solely driven by the large and persistent variation is the disaster probability, which is unobservable." We bring new data to bear on this question. We find that the overall variation in disaster probabilities used in calibrations such as Wachter (2013) line up well with our estimates. Our estimates, however, suggest substantially lower persistence than previously calibrated by Wachter (2013) and Gourio (2008, 2012). Moreover, we estimate that a 1 percentage point increase in the annual probability of a disaster increases risk premia by 1.16 percentage points. This effect on risk premia is remarkably close to the risk premia disaster sensitivity produced by Wachter (2013), where disaster magnitudes are calibrated to match the distribution of disasters in the Barro and Ursua (2008) cross-country data. We interpret this as evidence that the time-variation in disaster concerns measured by NVIX regards disasters of the same magnitude as studied in the rare disasters literature.

One motivation for our paper is the empirical fact that estimating aggregate risk-return trade-offs is a data intensive procedure. Indeed, Lundblad (2007) shows that the short samples used in the literature is the reason why research on the classic variance-expected return trade-off had been inconclusive. Testing the particular form of risk-return trade-off predicted by the time-varying disaster risk hypothesis is more challenging on two fronts; plausible measures of disaster risk are available for no more than two decades, and validation of these measures is even more challenging, since disasters are rare.

There is a large and fruitful literature that exploits the information embedded in option markets to learn about the structure of the economy. Drechsler (2008) proposes a theory where the VIX has information about degree of ambiguity aversion among investors. Drechsler and Yaron (2011) interpret it as a forward looking measure of risk. Bollerslev and Todorov (2011) use a model free approach to back out from option prices a measure of the risk-neutral distribution of jump sizes in the S&P 500 index. Bates (2012) shows that time-changed Lévy processes capture well stochastic volatility and substantial outliers in US stock market returns. Kelly and Jiang (2014) estimates a tail risk measure from a 1963-2010 cross-section of returns and find it is highly correlated with options-based tail risk measures. Backus, Chernov, and Martin (2011) present an important challenge to the idea that the "overpricing" of out-of-the-money put options can be explained by static rare disaster risk models. Seo and Wachter (2013) show, however, that this apparent inconsistency can be resolved in a model with time-varying disaster risk.

Our paper connects information embedded in VIX with macroeconomic disasters by extending it back a century, and by using cross-equation restrictions between disaster and return predictability regressions to estimate disaster probability variance and persistence. Importantly, by decomposing NVIX into word categories we add to this literature interpretable measures of distinct disaster concerns, and gain novel insights about the origins of risk premia variation.[3]

Broadly, our paper contributes to a growing body of work that applies text-based analysis to fundamental economic questions. Hoberg and Phillips (2010, 2011) use the similarity of company descriptions to determine competitive relationships. Tetlock (2007) documents that the fractions of positive and negative words in certain financial columns predict subsequent daily returns on the Dow Jones Industrial Average, and García (2013) shows that this predictability is concentrated in recessions. These effects mostly reverse quickly, which is more consistent with a behavioral investor sentiment explanation than a rational compensation for risk story. By contrast, we examine lower (monthly) frequencies, and find strong return and disaster predictability consistent with a disaster risk premium by funneling front-page appearances of all words through a first-stage text regression to predict economically interpretable VIX. The support vector regression we employ offers substantial benefits over the more common approach of classifying words according to tone (e.g. Loughran and McDonald, 2011). It has been used successfully by Kogan, Routledge, Sagi, and Smith (2010) to predict firm-specific volatility from 10-K filings.

The paper proceeds as follows. Section 2 describes the data and methodology used to construct NVIX. Section 3 tests the hypothesis that time-variation in uncertainty is an important driver of variation in expected returns in post-war US data, reports our main results and identifies time-varying disaster concerns as a likely explanation. Section 4 uncovers which concerns drive risk premia. Section 5 extends our analysis back to 1896 to directly test whether NVIX predicts economic disasters. Section 6 concludes.

[3] Sample size is especially important for studying rare events. An alternative approach to our long time-series is to study a large cross-section of countries (e.g. Gao and Song, 2013).

2 Data and Methodology

We begin by describing our unique news dataset and how we use it to construct news-based measures of option implied volatility. We then describe the standard asset pricing data we rely on to investigate the hypothesis that disaster concerns are priced in the US stock market.

We assume throughout that the choice of words by the business press provides a good and stable reflection of the concerns of the average investor. This assumption is quite natural and consistent with a model of a news firm which observes real-world events and then chooses what to emphasize in its report, with the goal of building its reputation. Gentzkow and Shapiro (2006) build a model along these lines and present a variety of empirical evidence consistent with its predictions. The idea that news media reflects the interests of readers is suggested in Tetlock (2007), empirically supported by Manela (2011), and used for estimation of the value of information in Manela (2014).

2.1 News Implied Volatility (NVIX)

Our news dataset includes the title and abstract of all front-page articles of the *Wall Street Journal* from July 1889 to December 2009. We focus on front-page titles and abstracts to make the data collection feasible, and because these are manually edited and corrected following optical character recognition, which improves their earlier sample reliability. We omit titles that appear daily.[4] Each title and abstract are separately broken into one and two word n-grams using a standard text analysis package that replaces highly frequent words (stopwords) with an underscore, and removes n-grams containing digits.[5]

We combine the news data with our estimation target, the implied volatility indices (VIX and VXO) reported by the Chicago Board Options Exchange. We use the older VXO implied volatility index that is available since 1986 instead of VIX that is only available since 1990 because it grants us more data and the two indices are 0.99 correlated at the monthly frequency.

We break the sample into three subsamples. The *train* subsample, 1996 to 2009, is used to

[4]We omit the following titles keeping their abstracts when available: 'business and finance', 'world wide', 'what's news', 'table of contents', 'masthead', 'other', 'no title', 'financial diary'.

[5]For example, the sentence "The Olympics Are Coming" results in 1-grams "olympics" and "coming"; and 2-grams "_ olympics", "olympics _", and "_ coming". We use ShingleAnalyzer and StandardAnalyzer of the open-source Apache Lucene Core project to process the raw text into n-grams. We have experimented with stemming and considering different degree n-grams and found practically identical results, but since this is the procedure we first used, we report its results throughout to get meaningful out-of-sample tests.

estimate the dependency between news data and implied volatility. The *test* subsample, 1986 to 1995, is used for out-of-sample tests of model fit. The *predict* subsample includes all earlier observations for which VIX is not available.[6]

We aggregate n-gram counts to the monthly frequency to get a relatively large body of text for each observation. Since there are persistent changes over our sample in the number of words per article, and the number of articles per day, we normalize n-gram counts by the total number of n-grams each month. We omit those n-grams appearing less than 3 times in the entire sample. Each month of text is therefore represented by $\mathbf{x}_t$, a $K = 468,091$ vector of n-gram frequencies, i.e.

$$x_{t,i} = \frac{\text{appearances of n-gram } i \text{ in month } t}{\text{total n-grams in month } t}.$$

We use n-gram frequencies to predict VIX v_t with a linear regression model

$$v_t = w_0 + \mathbf{w} \cdot \mathbf{x}_t + \upsilon_t \qquad t = 1 \ldots T \tag{1}$$

where $\mathbf{w}$ is a K vector of regression coefficients. Clearly $\mathbf{w}$ cannot be estimated reliably using least squares with a training time series of $T_{train} = 168$ observations.

We overcome this problem using Support Vector Regression (SVR), an estimation procedure shown to perform well for short samples with an extremely large feature space K.[7] While a full treatment of SVR is beyond the scope of this paper, we wish to give an intuitive glimpse into this method, and the structure that it implicitly imposes on the data. SVR minimizes the following objective

$$H(\mathbf{w}, w_0) = \sum_{t \in train} g_\epsilon (v_t - w_0 - \mathbf{w} \cdot \mathbf{x}_t) + c (\mathbf{w} \cdot \mathbf{w}),$$

[6] A potential concern is that since the *train* sample period is chronologically after the *predict* subsample, we are using new information, unavailable to those who lived during the *predict* subsample, to predict future returns. While theoretically possible, we find this concern empirically implausible because the way we extract information from news is indirect, counting n-gram frequencies. For this mechanism to work, modern newspaper coverage of looming potential disasters would have to use *less* words that describe old disasters. By contrast, suppose modern journalists now know the stock market crash of 1929 was a precursor for the great depression. As a result, they give more attention to the stock market and the word "stock" gets a higher frequency conditional on the disaster probability in our *train* sample than in earlier times. Such a shift would cause its regression coefficient to *underestimate* the importance of the word in earlier times. Such measurement error works against us finding return and disaster predictability.

[7] See Kogan, Levin, Routledge, Sagi, and Smith (2009); Kogan, Routledge, Sagi, and Smith (2010) for an application in finance or Vapnik (2000) for a thorough discussion of theory and evidence. We discuss alternative approaches in Section **??**.

where $g_\epsilon(e) = \max\{0, |e| - \epsilon\}$ is an "ϵ-insensitive" error measure, ignoring errors of size less than ϵ. The minimizing coefficients vector $\mathbf{w}$ is a weighted-average of regressors

$$\hat{\mathbf{w}}_{SVR} = \sum_{t \in train} (\hat{\alpha}_t^* - \hat{\alpha}_t)\, \mathbf{x}_t \tag{2}$$

where only some of the T_{train} observations' (dual) weights α_t and α_t^* are non-zero.[8]

SVR works by carefully selecting a relatively small number of observations called support vectors, and ignoring the rest. The trick is that the restricted form (2) does not consider each of the K linear subspaces separately. By imposing this structure, we reduce an over-determined problem of finding $K \gg T$ coefficients to a feasible linear-quadratic optimization problem with a relatively small number of parameters (picking the T_{train} dual weights α_t). The cost is that SVR cannot adapt itself to concentrate on subspaces of $\mathbf{x}_t$ (Hastie, Tibshirani, and Friedman, 2009). For example, if the word "peace" is important for VIX prediction independently of other words that appear frequently on the same low VIX months (e.g. "Tolstoy"), SVR would assign similar weight to both. Ultimately, success or failure of SVR must be evaluated by out-of-sample fit which we turn to next.

Figure 1 shows estimation results. Looking at the *train* subsample, the most noticeable observations are the LTCM crisis in August 1998, September 2002 when the US made it clear an Iraq invasion is imminent, the abnormally low VIX from 2005 to 2007, and the 2008 financial crisis. In-sample fit is quite good, with an $R^2(train) = 91\%$. The tight confidence interval around $\hat{v}_t$ suggests that the estimation method is not sensitive to randomizations (with replacement) of the *train* subsample. This gives us confidence that the methodology uncovers a fairly stable mapping between word frequencies and VIX, but with such a large feature space, one must worry about over-fitting.

However, as reported in Table 1, the model's out-of-sample fit over the *test* subsample is quite good, with $RMSE(test)$ of 7.48 percentage points ($R^2(test)$ of 19%). In addition to these statistics, we also report results from a regression of *test* subsample actual VIX values on news-based values.

[8] SVR estimation requires us to choose two hyper-parameters that control the trade-off between in-sample and out-of-sample fit (the ϵ-insensitive zone and regularization parameter c). Rather than make these choices ourselves, we use the procedure suggested by Cherkassky and Ma (2004) which relies only on the *train* subsample. We first estimate using k-Nearest Neighbor with $k = 5$, that $\sigma_v = 6.65$. We then calculate $c_{CM2004} = 50.74$ and $\epsilon_{CM2004} = 3.49$. We numerically estimate $\mathbf{w}$ by applying with these parameters the widely used SVM^{light} package (available at `http://svmlight.joachims.org/`) to our data.

We find that NVIX is a statistically powerful predictor of actual VIX. The coefficient on $\hat{v}_t$ is statistically greater than zero ($t = 4.01$) and no different from one ($t = -0.88$), which supports our use of NVIX to extend VIX to the longer sample.

2.2 NVIX is a Reasonable Proxy for Uncertainty

NVIX captures well the fears of the average investor over this long history. Noteworthy peaks in NVIX include the stock market crash of October and November 1929 and other tremulous periods which we annotate in Figure 2. Stock market crashes, wars and financial crises seem to play an important role in shaping NVIX. Noteworthy in its absence is the "burst" of the tech bubble in March 2000, thus not all market crashes indicate rising concerns about economic disasters. Our model produces a spike in October 1987 when the stock market crashed and a peak in August 1990 when Iraq invaded Kuwait and ignited the first Gulf War. This exercise gives us confidence in using the model to predict VIX over the entire *predict* subsample, when options were hardly traded, and actual VIX is unavailable.

We find it plausible that spikes in uncertainty perceived by the average investor coincide with stock market crashes, world wars and financial crises. Because these are exactly the times when NVIX spikes due to each of these concerns, we find it is a plausible proxy for investor uncertainty.

It is perhaps surprising that NVIX is relatively smooth during the Great Depression when NVIX increases from about 25% to 30%, peaking at 40% on October 1929. We note, however, that like options-implied volatility, NVIX is a forward-looking measure of uncertainty and will be naturally smoother than backward-looking realized volatility, which mechanically spikes during disaster realizations. Alternatively, this could happen because measurement error attenuates NVIX, a concern we turn to next.

2.3 Word-choice Stability and Measurement Error

We assume throughout that the choice of words by the business press provides a good and stable reflection of the concerns of the average investor. Otherwise, the type of machine learning techniques we use to interpret text would produce noisy estimates of implied volatility. Such measurement error would bias our predictability results toward zero.

One concern is that the issues worrying investors change over time. For example, the "Dust

Bowl" was a uniquely salient feature of the 1930s, which featured severe dust storms, drought, and agricultural damage. Since this type of event is unlikely to concern modern day investors enough to make front page news during our training sample, we might measure with error the perception of uncertainty that prevailed during the thirties. Technically, to estimate reliably the relationship between specific sources of aggregate uncertainty and word usage of the business press, we require variation in both during our train subsample. We choose to train on the recent sample, and test on the earlier one, so we can get a sense of out-of-sample fit when we go even further back in time. This choice is not innocuous. If we were to reverse the order and train on the earlier sample, our text regression would miss important variation due to the financial crisis of 2008, and instead focus on the stock market crash of 1987.

A related concern is that the meaning of certain words or phrases used by the business press has changed considerably over our long sample. For example, the mapping from the 2-gram "Japanese navy" to investor concerns about disaster risks in the 1940s is likely different than in the 2000s. Ideally, we would only consider more common phrases with a stable meaning, such as "war". The techniques we use are, however, designed to avoid such overfitting pitfalls, and proved successful in related settings (Antweiler and Frank, 2004; Kogan, Routledge, Sagi, and Smith, 2010).

Nonetheless, we wish to quantify how measurement error changes when moving from the *test* subsample to the *predict* subsample, but VIX is unavailable during this earlier period. Instead, we repeat the same estimation procedure over the same *train* subsample as before, but replace VIX with realized volatility as the dependent variable of the SVR in Equation (1).

We find that our predictive ability over the long sample is quite stable. Table 2 reports several different measures of realized volatility fit to news data over the three subsamples. The most natural measure of fit is root mean squared error of the text regression (*RMSE SVR*), according to which, measurement error in the *predict* subsample is only slightly higher than in the *test* subsample. RMSE increases from 9.6 percent to 10.7 percent annualized volatility. These results suggest only a modest increase in measurement error of NVIX as we extend VIX further back to times the index did not exist.

2.4 Asset Pricing Data

We use two different data sources for our stock market data. We use the CRSP total market portfolio for the period from 1926 to 2009 and the Dow Jones index from Global Financial Data, available monthly from July 1896 to 1926. We refer to this series as "market" returns. Results are similar if we use the Dow Jones index throughout. We also use Robert Shiller's time series of aggregate S&P 500 earnings from his website. We chose to use this data to run our predictability tests because this index is representative of the overall economy and goes back a long way. We use daily returns on the CRSP total market portfolio and the Dow Jones index to construct proxies for realized volatility, which we use to explore alternative explanations. To compute excess returns we use the one-month t-bill rate to measure the risk free rate, and when it is unavailable we use yields on 10 year US government bonds from Global Financial Data. We use the difference between Moody's Baa and Aaa yields to measure credit spreads. This data is only available after 1919. We use the VXO and VIX indices from the CBOE. They are implied volatility indices derived from a basket of option prices on the S&P 500 (VIX) and S&P 100 (VXO) indices. The VIX time series starts in January 1990 and VXO starts in January 1986. The LT measure of Bollerslev and Todorov (2011) was kindly provided to us by the authors. We use Option Metrics data to construct a measure of the slope of the implied volatility curve for the S&P 500 index.

3 Post-War Compensation for Risks Measured by NVIX

In this section we test the hypothesis that time-variation in uncertainty is an important driver of variation in expected returns on US equity over the post-World War II, 1945 to 2009 sample. During this period, commonly studied in the literature, the US experienced no economic disasters of the magnitude of a Great Depression or a World War. High-quality stock market data is available for this period. We start with our main findings that NVIX predicts returns. We then show that stochastic volatility is not behind this result, and that our results survive the inclusion of several predictors of stock market returns, and alternative text-based measures of uncertainty. Finally, we extend our methodology to other alternative measures of tail risk and find similar results.

3.1 NVIX Predicts Returns

Asset pricing models with time-varying risk premia predict that times when risk is relatively high would be followed by above average returns on the aggregate market portfolio. For example, the dynamic risk-return tradeoff of Merton (1973) predicts a linear relation between the conditional expected excess return on the market and its conditional variance, as well as its conditional covariance with other priced risk factors. The more recent time-varying rare disaster models, predict a linear relationship between expected excess returns and the variance premium, which is linear in the time-varying probability of a rare disaster (e.g. Gabaix, 2012). Therefore, our main tests try to explain future excess returns on the market portfolio at various horizons with lagged forward-looking measures of risk as measured by NVIX squared. We place our measure in variance space because in all the above-mentioned models, risk premia are linear in variances as opposed to standard deviations.[9] To alleviate any concerns about news-based measures that rely on weekend news coverage not yet priced in the stock market, we skip a month to err on the side of caution. Throughout the paper we report both Newey and West (1987) standard errors with the same number of lags as the forecasting window, and bootstrapped standard errors based on Murphy and Topel (2002) that further account for the fact that our main regressor NVIX is estimated in a first stage. For a complete discussion of standard errors see Appendix A.1.

The last two columns of Table 3 show that in the short-sample for which option prices are available, the ability of VIX to predict returns is statistically rather weak. In the sample for which VIX is available, the implied volatility index predicts excess returns in the six months to twelve months horizons. If we consider a slightly longer sample for which the VXO implied volatility index on the S&P 100 is available, the evidence for return predictability becomes weaker. Would these results change if we had a longer sample of such forward-looking measures of uncertainty?

While we do not have new options data to bring to bear, we use NVIX to extrapolate these options-based measures of uncertainty back in time. NVIX largely inherits the behavior of VIX and VXO in sample periods where both are available. Point estimates and standard errors are quite similar, especially for the VIX sample. This is hardly surprising, because NVIX was constructed to fit these implied volatility indices, though we only use post 1995 data for NVIX estimation.

[9] The results are very similar in terms of statistical significance and economic magnitude if we use NVIX instead.

The main advantage of using NVIX, however, is the ability to consider much longer samples. The first two columns of Table 3 reports our main results for two alternative extended sample periods. In the first column we see that return predictability for the entire post-war period going from 1945 to 2009 is well estimated with larger point estimates relative to the VIX sample. From six months to twenty-four months horizons the coefficients are statistically significant at the 1 to 5 percent level, unlike for the VIX sample. The second column reports results for the sample period for which we did not use any in-sample option price data. Out-of-sample, estimates are even larger and statistically significant at one to twenty-four months horizons.

We interpret the extended sample results as strong evidence for the joint hypothesis that NVIX measures investors' uncertainty and that time-variation in uncertainty drives expected returns. The coefficient estimates imply substantial predictability with a one standard deviation increase in $NVIX^2$ leading to $\sigma_{NVIX^2} \times \beta_1 = 20.5 \times 0.16 = 3.3\%$ higher excess returns in the following year. Unsurprisingly, R-squares are small and attempts to exploit this relationship carry large risks even for a long-run investor. Annualized forecasting coefficients are stable across forecasting horizons.

3.2 Alternative Text-based Approaches

We estimate the relationship between news coverage, volatility, and returns using support vector regression (1). SVR overcomes the main challenge, which is the large dimensionality of the feature space (number of unique n-grams). Our approach lets the data speak without much human interaction. Two alternative approaches have been suggested by previous literature.

One popular approach is to create a topic-specific compound full-text search statement and counts the resulting number of articles normalized by a measure of normal word count. The result is a univariate time-series that can be used in a least squares regression. An advantage of this approach is that resulting articles are highly likely to be related to the specific topic. However, this approach relies on the econometrician's judgment, unlike our approach which relies on an objective measure of success (VIX). Since out-of-sample fit is paramount in our paper, we find the text regression superior for our purposes. A leading example of this approach is the news-based economic policy uncertainty index (EPU) proposed in Baker, Bloom, and Davis (2013). In Column 2 of Table 4 we horse race our measure against their EPU measure. Comparing with the univariate specification (Column 1) we see that EPU does not increase the regression fit, and the predictability coefficients on NVIX are virtually unchanged. In unreported results we find that EPU does not predict returns even in a univariate specification. Evidently, these measures capture distinct pieces of information. While NVIX captures variation in uncertainty that is priced by the aggregate stock market and relevant for expected returns, EPU does not.

A second approach classifies words into dictionaries or word lists that share a common tone. One then counts all occurrences of words in the text belonging to a particular word list, again normalized by a measure of normal word count.[10] An advantage of this approach is that it reduces the feature space from the number of n-grams to the number of word lists. One disadvantage is that words within a list are equally-weighted. Thus the words 'war' and 'yawn' might count the same, even if the importance of their appearance on the front page of a newspaper is quite different.

A recent contribution by Loughran and McDonald (2011) develops a negative word list, along with five other word lists, that reflect tone in financial text better than the widely used Harvard Dictionary and relate them to 10-K filing returns. We applied the Loughran and McDonald (2011) methodology to our sample of articles. We tried both tf (proportional weights) and tf.idf weights of words appearing in their Negative, Positive, Uncertainty, Modal Strong, and Modal Weak word lists. Table 4 reports return predictability regressions on the scores of each word list together with NVIX.[11] Most lists add no predictive power. Only Uncertainty and Modal Weak using proportional weights improve on the univariate NVIX specification. The NVIX regression coefficient barely changes across specifications. We conclude that SVR is better for our purposes given our data.

3.3 Stochastic Volatility Does Not Explain These Results

We next dig deeper into the nature of the priced uncertainty captured by NVIX. One potential explanation for the ability of NVIX to predict returns is that NVIX measures variation in current stock market volatility (Merton, 1973). According to this hypothesis, NVIX predicts returns because investors demand higher expected returns during more volatile periods.

We test this hypothesis using lagged realized variance as well as five alternative variance forecasting models, gradually adding additional predictors, such as additional realized variance lags,

[10]Examples of this approach can be found in Antweiler and Frank (2004), Tetlock (2007), Engelberg (2008), and Tetlock, Saar-Tsechansky, and Macskassy (2008).

[11]The intermediate step of regressing VIX on the scores is unnecessary here because the predicted value of VIX would just be a constant multiplying the raw word list score.

the price-to-earnings ratio, NVIX, and the credit spread. The last line of Table 5 compares the ability of the alternative variance forecasting models to predict future variance.

Table 5 shows that the coefficient on NVIX is about the same and that standard errors either decrease or only slightly increase when we control for realized or expected variance. Note that its coefficient does not change even after we add NVIX to the variance forecasting model (model 4). This establishes that NVIX embeds priced information that is largely orthogonal to any information NVIX or other standard predictor variables contain regarding future volatility. In Appendix A.4, we also find that the predictive power of NVIX is robust to the inclusion of previously suggested return predictors like the price-to-earnings ratio and credit spreads.

3.4 Alternative Measures of Uncertainty Focused on Tail Risk

In Table 6 we replicate our analysis using alternative measures of uncertainty, which are more focused on left tail risk, controlling for expected future variance. For each of these measures we reproduce the methodology we applied to VIX. The first column reproduces our main results. In the second column is VIX premium ($= VIX_t^2 - E_t[Var(R_{t+1})]$), where $E_t[Var(R_{t+1})]$ is constructed using an AR(1) for realized variance (Bollerslev, Tauchen, and Zhou, 2009). In the third column is the options-based and model free Left-Tail (LT) measure of Bollerslev and Todorov (2011). In the fourth column is the slope of the option implied volatility curve, constructed using 30-day options from Option metrics. Table A.1 reports raw correlations between these measures of tail risk.

These alternative measures of news implied uncertainty yield similar predictability results. The direction of return predictability is consistent with the hypothesis that the predictability is driven by time-varying disaster concerns. When tail-risk is high, as measured by any of the four alternative measures, average returns are higher going forward.

All of these measures in one way or another take higher values when options that pay off in bad states of the world are relatively expensive. These options can be expensive because investors' attitudes towards these states take a turn for the worse, as in the time-varying Knightian uncertainty model of Drechsler (2008), or because the objective probability that these states occur increases, as in time-varying rare disaster models (Gourio, 2008, 2012; Gabaix, 2012; Wachter, 2013). In either case, NVIX appears to capture concerns related to tail risk.

4 Origins of Uncertainty Fluctuations

In this section, we lever the text-based feature of our uncertainty measure to gain novel insights into the origins of uncertainty fluctuations. The results in Section 3 imply that priced variation in NVIX is unrelated with standard measures of stock market risk, and likely to be related to fluctuations in tail risk. Guided by this evidence, we decompose our uncertainty measure into five interpretable categories meant to capture different types of shocks: Government, Financial Intermediation, Natural Disasters, Stock Markets and War. We find that a substantial amount of risk premia variation is driven by war and government related concerns.

4.1 Important Words

We calculate the fraction of NVIX variance that each word drives over the *predict* subsample. Define $\hat{v}_t(i) \equiv x_{t,i}w_i$ as the value of VIX predicted only by n-gram $i \in \{1..K\}$. We construct

$$h(i) \equiv \frac{Var(\hat{v}_t(i))}{\sum_{j\in K} Var(\hat{v}_t(j))} \tag{3}$$

as a measure of the n-gram specific variance of NVIX.[12] Table 7 reports $h(i)$ for the top variance driving n-grams and the regression coefficient w_i from the model (1) for the top variance n-grams. Note that the magnitude of w_i does not completely determine $h(i)$ since the frequency of appearances in the news interacts with $\mathbf{w}$ in (3).

Clearly, when the stock market makes an unusually high fraction of front page news it is a strong indication of high implied volatility. The word "stock" alone accounts for 37 percent of NVIX variance. Examining the rest of the list, we find that stock market-related words are important as well. This should not be surprising since when risk increases substantially, stock market prices tend to fall and make headlines. War is the fourth most important word and accounts for 6 percent.

4.2 Word Categorization

We rely on the widely used WordNet and WordNet::Similarity projects to classify words.[13] WordNet is a large lexical database where nouns, verbs, adjectives and adverbs are grouped into sets of

[12] Note that in general $Var(\hat{v}_t) \neq \sum_{j\in K} Var(\hat{v}_t(j))$ due to covariance terms.

[13] WordNet (Miller, 1995) is available at http://wordnet.princeton.edu. WordNet::Similarity (Pedersen, Patwardhan, and Michelizzi, 2004) is available at http://lincoln.d.umn.edu/WordNet-Pairs.

cognitive synonyms (synsets), each expressing a distinct concept. We select a number of root synsets for each of our categories, and then expand this set to a set of similar words which have a path-based WordNet:Similarity of at least 0.5.

Table 8 reports the percentage of NVIX variance ($= \sum_{i \in C} h(i)$) that each n-gram category drives over the *predict* subsample. Stock market related words explain over half the variation in NVIX. War-related words explain 6 percent. Unclassified words explain 36 percent of the variation. Clearly there are important features of the data, among the $467,745$ unclassified n-grams that the automated SVR regression picks up. While these words are harder to interpret, they seem to be important in explaining VIX behavior in-sample, and predicting it out-of-sample.

Each NVIX component can be interpreted as a distinct type of disaster concern. Figure 3 plots each of the four NVIX categories responsible for more of its variation to provide some insight into their interpretation. We omit the easily interpretable Natural Disasters category because it generates a negligible amount of NVIX variation.

The NVIX Stock Markets component has a lot to do with stock market volatility as shown in Figure 3a. Attention to the stock market as measured by this component seems to spike at market crashes and persist even when stock market volatility declines. This component likely captures proximate concerns about the stock market that have other ultimate causes, but can also capture concerns with the market itself.

Wars are clearly a plausible driver of disaster risk because they can potentially destroy a large amount of both human and physical capital and redirect resources. Figure 3b plots the NVIX War component over time. The index captures well the ascent into and fall out of the front-page of the *Journal* of important conflicts which involved the US to various degrees. A common feature of both world wars is an initial spike in NVIX when war in Europe starts, a decline, and finally a spike when the US becomes involved.

The most striking pattern is the sharp spike in NVIX in the days leading up to US involvement in WWII. The newspaper was mostly covering the US defensive buildup until the Japanese Navy's surprise attack at Pearl Harbor on December 1941. Following the attack, the US actively joined the ongoing War. NVIX War jumps from 0.75 in November to 2.47 in December and mostly keeps rising. The highest point in the graph is the Normandy invasion on June 1944 with the index reaching 3.83. The *Journal* writes on June 7, 1944, the day following the invasion: "Invasion of the continent of Europe signals the beginning of the end of America's wartime way of economic life." Clearly a time of elevated disaster concerns. Thus, NVIX captures well not only whether the US was engaged in war, but also the degree of concern about the future prevalent at the time.

Policy-related uncertainty as captured by our Government component tracks well changes in the average marginal tax rate on dividends as shown in Figure 3c. An important potential disaster from a stock market investor perspective is expropriation of ownership rights through taxation. While in retrospect, a socialist revolution did not occur in the US over this period, the probability of major redistributive policy changes could have been elevated at times.

Financial Intermediation-related NVIX spikes when expected, mostly during financial crises. Figure 3d shows that the Intermediation component is high during banking crises identified by Reinhart and Rogoff (2011), but also during other periods when bank failures were high, such as the late 1930s and early 1970s. Apparent in the figure are the panic of 1907, the Great Depression of the 1930s, the Savings & Loans crisis of the 1980s and the Great Recession of 2008.

4.3 Which Concerns Drive Risk Premia Variation?

We report a text-based decomposition of risk premia variation in Table 9. The shares of risk premia variation due to each of the categories is in parentheses. At the yearly horizon, Government (57%) and War (17%) related concerns capture the bulk of the post-war variation in risk premia. Both categories have a statistically reliable relation with future market excess returns. Concerns related to Financial Intermediation (0.7%), Stock Markets (0.3%), and Natural disasters (5%), account for some of the variation in expected returns, but the relationship is statistically unreliable. The harder to interpret orthogonal residual accounts for 19% of the variation.

During the post-war sample, war-related concerns explain a substantial part of the variation in risk premia. This is somewhat surprising to a 21st century economist who knows that the US economy did not contract sharply during any of its 1945–2009 military conflicts. We stress again, however, that NVIX captures the concerns prevalent at the time, without the benefit of hindsight. During the 1896–1944 period, which included two world wars, war-related concerns explain a much larger 67 percent share of this variation, or 53 percent in the full sample. A substantial part of the variation in risk premia is therefore unequivocally related to disaster concerns.

Government-related concerns allows for a wider range of potential interpretations. Work by

Pastor and Veronesi (2012), Croce, Nguyen, and Schmid (2012), and Baker, Bloom, and Davis (2013) emphasizes the role of policy-related uncertainty in inducing volatility and reducing asset prices in the recent period. We find that policy-related uncertainty explains a substantial part of risk-premia variation, but not in the early sample. This finding is consistent with an increasing role for the government in the aftermath of the Great Depression and World-War II.

One might argue that policy-related uncertainty is a very different type of risk than the rare disaster risk that the macro-finance literature has in mind. However, we find the tight relation between our government concerns measure and the evolution of US capital taxation shown in Figure 3c suggestive that our measure captures concerns related to expropriation risk. Not the typical cash-flow shock we use to model risk, but from the average capital holder perspective, a sudden sharp rise in taxes is a disaster. These results suggest that we may need to go beyond representative agent models to fully account for variation in risk premia.

Just as important, this result shows that most of variation in news implied volatility that is *priced* in the stock market is due to disaster concerns. The fact that a substantial fraction of the variation in risk premia over the last century is due to concerns related to wars and taxation, strongly suggests that risk premia estimates likely reflect the special realization of history the US happened to experience during this period (Brown, Goetzmann, and Ross, 1995).

Stock Markets-related concerns are not reliably related to future returns. Figure 3a shows that these concerns track well the time-series of realized volatility. Common sense and theory predicts that investors pay more attention to the stock market in periods of high volatility (Abel, Eberly, and Panageas, 2007; Huang and Liu, 2007). While these concerns of the stock market with itself explain about half the variation in NVIX, we find that this variation is not priced.

We were surprised to find that Financial Intermediation does not account for much of the time-variation in risk premia in our data. This was puzzling to us because the largest event in the sample we estimate NVIX is the 2007–2008 financial crisis. We think there are different possible conclusions from this evidence: it could be that our measure of uncertainty fails to pick up concerns related to the intermediary sector appropriately. However, Figure 3d strongly suggests that our measure gets the timing of the major financial events right. For example, during the great depression the intermediation measure peaks in 1933, three years after NVIX peaks. This timing lines up exactly with the the declaration of a national banking holiday and with the peak in bank failures (Cole and

Ohanian, 1999). A second possibility is that financial crises are intrinsically different since they are liability crises, essentially credit booms gone bust (Schularick and Taylor, 2009; Krishnamurthy and Vissing-Jorgensen, 2012). Reinhart and Rogoff (2011) suggests that financial crises are the result of complacency and a sense of "this time is different" among investors, i.e. financial crises happen only when investors are not concerned about financial intermediaries. Moreira and Savov (2013) build a macroeconomic model consistent with the notion that financial crises happen when investors perceive risk to be low, and predict that times of high intermediary activity are periods of low risk premia. The fact that Financial Intermediation does not account for much of the time-variation in risk premia in our data is consistent with our measure picking up financial intermediary activity during normal times, and concerns related to financial intermediaries during financial crises.

Our fifth category, Natural Disasters, also fails to predict returns. This is somewhat expected because we perceive as unlikely that there is time-variation in the likelihood of natural disasters at the frequencies we examine. Even though a large fraction of NVIX variation is not interpretable, as the overwhelming majority of words are unclassified, this residual component explains at most 20% of the variation in risk premia at annual frequencies. Our ex-ante chosen categories seem to do a good job of capturing the concerns that impact risk premia, but there is still a non-trivial fraction of risk premia left unexplained.

Taken together these results paint a novel picture of the origins of aggregate fluctuations. Of the roughly 4% ($= \sqrt{R^2 \sigma^2_{Retruns}} = \sqrt{0.063 \times 0.16^2}$) a year variation in risk premia news implied volatility can measure (Table 9, column 6), about half is driven by war concerns, tightly related to the type of disasters that motivates the rare disaster literature. An additional 27 percent of this variation is plausibly related to expropriation risk, which is quite different from the cash-flow shocks usually studied in rare disaster models.

5 A Century of Disaster Concerns

We extend our analysis to include the earlier 1896–1944 sample to further evaluate the time-varying disaster risk hypothesis. The occurrence of the Great Depression and the two world wars allows us to directly test whether NVIX has information regarding the likelihood of a disaster, and how disaster realizations impact the predictability pattern documented above.

We develop a formal methodology to empirically identify economic disasters. We then test whether NVIX encodes forward-looking information regarding disaster realizations, and whether a similar relationship between NVIX and future returns exists in the pre-war sample.

Consistent with time-varying rare disaster models, we find that NVIX is abnormally high up to 12 months before a disaster. Moreover, once we adjust our estimation to take into account disaster realizations in the pre-war sample and their persistence, the relationship between NVIX and future returns is strikingly similar to our post-war estimates.

5.1 Identifying Rare Disasters

Before we can say anything about the ability of NVIX to predict disasters, we need to identify disasters and their exact timing. We formally measure disasters using a Bayesian framework in the spirit of Nakamura, Steinsson, Barro, and Ursúa (2013), which generates endogenous estimates of the timing and length of disasters. The Nakamura, Steinsson, Barro, and Ursúa model can be viewed as a disaster filter. Just like a business-cycle filter isolates business cycle movements in output, their model isolates consumption movements attributable to disasters. Here, we provide only an intuitive description and relegate its details to Appendix A.2.

Our main contribution over Nakamura, Steinsson, Barro, and Ursúa (2013) is to extend their consumption growth-based framework to include also stock market returns as an additional signal about the state of the economy, to more precisely determine disaster arrival times. Importantly, stock market drops are necessary but not sufficient for disaster identification. Specifically, the model interprets large negative returns as more likely to indicate transitions into disaster, if preceding volatility is low and future consumption growth is persistently negative. Large negative returns that are not followed by drops in economic activity are interpreted as a mix of increases in volatility and unusually large negative return realizations. This extension allows us to pinpoint the timing of regime changes even in the earlier part of our sample, when consumption is only available annually. For example, annual consumption growth in 1929 was a healthy 3% followed by a contraction of 6.4% in 1930. Without stock market data and more coarse consumption data it is not possible to time the start of the Great Depression. Our framework interprets the sharp drop in the stock market in October 1929, together with the fact that consumption growth decreases sharply, as a high likelihood that the economy transitioned to a disaster in October 1929.

We use the filtered probability of a disaster that emerges from this Bayesian framework to evaluate whether NVIX predicts disasters. Figure 4a shows the posterior probability that the economy is in a disaster state from the econometrician's perspective. We identify three distinct disaster periods: two disasters during the period known as the Great Depression, October 1929 to January 1933, and then June 1937 to 1939, as well as a four year period that starts with the US entry into WWI in 1917 and lasts until the end of the 1920–1921 depression. Other periods stand out as near misses, like the 2007–2009 financial crisis, the Volcker recession of the early 1980s, the oil shock of the 1970s, and the US entry into WWII.

Figure 4b shows the probability that the economy transitioned into a disaster state in a particular month. This probability of state transition is our empirical proxy for a disaster realization. Formally, the disaster transition probability over an interval $[t,t+\tau]$ is $I^{N\to D}_{t\to t+\tau} = Pr(\sum_{u=1}^{\tau} s_{t+u} \geq 1 | s_t = 0, y^T)$.[14] This continuous measure allows better inference because it relies not only on clear transitions into disaster, but also on near misses, like the 2007–2009 period.

We deliberately focus on a framework where the probability of a disaster transition is constant, because it identifies disasters exclusively from the ex-post behavior of consumption and stock-market data. By contrast, had we allowed the disaster probability to vary over time, the Bayesian filter applied to the data would be more likely to infer a high disaster probability just before disaster realizations, or conversely, if we introduced a disaster probability signal (NVIX) in the estimation, the Bayesian procedure would be likely to find disasters in periods when the disaster signal is high.

5.2 Disaster Predictability

A robust prediction of rational time-varying disaster risk theories is that abnormally high disaster concerns precede disasters. This prediction does not say economic disasters are fully predictable, but rather that in a long enough sample, disasters occur more often when disaster concerns are elevated. We test whether our proxy for the disaster predictability, NVIX, predicts disasters using a simple linear probability model.

Our main specification tests if NVIX predicts disaster transitions as proxied by $I^{N\to D}_{t\to t+\tau}$, controlling for the contemporaneous disaster state $s_t = I^D_t > 0.5$ and the interaction of the current

[14] This expression can be written as $I^{N\to D}_{t\to t+\tau} = Pr(s_{t+1} = 1|s_t = 0, y^T) + \prod_{j=1}^{\tau-1}(1 - Pr(s_{t+j} = 1|s_{t+j-1} = 0, y^T))Pr(s_{t+j+1} = 1|s_{t+j} = 0, y^T) = I^{N\to D}_{t+1} + \prod_{j=1}^{\tau-1}(1 - I^{N\to D}_{t+j})I^{N\to D}_{t+j+1}$.

disaster state and NVIX. Intuitively, the interaction controls for the mechanical effect that NVIX cannot predict a transition into disaster when the economy is already in a persistent disaster state. We control for expected stock market variance and its interaction with the current disaster state, using the same variance models of Section 3.3.

Table 10 reports the coefficient on NVIX for different horizons and subject to alternative controls for expected stock market risk. As in the return predictability regressions, we run the regression in variance space consistent with the theory (e.g. Gabaix, 2012). We find that, in the full sample, NVIX is high just before disaster transitions. When the filtered disaster probability is zero and $NVIX^2$ is one standard deviation above its mean, the probability of a disaster over the next twelve months increases by 2.5% (column 3 times the NVIX standard deviation). Columns 4 to 8 use various models to control for expected stock market variance. Coefficients and statistical significance are stable across specifications. Column 7 (model 4) is of special interest as it uses NVIX in the variance forecasting model. Similar to what we found for returns in Table 5, we find that the disaster forecasting ability of NVIX is orthogonal to its ability to forecast variance. Only when we add credit spreads (model 5) to the variance forecasting model, which is also a disaster sensitive measure and is available in a much shorter sample, the coefficients become less precisely estimated and lose their statistical significance (column 8). Coefficients, however, barely change.

These results show that disaster risk is quite different from volatility risk. Even though disasters are periods of elevated volatility, realized financial volatility has little forecasting power about transitions into an economic disaster. This feature is especially evident at medium term horizons, 3 to 24 months, where volatility forecasts barely impact the regression R-squared.

Figure 5 illustrates this predictability result by showing the average behavior of NVIX and a realized variance-based measure of VIX around disasters. We see that up to 15 months before a disaster NVIX is consistently above its long-run mean, while the variance-based measure remains close to its long-run mean. During a disaster transition, realized variance mechanically spikes up, and as time passes differences in their behavior disappears.

Overall these results reinforce the hypothesis that NVIX captures concerns about disaster risk. These results also tell us that these concerns are rational in the weak sense that disaster concerns are associated with future transitions into a disaster regime. The magnitude of disaster risk variation is also reasonable; our estimates imply that the probability of the economy transitioning into a disaster within a year has a standard deviation of 2.6% (Table 10, average across specifications, $\sigma(E[I^{N\to D}_{t\to t+12}|NVIX^2_{t-1}]) = \beta_1 \times \sigma(NVIX^2) = 0.13 \times 20.20)$. Since in our sample the annual unconditional probability of transition into a disaster is 3.78%, our estimates imply that that the annual probability of a disaster arrival is below 9.5% more than 95% of the time. The probability of being in a disaster state is substantially higher, because some disasters are quite persistent.

5.3 Return Predictability

Figure 6 shows that the inclusion of either the Great Depression or World War II has a large impact on our estimates. The figure depicts how the return predictability estimates evolve over our sample, with the date on the x-axis denoting the beginning of the estimation sample. Once each of these rare events drops out of the estimation sample, the coefficient increases sharply.

From the perspective of a rare disaster risk model, two plausible mechanisms can reconcile the full sample with our findings about the post-war sample: (i) disaster realizations could statistically attenuate the return predictability relation if NVIX successfully forecasts disasters as predicted by the theory; (ii) long-lasting disaster periods, a salient feature of the data (Nakamura, Steinsson, Barro, and Ursúa, 2013), could have a similar effect as time-varying rare disaster models (e.g. Gourio, 2012) predict that the link between the disaster probability, option implied volatility, and expected returns breaks down when the economy is already in a disaster state. Intuitively, options have little additional information about a disaster when the economy is already in this state.

We next investigate whether the large macroeconomic events of the early sample are indeed behind the sharp break in return predictability. If the probability of a disaster per period is low enough, a time-varying rare-disaster model predicts that realized excess returns can be written in terms of the expected probability of a disaster event and actual disaster realizations as follows,

$$r^e_{t\to t+\tau} = \beta_0 + \beta_1 E_t[I^{N\to D}_{t\to t+\tau}] + \left(\beta_2 + \epsilon^D_{t\to t+\tau}\right) I^{N\to D}_{t\to t+\tau} + \epsilon^N_{t\to t+\tau}, \tag{4}$$

where $\beta_2 = E_t[r^e_{t\to t+\tau}|I^{N\to D}_{t\to t+\tau} = 1]$ is the expected excess return conditional on a disaster event, a large negative number. In models like Gabaix (2012), β_1 is the expected disaster loss under the risk-neutral measure. If $M_{t,t+\tau}$ is the stochastic discount factor that prices cash-flows between t and $t+\tau$, $\beta_1 = -E_t[M_{t,t+\tau} r^e_{t\to t+\tau}|I^{N\to D}_{t\to t+\tau} = 1]$. In models with recursive utility (Wachter, 2013; Gourio, 2012), β_1 also includes the risk premia associated with disaster probability risk, which compensate investors for the risk associated with changes in the probability of a disaster. If there is no risk premia associated with disaster or disaster-probability risks, then $\beta_1 = -\beta_2$. In general we expect $\beta_1 > -\beta_2$ if investors require a premium to be exposed to disaster risk.

In samples without disasters, a univariate regression of excess returns on the disaster probability recovers a consistent estimate of β_1, and that is how we interpret our post-war results. Generally, however, the estimates depend on the number of realized disasters in finite samples, which renders the coefficient estimates not directly interpretable.

A regression of realized returns on the disaster probability that excludes disaster realizations recovers consistent estimates of β_1 as long as $E\left[I^{N\to D}_{t\to t+\tau}\epsilon^N_{t\to t+\tau}\right] = 0$. This condition is satisfied in the time-varying rare disaster model, but may not hold under a plausible alternative model where stock market variance is variable and predictable. We pursue here the strategy of excluding disasters, which is the right approach under the time-varying rare disaster model. In Appendix A.3 we estimate a truncated regression model with time-varying volatility and find that the bias adjustment is modest and has no material effect on the return predictability coefficients or their statistical significance.

Formally, we follow Schularick and Taylor (2009) and Krishnamurthy and Vissing-Jorgensen (2012) and construct $I^R_{t\to t+\tau} = 1_{\left\{I^{N\to D}_{t\to t+\tau} > 0.5\right\}}$, which is an indicator variable that turns on whenever the probability of a disaster transition in the forecasting window is above 50%. Following the same logic behind the disaster predictability regressions of Section 5.2, we add controls for the contemporaneous disaster state $s_t = 1_{\left\{I^D_t > 0.5\right\}}$ and interactions of the state with NVIX.

Table 11 reports the normal times predictability coefficient of news implied variance, its t-statistic, and the R-squared for alternative horizons and controls. Column 1 presents full sample estimates including disasters. Consistent with Figure 6, the positive relationship between NVIX and future returns is statistically weak and only statistically significant at the 12 month horizon. Columns 2 to 9 exclude periods where the forecasting window has a disaster transition ($I^R_{t\to t+\tau} = 1$). This procedure removes a very small number of months; for one-month (twelve-month) ahead forecast it excludes 3 (46) observations. Consistent with the results for the post-war sample, the predictability coefficient β_1 is positive and statistically significant at the 6, 12, and 24 months.

Columns 3 to 8 show that these results are robust to the inclusion of various measures of expected variance. Note that in the columns where we use the credit spread in the variance model, the sample starts in 1919, and as a result the coefficients increase quite a bit and become better measured. The most conservative specification is in column 6, which includes NVIX in the variance model, and controls for any variance forecasting ability NVIX might have.

These results reinforce the time-varying rare disaster risk interpretation of our findings. We find in the full sample a relation between NVIX and future returns that is strikingly similar to the one we found in the post-war sample. Consistent with this interpretation, the relationship between NVIX and future returns is mostly present during normal times, and implies a large amount of disaster risk premia variation in frequencies from 6 months up to two years.

Quantitatively, the coefficients are in line with the post-war results, with a one standard deviation increase in $NVIX^2$ leading to $\sigma_{NVIX^2} \times \beta_1 \in [2.9\%, 5.4\%]$ higher excess returns in the following year depending on the model we use to control for stock market risk. This compares with $\sigma_{NVIX^2} \times \beta_1 \in [3.4\%, 5.3\%]$ over the post-war sample.

5.4 A Quantitative Evaluation of Time-varying Rare Disaster Models

Time-varying rare disaster risk models were developed as a candidate explanation for the excess volatility puzzle. Their quantitative success as an explanation for the puzzle hinges on the pattern of time-variation in disaster risk, and the sensitivity of excess returns to disaster probability risk. Calibrations such as Gourio (2012) and Wachter (2013) use cross-country estimates such as Barro and Ursua (2008) to determine the severity of disasters terms of drops in consumption and losses in the financial claims of interest. Together with assumptions about investors preferences, this disciplines the model-implied relationship between time-variation in the disaster probability and risk premia. However, in the absence of direct measurement of the disaster probability, the modeler is free to pick a disaster probability process that fits the pattern of return predictability and excess volatility observed in the data. Our disaster probability estimates can inform such calibrations. Our return probability estimates are also useful in providing a check if the cross-country data extrapolates well to the US. In particular, by analyzing the sensitivity of excess returns to disaster probability shocks we can evaluate if the disaster concerns that we measure are related to events of the same magnitude as the ones implied by the cross-country data.

We compare our estimates with Wachter (2013) because it provides direct counterparts to the quantities of interest. In that model, unconditionally, the disaster probability spends 95% of the time in values lower than 10%. This lines up surprisingly well with our estimates, where the disaster probability spends 95% of time in values below 9.5% (see Section 5.2). Thus, in terms of overall variation in the disaster probability, the Wachter (2013) calibration is in-line with our estimates. However, it achieves this disaster probability distribution by considering a more persistent disaster probability process than we recover from the data. Its assumed (annualized) disaster probability has a persistence of 0.9934 at the monthly frequency, and a standard deviation of 0.36%. We estimate a 1.55% standard deviation and 0.8 persistence at the same frequency. At the yearly frequency, that model implies a volatility of disaster probability of 1.21% and persistence of 0.92, while our estimates indicate a standard deviation of 2.26% and persistence of 0.5. The lower persistence of the disaster predictability detected in the data, implies the disaster risk model can explain much less of the very long-run movements observed in risk premia.

The Wachter (2013) setup produces a sensitivity of risk-premia to the disaster probability of 1.8 in the calibration with recursive utility.[15] For example, if the instantaneous disaster probability increases by one percentage point, the instantaneous risk premium increases by 1.8 percentage points.[16] In its CRRA utility specification, the sensitivity is slightly above 1, with a 1 pp increase in the probability of a disaster mapping into a 1 pp higher risk premia. We estimate that a one-standard deviation movement in NVIX increases the probability of a disaster by 2.5% and expected returns by 2.9%, implying a risk premia sensitivity to disaster of 1.16, very close to the Wachter (2013) CRRA specification. This indicates that the type of disaster risk that our measure is capturing is related to events of similar severity as the ones implied by the cross-country data.

Our estimates suggest that the disaster probability process and the risk premia variation it induces are consistent with a leading calibration of the rare disaster risk model. While our estimates and the Wachter (2013) calibration agree on the unconditional distribution of disaster risk shocks, our estimates point to shocks (to the disaster probability) that are larger, but less persistent. The data suggests that disaster concerns produce large, but relatively short-lived spikes in risk premia.

[15]There the coefficient of relative risk aversion is 3 and the intertemporal elasticity of substitution is 1.

[16]This quantity can be computed directly from Wachter (2013), Figure 3.

6 Conclusion

We use a text-based method to extend options-implied measures of uncertainty back to the end of the 19th century. We find that our news-based measure of implied volatility, NVIX, predicts returns at frequencies from 6 up to 24 months. Four pieces of evidence suggest that these return predictability results are driven by variation in investors' concerns regarding rare disasters. First, we find that the predictive power of NVIX is orthogonal to contemporaneous or forward-looking measures of stock market volatility. Second, we use alternative options-based measures, which are more focused on left tail risk, to estimate their news-based counterparts and find similar return predictability results. Third, using content analysis we trace a large part of the variation in risk premia to concerns about wars and government policy, which are tightly related to the types of events discussed in the rare disasters literature. Lastly, we show that our measure predicts disasters, even after controlling for stock market volatility. Importantly, the amounts of predictability detected in stock returns and disasters are quantitatively consistent with disasters of the same magnitude as documented by Barro and Ursua (2008) using cross-country data.

A Appendix

A.1 Inference

Our main specification poses two statistical challenges for inference: the use of overlapping observations and the use of generated regressors. The issue of overlapping data can be appropriately addressed with off-the-shelf adjustments in our empirical design. Specifically we adjust standard errors to reflect the dependence that this introduces into forecast errors using four different ways: Newey and West (1987), Hansen and Hodrick (1980), Hodrick (1992), and bootstrap. For the first three standard errors we use the same number of lags as the forecasting window. In our empirical analysis, results for all of these test statistics are similar, and robust to the use of somewhat longer lags. We report Newey and West (1987) standard errors throughout.

The second issue is that NVIX (and other news implied measures) are estimated in a first stage, which could add to the estimation uncertainty of coefficients in the second stage Murphy and Topel (2002). Before describing how we adjust our standard errors for the first stage uncertainty, it is important to note that under the null of no return predictability, there is no adjustment to the standard errors. Thus, if one is testing whether NVIX can predict returns, one should not adjust the standard errors (Wooldridge, 2010, ch. 6, pp. 115–116).

While theoretically these standard errors should not be used to construct tests of the null that a return predictability coefficient is zero, in order to be conservative we also develop a procedure to quantify the estimation uncertainty around our point estimates that is introduced by the first stage regression. These standard errors are not useful for statistical tests where the null is zero, but they are useful for evaluating the overall uncertainty associated with our estimates.

Because we use a Support Vector Regression in the first stage, a machine learning methods for which standard inference tools have not yet been developed, we cannot apply an off-the-shelf methodology here. Instead, we merge the Murphy and Topel (2002) methodology for computing standard errors when regressors are estimated with a bootstrap methodology to estimate the estimation uncertainty in the first stage. We now describe this procedure in detail.

A.1.1 Standard Errors with Generated Regressors

The second-stage regression model can be written as

$$y = \beta_0 x_0 + \beta_1 f(\mathbf{w}, \mathbf{x_1}) + \epsilon_t, \tag{5}$$

where in our setting y is stock market excess returns, x_0 are the set of regressors that do not feature a generated regressor problem, and $f(\mathbf{w}, \mathbf{x_1})$ is the function where parameters w are estimated in a first stage. In general, f can be multivariate, so let it be an $m \times 1$ vector of functions. In our main specification $m = 1$ as $f(\hat{\mathbf{w}}, \mathbf{x_1}) = \frac{(\hat{\mathbf{w}} \cdot \mathbf{x}_1)^2}{12}$, where $\hat{w}' x_1 = NVIX$, that is $f(\hat{\mathbf{w}}, \mathbf{x_1})$ is NVIX in variance space and in monthly units, where the vector w is estimated in the first stage to fit VIX, and x_1 is the vector of word counts.

We apply Theorem 1 of Murphy and Topel (2002) to our setting. Define $Z = [x_0, f(\hat{\mathbf{w}}, \mathbf{x_1})]$, and F^* as the matrix whose individual entries are given by $F^*_{tj} = \sum_{k=1}^m \beta_{1,k} \frac{\partial f_k}{\partial w_j}(\hat{\mathbf{w}}, \mathbf{x_1})$, where j indexes the vector of n-gram weights $\mathbf{w}$. Let $Q_1 = \lim_{T\to\infty} \frac{\sum_{t=1}^T Z_t \otimes F_t^*}{T}$ and $Q_0 = \lim_{T\to\infty} \frac{\sum_{t=1}^T Z_t' Z_t}{T}$. Then the variance-covariance matrix of the two-stage OLS estimator $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1]$ is given by

$$\Sigma = \Sigma_{OLS} + Q_0^{-1} Q_1 V(\hat{\mathbf{w}}) Q_1' Q_0^{-1}, \tag{6}$$

where Σ_{OLS} is the standard variance-covariance matrix of the second-stage, the one that ignores the fact that $\hat{\mathbf{w}}$ has to be estimated in the first stage. In our application $V(\hat{\mathbf{w}})$ is a variance-covariance matrix of n-gram weights, which has the same dimension as our dictionary, $N \times N$, where $N \approx 400,000$. Since we only have 1,368 months in our full-sample, we cannot directly estimate the estimation uncertainty related to the weights. However, our application only requires estimating the uncertainty associated with our index, which is a linear combination of words. Formally,

$$\Sigma = \Sigma_{OLS} + Q_0^{-1} V(Q_1' \hat{\mathbf{w}}) Q_0^{-1}, \tag{7}$$

where $V(Q_1 \cdot \hat{\mathbf{w}}) = V(\frac{\sum_{t=1}^T Z_t \otimes F_t^*}{T} \cdot \hat{\mathbf{w}})$. In our main specification, $F^*_{tj} = \frac{2}{12}\beta_1 NVIX_t \mathbf{x_{1,t}}$. Thus it follows that

$$V(Q_1 \cdot \hat{\mathbf{w}}) = V\left(\frac{\sum_{t=1}^T Z_t \otimes \frac{2}{12}\beta_1 NVIX_t \mathbf{x}_{1,t} \cdot \hat{\mathbf{w}}}{T}\right) = \left(\frac{2}{12}\beta_1\right)^2 Q_2, \tag{8}$$

where $Q_2 \equiv V\left(\frac{\sum_{t=1}^{T} Z_t \otimes NVIX_t \mathbf{x}_{1,t} \cdot \hat{\mathbf{w}}}{T}\right)$.

This object is much simpler and has the same dimensions as the total number of coefficients being estimated in the *second-stage.* In our main specification, $V(Q_1 \cdot \hat{\mathbf{w}})$ is a 2×2 matrix. As mentioned, if $\beta_1 = 0$, the generated regressor standard error adjustment is trivially zero.

We estimate $V(Q_1 \cdot \hat{\mathbf{w}})$ using bootstrap as follows. Recall that $NVIX_t$ is the predicted value of VIX on month t based on the vector of word counts $\mathbf{x}_{1,t}$ and weight vector $\hat{\mathbf{w}}$. We draw from the *train* subsample with replacement, $B = 1,000$ bootstrap samples of the same size. We then estimate alternative $NVIX_{b,t} = \hat{\mathbf{w}}_\mathbf{b} \cdot \mathbf{x}_{1,t}$, $b = 1 \ldots B$, using a support vector regression with the same hyper-parameters as in footnote 8.

We estimate Q_2 by computing the variance-covariance matrix

$$\widehat{Q}_2 = \sum_{b=1}^{B} \frac{1}{B} \left(\frac{\sum_{t=1}^{T} Z_t \otimes NVIX_t \hat{\mathbf{w}}_\mathbf{b} \cdot \mathbf{x}_{1,t}}{T} - \sum_{b=1}^{B} \frac{1}{B} \frac{\sum_{t=1}^{T} Z_t \otimes NVIX_t \hat{\mathbf{w}}_\mathbf{b} \cdot \mathbf{x}_{1,t}}{T} \right)^2 . \tag{9}$$

This analysis can be extended for the case where $f(\mathbf{w}, \mathbf{x_1})$ is multivariate. For example, in Section 4, where we decompose $NVIX$ into categories, each category is estimated, therefore

$$f(\mathbf{w}, \mathbf{x_1}) = \left[\sum_{j=1}^{N} w_j x_{1,tj}, \sum_{j=1}^{N} w_{N+j} x_{1,tj}, \ldots, \sum_{j=1}^{N} w_{(m-1)N+j} x_{1,tj} \right], \tag{10}$$

where $f(w, x_1)$ is $m \times 1$, the number of text categories in the regression, and w is an $mN \times 1$ vector of the stacked individual category weights. In this case,

$$\begin{aligned} F^*_{t,j} = \sum_{k=1}^{m} \beta_{1,k} \frac{\partial f_k}{\partial w_j} &= \beta_{1,1} [x_{1,tj}, \ldots, 0] + \beta_{1,2} [0, x_{1,tj}, \ldots, 0] + ... + \beta_{1,m} [0, \ldots, x_{1,tj}] \qquad (11) \\ &= [\beta_{1,1} x_{1,tj}, \beta_{1,2} x_{1,tj}, \ldots, \beta_{1,m} x_{1,tj}], \end{aligned}$$

and therefore,

$$Q_1 \cdot \hat{\mathbf{w}} = \frac{\sum_{t=1}^{T} Z_t \otimes F^*_t}{T} \cdot \hat{\mathbf{w}} = \frac{\sum Z_t \left[\sum_{k=1}^{m} \beta_{1,k} \sum_{j=1}^{N} w_{(k-1)N+j} x_{1,tj} \right]}{T}, \tag{12}$$

which can again be estimated via bootstrap.

A.2 Structural Disaster Identification

We identify disasters using a statistical model of rare disasters in the spirit of Nakamura, Steinsson, Barro, and Ursúa (2013), who use a Bayesian framework to statistically distinguish disaster periods from normal periods by measuring the behavior of annual consumption for a large cross-section of countries. As mentioned, we deliberately focus on a framework where the probability of a disaster transition is constant to avoid biasing the predictability regressions.

We extend the Nakamura, Steinsson, Barro, and Ursúa (2013) framework to include stock market returns as an additional signal about the state of the economy. Our model interprets large negative returns as more likely to be transitions into a disaster, if volatility has been previously low, and if future periods exhibit a substantial and persistent reduction in consumption growth. Large negative returns that are not followed by drops in economic activity are interpreted as a mix of increases in volatility and unusually large negative return realizations.

A.2.1 State Dynamics

Consumption of the representative agent follows a two state Markov chain. Let states be $s_t \in \{0, 1\}$, where $s_t = 1$ denotes a disaster state. Log consumption and dividend growth follow,

$$\begin{aligned} \Delta c_{t+1} &= \mu_c(s_t) + \sigma_c \epsilon^c_{t+1}, \\ \Delta d_{t+1} &= \mu_d(s_t) + \sigma_{d,t} \epsilon^d_{t+1}, \end{aligned} \tag{13}$$

with state transitions $p \equiv Pr(s_{t+1} = 1 | s_t = 0) < Pr(s_{t+1} = 1 | s_t = 1) \equiv q$. Disasters are persistent. Once in a disaster, the economy is likely to stay in a disaster for a while. For simplicity, we abstract from high-frequency covariation between dividends and consumption, which is very low in the data. For our purposes this assumption is harmless.

In addition to the state s_t which is the only driver of time-variation in consumption growth, the economy features variation in non-priced shocks to dividend volatility $\sigma_{d,t}$ that evolves as follows

$$\sigma^2_{d,t+1} = (1 - \rho_v)\overline{\sigma}^2_d + \rho_v \sigma^2_{d,t} + \sigma_v \epsilon^v_{t+1}. \tag{14}$$

Stochastic volatility provides an important competing mechanism for the model to account for large return surprises.

We are interested in estimating $I_t^D \equiv Pr(s_t = 1|y^T)$, the probability the economy was in a disaster regime at time t, and $I_{t+1}^{N \to D} \equiv Pr(s_{t+1} = 1, s_t = 0|y^T)$, the probability that economy transitioned into a disaster regime from t to $t+1$.

A.2.2 Asset Pricing Framework

We assume that the representative consumer in our model has a stochastic discount factor given by $M_{t+1} = M(\Delta c_{t+1}, s_{t+1}, s_t)$, which is consistent with both CRRA-type preferences and with recursive preferences (Epstein and Zin, 1989; Weil, 1990). This specification implies that the price-dividend ratio of the dividend process (13) can be written as a function of the disaster state $\pi(s_t) = \bar{\pi} e^{\psi(s_t)}$, with the normalization $\psi(0) = 0$. Substituting this expression into the log return of the dividend claim we obtain the realized return dynamics. Using the standard log-linearization around the average price-dividend ratio we obtain,

$$log(R_{t+1}^e) = \mu_d(s_t) + \sigma_{d,t}\epsilon_{t+1}^d + \kappa_1 \psi(s_{t+1}) - \psi(s_t) + \kappa_0, \tag{15}$$

where κ_1 and κ_0 are log-linearization constants. Realized returns reflect permanent shocks to dividends ϵ_{t+1}^d and transitions into and out of disaster states $\kappa_1 \psi(s_{t+1}) - \psi(s_t)$.

Realized returns are informative about regime transitions to the extent that the price-dividend ratio sensitivity to the disaster state $\psi(1) - \psi(0)$ is large relative to the volatility $\sigma_{d,t}$. The model interprets large negative returns as more likely to be a transition into a disaster, if volatility has been previously low, and if future periods exhibit a substantial and persistent reduction in consumption growth. Large negative returns that are not followed by drops in economic activity are interpreted as a mix of increases in volatility and unusually large negative dividend innovations $\sigma_{d,t}\epsilon_{t+1}^d$.

A.2.3 Measurement

We use a "mixed-frequency" approach adapted from Schorfheide, Song, and Yaron (2013) to simultaneously use economic data measured at different frequencies, which allows us to use the best consumption growth data that is available in each sample period. We model the true monthly consumption growth as hidden to the econometrician, and use annual consumption growth (Barro and Ursua (2008), 1896–1959) as signals. Whenever data on monthly consumption growth (NIPA, 1960–2009) is available we assume it measures monthly consumption growth without error.

We represent monthly time subscript t as $t = 12(j-1) + m$, where $m = 1, ..., 12$, j indexes the year and m the month within the year. Annual consumption is the sum of monthly consumption over the span of a calendar year, $C^a_{(j)} = \sum_{m=1}^{12} C_{12(j-1)+m}$. Following Schorfheide, Song, and Yaron (2013) we represent annual consumption growth rates as a function of monthly ones. We log-linearize this relationship around an average monthly growth rate C^* and define c as the percent deviations from C^*:

$$c^a_{(j)} = \frac{1}{12} \sum_{m=1}^{12} c_{12(j-1)+m}. \tag{16}$$

Because monthly consumption growth can be written $g_{c,t} = c_t - c_{t-1}$, annual growth rates are

$$g^a_{c,(j)} = c^a_{(j)} - c^a_{(j-1)} = \sum_{\tau=1}^{23} \left(\frac{12 - |\tau - 12|}{12} \right) g_{c,12j-\tau+1}. \tag{17}$$

We measure realized variance using daily stock market returns within month t, which satisfies,

$$rvar_t = \sigma^2_{d,t} + \sigma_{rvar} w^{rvar}_t, \tag{18}$$

wherew^{rvar} represents measurement error, the noise in realized volatility due to the volatility of realized returns.

A.2.4 State Space Representation

We now construct the system state evolution and measurement equations. Let us define consumption growth shocks as deviations from the conditional (on the economic regime s) expected growth rate, $\epsilon^c_{t+1} = g_{c,t+1} - \mu_c(s_t)$, and define the hidden state x_t as

$$x_t = \begin{bmatrix} \sigma^2_{d,t} - \overline{\sigma}^2_d \\ \epsilon^c_t \\ \epsilon^c_{t-1} \\ \cdots \\ \epsilon^c_{t-22} \end{bmatrix}. \tag{19}$$

The hidden state's evolution can be represented as an auto-regressive process given by

$$x_{t+1} = Ax_t + C\epsilon_{t+1}, \tag{20}$$

where $\epsilon = [\epsilon^c, \epsilon^d, \epsilon^v]$. The measurement vector

$$y_{t+1} = \begin{bmatrix} log(R^e_{t+1}) - log(R^f_{t+1}) \\ rvar_{t+1} \\ \Delta c^m_{t+1} \\ \Delta c^a_{t+1} \end{bmatrix} \tag{21}$$

can be represented as a function of the hidden states and the hidden disaster regimes as follows

$$H_{t+1} \times y_{t+1} = H_{t+1} \times F\left(\{s_{t-j}\}_{j=0}^{11}\right) + Gx_t + B(x_t)\epsilon_{t+1} + Dw_{t+1}, \tag{22}$$

where the matrix H_{t+1} selects the components of the measurement vector that are observed in a particular sample period. For example, annual consumption growth is only observed at the end of the year, so H_{t+1} selects the fourth row only when $t+1$ is a December month and the annual consumption growth data is available. Monthly consumption is only available after 1959, so the matrix H_{t+1} selects the third row if $t+1$ is in a year after 1959. The vector $F(\{s_{t-j}\}_{j=0}^{11})$ adds the expected value of each of the measurement variables as function of the hidden economic states s_t. The matrix G maps hidden state variables into the observable variables, the vector ϵ groups economic shocks, and the vector w_{t+1} groups measurement errors.

A.2.5 Bayesian Filtering

Our goal is to filter the time-series of realized disasters. We keep the estimation simple by calibrating the parameters and using a Bayesian approach to infer state transitions. Given the calibrated parameters and the observed data $Y = \{y_t\}_{t=1}^T$, we estimate the most likely trajectory of the disaster state $S = \{s_t\}_{t=1}^T$ and the hidden variables $X = \{x_t\}_{t=1}^T$,

$$p(S, X|Y) \propto p(Y|X, S)p(X|S)p(S) \tag{23}$$

Bayesian inference requires the specification of a prior distribution $p(S)$ which we choose consistent with the 2% per year probability of a disaster event estimated by Barro and Ursua (2008) using cross-country data.

We use a Gibbs sampler to construct the posterior by repeating the following two steps:

1. Draw $S^{(i)} \sim p(S|X^{(i-1)}, Y)$

2. Draw $X^{(i)} \sim p(X|S^{(i)}, Y)$

Where we construct $p(X|S^{(i)}, Y)$ using a Kalman smoother. The Gibbs sampler generates a sequence of random variables which converges to the posterior $p(S, X|Y)$.

A.2.6 Calibration

Table A.2 summarizes the calibrated parameters. Most of these are easily estimated from the data. We estimate the parameters driving the hidden volatility process by first fitting an AR(3) to realized variance and then estimating an AR(1) on the one step ahead variance predictor. The realized variance measurement error σ_{rvar} is constructed from the forecasting error of this specification. Consumption growth is calibrated to have annual volatility of 2%, and annual growth rate of 3.5% in good times and -2% during disasters. Disasters are assumed to strike with a 2% probability per year, and disasters end within a year with a 10% probability (this number pins down q). The log-linearization constant κ_1 is constructed using the average price-dividend ratio in the post-war sample. We set $\psi(s = 1)$ to be consistent with a stock market drop on a normal times to disaster transition of -25%, and set the quantity $\kappa_0 + \mu_d(s_t = 0) - log(R_t^f)$ to fit the equity premium during the post war period. The change in dividend growth is chosen so that $\mu_{d,t}(s = 1) - \mu_{d,t}(s_t = 0)$ lines up with the consumption drop during a disaster.

A.3 Truncation

Excluding disasters can lead to biases if our predictor forecasts stock market variance. Because the timing of disaster realizations are greatly influenced by the realization of abnormally low returns, a plausible alternative model might feature time-varying volatility (at least in the pre-war sample), but not time-varying disaster risk. Under this alternative model, our procedure would be classifying as disasters, periods of high variance that turn out to have low return realizations, and a variable that predicts variance (but not returns), could show up as predicting returns if we exclude "disaster periods" from the regressions.

According to this truncation mechanism it is enough to control for the forecast of the truncated mean of the returns distribution (the Mills ratio). If return predictability excluding "disasters" is only a result of time-varying truncation, a predictor of the Mills ratio would completely drive out NVIX. To be consistent with the specification we used when forecasting disasters, our specification controls for the predicted Mills ratio using several alternative variables. Table A.3 presents the results. Neither the coefficients or their statistical significance are impacted by including the relevant Mills ratio forecast. As before, the most conservative specification is model (4), which includes NVIX, price-to-earning ratio and three lags of realized variance. As before, including credit spreads make the results stronger.

Here we describe formally the logic of this truncation adjustment. Consider the following alternative model featuring time-varying volatility and constant expected returns,

$$\begin{aligned} \sigma_{t+1}^2 &= \mu_\sigma + \rho_\sigma \sigma_t^2 + \omega\sqrt{\sigma_t^2} H_\sigma w_{t+1} \\ r_{t+1} &= \mu_r + \sigma_{t+1} H_r w_{t+1} \end{aligned}$$

In this counter-factual economy there is no predictability, and there is no sense that very low returns are special as there are no special compensation for disasters. But suppose in this environment we use threshold $\underline{r}$ to split the sample in disaster periods and normal times. In this case we would have average returns in normal-periods periods given by:

$$E[r_{t+1}|r_{t+1} \geq \underline{r}, \sigma_{t+1}] = \mu_r + \sigma_{t+1} E\left[w_{r,t+1}|w_{r,t+1} \geq \frac{\underline{r} - \mu_r}{\sigma_{t+1}}\right] = \mu_r + \sigma_{t+1}\lambda(\frac{\underline{r} - \mu_r}{\sigma_{t+1}}),$$

where $\lambda(x)$ is known as the Mills ratio. In the context of our exercise we know exactly the threshold $\underline{r}$. If $NVIX_t$ predicts future volatility the truncation effect will lead us to find that NVIX predicts returns when in fact it does not. In this case, conditional expectations are given by:

$$E[r_{t+1}|r_{t+1} \geq \underline{r}|NVIX_t^2, \sigma_t^2] = \mu_r + E[\sigma_{t+1}\lambda(\frac{\underline{r} - \mu_r}{\sigma_{t+1}})|NVIX_t^2, \sigma_t]$$

The above expression tells us that in order to test the time-varying rare disaster story against the truncation story it suffices to control for the best predictor of the quantity $\sigma_{t+1}\lambda(\frac{\underline{r}-\mu_r}{\sigma_{t+1}})$. The essence of this test is the restriction imposed by the truncation hypothesis that any return predictability has to happen through the prediction of the Mills ratio multiplied by the volatility. That is, we first estimate,

$$\sigma_{t+1}\lambda(\frac{\underline{r}-\mu_r}{\sigma_{t+1}}) = \Gamma X_t + \epsilon_t,$$

where X_t is a set of predictors (NVIX inclusive) and the constant, we then run

$$r^e_{t+1} = \beta_0 + \beta_1 NVIX^2_t + \beta_2 \hat{\Gamma} X_t.$$

Under the null that all predictability is driven by truncation, we have $\beta_1 = 0$ and $\beta_2 = 1$.

For multi-period return forecasts, a observation is excluded as long there is at least one disaster transition in the forecasting window. To derive the truncation bias formally write multi-period expected returns as,

$$E[\frac{\sum_{i=1}^{\tau} r_{t+i}}{\tau}|\{r_{t+z} \geq \underline{r}|1 \leq z \leq \tau\}, X_t] = \frac{1}{\tau}\sum_{i=1}^{\tau} E[E[r_{t+i}|r_{t+i} \geq \underline{r}]|X_t] = \frac{1}{\tau}\sum_{i=1}^{\tau} E[\sigma_{t+i}\lambda(\frac{\underline{r}-\mu_r}{\sigma_{t+i}})|X_t].$$

We implement this by constructing multi-period forecasts of the Mills ratio,

$$\frac{1}{\tau}\sum_{i=1}^{\tau}\sigma_{t+i}\lambda(\frac{\underline{r}-\mu_r}{\sigma_{t+i}}) = \Gamma_\tau X_t + \epsilon_{t+\tau},$$

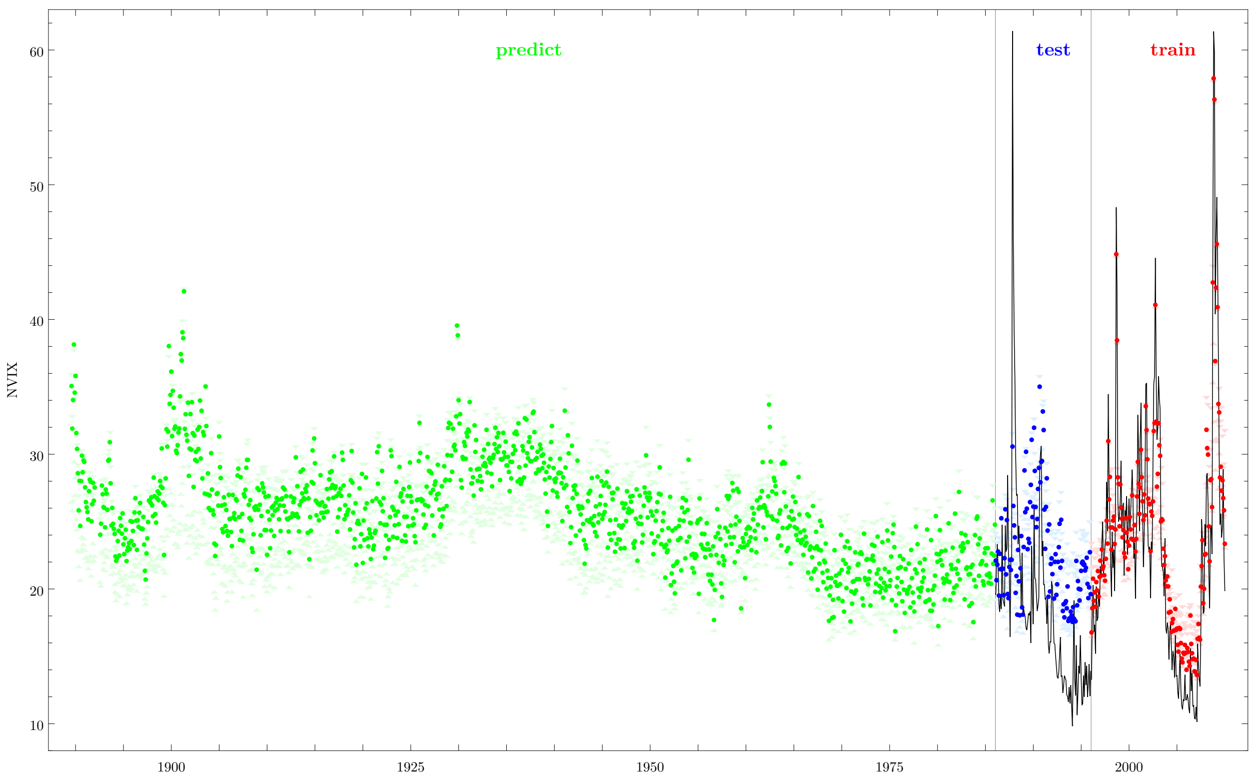
and using $EMILLS_{t-1,\tau} = \hat{\Gamma}_\tau X_{t-1}$ as a control variable.

A.4 Horse Races with Financial Variables

We horse race NVIX directly against different predictors. If the concerns encoded in NVIX are the same concerns reflected in the other predictors, then the predictor measured with more noise should be driven out of the regression, and if not driven completely out we would expect the coefficient magnitude to decrease.

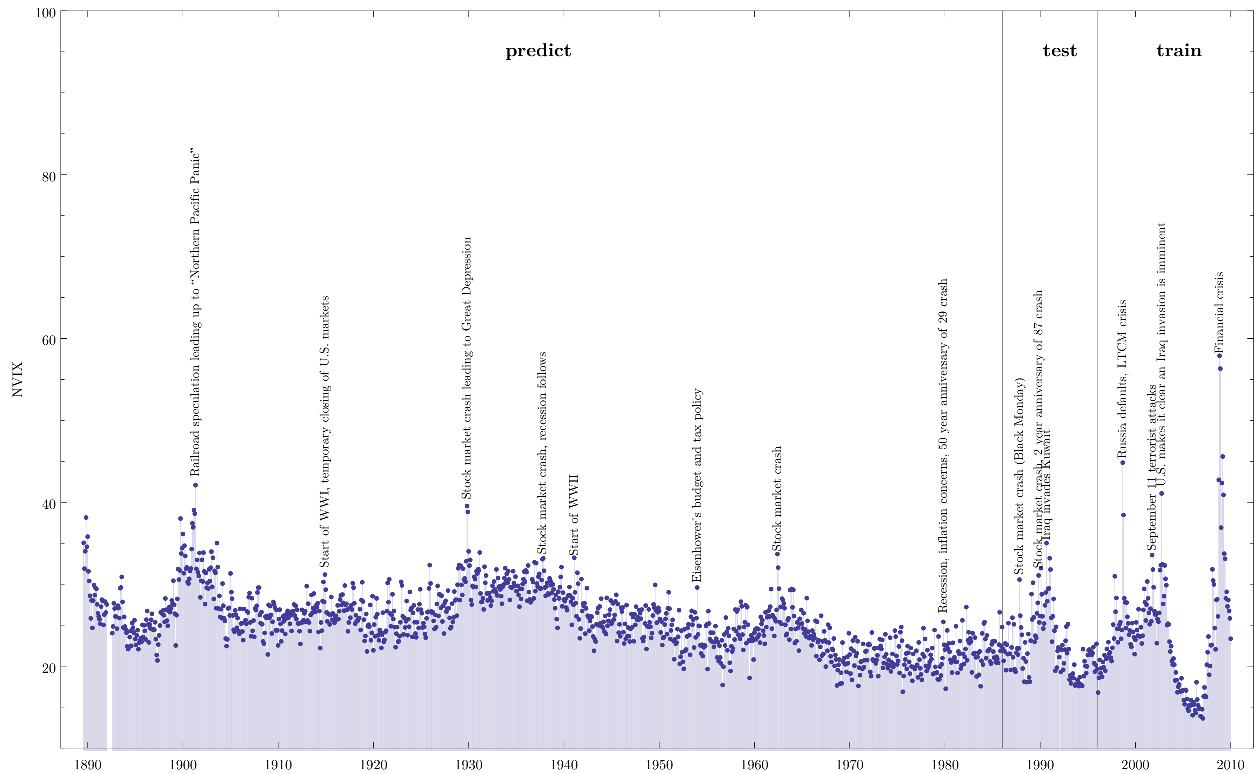
The results in Table A.4 show remarkably stable coefficients across specifications, suggesting that NVIX captures additional information relative to what is reflected in a variance based measured of VIX, the credit spread, or the price to earnings ratio. Comparing R-squared across horizons we see that the predictive power of NVIX and the other variables roughly add up. At the yearly horizon, NVIX has a (univariate) R-squared of 3.3% and a marginal contribution of 2.3% (column 4 minus column 5 is 8.6%–6.3%). All the other variables together have an R-squared of 6.3% with marginal contribution of 5.3% (8.6%–3.3%). This pattern strongly suggests that these variables measure different things.

Figure 1: News-Implied Volatility 1890–2009



Solid line is end-of-month CBOE volatility implied by options VIX_t. Dots are news implied volatility (NVIX) $VIX_t = w_0 + \mathbf{w} \cdot \mathbf{x}_t$, where $x_{t,i}$ are appearances of n-gram i in month t scaled by total month t n-grams, and $\mathbf{w}$ is estimated with a support vector regression. The *train* subsample, 1996 to 2009, is used to estimate the dependency between news data and implied volatility. The *test* subsample, 1986 to 1995, is used for out-of-sample tests of model fit. The *predict* subsample includes all earlier observations for which options data, and hence VIX is not available. Light-colored triangles indicate a nonparametric bootstrap 95% confidence interval around $\widehat{VIX}$ using 1000 randomizations. These show the sensitivity of the predicted values to randomizations of the *train* subsample.

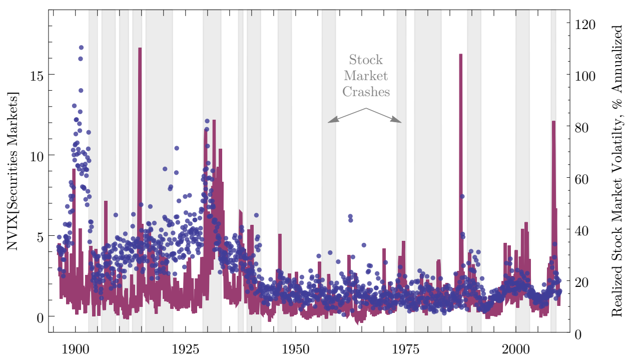
Figure 2: News-Implied Volatility Peaks by Decade



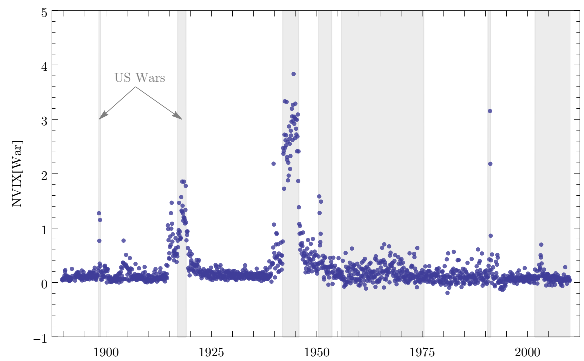
We describe NVIX peak months each decade by reading the front page articles of *The Wall Street Journal* and cross-referencing with secondary sources when needed. Many of the market crashes are described in Mishkin and White (2002). See also Noyes (1909) and Shiller and Feltus (1989).

Figure 3: News Implied Volatility due to Different Word Categories

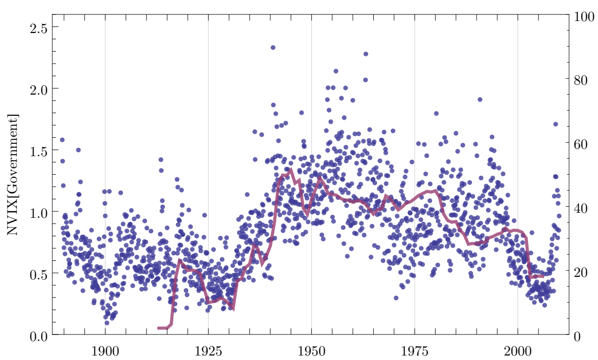
(a) Stock Markets



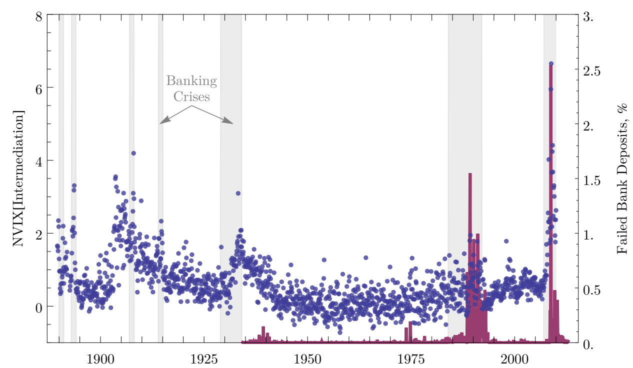
(b) War



(c) Government



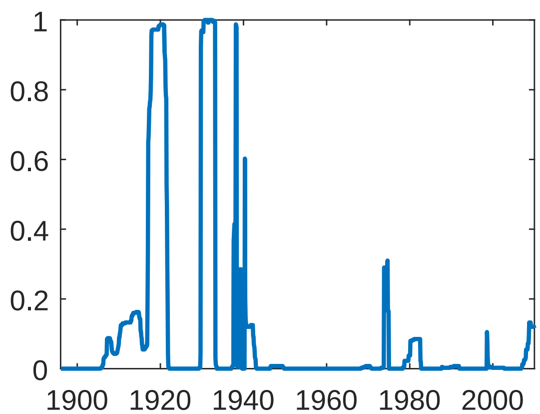
(d) Intermediation



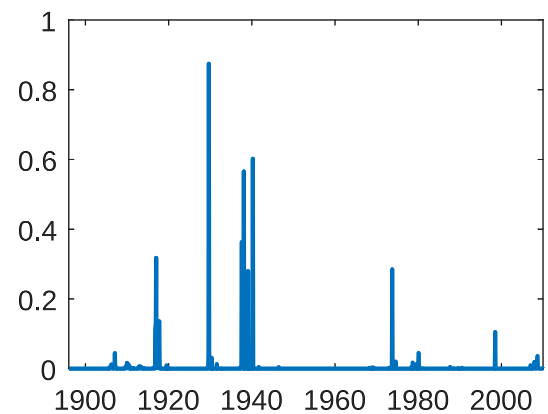
In all panels dots are monthly NVIX due only to category C-related words $\hat{v}_t(C) = \mathbf{x}_t \cdot \mathbf{w}(C)$. Panel (a): Solid line is annualized realized stock market volatility. Shaded regions indicate stock market crashes identified by Reinhart and Rogoff (2011). Panel (b): Shaded regions are US wars, specifically the American-Spanish, WWI, WWII, Korea, Vietnam, Gulf, Afghanistan, and Iraq wars. Panel (c): Solid line is the annual average marginal tax rate on dividends from Sialm (2009). Panel (d): Solid line is percent of total insured deposits held by US banks that failed each month, from the FDIC starting April 1934. Shaded regions indicate banking crises identified by Reinhart and Rogoff (2011).

Figure 4: Filtered Disaster Probability

(a) Probability that US economy is in a disaster

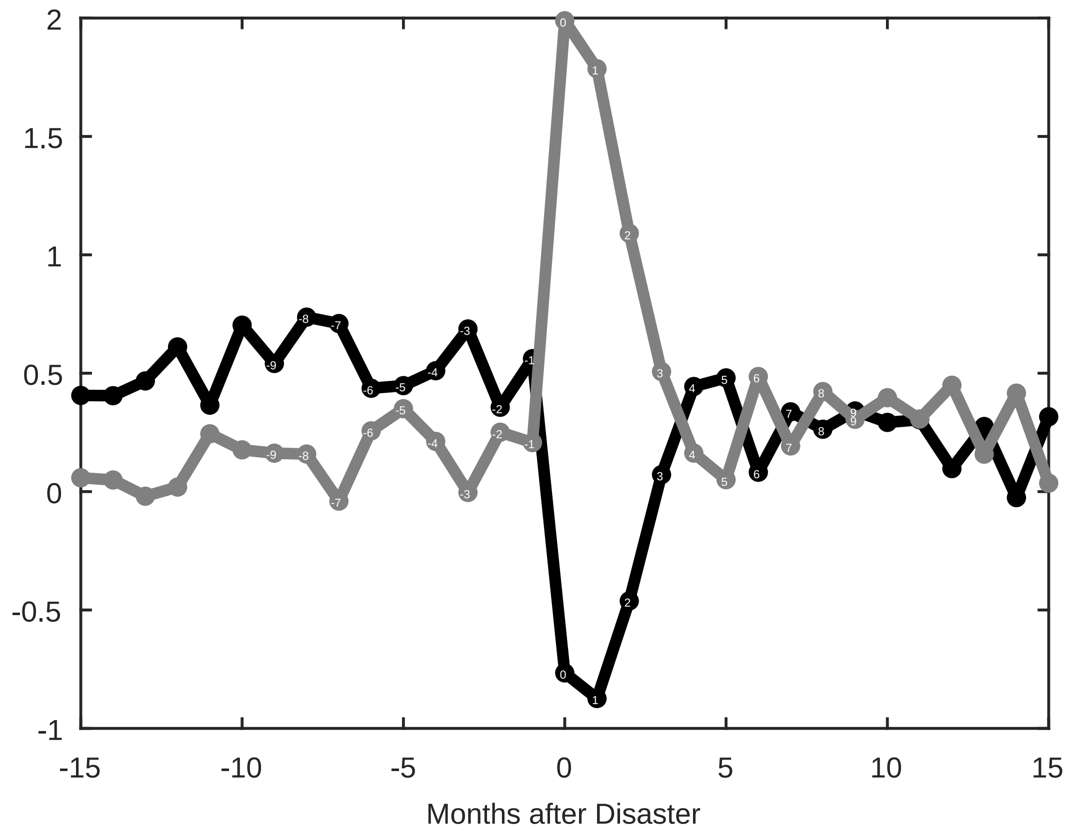


(b) Probability that US economy transitions into a disaster



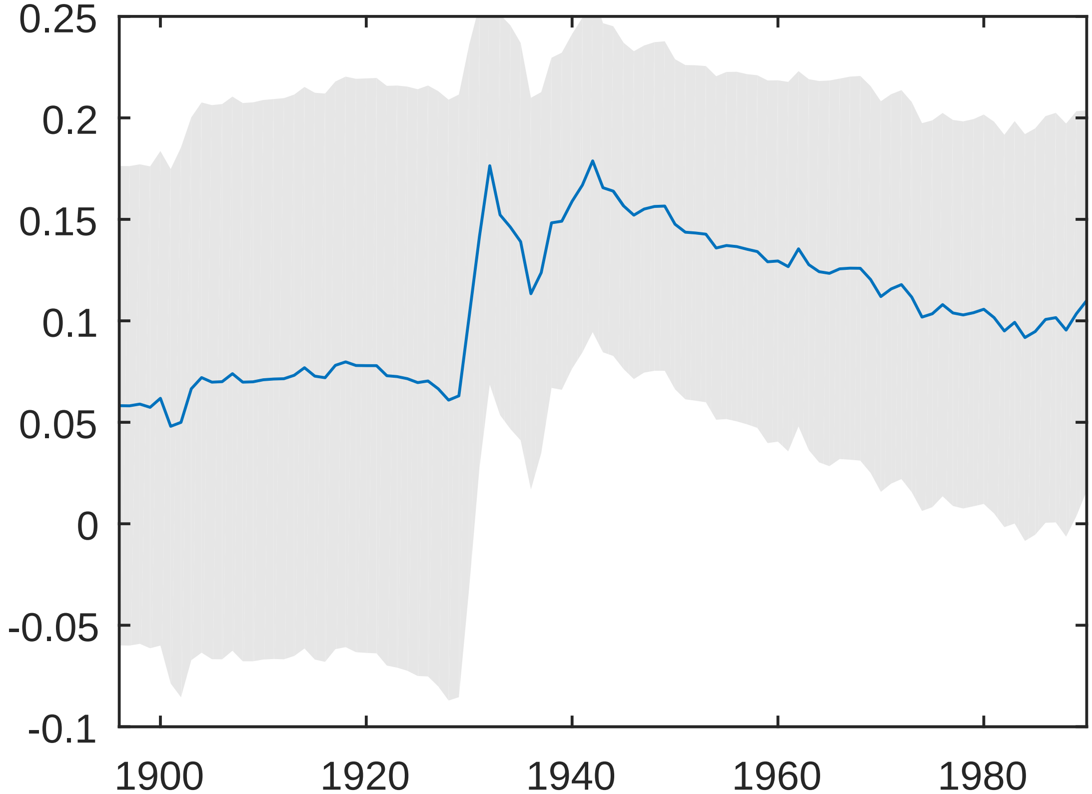
Panel (a) depicts the posterior probability that the US economy is in a disaster regime, $I_t^D = Prob\left(s_t = 1|y^T\right)$. Panel (b) depicts the probability that the economy transitions into a disaster regime during a particular month, $I_{t\to t+1}^{N\to D} = Prob\left(s_{t+1} = 1, s_t = 0|y^T\right)$. Both measures are posterior distributions implied by aggregate consumption data and aggregate stock market return data. Estimation details are in Appendix A.2.

Figure 5: NVIX and Variance forecasts Before and After Transitions into Disaster



The black line is the component of $NVIX^2$ orthogonal to the variance-based forecast of VIX^2. The gray line is the realized variance-based forecast of VIX^2 (Model 6 in Table 10). Both measures are demeaned and standardized using their sample standard deviation. Reported are averages across disaster transitions organized in event time, where a disaster transition is defined as a month t where the disaster transition probability is higher than 0.1 ($I^{N\to D}_{t-1\to t} > 0.1$).

Figure 6: Rolling Return Predictability Regression Estimates



Rolling window coefficient β_1 estimates for the excess return predictability regression $r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \epsilon_t$ at the $\tau = 12$ months horizon. Years on the x-axis represent the start of the estimation window. All windows run until 2009, the end of our sample. The shaded region represents the 95% confidence interval.

Table 1: Out-of-Sample VIX Prediction

Panel (a) Out-of-Sample Fit		Panel (b) Out-of-Sample OLS Regression $v_t = a + b\hat{v}_t + e_t, \quad t \in test$		
$R^2\,(test) = 1 - Var\,(v_t - \hat{v}_t)\,/Var\,(v_t)$	18.53	a	0.17	[4.37]
$RMSE\,(test) = \sqrt{\frac{1}{T_{test}} \sum_{t \in test} (v_t - \hat{v}_t)^2}$	7.48	b	0.82	[0.20]
Obs	119	R^2	19.46	

Reported are out-of-sample model fit statistics using the *test* subsample. Panel (a) reports variance of the predicted value (NVIX) as a fraction of actual VIX variance, and the root mean squared error. Panel (b) reports a univariate OLS regression of actual VIX on NVIX. Robust standard errors are in brackets.

Table 2: Out-of-Sample Realized Volatility Prediction Using News

Subsample	$RMSE$ SVR	R^2 SVR	$RMSE$ Reg	R^2 Reg	Correlation	Obs.
train	3.38	90.69	2.62	92.70	96.28	168
test	9.61	20.24	9.08	20.35	45.11	119
predict	10.68	13.58	8.50	15.99	39.98	1150

Reported are model fit statistics repeating the estimation procedure over the same *train* subsample as before, only replacing VIX with realized volatility as the dependent variable of the SVR (1). The *train* subsample, 1996 to 2009, is used to estimate the dependency between monthly news data and *realized* volatility. The *test* subsample, 1986 to 1995, is used for out-of-sample tests of model fit. The *predict* subsample includes all earlier observations for which options data, and hence VIX is not available. $RMSE$ SVR is root mean squared error of the SVR. R^2 SVR is one less the prediction error's variance as a fraction of actual realized volatility's variance. $RMSE$ Reg and R^2 Reg are estimated from a subsequent univariate OLS regression of actual realized volatility on realized volatility implied by news.

Table 3: NVIX Predicts Post-war Stock Market Returns

$r^e_{t\to t+\tau} = \beta_0 + \beta_1 X^2_{t-1} + \epsilon_{t+\tau}$							
	X:	NVIX				VXO	VIX
		1945-2009	1945-1995	1986-2009	1990-2009	1986-2009	1990-2009
τ		(1)	(2)	(3)	(4)	(5)	(6)
1	β_1	0.15	0.37**	0.10	0.09	0.12	0.12
	t_{NW}	[0.99]	[2.21]	[0.58]	[0.53]	[1.05]	[0.79]
	t_{GR}	[0.98]	[1.84]	[0.58]	[0.53]		
	R^2	0.35	0.74	0.28	0.30	0.82	0.68
3	β_1	0.12	0.45***	0.04	0.03	0.08	0.08
	t_{NW}	[0.81]	[3.62]	[0.25]	[0.20]	[0.79]	[0.58]
	t_{GR}	[0.80]	[2.45]	[0.25]	[0.20]		
	R^2	0.56	2.96	0.11	0.09	0.99	0.78
6	β_1	0.18**	0.43***	0.11	0.10	0.09*	0.13**
	t_{NW}	[2.48]	[3.71]	[1.42]	[1.35]	[1.87]	[1.97]
	t_{GR}	[2.41]	[2.48]	[1.42]	[1.35]		
	R^2	2.37	4.81	1.91	2.00	2.49	3.72
12	β_1	0.16***	0.31***	0.10*	0.11*	0.08*	0.11*
	t_{NW}	[3.21]	[2.77]	[1.65]	[1.93]	[1.67]	[1.94]
	t_{GR}	[3.05]	[2.13]	[1.65]	[1.92]		
	R^2	3.31	4.69	3.01	3.97	3.28	4.46
24	β_1	0.14***	0.21**	0.11**	0.11**	0.06	0.08
	t_{NW}	[3.58]	[2.16]	[2.18]	[2.04]	[1.44]	[1.36]
	t_{GR}	[3.37]	[1.81]	[2.17]	[2.03]		
	R^2	5.02	4.25	6.25	5.99	4.18	4.06
Obs		779	611	287	239	287	239

Reported are monthly return predictability regressions based on news implied volatility (NVIX), S&P 100 options implied volatility (VXO), and S&P 500 options implied volatility (VIX). The dependent variables are annualized log excess returns on the market index. Each row and each column represents a different regression. The first column examines the entire post-war period, while the second focuses on a sample that was not used in fitting NVIX to options implied volatility. The third and fourth columns report results for the sample period for which VXO and VIX are available. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table 4: Alternative Text-based Approaches

		$r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \beta_2 X_{t-1} + \epsilon_{t+\tau}$											
Controls:		None	EPU	Negative		Uncertainty		Positive		Modal Strong		Modal Weak	
				tf	tf.idf	tf	tf.idf	tf	tf.idf	tf	tf.idf	tf	tf.idf
τ		(1)	(2)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)
1	β_1	0.15	0.13	0.14	0.15	0.14	0.15	0.14	0.14	0.16	0.15	0.13	0.15
	t_{NW}	[0.99]	[0.86]	[0.93]	[0.98]	[0.95]	[0.98]	[0.96]	[0.94]	[1.03]	[0.99]	[0.89]	[0.98]
	t_{GR}	[0.98]	[0.86]	[0.92]	[0.98]	[0.94]	[0.98]	[0.95]	[0.94]	[1.03]	[0.99]	[0.89]	[0.97]
	R^2	0.35	0.44	0.44	0.35	0.85	0.35	0.40	0.40	0.43	0.39	0.91	0.35
3	β_1	0.12	0.11	0.11	0.12	0.11	0.12	0.12	0.11	0.11	0.12	0.11	0.12
	t_{NW}	[0.81]	[0.79]	[0.76]	[0.80]	[0.78]	[0.80]	[0.81]	[0.79]	[0.80]	[0.80]	[0.76]	[0.80]
	t_{GR}	[0.80]	[0.78]	[0.75]	[0.80]	[0.78]	[0.80]	[0.81]	[0.79]	[0.80]	[0.80]	[0.76]	[0.80]
	R^2	0.56	0.58	0.67	0.56	1.10	0.56	0.56	0.56	0.56	0.57	0.91	0.56
6	β_1	0.18**	0.19***	0.18***	0.18**	0.18**	0.18**	0.18**	0.18**	0.18**	0.18**	0.17**	0.18**
	t_{NW}	[2.48]	[2.66]	[2.59]	[2.45]	[2.46]	[2.45]	[2.57]	[2.41]	[2.55]	[2.47]	[2.47]	[2.45]
	t_{GR}	[2.41]	[2.55]	[2.52]	[2.38]	[2.39]	[2.38]	[2.50]	[2.34]	[2.47]	[2.40]	[2.40]	[2.38]
	R^2	2.37	2.43	2.37	2.37	2.94	2.37	2.38	2.37	2.38	2.41	2.77	2.38
12	β_1	0.16***	0.17***	0.16***	0.16***	0.15***	0.16***	0.16***	0.16***	0.16***	0.16***	0.15***	0.16***
	t_{NW}	[3.21]	[3.41]	[3.18]	[3.14]	[3.22]	[3.13]	[3.30]	[3.04]	[3.29]	[3.21]	[3.20]	[3.11]
	t_{GR}	[3.05]	[3.19]	[3.04]	[2.99]	[3.07]	[2.99]	[3.15]	[2.90]	[3.13]	[3.05]	[3.05]	[2.96]
	R^2	3.31	3.66	3.31	3.31	4.43	3.35	3.32	3.33	3.32	3.45	4.13	3.42
24	β_1	0.14***	0.16***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***
	t_{NW}	[3.58]	[3.50]	[3.01]	[3.65]	[3.60]	[3.64]	[3.50]	[3.63]	[3.49]	[3.64]	[3.53]	[3.67]
	t_{GR}	[3.37]	[3.27]	[2.89]	[3.43]	[3.40]	[3.42]	[3.33]	[3.41]	[3.31]	[3.42]	[3.33]	[3.44]
	R^2	5.02	5.79	5.03	5.02	7.04	5.11	5.46	5.23	5.55	5.05	6.23	5.27
	Obs	779	779	779	779	779	779	779	779	779	779	779	779

This table presents return predictability regressions based on our constructed NVIX series, the Economic Policy Uncertainty measure (EPU) from Baker, Bloom, and Davis (2013), and the different "language tone" word lists. We report both tf (proportional weights) and tf.idf weights of words appearing in the Loughran and McDonald (2011) Negative, Positive, Uncertainty, Modal Strong, and Modal Weak word lists. All columns feature NVIX as a predictor, and each column has as a control a different measure. The dependent variables are annualized log excess returns on the market index. The sample period is 1945 to 2009. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table 5: Stochastic Volatility Does Not Explain the Return Predictability Results

$r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \beta_2 EVAR_{t-1} + \epsilon_t$						
τ		(1)	(2)	(3)	(4)	(5)
1	β_1	0.20	0.20	0.22	0.20	0.25
	t_{NW}	[1.48]	[1.30]	[1.43]	[1.49]	[1.46]
	t_{GR}	[1.45]	[1.26]	[1.37]	[1.44]	[1.27]
	R^2	0.49	0.40	0.45	0.44	0.43
3	β_1	0.13	0.15	0.17	0.16	0.21
	t_{NW}	[0.94]	[1.12]	[1.28]	[1.32]	[1.45]
	t_{GR}	[0.93]	[1.10]	[1.25]	[1.29]	[1.27]
	R^2	0.59	0.61	0.71	0.69	0.74
6	β_1	0.18**	0.21**	0.23***	0.22***	0.27**
	t_{NW}	[2.34]	[2.38]	[2.65]	[2.71]	[2.22]
	t_{GR}	[2.21]	[2.14]	[2.35]	[2.46]	[1.70]
	R^2	2.37	2.45	2.65	2.57	2.65
12	β_1	0.17***	0.18***	0.21***	0.20***	0.26**
	t_{NW}	[3.04]	[2.63]	[2.85]	[2.81]	[2.25]
	t_{GR}	[2.78]	[2.33]	[2.49]	[2.53]	[1.71]
	R^2	3.34	3.46	3.82	3.82	4.06
24	β_1	0.15***	0.17***	0.19***	0.21***	0.31***
	t_{NW}	[3.34]	[2.79]	[2.78]	[2.99]	[2.64]
	t_{GR}	[3.00]	[2.43]	[2.45]	[2.66]	[1.87]
	R^2	5.07	5.34	6.04	7.19	8.57
Obs		779	778	778	778	778
EVAR Model R^2		9.21	25.53	25.87	28.22	31.83

Return predictability regressions controlling for expected variance, where the dependent variables are market annualized log excess returns, over the post-war 1945–2009 period. Each row and each column represents a different regression. Rows show different forecasting horizons. EVAR is predicted variance using the following variables: model 1 uses lagged lagged variance, model 2 uses an three lags of realized variance, model 3 adds price to earning ratio to model 2, model 4 ads $NVIX^2$ to model 3, and model 5 adds credit spread to model 4. The last row reports the percent R-squared from the variance predictability regression used to estimate EVAR. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table 6: Alternative Measures of Uncertainty Focused on Tail Risk

		$r^e_{t\to t+\tau} = \beta_0 + \beta_1 \widehat{X_{t-1}} + \beta_2 EVAR_{t-1} + \epsilon_{t+\tau}$			
τ	$X:$	VIX^2	VIX premium	LT	Slope
1	β_1	0.20	0.43**	1.87	125.49*
	t_{NW}	[1.30]	[2.56]	[1.55]	[1.85]
	t_{GR}	[1.26]	[1.79]	[1.43]	[1.19]
	R^2	0.40	1.34	0.47	0.51
3	β_1	0.15	0.17	1.51	95.12*
	t_{NW}	[1.12]	[1.36]	[1.32]	[1.71]
	t_{GR}	[1.10]	[1.19]	[1.24]	[1.15]
	R^2	0.61	0.67	0.82	0.80
6	β_1	0.21**	0.17**	2.11***	75.39*
	t_{NW}	[2.38]	[2.01]	[3.01]	[1.82]
	t_{GR}	[2.14]	[1.56]	[2.34]	[1.19]
	R^2	2.45	1.59	3.13	1.39
12	β_1	0.18***	0.12*	1.65***	54.99
	t_{NW}	[2.63]	[1.80]	[2.91]	[1.64]
	t_{GR}	[2.33]	[1.46]	[2.29]	[1.13]
	R^2	3.46	1.71	3.60	1.57
24	β_1	0.17***	0.11**	1.47***	53.68**
	t_{NW}	[2.79]	[2.18]	[2.73]	[2.24]
	t_{GR}	[2.43]	[1.64]	[2.20]	[1.28]
	R^2	5.34	2.44	5.21	2.41
T		778	778	778	778

This table replicates our main results of Table 3 for alternative tail risk measures, over the post-war 1945–2009 period. For each of these measures we reproduce the methodology we applied to VIX. The symbol $\widehat{X_t}$ denotes the text based estimator of variable X_t. The first column reproduces our main results. In the second column is VIX premium ($= VIX_t^2 - E_t[Var(R_{t+1})]$), where $E_t[Var(R_{t+1})]$ is constructed using an AR(1). In the third column is the Left-Tail (LT) measure from Bollerslev and Todorov (2011). In the fourth column is the slope of option implied volatility curve, constructed using the 30-day volatility curve from Option metrics. We use puts with delta of -0.5 and -0.8 to compute the slope. The variable $EVAR_{t-1}$ is included in all columns to control for expected future variance using an AR(3) model (Model 2 of Table 5). t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table 7: Top Variance Driving n-grams

n-gram	Variance Share, %	Weight, %
stock	36.61	0.09
market	7.03	0.06
stocks	6.93	0.07
war	5.57	0.03
u.s	3.66	0.06
tax	2.10	0.04
special	1.85	0.02
washington	1.61	0.01
banks	1.56	0.06
financial	1.41	0.10

n-gram	Variance Share, %	Weight, %
treasury	1.40	0.05
gold	1.38	-0.04
oil	1.34	-0.03
_ u.s	0.85	0.04
bonds	0.82	0.04
house	0.80	0.05
_ stock	0.78	0.03
billion	0.69	0.05
economic	0.67	0.05
like	0.51	-0.04

We report the fraction of NVIX variance $h(i)$ that each n-gram drives over the *predict* subsample as defined in (3), and the regression coefficient w_i from (1), for the top 20 n-grams.

Table 8: Categories Total Variance Share

Category	Variance Share, %	n-grams	Top n-grams
Government	2.75	86	tax, money, rates, government, plan
Intermediation	3.99	75	banks, financial, business, bank, credit
Natural Disaster	0.01	71	fire, storm, aids, happening, shock
Stock Markets	51.75	61	stock, market, stocks, industry, markets
War	5.62	54	war, military, action, world war, violence
Unclassified	35.90	467,754	u.s, special, washington, treasury, gold

We report the percentage of NVIX variance ($= \sum_{i \in C} h(i)$) that each n-gram category C drives over the *predict* subsample.

Table 9: Risk Premia Decomposition

$r^e_{t\rightarrow t+\tau} = \beta_0 + \sum_{j=1}^{N} \beta_j X_{j,t-1} + \epsilon_{t+\tau}$						
	1945–2009		1896–1945		1896–2009	
τ:	6	12	6	12	6	12
Government	4.27***	4.17***	-0.96	-0.94	2.54**	2.47**
t_{NW}	[3.15]	[2.90]	[0.37]	[0.44]	[2.20]	[2.07]
t_{GR}	[2.98]	[2.79]	[0.37]	[0.43]	[2.11]	[2.03]
var share	(45.77)	(56.76)	(5.34)	(5.32)	(27.33)	(27.18)
War	4.72***	2.97**	3.37**	3.98***	3.86***	3.71***
t_{NW}	[3.15]	[2.26]	[2.08]	[2.74]	[3.80]	[4.41]
t_{GR}	[2.74]	[2.05]	[1.98]	[2.54]	[3.24]	[3.65]
var share	(30.70)	(17.37)	(57.98)	(66.75)	(53.98)	(52.96)
Intermediation	0.26	0.67	0.86	1.74	1.04	1.48
t_{NW}	[0.10]	[0.37]	[0.32]	[0.78]	[0.61]	[1.03]
t_{GR}	[0.10]	[0.36]	[0.32]	[0.77]	[0.60]	[1.00]
var share	(0.01)	(0.71)	(0.26)	(1.45)	(0.27)	(2.10)
Stock Markets	0.87	0.55	3.27	3.06	0.93	1.08
t_{NW}	[0.26]	[0.20]	[1.42]	[1.13]	[0.57]	[0.59]
t_{GR}	[0.26]	[0.20]	[1.36]	[1.11]	[0.56]	[0.58]
var share	(0.76)	(0.33)	(42.47)	(37.47)	(6.30)	(7.61)
Natural Disaster	0.95	1.03	1.90	0.05	0.69	1.02
t_{NW}	[1.08]	[1.63]	[0.97]	[0.03]	[0.80]	[1.55]
t_{GR}	[1.07]	[1.61]	[0.97]	[0.03]	[0.79]	[1.54]
var share	(3.21)	(5.30)	(4.74)	(0.06)	(0.63)	(2.32)
Residual	2.26**	2.02*	1.35	1.20	1.90*	1.48
t_{NW}	[2.03]	[1.87]	[0.37]	[0.36]	[1.81]	[1.49]
t_{GR}	[1.87]	[1.74]	[0.37]	[0.36]	[1.69]	[1.42]
var share	(19.54)	(20.19)	(0.41)	(0.40)	(12.03)	(7.83)
R^2	6.60	8.87	3.87	6.88	4.01	6.35
Obs	779	779	588	588	1367	1367

Reported are monthly return predictability regressions based on six word categories constructed from news implied volatility (NVIX). The dependent variables are annualized log excess returns on the market index. All six variables are normalized to have unit standard deviation over the entire sample. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.*var share* is the percent of risk premia variation due to each category, i.e. $cov\left(\beta_j X^j_{t-1}, \sum_{j=1}^{N} \beta_j X^j_{t-1}\right) / var\left(\sum_{j=1}^{N} \beta_j X^j_{t-1}\right)$. The residual category is the orthogonal component of NVIX that is not explained by the five interpretable word categories.

Table 10: NVIX Predicts Disasters

$$I^{N\to D}_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \beta_2 s_{t-1} + \beta_3 s_{t-1} \times NVIX^2_{t-1} + \beta_4 EVAR_{t-1} + \beta_5 s_{t-1} \times EVAR_{t-1} + \epsilon_{t+\tau}$$

τ		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	$\beta_1 \times 100$		0.01*	0.01*	0.01	0.01	0.01	0.01	0.00	0.01
	t_{NW}		[1.74]	[1.78]	[1.40]	[1.24]	[1.25]	[0.88]	[0.35]	[1.26]
	t_{GR}		[1.70]	[1.73]	[1.36]	[1.21]	[1.21]	[0.85]	[0.35]	[1.23]
	R^2	0.01	0.23	0.25	0.28	0.30	0.30	0.31	0.54	0.30
3	$\beta_1 \times 100$		0.03*	0.03*	0.03*	0.03	0.03	0.03	0.02	0.03
	t_{NW}		[1.92]	[1.93]	[1.80]	[1.52]	[1.52]	[1.23]	[0.85]	[1.48]
	t_{GR}		[1.86]	[1.87]	[1.70]	[1.46]	[1.46]	[1.15]	[0.83]	[1.42]
	R^2	0.01	0.79	0.88	0.93	0.94	0.94	0.94	1.62	0.95
6	$\beta_1 \times 100$		0.06*	0.07*	0.06*	0.06*	0.06*	0.06	0.04	0.06*
	t_{NW}		[1.89]	[1.89]	[1.80]	[1.67]	[1.67]	[1.37]	[0.86]	[1.66]
	t_{GR}		[1.84]	[1.83]	[1.70]	[1.59]	[1.58]	[1.27]	[0.84]	[1.57]
	R^2	0.00	1.77	1.95	1.99	2.03	2.04	2.04	3.84	2.03
12	$\beta_1 \times 100$		0.12*	0.12*	0.13*	0.13*	0.13*	0.13*	0.12	0.13*
	t_{NW}		[1.81]	[1.77]	[1.83]	[1.79]	[1.78]	[1.68]	[1.23]	[1.79]
	t_{GR}		[1.76]	[1.72]	[1.73]	[1.69]	[1.68]	[1.51]	[1.17]	[1.68]
	R^2	0.00	3.04	3.17	3.25	3.26	3.23	3.23	5.56	3.26
24	$\beta_1 \times 100$		0.17	0.18	0.19*	0.19	0.19	0.20	0.18	0.19
	t_{NW}		[1.56]	[1.53]	[1.65]	[1.63]	[1.62]	[1.57]	[1.17]	[1.63]
	t_{GR}		[1.53]	[1.50]	[1.57]	[1.55]	[1.54]	[1.43]	[1.12]	[1.55]
	R^2	0.15	3.31	3.42	3.55	3.53	3.49	3.49	6.25	3.53
	Obs	1368	1367	1367	1364	1364	1364	1364	1091	1365
Variance model		-	-	-	(1)	(2)	(3)	(4)	(5)	(6)

Reported are monthly disaster predictability regressions based on news implied volatility (NVIX). The dependent variable is $I^{N\to D}_{t\to t+\tau} = I^{N\to D}_{t+1} + \prod_{j=1}^{\tau-1}(1 - I^{N\to D}_{t+j}) I^{N\to D}_{t+j+1}$, which reflects the probability that the economy transitions into a disaster between t and $t+\tau$. Each row and each column represents a different regression. Rows show different forecasting horizons. Column (1) shows disaster predictability only controlling for the regime of the economy $s_t = I^D_t > 0.5$. Column (2) adds $NVIX^2$, and column (3) adds the interaction $NVIX^2 \times s_t$. Columns (4-8) control for alternative measures of expected stock market variance: (1) past realized variance; (2) AR(3) forecasting model; (3) adds price to earnings ratio to model (2) ;(4) adds $NVIX^2$ to model (3); (5) adds credit spread to model (4); model (6) use as a variance proxy $E\left[VIX^2_{t-1}|VAR\right]$ from Table A.4, the forecast of VIX^2 using contemporaneous and two lags of realized variance. The sample is Jan/1896 to Dec/2009 for columns 1-7 and 9, and Jan/1919 to Dec/2009 for column 8. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table 11: Return Predictability in the Full Sample

$r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \beta_2 s_{t-1} + \beta_3 s_{t-1} \times NVIX^2_{t-1} + \beta_4 EVAR_{t-1} + \beta_5 s_{t-1} \times EVAR_{t-1} + \epsilon_{t+\tau}$									
τ		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	β_1	0.13	0.15	0.14	0.14	0.15	0.14	0.16	0.14
	t_{NW}	[1.23]	[1.46]	[1.33]	[1.31]	[1.38]	[1.10]	[1.03]	[1.30]
	t_{GR}	(1.22)	(1.43)	(1.29)	(1.27)	(1.33)	(1.05)	(1.00)	(1.26)
	R^2	0.49	0.61	0.62	0.63	0.62	0.62	1.05	0.62
	excl./obs	0/1367	3/1364	3/1361	3/1361	3/1361	3/1361	3/1088	3/1362
3	β_1	0.05	0.08	0.08	0.07	0.07	0.07	0.14	0.07
	t_{NW}	[0.58]	[0.90]	[0.88]	[0.78]	[0.85]	[0.74]	[1.26]	[0.76]
	t_{GR}	(0.58)	(0.89)	(0.87)	(0.77)	(0.84)	(0.72)	(1.20)	(0.75)
	R^2	0.48	0.65	0.66	0.67	0.66	0.66	1.86	0.67
	excl./obs	0/1367	9/1358	9/1355	9/1355	9/1355	9/1355	9/1082	9/1356
6	β_1	0.09	0.12**	0.12*	0.12*	0.12*	0.13	0.22***	0.12*
	t_{NW}	[1.40]	[1.98]	[1.77]	[1.68]	[1.78]	[1.63]	[2.87]	[1.66]
	t_{GR}	(1.37)	(1.91)	(1.68)	(1.59)	(1.68)	(1.47)	(2.33)	(1.57)
	R^2	1.37	1.82	2.11	2.48	2.45	2.43	5.10	2.51
	excl./obs	0/1367	20/1347	20/1344	20/1344	20/1344	20/1344	20/1073	20/1345
12	β_1	0.10*	0.14***	0.14**	0.14**	0.15**	0.15*	0.27***	0.14**
	t_{NW}	[1.72]	[2.71]	[2.20]	[2.10]	[2.21]	[1.95]	[3.46]	[2.08]
	t_{GR}	(1.68)	(2.54)	(2.03)	(1.94)	(2.02)	(1.69)	(2.61)	(1.92)
	R^2	2.29	3.46	4.94	5.79	5.72	5.67	13.64	5.86
	excl./obs	0/1367	46/1321	46/1318	46/1318	46/1318	46/1318	46/1055	46/1319
24	β_1	0.08	0.14***	0.15**	0.15**	0.15**	0.16**	0.28***	0.15**
	t_{NW}	[1.56]	[2.85]	[2.42]	[2.32]	[2.39]	[2.14]	[3.95]	[2.30]
	t_{GR}	(1.53)	(2.66)	(2.20)	(2.11)	(2.17)	(1.82)	(2.82)	(2.09)
	R^2	3.54	6.57	9.09	10.05	10.00	9.95	21.22	10.13
	excl./obs	0/1367	94/1273	94/1270	94/1270	94/1270	94/1270	94/1019	94/1271
Variance model		-	-	(1)	(2)	(3)	(4)	(5)	(6)
Exclude disasters		no	yes	yes	yes	yes	yes	yes	yes

Reported are monthly return predictability regressions based on news implied volatility ($NVIX^2$). The dependent variables are annualized log excess returns on the market index. Each row and each column represents a different regression. Rows show different forecasting horizons. Column (1) controls for the state of the economy $s_t = I^D_t > 0.5$, and the interaction with $NVIX^2$. Column (2-8) excludes disasters. A observation t is excluded as a disaster month if $I^R_{t\to t+\tau} = \left(I^{N\to D}_{t\to t+\tau} > 0.5\right) = 1$, that is, if the filtered probability implies a higher than 50% probability that the economy transitioned into a disaster during the return forecasting window. Columns (3-8) control for alternative measures of expected stock market variance: (1) past realized variance; (2) AR(3) forecasting model; (3) adds price to earnings ratio to model (2) ;(4) adds $NVIX^2$ to model (3); (5) adds credit spread to model (4); model (6) uses as a variance proxy $E\left[VIX^2_{t-1}|VAR\right]$ from Table A.4, the forecast of VIX^2 using contemporaneous and two lags of realized variance . The sample is Jan/1896 to Dec/2009 for columns 1-7 and 9, and Jan/1919 to Dec/2009 for column 8. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table A.1: Correlations Between Alternative Measures of Tail Risk

Panel A: Option based measures				
	VIX^2	VIX premium	LT	Slope
VIX^2	1.00	0.95	0.98	0.85
VIX premium		1.00	0.91	0.86
LT			1.00	0.86
Slope				1.00

Panel B: News Based Measures				
	$NVIX^2$	VIX premium	LT	Slope
VIX^2	1.00	0.83	0.92	0.82
VIX premium		1.00	0.86	0.87
LT			1.00	0.91
Slope				1.00

The options-based measured correlations are for the period Jan/1996 to Dec/2008 for which we have all four quantities. The News based measure correlations are for the full sample, Jan/1896 to Dec/2009.

Table A.2: Filtering Disasters: Calibration Parameters

Description	Parameter	Value
Average dividend volatility	$\overline{\sigma}_d^2$	0.0025
Volatility of dividend volatility	σ_v	0.0018
Persistence of dividend volatility	ρ_v	0.7300
Measurement error in realized dividend volatility	σ_{rvar}	0.0041
Log-linearization constant (average price-dividend ratio)	κ_0	0.9948
Equity premium during normal times	$\kappa_0 + \mu_d(0) - log(R_t^f)$	0.0050
Drop in dividend growth during disasters	$\mu_d(1) - \mu_d(0)$	-0.0058
Consumption growth in normal times	$\mu_c(0)$	0.0029
Drop in consumption growth during disasters	$\mu_c(1) - \mu_c(0)$	-0.0058
Volatility of consumption growth	σ_c	0.0058
Price-dividend ratio drop in a normal times to disaster transition	$\psi(1)$	-0.2500
Probability of normal times to disaster transition	p	0.0017
Probability of disaster to normal times transition	$1 - q$	0.0285

All quantities are at the monthly frequency. Discussion of parameter choice in the Appendix A.2.6.

Table A.3: Return Predictability: Controlling for Truncation Effects

$$r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \beta_2 s_{t-1} + \beta_3 s_{t-1} \times NVIX^2_{t-1} + \beta_4 EVAR_{t-1} + \beta_5 s_{t-1} \times EVAR_{t-1} + \beta_6 EMILLS_{t-1,\tau} + \epsilon_{t+\tau}$$

τ		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	$\beta_1 \times 100$	0.13	0.15	0.14	0.14	0.15	0.15	0.16	0.14
	t_{NW}	[1.23]	[1.46]	[1.26]	[1.36]	[1.43]	[1.15]	[1.02]	[1.35]
	t_{GR}	[1.22]	[1.43]	[1.22]	[1.31]	[1.38]	[1.09]	[0.99]	[1.31]
	R^2	0.49	0.61	0.67	0.63	0.63	0.63	1.05	0.63
	excl./obs	0/1367	3/1364	3/1361	3/1361	3/1361	3/1361	3/1088	3/1362
3	$\beta_1 \times 100$	0.05	0.08	0.06	0.07	0.07	0.03	0.25*	0.07
	t_{NW}	[0.58]	[0.90]	[0.73]	[0.76]	[0.84]	[0.23]	[1.80]	[0.82]
	t_{GR}	[0.58]	[0.89]	[0.72]	[0.75]	[0.83]	[0.23]	[1.61]	[0.81]
	R^2	0.48	0.65	0.97	0.77	0.75	0.74	3.34	0.74
	excl./obs	0/1367	3/1364	3/1361	3/1361	3/1361	3/1361	3/1088	3/1362
6	$\beta_1 \times 100$	0.09	0.12**	0.12*	0.11	0.12*	0.16*	0.29***	0.11
	t_{NW}	[1.40]	[1.98]	[1.69]	[1.55]	[1.72]	[1.80]	[2.80]	[1.59]
	t_{GR}	[1.37]	[1.91]	[1.61]	[1.48]	[1.63]	[1.54]	[2.17]	[1.52]
	R^2	1.37	1.82	2.19	2.95	2.59	2.59	5.75	3.01
	excl./obs	0/1367	3/1364	3/1361	3/1361	3/1361	3/1361	3/1088	3/1362
12	$\beta_1 \times 100$	0.10*	0.14***	0.13**	0.13**	0.15**	0.15*	0.31***	0.14**
	t_{NW}	[1.72]	[2.71]	[2.04]	[2.01]	[2.21]	[1.95]	[3.45]	[2.03]
	t_{GR}	[1.68]	[2.54]	[1.90]	[1.86]	[2.02]	[1.67]	[2.53]	[1.88]
	R^2	2.29	3.46	5.22	6.07	5.73	5.67	14.88	6.13
	excl./obs	0/1367	46/1321	46/1318	46/1318	46/1318	46/1318	46/1055	46/1319
24	$\beta_1 \times 100$	0.08	0.14***	0.14**	0.14**	0.16**	0.15**	0.29***	0.14**
	t_{NW}	[1.56]	[2.85]	[2.28]	[2.22]	[2.44]	[2.10]	[4.08]	[2.23]
	t_{GR}	[1.53]	[2.66]	[2.09]	[2.03]	[2.20]	[1.78]	[2.82]	[2.04]
	R^2	3.54	6.57	9.64	10.41	10.55	10.41	21.50	10.46
	excl./obs	0/1367	94/1273	94/1270	94/1270	94/1270	94/1270	94/1019	94/1271
Variance model		-	-	(1)	(2)	(3)	(4)	(5)	(6)
Exclude disasters		no	yes	yes	yes	yes	yes	yes	yes

Reported are monthly return predictability regressions based on news implied volatility ($NVIX^2$). The dependent variables are annualized log excess returns on the market index. Each row and each column represents a different regression. Rows show different forecasting horizons. Column (1) controls for the state of the economy $s = I^D > 0.5$, and the interaction of the state s with $NVIX^2$. Column (2-8) excludes disasters. Observation t is excluded as a disaster month if $I^R_{t\to t+\tau} = 1 = \left(I^{N\to D}_{t\to t+\tau} > 0.5\right)$, i.e. if the filtered probability implies a higher than 50% probability that the economy transitioned into a disaster during the return forecasting window. Columns (3-8 control for the predicted stock market variance and the predicted mills ratio using the following variables: (1) past realized variance; (2) three lags of realized variance (3) adds price to earnings ratio to model (2); (4) adds $NVIX^2$ to model (3); (5) adds credit spread to model (4); and (6) use the variance implied VIX $E\left[VIX^2_{t-1}|VAR\right]$ from Table A.4 as a control . The sample goes from Jan/1896 to Dec/2009 for columns (1-6,8) , and Jan/1919 to Dec/2009 for column 7. t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

Table A.4: Horse Races with Financial Predictors

$r^e_{t\to t+\tau} = \beta_0 + \beta_1 NVIX^2_{t-1} + \sum_{j=2}^{N} \beta_j X_{j,t-1} + \epsilon_{t+\tau}$						
τ		(1)	(2)	(3)	(4)	(5)
1	β_1	0.15	0.20	0.20	0.19	
	t_{NW}	[0.99]	[1.29]	[1.28]	[1.19]	
	t_{GR}	[0.98]	[1.26]	[1.24]	[1.16]	
	R^2	0.35	0.40	0.45	0.81	0.50
3	β_1	0.12	0.14	0.14	0.13	
	t_{NW}	[0.81]	[1.10]	[1.08]	[0.99]	
	t_{GR}	[0.80]	[1.08]	[1.06]	[0.97]	
	R^2	0.56	0.60	0.84	1.80	1.38
6	β_1	0.18**	0.21**	0.21**	0.20**	
	t_{NW}	[2.48]	[2.38]	[2.35]	[2.19]	
	t_{GR}	[2.42]	[2.17]	[2.16]	[2.03]	
	R^2	2.37	2.44	3.11	5.06	3.42
12	β_1	0.16***	0.18***	0.18***	0.17**	
	t_{NW}	[3.21]	[2.64]	[2.62]	[2.47]	
	t_{GR}	[3.08]	[2.37]	[2.36]	[2.24]	
	R^2	3.31	3.44	4.08	8.56	6.30
24	β_1	0.14***	0.17***	0.17***	0.15***	
	t_{NW}	[3.58]	[2.80]	[2.81]	[3.05]	
	t_{GR}	[3.40]	[2.49]	[2.49]	[2.65]	
	R^2	5.02	5.32	5.33	16.45	13.02
Controls\Obs		779	779	779	779	779
$NVIX^2_{t-1}$		yes	yes	yes	yes	no
$E\left[VIX^2_{t-1} \mid VAR\right]$		no	yes	yes	yes	yes
$Creditspread_{t-1}$		no	no	yes	yes	yes
$(\frac{P}{E})_{t-1}$		no	no	no	yes	yes

Reported are monthly return predictability regressions based on news implied volatility (NVIX) and controls. The dependent variables are annualized log excess returns on the market index. Each row and each column represents a different regression. Rows show different forecasting horizons. The sample is Jan/1945 to Dec/2009. The variable $E\left[VIX^2_{t-1} \mid VAR\right]$ is the variance-based VIX, the predicted value of VIX^2 using the contemporaneous variance plus two additional lags. The model is estimated in the sample where VIX is available (1990–2009). t_{NW} are Newey-West corrected t-statistics with number of lags/leads equal to the size of the return forecasting window. t_{GR} t-statistics additionally correct for the fact that the regressors are generated. *, **, and *** indicate 10, 5, and 1 percent significance levels.

References

Abel, Andrew B, Janice C Eberly, and Stavros Panageas, 2007, Optimal inattention to the stock market, *American economic review* 97, 244–249.

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? the information content of internet stock message boards, *Journal of Finance* 59, 1259–1293.

Backus, D., M. Chernov, and I. Martin, 2011, Disasters implied by equity index options, *Journal of Finance* 66, 1969–2012.

Baker, Scott, Nicholas Bloom, and Steven Davis, 2013, Measuring economic policy uncertainty, Working Paper 13-02 Chicago Booth.

Barro, R.J., 2006, Rare disasters and asset markets in the twentieth century, *Quarterly Journal of Economics* 121, 823–866.

——— , and J.F. Ursua, 2008, Consumption disasters in the twentieth century, *American Economic Review* 98, 58–63.

Barro, Robert J., 2009, Rare disasters, asset prices, and welfare costs, *American Economic Review* 99, pp. 243–264.

Bates, David S., 2012, U.s. stock market crash risk, 1926–2010, *Journal of Financial Economics* 105, 229–259.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Bollerslev, T., and V. Todorov, 2011, Tails, fears, and risk premia, *Journal of Finance* 66, 2165–2211.

Brown, Stephen J, William N Goetzmann, and Stephen A Ross, 1995, Survival, *Journal of Finance* 50, 853–873.

Cherkassky, V., and Y. Ma, 2004, Practical selection of svm parameters and noise estimation for svm regression, *Neural networks* 17, 113–126.

Cole, Harold L, and Lee E Ohanian, 1999, The great depression in the united states from a neoclassical perspective, *Federal Reserve Bank of Minneapolis Quarterly Review* 23, 2–24.

Croce, Mariano M., Thien T. Nguyen, and Lukas Schmid, 2012, The market price of fiscal uncertainty, *Journal of Monetary Economics* 59, 401 – 416.

Drechsler, I., 2008, Uncertainty, time-varying fear, and asset prices, in *AFA 2010 Atlanta Meetings Paper*.

——— , and A. Yaron, 2011, What's vol got to do with it, *Review of Financial Studies* 24, 1–45.

Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, *Working paper*.

Epstein, Larry G, and Stanley E Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* pp. 937–969.

Gabaix, X., 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.

Gao, George P., and Zhaogang Song, 2013, Rare disaster concerns everywhere, Working paper.

García, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267–1300.

Gentzkow, Matthew, and Jesse M. Shapiro, 2006, Media bias and reputation, *Journal of Political Economy* 114, pp. 280–316.

Gourio, Francois, 2008, Time-series predictability in the disaster model, *Finance Research Letters* 5, 191–203.

——— , 2012, Disaster risk and business cycles, *American Economic Review* 102, 2734–2766.

Hansen, Lars Peter, and Robert J Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *The Journal of Political Economy* pp. 829–853.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2009, *The elements of statistical learning* (Springer) second edition edn.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *Review of Financial Studies* 23, 3773–3811.

——— , 2011, Text-based network industries and endogenous product differentiation, *Working paper*.

Hodrick, Robert J, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial studies* 5, 357–386.

Huang, Lixin, and Hong Liu, 2007, Rational inattention and portfolio selection, *Journal of Finance* 62, 1999–2040.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.

Kogan, S., D. Levin, B.R. Routledge, J.S. Sagi, and N.A. Smith, 2009, Predicting risk from financial reports with regression, in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* pp. 272–280. Association for Computational Linguistics.

Kogan, S., B. Routledge, J. Sagi, and N. Smith, 2010, Information content of public firm disclosures and the sarbanes-oxley act, *Working paper*.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2012, Short-term debt and financial crises: What we can learn from us treasury supply, *Unpublished Working Paper*.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *Journal of Finance* 66, 35–65.

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

Manela, Asaf, 2011, Spreading information and media coverage: Theory and evidence from drug approvals, Ph.D. thesis University of Chicago.

——— , 2014, The value of diffusing information, *Journal of Financial Economics* 111, 181–199.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, pp. 867–887.

Miller, G.A., 1995, Wordnet: a lexical database for english, *Communications of the ACM* 38, 39–41.

Mishkin, F.S., and E.N. White, 2002, Us stock market crashes and their aftermath: implications for monetary policy, *NBER Working paper*.

Moreira, Alan, and Alexi Savov, 2013, The macroeconomics of shadow banking, *Available at SSRN 2310361*.

Murphy, Kevin M, and Robert H Topel, 2002, Estimation and inference in two-step econometric models, *Journal of Business & Economic Statistics* 20, 88–97.

Nakamura, Emi, Jón Steinsson, Robert Barro, and José Ursúa, 2013, Crises and recoveries in an empirical model of consumption disasters, *American Economic Journal: Macroeconomics* 5, 35–74.

Newey, WK, and KD West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Noyes, A.D., 1909, A year after the panic of 1907, *Quarterly Journal of Economics* 23, 185–212.

Pastor, Lubos, and Pietro Veronesi, 2012, Uncertainty about government policy and stock prices, *Journal of Finance* 67, 1219–1264.

Pedersen, Ted, Siddharth Patwardhan, and Jason Michelizzi, 2004, Wordnet::similarity - measuring the relatedness of concepts, in Daniel Marcu Susan Dumais, and Salim Roukos, ed.: *HLT-NAACL 2004: Demonstration Papers* pp. 38–41 Boston, Massachusetts, USA. Association for Computational Linguistics.

Reinhart, Carmen M., and Kenneth S. Rogoff, 2011, From financial crash to debt crisis, *American Economic Review* 101, pp. 1676–1706.

Rietz, T.A., 1988, The equity risk premium a solution, *Journal of monetary Economics* 22, 117–131.

Schorfheide, Frank, Dongho Song, and Amir Yaron, 2013, Identifying long-run risks: A bayesian mixed-frequency approach, Discussion paper Federal Reserve Bank of Philadelphia.

Schularick, Moritz, and Alan M Taylor, 2009, Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870–2008, Discussion paper National Bureau of Economic Research.

Seo, Sang Byung, and Jessica A Wachter, 2013, Option prices in a model with stochastic disaster risk, Discussion paper National Bureau of Economic Research.

Shiller, R., and W. Feltus, 1989, Fear of a crash caused the crash, *The New York Times* p. F3.

Sialm, Clemens, 2009, Tax changes and asset pricing, *American Economic Review* 99, pp. 1356–1383.

Tetlock, Paul, Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Vapnik, N. Vladimir, 2000, *The Nature of Statistical Learning Theory* (Springer-Verlag, New York.).

Wachter, Jessica A., 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Weil, Philippe, 1990, Nonexpected utility in macroeconomics, *The Quarterly Journal of Economics* pp. 29–42.

Wooldridge, Jeffrey M, 2010, *Econometric analysis of cross section and panel data* (MIT press).