Policy News and Stock Market Volatility

Scott R. Baker,[a] Nicholas Bloom,[b] Steven J. Davis[c] and Kyle Kost[d]

December 2021

Abstract: We exploit the text in newspapers and 10-K filings to quantify the drivers of aggregate and firm-level stock market volatility. We first create a newspaper-based Equity Market Volatility (EMV) tracker that moves closely with the VIX and the volatility of returns on the S&P 500. Parsing the underlying text, we then construct forty category-specific EMV trackers. News about commodity markets, interest rates, real estate markets, aggregate activity and inflation figure prominently in EMV articles, with large category-specific variation over time. Policy news is another major source of market volatility: 30 percent of EMV articles discuss tax policy, 30 percent discuss monetary policy, and 25 percent refer to some form of regulation. Trade policy news went from a virtual nonfactor in market volatility to a leading source after U.S.-China trade tensions escalated. Next, we use 10-K filings to quantify firm-level exposures to the same forty risk categories. Combining our newspaper-based measures with our textual analysis of 10-K filings, we obtain monthly firm-level risk exposure measures. Finally, we show that these measures are highly statistically significant in explaining the firm-level structure of realized equity market volatilities at the monthly frequency, even after conditioning on firm effects, time effects and industry-time effects.

JEL No. D80, E22, E66, G18, L50

Keywords: stock market, equity returns, volatility, uncertainty, government policy

Acknowledgements: Many thanks to John Cochrane, Bill Schwert (the editor), two anonymous referees and seminar participants at the Monetary Authority of Singapore for helpful comments on an earlier draft. We gratefully acknowledge financial support from the U.S. National Science Foundation and the University of Chicago Booth School of Business.

[a] Kellogg School of Management; s-baker@kellogg.northwestern.edu
[b] Stanford; nbloom@stanford.edu
[c] University of Chicago Booth School of Business and the Hoover Institution; steven.davis@chicagobooth.edu
[d] University of Chicago; kkost84@gmail.com

> The history of thought in financial markets has shown a surprising lack of consensus about a very fundamental question: what ultimately causes all those fluctuations in the price of speculative assets like corporate stocks...? One might think that so basic a question would have long ago been confidently answered.
>
> Robert Shiller, 2014

1. Introduction

Volatility in aggregate equity returns is resistant to convincing interpretation. Shiller's classic 1981 contribution shows that stock market ups and downs cannot be rationalized by realized future dividends discounted at a constant rate.[1] Partly motivated by Shiller's demonstration, one major line of research stresses time-varying expected returns in asset-pricing models with rational agents. Another prominent line, also partly motivated by Shiller, stresses non-rational beliefs, limits to arbitrage, and fads that move equity prices in ways not fully tethered to real investment opportunities.[2] See Cochrane (2017) and Barberis (2018) for recent reviews.

We develop new data and evidence that inform rational and behavioral interpretations of the volatility in equity returns. In a first step, we identify articles about stock market volatility in leading U.S. newspapers and use them to construct an Equity Market Volatility (EMV) tracker. Figure 1 displays the resulting measure, which runs from 1985 to 2018 and is scaled to match the mean value of the VIX from 1985 to 2015. Our EMV tracker moves closely with the VIX and the realized volatility of daily returns on the S&P 500, with correlations of about 0.8 (0.85) in monthly (quarterly) data.

In a second step, we parse the text in the EMV articles to quantify journalist perceptions about the news items, developments, concerns, and anticipations that drive volatility in equity returns. We classify these proximate drivers into about forty categories, many of which pertain to particular types of policy. This approach lets us assess the importance of each category to the average level of stock market volatility and its movements over time. An immediate result is the importance of news about Commodity Markets, which receives attention in 44% of all articles that enter into our EMV tracker. Most EMV articles discuss multiple topics. Thus, we also find that 31% mention

[1] See, also, LeRoy and Porter (1981), Campbell and Shiller (1987, 1988), West (1988), Schwert (1989), Cochrane (1992) and Barberis, Huang and Santos (2001), among many others. Cochrane (1991) stresses the equivalence of excess volatility to return predictability.

[2] On the difficulty of drawing confident inferences about the presence of such fads, see Summers (1986), Fama and French (1988) and Poterba and Summers (1988).

Interest Rates, 29% mention Inflation, 27% mention GDP and other Broad Quantity Indicators, and 8% mention Financial Crises.

As we show below, a narrower EMV tracker tailored to news about petroleum markets correlates well with the implied and realized volatility of oil prices. Another EMV tracker, which we tailor to macroeconomic news, surges in the wake of episodes that involve unusually high uncertainty about the near-term macroeconomic outlook – e.g., the October 1987 stock market crash, the 9-11 terrorist attacks, the March 2003 invasion of Iraq, the Global Financial Crisis, and the U.S. debt-ceiling crisis in summer 2011. These results suggest that our EMV trackers capture important drivers of fluctuations in equity market volatility.

The share of EMV articles that discusses government policy rises over time, reaching peaks in the 2001-03 period (9/11 and Iraq Invasion), the 2011-12 period (U.S. debt-ceiling crisis and the "fiscal cliff"), and the period since Donald Trump's election in November 2016. Parsing the role of policy more finely, we find that 35 percent of EMV articles refer to Fiscal Policy (mostly Tax Policy), 30 percent mention Monetary Policy, 25 percent mention some form of Regulation, and 13 percent mention National Security matters. We also construct EMV trackers tailored to these policy categories and find that each one fluctuates markedly over time. For example, our National Security EMV tracker is low in most periods but highly elevated after the 9/11 terrorist attacks and around Gulf Wars I and II. Trade Policy matters went from a virtual nonfactor for equity market volatility in the twenty years before Donald Trump's election to a leading source afterwards, especially since the intensification of U.S-China trade tensions from March 2018.

How should we interpret these findings? According to the efficient markets view, equity price movements reflect genuine news that alters rationally grounded forecasts of future earnings and discount factors. Under this view, it's natural to interpret news reports as a catalog of the rational forces that drive the volatility of equity returns. Some recent work has provided support for news-related drivers of firm stock price movements. For instance, Griffin, Hirschley, and Kelly (2011) note that firms' stock prices move much more on days where information is released about that firm. Boudoukh, Feldman, Kogan, and Richardson (2018) push further, showing that news events related to firms increase equity volatility most prominently in overnight returns, where *public* information is the main driving force.

Shiller (2014, 1496-97) articulates a rather different view: "The market fluctuates as the sweep of history produces different mindsets at different points of time, different zeitgeists….

[A]ggregate stock market price changes reflect inconstant perceptions, changes that Keynes referred to with the term 'animal spirits.'" Under this view, we expect newspaper articles to (imperfectly) mirror these mindsets and their shifts over time.[3] Under either view, we see our methods and measures as helpful in efforts to address the "basic question" posed in the epigraph: what drives corporate stock fluctuations?

Our EMV trackers have several noteworthy attributes: First, their construction is straightforward, transparent, easy to refine, and simple to replicate. Second, the frequency and volume of newspaper text affords much scope for granular characterizations of the forces that underlie equity market volatility and its movements over time. We develop several tailored EMV trackers that exploit this granular richness. Third, our text-based approach is useful for assessing the role of wars, policy risks, and other hard-to-quantify sources of stock market volatility. Fourth, we update our EMV trackers monthly in real time.[4] These real-time updates facilitate efforts to assess the out-of-sample performance of our measures.

Fourth, our measurement methods are highly scalable across countries, over time, and to new topics. Although we focus on the volatility of aggregate U.S. equity markets from 1985 onwards, we also extend our analysis of equity market volatility back to 1928. Our methods extend readily to any country or time period with digital newspaper archives and data on aggregate equity returns. As one example of such extensions, we build an Infectious Disease EMV index in response to the COVID-19 outbreak (see Baker, et al (2020)). We use this index to provide quantitative comparisons of the impacts of COVID-19 on stock market volatility relative to previous periods of severe disease outbreaks throughout recent history such as SARS, MERS, and Swine Flu.

There is a vast literature on equity returns and stock market volatility. Fama (1981), Chen, Roll and Ross (1986), and Fama and French (1989) are influential early studies that relate equity returns to macroeconomic forces. More recent contributions include Boyd et al. (2005) on stock market reactions to unemployment news, Killian and Park (2009) on the role of oil price shocks, and Bekaert et al. (2013) on the relationship between monetary policy and stock market volatility.

In one of the first studies to use newspaper text, Niederhoffer (1971) considers "world events" from 1950 to 1966 – as indicated by large headlines in the *New York Times* – and relates them to

[3] Shiller (2014, page 1497) also writes "News media tend to slant their stories toward ideas of current interest, rather than useful facts that readers no longer find interesting." Our results help in forming a judgement regarding that claim as well.

[4] Our EMV trackers are available at policyuncertainty.com/EMV_monthly.html.

U.S. stock market movements. Cutler, Poterba and Summers (1989) relate returns on U.S. equities to macroeconomic data and news accounts of "political and world events." They conclude that it's hard to explain more than half the variation in aggregate stock prices by information in these sources about discount rates and future cash flows. Baker, Bloom, Davis, and Sammon (2019) consider thousands of daily stock market moves greater than |2.5%| in fourteen national markets. Based on systematic human readings of next-day newspaper accounts, they find that journalists attribute 37% of large daily moves in the United States to news about government policy. Evidence that policy developments move stock markets resonates with the theoretical work of Pastor and Veronesi (2012, 2013), who model the role of government policy as a source of economic uncertainty and the resulting implications for risk premia and equity prices.

Another line of research explores the usefulness of stock market volatility, as measured by the VIX, for predicting and assessing other important financial and economic variables. Nagel (2012) shows the VIX to be highly predictive of the return on liquidity provision. Dreschler and Yaron (2011) show that the equity variance premium – the squared VIX minus the expected realized variance – has predictive power for stock returns. Forbes and Warnock (2012) and Rey (2018) document global patterns in capital flows, asset prices and credit growth that are closely tied to the VIX. Our EMV trackers offer a new means to identify which developments underlie the relationships of stock market volatility to other outcomes of interest uncovered in earlier works.

Finally, we contribute to the rapidly growing body of research in economics and finance that applies text-based methods. Gentzkow, Kelly, and Taddy (2018) offer an excellent survey of research in this area. Here, we mention a few papers that are closest to ours. Baker, Bloom and Davis (2016) construct newspaper-based indices of economic policy uncertainty. They find that stock price volatility reacts more strongly to their policy uncertainty indices in firms with greater exposure to policy risks. Hassan et al. (2019) apply tools from computational linguistics to conference calls about earnings announcements to construct time-varying, firm-level measures of political risks. Their text-based measures also have explanatory power for firm-level variation in stock price volatility.

Davis and Seminario (2019) quantify firm-level policy risk exposures using the text in 10-K filings. Their measures account for much of the huge dispersion in firm-level stock returns in the wake of Donald Trump's surprise victory in the 2016 presidential election. Kelly, Manela, and Moreira (2018) develop an econometric model of text usage, estimate the model on multiple text sources, and use the estimates to backcast, nowcast and forecast financial variables. Manela and Moreira (2017) apply machine-learning methods to front-page articles in the *Wall Street Journal* to develop an "NVIX" measure of stock market uncertainty and the perceived risk of rare disasters. They conclude that policy risks and especially war-related concerns are a major source of variation in equity risk premia, broadly in line with the literature on rare disasters and asset prices.[5]

2. Methodology

2.1 Constructing an Equity Market Volatility Tracker

In constructing our Equity Market Volatility (EMV) tracker, we follow Baker, Bloom and Davis (BBD) in using scaled frequency counts of newspaper articles that contain selected terms. We differ in our approach to term selection. They rely on human readings of 12,000 randomly sampled articles to populate a list of candidate terms. They then select the permutation of candidate terms that minimizes the sum of false positives and false negatives in computer-automated classifications compared to human classifications.[6] Their approach makes sense in developing a measure of economic policy uncertainty, for which there is no obvious observable counterpart. We exploit the observability of stock market volatility to take a much less labor-intensive approach.

We first specify terms in three sets, as follows:

E: {economic, economy, financial}

M′: {"stock market", stock OR stocks, "equity market", equity OR equities, S&P OR "S & P", "Standard and Poors" OR "Standard and Poor's" OR "Standard and Poor" OR "Standard & Poors" OR "Standard & Poor's}

V′: {volatility OR volatile, "realized volatility", uncertain OR uncertainty, risk OR risky, variance, VIX}

Second, we randomly select a 30% sample of articles that contain at least one element in each of **E, M′** and **V′** from 1990 to 2015.[7] Third, using the sampled articles, we construct a candidate

[5] See Rietz (1988), Barro (2006), Gourio (2008), Gabaix (2012) and Wachter (2013), among others.
[6] BBD use this procedure to select the "Policy" terms for their newspaper-based Economic Policy Uncertainty Index. Their approach to selecting terms in "Economy" and "Uncertainty" is similar in spirit but much less formal.
[7] Here, we use four newspapers for which we could download many articles that meet our criteria: the Miami Herald, Dallas Morning News, San Francisco Chronicle, and Houston Chronicle.

EMV tracker for each permutation of elements in $\mathbf{M'}$ and $\mathbf{V'}$.[8] Specifically, we count articles that contain the candidate permutation, scale that count by the number of all articles in the same paper and month, standardize the scaled frequency counts to unit standard deviation for each paper, and then average the resulting standardized, scaled counts over papers by month.[9] Finally, we select the permutation that achieves the highest R-squared value in an OLS regression of the 30-day VIX on the candidate EMV tracker using monthly data from 1990 to 2015.

Log, level, and level specifications with quadratic and cubic terms yield the same best-fit permutation, given by:

Economic terms (E): {economic, economy, financial}

Equity Market terms (M): {"stock market", equity, equities, "Standard and Poors" (and variants)}

Volatility terms (V): {volatility, volatile, uncertain, uncertainty, risk, risky}

In the analyses below, our EMV tracker is based on this best-fit term set.

In assessing our term sets and our selection procedure, a few additional remarks will be helpful. We start with parsimonious **E,** $\mathbf{M'}$ and $\mathbf{V'}$ sets to reduce the danger of overfitting. While each regression in our selection procedure has few explanatory variables (just one, except when we add quadratic and cubic terms), we consider many such regressions.

We eschew terms like "Lehman Brothers," "Bernanke" and "Iraq war" that might improve in-sample performance but perform poorly out of sample.[10] And we prefer terms that extend easily to other countries and settings. Terms like "economy," "stock market," "volatility" and "uncertainty" translate readily, while terms like "Standard and Poors" have obvious counterparts for other national stock markets. In this respect, we regard it as fortuitous that "VIX" did not make the cut for our best-fit permutation, because there is no VIX counterpart for many national stock markets.

---

[8] We consider all permutations in $P(\mathbf{M'}) \times P(\mathbf{V'})$, where $P(\bullet)$ denotes the power set and $\times$ is the Cartesian product. "Equity market" never appears in our sample of articles, so we drop it. That leaves five elements in $\mathbf{M'}$ and six in $\mathbf{V'}$, which yields $2^5 \times 2^6 = 2048$ permutations.

[9] These mechanics follow Baker, Bloom and Davis (2016) exactly.

[10] In this way, our method has parallels to the rich literature on feature or subset selection. Much of the machine learning literature may begin with a richer feature set (eg. all words ever seen in newspapers) and reduces this feature space by penalizing terms that add complexity but do not meaningfully increase performance in-sample (see Cherkassky and Ma 2004). We instead choose to first reduce our set of terms to a parsimonious one by hand and through consideration of each term independently. Only after selecting such a parsimonious set of terms do we allow performance in-sample to govern further reductions in the feature space.

Armed with our best-fit term set, we obtain monthly counts of articles that contain at least one term in each of **E, M** and **V** for eleven major U.S. newspapers: the Boston Globe, Chicago Tribune, Dallas Morning News, Houston Chronicle, Los Angeles Times, Miami Herald, New York Times, San Francisco Chronicle, USA Today, Wall Street Journal, and Washington Post. At this stage, we use counts from the full set of articles published in each newspaper, not a sample, and we again scale by the count of all articles in the same paper and month.[11] We then standardize the scaled counts and average over newspapers by month. In a final step, we multiplicatively rescale our best-fit EMV tracker to match the mean value of the VIX from 1985 to 2015.

Figure 1 displays our EMV tracker from January 1985 to October 2018.[12] The series exhibits pronounced upward spikes in reaction to the 1987 stock market crash, the 1998 Russian financial crisis, the Enron and WorldCom accounting scandals and bankruptcies in 2001-2002, the full-force eruption of the financial crisis in September 2008, and the U.S. debt-ceiling crisis in the summer of 2011. Several other episodes triggered smaller spikes. We validate our EMV tracker, assess its performance in various ways, and consider robustness checks in Section 3 below. Before doing so, we explain how to construct our category-specific trackers.

To demonstrate the ability of our news-based EMV methodology to push back further in time, we also construct a historical EMV index that extends backwards from 1984 to 1928, augmenting our more contemporary index. For this index, we utilize data from the Proquest Historical Archive for the New York Times, Wall Street Journal, Boston Globe, Chicago Tribune, Washington Post, and Los Angeles Times.

2.2 Parsing the Text and Constructing Category-Specific Trackers

We parse the text in our best-fit EMV articles to quantify journalist perceptions about the particular forces that drive volatility in equity returns. As a first step, we classify these forces into 20 general economic categories and about 20 policy-related categories, including subcategories.

[11] The reader might wonder why we don't use all eleven papers in the term set selection procedure. The answer is purely one of feasibility. We cannot obtain a large sample of machine-readable articles for most newspapers. Nor can we put millions of queries to digital newspaper archives to cover all the permutations of **M′** and **V′**. Given the **E, M** and **V** sets, however, we need only two article counts per paper per month – the EMV count and the "all" count.

[12]Data for the CBOE 30-day VIX starts in 1990. After selecting our best-firm term set using data from 1990 to 2015, we obtained the VIX data developed in Berger et al. (2019) back to 1983. Thus, our EMV tracker data before 1990 and after 2015 are "out of sample" in the sense that they are outside the period used in our term selection procedure.

These classifications provide a basis for assessing the importance of each category for the average level of stock market volatility and its movements over time.

Our classification approach is conceptually simple: If certain category-relevant terms appear in an EMV article, we infer that the article discusses one or more topics covered by the category in question. For example, consider our term sets for **Interest Rates** (one of our general categories) and **Monetary Policy** (one of our policy categories):

Interest Rates: {interest rates, yield curve, fed funds rate, overnight rate, repo rate, T-bill rate, bond rate, bond yield}

Monetary Policy: {monetary policy, money supply, open market operations, fed funds rate, discount window, quantitative easing, forward guidance, interest on reserves, taper tantrum, Fed chair, Greenspan, Bernanke, Volker, Yellen, Draghi, Kuroda, Jerome Powell, lender of last resort, central bank, federal reserve, the fed, European Central Bank, ecb, Bank of England, bank of japan, people's bank of china, pboc, pbc, central bank of china, Bank of Italy, Bundesbank}

If an EMV article contains one or more terms in **Interest Rates,** we infer that the article includes a discussion of interest rates; likewise, if it contains one or more terms in **Monetary Policy,** we infer that it discusses monetary policy. As these examples suggest, many EMV articles contain terms in more than one category. That is by design. We do not draw overly sharp boundaries between overlapping categories, nor do we aim to draw distinctions that are too fine for our text sources and methods. Appendix B sets forth a complete listing of our category-specific term sets.

Next, we calculate the share of EMV articles in each category and multiply by the EMV tracker value to obtain category-specific trackers. For example, to measure the importance of monetary policy considerations in equity market volatility during month *t*, we calculate

$$\left(\frac{\#\,\{\boldsymbol{E} \cap \boldsymbol{M} \cap \boldsymbol{V} \cap \textbf{Monetary Policy}\}_t}{\#\,\{\boldsymbol{E} \cap \boldsymbol{M} \cap \boldsymbol{V}\}_t}\right) EMV_t,$$

where # denotes the count of newspaper articles in the indicated set, and EMV_t is the value of our overall EMV tracker in month *t.* We use this same approach for all categories.

As before, a few additional remarks will be helpful in assessing our method. First, the overfitting concern that led us to start with parsimonious **E, M′** and **V′** sets in developing our overall EMV tracker is no longer germane, because we have already identified our best-fit EMV articles. At this point, our goal is to capture and classify the full set of topics and concerns that animate discussions of stock market volatility in the EMV articles. Thus, several of our category-specific sets contain many terms. **Monetary Policy,** for example, has more than 25 terms. Other categories with lengthy term sets include **Macroeconomic News & Outlook, Commodity Markets, Taxes,** and **Financial Regulation.**

Second, while we deliberately avoid particularistic terms like “Brexit” or “Bernanke” and “Northern Rock” in constructing our overall EMV tracker, we embrace them in devising our category-specific term sets. The difference in approach reflects a difference in objectives. In developing our overall EMV tracker, we seek a measure with good prospects for fitting well out of sample and ready portability to other national stock markets and eras. In contrast, we design the category-specific term sets to characterize and quantify the specific forces that underlie stock market volatility and its variability over time and space.

We recognize that our category-specific sets require considerable modification when applied to other countries and time periods. In essence, these more specific categorical indexes act as an accounting exercise to apportion EMV articles to various topics, even when such topics are highly local to a particular setting. Still, our roughly 40 categories are portable over time and space, even when many of the category-specific terms are not.

Third, our term sets for the policy-related categories extend Baker, Bloom and Davis (2016) and Davis (2017). They populate their category-specific term sets by consulting textbooks, newspapers, “risk factor” discussions in 10-K filings, and other sources – including their own knowledge of economic matters and input from other economists in seminars and personal communications. We extend the policy-related term sets of BBD and Davis and build term sets for the general economic categories using the same basic approach. Thus, our classification approach is expert-driven and judgmental, in contrast to the algorithmic use of external libraries to classify *n*-grams as in Hassan et al. (2019), who borrow methods from computational linguistics.

Table 1 considers all EMV articles from 1985 to 2018 and reports the percent that contain terms in each category.[13] The top row says that news and other remarks about the Macroeconomic Outlook feature very prominently, appearing in 72% of all EMV articles. News about Commodity Markets appear in 44% of EMV articles, while news about Interest Rates figures in 31%. Panel B in Table 1 considers our policy-related categories, including aggregated categories for Fiscal Policy and Regulation. Tax Policy and Monetary Policy each receive attention in 30% of EMV

[13] The column entries sum to more than 100 percent for two reasons: First, because certain terms such as “Fed funds rate” appear in the term set for more than one category. Second, because many EMV articles refer to multiple sources of equity market volatility.

articles, the aggregated Regulation category features in 25%, and National Security matters figure in 13%. Most other categories play a small role over the 1985-2018 period as a whole, although they can become prominent in certain episodes, as we show below.

3. Validation and Robustness Checks

3.1 EMV Tracking Performance

Table 2 provides information about how well our EMV measure tracks monthly movements in stock market volatility from 1985 to 2018. As reported in column (1), regressing the VIX on contemporaneous EMV values yields a highly statistically significant slope coefficient of 0.76 and an R-squared value of 0.61. The first two lags of EMV are also statistically significant, and their inclusion raises the R-squared to 0.70. Adding lagged VIX pushes the R-squared value well above 0.8 and knocks out the statistical power of the lagged EMV terms, but the contemporaneous EMV term remains highly significant. Log-log specifications and regressions of realized stock market volatility on EMV yield similar results.

Figure 2 plots the VIX and the fitted values for the column (1) specification. For the most part, fitted values – and the underlying EMV values – move closely with the VIX. There are some exceptions: (i) fitted VIX jumps less than actual VIX in reaction to the October 1987 stock market crash, (ii) fitted VIX largely misses the VIX reaction to the Iraqi invasion of Kuwait in August 1990, (iii) fitted VIX persistently exceeds the VIX from 1993 to 1996 and 2005 to early 2007, and (iv) fitted VIX reverts to the mean more quickly than actual VIX after major upward spikes, a pattern that is most evident for the cataclysmic events of September-November 2008.

We could address (i) and (ii) by incorporating episode-specific terms like “Black Monday” and “Kuwait invasion” into our EMV term sets. We refrain from that approach for reasons discussed in Section 2.1. Fit errors of type (iv) reflect how press coverage evolves after surprise events that jolt financial markets. In the immediate wake of events like 9-11 and the 2011 U.S. debt-ceiling crisis, an outpouring of newspaper articles discusses the event and its bearing on stock market volatility. Elevated volatility levels persist, but press coverage abates as the event loses its newness. As a result, our EMV tracker drops relative to the VIX in the near-term aftermath of such events. A closer examination of regression residuals in Appendix Figure C.1 reinforces this interpretation. Appendix Figure C.2 shows that adding lagged VIX to the regression specification greatly dampens fit errors of type (iv), except for the stock market crash of 1987. While this temporal pattern in the residuals is an interesting commentary on press coverage, it does not undercut the usefulness of EMV for our purposes. In any event, adding lagged VIX to the regression specification largely resolves this type of tracking error as well as errors of type (iii).

3.2 EMV Performance at Different Horizons

In Table 3, we extend our analysis of the performance of our EMV tracker against the VIX. In particular, we test the predictive power of EMV against a range of horizons of market implied volatility. For each regression, we include contemporaneous monthly EMV (t), the average of three months lagged EMV (eg. average of t-1, t-2, and t-3), and the average of twelve months lagged EMV. The independent variable in each regression is the implied market volatility at various horizons, from one month to 10 years. In Table 2, the baseline VIX measure uses a one-month horizon.

In general, we find that our EMV tracker offers substantial predictive power for implied volatility at all horizons. However, as the horizon lengthens, the average lagged values of EMV provide much more explanatory power relative to the contemporaneous monthly value of EMV. That is, the lower frequency fluctuations in EMV predict longer horizons of implied volatility.

For all horizons out to 10 years, the average twelve month lagged EMV has a significant correlation with implied volatility, while contemporaneous monthly EMV offers a significant correlation only for horizons of a year or less. In addition, R-squared declines monotonically as the horizon increases in length.

We pursue this line of analysis further in Table C.1. Here we regress monthly EMV, one and two month lagged EMV and lagged VIX against the same set of horizons of implied volatility. In all cases, EMV has predictive power for contemporaneous levels of VIX, even controlling for lagged VIX. As in Table 3, as the horizon of implied volatility lengthens, the explanatory power of contemporaneous EMV is reduced substantially.

Beyond the composite EMV index, we examine the relationship between our more detailed categorical EMV trackers and the varying horizons of the VIX. We use a LASSO approach to select categories and display the post-selection regressions in Table 3b. We find substantial variation in the categories that correlate well with the different horizons of VIX. While Financial

Regulation and Macro – Broad Quantity Indicators seem to be important components of variation across all horizons, other subcategories are selected for only one horizon. For instance, EMV related to National Security or Real Estate tend to be highly linked to short horizons of VIX while EMV related to Regulation tends to be more closely related to longer term VIX horizons. This exercise highlights the benefits of our categorical decompositions which tend to have substantial differences in their overarching trends as well as quite different linkages to important financial market constructs.

3.3 Correlations with Future Equity Market Returns

We investigate whether our EMV indexes track volatility that is priced by the market in Table 4. We regress our composite EMV index as well as some of the categorical EMV trackers against total market returns over a range of horizons from 1 month to 2 years. We use one-month lagged values of the EMV indexes as dependent variables to alleviate concerns that our measures are picking up variation not yet priced by markets, as in Manela and Moreira (2017). In the first row in Table 4, we see that our composite EMV index has a positive associate with future returns across all horizons, consistent with EMV tracking uncertainty priced by markets, though is only significant at a longer horizon of 2 years.

The remaining three rows in Table 4 describe an equivalent exercise using categories of EMV rather than the composite EMV index. We see that each of these EMV trackers has positive correlations with future equity returns across all horizons (other than a single small and insignificant coefficient at a 1 month horizon for EMV – Regulation). However, some of our subcategories have substantially different patterns of correlation with excess returns over these horizons. For instance, EMV related to Macroeconomic News has some predictive power across a range of horizons. EMV linked to Regulation only has predictive power over longer horizons while EMV related to National Security is the reverse: only yielding predictive power over market returns in the short run. Such variation is consistent with the uncertainty underlying each EMV category being resolved at different. So, for instance the impacts of regulation may only become apparent in the longer run so the priced equity premium for holding stocks during high regulatory EMV may not be high in the short run. Conversely, national security EMV may be resolved more quickly so the equity premium disappears after a shorter horizon.

3.4 Comparison to NVIX

Manela and Moreira (2017) construct a monthly news-based implied volatility (NVIX) measure using abstracts and headlines of front-page articles in the *Wall Street Journal.* From this text source, they create large "feature sets" of *n*-grams that serve as explanatory variables in support vector regressions fit to the VIX. While their method and text source differ from ours, the spirit of their statistical undertaking is similar. As another check on EMV, we now assess how it fares – with respect to similarity to the VIX – in comparison to NVIX. We use monthly data from January 1985 to March 2016 for this purpose, the longest overlap period for the three measures.

Table 5 reports summary statistics from this analysis. EMV correlates with the VIX at 0.78, which compares to 0.70 for NVIX in the modern period. The mean absolute monthly difference between EMV and VIX is 3.7 points, as compared to 4 points for NVIX. The EMV standard deviation, skewness, and kurtosis are much closer to the corresponding VIX statistics. Turning to Figure 3, we see that NVIX underperforms EMV in tracking the VIX during the second half of the 1980s and from 2012 to 2015. NVIX performs better than EMV in 1990 around the time of the Iraqi invasion of Kuwait.

The scaling factor are 0.81 (NVIX) and 1 (EMV). In Panels B and C, we scale NVIX and EMV to match the mean value of RVol from 1960 to 1994 and 1928 to 1959, respectively. The scaling factors required for this are 0.52 (NVIX) and 2.79 (EMV) in Panel B and 0.72 (NVIX), 10.51 (EMV) in Panel C. Appendix Figures A2 and A3 display our Historical EMV index plotted alongside the NVIX and realized volatility for these two periods.

In summary, each measure has weaknesses and strengths, but EMV generally outperforms NVIX in matching the moments of the VIX and realized volatility. A big reason for EMV's superior performance is its reliance on a much larger text corpus – the full text of eleven major newspapers, as compared to abstracts and headlines of front-page articles in a single paper. In fact, when we rerun specification (1) in Table 2 using an EMV measure based on a single paper, the R-squared value drops drastically – by 17 to 38 percentage points. See Appendix Table C.2.[14]

[14] In Appendix Table C.2, we also show that EMV and NVIX each have independent explanatory power in VIX regressions, regardless of whether we control for lagged VIX. This pattern suggests that applying machine-leaning methods to the full text of our eleven newspapers would materially improve on our EMV tracker. Of course, implementing such an approach requires direct access to the full text of each paper.

3.5 Robustness to Alternative Newspaper Weightings

We also assess the assumption, implicit in our method, that each newspaper is equally useful (on the margin) in tracking equity market volatility. To do so, we double the weight on each newspaper, one at a time, in constructing EMV. Then we rerun specification (1) in Table 2 using the EMV tracker based on the modified newspaper-level weights. Appendix Table C.3 reports the results. Doubling the weight on the *Wall Street Journal* or the *Miami Herald* yields an incremental R-squared gain of .002 to .004. Doubling the weight on the *San Francisco Chronicle* leaves the R-squared unchanged, and doubling the weight on any other paper lowers the R-squared, with a maximal drop of 0.011. We also drop each newspaper, one at a time, and repeat the exercise. In two cases, dropping the paper yields a modest fit improvement, in one case it has no effect, and in the other eight cases fit deteriorates modestly. The largest absolute change in the R-squared value from dropping newspapers is only 0.013.

We draw three conclusions from these results. First, tracking performance improves greatly by drawing on multiple newspapers. Second, the performance of our preferred EMV measure is robust to alternative newspaper weightings on the margin (i.e., given eleven papers in our baseline). Third, while using multiple newspapers yields huge performance gains, the gains are subject to strong diminishing returns. Eleven papers appear sufficient to exhaust the largest gains. Of course, we cannot preclude the possibility that an untried newspaper would materially improve EMV tracking performance. However, even the financially-oriented *Wall Street Journal* matters little on the margin, which casts doubt on the notion that an untried paper would add a lot.

3.6 Petroleum Markets EMV Tracker

We now subject our method to a different type of assessment, one that is especially pertinent for our category-specific measures. Specifically, we construct a Petroleum Markets EMV tracker and compare it to observed measures of oil price volatility. To that end, define a **Petroleum Markets** term set, {oil, petroleum, crude, gas}, and compute

$$\left(\frac{\#\,\{\boldsymbol{E} \cap \boldsymbol{M} \cap \boldsymbol{V} \cap \textbf{Petroleum Markets}\}_t}{\#\,\{\boldsymbol{E} \cap \boldsymbol{M} \cap \boldsymbol{V}\}_t}\right) EMV_t.$$

This Petroleum Markets EMV tracker correlates at 0.60 with the CBOE Crude Oil Volatility Index (0.69 in quarterly data) from 2007 to 2018 and at 0.50 with the CBOE Crude Oil Realized Volatility (0.56 in quarterly data) from 1986 to 2018. Inspecting Figure 4 confirms that our measure mirrors many of the movements in oil price volatility. It also misses badly in certain episodes, e.g., after the stock market crash of 1987 and during the Global Financial Crisis. These episodes involve much larger jumps in stock price volatility than oil price volatility. Hence, it's no surprise that our measure, with its focus on equity markets, remains highly sensitive to these events even when we narrow its scope to petroleum markets. Nor is this sensitivity a problem for our purposes, given that we aim to characterize the sources of *equity* market volatility.

In summary, Figure 4 gives assurance that our category-specific EMV trackers capture variation in the role of the corresponding topics and concerns as drivers of equity market volatility. We interpret our category-specific EMV trackers accordingly.

4. What Drives Fluctuations in Aggregate Stock Market Volatility?

4.1 News About the Economic Outlook

Figure 5 displays our EMV tracker for **Macroeconomic News and Outlook,** which contains about 80 terms and reflects an expansive conception of the category. Since topics covered by this category appear in 72 percent of EMV articles (Table 1), Macro EMV moves similarly to overall EMV and to the VIX. For example, the Macro EMV tracker jumps in reaction to the October 1987 stock market crash, the Russian Financial Crisis, the Global Financial Crisis, and the debt-ceiling crisis in summer 2011 – episodes that involved major upsurges in uncertainty about the macroeconomic outlook. In contrast, the Enron and WorldCom scandals – which arguably injected little uncertainty about the macro outlook – generated a muted response in Macro EMV relative to overall EMV (Figure 1) and relative to the VIX (Figure 2).

Our method is easily adapted to more tightly focused EMV trackers. As an illustration, Figure 6 plots a Financial Crisis EMV tracker based on the following term set: {financial crisis, financial crises, Northern Rock failure, Lehman failure, Lehman Brothers failure, AIG Takeover, euro crisis, Eurozone crisis, Greek crisis}. Two events stand out in the evolution of this EMV tracker: the Global Financial Crisis, and the U.S. debt-ceiling crisis of 2011. The Mexican Peso Crisis of 1994, the Asian and Russian Financial Crises of 1997-98, and concerns related to Greece and China in 2015 also leave clear marks on our Financial Crisis EMV tracker. Interestingly, the tracker's baseline level is consistently higher after the Global Financial Crisis, which suggests that the GFC prompted a persistent shift in perceptions about the relevance of financial crises to U.S. stock market volatility.

4.2 How Big a Role for Animal Spirits?

We find a rather modest role for consumer sentiment and business confidence measures in explaining the lower frequency time series variation in equity market volatility. Figure 16 and 17 plot two of our Macroeconomic News and Outlook subcategory EMV Trackers: Business Investment and Sentiment, Consumer Spending and Sentiment. We see relatively large roles for both business and consumer sentiment in the early 2000's following the dot-com bubble crash. In addition, there are notable spikes in the Macroeconomic Sentiment tracker in 1987-1988, 1998, the Great Recession, and in 2011. The Business Investment and Sentiment tracker mirrors some of these patterns, but has a quite muted response to the Great Recession.

We also construct a more general 'Sentiment' term set that combines some terms from the business and consumer sentiment topics along with more general sentiment related terms. Figure 18 plots the share of EMV articles that also has a term from this broader 'Sentiment' term set. This time series is stable over the long-run, hovering around 0.45 over the entire time period. While this share is quite large at all times, there are no striking time series patterns or particular periods where sentiment and animal spirits consistently reigned supreme.

4.3 The Growing Role of Policy Matters

Figure 7 reveals an upward drift in the fraction of EMV articles that devote attention to policy matters, with peaks in the 2001-03 period (9/11 and Iraq Invasion), the 2011-12 period (U.S. debt-ceiling crisis and the "fiscal cliff"), and the period since Donald Trump's presidential election in November 2016. To construct Figure 7, we sum EMV article counts over the policy-related categories in Table 1 and divide by the EMV article count summed over all categories – both general economic and policy-related categories.[15] We take this approach, because limits on the number of terms per search query prevent us from directly computing the share of EMV articles that contain one or more of our policy-related terms. As a robustness check, we performed the direct calculation using the much smaller set of "Policy" terms in the Economic Policy Uncertainty Index.[16] This alternative calculation, reported in Appendix Figure C.3, also shows an upward drift in the policy fraction of EMV articles, broadly in with Figure 7.

---

[15] For Fiscal Policy and Regulation, we use article counts for the more disaggregated categories.

[16] Their Policy term set is {regulation, regulations, regulatory, deficit, deficits, legislation, legislative, legislature, white house, federal reserve, the fed, congressional, congress, war, tariff.}

This upward drift suggests a growing role for policy concerns in U.S. stock market volatility. It resonates with other evidence of an expanding government role in the economy and an upward trend in policy-related economic uncertainty, as discussed in Baker et al. (2014) and Davis (2017): secular growth in government expenditures as a share of GDP, the increasing scale and complexity of the regulatory system, the increasing complexity of the federal tax code, the growing share of business "risk factors" that U.S. firms attribute to government policy in their 10-K filings, a secular rise in the newspaper-based Economic Policy Uncertainty Index of Baker, Bloom and Davis (2016), an upward drift in the frequency with which the Federal Reserve System's Beige Books refer to policy uncertainty, and a secular rise in U.S. political polarization that has drawn enormous attention from political scientists. On this last point, see, e.g., McCarty, Poole and Rosenthal (2016). Since these long-term developments show little sign of reversal, policy concerns are likely to remain a major source of stock market volatility for many years.

As suggested by the annotations in Figure 7, the mix of policy-related factors in stock market volatility varies over time. We can use our category-specific term sets to develop this point in a systematic, quantitative manner. As an illustration, Figure 8 displays the percent of EMV articles by month that contain one or more terms in **Trade Policy.** The figure shows a dramatic upsurge in trade policy concerns as a source of stock market volatility after Donald Trump's election. Days after his inauguration, President Trump pulled the United States out of the Trans-Pacific Partnership, which had yet to be ratified. He threatened to jettison the North American Free Trade Agreement, triggering contentious trade negotiations with Canada and Mexico. The Trump administration also imposed tariff hikes on steel, aluminum and other goods and has threatened to impose many more. Since March 2018, the United States has ratcheted up tariffs and tariff threats with China, and the China has responded in kind. These are among the developments that took trade policy concerns from a virtual nonfactor in U.S. stock market volatility to a leading source.[17]

4.4 Policy-Related EMV Compared to Economic Policy Uncertainty

We now construct a Policy-Related EMV tracker and compare it to the Economic Policy Uncertainty Index of Baker, Bloom and Davis (2016). While both measures rely on scaled frequency counts of newspaper articles, they are conceptually distinct. The EPU Index aims to

[17] For a fuller account of trade policy developments since the November 2016 election, see the Trade War Timeline in Brown and Kolb (2019).

quantify policy-related uncertainty for the economy as a whole. The Policy-Related EMV tracker aims to quantify the full range of policy-related volatility sources for the stock market in particular. To obtain our Policy-Related EMV tracker, we multiply the overall EMV tracker in Figure 1 by the policy-related fraction in Figure 7. We then multiplicatively rescale to match the mean EPU value from 1985 to 2009, so that we can readily compare the two series.

Figure 8 displays the comparison. The two measures respond to many of the same developments but also show systematic differences. Stock market crashes and financial crises leave larger marks on Policy-Related EMV. National security developments, national elections, and fiscal policy conflicts are more visible in the EPU Index.

Panel B in Table 1 makes a closely related point. Financial Regulation receives attention in 25% of EMV articles as compared to 6% of EPU articles. In contrast, National Security, Healthcare Policy, and Entitlement and Welfare Programs are among the policy-related categories that loom larger for Economic Policy Uncertainty than Equity Market Volatility. Reassuringly, policy-related discussions appear more frequently in EPU than in EMV articles as we use "policy" terms in identifying EPU articles.

4.5 A Suite of Policy-Related EMV Trackers

We now implement the method in Section 2.2 to construct a suite of policy-related EMV trackers. Figures 10, 11 and 12 display EMV trackers for monetary and fiscal policy categories. Certain events loom large in all three trackers: the stock market crash of October 1987 and the debt-ceiling crisis of 2011.

However, other events are particularly distinct when comparing different policy categories. For instance, in the Monetary Policy EMV tracker, several unique events are noticeable spikes: the start of QE1 and QE2, the Taper Tantrum, and the July 2015 Greek Referendum that shook the Eurozone. Other events are more prominent in the Tax Policy EMV tracker: the Bush Tax Cuts of 2001 and 2003, the Fiscal Cliff episode in late 2012 and early 2013, and the Tax Cut and Jobs Act enacted in November 2017. Yet other events are prominent in the EMV tracker for Government Spending, Deficits and Debt: the Government Shutdowns of 1995-96 and 2013 and the Fiscal Cliff.

The EMV tracker for Financial Regulation in Figure 13 shows large upward spikes around the enactment of the Sarbanes-Oxley Act of 2002, during the Global Financial Crisis, and around the time of the Dodd-Frank Act of 2010. The EMV tracker for Elections and Political Governance in

Figure 14 fluctuates at low levels except for short time windows around the U.S. presidential elections of 2000, 2016 and, to a lesser extent, 1992. The National Security EMV tracker in Figure 15 exhibits large upward spikes around Gulf War I, the 9-11 attacks, and the early stages of Gulf War II. EMV trackers for Healthcare Policy and Trade Policy (Figures C.4 and C.5) also show distinctive fluctuations. All of the underlying data for these figures, and more, are available at http://www.policyuncertainty.com/EMV_monthly.html, with regular monthly updates.

To summarize, these figures show highly distinctive temporal movements in the category-specific EMV trackers. Certain events, most notably the market crash of 1987, leave a strong mark in most or all of the category-specific trackers. Many other events, however, leave a strong mark in only one or a few of the category-specific trackers. The distinctiveness of the temporal patterns in the category-specific trackers is potentially quite useful in downstream econometric work that seeks to explain firm-level outcomes. Moreover, the quantification of the relative size of such events is an important contribution of our approach. Not only can this methodology identify periods or events of particular importance to market volatility, but it can help to quantify the relative impacts of qualitatively very different news.

5. Do Our EMV Trackers Help Explain Firm-Level Stock Price Volatilities?

5.1 Using Part 1A in 10-K Filings to Quantify Firm-Level Risk Exposures

In 2005, the U.S. Securities and Exchange Commission (SEC) issued a regulation that requires most publicly held firms to include a discussion of "Risk Factors" in Part 1A of their annual 10-K filings. In "How to Read a 10-K" at www.sec.gov/answers/reada10k.htm, the SEC describes Part 1A as follows:

> **Item 1A - "Risk Factors"** includes information about the most significant risks that apply to the company or to its securities. Companies generally list the risk factors in order of their importance. In practice, this section focuses on the risks themselves, not how the company addresses those risks. Some risks may be true for the entire economy, some may apply only to the company's industry sector or geographic region, and some may be unique to the company.

See Campbell et al. (2014) for an extended discussion and analysis of this regulatory development.

Baker, Bloom and Davis (2016) use the text in Part 1 to quantify firm-level policy risk exposures, which they combine with their newspaper-based economic policy uncertainty index to explain firm-level stock price volatilities, investment rates and employment growth rates in a panel regression setting. Davis and Seminario (2019) also use the text in Part 1A to quantify firm-level policy risk exposures.

We build on this approach. In particular, we use computer-automated methods to read the 10-K filings and measure the frequency of sentences in Part 1A that contain particular terms of interest. Our text analysis of Part 1A covers 10-K filings in calendar years 2006 to 2019 (fiscal years 2005 to 2018). Most 10-K filings occur in or around March.[18]

From each of these reports, we then use automated methods to count the number of sentences in each Part 1A section that pertain to each of the Equity Market Volatility topics. After obtaining this count for each firm-year observation, we divide by the total number of sentences in the same Part 1A.

$$Firm_{i,y}^{b} = F_{i,y}^{b} = \frac{(\#\ of\ sentences\ about\ EMV\ topic\ b)_{i,y}}{(\#\ of\ total\ sentences\ in\ Part\ 1A\ of\ 10K)_{i,y}}$$

Table 4 reports summary statistics of the denominator of our 10-K F measures. One major takeaway is the overall increasing trend in the total number of Part 1A sentences over time.

5.2 Combining the Newspaper-Based EMV Indices with 10-K Information

We use the constructed firm-year level topic weights as measures of the exposure of firms to each topic. We employ them as interaction terms in our panel regression of firm level monthly realized volatility on a composite EMV topic index as follows:

$$\sigma_{i,t} = \alpha_i + \gamma_t + \beta \sum F_{i,y}^{b}\, EMV_t^{b} + \epsilon_{i,t}$$

where t represents the monthly time level, $\sigma_{i,t}$ stands for the firm monthly realized volatility, and α_i and γ_t are firm and time fixed effects respectively.[19] That is, we seek to identify the extent to

[18] We drop filings for which the automated sentence counter returns a value of less than nine for the Part 1A section as these, upon inspection, typically represent routine headings and section separators of 10-K filings with an empty Part 1A. When the same firm filed multiple 10-K files on the same date, we retain the one with the longer Part 1A. When a firm has more than one 10-K filing in the same calendar year, we retime the "early" ("late") filing to the prior (next) calendar year provided the firm has no filing in the prior (next) calendar year. If a firm still has multiple 10-K filings in the same calendar year, we retain the file with the longer Part 1A.

[19] See the Appendix D - Computing Firm-Level Stock Returns for discussion of the calculation of firm-level monthly realized volatility. Data is obtained from CRSP and includes 508,420 observations with an average of 36,316 firms per year of coverage.

which a firm's exposure to different components of overall risk drives the relationship between their firm-level volatility and our EMV index.

When we assign the 10-K F measures to the given EMV months, we use the most recent 10-K filing with a non-empty Part 1A. For example, suppose the firm files its 10-K in March of each calendar year. Then to construct its F values from, say, April 2013 to March 2014, we combine its 10-K filing in March of 2013 with the topic-specific EMV tracker values from April 2013 to March 2014.

Note, that this regression has one coefficient for the entire composite EMV topics index weighted by our 10-K exposure weights. We will also consider specifications that break apart this composite variable into various topic subsets. In some cases, given the large number of variables that we consider, we use LASSO methods to select variables and then report the resulting OLS specification using the LASSO-selected variables.[20]

Before we report results from those regression specifications, we turn to some summary statistics of the dependent variable, firm-level realized volatility that are reported in Table 5. The summary statistics are relatively stable across the time period in consideration with the one notable exception being the much higher mean value during the Great Recession.

Table 6 presents results of our realized volatility regressions. Column (1) considers the baseline regression of realized volatility on our EMV Topics Composite variable. We see a strongly significant positive relationship between the two. Columns (2)-(4) break this relationship down between non-policy and policy categories respectively. In particular, column (2) considers the EMV Topics Composite variable only constructed using the non-policy EMV topics, column (3) considers only the policy EMV topics in the variable construction, and column (4) includes both the non-policy and policy EMV Topics Composite variables.

Despite both non-policy and policy factors being strongly significant independently, when considered together, the non-policy composite variable carries most of the weight. This is reinforced by the specification in column (5). In this column, we consider a LASSO specification using the full set of EMV Topics Composite variables constructed using each of the various EMV topics separately giving us a set of 38 variables to consider. Column (5) then reports the OLS

[20] We cluster standard errors at a firm level and include firm and period level fixed effects so that our LASSO selection procedure corresponds directly with our panel OLS specifications.

results using the LASSO-selected variables. The three selected topics are each non-policy topics: Interest Rates, Real Estate Markets, and Commodity Markets.

5.3 Incorporating Firm Characteristics

Table 7 reports results of our realized volatility on EMV Topics Composite regressions considering various firm characteristics and their interactions with our EMV indices. Columns (1)-(5) report the same specifications from Table 6 but now using the subsample of observations that have data on the 62 firm characteristics we consider. These 62 firm characteristics are borrowed from Freyberger et al. (2019). The results are comparable to Table 6 with the only notable difference being that the Macro – Real Estate Markets sub-index is not selected in the EMV categories LASSO specification in column (5). Columns (6), (7), and (8) extend the LASSO specification to include the set of firm characteristics and their interactions with the 10-K EMV Composite variable, both the policy and non-policy 10-K EMV Composite variables, and each of the 38 individual EMV categories respectively.

The variables selected from the specification in column (6) are as follows: ratio of book value of equity to market value of equity, return-on-equity, sales-to-price ratio, assets-to-market cap, cash flow to price ratio, return on invested capital, the average bid-ask spread and its interaction, closeness to 52-week high and its interaction, momentum, long-term reversal, CAPM beta and its interaction, daily CAPM beta, total volatility and its interaction, standard deviation of daily turnover regression residuals, the interaction with cash and short-term investments ratio to total assets, the interaction with short-term reversal, and the interaction with the cumulative return from 6 months to two months before.

The extended specifications in columns (7) and (8) see many of the same firm characteristics being selected with the additional granularity in 10-K EMV topics adding to the number of EMV-related variables selected. We see a noticeable increase in the within R^2 when adding in the firm characteristic interactions from a value of 0.0039 in column (5) of Table 6 to a value of 0.146 in column (8) of Table 7. In addition, we see substantial policy-related EMV components selected in the LASSO approach, with about 20% of the selected variables in column (8) being policy-related EMV variables.

With these various specifications in hand, we can turn our attention to how well our EMV trackers explain firm-level stock price volatilities. First, we consider how well our preferred specification in column (6) of Table 7 fits the time series movements in the cross-sectional standard deviation of firm-level realized volatility. Figure 19 plots this standard deviation time series for both the actual realized volatility data along with the predicted values from our preferred specification. In both cases, time and firm fixed effects have been swept out prior to the analysis. The figure also displays the ratio of the two time series. Over time, the fit remains relatively stable with the ratio hovering around 0.4. Interestingly, we see the best fit around the Great Recession with the cross-sectional standard deviation of actual realized volatility increasing, but our predicted values increasing even more to compensate.

We consider the fit of our model in another way as well by looking at the firm-level time series correlations between the actual and fitted realized volatility series. For each firm, we can calculate the correlation we get from their actual realized volatility series and the predicted series produced by our preferred specification. Figure 20 displays the histogram of these firm-level correlation coefficients. Again, we see a similar story to before with the median coefficient around 0.34.[21] Most firms have a positive correlation between the actual and predicted series.

Table 8 breaks down this correlation structure across a few of the more notable firm characteristics that we consider. Specifically, we report the average correlation for firms in the five quintiles of each of the six firm characteristics presented. These characteristics were chosen because of the monotone relationship we see in the average correlation across quintiles. For characteristics that represent firm size mostly such as total assets and the assets to market capitalization ratio, we see a positive trend of increasing average correlations across quintiles. This is sensible as we expect better coverage and discussion of relevant factors in the 10-K reports of larger firms, and these large firms may be the most susceptible to aggregate volatility movements that our EMV trackers capture.

6. Summary and Directions for Research

We develop a simple, transparent, scalable method for constructing newspaper-based Equity Market Volatility (EMV) trackers. Implementing the method using eleven major U.S. newspapers, our EMV tracker moves closely with the VIX and with realized volatility on the S&P 500.

[21] This median value increases to 0.41 when considering only a balanced panel of firms. Also, these correlations all increase substantially when considering quarterly or yearly average values. The median for the quarterly and yearly versions are 0.69 and 0.87 respectively.

We also parse the text in the EMV articles to quantify journalist perceptions about the forces that underlie stock market volatility and its movements over time. We classify these forces into about forty categories – including Macroeconomic News, Monetary Policy, Tax Policy and Financial Regulation – and construct a tailored EMV tracker for each category.

This exercise reveals an upward drift over time in the role of policy as a source of stock market volatility, as measured by the share of EMV articles that discuss policy-related matters. It also reveals Monetary Policy and Tax Policy to be the most important policy-related sources of stock market volatility, followed by our aggregated Regulation category. The contribution of specific policy categories to stock market volatility fluctuates markedly over time. We then use our category-specific EMV trackers to explain and interpret the distribution of firm-level stock price volatilities and its movements over time.

There are several natural directions for future research. One primary direction is in extending this measurement approach globally. Our methods extend readily to any country or time period with digital newspaper archives and data on aggregate equity returns. By developing EMV trackers for multiple countries, one can explore the specific global and national forces that underlie stock market volatilities around the world. In addition, our basic approach could be usefully applied to construct and parse newspaper-based trackers for other concepts. It would be straightforward, for example, to adapt our methods to construct newspaper-based trackers of consumer confidence or other concepts and to delve into the forces that drive their movements.

References

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold, 2002. "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1092.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega, 2003. "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38-62.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega, 2007. "Real-Time price discovery in global stock, bond, and foreign exchange markets," *Journal of International Economics*, 73, 251-277.

Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims, 2013. "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5, 2, 217-49.

Baker, Scott R, Nicholas Bloom, Brandice Canes-Wrone, Steven J. Davis and Jonathan Rodden, 2014. "Why Has Policy Uncertainty Risen Since 1960," *American Economic Review, Papers & Proceedings,* 104, no. 5 (May), 56-60.

Baker, Scott R, Nicholas Bloom and Steven J. Davis, 2016. "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 131, no. 4 (November), 1593-1636.

Baker, Scott R, Nicholas Bloom, Steve J. Davis and Marco Sammon, 2019. "What triggers stock market jumps?" work in progress presented at the January 2019 ASSA meetings.

Barro, Robert J., 2006. "Rare Disasters and Asset Markets in the Twentieth Century," *American Economic Review,* 121, no. 3, 823-856.

Barberis, Nicholas, 2018. "Psychology-Based Models of Asset Prices and Trading Volumes," forthcoming in D. Bernheim, S DellaVigna and D. Laibson, editors, *Handbook of Behavioral Economics.* North Holland, Amsterdam.

Barberis, Nicholas, Ming Huang and Tano Santos, 2001. "Prospect Theory and Asset Prices," *Quarterly Journal of Finance,* 116, no. 1, 1-53.

Bekaert, Geert, Marie Hoerova and Marco Lo Duca, 2013. "Risk, Uncertainty and Monetary Policy," *Journal of Monetary Economics,* 60, 771-778.

Berger, David, Ian Dew-Becker, and Stefano Giglio, 2019. "Uncertainty shocks as second-moment news shocks," forthcoming in the *Review of Economic Studies.*

Bialkowski, J., K. Gottschalk and T. Wisniewski, 2008. "Stock Market Volatility Around National Elections," *Journal of Banking and Finance*, 32, 1941-53.

Bittlingmayer, George, 1998. "Output, Stock Volatility, and Political Uncertainty in a Natural Experiment: Germany, 1880-1940," *Journal of Finance,* 53, no. 6 (December), 2243-2257.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, Matthew Richardson, Information, Trading, and Volatility: Evidence from Firm-Specific News, The Review of Financial Studies, Volume 32, Issue 3, March 2019, Pages 992–1033Boutchkova, Maria, Hitesh Doshi, Art Durnev and Alexander Molchanov, 2012. "Precarious Politics and Return Volatility," *Review of Financial Studies,* 25, no. 4 (April), 1111-1154.

Boyd, John H., Jian Hu and Ravi Jagannathan, 2005. "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks," *Journal of Finance,* 60, no. 2 (April), 649-672.

Brown, Chad P. and Melina Kolb, 2019. "Trump's Trade War Timeline: An Up-to-Date Guide," https://piie.com/blogs/trade-investment-policy-watch/trump-trade-war-china-date-guide.

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu and Logan B. Steele, 2014. "The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings," *Review of Accounting Studies,* 19, 396-455.

Campbell, John Y. and Robert J. Shiller, 1987. "Cointegration and Tests of Present-Value Models," *Journal of Political Economy,* 95, 1062-1088.

Campbell, John Y. and Robert J. Shiller, 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies,* 1, no. 3 (July), 195-228.

Chen, Nai-Fu, Richard Roll and Stephen A. Ross, 1986. "Economic Forces and the Stock Market," *Journal of Business,* 59, no. 3 (July), 383-403.

Christensen, B.J. and N.R. Prabhala, 1998. "The relation between implied and realized volatility," *Journal of Financial Economics*, 50, 2, 125-150.

Cochrane, John H., 1991. "Volatility Tests and Efficient Markets: A Review Essay," *Journal of Monetary Economics,* 27, 463-485.

Cochrane, John H., 1992. "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies,* 5, no. 2 (April), 243-280.

Cochrane, John H., 2017. "Macro-Finance," *Review of Finance,* 21, no. 3 (May), 945-985.

Cutler, David M., James M. Poterba, and Lawrence H. Summers, 1989. "What moves stock prices?" *Journal of Portfolio Management*, 15, no. 3 (Spring), 4-12.

Davis, Steven J., 2017. "Regulatory Complexity and Policy Uncertainty: Headwinds of Our Own Making," paper prepared for the Hoover Institution Conference on Restoring Prosperity: Contemporary and Historical Perspectives.

Davis, Steven J. and Cristhian Seminario, 2019. "Diagnosing the Stock Market Reaction to Trump's Surprise Election Victory," University of Chicago, in progress.

Drechsler, Itamar and Amir Yaron, 2011. "What's Vol Got to Do with It?" *The Review of Financial Studies*, 24, 1, 1–45.

Fama, Eugene F., 1981. "Stock Returns, Real Activity, Inflation, and Money," *American Economic Review,* 71, no. 4 (September), 545-565.

Fama, Eugene F. and Kenneth R. French, 1988. "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy,* 96, 246-273.

Fama, Eugene F. and Kenneth R. French, 1989. "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics,* 25, 23-49.

Forbes, Kristin J. and Francis E. Warnock, 2012. "Capital Flow Waves: Surges, Stops, Flight and Retrenchment," *Journal of International Economics,* 88, no. 2, 235-251.

Gabaix, Xavier, 2012. "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics,* 127, no. 2, 645-700.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy, 2018. "Text as data," *Journal of Economic Literature* (Forthcoming).

Gerlach, Stefan, Srichander Ramaswamy and Michela Staigna, 2006. "150 Years of Financial Market Volatility," *BIS Quarterly Review* (September), 77-91.

Gorodnichenko, Yuriy and Michael Weber, 2016, "Are Sticky Prices Costly? Evidence from the Stock Market," *Economic Review*, 106, 1, 165-199.

Gourio, Francois, 2008. "Time Series Predictability in the Disaster Model," *Finance Research Letters,* 5, no. 4, 191-203.

Grennan, Jillian, 2019. "A Corporate Culture Channel: How Increased Shareholder Governance Reduces Firm Value," working paper, Duke University.

Griffin, J. M., N. H. Hirschey, and P. J. Kelly. 2011. How important is the financial media in global markets? Review of Financial Studies 24:3941–92

Hassan, Tarek A., Stephan Hollander, Laurence van Lent and Ahmed Tahoun, 2019. "Firm-Level Political Risk: Measurement and Effects," working paper.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman, 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* second edition. Springer.

Jiang, Liangliang, Jeffrey A. Pittman and Walid Saffar, 2019. "Policy Uncertainty and Textual Disclosure," working paper at http://dx.doi.org/10.2139/ssrn.3015420.

Jurado, Kyle, Sydney Ludvigson, and Serena Ng, 2015. "Measuring Uncertainty," *American Economic Review*, 105 (3), 1177–1216.

Kelly, Bryan, Asaf Manela, and Alan Moreira, 2018. "Text Selection," working paper, October.

Killian, Lutz and Cheolbeom Park, 2009. "The Impact of Oil Price Shocks on the U.S. Stock Market," *International Economic Review,* 50, no. 4 (November), 1267-1287.

Londono, Juan M. and Beth Anne Wilson, 2018. "Understanding Global Volatility," *IFDP Notes*, January 2018.

Manela, Asaf and Alan Moreira, 2017. "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123, no.1, 137-162.

McCarty, Nolan, Keith T. Poole and Howard Rosenthal, 2016. *Polarized America: The Dance of Ideology and Unequal Riches,* second edition. Cambridge, MA: MIT Press.

Nagel, Stefan, 2012. "Evaporating Liquidity," *Review of Financial Studies*, 25, 7, 2005-2039.

Niederhoffer, Victor, 1971. "The Analysis of World Events and Stock Prices," *Journal of Business,* 44 (April), 193-319.

Pastor, Lubos and Pietro Veronesi, 2012. "Uncertainty about Government Policy and Stock Prices," *Journal of Finance*, 64, 4, 1219-1264.

Pastor, Lubos and Pietro Veronesi, 2013. "Political Uncertainty and Risk Premia," *Journal of Financial Economics*, 110, 3, 520-545.

Poterba, James and Lawrence Summers, 1988. "Mean Reversion in Stock Prices," *Journal of Financial Economics,* 22, 27-59.

Rey, Helene, 2018. "Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence," NBER Working Paper 21162, revised.

Rietz, Thomas A., 1988. "The Equity Risk Premium: A Solution," *Journal of Monetary Economics,* 22, no. 1, 117-131.

Schwert, G. William, 1989. "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance,* 44, no. 5 (December), 1115-1153.

Shiller, Robert J., 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review,* 71, no. 3 (June), 421-436.

Shiller, Robert J., 1984. "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity,* 1984:2, 457-510.

Shiller, Robert J., 2014. "Speculative Asset Prices," *American Economic Review,* 104, no. 6 (June), 1486-1517.

Summers, Lawrence H., 1986. "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance,* 41, no. 3 (July), 591-601.

Wachter, Jessica A., 2013. "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *Journal of Finance,* 68, no. 3, 987-1035.

Table 1: Percent of EMV Articles in Each Category, 1985-2018

Panel A. General Economic Categories	**Percent of EMV Articles**	
Macroeconomic News and Outlook	71.8	
Broad Quantity Indicators	26.8	
Inflation	28.7	
Interest Rates	30.7	
Other Financial Indicators	3.5	
Labor Markets	22.9	
Real Estate Markets	30.7	
Trade	2.4	
Business Investment and Sentiment	1.8	
Consumer Spending and Sentiment	9.2	
Commodity Markets	43.7	
Financial Crises	8.1	
Exchange Rates	2.0	
Healthcare Matters	6.4	
Litigation Matters	4.7	
Competition Matters	3.8	
Labor Disputes	4.0	
Intellectual Property Matters	3.3	
Panel B. Policy-Related Categories	**Percent of EMV Articles**	**Percent of EPU Articles**
Fiscal Policy:	34.7	44.6
Taxes	29.8	36.1
Government Spending, Deficits, and Debt	6.1	15.3
Entitlement and Welfare Programs	7.1	12.0
Monetary Policy	29.5	34.9
Regulation (generic + 4 big regulation categories)	24.9	27.1
Financial Regulation	14.7	6.3
Competition Policy	2.4	1.1
Intellectual Property Policy	0.1	0.3
Labor Regulations	2.0	3.3
Immigration	0.2	1.5
Energy and Environmental Regulation	1.3	5.5
Lawsuit and Tort Reform, Supreme Court	1.4	4.2
Housing and Land Management	1.2	1.5
Other Regulation: Education, Communication, Consumer Safety, etc	1.0	1.7
National Security Policy	13.1	28.6
Government-Sponsored Enterprises (e.g., Fannie Mae)	4.9	2.7
Trade Policy	2.8	6.0
Healthcare Policy	3.6	8.5
Food and Drug Policy	1.3	1.0
Transportation, Infrastructure, and Public Utilities	1.3	2.6
Elections and Political Governance	3.0	8.2
Agricultural Policy	0.2	0.6

Notes to Table 1: The second column reports the share of EMV articles with one or more terms in the indicated category-specific term set. See Appendix B for the term sets. The rightmost column in Panel B reports the share of EPU articles that contain one or more terms in the category-specific set, where EPU articles are those that meet the criteria of Baker, Bloom and Davis (2016) for policy-related economic uncertainty.

Table 2: Regressions of Stock Market Volatility Measures on the EMV Tracker

	(1) VIX_t	(2) VIX_t	(3) VIX_t	(4) VIX_t	(5) $Log(VIX_t)$	(6) $RVol_t$	(7) $RVol_t$
EMV_t	0.75 (0.06)	0.53 (0.10)	0.43 (0.06)	0.47 (0.08)		0.95 (0.09)	0.79 (0.12)
EMV_{t-1}		0.20 (0.10)		-0.11 (0.11)			
EMV_{t-2}		0.20 (0.06)		-0.02 (0.08)			
$Log(EMV_t)$					0.78 (0.04)		
VIX_{t-1}			0.58 (0.07)	0.66 (0.08)			
$RVol_{t-1}$							0.24 (0.15)
R^2	0.61	0.70	0.83	0.84	0.59	0.65	0.89
Obs.	408	406	407	406	408	408	407

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 1985 to December 2018. The sample for Columns (2) and (4) starts in March 1985. EMV is Equity Market Volatility tracker developed in Section 2.1. VIX is the monthly average of daily closing values on the CBOE 30-day implied volatility index from January 1990 onwards, appended to data from Berger et al. (2019) in earlier years. RVol is the standard deviation of daily returns on the S&P500 in the month. Robust standard errors in parentheses.

Table 3: Regressions of VIX Term Lengths on the EMV Tracker – All Terms

	(1) VIX_t	(2) VIX_t	(3) VIX_t	(4) VIX_t	(5) VIX_t	(6) VIX_t	(7) VIX_t
Term Structure	1 Month	3 Month	6 Month	1 Year	3 Year	5 Year	10 Year
EMV_t	0.45	0.30	0.21	0.14	0.08	0.06	0.04
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)
EMV 3 Lag Avg.	0.61	0.57	0.48	0.38	0.26	0.21	0.11
	(0.11)	(0.09)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)
EMV 12 Lag Avg.	0.39	0.50	0.58	0.61	0.58	0.53	0.38
	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)	(0.11)
R^2	0.86	0.86	0.83	0.78	0.67	0.60	0.38
Obs.	151	151	151	151	151	151	151

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 2004 to July 2016. EMV is Equity Market Volatility tracker developed in Section 2.1. VIX is the monthly average of the different term length VIX measures. Newey-West standard errors with maximum autocorrelation lag of 2 in parentheses.

Table 3b: Regressions of VIX Term Lengths on Categorical EMV Trackers

	(1) VIX_t	(2) VIX_t	(3) VIX_t
Term Structure	1 Month	1 Year	10 Year
Regulation - Financial	0.268***	0.243***	0.605***
	(0.0808)	(0.0788)	(0.149)
Macro - Broad Quantity Indicators	0.224**	0.161***	0.216***
	(0.0879)	(0.0569)	(0.0533)
National Security	0.187***		
	(0.0595)		
Macro - Other Financial Indicators	0.239		
	(0.313)		
Macro - Consumer Spending	0.287		
	(0.19)		
Macro - Real Estate	0.140**		
	(0.0662)		
Macroeconomic News and Outlook	-0.199***		
	(0.0647)		
Macro - Labor Markets			0.149
			(0.101)
Regulation - Competition			-0.0743
			(0.416)
Regulation - All			-0.459***
			(0.168)
Financial Crises			0.0655
			(0.123)
R^2	0.518	0.428	0.410
Obs.	163	163	163

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 2004 to December 2018. Each row variable is a categorical EMV tracker as defined in Section 2. VIX is the monthly average of the different term length VIX measures. Newey-West standard errors with maximum autocorrelation lag of 2 in parentheses.

Table 4: Predicting Stock Market Returns Using Equity Market Volatility

	(1) $r(t \rightarrow t+\tau)$	(2) $r(t \rightarrow t+\tau)$	(3) $r(t \rightarrow t+\tau)$	(4) $r(t \rightarrow t+\tau)$	(5) $r(t \rightarrow t+\tau)$
Returns Horizon	1 Months	3 Months	6 Months	1 Year	2 Years
EMV_{t-1}	0.052	0.035	0.022	0.016	0.026**
	(0.063)	(0.042)	(0.028)	(0.017)	(0.012)
R^2	0.0024	0.0026	0.0018	0.0017	0.0071
Obs.	396	396	396	396	396
	0.174	0.129*	0.104**	0.030	0.059**
EMV_{t-1} – Macroeconomic News	(0.115)	(0.076)	(0.052)	(0.043)	(0.026)
R^2	0.0063	0.0079	0.0091	0.0014	0.0084
Obs.	396	396	396	396	396
EMV_{t-1} – Regulation	-0.030	0.015	0.099	0.199**	0.189***
	(0.297)	(0.185)	(0.139)	(0.083)	(0.064)
R^2	0.0001	0.0001	0.0013	0.0097	0.0140
Obs.	396	396	396	396	396
	0.359*	0.243**	0.128	0.099	0.080
EMV_{t-1} – National Security	(0.189)	(0.123)	(0.090)	(0.097)	(0.052)
R^2	0.0081	0.0086	0.0042	0.0045	0.0047
Obs.	396	396	396	396	396

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 1985 to December 2018. EMV is Equity Market Volatility tracker developed in Section 2.1. Returns is the total return of the S&P 500 index during the listed horizon. Each cell reports a results from a separate regression. Newey-West standard errors with maximum autocorrelation lag of the horizon length in parentheses.

Table 5: Summary Statistics for the VIX, Realized Volatility, EMV and NVIX

A. January 1985 to March 2016

	RVol	VIX	EMV	NVIX
Standard Deviation	9.57	7.81	8.14	4.83
Skewness	3.67	2.19	2.40	1.27
Kurtosis	24.34	10.76	11.41	7.43
Pairwise Correlation with VIX			0.78	0.70
Pairwise Correlation with VIX in 1st Differences			0.58	0.48
Mean Absolute Distance from VIX			3.69	4.03
Pairwise Correlation with RVol			0.80	0.65
Pairwise Correlation with RVol in 1st Differences			0.66	0.49
Mean Absolute Distance from RVol			5.51	6.41

B. January 1960 to December 1984

	RVol	EMV	NVIX
Standard Deviation	5.18	5.52	1.38
Skewness	1.42	0.84	0.74
Kurtosis	5.88	3.45	3.90
Pairwise Correlation with RVol		0.42	-0.02
Pairwise Correlation with RVol in 1st Differences		0.37	0.17
Mean Absolute Distance from RVol		3.08	3.10

C. January 1928 to December 1959

	RVol	EMV	NVIX
Standard Deviation	13.80	9.22	2.35
Skewness	1.93	0.81	0.15
Kurtosis	6.75	3.78	3.09
Pairwise Correlation with RVol		0.53	0.56
Pairwise Correlation with RVol in 1st Differences		0.14	0.16
Mean Absolute Distance from RVol		7.34	6.19

Notes: The NVIX measure developed by Manela and Moreira (2017) runs through March 2016 and is downloadable at http://apps.olin.wustl.edu/faculty/manela/data.html. The VIX, RVol and EMV measures are as defined in Table 2. We multiplicatively scale the NVIX and EMV measures to match the mean VIX value from 1985 to 2015 in Panel A. We scale NVIX and EMV to match the mean RVol value from January 1960 to December 1984 in Panel B and from January 1928 to 1959 in Panel C. As discussed in the text, the EMV tracker in Panels B and C relies on six newspapers, whereas the version in Panel A relies on eleven papers.

Table 6: Total Number of 10-K Part 1A Sentences Summary Statistics

	# Sentences				
Filing Year	Mean	Median	SD	Min	Max
2006	220	158	221	9	3633
2007	243	176	245	9	4765
2008	255	192	240	9	4575
2009	263	205	217	9	2793
2010	267	211	209	9	2788
2011	265	212	201	9	2741
2012	279	219	216	9	2743
2013	287	233	225	9	3608
2014	308	247	231	9	2490
2015	327	263	238	9	2181
2016	340	274	245	11	2025
2017	367	297	255	9	2147
2018	383	312	257	12	2197
2019	418	345	280	9	2588

Notes: From each 10-K filing, we use automated methods to count the number of sentences in each Part 1A section. We drop filings for which the automated sentence counter returns a value of less than nine for the part 1A section. This cutoff seems to be the appropriate one based off visual inspection of 10-K filings where sections that are less than 9 sentences typically represent routine headings and section separators in 10-K filings with an empty Part 1A. When the same firm filed multiple 10-K files on the same date, we retain the one with the longer Part 1A. When a firm has more than one 10-K filing in the same calendar year, we retime the “early” (“late”) filing to the prior (next) calendar year provided the firm has no filing in the prior (next) calendar year. If a firm still has multiple 10-K filings in the same calendar year, we retain the file with the longer Part 1A.

Table 7: Realized Volatility Summary Statistics

	Realized Volatility					Realized Volatility – Weighted by Previous Month Market Capitalization				
Filing Year	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max
2006	34.4	29.3	24.2	0	1026.8	21.8	18.6	12.3	0	615.7
2007	37.0	32.1	23.3	0	542.4	25.5	22.4	13.6	0	542.4
2008	68.9	55.7	47.5	0	1133.3	49.3	38.6	35.0	0	1133.3
2009	65.9	53.9	49.9	0	1863.3	37.4	30.9	25.5	0	1863.3
2010	43.2	36.8	32.8	0	1412.0	26.1	23.3	13.9	0	1412.0
2011	46.8	38.7	38.6	0	1821.7	29.6	24.9	17.1	0	1821.7
2012	39.2	30.9	35.5	0	1180.8	22.2	19.5	12.3	0	1075.9
2013	35.1	26.9	32.7	0	1946.6	21.0	18.6	11.1	0	1946.6
2014	35.3	27.2	40.5	0	5288.2	21.2	18.1	12.3	0	5288.2
2015	39.6	30.3	100.1	0	13666.5	24.9	21.9	26.5	0	13666.5
2016	41.7	32.7	35.8	0	1523.0	24.4	20.5	15.1	0	1523.0
2017	36.0	27.4	40.9	0	5163.4	18.9	16.3	11.8	0	1152.6
2018	40.7	32.6	34.1	0	1146.8	27.0	24.1	13.7	0	1146.8
2019	40.8	30.9	36.8	0	889.7	23.4	20.3	13.1	0	889.7

Notes: Firm-level measures are computed at a firm-month level by pooling all monthly firm-level observations within a given year. 508,420 firm-month observations. Average number of firms per year is 36,317.

Table 8: Realized Volatility on 10-K EMV Topics Composite Regression

	(1) *Realized Volatility*$_{i,t}$	(2) *Realized Volatility*$_{i,t}$	(3) *Realized Volatility*$_{i,t}$	(4) *Realized Volatility*$_{i,t}$	(5) *Realized Volatility*$_{i,t}$
EMV Topics Composite$_{i,t}$	2.16*** (0.22)				
EMV Non-Policy Topics$_{i,t}$		2.50*** (0.25)		2.46*** (0.25)	
EMV Policy Topics$_{i,t}$			1.35*** (0.48)	0.83* (0.49)	
EMV Topics Macro-Interest Rates$_{i,t}$					-9.33*** (1.01)
EMV Topics Macro – Real Estate Markets$_{i,t}$					7.90*** (0.81)
EMV Topics Commodity Markets$_{i,t}$					2.50*** (0.29)
R^2	0.546	0.546	0.545	0.546	0.547
R^2 - Within	0.0015	0.0016	0.0001	0.0016	0.0039
Observations	508,447	508,447	508,447	508,447	508,447

Notes: The sample period for the regressions is 2006-2019. The EMV Topics Composite variable is constructed as discussed in the text as $\sum F_{i,y}^{b} EMV_{t}^{b}$. The policy and non-policy composite variables for columns (2)-(4) are constructed the same as the overall composite variable but only using the policy or non-policy categories respectively. Column (5) reports the OLS coefficients of the LASSO selected variables when considering a LASSO specification using all of the separate category EMV indices. Realized volatility is winsorized at the 1% and 99% levels. All regressions include firm and time fixed effects. All regressions are weighted by the product of the square root of the number of sentences in the 10-K Section 1A and the lagged value of log market capitalization. Standard errors clustered at the firm-level are reported in parentheses. $p < 0.01$ ***, $p < 0.05$ **, $p < 0.10$ *

Table 9: Realized Volatility on EMV Topics Composite Regression – Firm Characteristics

	(1) *Realized Volatility*$_{i,t}$	(2) *Realized Volatility*$_{i,t}$	(3) *Realized Volatility*$_{i,t}$	(4) *Realized Volatility*$_{i,t}$	(5) *Realized Volatility*$_{i,t}$	(6) *Realized Volatility*$_{i,t}$	(7) *Realized Volatility*$_{i,t}$	(8) *Realized Volatility*$_{i,t}$
EMV Topics Composite$_{i,t}$	1.63*** (0.27)							
EMV Non-Policy Topics$_{i,t}$		1.75*** (0.30)		1.73*** (0.30)				
EMV Policy Topics$_{i,t}$			1.04 (0.84)	0.86 (0.84)				
EMV Topics Macro-Interest Rates$_{i,t}$					-5.13*** (1.18)			
EMV Topics Commodity Markets$_{i,t}$					2.08*** (0.34)			
# of vars	1	1	1	2	38	125	188	2,456
# LASSO-selected vars	-	-	-	-	2	21	28	46
# LASSO EMV vars	-	-	-	-	2	7	14	35
# LASSO EMV Policy	-	-	-	-	-	-	6	9
R^2	0.593	0.593	0.593	0.593	0.597	0.650	0.650	0.652
R^2 - Within	0.0010	0.0010	0.00005	0.0010	0.0013	0.141	0.142	0.146
Observations	214,943	214,943	214,943	214,943	214,943	214,943	214,943	214,943

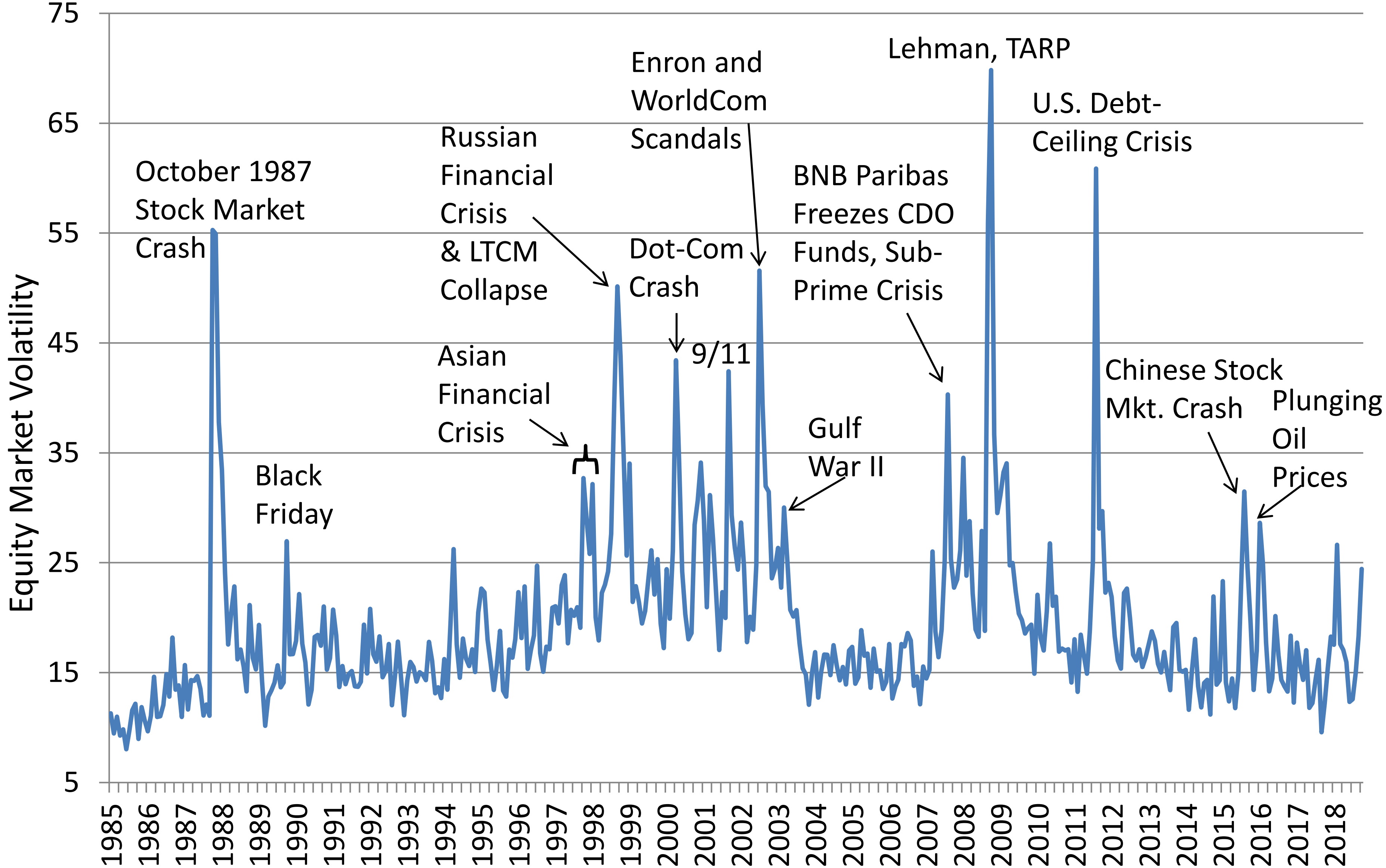
Notes: The sample period for the regressions is 2006-2014. Sample is limited to firms that have data on all 62 firm financial characteristics considered for the analysis. This list of firm characteristics and the corresponding data come from Freyberger et al. (2019). The EMV Topics Composite variable is: $\sum F_{i,y}^{b} EMV_{t}^{b}$. The policy and non-policy composite variables for columns (2)-(4) mirror the overall composite variable but using the policy or non-policy categories. Column (5) reports the OLS coefficients of the LASSO selected variables when considering a LASSO specification using all separate category EMV indices. Realized volatility is winsorized at 1% and 99%. All regressions include firm and time fixed effects. All regressions are weighted by the product of the square root of the number of sentences in the 10-K Section 1A and the lagged value of log market cap. Standard errors clustered at the firm-level are reported in parentheses. p < 0.01 ***, p < 0.05 **, p < 0.10 *

Table 10: Firm-Level Time Series Correlations by Firm Characteristic

	(1) *Quintile 1*	(2) *Quintile 2*	(3) *Quintile 3*	(4) *Quintile 4*	(5) *Quintile 5*
Assets-to-Market Cap	0.29	0.29	0.30	0.35	0.40
Total Assets	0.23	0.26	0.33	0.36	0.46
Book to Market Value Ratio	0.31	0.30	0.32	0.34	0.37
Log Change in Split Adj Shares Outstanding	0.37	0.33	0.32	0.31	0.31
Debt to Price Ratio	0.24	0.30	0.32	0.36	0.41
Tobin's Q	0.37	0.36	0.31	0.30	0.30

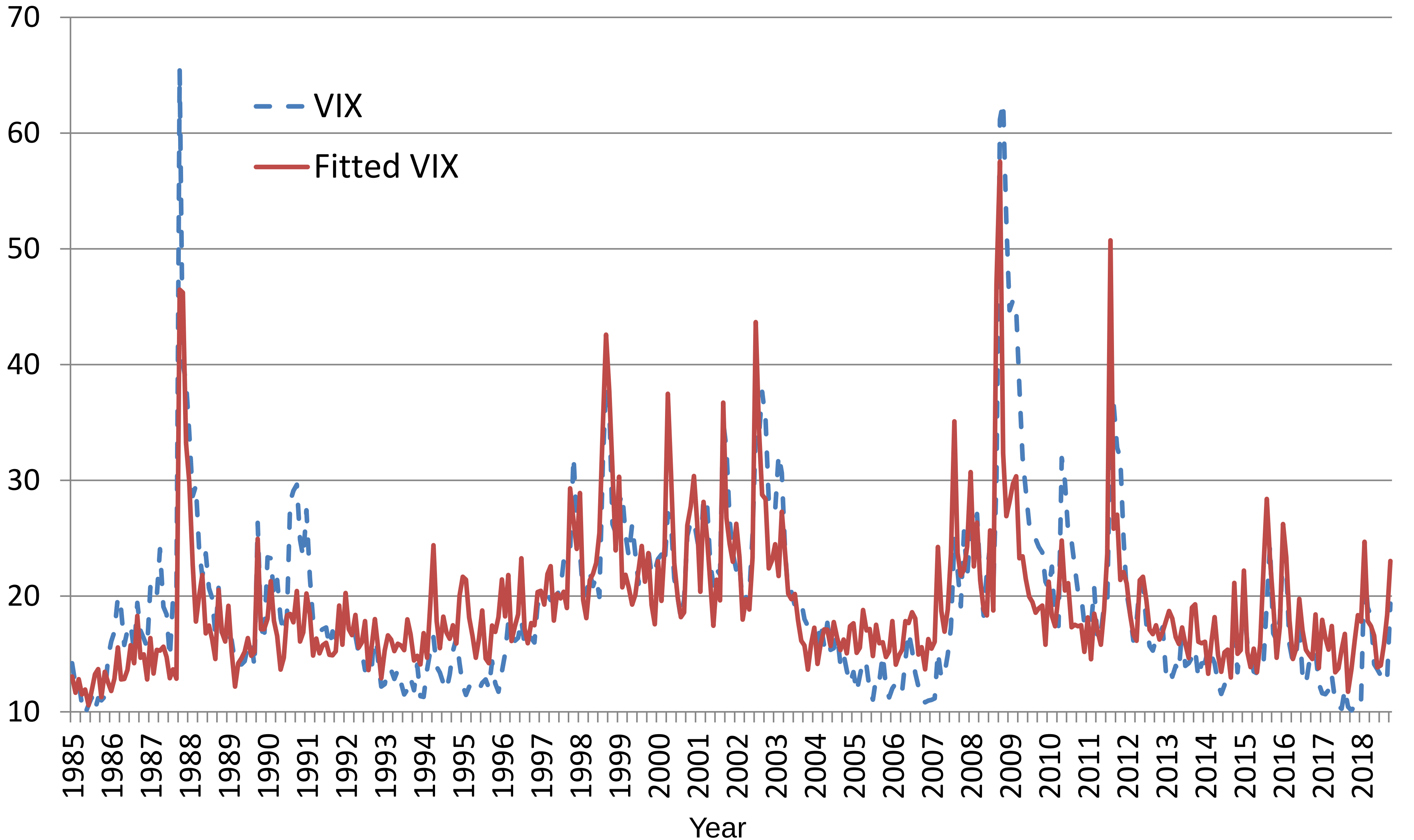
Notes: Each column reports the average firm-level time series correlation between actual and fitted realized volatility for the firms in the firm characteristic quintile reported by the row and column headings. Each row considers a different firm characteristic and each column splits the firms in the sample into quintiles based off that given firm characteristic. Firms with less than 12 monthly observations were dropped.

Figure 1: Newspaper-Based Equity Market Volatility Tracker, 1985-2018



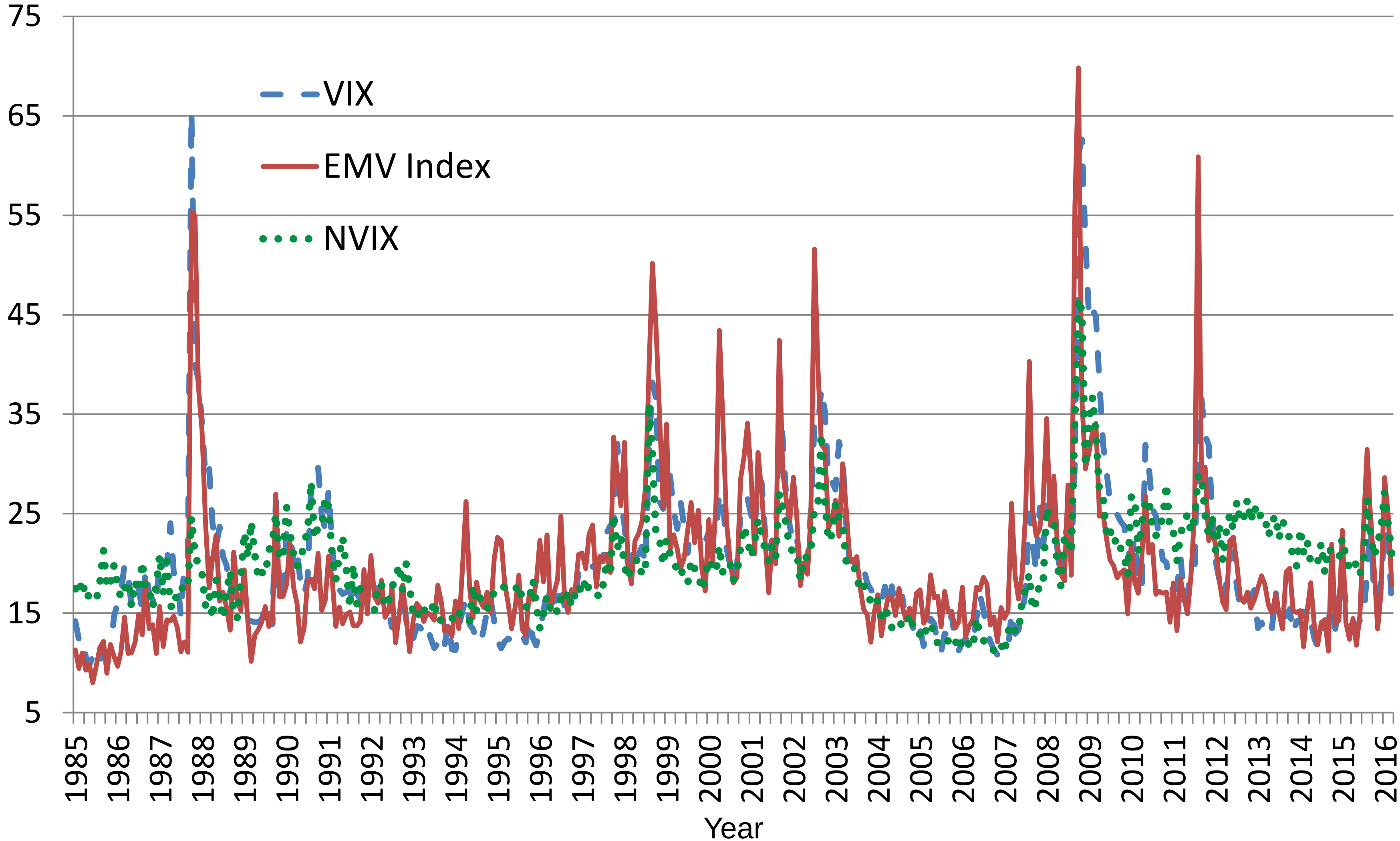
Notes: The Equity Market volatility (EMV) tracker runs from January 1985 to October 2018. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in 11 leading U.S. newspapers, as detailed in Section 2.1. We scale the EMV tracker to match the mean value of the VIX from 1985 to 2015.

Figure 2: VIX and Fitted VIX from a Regression on EMV, 1985-2018



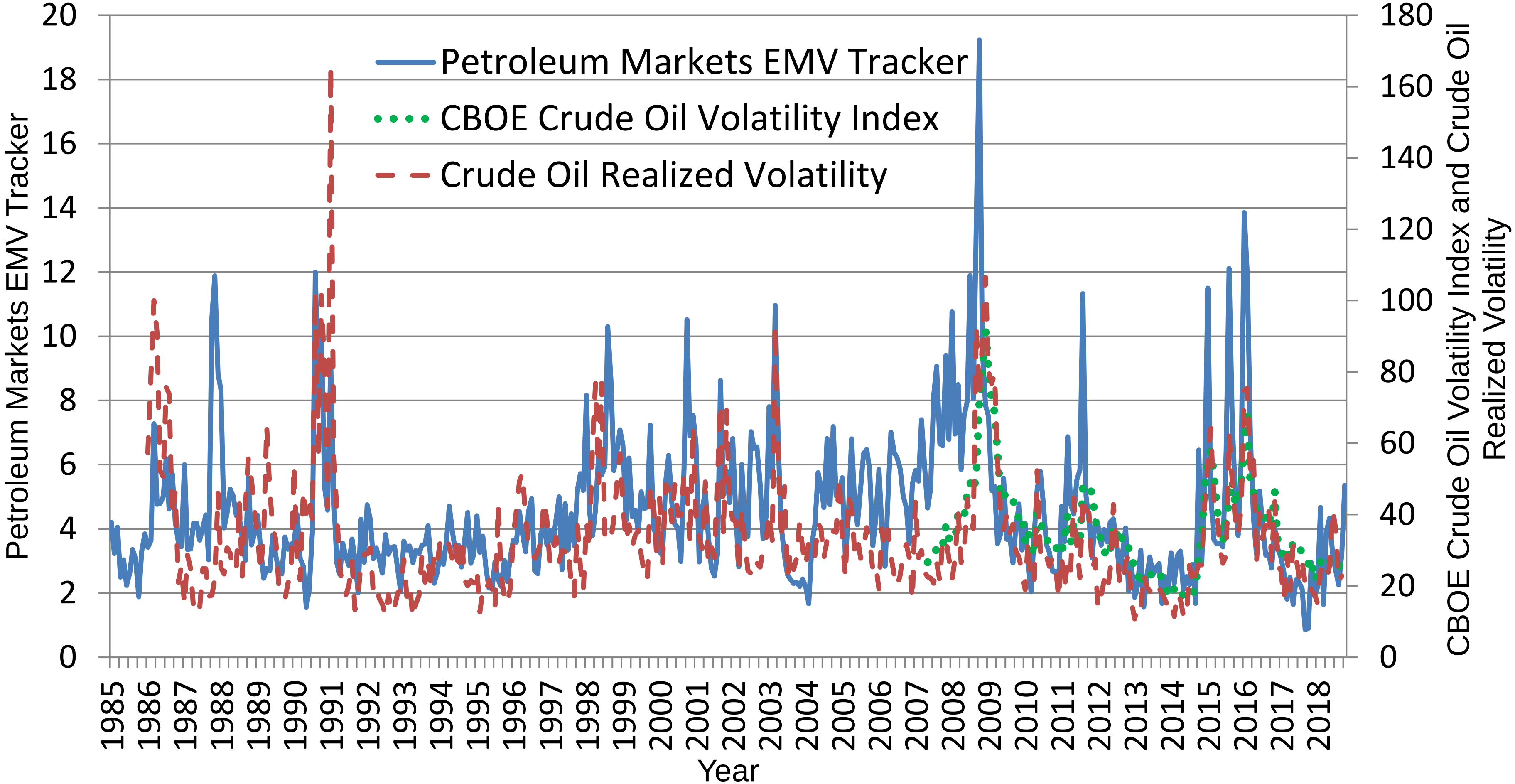
Notes: Data for the CBOE 30-Day VIX data from 1990 to 2017 appended to the VIX series in Berger et al. (2019) from 1985 to 1989. "Fitted VIX" values are from the regression VIX on EMV reported in Table 2, column (1) Both series run from January 1985 to October 2018.

Figure 3: VIX, EMV and NVIX, January 1985 to March 2016



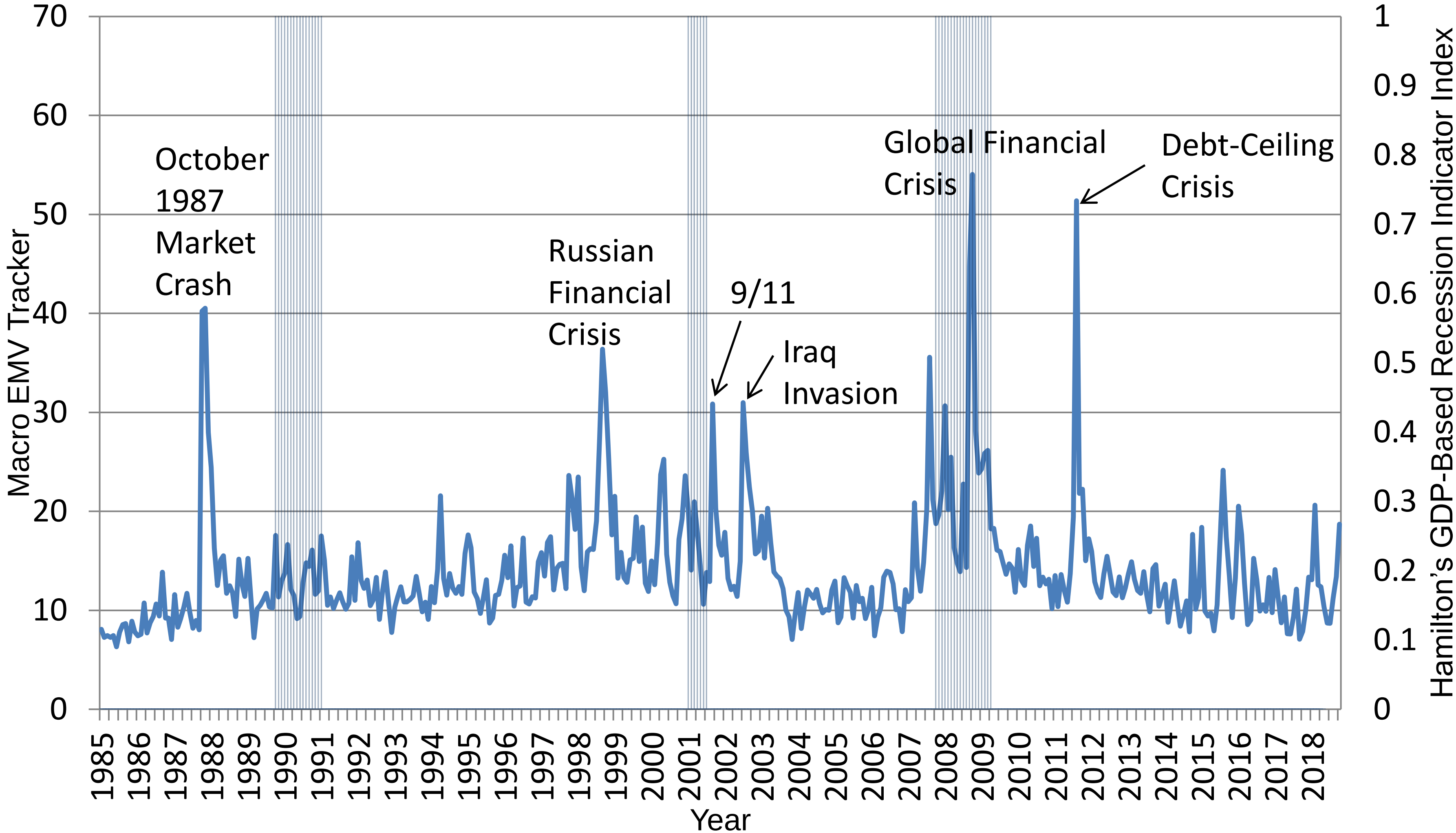
Notes: The NVIX measure is from Manela and Moreira (2017) and runs through March 2016. See the notes to Figure 2 for the VIX and NVIX. We multiplicatively scale NVIX and EMV to match the mean value of the VIX from 1985 to 2015.

Figure 4: Petroleum Markets EMV Compared to Oil Price Volatility, Monthly, 1985 to 2018



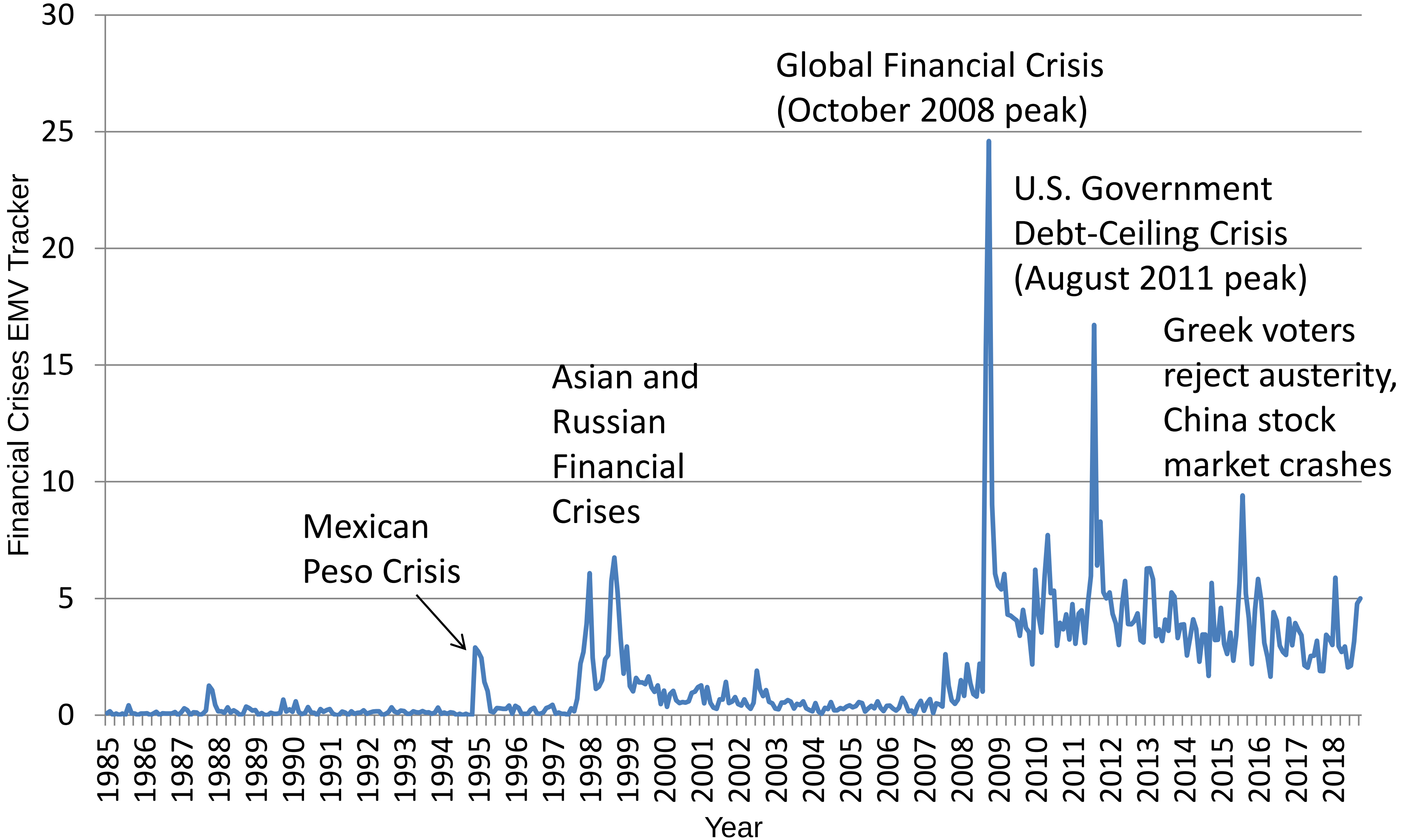
Notes: CBOE Crude Oil Volatility Index is the monthly mean of daily CBOE Crude Oil ETF Volatility Index values. Crude Oil Realized Volatility reflects daily price data for West Texas Intermediate. We extract both series from the St. Louis Federal Reserve FRED database. The Petroleum Markets EMV tracker is constructed from scaled frequency counts of newspaper articles. See Sections 2.1 and 3.4 in the text for details.

Figure 5: Macroeconomics EMV Tracker



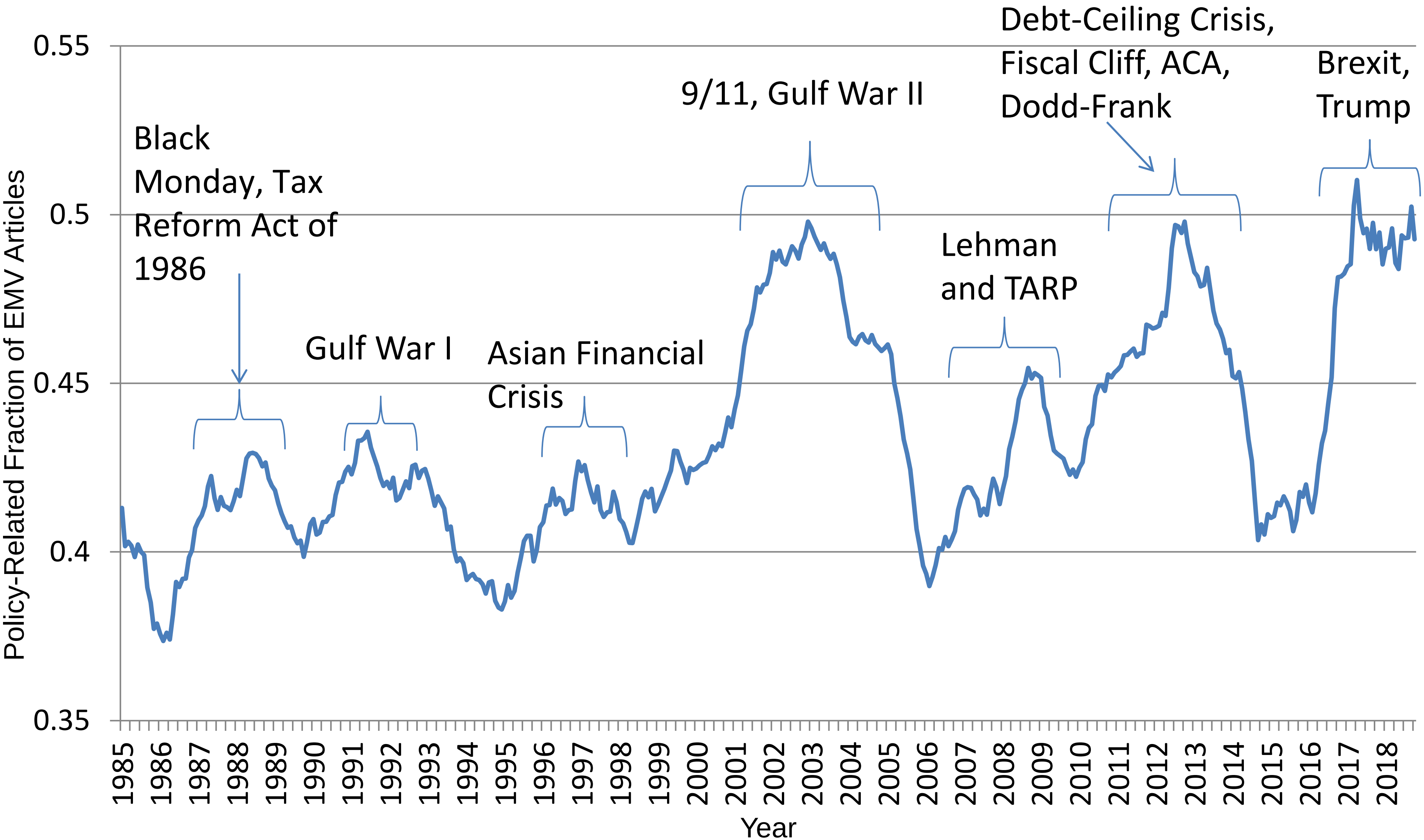
Notes: We construct the Macroeconomics EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Macroeconomic News and Outlook.** See Appendix B for the list of terms.

Figure 6: Financial Crisis EMV Tracker, 1985-2018



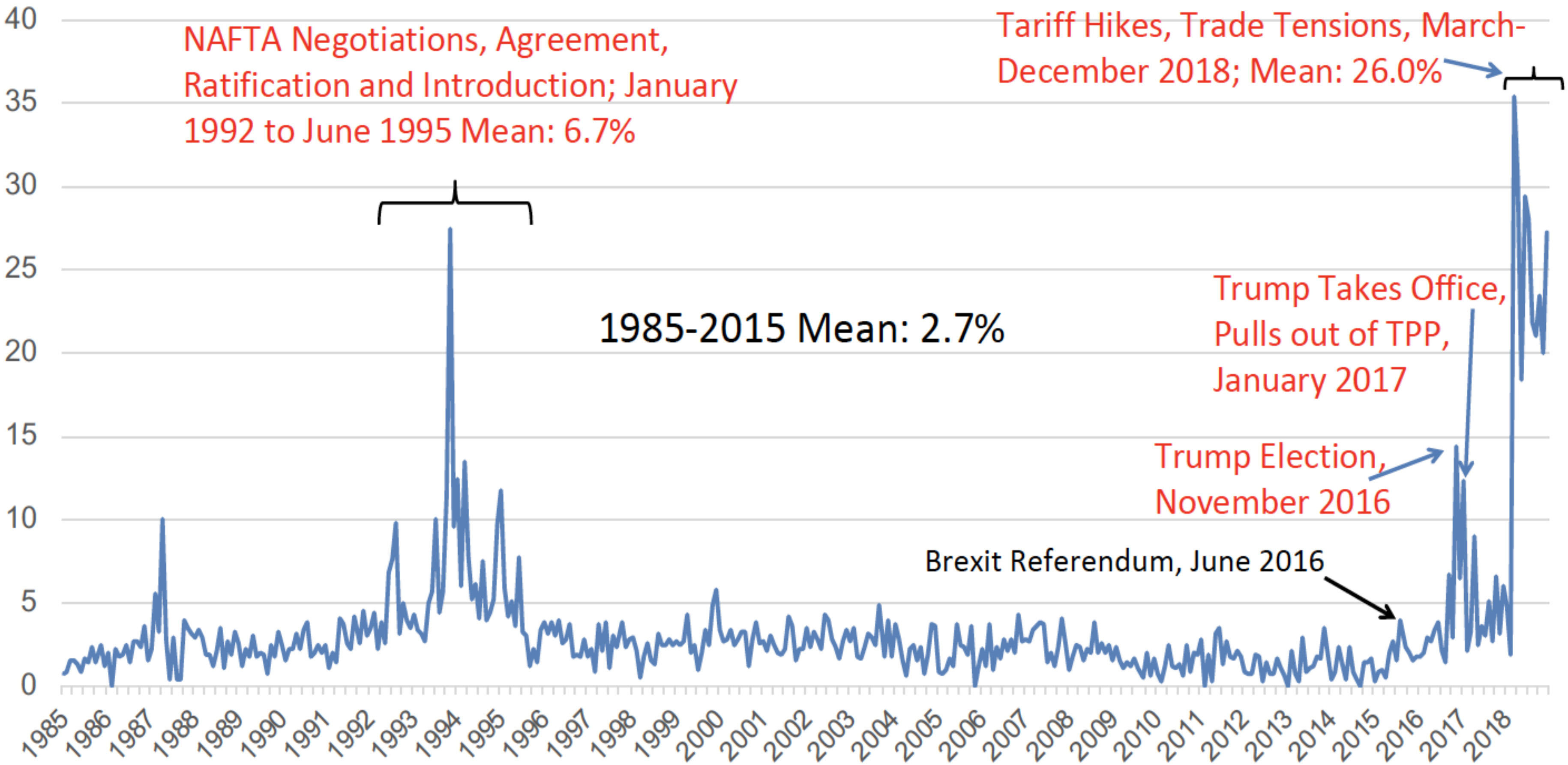
Notes: We construct the Financial Crisis EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Financial Crises.** See Appendix B for the list of terms.

Figure 7: Fraction of EMV Articles that Discuss Policy Matters, 12-Month Moving Average, 1985-2018.



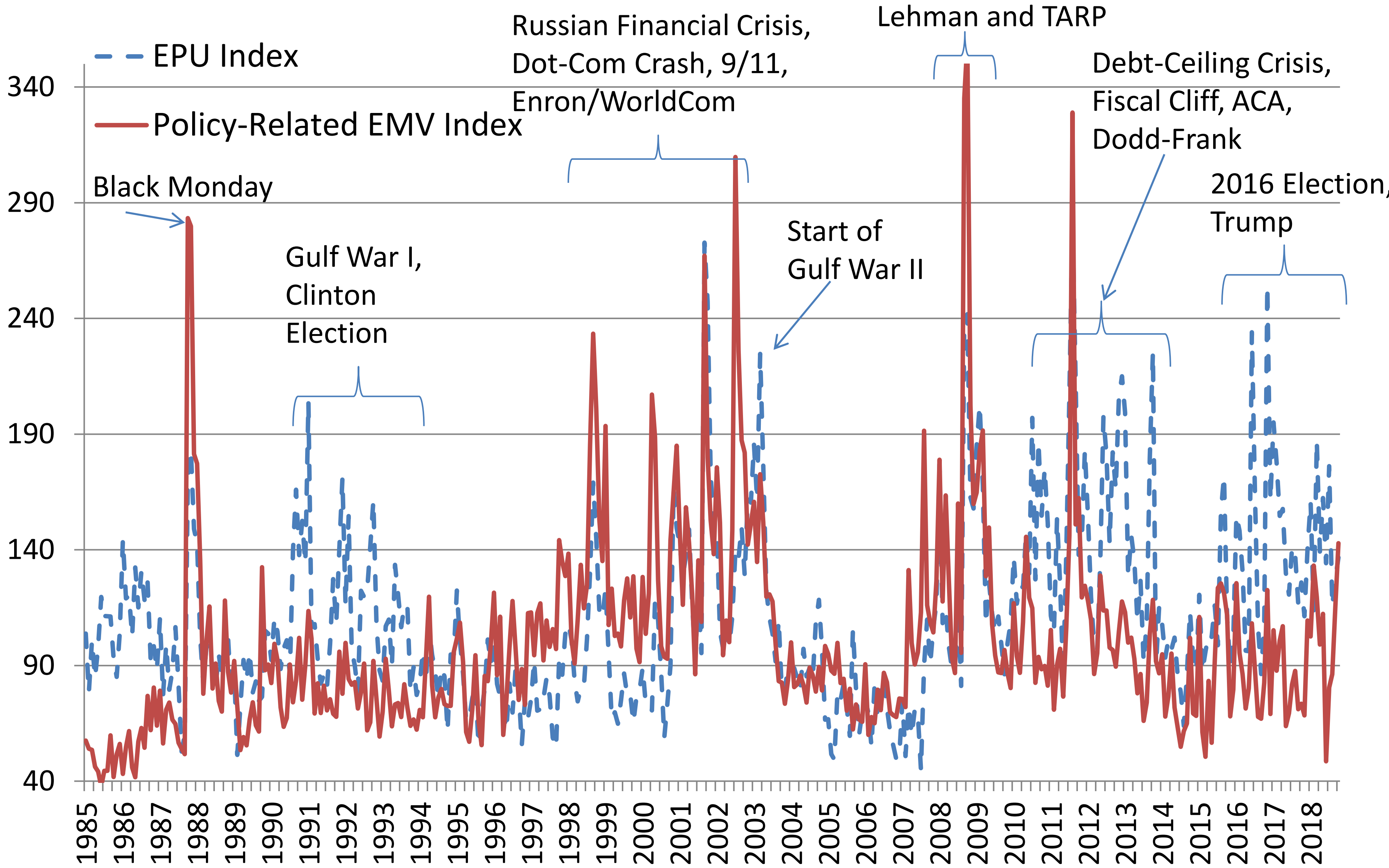
Notes: We sum EMV article counts over policy-related categories and divide by the sum of EMV article counts over all categories (general and policy-related). We compute this ratio for each newspaper and month, average over papers by month and then compute a moving average with six lags and leads, truncating lags (leads) near the sample start (end).

Figure 8: Percent of EMV Articles that Discuss Trade Policy Matters, January 1985 to December 2018



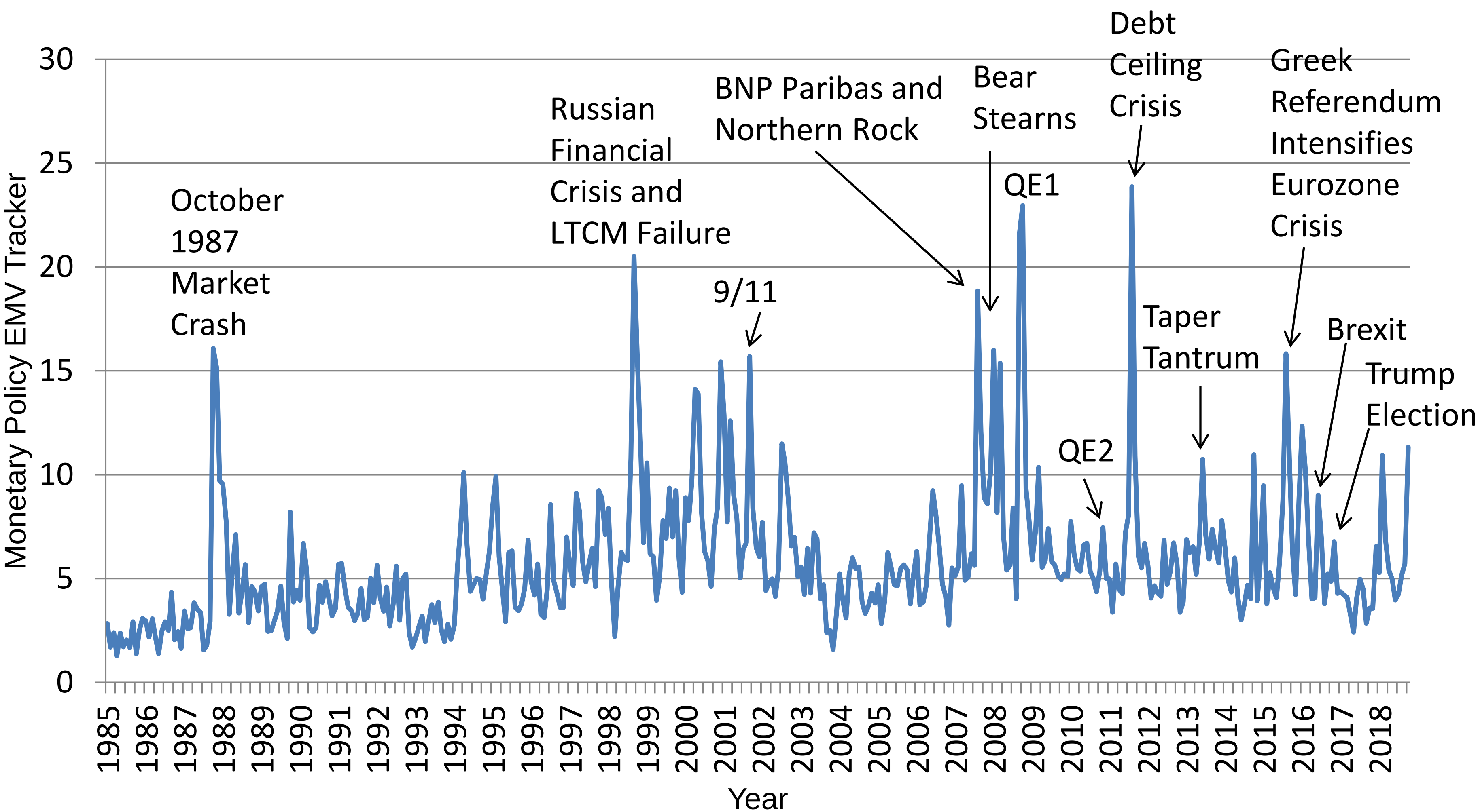
Note: This chart shows the percent of EMV articles that contain one more terms in **Trade Policy** by month**.** See Appendix B for a specification of the terms in **Trade Policy.**

Figure 9: Policy-Related EMV Tracker and BBD EPU Index, 1985-2018



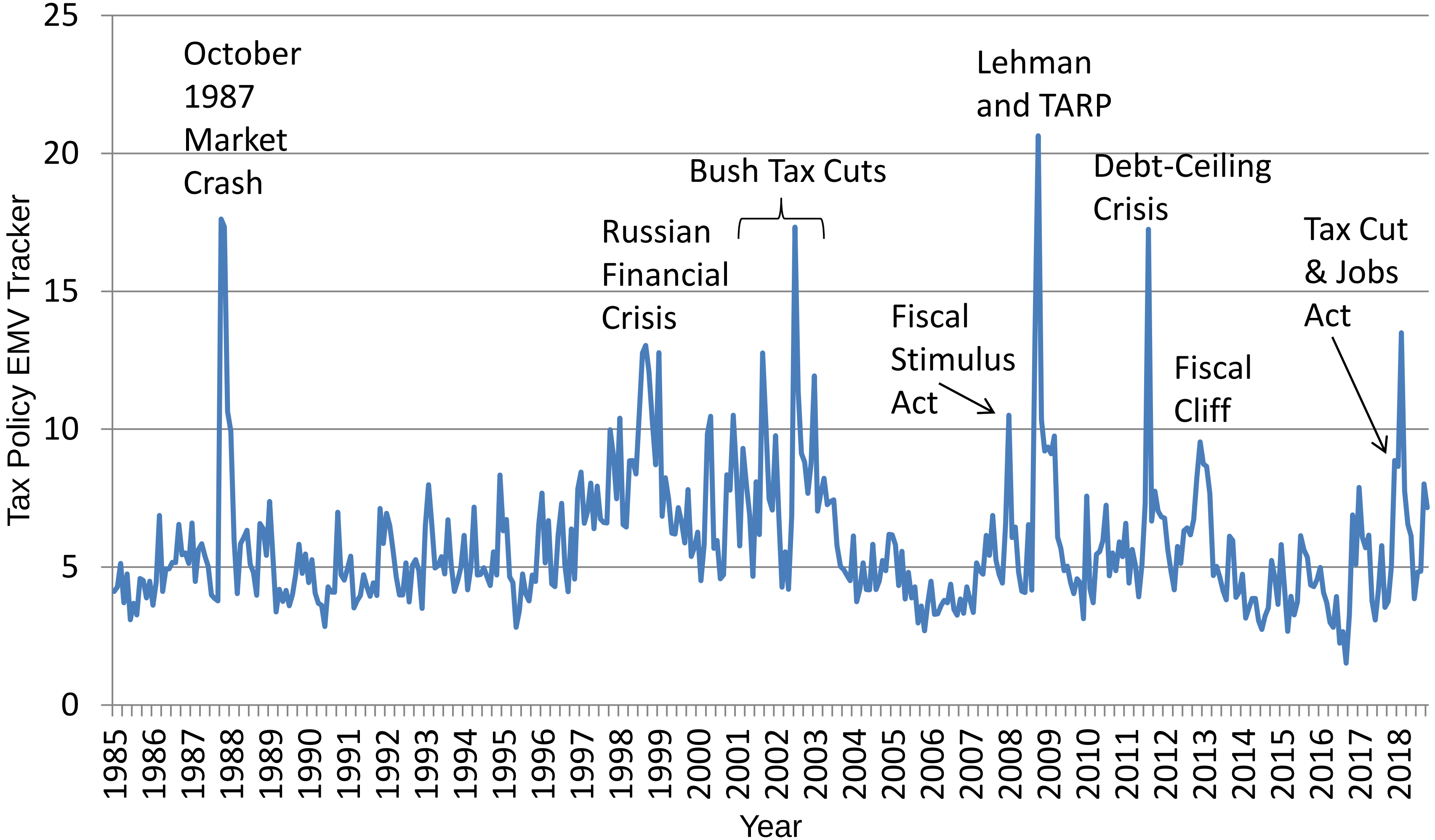
Notes: The BBD EPU Index is from Baker Bloom and Davis (2016). To construct the Policy-Related EMV tracker, we multiply our overall EMV tracker by the fraction of EMV articles the discuss policy matters. We multiplicatively rescale Policy-Related EMV to match mean of the BBD EPU Index from 1985 to 2009.

Figure 10: Monetary Policy EMV Tracker, 1985-2018



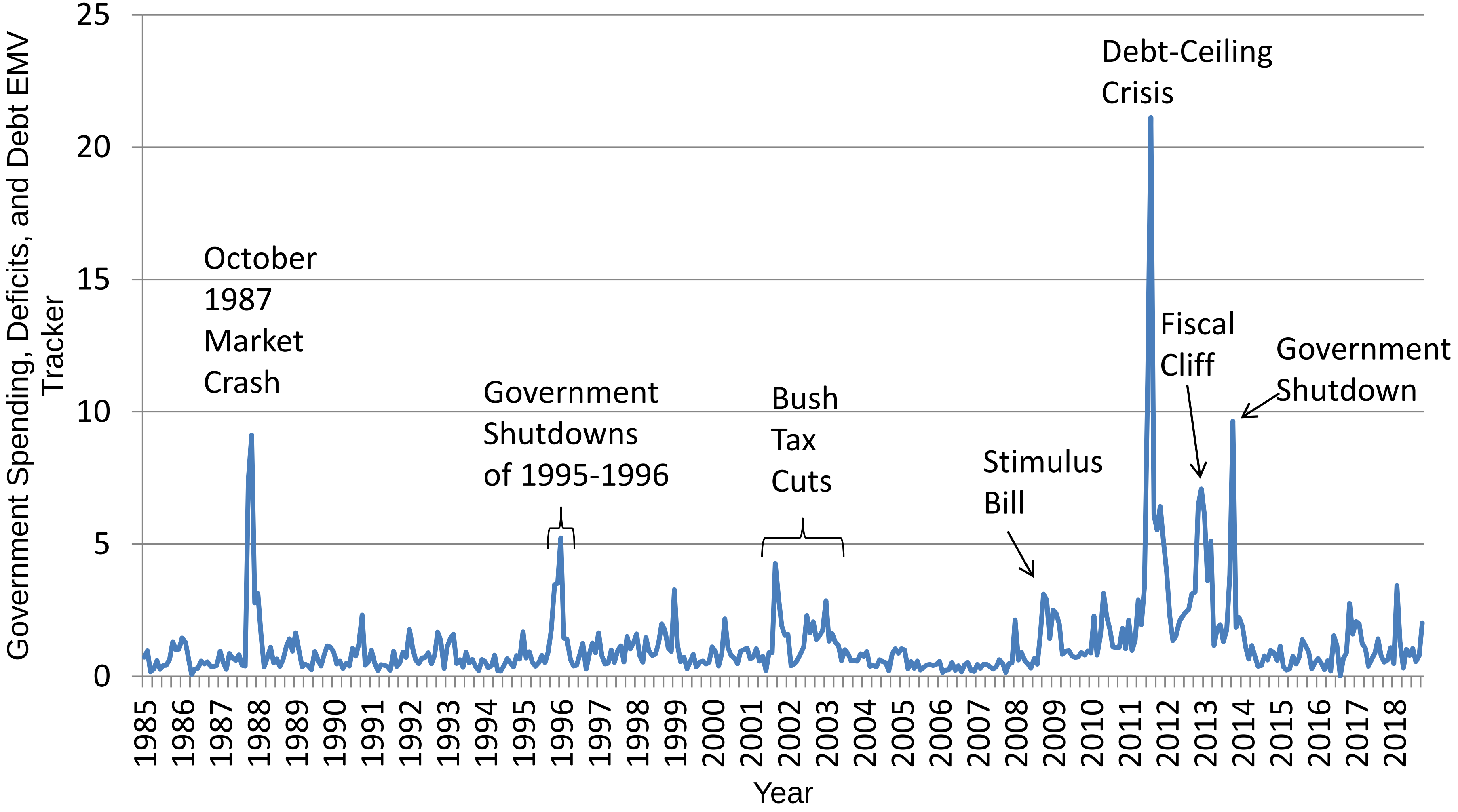
Notes: We construct the Monetary Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Monetary Policy.** See Appendix B for the list of terms.

Figure 11: Tax Policy EMV Tracker, 1985-2018



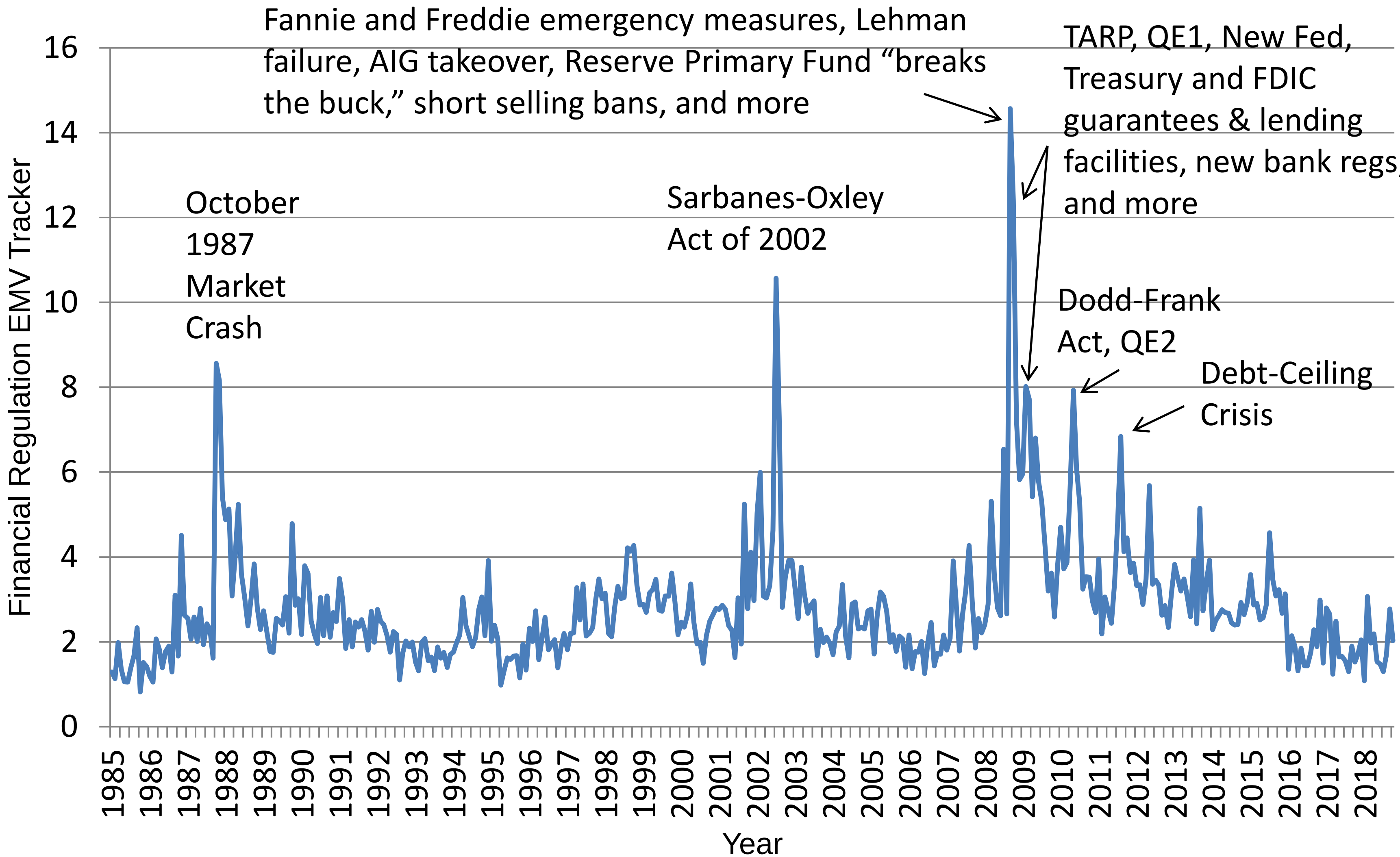
Notes: We construct the Tax Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Taxes.** See Appendix B for the list of terms.

Figure 12: Government Spending, Deficits and Debt EMV Tracker, 1985-2018



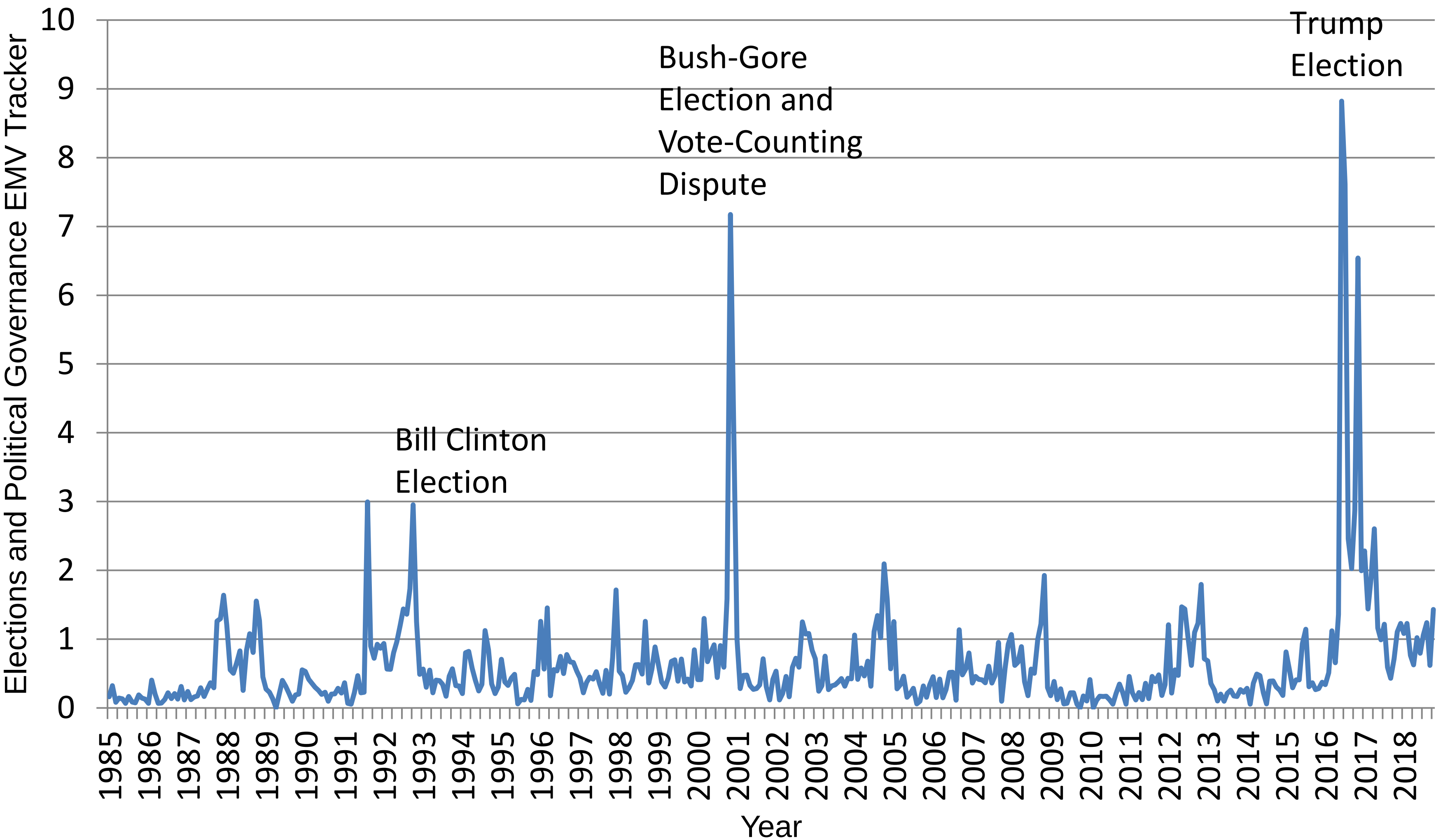
Notes: We construct the Government Spending, Deficits and Debt EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Government Spending, Deficits and Debt.** See Appendix B for the list of terms.

Figure 13: Financial Regulation EMV Tracker, 1985-2018



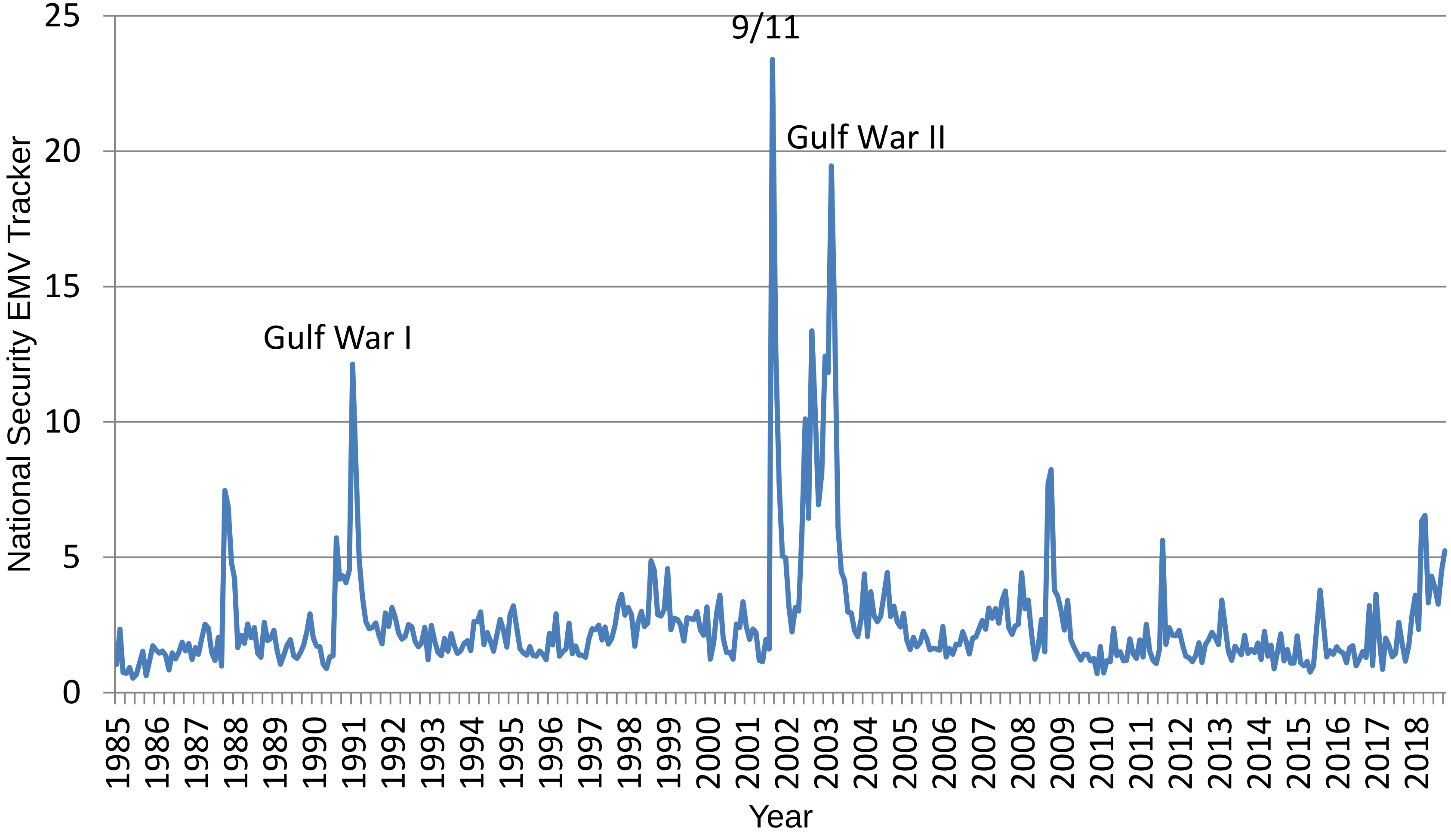
Notes: We construct the Financial Regulation EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Financial Regulation.** See Appendix B for the list of terms.

Figure 14: Elections and Political Governance EMV Tracker, 1985-2018



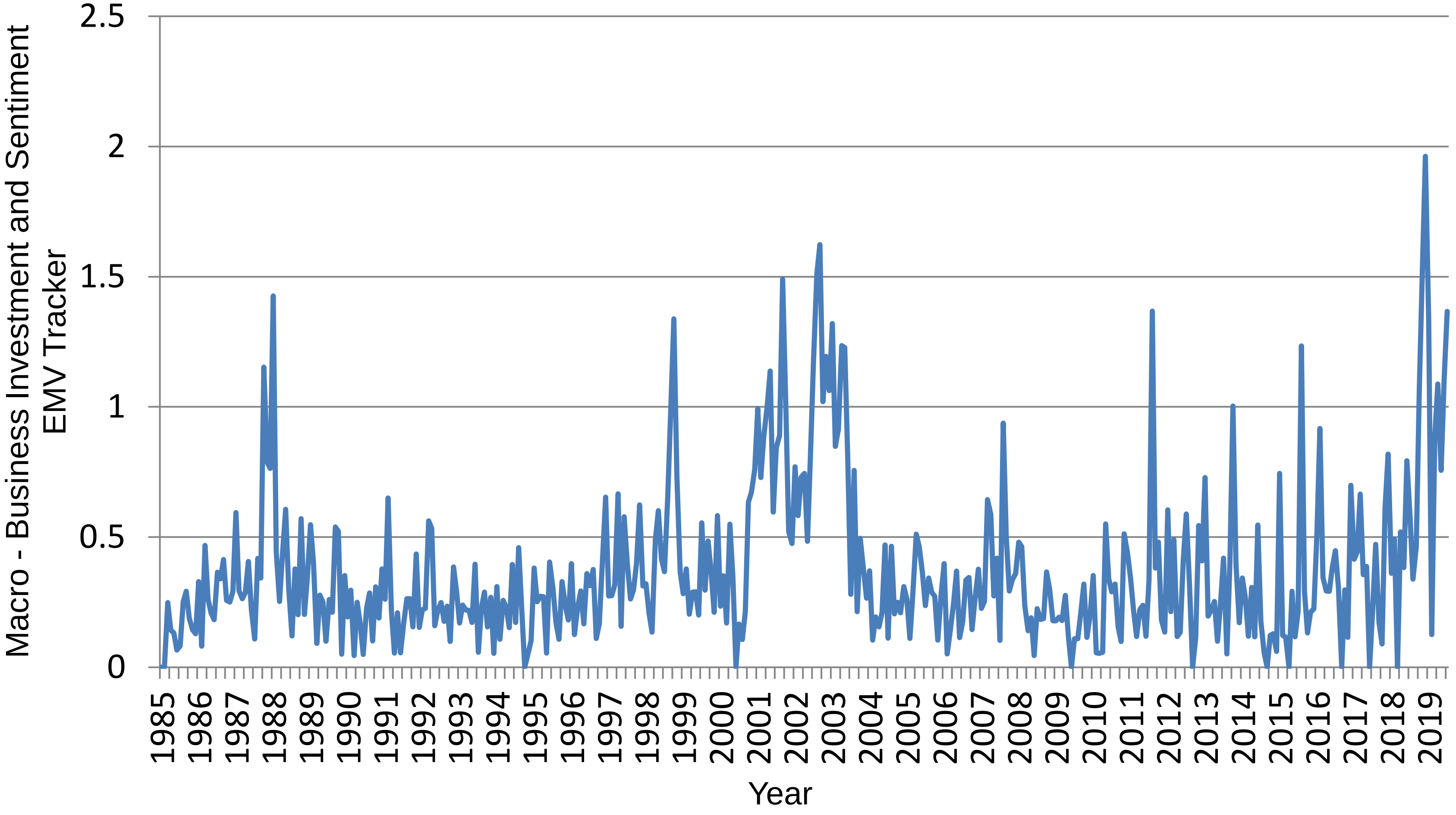
Notes: We construct the Elections and Political Governance EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Elections and Political Governance.** See Appendix B for the list of terms.

Figure 15: National Security EMV Tracker, 1985-2018



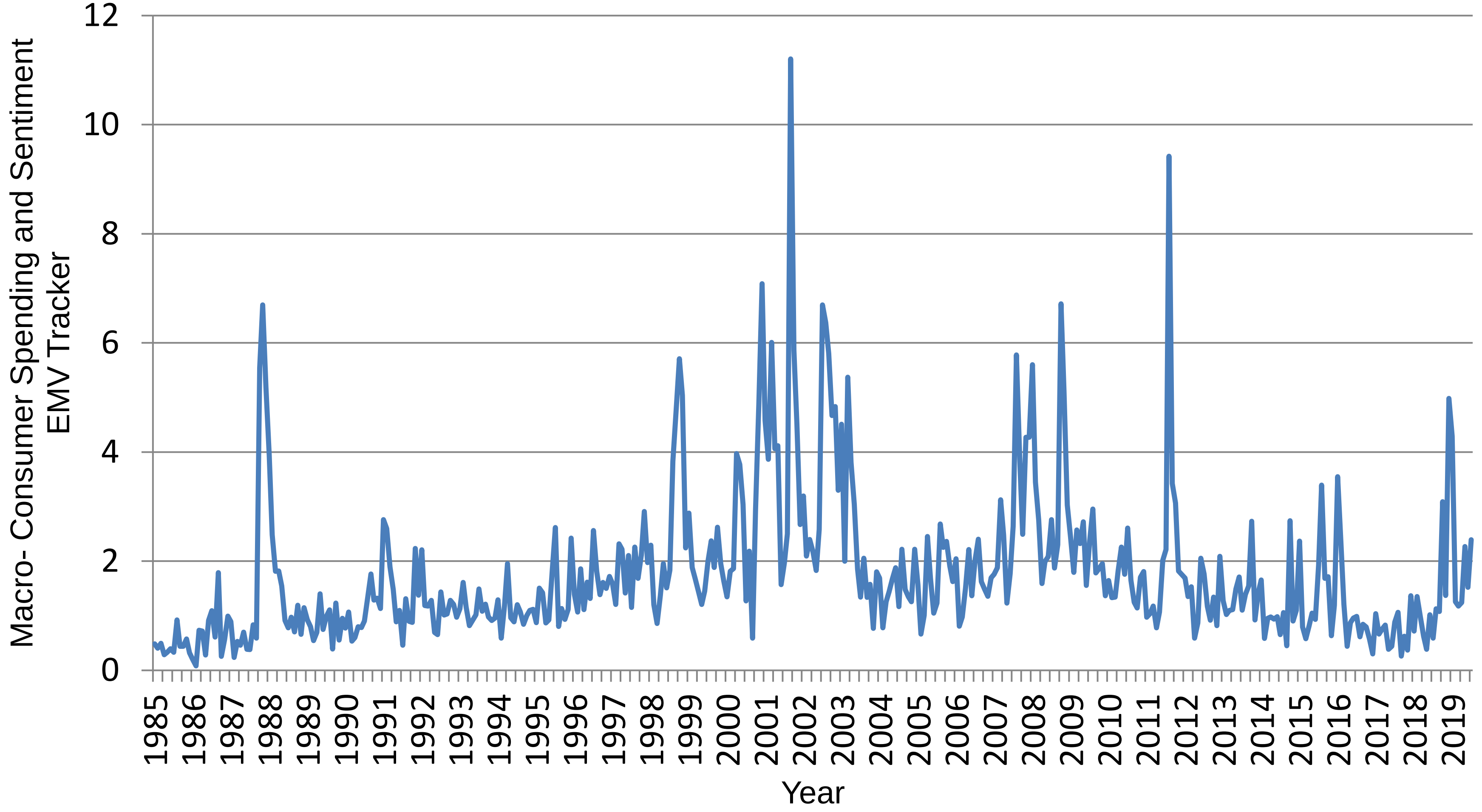
Notes: We construct the National Security EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **National Security.** See Appendix B for the list of terms.

Figure 16: Macro – Business Investment and Sentiment EMV Tracker, 1985-2019



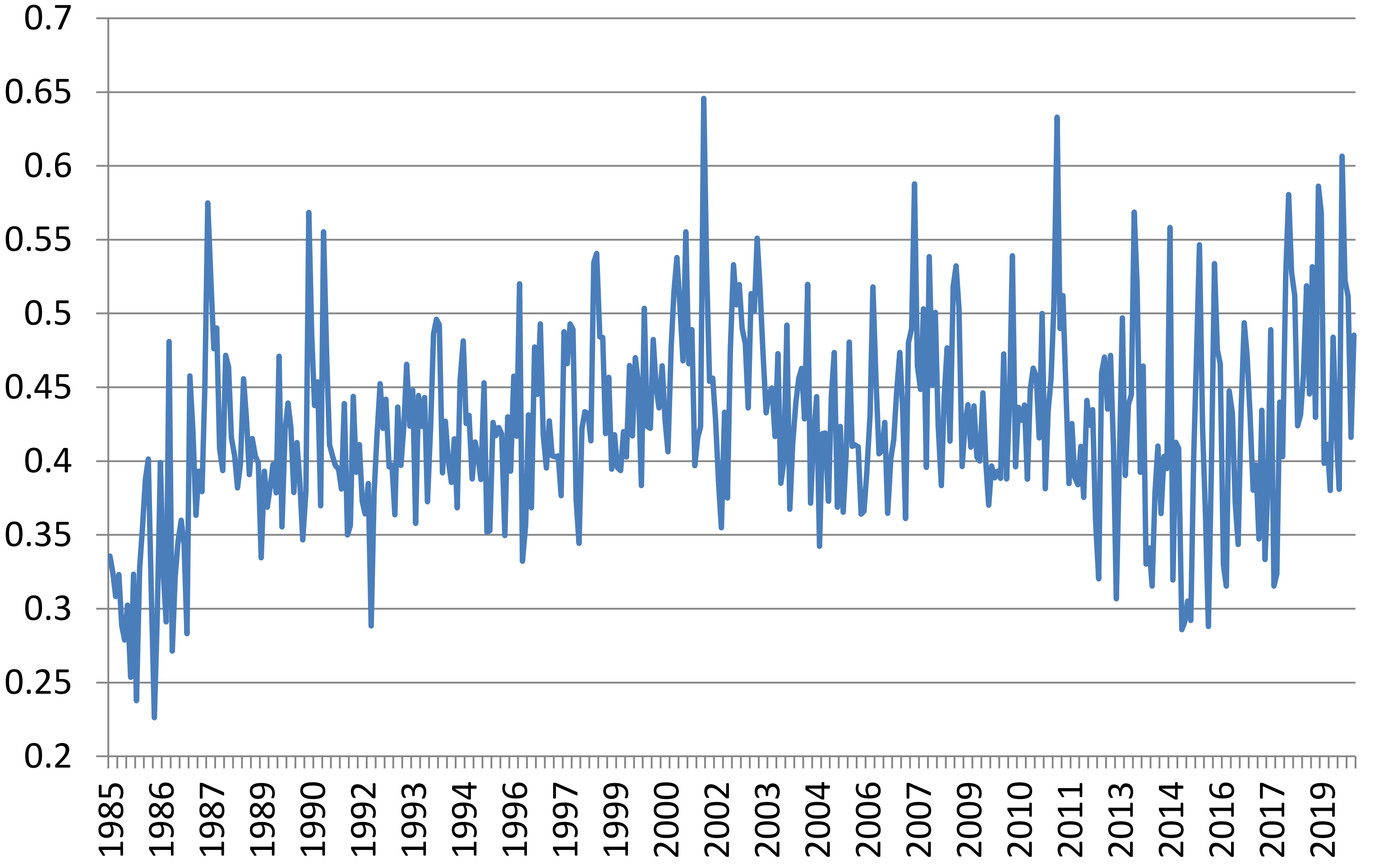
Notes: The Macro – Business Investment and Sentiment EMV Tracker is constructed as our EMV Index multiplied by the share of EMV Articles that contain one or more terms in the "Macro – Business Investment and Sentiment" termset which can be found in the Appendix.

Figure 17: Macro – Consumer Spending and Sentiment EMV Tracker, 1985-2019



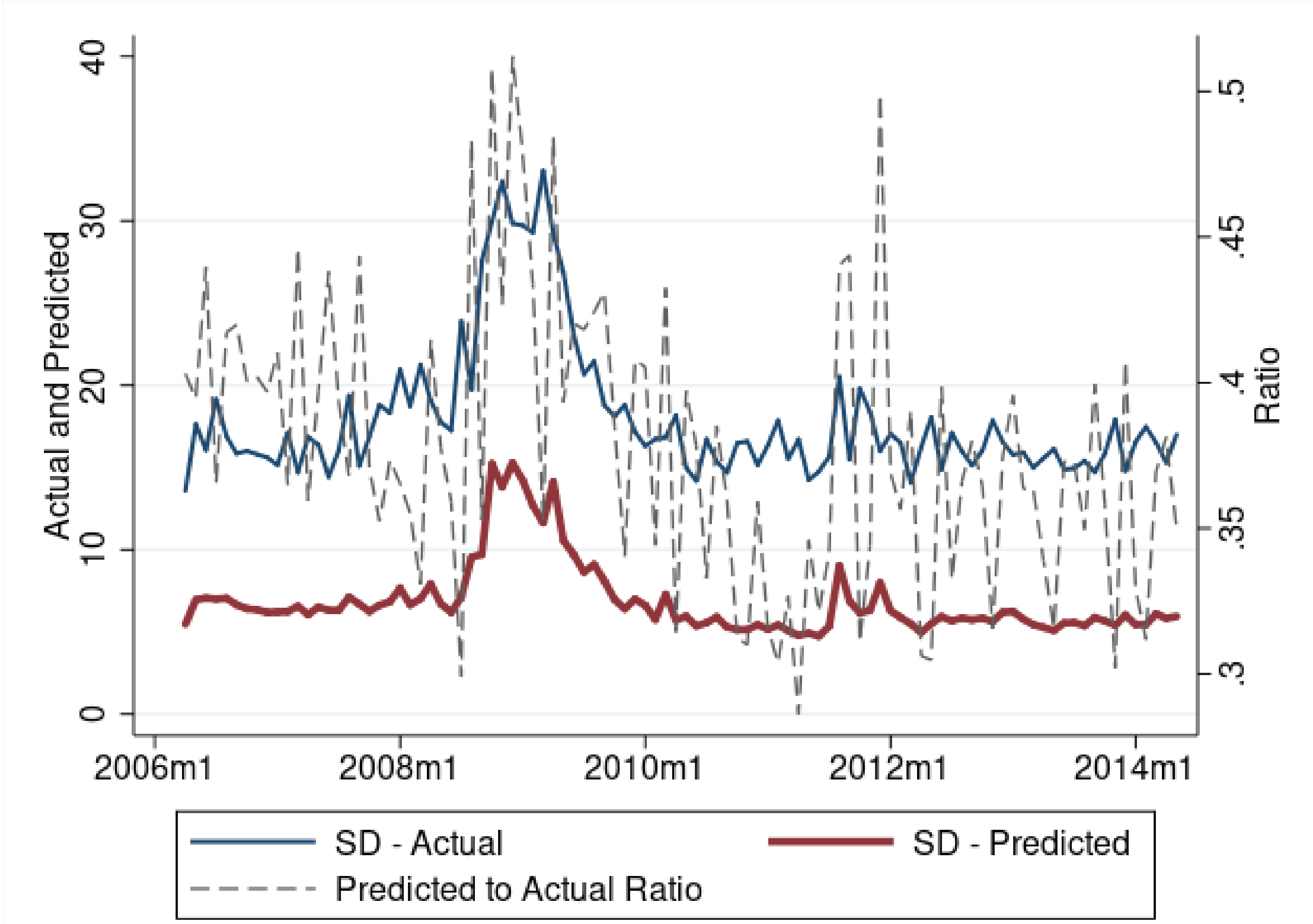
Notes: The Macro – Consumer Spending and Sentiment EMV Tracker is constructed as our EMV Index multiplied by the share of EMV Articles that contain one or more terms in the "Macro – Consumer Spending and Sentiment" termset which can be found in the Appendix.

Figure 18: Sentiments EMV Share, 1985-2019



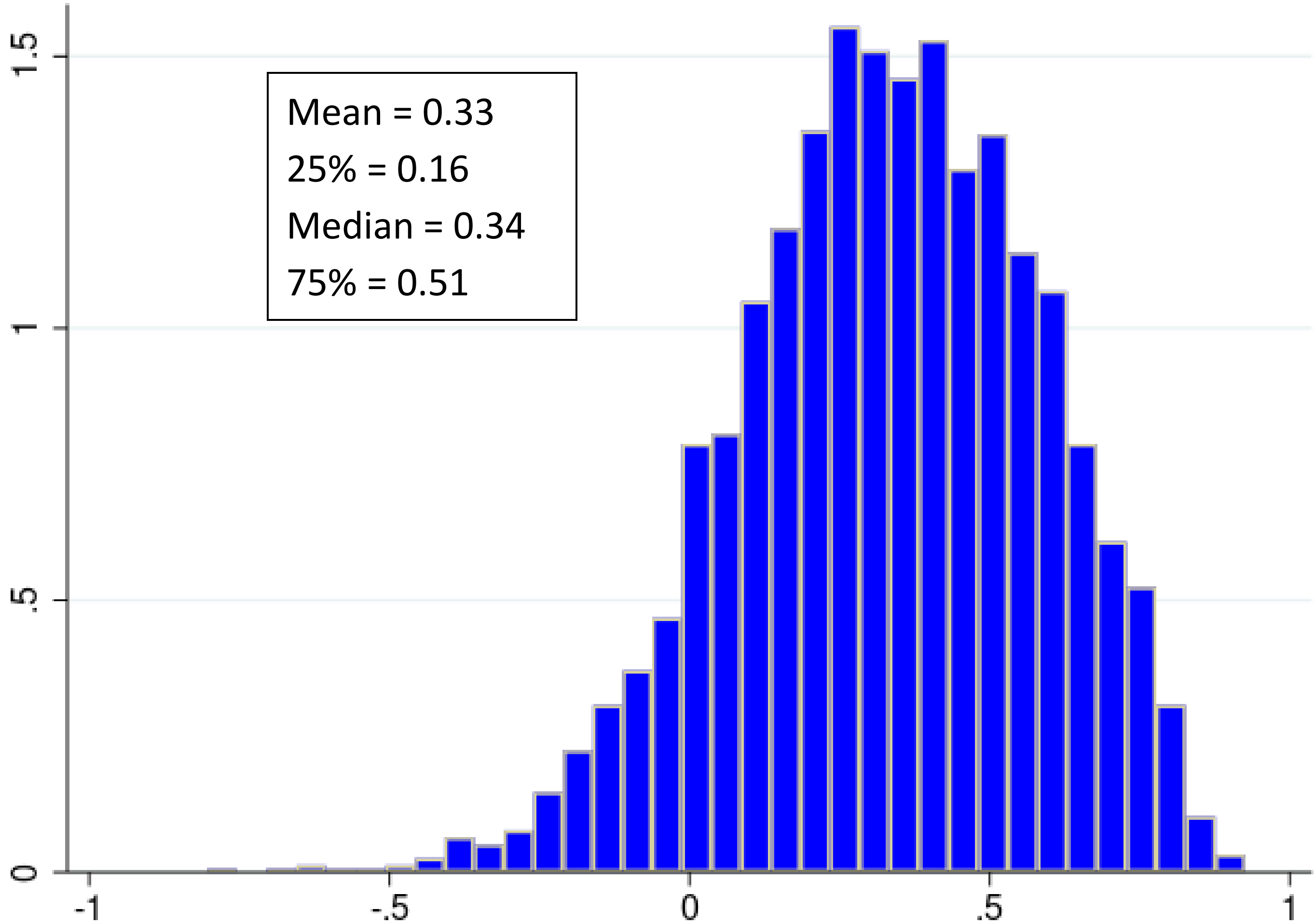
Notes: This time series is constructed as the share of EMV Articles that contain one or more terms in the "Sentiments" termset which can be found in the Appendix.

Figure 19: Cross-Firm Standard Deviation of Realized Volatility – Actual versus Fitted Values (2006-2014)



Notes: Firm and time fixed effects were swept out in a prior step. Then, the specification from Table 2 column (6) was run and predicted values for realized volatility at the firm-month level were constructed. The figure plots the standard deviation across firms for each point in time for both the residualized realized volatility and the predicted values.

Figure 20: Firm-Level Correlations between Actual and Fitted Realized Volatility



Notes: Firm and time fixed effects were swept out in a prior step. Then, the specification from Table 2 column (6) was run and predicted values for realized volatility at the firm-month level were constructed. The figure plots the histogram of time series correlation coefficients between the actual residualized realized volatility and the predicted values for each firm. Firms with less than 12 monthly observations were dropped.

Appendix A. Additional Information About Our Text Sources

Figure A.1 plots the total number of articles in the newspapers we draw on in constructing our Equity Market Volatility (EMV) tracker and related measures. The total article counts fluctuate in the range of 60-90 thousand per month in the first 16 years of our sample period, then drift down, reaching lows of about 35,000 per month.

The rightmost column of Table A.1 reports average daily article counts by newspaper from 1985 to 2017. The remaining columns report average daily counts and percentages of all articles that satisfy various criteria defined by our **E, M** and **V** term sets. Not surprisingly, the *Wall Street Journal* stands out for percent of articles devoted to topics encompassed by our term sets.

Five newspapers are not available to us for the entire 1985-2018 time period. Access World News discontinued coverage of the *Dallas Morning News* from July 2016. The ProQuest newspaper archive covers the *New York Times* through 2015 only, as of this writing. Access World News coverage of *USA Today* from Access World News is missing in 1985, 1986 and the first half of 1987. roquest archive coverage of the *Houston Chronicle* is missing for most of 1985, and its coverage of the *Washington Post* is missing in 1985 and 1986.

When missing, we impute scaled counts using fitted values from the regressions,

$$SC_{jt} = \alpha^j + \sum_{i \in N} \beta_i^j SC_{it} + \varepsilon_{it}^j, \quad \text{for } j \in N^{Miss}$$

where N^* is the set of newspapers with complete coverage (*Boston Globe, Chicago Tribune, Los Angeles Times, Miami Herald, San Francisco Chronicle,* and *Wall Street Journal*), N^{Miss} is the set of newspapers with missing coverage, and SC_{it} is the scaled EMV frequency count for newspaper *i* in month *t.* We run this regression from 1988 to 2015 for each paper ion N^{Miss} and use it to impute missing SC_{jt} values in other months.

Table A1: Articles per Day by Term Set Category, 1985-2017

	Articles in Set ***E***		Articles in **E∩V**		Articles in **E∩M**		Articles in **E∩M∩V**		All Articles per Day
	Per Day	Percent	Per Day	Percent	Per Day	Percent	Per Day	Percent	
Dallas Morning News	19.33	11.3	2.24	1.3	1.39	0.8	0.38	0.22	171.2
Houston Chronicle	18.97	11.2	2.31	1.4	1.50	0.9	0.38	0.23	169.6
Miami Herald	20.03	10.6	2.23	1.2	1.30	0.7	0.33	0.18	189.5
San Francisco Chronicle	12.44	13.2	1.56	1.7	1.02	1.1	0.26	0.28	94.1
USA Today	18.35	13.5	2.89	2.1	2.18	1.6	0.70	0.51	135.8
Boston Globe	20.75	14.0	3.14	2.1	1.64	1.1	0.51	0.35	147.9
Chicago Tribune	27.43	9.7	4.29	1.5	2.74	1.0	0.92	0.32	283.9
Wall Street Journal	44.17	39.7	10.58	9.5	9.53	8.6	3.62	3.25	111.3
New York Times	54.32	13.4	9.67	2.4	6.93	1.7	2.16	0.54	412.0
Los Angeles Times	48.90	17.8	6.75	2.5	3.60	1.3	1.14	0.41	274.8
Washington Post	41.34	20.4	7.34	3.6	3.66	1.81	1.18	0.58	202.6

Notes: See main text, Section 2.1 for definitions of the **E, M** and **V** term sets. The last column reports articles per day based on a count of weekdays per year. The Dallas Morning News coverage stops in May 2016, the New York Times coverage stops at the end of 2015, the USA Today coverage begins in the middle of 1987, the Houston Chronicle coverage begins near the end of 1985, and the Washington Post coverage begins in 1987, so the days are adjusted for those newspapers.

Appendix B. Category-Specific Term Sets

Our term sets for the Policy-Related Categories build on Baker, Bloom and Davis (2016) and Davis (2017). We developed terms sets for the General Economic Categories for this paper. We group related terms into topics within categories, as indicated by { }. These topical groupings play no role in counting methods or analysis, but we find them helpful in conceptualizing the boundaries of each category. In defining our **Regulation** term set, we hit a ceiling on the number of terms per search query. Given this constraint, we limit our **Regulation** term set to the union of terms in the most common regulation categories plus a few generic terms indicative of government regulation.

General Economic Categories

- **Macroeconomic News and Outlook – the union of the following subcategories:**
  - **Broad Quantity Indicators:** {gdp, economic growth}, {depression, recession, economic crisis}, {macroeconomic indicators, macroeconomic news, macroeconomic outlook}, {industrial production, ism report, manufacturing index}, {rail loadings, railroad loadings}
  - **Inflation:** {cpi, inflation, consumer prices, ppi, producer prices}, {gold, silver}
  - **Interest Rates:** {interest rates, yield curve, fed funds rate, overnight rate, repo rate, T-bill rate, bond rate, bond yield}
  - **Other Financial Indicators:** {bank loans, mortgage loans}, {credit spread}, {household credit, household savings, household debt, household borrowing, consumer credit}, {business credit, business borrowing, business debt}
  - **Labor Markets:** {labor force, workforce, unemployment, employment, unemployment insurance, ui claims, jobs report, jobless claims, payroll, underemployment, quits, hires, weekly hours, labor strike}, {wages, labor income, labor earnings}
  - **Real Estate Markets:** {housing prices, home prices, homebuilding, homebuilders, housing starts, home sales, building permits, residential sales, mortgages, residential construction, commercial construction, commercial real estate, real estate}
  - **Trade:** {trade news, trade surplus, trade deficit, national exports, national imports}
  - **Business Investment and Sentiment:** {business investment, business inventories}, {business sentiment, business confidence}
  - **Consumer Spending and Sentiment:** {consumer spending, retail sales, consumer purchases}, {consumer confidence, consumer sentiment}
- **Commodity Markets:** {wheat, corn, soy, sugar, cotton, beef, pork}, {petroleum, oil, coal, natural gas}, {biofuel, ethanol}, {steel, copper, zinc, tin, platinum, rare earth metals, gold, metal, silver, aluminum, lead}, {cme, commodity exchange, cbot, nymex, lme, London metal exchange, mercantile exchange, intercontinental exchange, board of trade}, {keystone pipeline, Alaska pipeline, gas pipeline}
- **Financial Crises:** {financial crisis, financial crises}, {Northern Rock failure, Lehman failure, Lehman Brothers failure, AIG Takeover}, {euro crisis, Eurozone crisis, Greek crisis}
- **Exchange Rate**: {exchange rate}, {currency crisis}, {currency devaluation, currency depreciation}, {currency revaluation, currency appreciation}, {crawling peg, managed float}, {currency manipulation, currency intervention}
- **Healthcare Matters**: {healthcare}, {health insurance}, {Medicaid}, {Medicare}, {Affordable care act, Obamacare}, {medical liability, medical malpractice}, {prescription drug}, {drug

policy}, {food and drug administration, fda}, {VA hospital, VA healthcare, Veterans Affairs hospital, Veterans Affairs healthcare, Veterans Health Administration}, {National Institutes of Health}

- **Litigation Matters: {**lawsuit, litigation, class action, tort}, {punitive damages}, {patent infringement, trademark infringement, copyright infringement}, {medical malpractice}, {Supreme Court}
- **Competition Matters:** {antitrust, competition policy, competition law}, {federal trade commission, ftc}, {unfair business practice}, {monopoly, monopolization}, {cartel}, {price fixing, price conspiracy}, {Sherman Act}, {Robinson Patman Act}, {Clayton Act}, {Hart-Scott-Rodino}, {European Commission}
- **Labor Disputes:** {labor dispute, labor unrest, strike}, {labor litigation, employee discrimination, wage and hour litigation, labor class action}
- **Intellectual Property Matters:** {patent}, {trademark}, {copyright}, {Patent and Trademark Office}, {International Trade Commission}, {federal trade commission, ftc}, {intellectual property}, {Hatch-Waxman}, {new drug application}

<u>**Policy-Related Categories**</u>

- **Fiscal Policy**: **Taxes** ∪ **Government Spending, Deficits and Debt** ∪ **Entitlement and Welfare Programs**
    - **Taxes:** {taxes, tax, taxation, taxed}, {income tax, tax on individuals, personal tax}, {capital gains tax, tax on capital gains}, {dividend tax}, {mortgage interest deduction, deduction for mortgage interest}, {IRA account, Roth IRA, traditional IRA, 401-k}, {state and local tax deduction, deductibility of state and local tax}, {payroll tax, social security tax, social security contributions, Medicare taxes, FICA, unemployment tax, FUTA}, {sales tax, excise tax, value added tax, vat, goods and services tax, gross receipts tax}, {carbon tax, energy tax}, {corporate tax, business tax, profit tax}, {investment tax credit, accelerated depreciation}, {R&D tax credit, research and development tax credit}, {tax credit for low-income housing, low-income housing credit}, {black liquor tax credit, black liquor credit}, {ethanol credit, ethanol credit, ethanol tax rebate}, {biofuel tax credit, biofuel producer tax credit, fuel excise tax rebate, fuel tax credit, alcohol fuel credit}, {property tax}, {fiscal cliff}, {Internal Revenue Service}
    - **Government Spending, Deficits and Debt:** {government spending, government outlays, government appropriations, government purchases}, {defense spending, military spending, defense purchases, military purchases, defense appropriations}, {entitlement spending}, {government subsidy}, {fiscal stimulus}, {government deficit}, {federal budget, government budget}, {Gramm Rudman, balanced budget, balance the budget, budget battle, debt ceiling}, {fiscal cliff, government sequester, budget sequestration, government shutdown}, {sovereign debt}
    - **Entitlement and Welfare Programs:** {entitlement program, entitlement spending, government entitlements}, {social security, Supplemental Security Income, ssi, disability insurance}, {Medicaid}, {Medicare}, {supplemental nutrition assistance program, food stamps, wic program}, {unemployment insurance, unemployment benefits, TAA program}, {welfare reform, aid to families with dependent children, afdc, temporary assistance for needy families, tanf, public assistance}, {earned income tax credit, eitc},

{head start program, early childhood development program}, {affordable housing, section 8, housing assistance, government subsidized housing}

- **Government-Sponsored Enterprises and Related Agencies:** {Federal Home Loan Mortgage Association, Freddie Mac}, {Fannie Mae, Federal National Mortgage Association}, {Federal Housing Finance Agency}, {Federal Housing Agency}, {Sallie Mae, Student Loan Marketing Association}, {Government National Mortgage Association, Ginnie Mae}, {Federal Home Loan Bank}, {Federal Farm Credit Bank, Federal Agricultural Mortgage Corporation, Farmer Mac}, {Resolution Funding Corporation, REFCORP}
- **Monetary Policy**: {monetary policy}, {money supply, open market operations}, {fed funds rate}, {discount window}, {quantitative easing}, {forward guidance}, {interest on reserves}, {taper tantrum}, {Fed chair, Greenspan, Bernanke, Volker, Yellen, Draghi, Kuroda, Jerome Powell}, {lender of last resort}, {central bank}, {federal reserve, the fed}, {European Central Bank, ecb}, {Bank of England}, {bank of japan}, {people's bank of china, pboc, pbc, central bank of china}, {Bank of Italy}, {Bundesbank}
- **Regulation**: {regulation, regulatory, regulate} ∪ **Financial Regulation** ∪ **Competition Policy** ∪ **Labor Regulations** ∪ **Lawsuit And Tort Reform, Supreme Court Decisions**
  - **Financial Regulation:** {bank supervision}, {thrift supervision}, {financial reform}, {truth in lending}, {firrea}, {Glass-Steagall}, {Sarbanes-Oxley}, {Dodd-frank}, {tarp, Troubled Asset Relief Program}, {Volcker rule}, {Basel}, {capital requirement}, {stress test}, {deposit insurance, fdic}, {federal savings and loan insurance corporation, fslic}, {office of thrift supervision, ots}, {comptroller of the currency, occ}, {commodity futures trading commission, cftc}, {Financial Stability Oversight Council}, {house financial services committee}, {securities and exchange commission, sec}, {Bureau of Consumer Financial Protection, Consumer Financial Protection Bureau, CFPB}, {SBA loan program}
  - **Competition Policy:** {antitrust policy, competition policy, competition law}, {federal trade commission, ftc}, {Sherman Act}, {Robinson Patman Act}, {Clayton Act}, {Hart-Scott-Rodino}, {European Commission}
  - **Intellectual Property Policy:** {patent policy, patent law}, {trademark policy, trademark law}, {copyright law}, {Patent and Trademark Office}, {International Trade Commission}
  - **Labor Regulations:** {Department of Labor}, {national labor relations board, nlrb}, {union rights, card check, right to work, closed shop}, {wages and hours, overtime requirements}, {minimum wage, living wage}, {workers' compensation}, {Occupational Safety and Health Administration, osha, Mine Safety and Health Administration}, {employment at will, advance notice requirement, at-will employment}, {affirmative action, equal employment opportunity, eeoc}, {trade adjustment assistance}, {Davis-Bacon}, {ERISA}, {Pension Benefit Guaranty Corporation, PBGC}
  - **Immigration:** {immigration policy, immigration reform, migration reform}, {Immigration and Customs Enforcement, immigration and naturalization service}, {immigrant workers, immigrant labor}, {farm worker jobs program, farm worker program farm worker program, farmworker program, guest worker program, guestworker program, H-2A program, H-2B program}, {H-1B program, H-1B visa}, {refugee crisis}, {Schengen}
  - **Energy and Environmental Regulation:** {energy policy}, {energy tax, carbon tax}, {cap and trade}, {cap and tax}, {drilling restrictions}, {offshore drilling}, {pollution controls,

environmental restrictions, clean air act, clean water act}, {environmental protection agency, epa}, {wetlands protection}, {Federal Energy Regulatory Commission, FERC}, {ethanol subsidy, ethanol tax credit, ethanol credit, ethanol tax rebate, ethanol mandate, biofuel tax credit, biofuel producer tax credit}, {corporate average fuel economy, CAFE standard}, {endangered species}, {Keystone pipeline}, {Alaska oil pipeline, Trans-Alaska pipeline}, {greenhouse gas regulation, climate change regulation}, {Nuclear Regulatory Commission}, {Pipeline and Hazardous Materials Safety Administration}
    - **Lawsuit and Tort Reform, Supreme Court Decisions:** {tort reform}, {class action reform}, {punitive damages reform}, {medical malpractice reform}, {lawsuit reform}, {Supreme Court}
    - **Housing and Land Management:** {Federal Housing Administration}, {Federal Housing Finance Agency}, {Department of Housing and Urban Development, HUD}, {Section 8 Housing}, {Office of Fair Housing and Equal Opportunity, FHEO}, {Bureau of Land Management}, {Department of Interior}, {zoning regulations, zoning laws}, {endangered species}, {US Forest Service, United States Forest Service}
    - **Other Regulation:** {Consumer Product Safety Commission}, {Department of Education}, {Small Business Administration}, {Federal Communications Commission, FCC}, {Fish and Wildlife Service}
- **National Security**: {national security}, {war, military conflict, military action}, {terrorism, terror, 9/11}, {defense spending, defense policy, military spending}, {Department of Defense}, {Department of Homeland Security}, {Defense Advanced Research Projects Agency, DARPA}, {armed forces}, {base closure}, {military procurement}, {no-fly zone}, {Syrian war}, {Iraq war}, {Libyan war}, {Ukraine conflict, Ukraine invasion, Crimean invasion, Crimean annexation}, {South China Sea conflict}, {naval blockade, military embargo}
- **Trade Policy**: {trade policy}, {tariff, import duty}, {import barrier, import restriction}, {trade quota}, {dumping}, {export tax, export duty}, {trade treaty, trade agreement, trade act}, {wto, world trade organization, Doha round, Uruguay round, gatt}, {export restriction}, {investment restriction}, {Nafta, North American Free Trade Agreement}, {Trans-Pacific Partnership, TransPacific Partnership}, {Federal Maritime Commission}, {International Trade Commission}, {Jones Act}, {trade adjustment assistance}
- **Healthcare Policy**: {healthcare policy}, {health insurance}, {Medicaid}, {Medicare}, {Affordable care act, Obamacare}, {malpractice tort reform, malpractice reform}, {VA hospital, VA healthcare, Veterans Affairs hospital, Veterans Affairs healthcare, Veterans Health Administration}, {National Institutes of Health}
- **Food and Drug Policy:** {prescription drug act}, {drug policy}, {food and drug administration, fda}
- **Transportation, Infrastructure and Public Utilities:** {Department of Transportation}, {Federal Highway Administration}, {federal highway fund}, {National Highway Traffic Safety Administration}, {U.S. Surface Transportation Board}, {Amtrak, National Railroad Passenger Corporation}, {Bonneville Power Administration, Tennessee Valley Authority, Southeastern Power Administration, New York Public Power Authority, Santee Cooper, South Carolina Public Service Authority, Salt River Project, Los Angeles Department of Water and Power}, {Corps of Engineers}, {Federal Aviation Administration, FAA}, {Federal Maritime Commission}, {National Aeronautics and Space Administration, NASA}, {Pipeline and Hazardous Materials Safety Administration}

- **Elections and Political Governance**: {presidential election}, {Congressional election}, {parliamentary election}, {presidential impeachment}, {Brexit}, {Scottish referendum}, {Grexit, Greek exit}, {Eurozone exit, Eurozone breakup}, {military takeover, coup}, {civil war}
- **Agricultural Policy:** {Department of Agriculture, USDA}, {ethanol subsidy, ethanol tax credit, ethanol credit, ethanol tax rebate, ethanol mandate, biofuel tax credit, biofuel producer tax credit}

Appendix C. Additional Analysis and Results

Figures C.1 and C.2 display the time series of residuals for the regressions reported in Table 2, Columns (1) and (3), respectively.

Table C.1: Regressions of Stock Volatility Measures on the EMV Tracker at Various Horizons

	3mon	6mon	1yr	2yr	5yr	10yr
	(1)	(2)	(3)	(4)	(5)	(6)
	VIX_t	VIX_t	VIX_t	VIX_t	VIX_t	VIX_t
EMV_t	0.27	0.20	0.14	0.10	0.07	0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
EMV_{t-1}	0.12	0.11	0.09	0.07	0.05	0.03
	(0.07)	(0.05)	(0.04)	(0.03)	(0.03)	(0.02)
EMV_{t-2}	-0.04	-0.03	-0.02	-0.01	-0.004	-0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
VIX_{t-1}	0.74	0.79	0.83	0.87	0.90	0.93
	(0.04)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
R^2	0.94	0.94	0.95	0.95	0.95	0.94
Obs.	161	161	161	161	161	161

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 2003 to July 2016. EMV is Equity Market Volatility tracker developed in Section 2.1. VIX is the monthly average of the VIX where the VIX is measured using different horizons as stated above each column. Newey-West standard errors with maximum autocorrelation lag of 2 in parentheses.

Table C.2 expands on the VIX regressions in Table 2 by using NVIX as an explanatory variable instead of, or in addition to, our EMV tracker. There are two main results in Table C.1: First, columns (1) to (4) show that EMV outperforms NVIX in tracking the VIX. Second, columns (5) and (6) show that EMV and NVIX have independent explanatory power in the sense that neither knocks out the statistical significance of the other. Moreover, including both explanatory variables substantially improves the goodness of fit.

Table C.2: Regressions of VIX on EMV and NVIX, January 1985 to March 2016

	(1)	(2)	(3)	(4)	(5)	(6)
EMV_t	0.75		0.43		0.55	0.36
	(0.06)		(0.07)		(0.07)	(0.06)
$NVIX_t$		1.12		0.53	0.61	0.32
		(0.12)		(0.11)	(0.12)	(0.07)
VIX_{t-1}			0.58	0.65		0.50
			(0.08)	(0.05)		(0.07)
R^2	0.61	0.48	0.83	0.77	0.71	0.85
Observations	374	374	372	372	374	372

Notes: Each column reports a regression of VIX on the indicated row variables, using monthly data from January 1985 to March 2016. VIX is the monthly average of daily closing values on the CBOE 30-day implied volatility index from January 1990 onwards, appended to data from Berger et al. (2019) in earlier years EMV is Equity Market Volatility tracker developed in Section 2.1. NVIX is the news-based volatility measure developed in Manela and Moreira (2017) using front-page abstracts and headlines in the *Wall Street Journal.*

Table C.3 explores the sensitivity to alternative newspaper weightings in regressions of VIX on EMV. Column (1) replicates our baseline specification reported in Column (1) of Table 2. The remaining rows adopt the same regression specification but double the weight on each newspaper, one at a time, in constructing the EMV tracker (Panel A), drop each newspaper one at a time (Panel B), or use a single newspaper in constructing EMV (Panel C).

Figures C.3. displays a time series for the fraction of EMV articles that contain one or more of the "Policy" terms that Baker, Bloom and Davis (2016) use in constructing their newspaper-based Economic Policy Uncertainty Index for the United States. Figures C.4 to C.7 display additional category-specific EMV trackers.

Table C.3: Fit Sensitivity to Alternative Newspaper Weightings in Regressions of VIX on EMV, 1985-2017

	(1) Baseline	(2) Dallas MN	(3) Houston Chronicle	(4) Miami Herald	(5) SF Chronicle	(6) USA Today	(7) Boston Globe	(8) Chicago Tribune	(9) WSJ	(10) NYT	(11) LAT	(12) Wash. Post
Panel A: Doubling the weight on the indicated newspaper												
EMV_t	0.76	0.76	0.74	0.75	0.74	0.75	0.75	0.77	0.78	0.78	0.75	0.76
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
R^2	0.611	0.607	0.604	0.615	0.611	0.606	0.609	0.604	0.613	0.607	0.600	0.608
Panel B: Dropping the indicated newspaper												
EMV_t	0.76	0.75	0.78	0.76	0.77	0.76	0.76	0.74	0.73	0.72	0.77	0.76
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
R^2	0.611	0.603	0.613	0.598	0.603	0.607	0.605	0.611	0.598	0.603	0.618	0.610
Obs.	396	396	396	396	396	396	396	396	396	396	396	396
Panel C: Using only the indicated newspaper												
EMV_t	0.76	0.29	0.39	0.39	0.35	0.36	0.40	0.53	0.52	0.45	0.45	0.59
	(0.06)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.09)	(0.09)	(0.06)	(0.06)
R^2	0.611	0.226	0.393	0.406	0.378	0.329	0.349	0.344	0.346	0.237	0.353	0.468
Obs.	396	396	396	396	396	396	396	396	396	396	396	396

Notes: All series are at the monthly level. EMV is the Equity Markets Volatility Index. The dependent variable is always the VIX where VIX refers to the monthly average of daily close of the VIX implied volatility index on the S&P500. Columns (2)-(12) of Panel A correspond to a different version of our EMV Index as the independent variable where the version is constructed such that the column title newspaper has twice the weight as the other newspapers. Columns (2)-(12) of Panel B correspond to a different version of our EMV Index as the independent variable where the version is constructed such that the column title newspaper has been removed from the index. Robust standard errors in parentheses. The slope coefficient is statistically significant at the 1% level in all regressions.

Appendix D. Computing Firm-Level Stock Returns

Let t index trading days, and let i index the firm (i.e., it's equity security). Compute percent daily equity returns as follows:

$$ER_{i,t,t+1} = \left[ln\left(\frac{PRCCD_{i,t+1} \times TRFD_{i,t+1}}{AJEXDI_{i,t+1}}\right) - ln\left(\frac{PRCCD_{i,t} \times TRFD_{i,t}}{AJEXDI_{i,t}}\right)\right] \times 100$$

Where PRCCD represents unadjusted closing equity prices, AJEXDI is a cumulative index accounting for stock splits, reverse stock splits and stock dividend payments implemented by companies over time, and TRFD[1] is a cumulative index accounting for cash dividend payments and other cash equivalent distributions. Drop observations with daily return, as measured above, outside the range of -100 to 100.

Let $n_{i,m}$be the number of trading days for firm $i$in month$m$. Given the previous calculation for daily firm-level stock returns, we calculate firm-level monthly realized stock volatility:

$$RVol_{i,m} = \sqrt{\frac{252}{n_{i,m}} \sum ER_{i,t,t+1}^2}$$

252 is just a constant representing the approximate number of trading days in a year that we use for the annualization factor. This formula is just the standard deviation of daily returns in the month for a zero mean return[2] and expressed on an annualized basis.

We use the same formula for calculating the monthly market level realized volatility based off S&P 500 returns. The only difference is that we simplify the daily returns calculation:

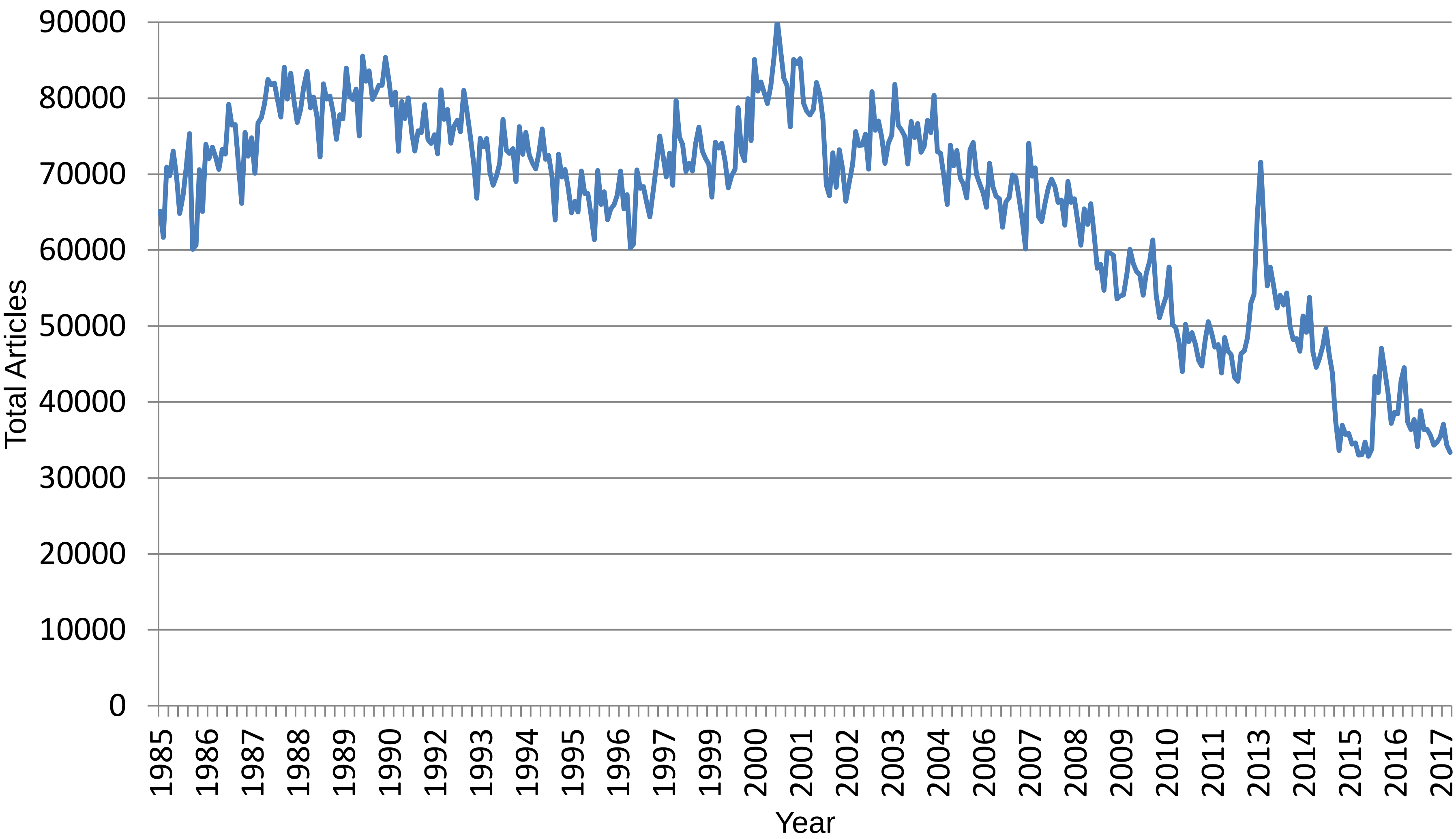
$$ER_{mkt,t,t+1} = \left[ln\left(Price_{mkt,t+1}\right) - ln\left(Price_{mkt,t}\right)\right] \times 100$$

---

[1] Many firms (38.77 percent) have missing TRFD for the entire period. In such cases, we impute TRFD=1 (i.e., we assume these companies did not implement stock splits or stock dividend payments during the considered period).

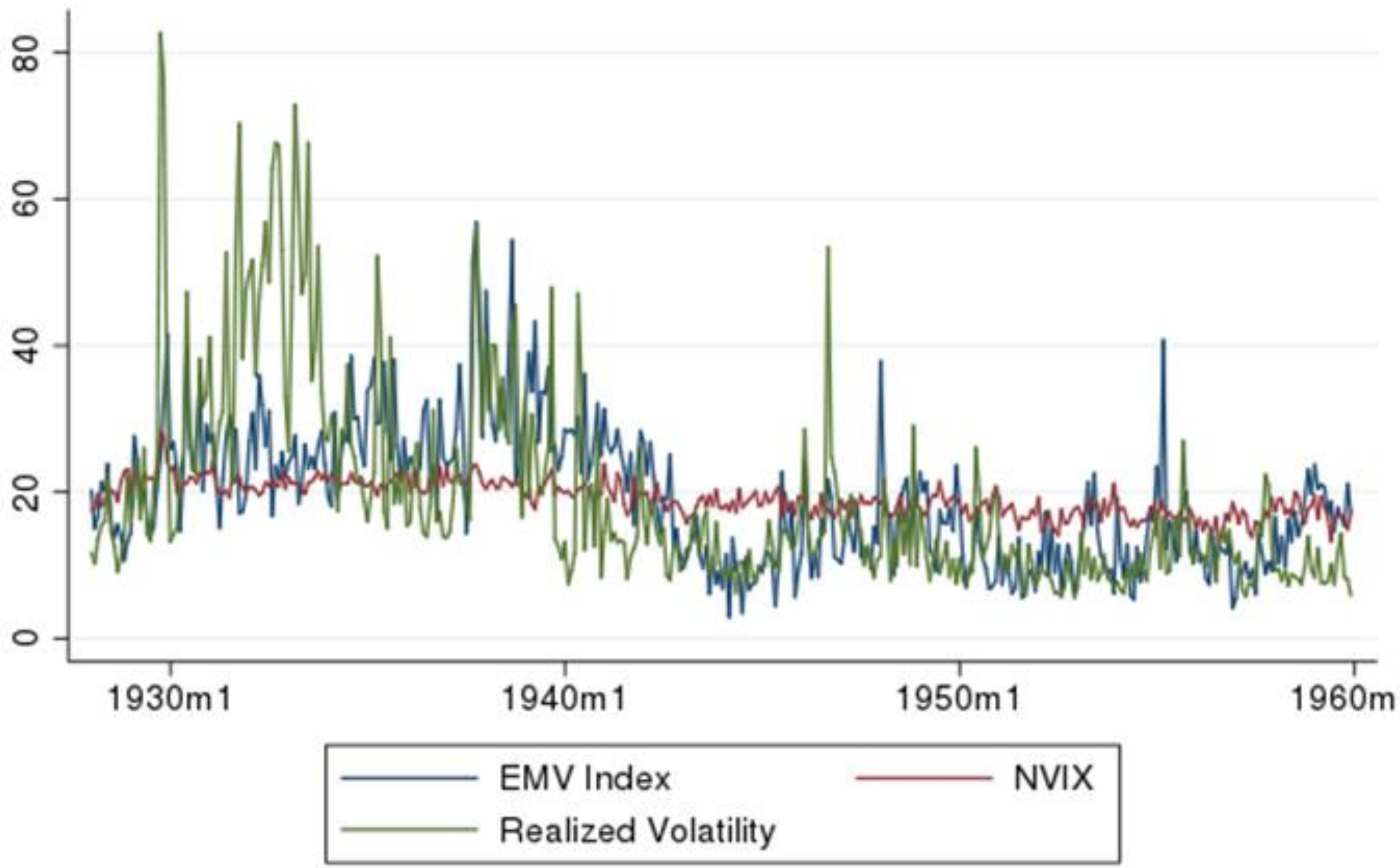
[2] The zero mean return assumption is common in practice since daily mean returns are usually very small.

Figure A.1: Total Number of Newspaper Articles, Monthly, 1985-2017



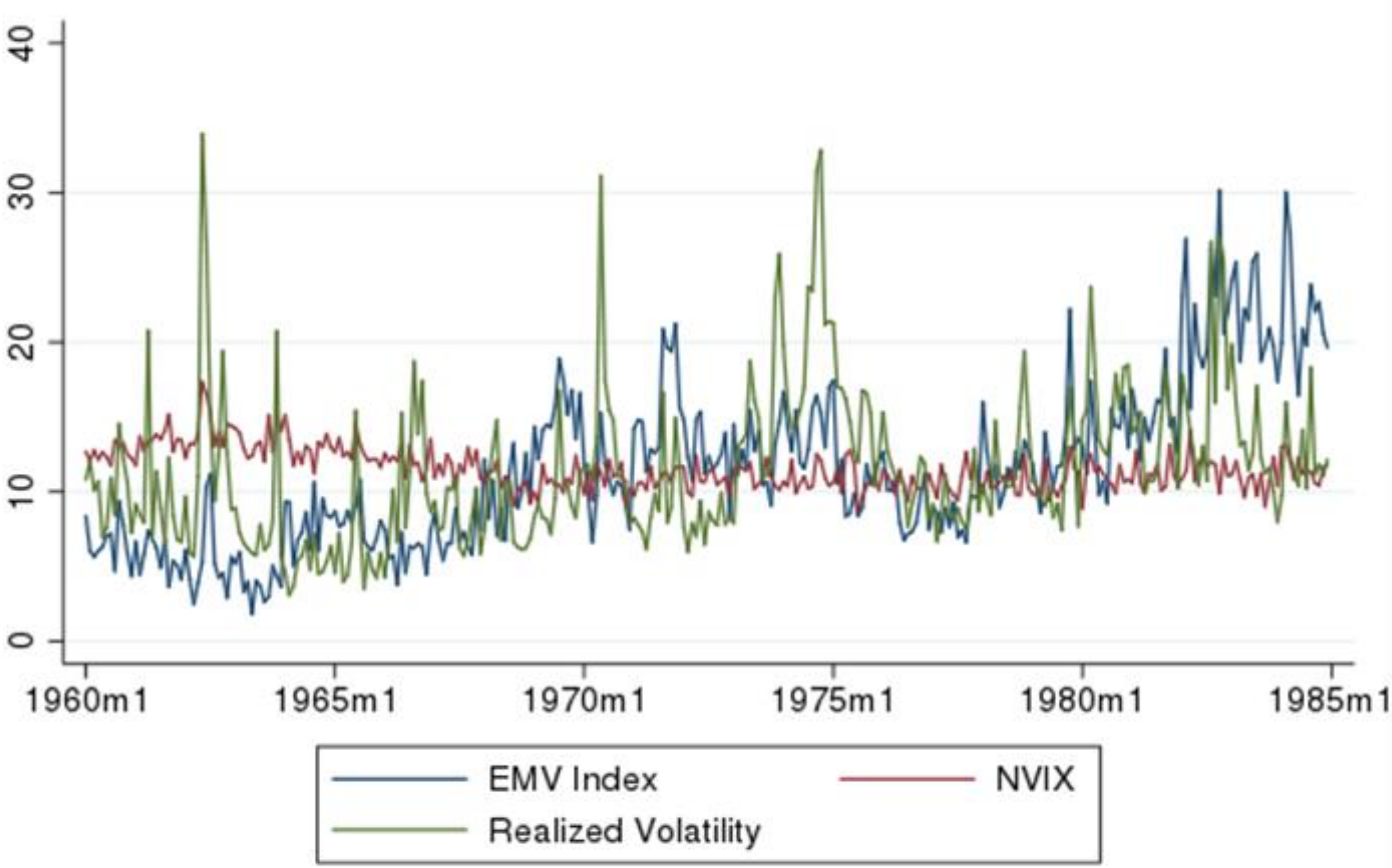
Notes: This charts shows the total number of articles in the eleven newspapers that enter into our EMV tracker. As discussed in Appendix A, digital archives for certain of our newspapers are unavailable near the beginning or end of our sample period. We scale up the article counts for non-missing papers to adjust for missing papers in certain periods.

Figure A.2: Historical EMV Index (1928-1959)



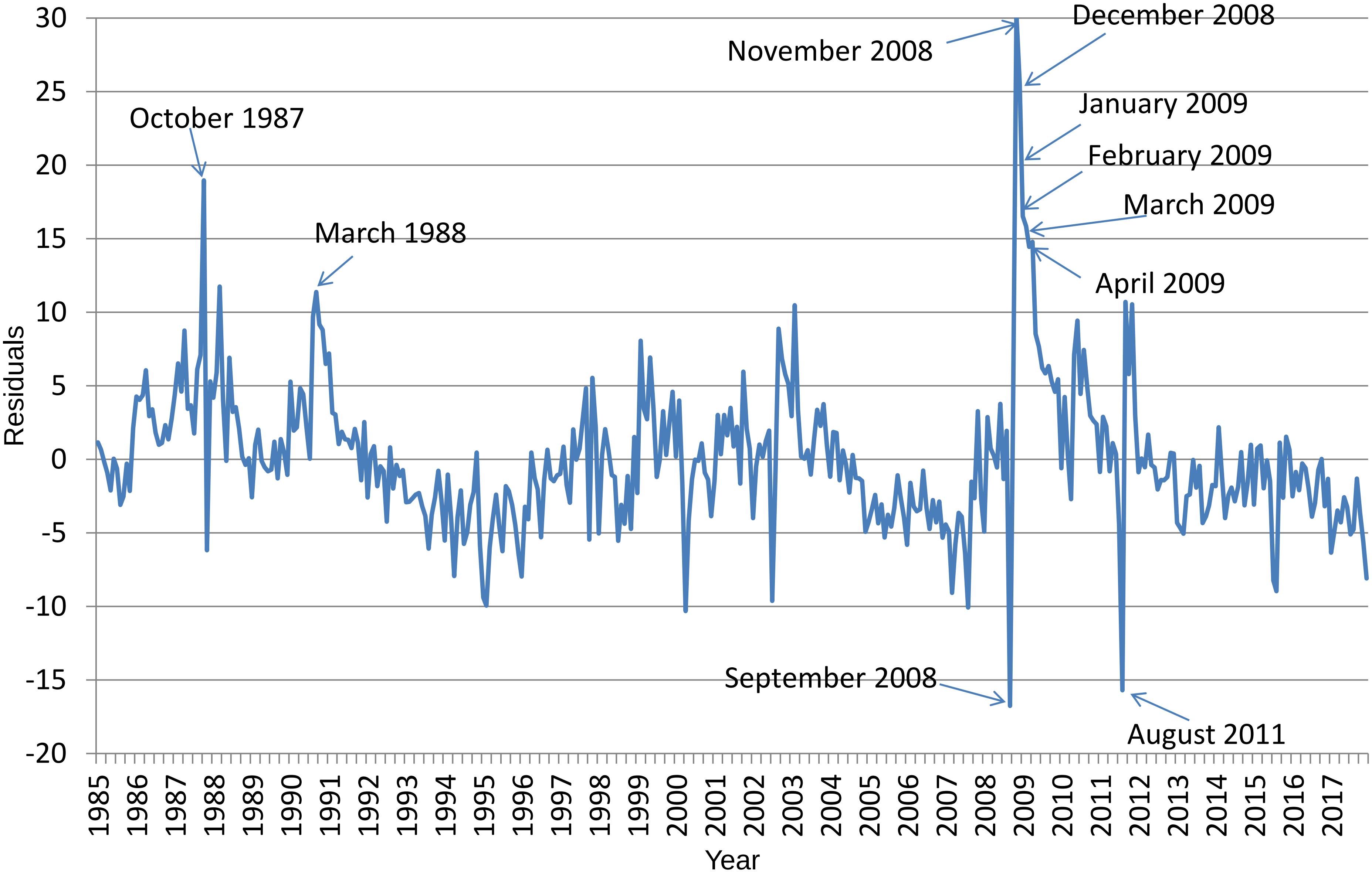
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe, Chicago Tribune, Washington Post, and Los Angeles Times.

Figure A.3: Historical EMV Index (1960-1984)



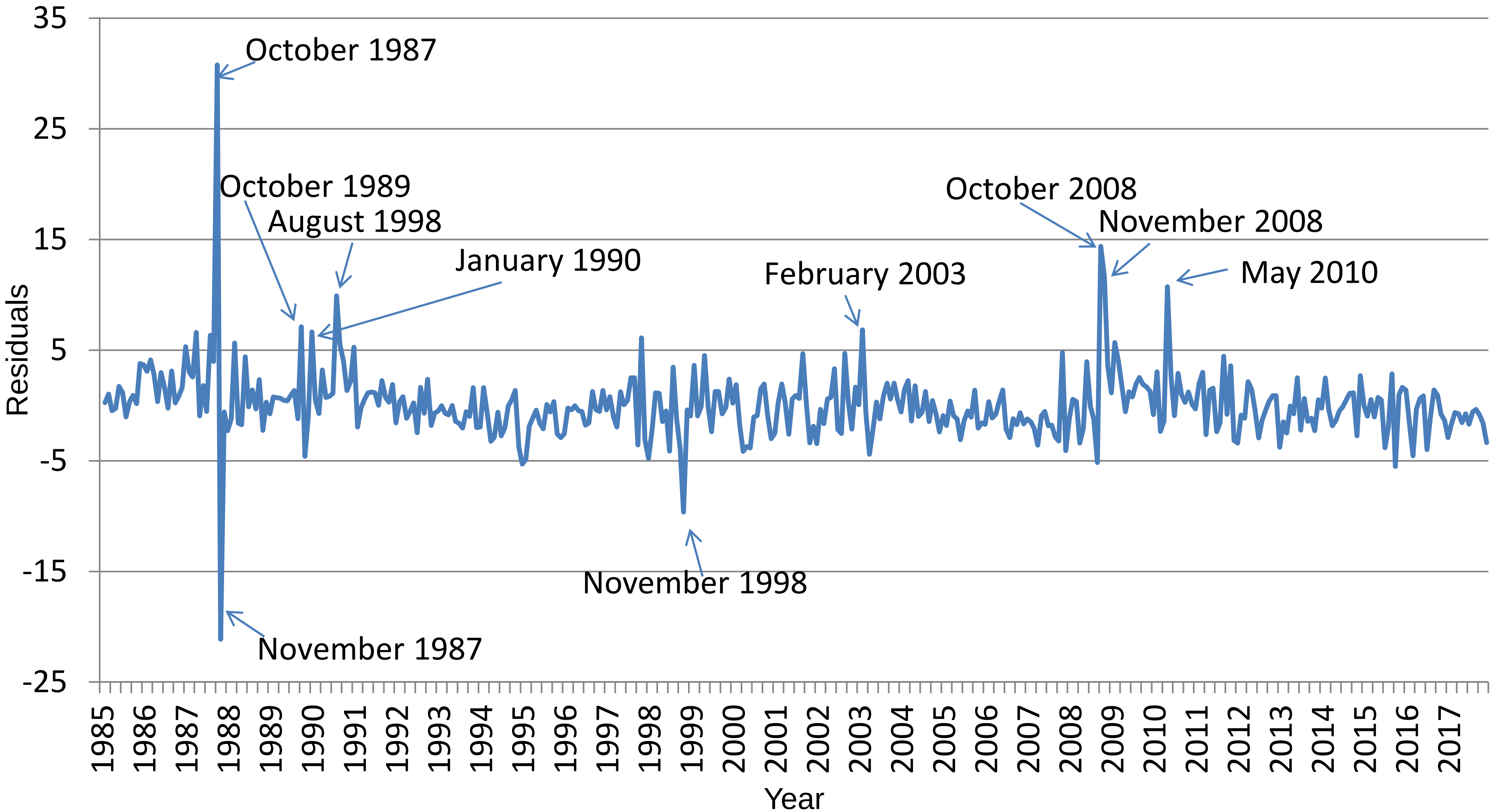
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe, Chicago Tribune, Washington Post, and Los Angeles Times.

Figure C.1: Residuals in Regression of VIX on EMV, 1985-2017



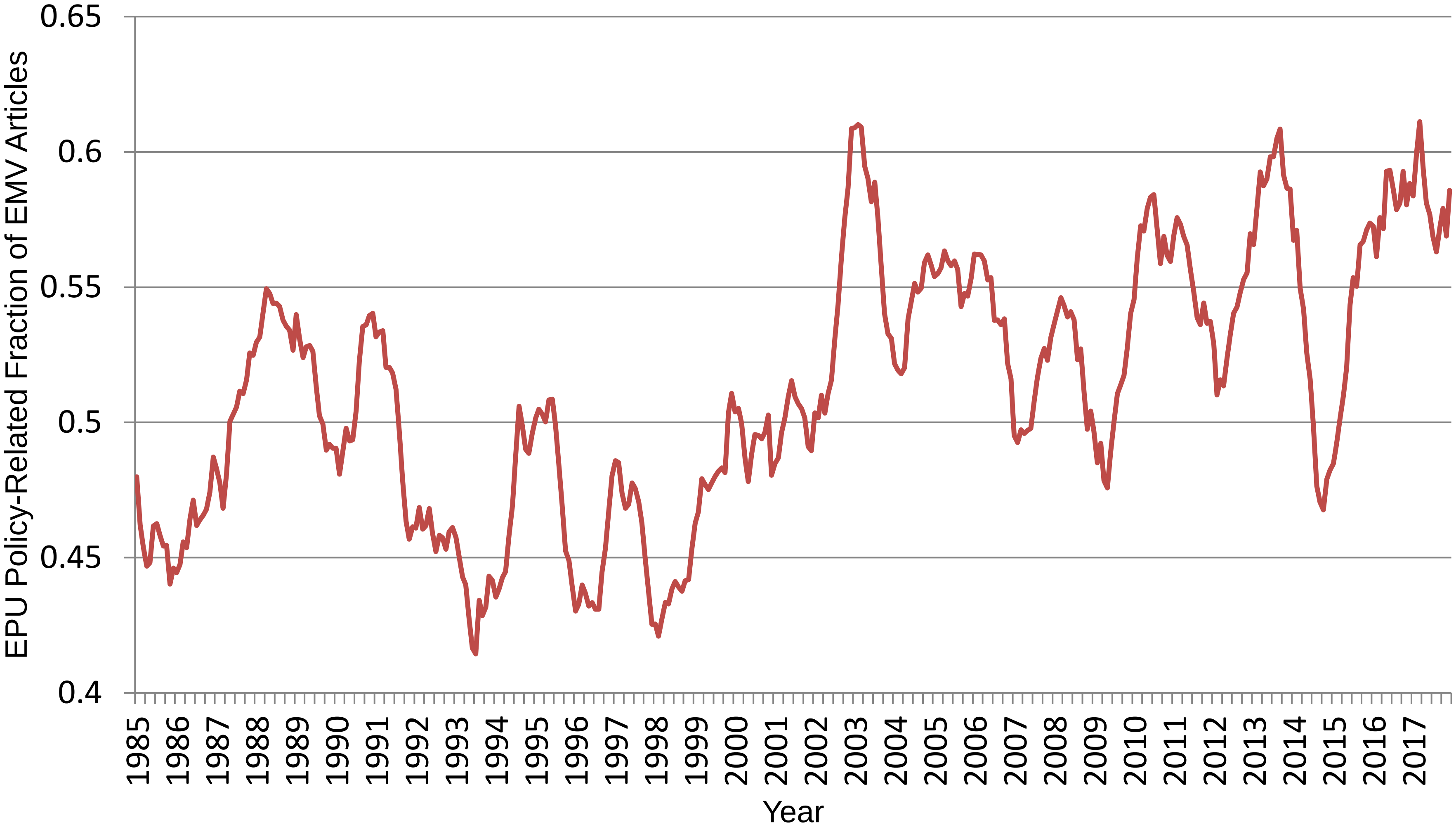
Notes: The residuals are for the specification in Column (1) of Table 2 and run from January 1985 to December 2017.

Figure C.2: Residuals in Regression of VIX on EMV and Lagged VIX, 1985 to 2017



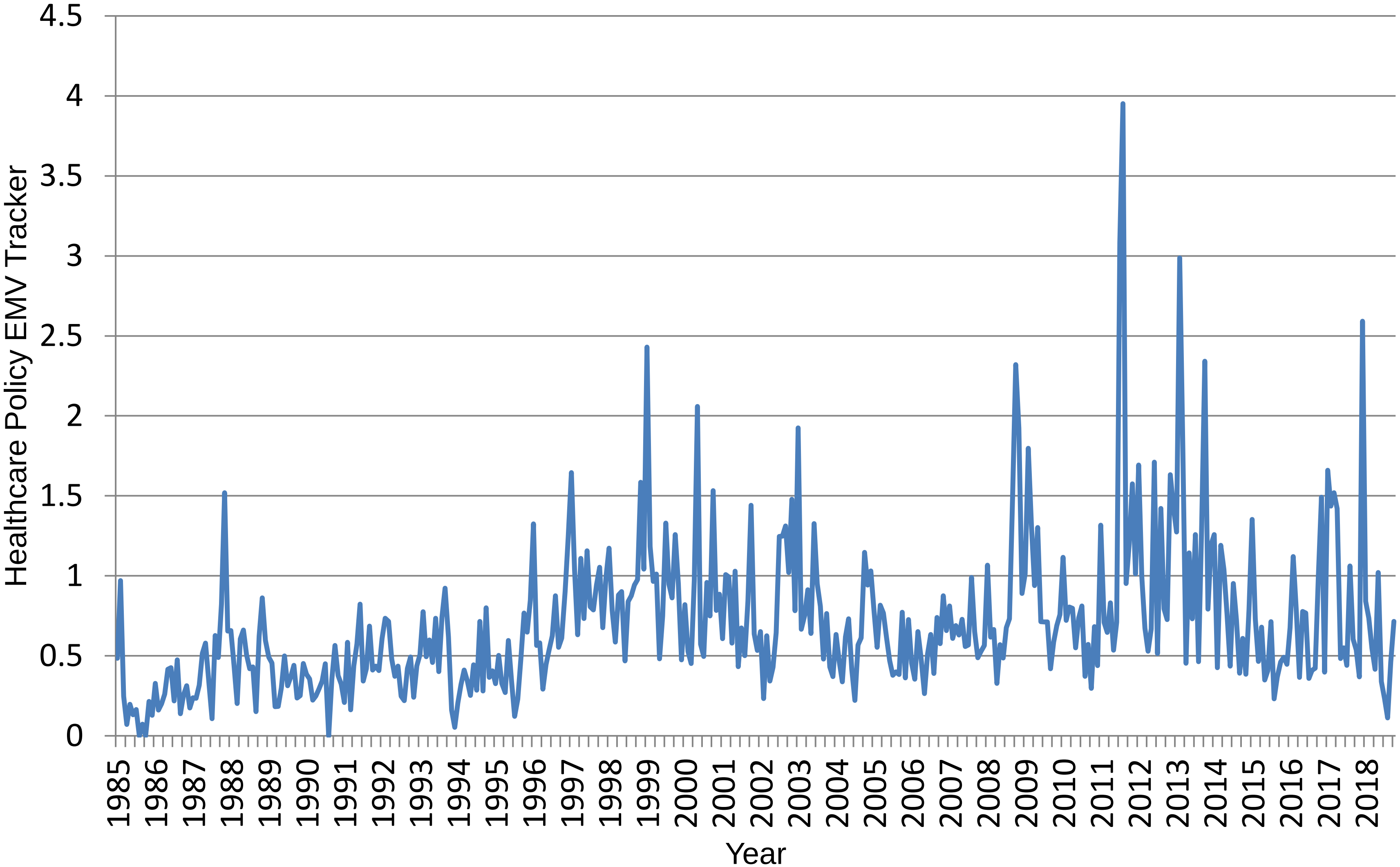
Notes: The residuals are for the specification in Column (3) of Table 2 and run from January 1985 to December 2017.

Figure C.3: Fraction of EMV Articles that Contain an EPU Policy Term, 1985-2017, 12-Month Moving Average



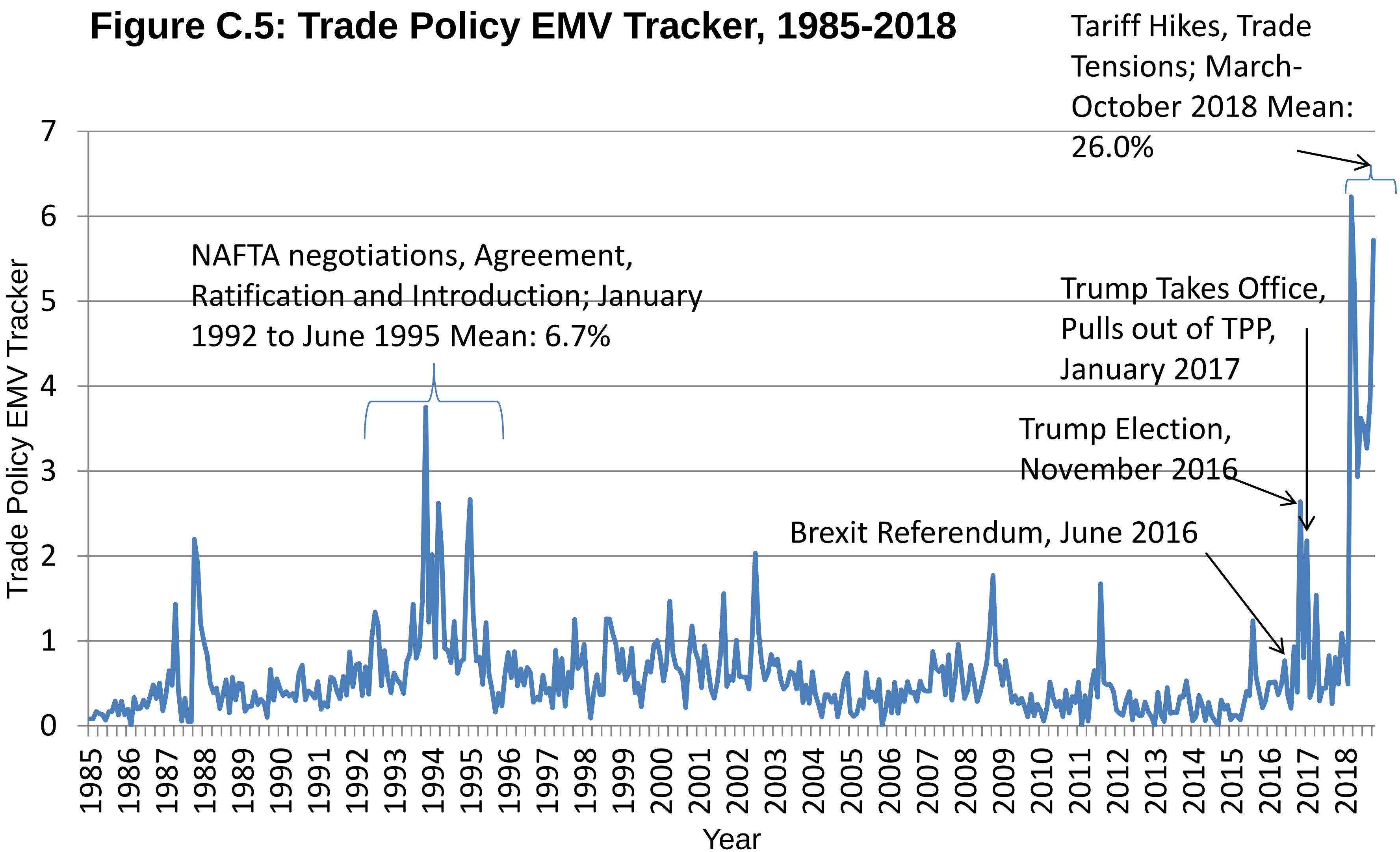
Notes: This chart shows the fraction of EMV articles that contain one or more of the policy terms used to construct the U.S. EPU Index of Baker, Bloom and Davis (2016). We compute this fraction for each newspaper and month, average over papers by month, and then compute a moving average with six lags and leads, truncating lags (leads) near the sample start (end).

Figure C.4: Healthcare Policy EMV Tracker, 1985-2018



Notes: We construct the Healthcare Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Healthcare Policy.** See Appendix B for the list of terms.

Figure C.5: Trade Policy EMV Tracker, 1985-2018



Notes: We construct the Trade Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Trade Policy.** See Appendix B for the list of terms.