NBER WORKING PAPER SERIES

POLICY NEWS AND STOCK MARKET VOLATILITY

Scott R. Baker
Nicholas Bloom
Steven J. Davis
Kyle J. Kost

Working Paper 25720
http://www.nber.org/papers/w25720

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2019, revised May 2025

Many thanks to John Cochrane, two anonymous referees, the editors Bill Schwert and Nikolai Roussanov, and participants at numerous seminars. We gratefully acknowledge financial support from the U.S. National Science Foundation and the University of Chicago Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w25720

Policy News and Stock Market Volatility
Scott R. Baker, Nicholas Bloom, Steven J. Davis, and Kyle J. Kost
NBER Working Paper No. 25720
March 2019, revised May 2025
JEL No. E44, G12, G18

ABSTRACT

We use newspapers to create Equity Market Volatility (EMV) trackers at daily and monthly frequencies. Our headline EMV tracker moves closely with the VIX and the S&P500 returns volatility in and out of sample. We exploit the volume of newspaper text to construct forty category-specific EMV trackers. News about commodity markets, interest rates, real estate markets, aggregate activity, and inflation figure prominently in EMV articles. 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. Combining our newspaper-based trackers with textual analysis of 10-K filings, we obtain monthly firm-level risk exposure measures. These measures help explain the cross-sectional structure of realized volatilities and its evolution over time, even after conditioning on firm and time fixed effects.

Scott R. Baker
Northwestern University
Kellogg School of Management
Department of Finance
and NBER
s-baker@kellogg.northwestern.edu

Nicholas Bloom
Stanford University
Department of Economics
and NBER
nbloom@stanford.edu

Steven J. Davis
Stanford University
and NBER
stevend5@stanford.edu

Kyle J. Kost
University of Chicago
kkost@uchicago.edu

A Data File is available at http://www.policyuncertainty.com/EMV_monthly.html

> 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 fluctuations 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 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 2023 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. For instance, one immediate result is the importance of news about Commodity Markets, which receives attention in over 40% of all articles that enter into our EMV tracker. Most EMV articles discuss multiple topics. Thus, we also find

[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).

that 31% mention 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 macro news, surges in the wake of episodes that involve 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, the U.S. debt-ceiling crisis in summer 2011, and the onset of the COVID-19 pandemic. 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 fluctuates over time, reaching peaks in the 2001-03 period (9/11 and Iraq Invasion), 2011-12 (U.S. debt-ceiling crisis and "fiscal cliff"), after the Brexit vote, and during the first Trump presidency. 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 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 first 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 forces that drive the volatility of equity returns. Prior research supports the relevance of news releases that drive firm-level stock price movements. For instance, Griffin, Hirschley and Kelly (2011) observe that firm-level stock prices move much more on days with information releases about the firm. Boudoukh et al. (2018) push further, showing that news items play an especially prominent role as drivers of firm-level moves that happen overnight, when there is less scope for private information or trading itself to drive returns.

Shiller (2014) 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? – by providing a means to catalog and quantify the drivers behind stock market fluctuations over time.

Our EMV trackers have several noteworthy attributes relative to AI-based methods of textual analysis. First, their construction is straightforward, transparent, easy to refine, and simple to replicate. The frequency and volume of newspaper text affords much scope for granular characterizations of the forces that underpin equity market volatility and its movements over time. Second, newspaper-based methods allow for timely and continuous updating using only news articles published on particular days. In practice, we update our EMV trackers daily in real time. To guard against look-ahead bias, we draw only from articles published on specified days and use a fixed set of terms. Finally, compared to machine-learning methods, our method affords easier, more assured access to the underlying source text. Its implementation requires only access to a search API that returns newspaper article counts. It does not require access to the full text of newspaper articles, which has become a contentious and legally contested matter.

Our real-time updates facilitate efforts to assess the out-of-sample performance of our measures, and they have had significant take-up in both academia and industry. Our EMV trackers garner direct traffic through our website (www.policyuncertainty.com) and via third-party hosting sites such as Bloomberg and FRED. We fixed our methodology for the EMV index in 2018, yielding over 5 years of out-of-sample data by 2023. As we show below, the out-of-sample performance of our indexes, when tracking the VIX and other targets, has not decreased relative to its predictive power prior to 2018. That is, our simple methodology has yielded durable out-of-sample performance despite the emergence of such dramatic developments as the COVID-19 pandemic, extraordinary policy responses to the pandemic, the escalation of the Russia-Ukraine war, and multiple conflicts in the Middle East.

[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.

Finally, 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 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, we built an Infectious Disease EMV index after the COVID-19 outbreak in Baker et al. (2020). We used that index to compare the contribution of COVID-19 to stock market volatility in 2020 to market volatility reactions to previous severe infectious disease outbreaks (SARS, MERS, Ebola Swine Flu and Bird 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 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 U.S. stock market movements. Cutler, Poterba and Summers (1989) relate returns on U.S. equities to macroeconomic data and news of "political and world events." They conclude that it's hard to explain over half the variation in aggregate stock prices through information in these sources about discount rates and future cash flows. Baker et al. (2024) consider thousands of global daily stock market moves greater than |2.5%|. 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 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.

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 (2013) document global patterns in capital flows, asset prices and credit growth that are 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 (2019) 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 policy uncertainty 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. Kelly, Manela, and Moreira (2021) 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 risk premia, broadly in line with the literature on rare disasters and asset prices.[4]

2. Methodology

2.1 Constructing an Equity Market Volatility Tracker

In constructing our Equity Market Volatility (EMV) tracker, we follow Baker, Bloom and Davis (2016) (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.[5] 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.

---

[4] See Rietz (1988), Barro (2006), Gourio (2008), Gabaix (2012) and Wachter (2013), among others.

[5] 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 less formal.

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.[6] Third, using the sampled articles, we construct a candidate EMV tracker for each permutation of elements in **M′** and **V′**.[7] 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 counts to unit standard deviation for each paper, and then average the resulting standardized, scaled counts over papers by month.[8] 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 and level specifications with quadratic and cubic terms yield the same best-fit permutation, which forms our EMV tracker utilized below:

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 assessing our term sets and our selection procedure, a few additional remarks will be helpful. We start with parsimonious **E, M′** and **V′** sets to reduce the danger of overfitting.[9] 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.

---

[6] 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.

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

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

[9] Machine-learning methods often start with an expansive feature set ("term set" in our language) and then shrink the set by penalizing terms that add complexity without materially improving in-sample performance. See, for example, Cherkassky and Ma (2004). We start with a limited set of terms, which we further shrink based on in-sample performance.

We eschew terms like "Lehman Brothers," "Bernanke" and "Iraq war" that might improve in-sample performance but perform poorly out of sample. 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, as there is no VIX counterpart for most markets.

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 and we again scale by the count of all articles in the same paper and month.[10] 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 December 2023.[11] 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, the U.S. debt-ceiling crisis in the summer of 2011, and the onset of the COVID-19 pandemic. Several other episodes trigger smaller spikes.

We validate our EMV tracker, assess its performance in various ways, and consider robustness checks in Section 3 below. It is important to note that the optimization process that yields our term set was finalized in 2018, with the index and term set first published in the March 2019 version of this paper (Baker et al., 2019). Since then, the data have been regularly updated using exactly the same method and posted to our website. As such, the data from 2019 onwards provides an ideal opportunity for out-of-sample testing, as implemented in Section 3.2 and Table 2.

[10] 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.

[11] 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. (2020) 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.

2.2 Two Extensions of the EMV Tracking Methodology

In addition to our headline monthly EMV tracker for the period from 1985 to the present, we undertake two extensions that demonstrate the flexibility of our method and provide additional tools for researchers looking to better understand equity market volatility. First, we construct an historical EMV tracker that runs back 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. Second, we build a daily EMV tracker from 1985 to the present using the thousands of US newspapers covered by the Access World News Newsbank database. This daily EMV tracker relies on counts of EMV articles summed across all English-language US newspapers in the Newsbank database.

2.3 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. 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," "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 sets of terms 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. We extend these policy-related term sets 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.

We find that news and other remarks about the Macroeconomic Outlook feature very prominently, appearing in 72% of all EMV articles.[12] News about Commodity Markets appear in 44% of EMV articles, while news about Interest Rates figures in 31%. Policy-related categories, including aggregated categories for Fiscal Policy and Regulation, also have high representation within these articles. Tax Policy and Monetary Policy each receive attention in 30% of EMV articles, the aggregated Regulation category features in 25%, and National Security matters figure in 13%. Most other categories play a small role over the 1985-2023 period as a whole, although they are prominent in certain episodes, as we show below.

3. Tracking Performance, Predictive Content, and Robustness Checks

3.1 EMV Tracking Performance

Table 1 provides information about how well our EMV measures track stock market volatility from 1985 to 2023. As reported in column (1), regressing monthly-average VIX values on contemporaneous EMV values yields a highly statistically significant slope coefficient of 0.75 and an R-squared value of 0.6. The first two lags of EMV are also statistically significant, and their inclusion raises the R-squared to 0.67. 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. Columns 4 and 5 examine the relationship using daily data. We mimic the specifications in columns 1 and 3, finding again that EMV is highly predictive of VIX. When controlling for one-day lagged VIX in column 5, we continue to find that daily EMV and its lags are highly correlated with contemporaneous VIX. Columns 6 to 8 show that log-log specifications and regressions of realized stock market volatility on EMV yield similar results.

[12] We report percentages for all categories and subcategories in Appendix Table A.1. The column entries sum to more than 100 percent for two reasons: First, because certain terms appear in the term set for more than one category. Second, because many EMV articles refer to multiple sources of equity market volatility.

Figure 2 plots the VIX and fitted values for the column (1) specification. For the most part, fitted values – and the underlying EMV values – move closely with 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 most evident for the cataclysmic events of September-November 2008.[13]

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 above. 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. Adding lagged VIX to the regression specification largely resolves this type of tracking error as well as tracking errors of type (iii).

3.2 Out-of-Sample Performance Assessments

We finalized the methodology and term sets that underlie our EMV trackers in 2018, and we first published them in a March 2019 NBER working paper. Thus, we can use data from 2019 to 2023 to subject our EMV trackers to clean out-of-sample performance assessments.

Table 2 provides information about the in-sample and out-of-sample performance of our overall EMV tracker. Columns 1 and 2 report univariate regressions of VIX on contemporaneous EMV values at daily and monthly frequencies for the "in-sample" period from January 1985 to December 2018. Similarly, column 3 reports a regression of realized market volatility in the month on the contemporaneous monthly VIX tracker. All three columns reveal a strong in-sample fit between implied or realized stock market volatility and the contemporaneous EMV tracker. In columns 4 to 6, we report corresponding results based on "out-of-sample" data from January 2019 to December 2023. In all cases, we continue to find strong tracking power of our EMV measures

[13] Appendix Figure A.1 displays a comparison of Realized Volatility alongside fitted Realized Volatility, as calculated from a regression of realized volatility on EMV (as reported in Table 1, column 7) during the main sample period of 1985-2023. Appendix Figures A.2 and A.3 display similar comparisons for the 1928-1959 and 1960-1984 periods.

for implied and realized volatility. These results show that EMV tracks stock market volatility out of sample. That's true despite the extraordinary shocks that struck the economy and stock markets from 2019 to 2023. This period involved the deepest contractions in the U.S. economy since the 1930s and enormous stock market gyrations in the United States and around the world (Baker et al., 2020, and Davis, Liu and Sheng, 2022). This strong out-of-sample performance highlights the value of our simple approach to the construction of tracking indexes.

3.3 EMV Tracker Performance at Longer Implied-Volatility Horizons

Table 3 assesses EMV performance in tracking implied stock market volatility at various horizons ranging from one month to ten years. In each column, we regress time-t implied volatility for the indicated horizon on contemporaneous and lagged EMV values in monthly data. Here, "EMV 3 Lag Average" at t is the simple mean of EMV_{t-1}, EMV_{t-2} and EMV_{t-3}. Analogously, "EMV 12 Lag Average" at t is the mean of EMV_{t-1}, …, EMV_{t-12}.

The results in Table 3a show that EMV tracks much of the variation in implied volatility at all horizons. R-squared values exceed 0.69 at horizons up to one year, and they exceed 0.52 at horizons up to five years. As the horizon lengthens, the lagged EMV averages provide more of the explanatory power. These lagged averages are better at capturing the low-frequency EMV movements that are relevant for tracking movements in long-horizon implied volatility.

3.4 Categorical EMV Trackers and Implied Volatility

We also consider the relationship of our categorical EMV trackers to implied volatility. The categorical EMV trackers differ greatly in their time-series features and properties. This fact is apparent by glancing at the categorical trackers displayed below and in the online appendix. Thus, we hypothesize that the categorical trackers differ in how they correlate with implied volatility measures for different horizons. To investigate this hypothesis, we regress each implied-volatility measure on many categorical EMV measures and use a LASSO approach to select the most informative categories. Table 3b displays the post-selection regressions.

There are indeed important differences across VIX horizons in the selected categorical trackers. The EMV tracker for Macro News about the Labor Market helps account for variation in shorter- and longer-horizon VIX measures. Other categorical trackers are useful at only shorter or longer horizons, but not both. For instance, the EMV tracker for Consumer Spending and Sentiment is informative about the one-month VIX but not the one-year or ten-year implied volatility measures. The EMV trackers for Financial Crises and for Macroeconomic News related to Trade are informative for the ten-year VIX but not for the shorter-horizon volatility measures. These results show that our categorical trackers differ in how they relate to implied volatility.

3.5 Correlations with Future Equity Returns

We now investigate whether our EMV trackers contain information about future equity returns. To do so, we regress annualized returns on the S&P 500 index from month t to $t + \tau$, for τ ranging from 3 to 24 months, on lagged values of our overall EMV tracker and selected category-level EMV trackers. Table 4 reports the results. The first row shows that our overall EMV tracker is predictive of future stock returns at all reported horizons. Higher EMV values foreshadow higher returns, which supports the view that EMV captures uncertainty that is priced in the market.

We also report results for two narrower EMV trackers. Each one is positively correlated with future equity returns. However, the EMV tracker for National Security Policy is significantly correlated with future returns only at shorter horizons of three and six months. In contrast, the EMV tracker for Macroeconomic News & Outlook has predictive content for future returns at shorter and longer horizons. In fact, it has stronger predictive content – at all horizons – than our composite EMV tracker. This pattern suggests that news related to the macroeconomic outlook is particularly pertinent to the types of uncertainty that are priced in equity markets.

3.6 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 relative to the NVIX in tracking implied and realized stock market volatility.

We start with monthly data from January 1985 to March 2016, as the VIX is unavailable before 1985 and the NVIX is unavailable after March 2016. EMV correlates with the VIX at 0.78 in this period, which compares to 0.70 for NVIX. The mean absolute monthly difference between EMV and VIX is 2.5 points, as compared to 3.5 points for NVIX. The standard deviation, skewness, and kurtosis of our EMV tracker are much closer to the corresponding VIX statistics (Table A.2).[14] A big reason for EMV's better performance is its reliance on a much larger corpus – the full text of

[14] Figure A.4 shows 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.

eleven major newspapers – as compared to the abstracts and headlines of front-page articles for a single paper that serve as the corpus for the NVIX. In fact, when we rerun specification (1) in Table 1 using an EMV measure based on a single newspaper, the R-squared value drops drastically – by 17 to 38 percentage points, depending on the paper.

Our historical EMV measure also outperforms the NVIX in tracking realized stock market volatility in monthly data from 1928 to 1984. See Figures A.5 and A.6 in the appendix. They show that the NVIX is essentially flat in this period, while our historical EMV tracks much of the variation in realized volatility. Here, we think the superior performance of our EMV tracker reflects the comparatively sparse nature of its "feature set" and our avoidance of terms that, while prominent in particular episodes, do not perform well over long time spans. These aspects of our methodology for constructing EMV trackers yield better "out-of-sample" performance.

3.7 Robustness to Alternative Newspaper Weightings

We now 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 1 using the EMV tracker based on the modified newspaper-level weights. Appendix Table A.3 reports the results. Doubling the weight on the *Wall Street Journal* or the *Miami Herald* yields an incremental R-squared gain of 0.002 to 0.004, respectively. 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 the 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 largely exhaust the 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.

3.8 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 from 2007 to 2023 and at 0.50 with the CBOE Crude Oil Realized Volatility from 1986 to 2023. Inspecting Figure 3 confirms that our measure mirrors many of the movements in oil price volatility, including in the out-of-sample period from 2019 to 2023. 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 3 provides some 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

Figures 4 displays three of our categorical EMV trackers for three categories. Topics covered by the Macroeconomic News and Outlook category appear in 72 percent of EMV articles, and the Macro EMV tracker moves similarly to overall EMV and 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 2011 debt-ceiling crisis – episodes that involved major upsurges in uncertainty about the macroeconomic outlook. Because the terms in our broad Macro category appear in such a large share of EMV articles, we also construct EMV trackers for nine distinct types of news about the Macro outlook. Appendix Figures B.7 and B.8 display two of them, one for Business Investment and Sentiment and another for Consumer Spending and Sentiment.

As an illustration of a more focused category, Panel (b) in Figure 4 displays the Financial Crisis EMV tracker. Three events stand out in the evolution of this EMV tracker: the Global Financial Crisis, the U.S. debt-ceiling crisis of 2011, and the COVID-19 pandemic. 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. Otherwise, the Financial Crisis concerns receive little attention in newspaper articles about market volatility.

4.2 The Role of Policy Matters in Equity Market Volatility

The EMV tracker for Government Spending, Deficits and Debt shown in Figure 4(c) is near zero most of the time from 1985 to 2023, but it jumps sharply during a few political clashes over government spending, especially the U.S. debt-ceiling crisis of 2011. This chart illustrates the power of policy conflicts and their then-uncertain resolution to drive market volatility.

Appendix B displays other policy-related EMV trackers. Several exhibit highly distinctive movements, illustrating how the sources of stock market volatility vary over time. For example, the Trade Policy EMV tracker hovers near zero throughout most of the period from 1985 through 2017 except for notable, but modest, upward moves from 1992 to 1995 (NAFTA) and late 2016 and early 2017 (Donald Trump's surprise election win and the U.S. withdrawal from the Trans-Pacific Partnership Program). It then skyrockets from March 2018 through late 2019 in reaction to the U.S.-China trade war. EMV trackers for Monetary Policy, Tax Policy, Elections and Political Governance, Healthcare Policy, and more also exhibit major upswings during certain episodes.

Figure 5 reveals a large and time-varying 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 "fiscal cliff"), and the period of Donald Trump's election and presidency from November 2016 to January 2021. To construct Figure 5, we sum EMV article counts over each of the policy-related categories listed in Panel B of Table A.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 that underlie the Economic Policy Uncertainty Index, finding alignment with Figure 5.[16]

[15] For Fiscal Policy and Regulation, we use article counts for the more disaggregated categories.
[16] That term set is {regulation, regulations, regulatory, deficit, deficits, legislation, legislative, legislature, white house, federal reserve, the fed, congressional, congress, war, tariff.}

Figure 5 highlights the role of policy concerns in U.S. stock market volatility, especially in the second half of the sample. It resonates with other evidence of an important and often 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 growing scale and complexity of the regulatory system, increasing complexity in the tax code, the growing share of business "risk factors" that U.S. firms attribute to government policy in their 10-K filings, and a secular rise in the Economic Policy Uncertainty Index of Baker, Bloom and Davis (2016). Since these long-term developments show little sign of reversal, policy concerns are likely to remain a major source of stock market volatility.

As suggested by the annotations in Figure 5, the mix of policy-related factors in stock market volatility varies over time. As an illustration, Appendix Figure A.7 displays the percentage 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 and the U.S. withdrawal from the Trans-Pacific Partnership, threats to quit the North American Free Trade Agreement, and tariff hikes on steel, aluminum and other goods.

4.2.1 Policy-Related EMV Compared to Economic Policy Uncertainty

We also compare the policy-related elements of our EMV index 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 quantify policy-related uncertainty for the economy as a whole.

For instance, Financial Regulation receives attention in 25% of EMV articles as compared to 6% of EPU articles.[17] In contrast, National Security, Healthcare Policy, and Entitlement and Welfare Programs are among the policy-related categories that loom larger for the Economic Policy Uncertainty (EPU) index than for our EMV tracker. Reassuringly, policy-related discussions are more common in EPU articles than in EMV articles.

We also construct a Policy-Related EMV tracker that aims to quantify the full range of policy-related volatility sources for the stock market. To obtain our Policy-Related EMV tracker, we multiply the overall EMV tracker in Figure 1 by the policy-related fraction in Figure 5. We then multiplicatively rescale to match the mean EPU value from 1985 to 2009, so that we can readily compare the two series. Appendix Figure A.8 displays the comparison. Stock market crashes and

[17] These breakdowns are enumerated in Panel B of Appendix Table A.1.

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.

4.3 How Big a Role for Animal Spirits?

To assess the role of animal spirits as a source of stock market volatility, we consider an EMV tracker for Consumer Spending and Sentiment (based on articles that contain "consumer spending," "retail sales," "consumer purchases," "consumer confidence" or "consumer sentiment") and one for Business Investment and Sentiment ("business investment," "business inventories," "business sentiment" and "business confidence"). As reported in Table A.1, terms in the consumer category appear in 9.2 percent of EMV articles, while terms in the business category appear in only 1.9 percent. These results reveal modest roles, on average, for consumer and business sentiment as sources of stock market volatility.

However, we find large roles for consumer sentiment as a source of volatility after the dot-com crash, 9-11 attacks, and Gulf War II (Appendix Figure B.8). The consumer sentiment EMV tracker also exhibits notable rises in reaction to the 1987 stock market crash, Great Recession, U.S. debt-ceiling dispute in 2011, and the early stages of the COVID pandemic. The business sentiment EMV tracker mirrors some of these patterns (Appendix Figure B.7) but is generally a small source of stock market volatility except during the early stages of the COVID pandemic.

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

We now combine our category-level EMV trackers with textual analysis of 10-K filings to construct monthly firm-level risk exposure measures. These measures help explain the firm-level structure of return volatilities and its evolution over time. Our category-level exposures derived from 10-K filings also help explain the cross-sectional structure of firm-level return correlations.

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

In 2005, the U.S. Securities and Exchange Commission (SEC) issued a regulation that requires most listed firms to discuss their "Risk Factors" in Part 1A of their 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 Part 1A to quantify firm-level policy risk exposures, which they combine with their EPU index to explain firm-level stock price volatilities, investment rates, and employment growth rates in a panel regression setting. Davis, Hansen and Seminario (2021) use Part 1A to explain the heterogeneity in firm-level reactions to macroeconomic shocks and to help predict firm-level investment and employment responses to those shocks. Our approach is similar in spirit to the ones in these earlier studies, but it differs greatly in the details.

We work with 10-K reports issued in calendar years 2006 to 2019, which typically pertain to fiscal years 2005 to 2018. Specifically, we count the number of sentences in Part 1A that pertains to each of our EMV categories. After obtaining this count for each firm-year observation, we divide by the total number of sentences in Part 1A for the same firm and year:

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

where i indexes firms, y indexes calendar years, and b designates an EMV category.[18] This expression quantifies each firm's self-reported exposure to each category.

5.2 Combining EMV Trackers with 10-Ks to Explain Firm-Level Return Volatilities

If the $F_{i,y}^{b}$ measures accurately capture category-level exposures at the firm level, and if our EMV trackers contain market-relevant information, the stock price volatility of firm i will be more responsive to the EMV tracker for category b when $F_{i,y}^{b}$ is larger. We test this hypothesis in a panel regression setting using the following type of specification:

$$\sigma_{i,t} = \alpha_i + \gamma_t + \beta \sum_{b} F_{i,y}^{b} EMV_t^{b} + \epsilon_{i,t} \tag{1}$$

where t denotes the monthly time period, $\sigma_{i,t}$ is the realized volatility of firm i in month t (constructed from daily firm-level returns), and α_i and γ_t are firm and time fixed effects.[19] The summation term on the right side of (1) weights each category-level EMV tracker by the corresponding $F_{i,y}^{b}$ value for the firm and year in question. When we apply these weights, we use

[18] We drop filings for which the automated sentence counter returns a value of less than nine for Part 1A. These cases typically contain routine headings and section separators of 10-K filings with an otherwise 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. See Appendix Table D.1 for summary statistics.

[19] Appendix D explains how we calculate the firm-level monthly realized volatilities and provides summary stats.

the firm's most recent Part 1A. For example, suppose the firm submits a 10-K filing in March 2013 (and again in March 2014). Then we apply its $F^{b}_{i,2013}$ values to the EMV^{b}_{t} trackers from t = April 2013 to t = March 2014. In this way, we ensure that the weights used to compute each composite index value pertain to category exposures reported by the firm in a prior period.

Table 5 presents results for our realized firm-level volatility regressions. Column (1) considers the baseline specification (1). It shows that the composite exposure measure is highly statistically significant in explaining the firm-level return volatilities, conditional on firm and time fixed effects. This result strongly supports the basic hypothesis we set out to assess.

The other columns in Table 5 unpack this result. Columns (2) to (4) isolate policy and non-policy sources of the firm-level risk exposures. Columns (2) and (3) show that each source is statistically significant when considered separately. Column (4) shows that the non-policy composite carries most of the weight when we include both composites in the regression. Column (5) reinforces and refines this result. To obtain the column (5) specification, we start with a LASSO regression that considers 38 separate exposure measures, one for each of our category-level EMV trackers.[20] Column (5) then reports an OLS regression on the LASSO-selected exposure measures. The three selected categories – Interest Rates, Real Estate Markets, and Commodity Markets – each pertain to non-policy categories.

To summarize, we combine our category-level EMV trackers with firm-level 10-K filings to build a composite firm-level risk exposure measure that varies over time at the monthly frequency. This composite measure helps explain the cross-sectional structure of firm-level return volatilities and its evolution over time in a parsimonious regression model that conditions on firm and time fixed effects. Our efforts to unpack this result suggest that news about Interest Rates, Real Estate Markets, and Commodity Markets are the most important sources of explained variation in the cross-sectional structure of firm-level return volatilities.

5.3 Using EMV Trackers to Explain the Cross-Sectional Structure of Returns

Lastly, we consider whether our EMV trackers help explain co-movements in equity returns across firms. To do so, we first assign each firm-month observations to a "leading" EMV category Specifically, we assign the firm to the category that accounts for the largest share of sentences in

[20] We cluster standard errors at the firm level and include firm and time fixed effects, so that our LASSO selection procedure corresponds directly to our OLS specification. Appendix Table D.2 reports summary statistics for the regression variables.

Part 1A of its same-year 10-K filing.[21] A firm's leading category can change from one year to the next as the emphasis in Part 1A of its 10-K filing changes. Second, for all firms assigned to the same leading category in month *t,* we compute the daily pairwise return correlations in that category and month. Third, we average these daily pairwise return correlations and assign that average value to the month-t observation for each firm in the category. Finally, we regress these firm-level average pairwise correlations on the natural log of the contemporaneous $EMV_t^{b=l}$ value, where $b=l$ refers to the leading EMV category for the firm-month observation.

The first column in Table 6 reports results for this regression when controlling for firm fixed effects, and Figure A.10 displays the corresponding bin scatter. The coefficient on $\ln(EMV_t^{b=l})$ is positive and highly statistically significant. Thus, the returns of firms assigned to the same leading category comove more strongly when the EMV tracker for that category is higher. According to column (1), doubling $\ln(EMV_t^{b=l})$ raises the average pairwise correlation of firms with the same leading category by 4.2 percentage points. That amounts to one-fifth of the dependent variable's sample mean value, a large effect.

The other columns in Table 6 consider other controls and sample splits. Column (2) controls for the contemporaneous VIX value, because market-wide volatility comoves with stock returns. Including the VIX shrinks the main coefficient of interest by more than half, but it remains highly statistically significant. Including lags of $\ln(EMV_t^{b=l})$ does not materially alter the main result. Adding time fixed effects further shrinks the main coefficient of interest, but it remains highly statistically significant. Finally, in columns (6) and (7), we split the sample into observations with a non-policy leading category and those with a policy leading category. The main result of interest continues to hold in each sub-sample, even with controls for firm and time fixed effects.

These results show that our EMV trackers help explain variation over time in the cross-sectional correlation structure of firm-level stock returns. At the same time, our results only scratch the surface of what is likely a much richer story of how and why the correlation structure of firm-level returns varies over time as certain risk categories become more or less salient. We hope that future research can build on this and other aspects of our analysis.

[21] We exclude compound categories such as 'Regulation' and 'Macro News' when making these assignments.

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. We extend the approach to encompass historical data back to 1928 and to construct a daily EMV tracker, finding similarly close relationships to implied and realized equity market volatility. Out-of-sample tests performed on data generated after we first developed our methods, proposed our EMV tracker, and circulated results confirm the utility of our approach.

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 large, time-varying roles for policy and non-policy developments as sources of stock market volatility. Monetary Policy and Tax Policy are the most important policy-related sources of stock market volatility, according to our analysis, followed by our aggregated Regulation category. The contribution of specific policy categories to stock market volatility fluctuates markedly over time. For example, National Security and Trade Policy matters contribute modestly to stock market volatility during much of our sample but occasionally emerge as major contributors to market volatility.

We also use our category-level EMV trackers in combination with self-reported risk exposures in 10-K filings to explain and interpret firm-level stock price volatilities and their movements over time. Finally, we show that our category-level EMV trackers help explain changes over time in the correlation structure of firm-level stock returns.

There are natural directions for future research. One is to extend our measurement approach to other countries and periods with digital newspaper archives and data on 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. Our basic approach could also be used 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 business sentiment and to delve into the forces that drive their movements.

References

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, no. 1, 38-62.

Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims, 2013. "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5, no. 2 (March), 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, Steven J. Davis and Kyle Kost, 2019. "Policy News and Stock Market Volatility," NBER Working Paper 25720.

Baker, Scott R., Nick Bloom, Steven J. Davis, Kyle Kost, Marco Sammon and Tasaneeya Viratyosin, 2020. "The Unprecedented Stock Market Reaction to COVID-19," *Review of Asset Pricing Studies,*10, no. 5, 742–758.

Baker, Scott R, Nicholas Bloom, Steve J. Davis and Marco Sammon, 2024. "What triggers stock market jumps?" NBER Working Paper 28687, revised.

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," in B. Douglas Bernheim, Stefano DellaVigna and David 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, 2020. "Uncertainty Shocks as Second-Moment News Shocks," *Review of Economic Studies,* 87, no. 1 (January), 40-76.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, Matthew Richardson, 2019. "Information, Trading, and Volatility: Evidence from Firm-Specific News, *Review of Financial Studies*, 32, no. 3 (March), 992–1033

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.

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.

Campbell, J. Y., M. Lettau, B. G. Malkiel and Y. Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic risk," *Journal of Finance*, 56, 1-43.

Chen, Qihui, Nikolai Roussanov and Xiaoliang Wang, 2023. "Semiparametric Conditional Factor Models: Estimation and Inference," NBER Working Paper 31817

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.

Cherkassky, Vladimir and Yunqian Ma, 2004. "Comparison of loss functions for linear regression," 2004 IEEE International Joint Conference on Neural Networks.

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., Dingqian Liu and Xuguang Simon Sheng, 2022. "Stock Prices and Economic Activity in the Time of Coronavirus," *IMF Economic Review,* 70, 32-67.

Davis, Steven J., Stephen Hansen and Cristhian Seminario, 2021. "Firm-Level Risk Exposures and Stock Return in the Wake of COVID-19," NBER Working Paper 27867.

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.

Fama, Eugene F. and Kenneth R. French, 1993. "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

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.

Freyberger, Joachim, Andreas Neuhierl and Michael Weber, 2020. "Dissecting Characteristics Nonparametrically," *Review of Financial Studies*, 33, no. 5, 2326-2377.

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.

Gagliardini, Patrick, Elisa Ossola and Olivier Scaillet, 2016. "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," Econometrica, 84, no. 3 (May), 985–1046.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy, 2019. "Text as data," *Journal of Economic Literature*, 57, no. 3 (September), 535-574..

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

Griffin, John M., Nicholas H. Hirschey and Patrick J. Kelly, 2011. "How Important Is the Financial Media in Global Markets?" *Review of Financial Studies,* 24, no. 12 (December), 3941–92.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016. "… and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29, no. 1 (January), 5–68.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent and Ahmed Tahoun, 2019. "Firm-Level Political Risk: Measurement and Effects," *Quarterly Journal of Economics*, 143, no. 4 (November), 2135-2202.

Jurado, Kyle, Sydney Ludvigson, and Serena Ng, 2015. "Measuring Uncertainty," *American Economic Review*, 105, no. 3, 1177–1216.

Kelly, Bryan, Asaf Manela, and Alan Moreira, 2021. "Text Selection," *Journal of Business and Economic Statistics,* 39, no. 4, 859-879.

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.

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, 2013. "Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence," Federal Reserve Bank of Kansas City Economic Policy Symposium.

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: Regressions of Stock Market Volatility Measures on the EMV Tracker

	(1) Monthly VIX_t	(2) Monthly VIX_t	(3) Monthly VIX_t	(4) Daily VIX_t	(5) Daily VIX_t	(6) Log(Monthly VIX_t)	(7) Monthly $RVol_t$	(8) Monthly $RVol_t$
EMV_t	0.745***	0.554***	0.496***	0.149***	0.0229**		0.971***	0.820***
	(0.0533)	(0.0822)	(0.0742)	(0.0117)	(0.0107)		(0.0935)	(0.128)
EMV_{t-1}		0.191**	-0.151*		-0.0129***			
		(0.08)	(0.0892)		(0.0046)			
EMV_{t-2}		0.156***	-0.0346		-0.00427			
		(0.058)	(0.063)		(0.0035)			
VIX_{t-1}			0.674***		0.960***			
			(0.0702)		(0.0108)			
Log(EMV_t)						0.765***		
						(0.0406)		
$RVol_{t-1}$								0.220*
								(0.120)
R^2	0.603	0.67	0.824	0.215	0.938	0.575	0.627	0.661
Observations	468	466	466	9,617	9,615	468	468	467

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using daily (columns 1-2) or monthly (columns 3-8) data from January 1985 to December 2023. EMV is daily or monthly Equity Market Volatility tracker developed in Section 2.1. Monthly 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. (2020) in earlier years. Daily VIX is extended backwards to 1985 using the VXO. RVol is the standard deviation of daily returns on the S&P500 in the month. Robust standard errors in parentheses.

Table 2: In- and Out-of-Sample Assessments of the Overall EMV Tracker

	In-Sample (1985-2018)			Out-of-Sample (2019-2023)		
	(1) VIX - Daily	(2) VIX - Monthly	(3) Realized Vol - Monthly	(4) VIX - Daily	(5) VIX – Monthly	(6) Realized Vol - Monthly
EMV_t	0.144***	0.752***	0.955***	0.180***	0.714***	1.14***
	(0.0031)	(0.030)	(0.089)	(0.00877)	(0.0835)	(0.343)
Obs. Count	8,317	408	408	1,300	60	60
R^2	0.210	0.606	0.649	0.245	0.558	0.543

Notes: Each column reports a regression of the indicated dependent variable on contemporaneous values of our Equity Market Volatility Tracker, using data for the indicated period. Daily VIX is extended backwards to 1985 using the VXO. Heteroskedasticity robust standard errors in parentheses.

Table 3a: Regressions of VIX, for Various Horizons, on Contemporaneous and Lagged EMV Trackers

	(1) VIX_t	(2) VIX_t	(3) VIX_t	(4) VIX_t	(5) VIX_t	(6) VIX_t	(7) VIX_t
VIX Horizon →	1 Month	3 Month	6 Month	1 Year	3 Year	5 Year	10 Year
EMV_t	0.457***	0.321***	0.227***	0.160***	0.0971**	0.0779*	0.0598
	(0.0492)	(0.0349)	(0.0274)	(0.0204)	(0.0432)	(0.044)	(0.0411)
EMV 3 Lag Avg.	0.292**	0.277***	0.228***	0.173**	0.299***	0.252***	0.159*
	(0.122)	(0.099)	(0.082)	(0.0739)	(0.0756)	(0.0762)	(0.0819)
EMV 12 Lag Avg.	0.321***	0.381***	0.400***	0.410***	0.419***	0.372***	0.248**
	(0.0966)	(0.0796)	(0.0725)	(0.0672)	(0.108)	(0.107)	(0.105)
R^2	0.728	0.746	0.721	0.691	0.607	0.534	0.334
Observation Count	314	314	314	314	165	165	165

Notes: We compute VIX_t as the average of daily VIX values in month *t* for the indicated VIX horizon. Each column reports a regression of the column variable on the indicated row variables. We fit the regressions to monthly data from January 1996 to February 2023 in columns 1-4 and from November 2002 to July 2016 in columns 5-7 (for reasons of data availability). EMV_t denotes the Equity Market Volatility tracker for month t. See Section for an explanation of how we construct this measure. "EMV 3 Lag Avg." at *t* is the simple mean of EMV_{t-1}, EMV_{t-2} and EMV_{t-3}. Analogously, "EMV 12 Lag Avg." at *t* is the mean of EMV_{t-1}, …, EMV_{t-12}. We report Newey-West standard errors in parentheses with maximum autocorrelation lag of 2.

Table 3b: Regressions of VIX, for Various Horizons, on Category-Level EMV Trackers

	(1) VIX_t	(2) VIX_t	(3) VIX_t
VIX Horizon →	1 Month	1 Year	10 Year
EMV_t Tracker for …			
Macro News: Labor Markets	0.862***	0.609***	0.594**
	(0.156)	(0.095)	(0.255)
Macro News: Consumer	1.195***		
Spending and Sentiment	(0.32)		
Commodity Markets	0.212		
	(0.143)		
Financial Regulation	1.560***	1.188***	
	(0.331)	(0.271)	
Competition Policy		1.671***	
		(0.476)	
Macro News: Trade			-6.144***
			(1.316)
Financial Crises			0.383*
			(0.201)
R^2	0.61	0.472	0.414
Observation Count	326	326	165

Notes: Each column reports a regression of the column variable on the indicated row variables, using monthly data from January 1996 to February 2023. See notes to Table 3a for definitions of the dependent variables. Each row variable is a category-level EMV tracker, as defined in Section 2.3. We report Newey-West standard errors in parentheses with maximum autocorrelation lag of 2.

Table 4: Predicting Stock Market Returns Using Equity Market Volatility Trackers

	(2) $r(t\to t+\tau)$	(3) $r(t\to t+\tau)$	(4) $r(t\to t+\tau)$	(5) $r(t\to t+\tau)$
S&P 500 Returns Horizon →	3 Months	6 Months	1 Year	2 Years
EMV_{t-1}	0.0857*	0.0590**	0.0470**	0.0298**
	(0.0439)	(0.0298)	(0.0216)	(0.0129)
R^2	0.0135	0.0117	0.0129	0.00909
Observation Count	431	431	431	431
EMV_{t-1} for Macroeconomic News	0.215***	0.157***	0.114**	0.109***
& Outlook	(0.0787)	(0.0547)	(0.0482)	(0.0259)
R^2	0.0195	0.0189	0.0175	0.0275
Observation Count	431	431	431	431
EMV_{t-1} for National Security Policy	0.334**	0.191*	0.0945	-0.0279
	(0.141)	(0.0987)	(0.112)	(0.0531)
R^2	0.0104	0.00616	0.00264	0.000402
Observation Count	431	431	431	431

Notes: Each cell reports results for a separate regression of the column variable on the indicated row variable, using monthly data from January 1985 to December 2023. EMV_{t-1} denotes the overall Equity Market Volatility tracker or the indicated category-level Equity Market Volatility Tracker, as described in Sections 2.1 and 2.3. Returns are the annualized total return of the S&P 500 index over the indicated horizon. Newey-West standard errors with maximum autocorrelation lag equal to the indicated returns horizon.

Table 5: Regressions of Firm-Level Realized Volatilities on Composite Exposure Indexes

Composite Index Based on:	(1)	(2)	(3)	(4)	(5)
All EMV Categories	2.16*** (0.22)				
Non-Policy Categories		2.50*** (0.25)		2.46*** (0.25)	
Policy Categories			1.35*** (0.48)	0.83* (0.49)	
Macro: Interest Rates					-9.33*** (1.01)
Macro: Real Estate Markets					7.90*** (0.81)
Commodity Markets					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

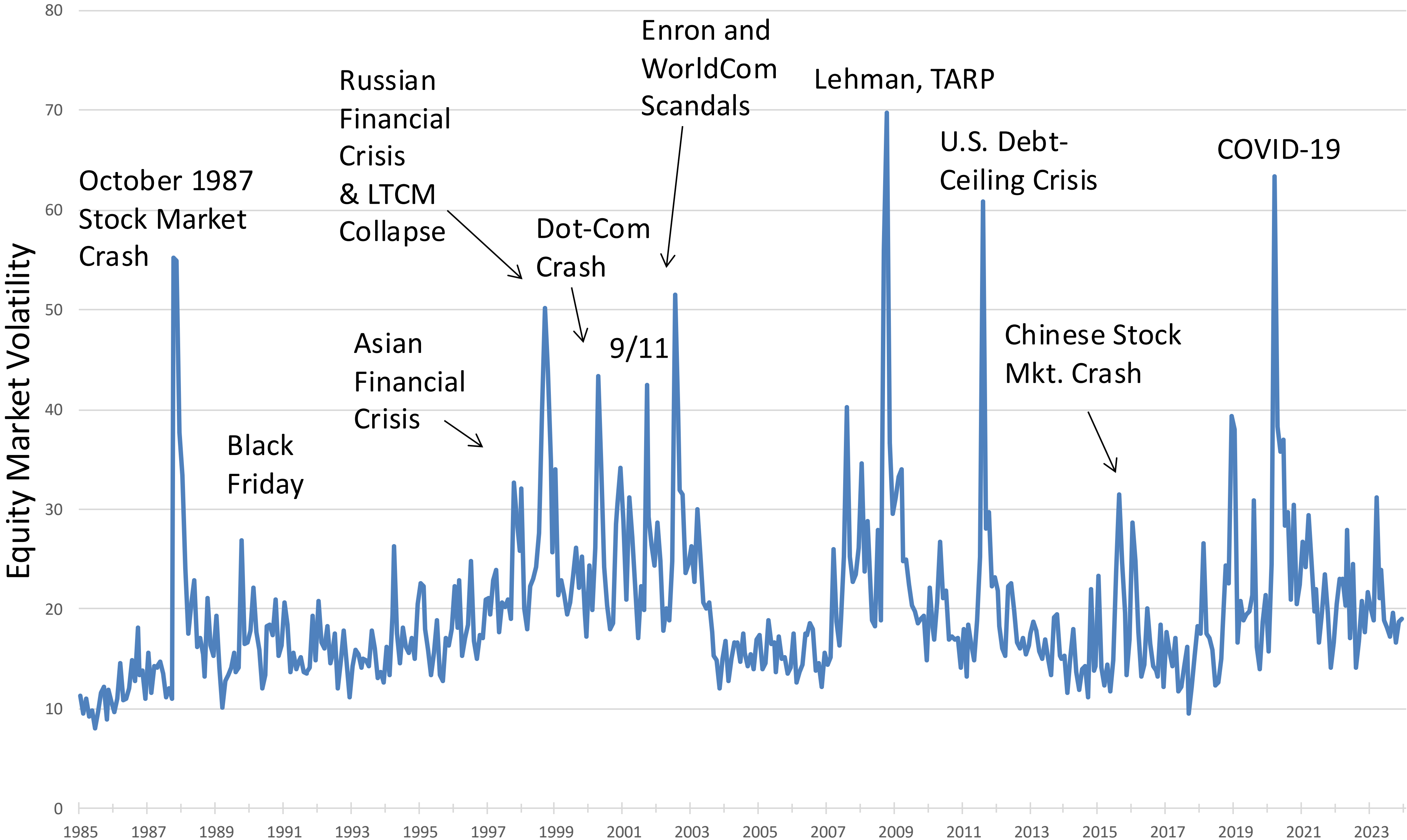
Notes: Each column corresponds to a regression of $\sigma_{i,t}$ (the realized volatility of daily equity returns for firm *i* in month *t*) on the indicated composite exposure index or indexes. We winsorize realized volatility at the 1% and 99% levels. Section 5.2 in the text describes the construction of the composite indexes. All specifications include a full set of firm and time fixed effects. The sample runs from 2006 to 2019. In fitting these regressions, we weight each observation by the lagged value of the firm's log market capitalization times the square root of the number of sentences in Part 1A of its current 10-K filing. This approach places more weight on firms with greater market values and firms with a more informative Part 1A in its 10-K filing. Column (5) reports the OLS regression results using LASSO-selected composite indexes. We cluster errors at the firm level when computing the standard errors reported in parentheses. $p < 0.01$ ***, $p < 0.05$ **, $p < 0.10$ *

Table 6: Regressions of Average Pairwise Return Correlations on Contemporaneous EMV Values of the Firms' Leading Category

	Average Pairwise Correlation of Daily Returns in Month *t* Among Firms with the Same Leading EMV Category for:						
	(1) All Firms	(2) All Firms	(3) All Firms	(4) All Firms	(5) All Firms	(6) Firm-Months with a Non-Policy Leading Category	(7) Firm-Months with a Policy Leading Category
$\ln(EMV_t^{b=l})$	4.24***	1.70***	3.90***	1.11***	0.871***	0.946***	0.718***
	(0.020)	(0.018)	(0.032)	(0.012)	(0.021)	(0.016)	(0.026)
$\ln(EMV_{t-1}^{b=l})$			0.171***		0.178***		
			(0.037)		(0.024)		
$\ln(EMV_{t-2}^{b=l})$			0.367***		0.116***		
			(0.033)		(0.022)		
VIX_t		0.608***					
		(0.015)					
Observations	407,479	407,479	390,917	407,479	390,917	295,874	111,576
R-squared	0.226	0.458	0.224	0.800	0.799	0.815	0.851
Time Fixed Effects	NO	NO	NO	YES	YES	YES	YES

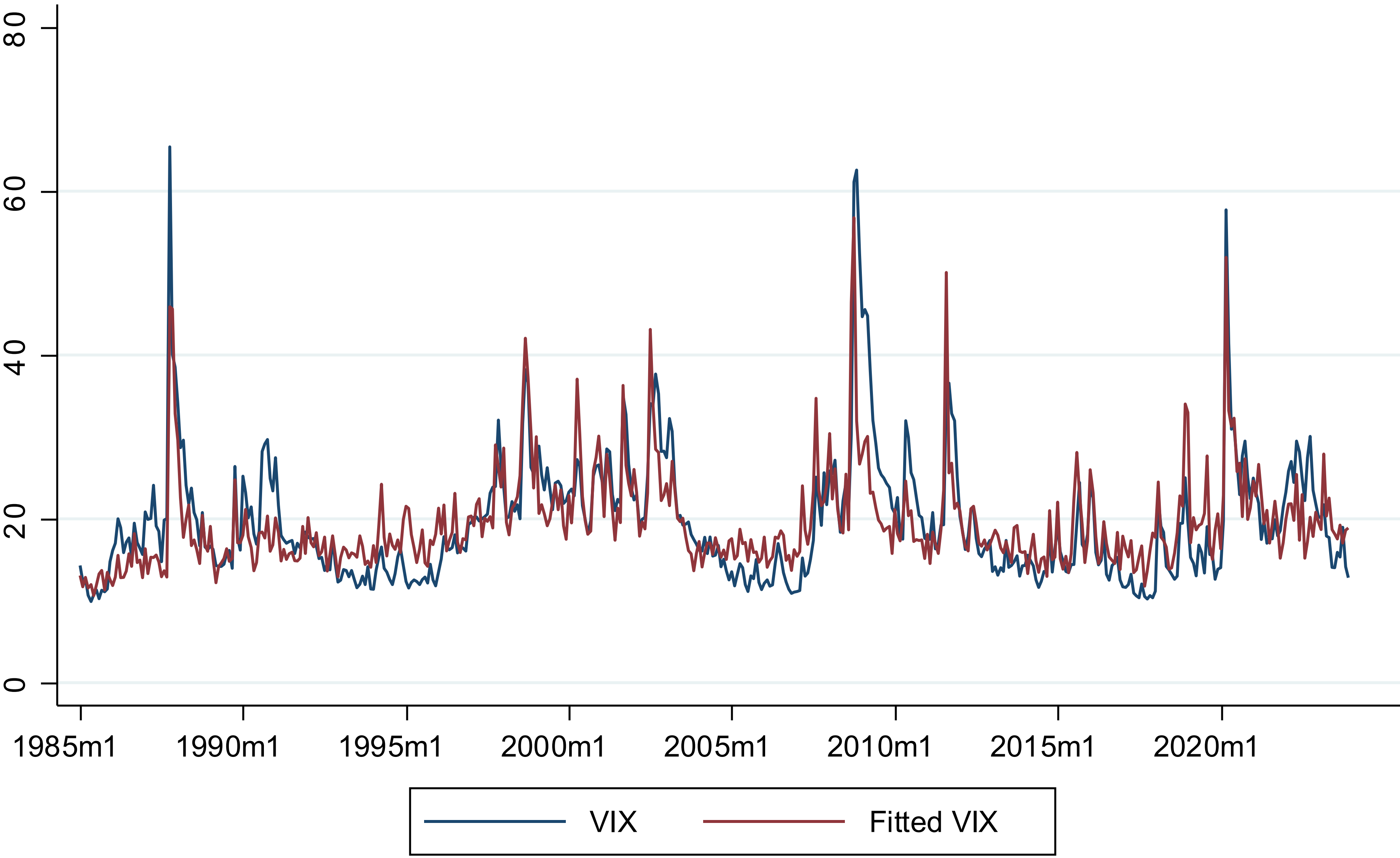
Notes: Each column reports a regression of the average month-*t* pairwise correlation of daily stock returns for firms assigned to the same leading EMV category on the indicated row variables. A firm's leading EMV category is the one most often discussed in Part 1A of its same-year 10-K filing. $EMV_t^{b=l}$ is the EMV tracker value in month *t* for *b=l,* where *l* denotes the leading category. All specifications include firm fixed effects. VIX_t is the mean of daily VIX values during month *t.* The sample in columns (1) to (5) covers all firms. Columns (6) and (7) restrict the sample to observations for which the leading category pertains to non-policy or policy matters, respectively. The mean (median) value of the dependent variable is 0.21 (0.19). We multiply all estimated coefficients by 100 for easier readability.

Figure 1: Newspaper-Based Equity Market Volatility Tracker, 1985-2023



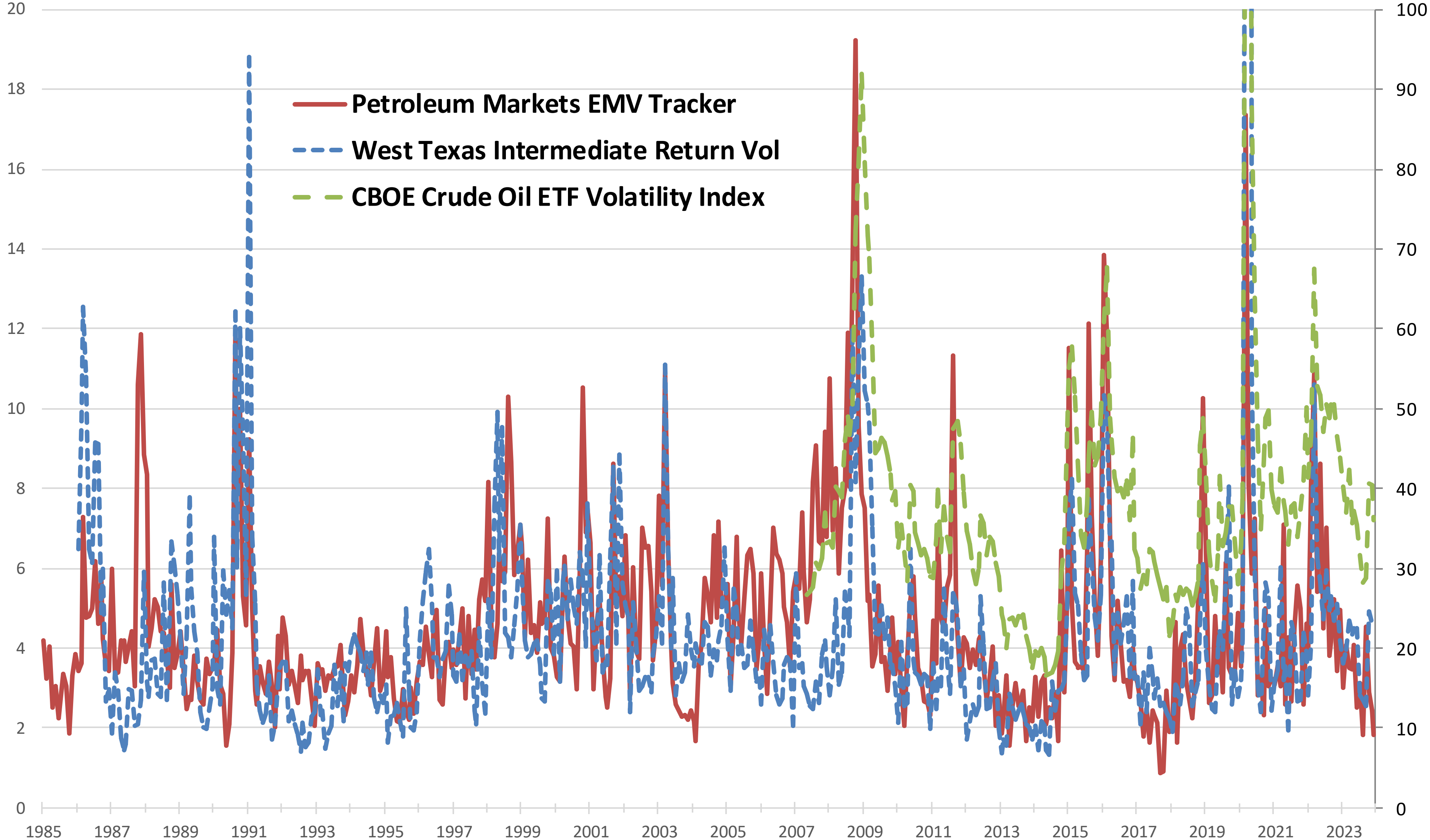
Notes: The Equity Market volatility (EMV) tracker runs from January 1985 to December 2023. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in 11 leading U.S. newspapers, as detailed in Section 2.1. We scale the EMV tracker to match the mean value of the VIX from 1985 to 2015.

Figure 2: VIX and Fitted VIX from a Regression on EMV, 1985-2023



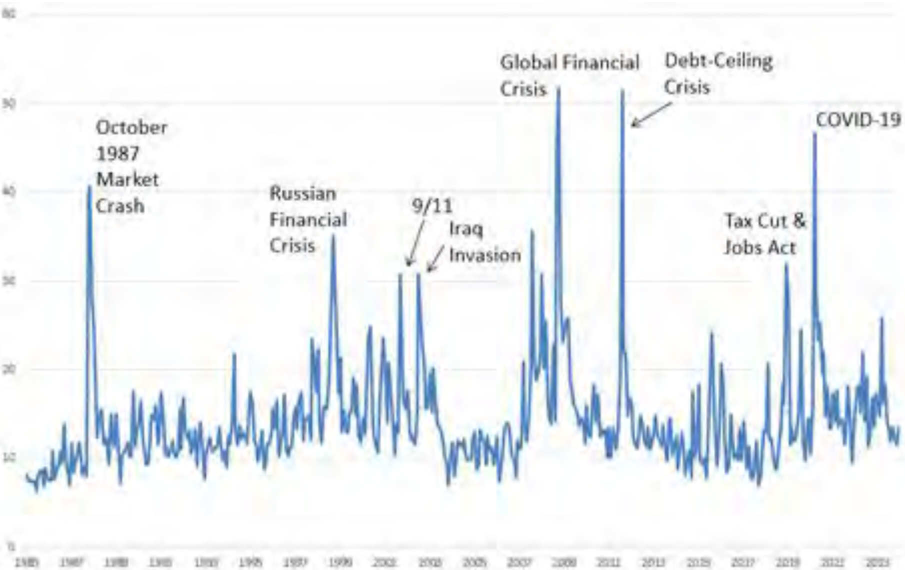
Notes: Data for the CBOE 30-Day VIX data from 1990 to 2023 appended to the VIX series in Berger et al. (2019) from 1985 to 1989. "Fitted VIX" values are from the regression VIX on EMV reported in Table 2, column (1) Both series run from January 1985 to December 2023.

Figure 3: Petroleum Markets EMV and Oil Price Volatility, 1985-2023

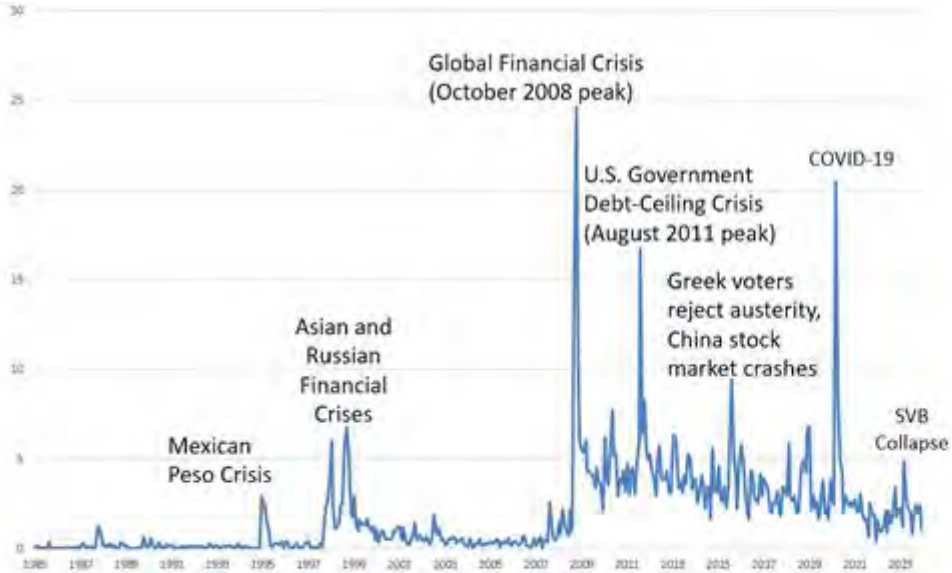


Notes: CBOE Crude Oil Volatility Index is the monthly mean of daily CBOE Crude Oil ETF Volatility Index values. Crude Oil Realized Volatility reflects daily price data for West Texas Intermediate. We extract both series from the St. Louis Federal Reserve FRED database. The Petroleum Markets EMV tracker is constructed from scaled frequency counts of newspaper articles. See Sections 2.1 and 3.6 in the text for details.

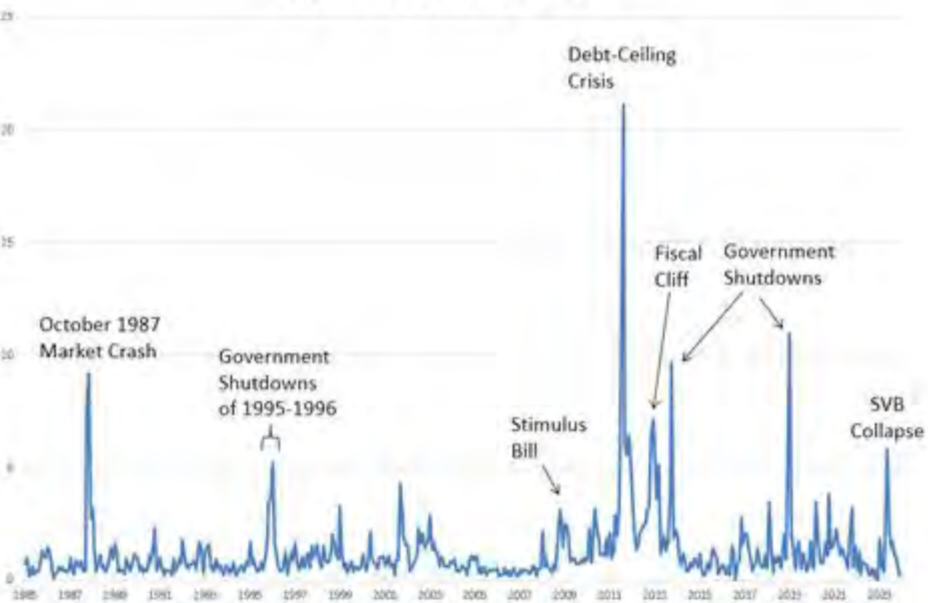
Figure 4: Categorical EMV Trackers, 1985-2023



(a) Macroeconomic EMV



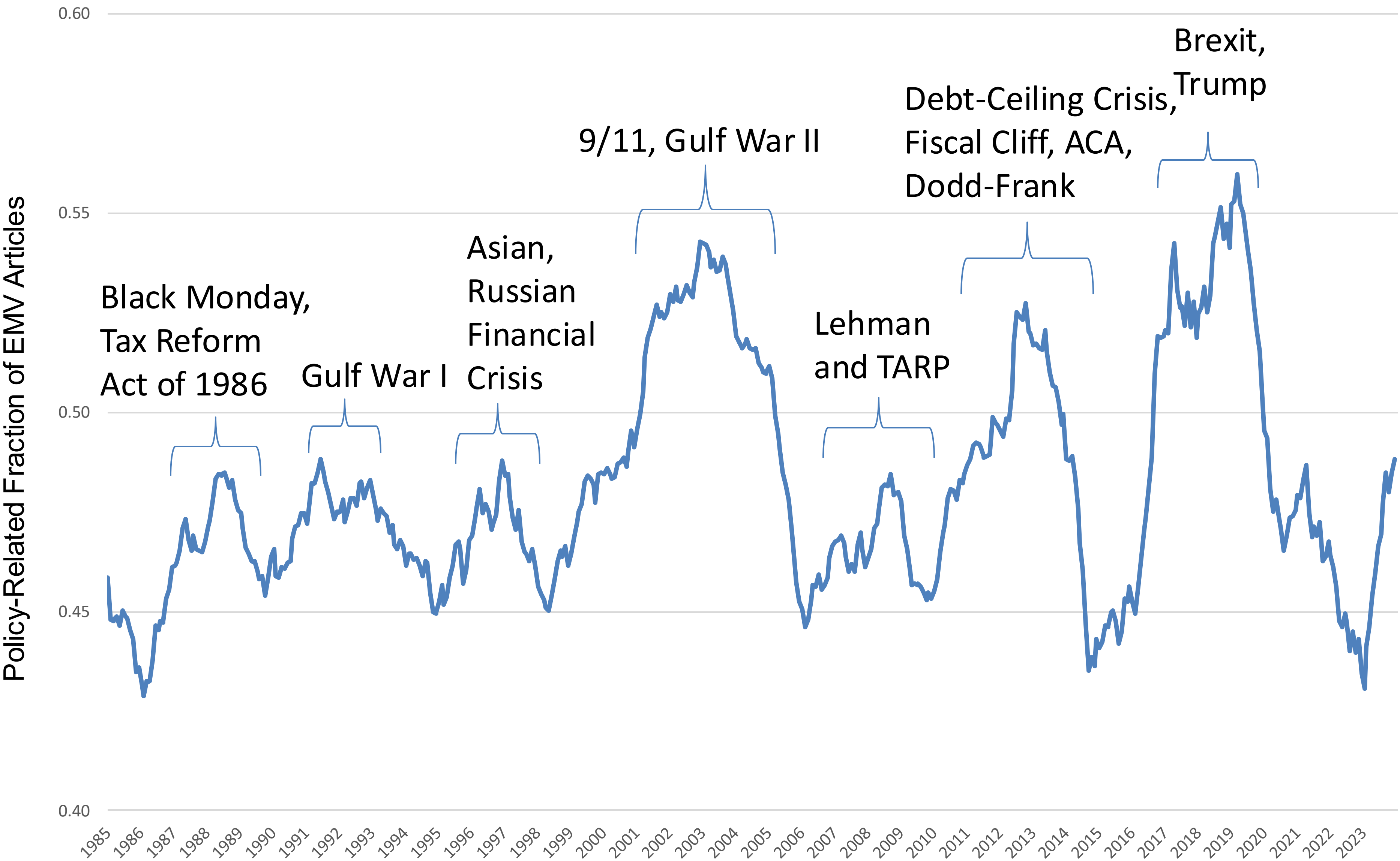
(b) Financial Crisis EMV



(c) Government Spending, Deficits and Debt EMV

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 **(a) Macroeconomic News and Outlook, (b) Financial Crises, and (c) Government Spending, Deficits and Debt.** See Appendix B for the list of terms.

Figure 5: Fraction of EMV Articles that Discuss Policy Matters, 12-Month Moving Average, 1985-2023



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

Appendix A. Additional Analysis and Results

Table A.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).

Table A.4 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 A.4: 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.

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. We pursue this line of analysis in more detail in Table A.5. 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 4, as the horizon of implied volatility lengthens, the explanatory power of contemporaneous EMV is reduced substantially.

Figure A.1 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.

Incorporating Firm Characteristics into 10-K Analysis

Table A.6 reports results of our realized volatility on EMV Topics Composite regressions while also incorporating various firm characteristics and their interactions with our EMV indices. Characteristics like the cyclicality of sales growth, rates of investment, fixed asset intensities may drive some of the relationship between firm-level volatility and the exposure to categorical EMV. While many of these relationships will be reflected in the topic loadings derived from firm 10-K reports, we can also explicitly measure a richer set of observable factors.

Columns (1)-(5) report the same specifications from Table 5 but now using the subsample of observations that have data on 62 firm characteristics. These 62 characteristics are borrowed from Freyberger et al. (2019). The results are comparable to Table 5 with the only notable difference being that the Macro – Real Estate Markets sub-index is not selected. 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 selected with the additional granularity in 10-K EMV topics adding to the number of EMV-related variables selected. We see an increase in the within R^2 when adding the firm characteristic interactions from 0.0039 in column (5) of Table 5 to 0.146 in column (8) of Table A.6. 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 A.6 fits the time series movements in the cross-sectional standard deviation of firm-level realized volatility. Regressing the actual cross-sectional standard deviation on the predicted analog, we get an R^2 value of 0.82.[1]

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, finding a median coefficient around 0.34.

[1] Appendix Figure A.9 displays the time series plot of these two series.

Table A.1: Percent of EMV Articles in Each Category, 1985-2023

Panel A. General Economic Categories	**Percent of EMV Articles**	
Macroeconomic News and Outlook	71.8	
Broad Quantity Indicators	27.3	
Inflation	28.8	
Interest Rates	30.7	
Other Financial Indicators	3.4	
Labor Markets	23.5	
Real Estate Markets	30.2	
Trade	2.3	
Business Investment and Sentiment	1.9	
Consumer Spending and Sentiment	9.2	
Commodity Markets	43.5	
Financial Crises	8.5	
Exchange Rates	1.9	
Healthcare Matters	6.6	
Litigation Matters	4.8	
Competition Matters	3.7	
Labor Disputes	3.9	
Intellectual Property Matters	3.2	
Panel B. Policy-Related Categories	**Percent of EMV Articles**	**Percent of EPU Articles**
Fiscal Policy:	34.7	44.6
Taxes	29.5	36.1
Government Spending, Deficits, and Debt	6.2	15.3
Entitlement and Welfare Programs	7.3	12.0
Monetary Policy	29.8	34.9
Regulation (generic + 4 big regulation categories)	24.9	27.1
Financial Regulation	14.3	6.3
Competition Policy	2.3	1.1
Intellectual Property Policy	0.1	0.3
Labor Regulations	2.0	3.3
Immigration	0.3	1.5
Energy and Environmental Regulation	1.3	5.5
Lawsuit and Tort Reform, Supreme Court	1.5	4.2
Housing and Land Management	1.2	1.5
Other Regulation	1.0	1.7
National Security Policy	13.6	28.6
Government-Sponsored Enterprises (e.g., Fannie Mae)	4.7	2.7
Trade Policy	3.2	6.0
Healthcare Policy	3.7	8.5
Food and Drug Policy	1.3	1.0
Transportation, Infrastructure, and Public Utilities	1.3	2.6
Elections and Political Governance	3.1	8.2
Agricultural Policy	0.2	0.6

Notes: 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 meeting the criteria of Baker, Bloom and Davis (2016) for policy-related economic uncertainty.

Table A.2: Summary Statistics for the VIX, Realized Volatility, EMV and NVIX

A. January 1985 to March 2016

	RVol	VIX	EMV	NVIX
Standard Deviation	9.58	7.81	8.12	4.83
Skewness	3.67	2.19	2.40	1.27
Kurtosis	24.30	10.74	11.38	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			2.47	3.50
Pairwise Correlation with RVol			0.80	0.65
Pairwise Correlation with RVol in 1st Differences			0.66	0.49
Mean Absolute Distance from RVol			4.08	4.21

B. January 1960 to December 1984

	RVol	EMV	NVIX
Standard Deviation	5.18	5.85	1.38
Skewness	1.42	0.93	0.74
Kurtosis	5.88	3.52	3.90
Pairwise Correlation with RVol		0.46	-0.02
Pairwise Correlation with RVol in 1st Differences		0.36	0.17
Mean Absolute Distance from RVol		3.19	3.10

C. January 1928 to December 1959

	RVol	EMV	NVIX
Standard Deviation	13.80	9.79	2.35
Skewness	1.93	0.84	0.15
Kurtosis	6.75	3.77	3.09
Pairwise Correlation with RVol		0.55	0.56
Pairwise Correlation with RVol in 1st Differences		0.14	0.16
Mean Absolute Distance from RVol		7.52	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 A.3: Fit Sensitivity to Alternative Newspaper Weightings in Regressions of VIX on EMV, 1985-2023

	(1) Baseline	(2) Dallas MN	(3) Houston Chron	(4) Miami	(5) SF Chronicle	(6) USA Tod	(7) Bost Glo	(8) Chicago Trib	(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
Obs.	468	468	468	468	468	468	468	468	468	468	468	468
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.	468	468	468	468	468	468	468	468	468	468	468	468
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.	468	468	468	468	468	468	468	468	468	468	468	468

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.

Table A.4: 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)
$NVIX_t$		0.91		0.43	0.49	0.26
		(0.10)		(0.09)	(0.10)	(0.05)
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	373	373	374	373

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 A.5: 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	0.281***	0.199***	0.142***	0.0974***	0.0672***	0.0475***
	(0.0446)	(0.0341)	(0.0226)	(0.0133)	(0.0128)	(0.0100)
EMV_{t-1}	-0.0248	-0.00919	-0.000863	0.00412	0.0501*	0.0259
	(0.0502)	(0.0376)	(0.0287)	(0.0215)	(0.0272)	(0.0189)
EMV_{t-2}	-0.0515**	-0.0374*	-0.0264	-0.0161	-0.00310	-0.0141
	(0.0250)	(0.0200)	(0.0165)	(0.0126)	(0.0115)	(0.0113)
VIX_{t-1}	0.780***	0.822***	0.856***	0.882***	0.894***	0.930***
	(0.0320)	(0.0261)	(0.0219)	(0.0189)	(0.0247)	(0.0249)
R^2	0.899	0.913	0.926	0.936	0.948	0.936
Obs.	324	324	324	324	324	324

Notes: Each column reports a regression of the indicated dependent variable on the indicated row variables, using monthly data from January 1996 to February 2023. 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 A.6: 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 2006-2019. 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 *

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.

<u>**General Economic Categories**</u>

- **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}

1.1 A Suite of Policy-Related EMV Trackers

We also implement the methodology in Section 2.2 to construct a suite of policy related EMV trackers. Figures B.2 and B.3 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 B.3 shows large upward spikes around the enactment of the Sarbanes-Oxley Act of 2002, during the Global Financial Crisis, around the time of the Dodd-Frank Act of 2010 and after the collapse of Silicon Valley Bank. The EMV tracker for Trade Policy (Figure B.4) also shows distinctive fluctuations, especially surrounding the 'trade war' with China during the Trump presidency. The EMV tracker for Elections and Political Governance in Figure B.5 fluctuates at low levels except for short time windows around the U.S. presidential elections of 2000, 2016, 2020, and, to a lesser extent, 1992. The Healthcare Policy EMV tracker in Figure B.6 shows an unprecedented surge at the onset of the COVID-19 pandemic and remains persistently elevated thereafter. All of the underlying data for these figures, and more, are available at http://www.policyuncertainty.com/EMV_monthly.html, with regular monthly updates.

While these categories focus on different topics, they also may feature substantial overlap. For instance, Monetary Policy EMV and Macro – Interest Rates EMV both highlight news that discusses interest rates and the drivers of interest rates. These two series exhibit substantial correlations, above 0.70. However, we can also learn from the differences between these two series and their behavior over time. For instance, in Figure B.9, we plot the ratio between these two series. We find that, prior to the Global Financial Crisis, Interest Rate EMV tended to be substantially higher than Monetary Policy EMV. In the aftermath, where the Fed took a commanding role in the financial system and pushed policy beyond setting rates, Monetary Policy EMV likewise increased not just in level terms, but also relative to Interest Rate EMV. Table B.1 also notes that, while the two series are highly correlated, they have significant independent explanatory power for bond yield volatility.

More broadly, Figure B.10 plots the ratio between the sum of all categorical EMV sub-indexes and the overall EMV index. Overall, there is significant overlap when considering all of the categorical indexes, yielding a total ratio of around 3.7, reflecting the fact that many articles discuss terms and events that can relate to many of our separate categorical indexes. The ratio mostly fluctuates around this value, with no substantial and persistent trend over our sample period.

To summarize, these figures show highly distinctive temporal movements in the category-specific EMV trackers. Certain events, most notably the market crash of 1987, the Global Financial Crisis, and the COVID-19 pandemic, 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.

Table B.1: Bond Yield Volatility and EMV Categories

	ln(Bond Yield Vol)	ln(Bond Yield Vol)	ln(Bond Yield Vol)	ln(Bond Yield Vol)	ln(Bond Yield Vol)	ln(Bond Yield Vol)
Monetary Policy EMV	0.0166**		-0.0225*	-0.0355**	-0.0245**	-0.0397***
	(0.00664)		(0.0116)	(0.0150)	(0.0116)	(0.0150)
Interest Rate EMV		0.0376***	0.0593***	0.0571***	0.0610***	0.0590***
		(0.00879)	(0.0147)	(0.0146)	(0.0147)	(0.0145)
Overall EMV				0.00708		0.00812
				(0.00607)		(0.00617)
Month FE	NO	NO	NO	NO	YES	YES
Observations	468	468	468	468	468	468
R-squared	0.014	0.048	0.057	0.060	0.071	0.075

Notes: Each column reports results from a separate regression on the natural log of bond yield volatility, as measured by the standard deviation of daily 10-year Treasury bond yields within a given month. Data extends from January 1985 to December 2023. Dependent variables are scaled by 100 for readability. Independent variables include Monetary Policy EMV, Macro – Interest Rate EMV, and the overall EMV index. Robust standard errors are reported in parentheses.

Appendix C. Additional Information About Our Text Sources

Figure C.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 20,000 per month.

The rightmost column of Table C.1 reports average daily article counts by newspaper from 1985 to 2023. 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 the percent of articles devoted to topics encompassed by our term sets.

Five newspapers are not available to us for the entire 1985-2023 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. Proquest 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 C.1: Articles per Day by Term Set Category, 1985-2023

	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 D. Firm 10K Data and 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[2] 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[3] 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:

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

[2] 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).

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

Table D.1: 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

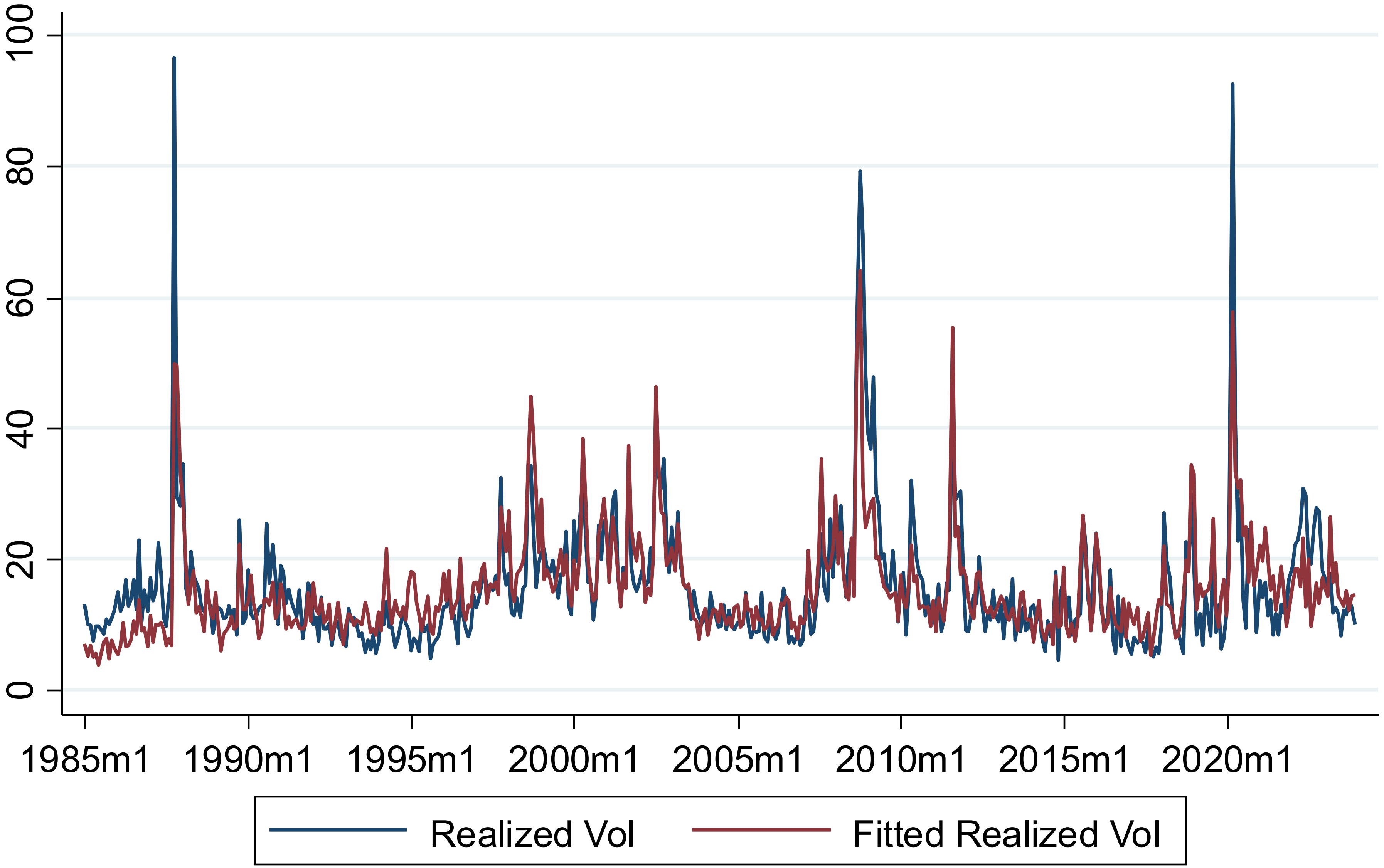
Notes: From each 10-K filing, we use automated methods to count the number of sentences in each Part 1A section. We drop filings for which the automated sentence counter returns a value of less than nine for the part 1A section. This cutoff seems to be the appropriate one based off visual inspection 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 D.2: 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

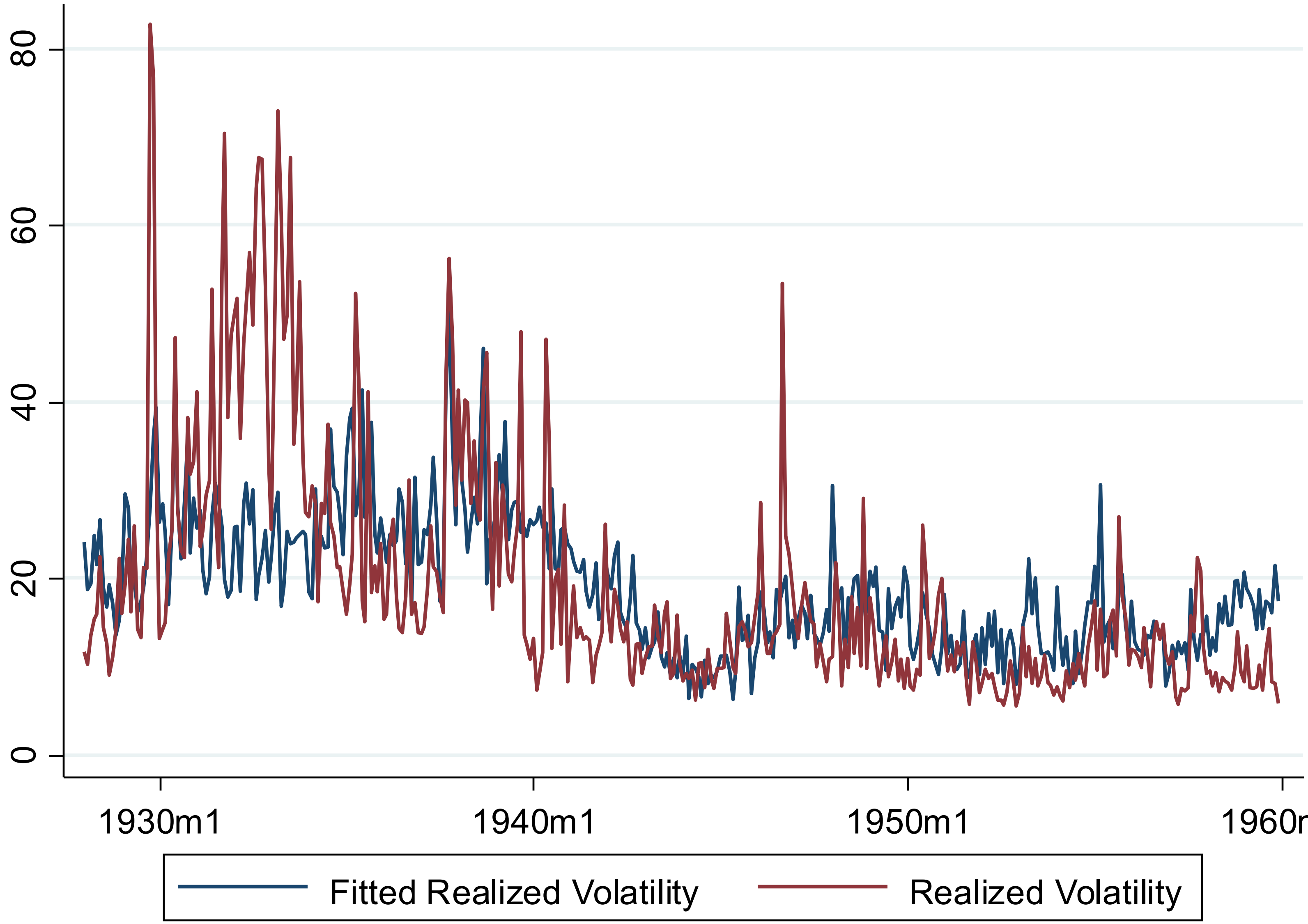
Notes: Firm-level measures are computed at a firm-month level by pooling all monthly firm-level observations within a given year. 508,420 firm-month observations. The average number of monthly firm-level observations per year is 36,317.

Figure A.1: RVol and Fitted RVol from a Regression on EMV, 1985-2023



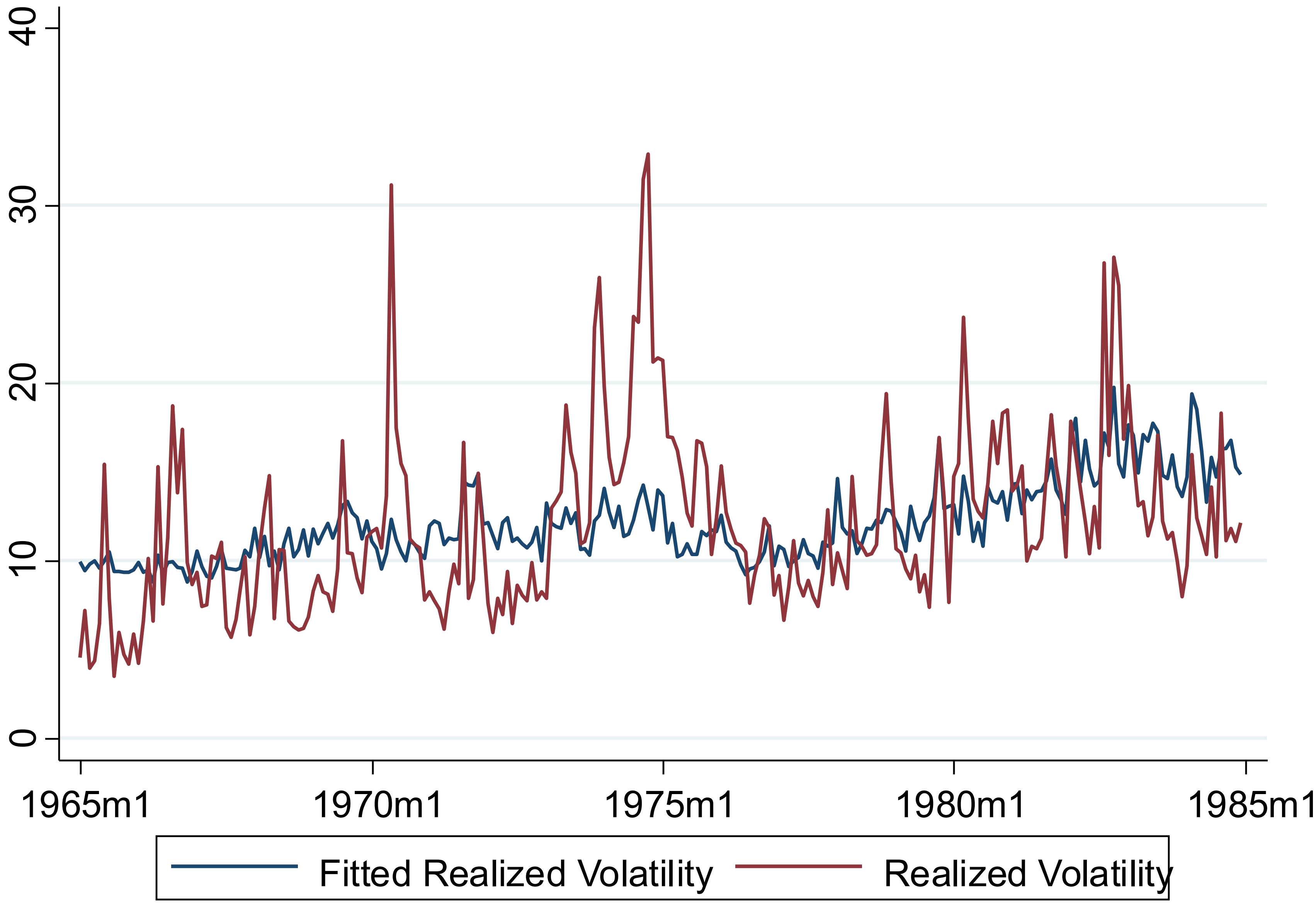
Notes: Realized vol represents the standard deviation of daily returns of the S&P-500 index for a given month. "Fitted RVol" values are from the regression of Realized Vol on EMV reported in Table 2, column (7) Both series run from January 1985 to December 2023.

Figure A.2: Historical Realized Vol and Fitted Realized Vol (1928-1959)



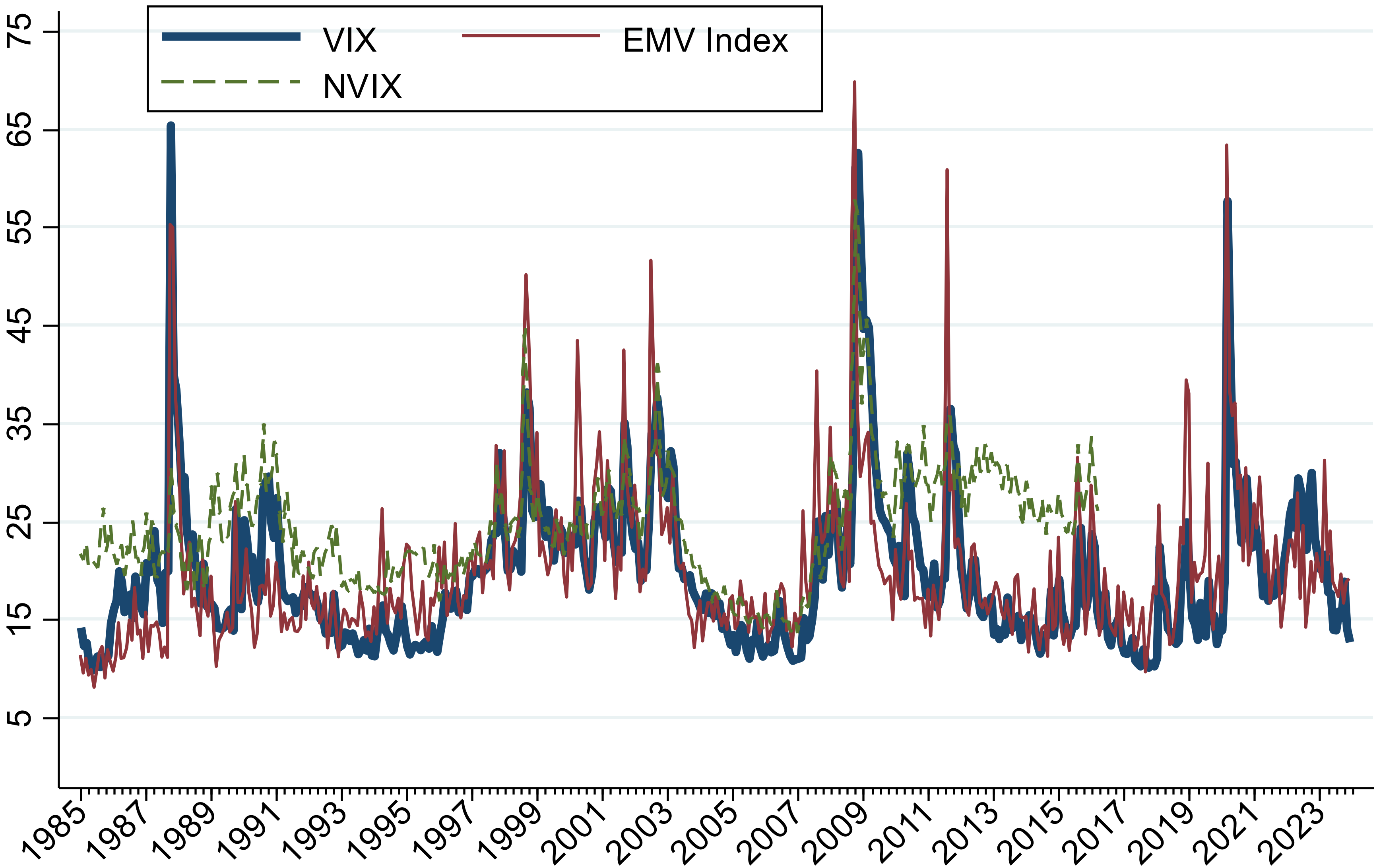
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe,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

Figure A.3: Historical Realized Vol and Fitted Realized Vol (1960-1984)



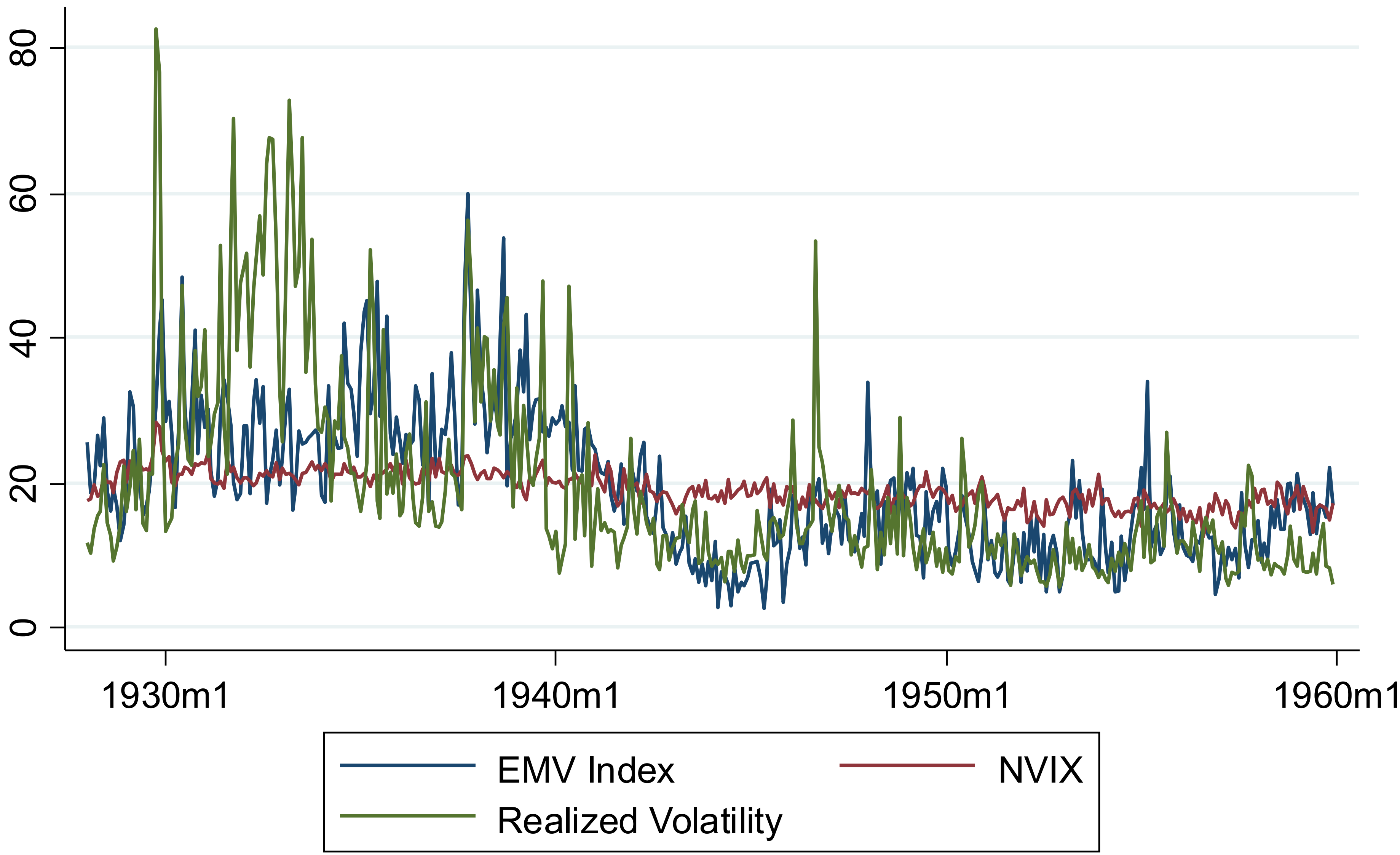
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe,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

Figure A.4: VIX, EMV and NVIX, 1985-2023



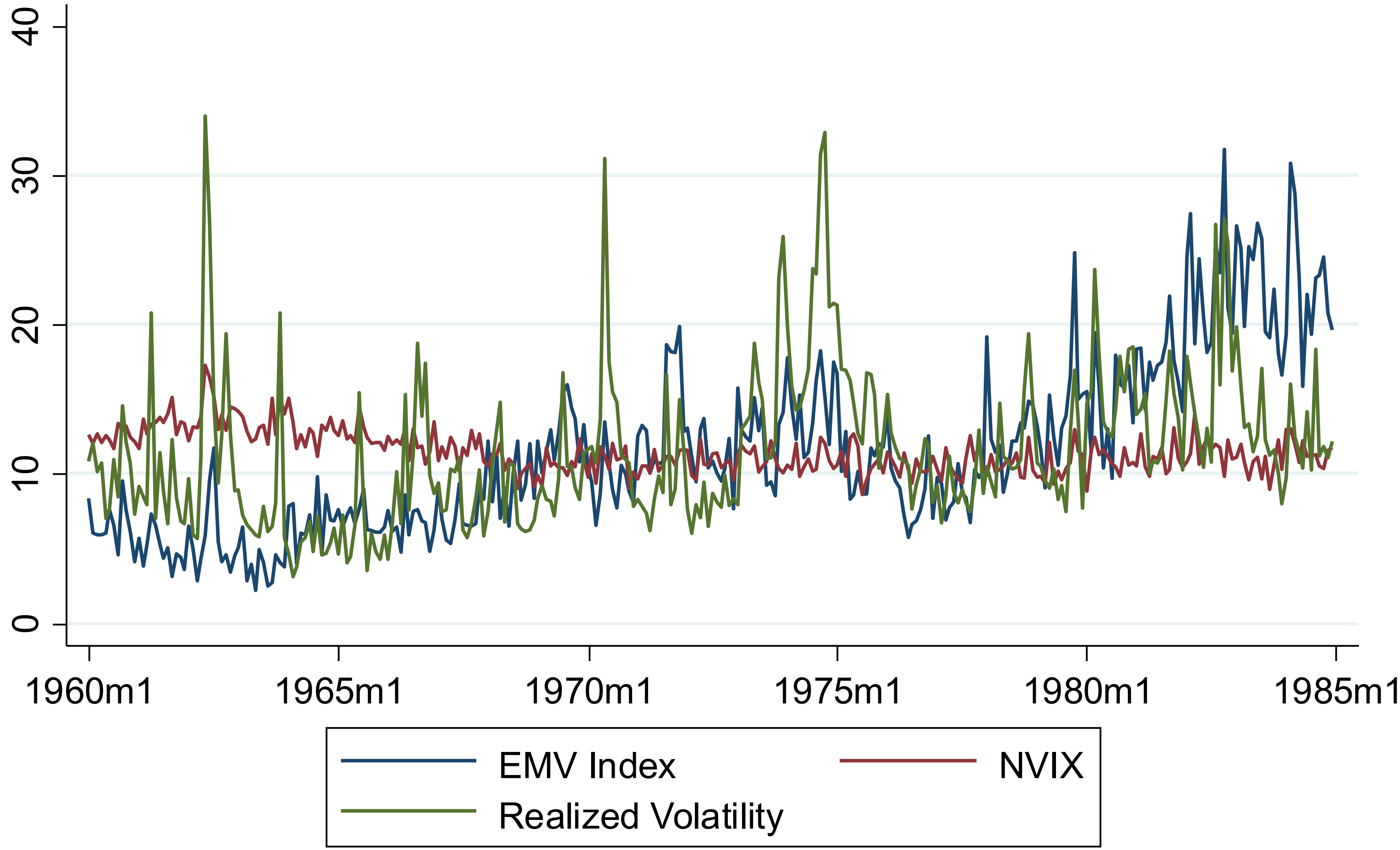
Notes: The NVIX measure is from Manela and Moreira (2017) and runs through March 2016. See the notes to Figure 2 for the VIX and NVIX. We multiplicatively scale NVIX and EMV to match the mean value of the VIX from 1985 to 2015.

Figure A.5: Historical EMV Index (1928-1959)



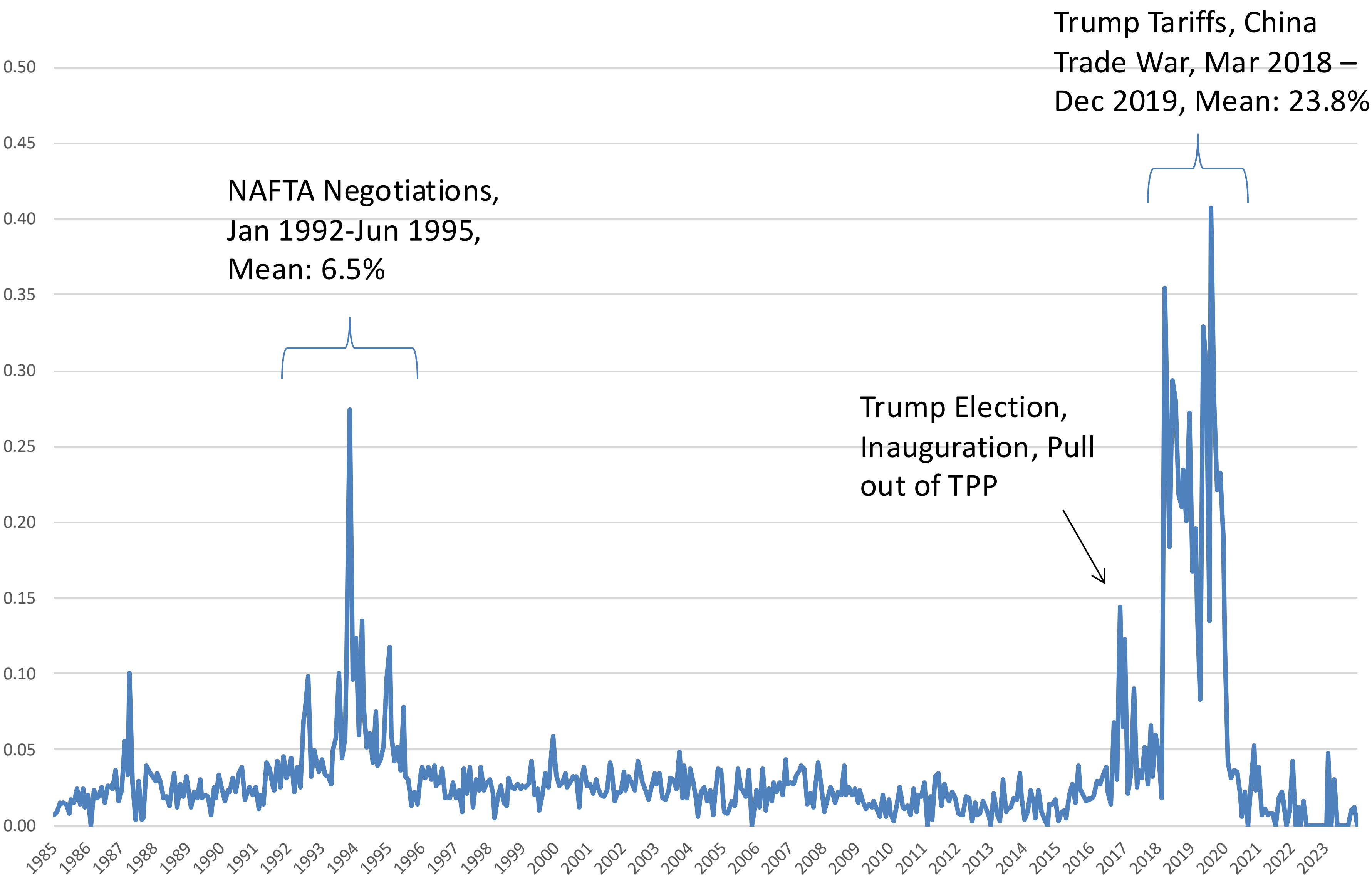
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe, Chicago Tribune, Washington Post, and Los Angeles Times. For this chart, we multiplicatively scale NVIX and EMV to match the mean value of the VIX from 1928 to 1959.

Figure A.6: Historical EMV Index (1960-1984)



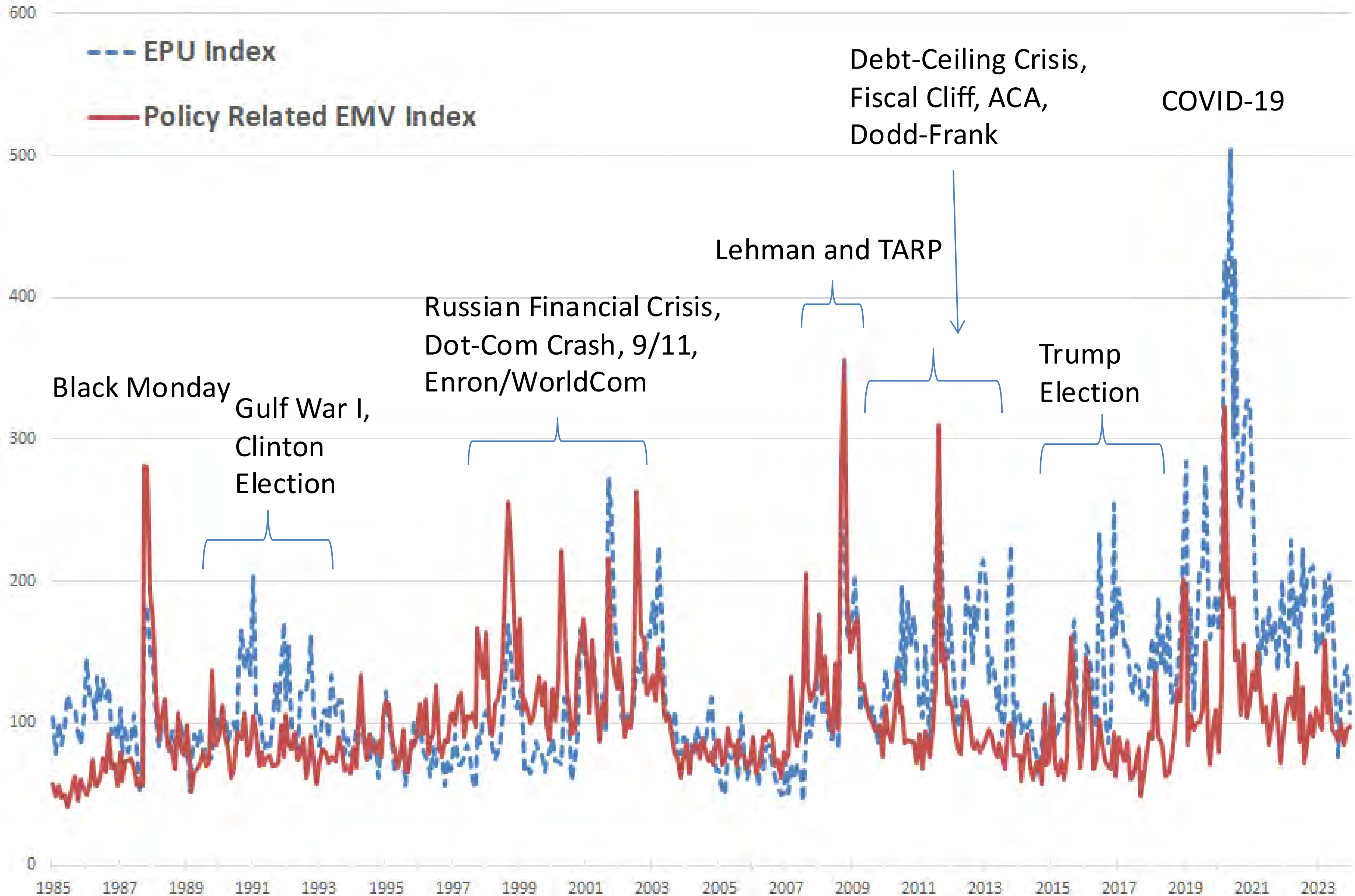
Notes: The Historical Equity Market volatility (EMV) tracker runs from January 1928 to December 1984. We construct it using scaled frequency of articles that contain terms about Economics, the Stock Market, and Volatility in leading U.S. newspapers: New York Times, Wall Street Journal, Boston Globe, Chicago Tribune, Washington Post, and Los Angeles Times. For this chart, we multiplicatively scale NVIX and EMV to match the mean value of the VIX from 1960 to 1984.

Figure A.7: Share of EMV Articles Discussing Trade Policy, 1985-2023



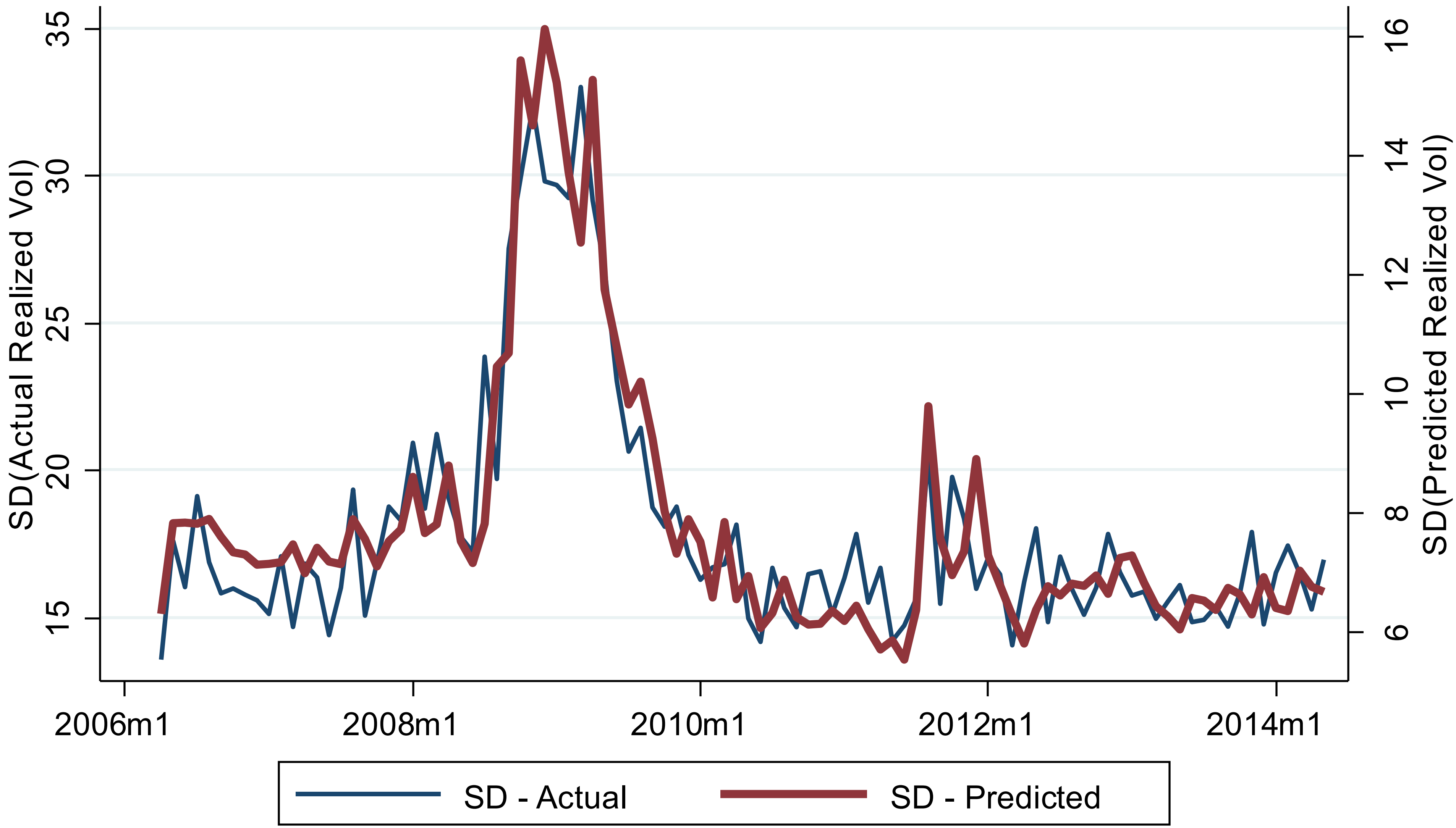
Note: This chart shows the share of EMV articles that contain one more terms in **Trade Policy** by month**.** See Appendix B for a specification of the terms in **Trade Policy.**

Figure A.8: Policy-Related EMV Tracker and BBD EPU Index, 1985-2023



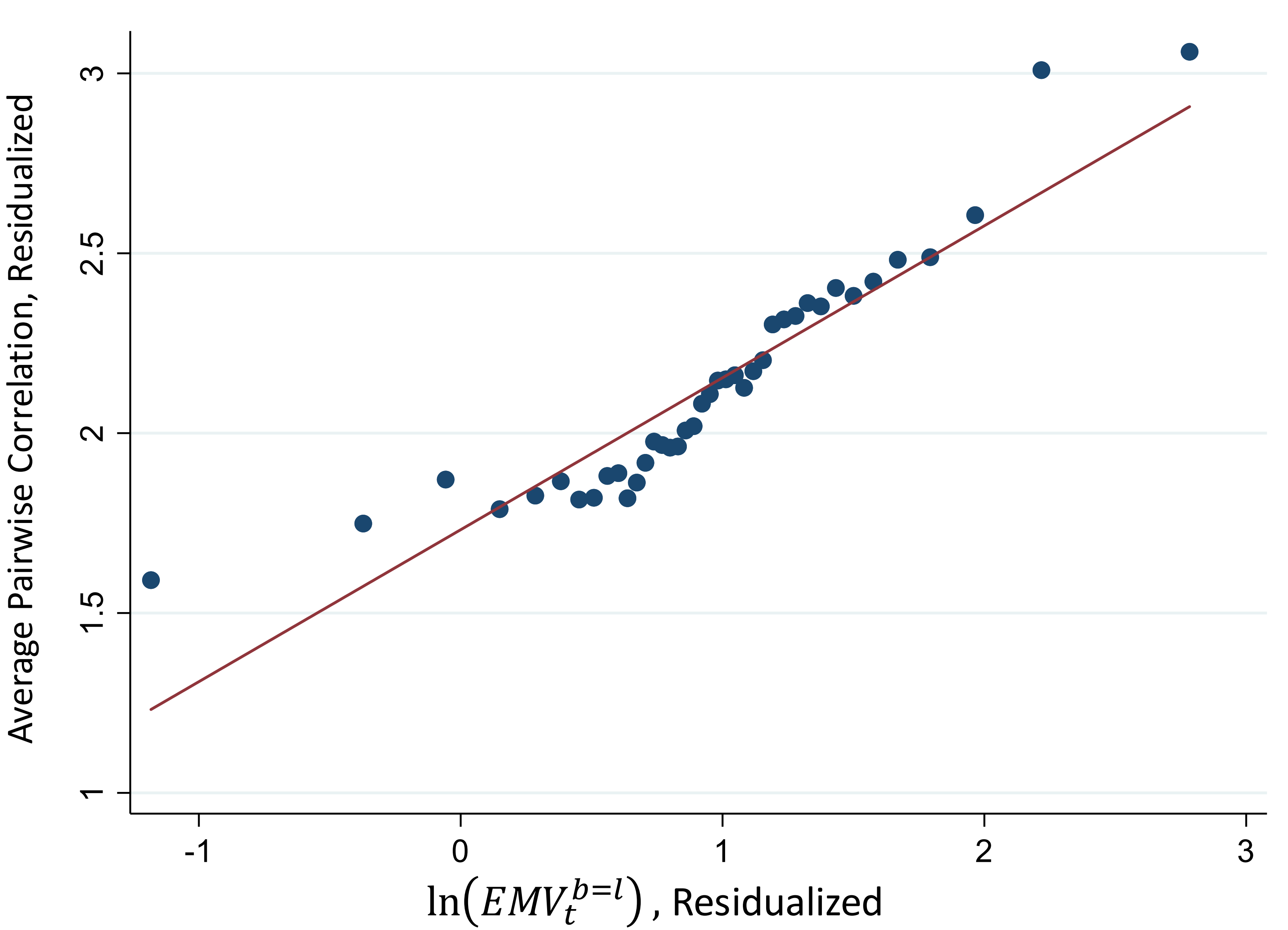
Notes: The BBD EPU Index is from Baker Bloom and Davis (2016). To construct the Policy-Related EMV tracker, we multiply our overall EMV tracker by the fraction of EMV articles the discuss policy matters. We multiplicatively rescale Policy-Related EMV to match mean of the BBD EPU Index from 1985 to 2009.

Figure A.9: Cross-Firm Standard Deviation of Realized Volatility – Actual versus Fitted Values (2006-2014)



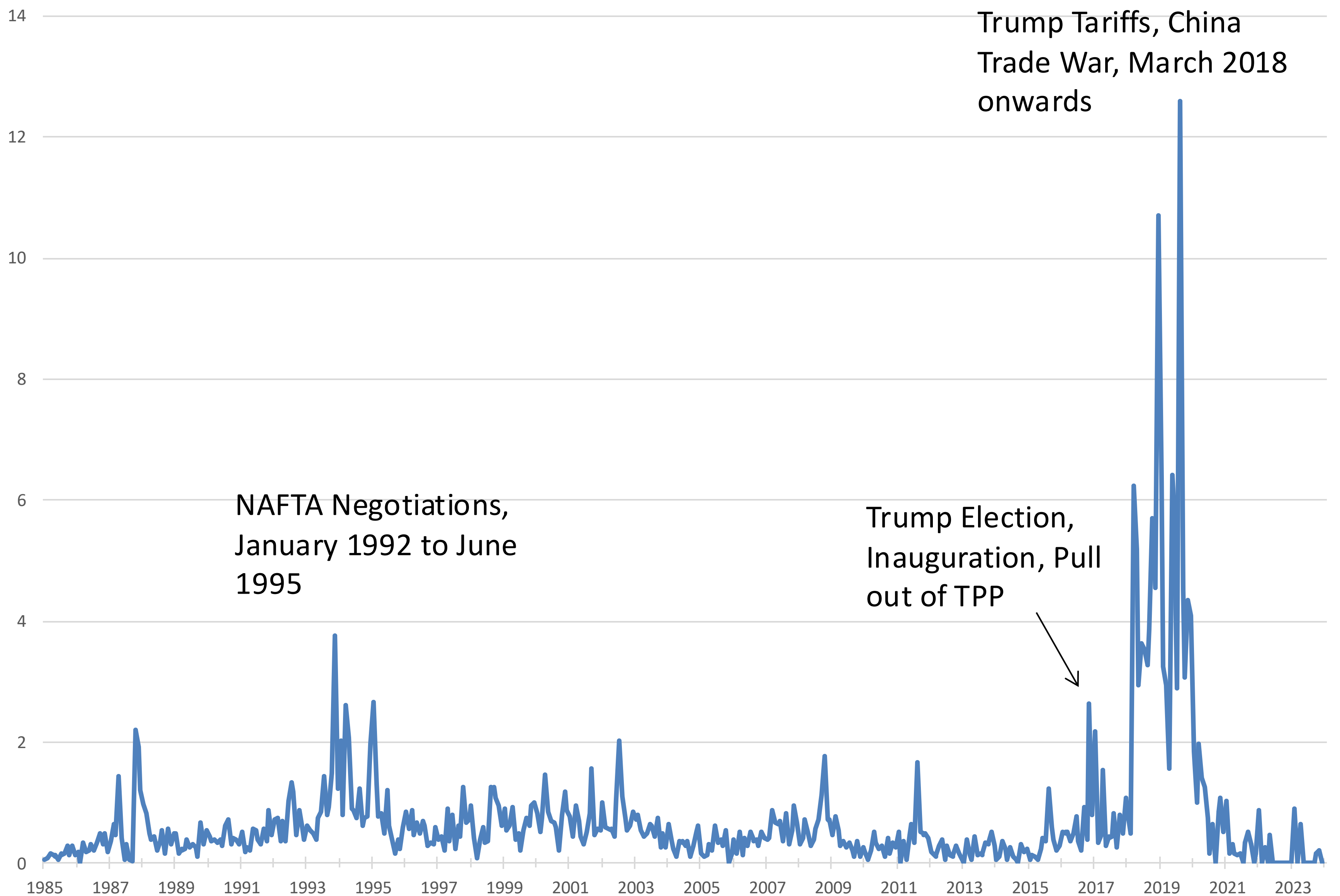
Notes: Firm and time fixed effects were swept out in a prior step. Then, the specification from Table 8 column (6) was run and predicted values for realized volatility at the firm-month level were constructed. The figure plots the standard deviation across firms for each point in time for both the residualized realized volatility and the predicted values.

Figure A.10: Bin Scatter Corresponding to Column 1 in Table 6



Note: This figure presents a bin scatter for the same observations and regression specification as Column (1) of Table 6 in the main text. Before constructing the bin scatter, we residualize each variable with respect to firm-level fixed effects.

Figure B.1: Trade Policy EMV Tracker, 1985-2023

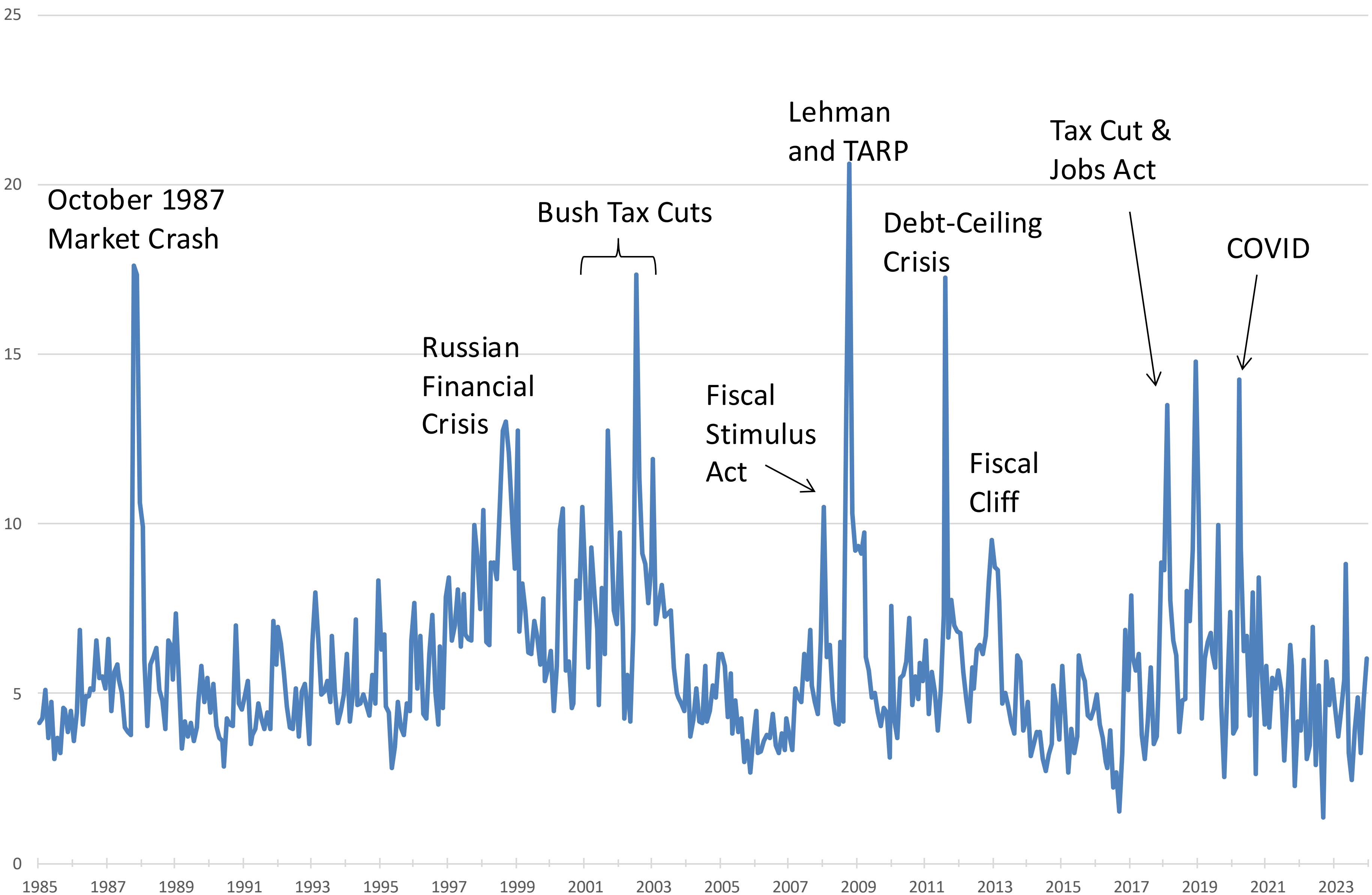


Notes: We construct the Trade Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Trade Policy.** See Appendix B for the list of terms.

Figure B.2: Monetary Policy EMV Tracker, 1985-2023

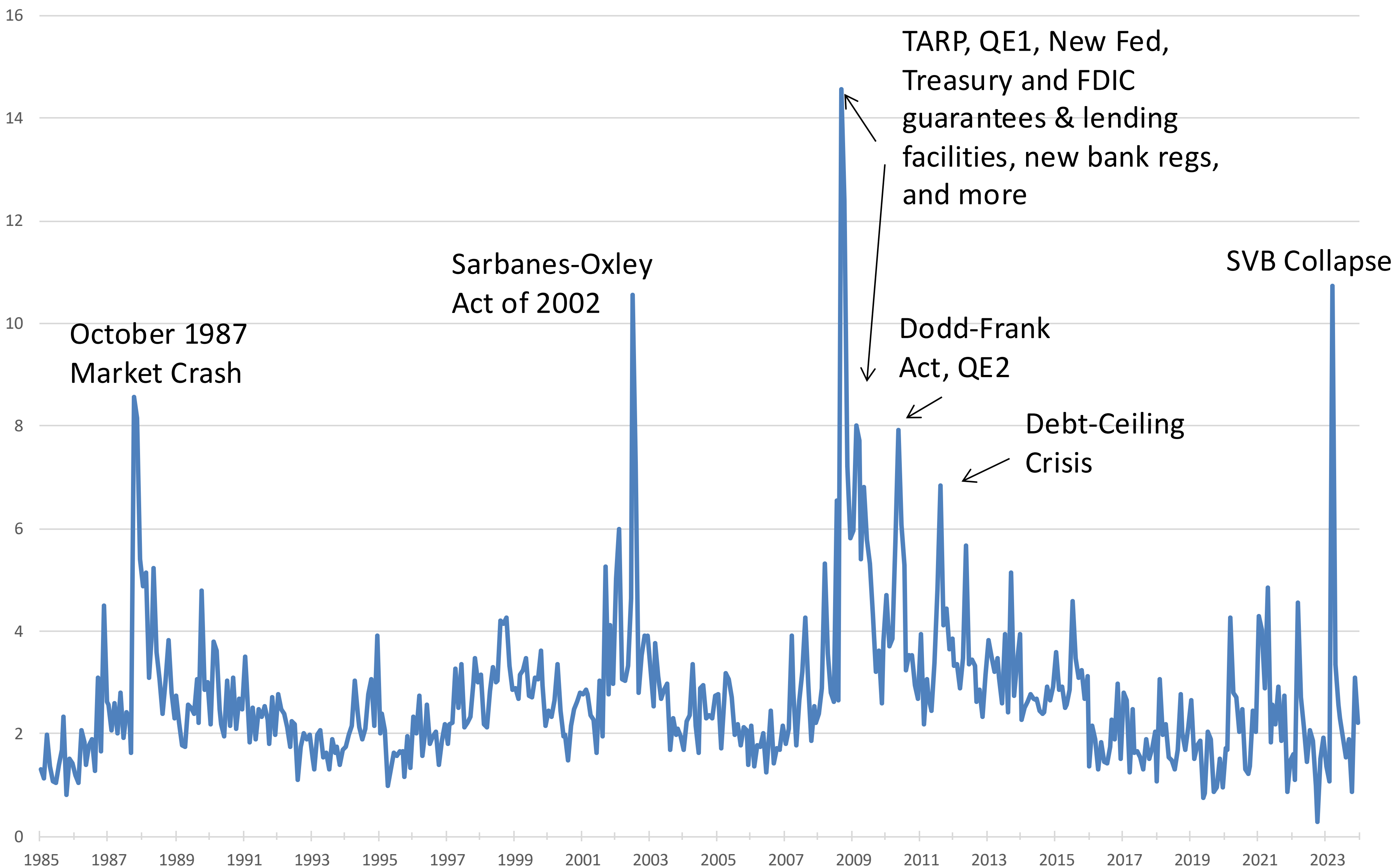
Notes: We construct the Monetary Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Monetary Policy.** See Appendix B for the list of terms.

Figure B.3: Tax Policy EMV Tracker, 1985-2023



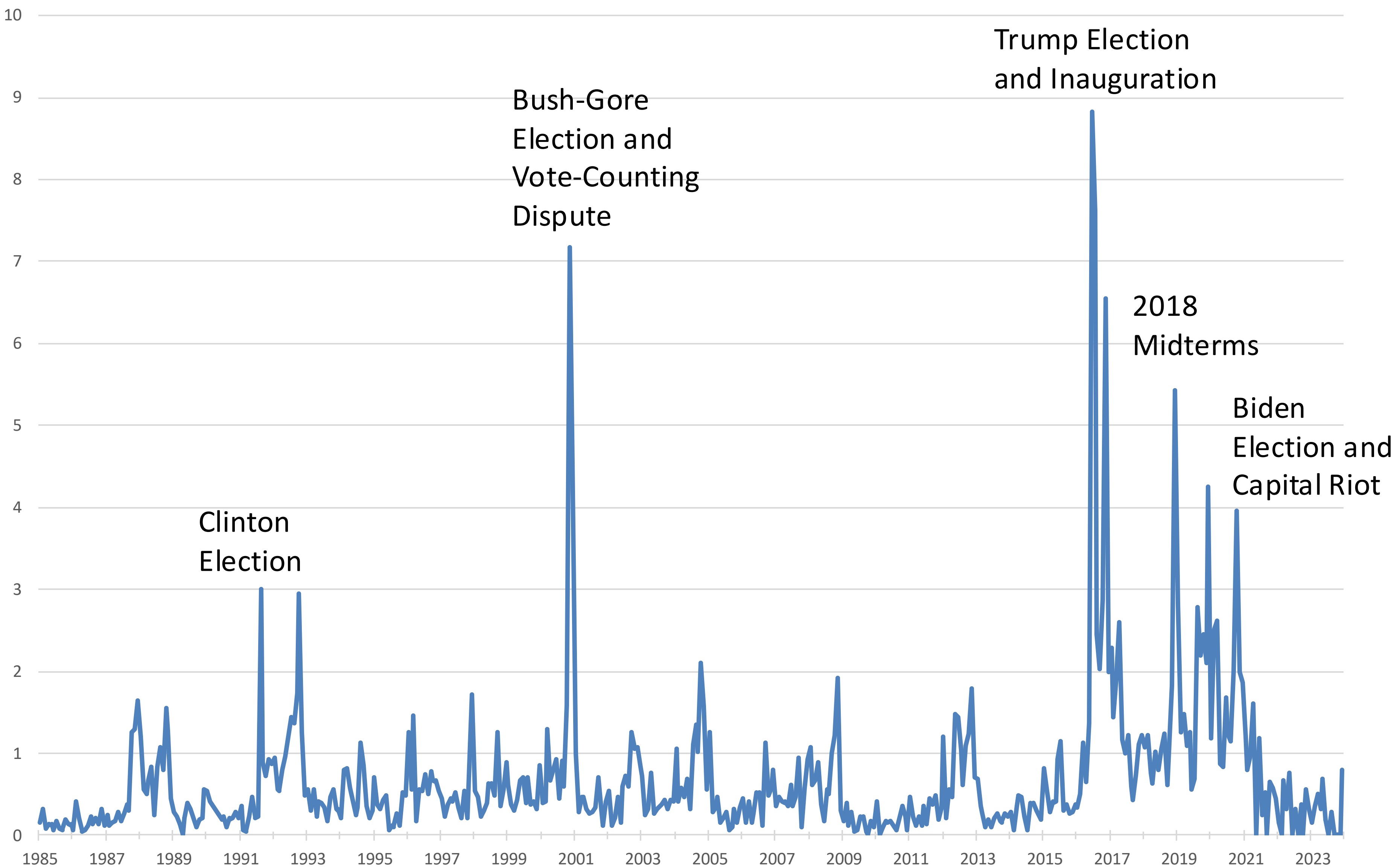
Notes: We construct the Tax Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Taxes.** See Appendix B for the list of terms.

Figure B.4: Financial Regulation EMV Tracker, 1985-2023



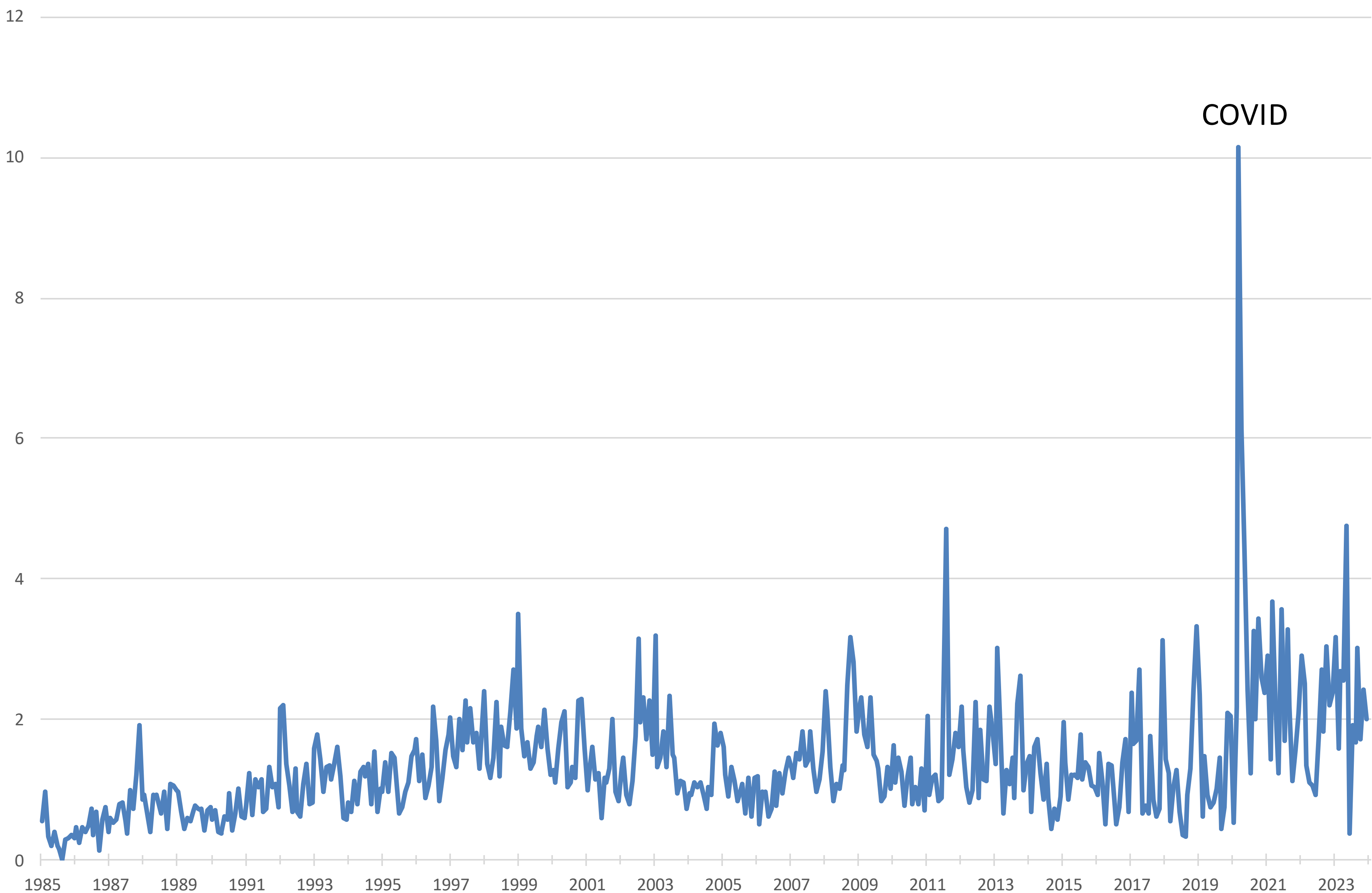
Notes: We construct the Financial Regulation EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Financial Regulation.** See Appendix B for the list of terms.

Figure B.5: Elections and Political Governance EMV Tracker, 1985-2023



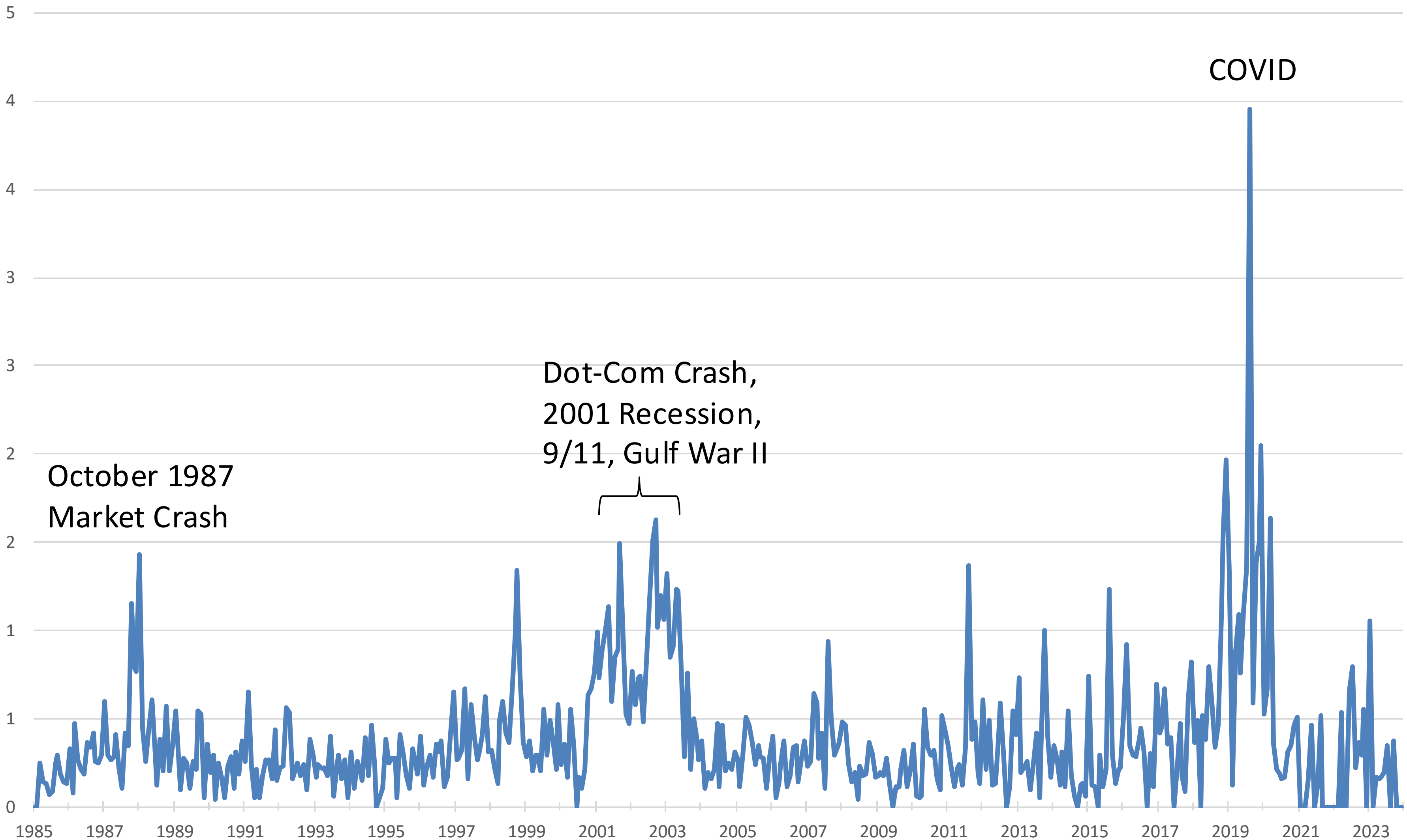
Notes: We construct the Elections and Political Governance EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Elections and Political Governance.** See Appendix B for the list of terms.

Figure B.6: Healthcare Policy EMV Tracker, 1985-2023



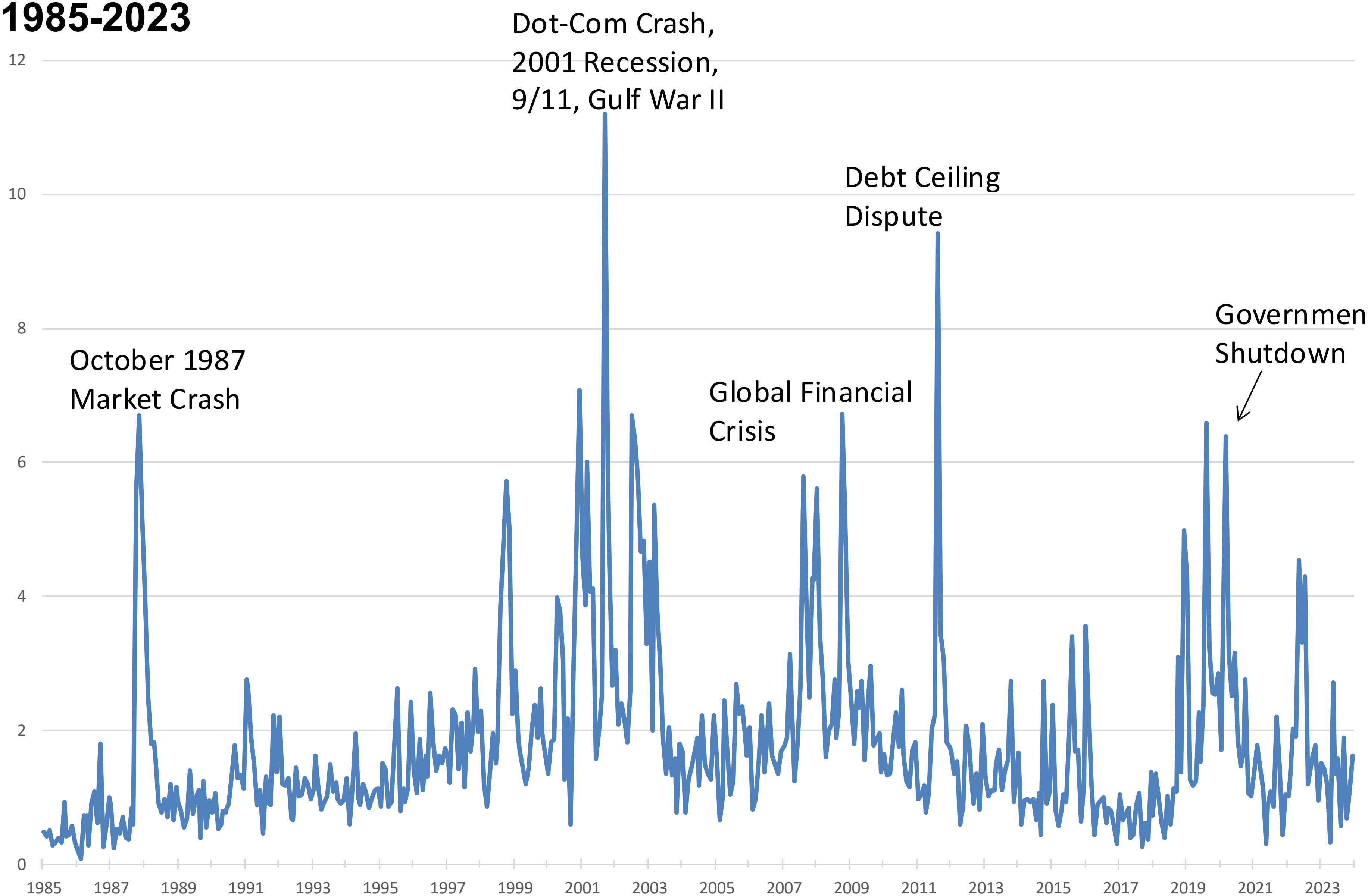
Notes: We construct the Healthcare Policy EMV tracker as the product of our overall EMV tracker and the share of EMV Articles that contain one or more terms in **Healthcare Policy.** See Appendix B for the list of terms.

Figure B.7: Macro – Business Investment and Sentiment EMV Tracker, 1985-2023



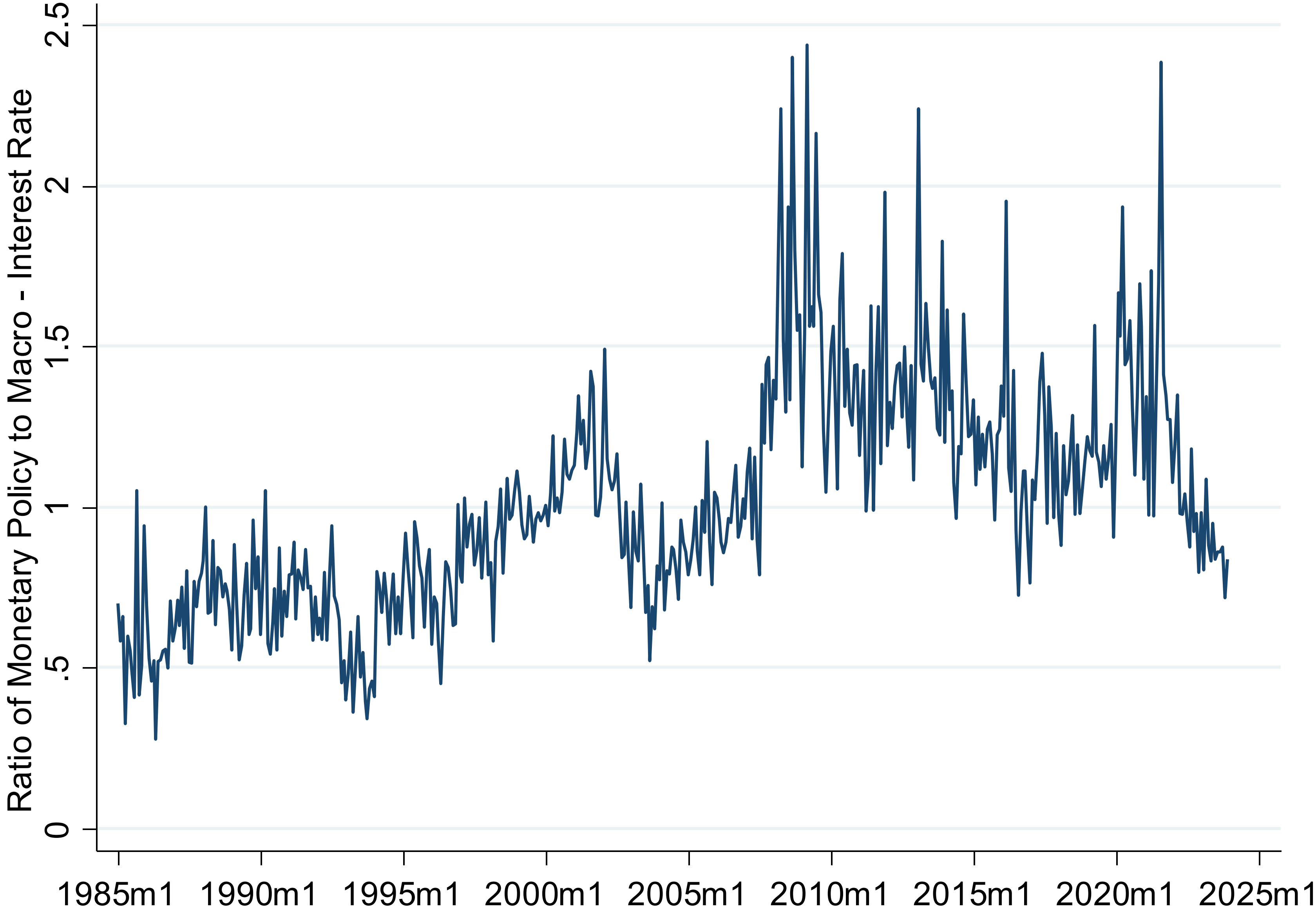
Notes: The Macro – Business Investment and Sentiment EMV Tracker is constructed as our EMV Index multiplied by the share of EMV Articles that contain one or more terms in the "Macro – Business Investment and Sentiment" termset which can be found in the Appendix.

Figure B.8: Macro – Consumer Spending and Sentiment EMV Tracker, 1985-2023



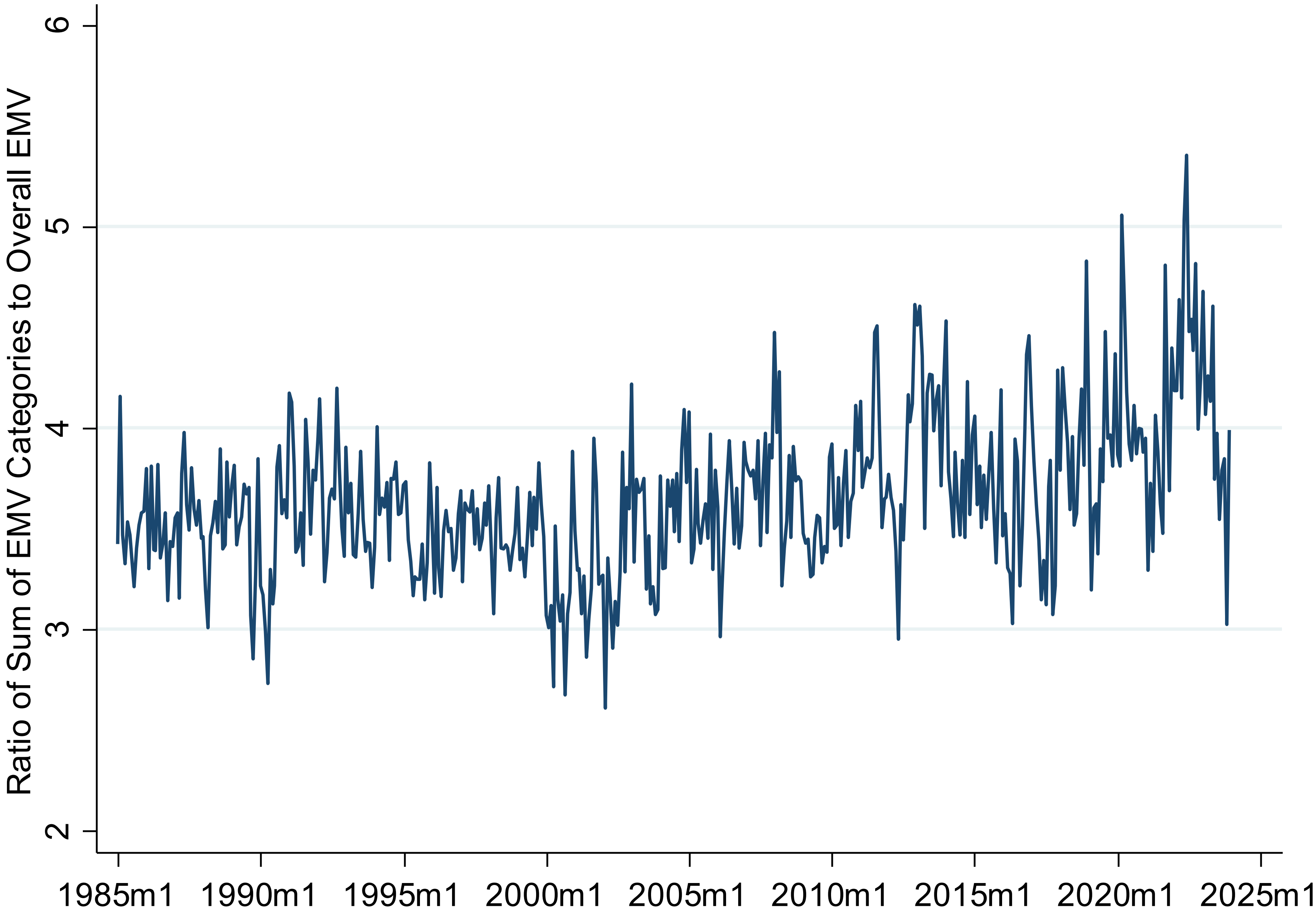
Notes: The Macro – Consumer Spending and Sentiment EMV Tracker is constructed as our EMV Index multiplied by the share of EMV Articles that contain one or more terms in the "Macro – Consumer Spending and Sentiment" termset which can be found in the Appendix.

Figure B.9: Ratio of Monetary Policy EMV to Macro – Interest Rate EMV



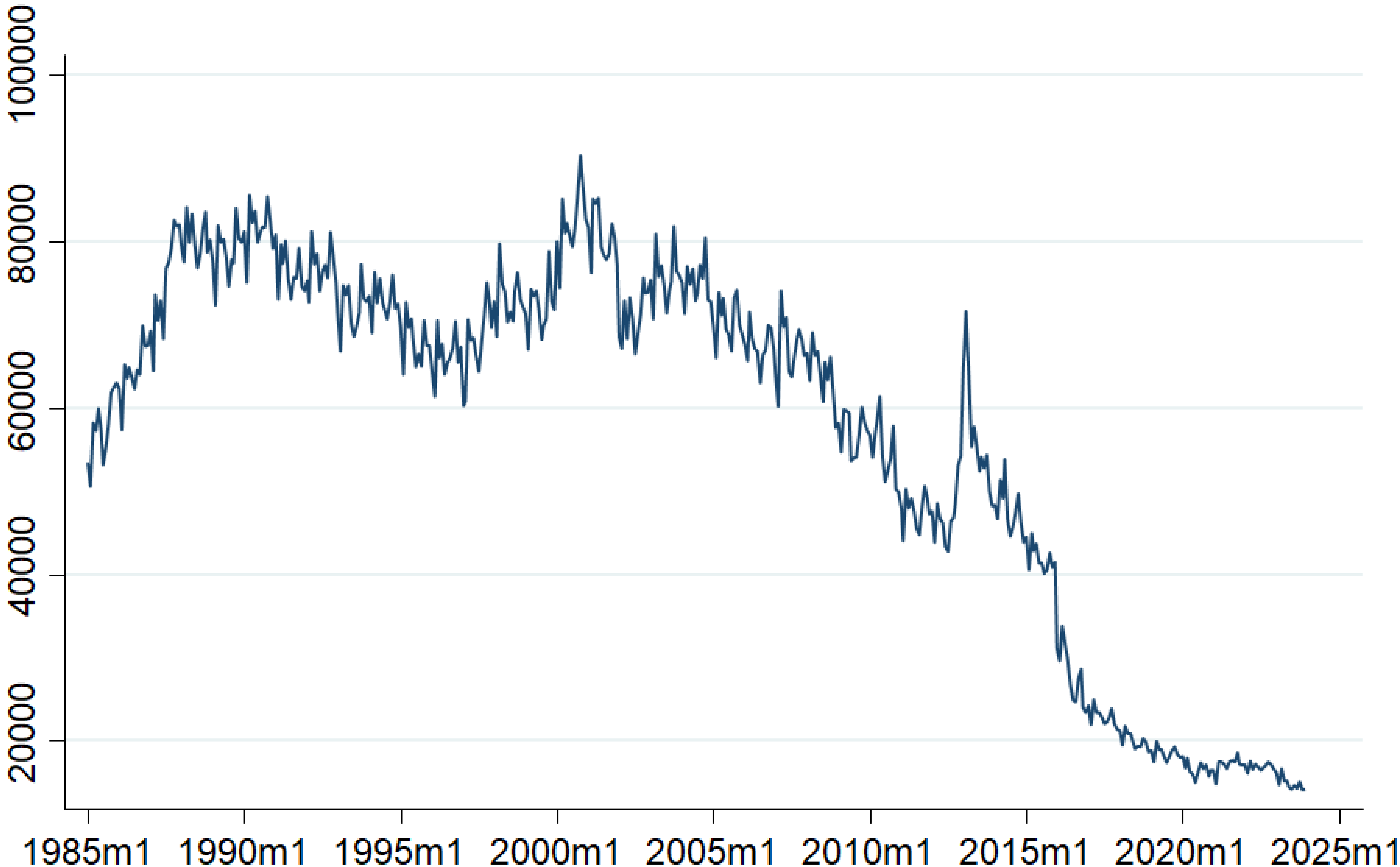
Notes: Plotted is the ratio of monthly values of the Monetary Policy EMV to monthly values of the Macro – Interest Rate EMV series. Data spans January 1985 to December 2023.

Figure B.10: Ratio of Sum of Categorical EMV Indexes to Overall EMV



Notes: Plotted is the ratio of the sum of all categorical EMV series to the overall EMV index. Data spans January 1985 to December 2023. Average of the ratio over the entire sample is 3.7. Ratio exceeds 1 in all months as some articles contain references to multiple categorical topics.

Figure C.1: Total Number of Newspaper Articles, Monthly, 1985-2023



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.