What Triggers Stock Market Jumps?

Scott R. Baker[a], Nicholas Bloom[b], Steve Davis[d] and Marco Sammon[a]

September 2019

Abstract: We examine newspapers the day after major stock-market jumps to evaluate the proximate cause, geographic source, and clarity of these events from 1900 in the US and 1980 (or earlier) in 13 other countries. We find three main results. First, the United States plays an outsized role in global stock-markets, accounting for 35% of jumps outside the US since 1980s, far above its 15% share of GDP. This matches other evidence on the dominance of the US in global finance. Second, the clarity of the cause of stock market jumps has been increasing notably since 1900, as news and financial markets have become more transparent. Jump clarity predicts future stock returns volatility: doubling the clarity index of a jump reduces future volatility by 68%. Third, jumps caused by non-policy events (particularly macroeconomics news) lead to higher future stock-volatility, while jumps caused by policy events (particularly monetary policy) reduce future stock-volatility. This suggests while monetary policy surprises lead to stock-market jumps, they may reduce future volatility.

JEL Codes:

Keywords: uncertainty, policy uncertainty, volatility, stock market

Acknowledgements: We would like to thank the National Science Foundation and the Sloan Foundation for their financial support.

a Kellogg School of Management, Northwestern University
b Stanford University and NBER
c Chicago Booth School of Business and NBER

1. Introduction

An old question in economics is "*what causes stock market jumps*"? At one extreme is the view that all stock price movements rationally incorporate news about stock returns or discount rates. As such, large jumps in national stock indices should be accompanied by news influencing future returns or discount rates. At the opposite extreme is the view that the stock-market fluctuations are driven by speculation, for example the well-known quotes by Keynes (1936) that investing is like a "beauty contest", where investors price stocks not based on their opinion of their fundamental valuation but what they think others currently value them for.

In this paper we tackle this question using examining the next day's newspaper after major stock market moves, covering over 1,100 jumps of +/- 2.5% since 1900 in the US and 2,500 jumps in 13 other countries. These jumps are large enough that they almost always attract newspaper coverage in major newspapers the following day, so we can analyze these articles using a team of 22 undergraduate and graduate auditors. And because a sizeable fraction of stock market movements occurs on these jumps days, understanding their determinants offers insights into financial market more broadly.[1]

Our auditor team categorizes stock market jumps into one of 16 categories according to the journalists' reporting, determines their geographic origin and evaluates measures of clarity of the attributed cause. In the US, we do this using five different newspapers for each jump – the Wall Street Journal, the New York Times, the Washington Post, the Boston Globe and the LA Times – while in other countries we use one or two leading papers.

We also test a range of machine learning and natural language models and discuss why these approaches are (at present) inferior to human auditors. We hope, however, that this large corpus of jump events and associated newspaper text will aid the ongoing development of text to data for financial moves.[2]

Of course, earlier studies have examined news reports to evaluate the drivers of stock-market moves. For example, classic studies by Niederhoffer (1971) and Cutler, Poterba, and Summers (1989) examined major US stock-market jumps in the past to see to what extent they could be explained by news events, coming to mixed conclusions. Our approach differs in its

[1] Between 1900 and 2018 about 20% of total daily variation (sum of absolute returns) and 50% of daily quadratic variation (sum of squared returns), happened on the 3% of trading days with the largest absolute returns.
[2] The jump dataset and full set of newspaper text is available at XXX.

scale in examining over 4,000 jumps, its breadth covering 14 countries and going back to 1900 in the US (and 1930 in the UK), and detail in measuring the causes, geographic source and clarity.

For some days, this attribution is simple. In Figure 1 plots the intraday movements (in 5-minute increments) of 4 days with daily stock market movements of greater than 2.5%. The top row contains two days with sharp, near-instantaneous, movements in the S&P 500 index which makes it easy for journalists to attribute the cause of movements on these days. In the top left the market jumped almost 3% after the Fed announced interest rate cuts while in the top right the market surged in opening following a European announcement to provide bailout support for Greece. In other cases, for example the two days in the bottom row, the market drifted by more than 2.5% during the day, but with no clear jump or event, leaving journalists were unclear about the cause.

This paper demonstrates four key results. First, the US has been and remains an extremely important driver of global stock-market volatility. Between 1980 and 2018 the share of jumps attributed to the US was 34%, substantially above its 20% share of global GDP. Moreover, this share of jumps attributed to the US has risen moderately since 1980 even though the US share of global GDP has fallen.

Second, the 'clarity' of stock market move attribution – measured by the share of articles within and across papers that agree on the cause of a jump, the share of "unknown" attributions, and the confidence of the journalists assertion over causality - has increased dramatically. From 1900 to 1945 news coverage of financial markets shows a steep rise in clarity, probably linked to the improvements in financial transparency, communications and news. Clarity also turns out to matter for future volatility – perhaps unsurprisingly, jumps which have unclear attribution are followed by significantly more volatility in future days.

Third, jumps caused by non-policy events (particularly macroeconomics news) lead to higher realized stock-market volatility, while jumps caused by policy events (particularly monetary policy) reduce realized and implied future stock-volatility. This suggests while monetary policy surprises lead to stock-market jumps, they may reduce future volatility.

Finally, the mix of jumps has itself changed over time. Most notably, comparing stock movements in the United States prior to 1945 to those following 1945, we find that Commodities, Regulation, and Sovereign Military Action were a significantly larger share of jump drivers in the pre-war period, while in the post-war period, Corporate Earnings, Macroeconomic News, Monetary Policy and Non-Sovereign Military Action (Terrorism) are more dominant.

Our work builds on several prior literatures. Many papers have shown that financial journalism affects the stock market, above and beyond the information contained in the articles. Tetlock (2007) shows that sentiment in the Wall Street Journal's *Abreast of the Market* column can predict returns, and extreme optimism or pessimism predicts high trading volume. We build on this, showing that different categories of news have different implications for volatility after the news is reported[3]. Engelberg and Parsons (2011) use differences in local media coverage of national events to show that differences in journalists' explanations are internalized by investors reading those articles. Our method covers multiple newspapers, and finds that when the reporters disagree, realized volatility is higher, consistent with Carlin, et al (2014). Manela and Moreira (2017) use machine learning to construct a measure of stock market uncertainty from newspaper data and find that news about wars and policy are important determinants of risk premia. We also find that policy is an important driver of stock market jumps, and discuss the potential pitfalls involved with machine classifications of newspaper articles.

We also contribute to the literature on how the clarity of financial writing affects stock returns. Li (2008) constructs a 'fog index', designed to measure the readability of SEC filings from document length and sentence complexity. Li finds that less 'fog' predicts better future firm performance. We construct a 'clarity' index based on subjective human assessment of article readability, and the strength of attribution of a cause to the jump of interest. We find high clarity predicts lower volatility after the jump. Shiller (2017) discusses how narratives can become widespread and affect global stock markets, even if they are not true. We find that jumps without a strong link to fundamental information on average lead to more volatility than jumps with clear connections to new economic developments.

A large body of work (eg. Shiller (1981), Roll (1988), etc.) has discussed the extent to which fluctuations in stock price movements, can be attributed to news about fundamentals like future cash flow and discount rates. In this vein, Cutler, Poterba, and Summers (1989) investigate the interaction of financial market returns with both macroeconomic news as well as 'qualitative news' regarding political or military events, by examining specific large movements of equity

[3] Note that our exercise differs from Tetlock (2007) and others, in that we are interested in the ex-post attribution of stock market jumps to causes by newspapers, rather than the effect of newspaper coverage on future stock-market behavior.

markets in the United States. We continue and expand upon their work, investigating what drives large stock market movements and how these causes may have important implications for the future path of asset prices and volatility[4]. This is consistent with Pastor and Veronesi (2012) where after bad fundamental news arises the government steps in to ameliorate the problem and with Kelly, Pastor and Veronesi (2016), where option prices drop after elections. Our results are consistent with the models' predictions when studying realized stock market volatility over the month following the jump. For example: monetary policy jumps are associated with relatively lower future abnormal realized volatility than macroeconomic news.

Many papers have measured the effect of news releases on the stock market. Boudoukh et. al. (2013) find that they can increase the R-squared measure in Roll (1988) by selecting 'relevant' news, and by conditioning on sentiment. We build on this in two dimensions: (1) By focusing on days with large stock market moves, there is almost always an article in the financial press offering a potential explanation (2) By having trained readers select the articles, we are more likely to be focusing on news relevant to each jump.

Birz and Lott (2011) identify news headlines following macroeconomic data releases and find that news about GDP and employment are especially important for predicting stock returns. We find that volatility is higher following jumps attributed to Macroeconomic News & Outlook than all other categories. Fernandez-Perez et. al. (2017) find that the VIX drops after FOMC announcements, consistent with our results that volatility is lower following jumps attributed to Monetary Policy than all other categories. Goldberg and Grisse (2013) conduct a high frequency analysis on days where Macroeconomic news is released and find that the stock market response to news depends on current economic conditions. Consistent with this, we find that the differences in future realized volatility across categories is stronger in recessions and is stronger when the initial jump is negative. Fisher et. al. (2017) find that media attention has predictive power for volatility even conditioning on information contained in the macro announcements. We find our results related to Monetary Policy are robust to conditioning on the monetary policy surprise contained in each FOMC announcement, as measured by Gurkaynak et. al. (2005).

[4] An extensive literature has more broadly documented and modeled the properties of stock market volatility. The Engle (1982) ARCH model allows previous shocks to influence current volatility. This was generalized in Bollerslev (1986), which allows for a general ARMA structure in the error variance. To account for the Black (1976) leverage effect, Glosten, Jagannathan, and Runkle (1993) allow for asymmetric effects of positive and negative innovations in the volatility process. For more related work, see Bollerslev, Engle, and Nelson's Chapter of the (1994) Handbook of Econometrics.

Many papers have documented the dominance of the United States in global financial markets. For example, Maggiori, et. al. (2018) find that dollar-denominated securities are an exception to home-bias puzzle in international investing. Boz et. al. (2017) find the dollar share of global invoicing is higher than the U.S. share of global GDP or global trade. Obstfeld (2015) finds that a large amount of credit intermediated outside the United States is denominated in U.S. Dollars. Gopinath and Stein (2018) argue the dominance the dollar can be explained by complimentary between a currency being used for invoicing and for being a safe store of value. We contribute this literature by recording the geographic origins of the jumps in our sample and confirming the dominant role of U.S. news developments as a driver of jumps globally.

Several papers have explored the links stock markets across countries. Mehl (2013) finds that shocks are transmitted across global stock markets, and these effects cannot be entirely explained by fundamentals. Consistent with this, we find that real links between economies (trade share of GDP) cannot explain the shares of jumps transmitted across countries. Ehrmann et. al. (2011) look at transmission of shocks both across countries and across asset classes. They find US has strong influence on Europe, but Europe has minimal effect on US. We find this is also true for stock market jumps.

Finally, there is a large literature linking the stock market to real economic outcomes. Fama (1981) finds a negative relationship between stock returns and inflation, Fischer and Merton (1984) show that stock returns are good predictors of business cycles and output, and Barro (1990) links stock returns to investment. Campbell et al. (2001) find that market-wide, industry-level and idiosyncratic volatility all have predictive power for GDP growth. We contribute to this by looking at the predictive power of different jump categories for GDP. We find that Macro-related news is positively related to future GDP growth, while for all other categories, jumps of any sign predict lower GDP going forward.

Section 2 describes the construction of the categorized stock market movement data as well as the other data sources utilized in the paper. Section 3 presents facts regarding composition of jump drivers over time and across countries. Section 4 contains several exercises taken to evaluate the accuracy of the categorizations. Section 5 discusses our measurement of the clarity of jump category attribution. Section 6 illustrates differential effects that jump categories have on returns and volatility. Section 7 notes the relationship between jump type and real economic effects. Section 8 concludes.

2. Data

2.1 US Stock Jumps Data

Using a large team of human readers, we categorize the cause of large daily stock market moves based on newspaper coverage the following day. For the United States, we first compile a list of all days where the CRSP Value-Weighted Index had an absolute return of 2.5% or more from 1926 to 2016. Prior to 1926, we utilize the GFD's DOW extension.

In the United States, we utilize the following procedure across five major newspapers: the Wall Street Journal, the New York Times, the Chicago Tribune, the Washington Post, and the LA Times.[5] For each newspaper and each day with a market move of more than 2.5%, human readers search the newspaper's archive for relevant articles published the following day. For example, for the large stock market jump on Tuesday the 29th of October 1974, the readers would search the archive on Wednesday the 30th of October, 1974 for articles. For large market movements that occur on a Friday, both the weekend edition of the newspaper and the Monday edition are searched.

The readers search the archive on a given date for articles the mention phrases like 'stock market', 'wall street', 'S&P', or 'Dow Jones'. The readers select the first article that features the search terms in the title and has relevant terms in the abstract/summary of the article or mentions the previous day's percentage rise or fall in the index in the title. Readers were instructed to avoid summaries, abstracts, digests, etc. (articles <300 words). In an article satisfied these requirements but did not directly discuss the cause of the previous day's movement, additional articles were checked using the procedure define above, excluding the original article.

If none of the search terms, index terms, mention of the rise or fall, or mention of the previous day's market action appeared, then a more in-depth search is undertaken where several articles are read in depth and the most appropriate article chosen. With this procedure, we were able to identify at least one relevant article for every day with a large stock market move in our US post-1926 sample.[6]

[5] For certain exercises, we limit our analysis to results from the Wall Street Journal. This newspaper has the most thorough coverage of financial news and has the most complete and consistent archive back to 1900.

[6] Especially in the earlier half of our sample, the most common article that is selected in the Wall Street Journal was the daily 'Abreast of the Market' column that has been utilized by other researchers for textual analysis. However, in most cases across our sample period, other articles do a more thorough job of highlighting causes of the previous

Readers are assigned to carefully review each article and categorize the article's attribution of the cause of the stock market movement on the previous day. A detailed approach to this coding is laid out in the detailed (110 page) online appendix "Coding Large Daily Financial Market Moves - Data Construction Guide".[7] For each category, a careful definition, as well as several examples from newspaper articles, are provided. In the Appendix, we display samples of the categorical examples from the Data Construction Guide taken from actual jump-day newspaper articles. Each notes the category that should be assigned to that day's article, highlights the relevant portion of the article's text, and gives the reasoning behind the category selection.

In addition, the Data Construction Guide goes on to further define the boundaries between pairs of related categories. As one example, the Data Construction Guide highlights that the 'Monetary Policy & Central Banking' category is distinguished from the 'Macroeconomic News & Outlook' category as follows:

> Some news articles that discuss market reactions to macro developments also discuss the Fed's normal response to the macro development. Generally, we code an article as Macro News & Outlook if it attributes the market move to news about the macro economy. We code it as Monetary Policy & Central Banking if the article attributes the market move to (a) shifts in how the Fed responds to a given macro development or (b) news about unexpected consequences of Fed actions. Take the following two examples:
>
> 1. Macroeconomic News & Outlook example: The market moves because it anticipates or speculates (or sees) that the Fed will respond in its usual manner to news about the macro economy. That is, the market anticipates or speculates that the Fed will respond to macro developments according to a Taylor Rule or other well-defined, well-understood description of the Fed's interest-rate setting behavior.

day's movement. Moreover, other newspapers rarely featured a consistent column across time that discussed market movements.

[7] The categories are: Commodities, Corporate Earnings and Profit, Elections and Political Transitions, Foreign Stock Markets, Government Spending, Macroeconomic News, Monetary Policy and Central banking, Non-Sovereign Military/Terror, Regulation, Sovereign Military/Terror, Taxes, Trade and Exchange Rate Policy, Other Policy, Other Non-Policy, and Unknown.

2. Monetary Policy & Central Banking example: The market moves because of a surprise change in the policy interest rate -- i.e., a surprise conditional on the state of the macro economy. From a Taylor Rule perspective, we can think of this change as a new value for the innovation term in the Taylor rule.

Each day in our sample is assigned a primary categorical cause for the day's large market movement. Many days also are coded with secondary causes, as determined by the weight put on each cause within the newspaper article. Causes that are emphasized in the title or sub-title of the article are given more weight, as are causes that are specifically noted to be the primary driver of the day's large movements. If an article mentions multiple causes but does not clearly denote a primary cause, the readers utilize the order in which the reasons are mentioned or discussed in the article as a tie breaker. Additional reasons (beyond primary and secondary) can be noted in the comment field.

For each primary cause of a market movement, the geographic source was also recorded. For instance, a large market movement in the US driven by a change in the Federal Funds Rate would be attributed to the United States, whereas a large market movement in the US caused by the decision of the UK to leave the gold standard would be attributed to the United Kingdom. Multiple countries may be cited if, for instance, a statement or action was taken by a multinational organization or coalition of countries.

Two additional measures for each article are recorded by the reader. The first is the 'Confidence' with which an article advances an explanation for a given day's market movements. This ranges from a Confidence score of 3 (high confidence) if the article's author directly states that the move was driven by a specific factor, to a score of 1 (low confidence) if the author gives multiple potential reasons, or states that investors and analysts were unsure of the reason for a market movement.

Readers also classify articles based on the 'Ease of Coding', which measures how difficult it was to assign a primary cause to the market movement. The score ranges from 3 (Easy to code) for articles that rapidly and clearly identify the cause of the jumps to 1 (Hard to code) for articles that meander, offer several explanations or are hard to understand. This related to Confidence but is not the same – a journalist may be confident that specific events drove markets on a given day but write an opaque article, or be unsure but state this clearly early in the article.

For the United States, we conducted a thorough cross-validation with multiple coders across multiple newspapers for each article. Each coder followed the coding procedure outlined above, as detailed in "Coding Large Daily Financial Market Moves - Data Construction Guide". After all articles were read, we re-examined days where coders disagreed about the primary and secondary cause of the market movement. This happened more often on days that were also coded as having a lower ease of coding and less confidence by the article's author regarding the driver of the market movement.

To resolve each disagreement, readers re-read the original article and referred to the Data Construction Guide to make sure that the guidelines were being carefully followed. Most disagreements were easily resolved as a reader may have misread an article or misapplied the guidelines from the Data Construction Guide. For articles which still produce disagreement, additional articles in the same newspaper were obtained through the same method as outlined above to seek clarity regarding the primary cause. After these steps were taken, readers still sometimes disagreed regarding some moves that were highly uncertain. For such days, readers could 'agree to disagree' regarding the causes of the stock move and our final dataset reflects such persistent disagreement.

Finally, before analysts started coding, they carefully read the audit guide, underwent a half-day training session and then coded 50 WSJ training articles. These WSJ training articles had already been coded up by us, enabling us to ensure our auditors were accurately coding (and to address any issues) before they coded the research sample.

2.2 Foreign Stock Jumps Data

For the US we choose a threshold of a 2.5% daily change in the stock market to define "jump" days to code. This threshold, which covers about 3% of trading days from 1900-2018, was chosen to be large enough to ensure the next day newspaper always contained articles discussing the prior days jump. When we extended to other countries we usually maintained a 2.5% daily return threshold to classify stock market moves as a significant event. For a subset of countries with more volatile stock-markets we increased the threshold as the stock markets there were more volatile, choosing these thresholds to cover approximately 2-3% of trading days. Appendix Table A1 lays out the threshold, start date, and primary newspaper utilized for each country.

Coders searched the archive of the newspaper of record for a given country (eg. the Globe and Mail for Canada or the Financial Times for the UK). This may take the form of English-language or non-English-language newspaper. If a non-English-language paper was used, a native speaker of that language was used as a coder. As with the coders for the United States, foreign country coders searched for articles on the day following each jump that mention the stock index in question or the stock market more generally. If the date is a Friday or Saturday, Monday's paper would be searched, as well.

3. Big Jumps Over Time and Across Countries

3.1 Stock Jumps Over Time

Using our human coders, we find a significant amount of variation over time in both jump frequency and in the categorical drivers of jumps. Figure 2 displays the evolution of large daily stock market jumps over time in the United States from 1900 to 2018. Also noted are the fraction of daily jumps that are driven by government policy rather than non-policy causes like news about the economy or corporate earnings, as categorized by coders reading the Wall Street Journal. For a relatively small fraction of articles, the cause of the market's movement for a given day cannot be determined by coders reading newspaper articles and a categorization of 'unknown' is utilized (shaded black).[8]

In the figure, we see two particularly notable spikes in the frequency of jumps: the first starting during the Great Depression from the late 1920s until the late 1930s and the second during the Great Recession from 2008-2012. There were also several periods of higher volatility during the early 1900s, with World War I, the Panic of 1907, and other financial panics playing a role. Almost surprisingly, other wars like World War II, the Korean War, and the War in Vietnam produced many fewer large daily jumps in the stock markets. During the post-war period, there are long periods with few daily movements large enough to cross the threshold of our sample.

3.2 Drivers of Stock Jumps

[8] For 5 days early in the sample (all pre-1926), we cannot find an article in the Wall Street Journal related to the previous day's large market movement. This may be driven by measurement error in daily market moves on the part of the DOW-extension pre-1926 when the market was composed of many fewer stocks than today.

Table 1 displays summary statistics regarding the distribution of the categorical causes of these large stock market movements in the United States in the pre-war and post-war period and also from our sample of foreign countries.[9] This shows that not only have the frequency of large stock market movements fluctuated substantially over time, but the causes of these jumps have changed, as well. For instance, in the pre-war period in the United States, agriculture made up a much larger portion of US GDP, driving a larger share of big stock movements. World Wars I and II contributed to the large number of sovereign military jumps. Finally, the New Deal was responsible for the high share of regulation jumps in the pre-war period. In the postwar period, we see that Monetary Policy was relatively more important, which is consistent with the start of regular FOMC meetings in 1981.

The direction of these large stock movements are not distributed evenly across categories: some categories are more likely to be "good news" than "bad news". For example, Monetary Policy jumps can often be attributed to the Federal Reserve responding to a crisis, which is one reason why they are positive on average. In contrast, jumps caused by Sovereign Military Action typically have negative implications for the stock market. These differences across categories contribute to the differential effects on stock market volatility that we explore in Section 6.

From the table we take away two important stylized facts. First, policy news drives a large portion of large daily stock market movements. Over 37% of US jumps are attributed to policy: more than macro (24%) or corporate earnings (11%). Globally, 26% of jumps are attributed to policy. Secondly, large stock moves driven by government policy tend to be positive. For instance, 57% of policy-jumps are positive in the US vas compared to just 42% for non-policy (56% vs 43% globally).

In Appendix Table A3, we investigate the extent to which one aspect of the categorization changes as jumps become larger. In particular, we look at the fraction of jump days that are caused by policy categories for progressively larger stock market movements. We split the days into four bins: 0-0.5% above the big jump threshold (which is 2.5% in the US and most countries), 0.5-1% above the threshold, 1-1.5% above the threshold, and 1.5% or more above the threshold. We find that for positive jumps the fraction of jump days driven by policy

[9] Australia, Canada, China, Germany, Greece, Ireland, Japan, New Zealand, Saudi Arabia, Singapore, South Africa, South Korea and UK. We utilize two separate sets of observations from China, one from the Hong Kong stock exchange and one from the Shanghai stock exchange as these indexes cover different portions of the Chinese economy.

news tends to increase substantially as the jumps get bigger. This reflects the fact that government often steps with stimulative policy action (eg. tax cuts, bailouts, monetary policy relaxation) following a large negative or financial economic shock, often causing large rebounds upwards in the stock market. For negative jumps there is no obvious trend.

3.3 Geographical Source of Jumps

Going beyond the categorical cause of large stock moves, we examine the geographic sources of large jump. Focusing on large stock jumps in non-US countries, we find that, on average, non-US newspapers attribute 27% of jumps to the US – well above the US's 11% share of global GDP. Although the US global share of GDP has been declining, the share of jumps attributed to the US has been increasing – with the time-trend significant at the 5% level.

A time series plot of the fraction of international stock market jumps attributed to the United States and to Europe is displayed in Figure 3. As we might expect, the US share is especially high during US-centric events, like the tech boom/bust, and relatively low during non-US events, like the European Sovereign Debt Crisis. Appendix Table 3 provides more detail about country-level sources of large stock market jumps, showing that the United States is a notable outlier in terms of the fraction of stock market jumps that are driven by domestic events relative to international ones.

3.4 Big Jumps in International Currency and Bond Markets

4. Validation of Human Coder Data

A potential concern is the reliability of human readers in consistently identifying the correct 'category' of the cause for a given large stock market movement. We test for consistency across coders who are investigating a given day's large stock movement by (a) reading articles in the same newspapers and (b) reading articles in different newspapers.

Table 2 examines various dimensions of cross-coder 'agreement' in categorization. First, we examine the average annual pairwise agreement in primary categorization across all pairs of coders who are reading the same newspaper. We find that coders who are reading the same newspaper largely agree on what is driving the large move in the stock market the day before. Overall 75% of coders agree on the precise category (across 16 distinct categories) of the movement's cause, and 89% agree on the policy/non-policy split. For the Wall Street Journal, which we feel has the highest quality financial reporting of all newspapers in our sample, these metrics rise to about 85% agreement in the granular categories, and over 90% agreement on policy versus non-policy explanations.

We extend this exercise to coders who are reading articles from different newspapers about the same large daily stock market movement. Here we see a decrease in the amount of agreement, to about 50% across the 16-categories and about 80% when considering only 'policy' or 'non-policy' as categories. This decrease is likely driven by the fact that, for a fraction of the days we study, the cause is ambiguous, leading to be significant differences in how different reporters write about the previous day's market movements.

Suggestive evidence for this is on days in which the articles have lower reported levels of journalist 'confidence' have lower rates of cross newspaper coder agreement. For example, an increase in average reporter confidence of 1 point (on a three-point scale) increases the rate of coder agreement by over 20%. An increase in the reported ease of coding has an effect of a similar magnitude.

4.1 Information Releases and Stock Jump Categorization

For a subset of categories, we expect that regular information releases drive large stock movements and can use this to test our coding. For instance, we would expect days to be categorized as 'Elections & Political Transitions' more often following elections than for the average jump day. Similarly, we would expect a relationship between Federal Reserve announcements and 'Monetary Policy & Central Banking' categorizations and high profile macroeconomic releases (eg. unemployment numbers and inflation reports) and "Macroeconomic News & Outlook' categorizations.

In Table 3, we demonstrate that these relationships hold statistically, despite coders not directly observing the dates of information releases (i.e. they read only the newspaper article in question and did not separately look up whether the Federal Reserve had made a statement during the previous day). In all cases, the expected categorization is substantially more likely to occur following the public information release.

4.2 High-Frequency Analysis

Another means of validating the accuracy a given day's categorization is to analyze intraday price patterns across the different sets of large stock movements where we have strong priors about what patterns should be observed. These sharp movements in intra-day stock prices tend to be associated with some categories more than others and certain categories tend to drive movements at predictable times *within* a day, as well.

Figure 4 demonstrates variation along these lines. In the top-left panel, we calculate the average fraction of daily returns that have occurred in each 30-minute window of the trading day for all days with more than a 1% return in the S&P 500 from 1986 to 2018. For example, about 28% of a day's total return occurs in the first 30 minutes of the trading day.

Each subsequent panel displays the deviation from these average returns, by 30-minute window, for a range of subsets of trading days. Clockwise from the middle of the top row, these subsets are: Monetary Policy jump days, Unknown jump days, Corporate Profit jump days, Macroeconomic News jump days, and jump days with a foreign (non-US) geographical source.

We note a number of interesting patterns. On average, returns are concentrated in the first 30 minutes of the day. Predictably, most of our subsets see even higher concentrations in these opening minutes. Corporate earnings releases and macroeconomic news are often published in the minutes before the markets open. In addition, most of the foreign-sourced jump days in our sample occur due to events in Europe and the Middle East that take place when markets in the United States are closed and are only incorporated into stock prices when markets open the following day.

Notably, the jumps that occur for Unknown reasons are less likely to have substantial movements at market open and are more evenly distributed throughout the day, making it more difficult to discern a singular cause for the day's return. Finally, we also see a larger than average portion of the day's returns occur in the afternoon for Monetary Policy jump days. This is likely driven by the fact that the Federal Reserve often announces rate changes at 2PM Eastern, yielding heavier trading and larger returns following these announcements.

5. Clarity of Stock Market Jumps

We also wanted to measure the clarity regarding the 'true' cause of a large daily stock market movement. For instance, some jumps are very clear – for example, the interest rate or

European bail-out events show in the top-half of Figure 1 – while others have no clear news that appeared to drive the jump (e.g. Black Monday in 1987 or the “Boxing Day” jumps in 2018). We propose four proxies of clarity, and combine these into an overall “Clarity Index”:

i. Confidence and Ease of Coding: When reading the newspaper, each coder reports (1) How confident the journalist was about the cause of the jump (2) How easy/difficult it was for them to code the article. On days with a clear cause, we expect both the journalist confidence, and the ease of coding to be high. On days driven by narratives, the journalist might list several possible explanations, and the coder might have trouble linking the explanations given to the stock move. On each day, we measure the average confidence, and the share of coders who gave the article a maximum confidence or ease of coding score.

ii. Agreement Across Newspapers: Consider all possible coding pairs for a given jump. (For example, if we have codings by persons 1, 2 and 3, then there are three pairwise codings: (1,2), (1,3) and (2,3). For each pairwise coding, set a measure of agreement $A_{ij}=1$ if i and j agree on the coding, and 0 otherwise. Then compute overall mean pairwise agreement = Sum A_{ij} / N, where the sum is over all i and j for i not equal to j, and N is the number of possible pairwise codings on the data. We expect agreement across newspapers to be lower if the cause of the jump is less clear – each paper may have their own narrative.

iii. Agreement Within Newspapers: Use the same agreement measure constructed above but calculate the average within each newspaper for each jump. Then average this value over newspapers to obtain the Average Newspaper Pairwise Agreement. We expect agreement within newspapers to be high if the cause of the jump is clear.

iv. Number of Unknown Codings: For each coder j, set $Un_j = 1$ if the primary category code is Unknown, zero otherwise. Compute the mean value of Un_j over coders to obtain the Unknown Cause rate for the jump. A higher unknown rate is less likely tied to discount rates or cash flow news.

Figure 4 plots these four measures over time, showing in all cases a rise in clarity over time (the “share of unknowns” is a reverse clarity measure). We can also combine these into a ‘clarity index’ by normalizing each measure to a z-score (mean zero and standard deviation one) and averaging and then re-normalizing.[10] Figure 5 plots this overall clarity index, showing a rise until about 1980 and then an approximately flat index thereafter. This rise presumably reflects in part the increase in the quality of economic data and financial reporting, but potentially also the increased enforcement of trading rules against market manipulation.

One notable contrast is seen in the two largest financial crises during our sample period. The Great Depression features some of the lowest levels of clarity of jump cause in our sample, while the Great Recession contains some of the highest levels of clarity. Despite both periods possessing extremely high levels of financial market volatility, most of the largest movements during the Great Recession were clearly attributable to a particular cause, while most of the largest movements in the Great Depression were fairly ambiguous. Intriguingly, clarity has also fallen rapidly post 2016 under the Trump administration.

At the level of an individual jump, the clarity index tends to be higher when we would have a strong prior about the cause of the large market movement due to a predictable release of information by a significant government body. For instance, we look at large daily jumps near days that occurred on the day of, or the day after an FOMC Meeting, were an election day, or had any data release of National Income and Product Accounts. In the post-1994 period, when the FOMC started issuing a press release after meetings indicating changes in the policy rate (Gurkaynak et. al. (2005)), the clarity index is approximately two standard deviations higher than average for jumps on FOMC announcement dates.

To account for the fact that most elections are decided after trading is over, we look at the clarity index on the day of, and the day following US House, Senate and Presidential Elections. Similarly, to account for releases that occur after trading hours, we look in two-day windows around NIPA releases. The clarity index is higher than average on these days, but it is not statistically significant owing to a small sample size (only 4 jumps near election days in our sample period and 15 NIPA data releases).

In Table 4, we perform a number of regressions spanning data from 1990-2018 that examine the correlation between our clarity index and aspects of the stock market on a jump day. In all columns, in addition to our clarity index, we include the absolute value of the daily return

[10] Our results are robust to using a principle component analysis on the complete time series, and take the first component, which explains almost 60% of the variance of the individual pieces.

interacted with indicators for the return being positive or negative, to allow for asymmetric effects. We also include year, month and day of the week fixed-effects to account for predictable differences in the dependent variables over time.

First, we examine the relationship between clarity and intra-day stock market volatility using high-frequency data from the S&P 500 from January 1990 to January 2015. For each day, we calculate returns in 5-minute intervals, with the first window being 9:30AM to 9:35AM. The final window the period between 3:55PM – 4:00PM. The 5-minute returns are calculated as the percentage change of the closing price in window *t* relative to the closing price in window *t-1*. We find that greater clarity is associated with lower intraday volatility, as measured by the sum of squared 5-minute returns for the S&P 500. In column 2, we find that days with higher levels of clarity also tend to have lower volume (here measured as the daily trading volume for the SPY, the largest S&P 500 ETF).

In column 3, we look how our clarity index is related to the fraction of total daily market movements (i.e. sum of total distance travelled in 5-minute increments) that occur in the single 5-minute window with the largest absolute return. We find that a higher level jump clarity is positively related to the relative concentration of the daily market's movement. Finally, column 4 shows that our clarity index also predicts the daily change in the VIX.

Overall, it seems that days with sudden bursts of trading in a single direction tend to be the most 'clear', while days that vacillate back and forth throughout the day in heavy trading tend to be difficult to code using our methodology. Moreover, as we demonstrate in the following section, these differences in stock market behavior are correlated with clarity not only on the day of a given large stock market jump, but are persistently different for weeks and months, as well.

5.1 Algorithmic Jump Classification

Given the costs and time involved with running large-scale human evaluations in order to accurately code hundreds or thousands of individual daily stock market movements, it may be natural to attempt to approach the question using automated textual-analysis tools.

To work towards an automated classification algorithm, we aim to 'rank' the most likely categories for each day in an automated fashion based on the raw text of the newspaper articles that were used by our human coders.[11]

We start by OCRing the full text of each Wall Street Journal (WSJ) article. Unlike our other newspapers, we only have 1 WSJ article per day, as part of an experiment to explicitly measure differences among coders reading the same articles in the same paper, rather than reading different articles from the same paper. For most supervised machine learning algorithms, we would like to have exactly one category per day in the training sample. For days where the WSJ coders agreed, this is straightforward. If they disagree, however, we take the category with the highest average score among categories, if the highest average score is above a certain threshold. In this subsection, we make that threshold 0.5, so at least one coder must assign it a lone primary and the other must assign it at least as a secondary category. If no category on a given day crosses this threshold, that day is dropped from the sample.

We then clean the articles by removing all (1) non-english words, which are usually OCR errors from early in the sample when the scanned articles are of lower quality (2) words that are parts of headers/footers generated by ProQuest when the articles are saved as PDFs (3) stop words using the NLTK toolbox in Python. We then do additional cleaning based on the algorithm in Loughran and McDonald (see https://sraf.nd.edu/textual-analysis/resources/ for detailed notes on their cleaning procedure) to make the punctuation meaningful, making it easier to break the document into sentences. Finally, we use the Porter Stemmer to reduce all words to their stems.

After cleaning the articles, we extract the first 200 words of each article. This has two main benefits: (1) It makes all the articles the same length, which is useful when doing tf-idf to prevent biases caused by differences in document length and (2) many articles, especially early in the sample, discuss several topics, including those unrelated to the jump. The first 200 words are usually the most relevant for categorizing the article. Finally, we require that words appear in a category at least 3 times, and overall at least 5 times.

Having cleaned the text data, we compute a tf-idf score for each word in each document. tf is computed within an article, while idf is computed across all articles that survive the filters

[11] For this exercise, we restrict our analysis to the Wall Street Journal, for which we can access the text of each article back to 1900.

described above. We then use these scores to perform a ‘leave-out-one’ classification of each article. To do this, we take the entire corpus, excluding the article we are trying to classify. We then take all the unique words in those articles, and sort on the average tf-idf score for these words across articles in each human-classified category. Finally, we take the top 100 words for each category from this sorting: these are the words we associate with each category. For example, for Commodities the top word is ‘wheat’, while for Sovereign Military Action the top word is 'germani' (stem of Germany).

Having identified the top words for each category, we add up the tf idf score for the words in each category for the article we are trying to classify, and rank categories by these sums. The category with the highest sum will be given rank 1, second highest rank 2, etc. Overall, our average ranking of the true category is 2.5 across our entire 1900-2018 sample. So, while we typically cannot identify the true category, it is generally ranked more highly than would be achieved through random guessing.

One primary concern with our mechanical approach is the substantial evolution in language utilized in newspaper articles across the years of our sample. To analyze the degree to which this issue decreases the accuracy of our mechanical classifiers, we split our sample into four periods, each containing one fourth of the total jump days in the United States since 1900: 1900-1931, 1932-1939, 1940-2007 and 2008-2017. Within each time period, chose categories that appeared at least 5 times. We repeat our ranking classifier on each sub-sample using a leave-one-out methodology for out of sample categorizations. We find that splitting the sample by period tends to improve fit significantly, despite losing the information that additional for each category articles can provide.

In addition, we find that the sub-samples are able to be categorized more accurately over time. As seen in Figure 7, while the oldest sub-sample tend to see an average ranking of approximately 3, the most recent sub-sample has an average ranking of approximately 1.5 (relative to a best-possible ranking of 1). This reflects the tendency for more recent articles to be written in a clearer and more focused fashion, allowing for greater differentiation between articles in terms of the cause for the day’s stock move. This tendency mirrors the evolution of our other measures of human-coded ‘clarity’ over time, showing that automated classification reveals a similar increase.

5.2 Barriers to Algorithmic Jump Classification

There are a number of reasons to be wary of an automated approach to jump day classification, at least when starting with the blank slate of a simple database of newspapers and stock market movements.

The first potential issue is simply that when aiming to categorize daily stock market movements into recognizable and detailed categories, the lack of a training sample already categorized in this way inhibits most standard machine learning approaches. That is, using no other input, Latent Dirichlet Allocation (LDA) (see Blei et. al. (2003)) can separate newspaper articles into N distinct topics composed of different weights on different sets of terms, but these may not be able to be mapped to categories that humans may find useful or applicable for further analysis. For instance, researchers may be interested specifically in understanding how trade policy drives large stock market movements, but a computer may not isolate this particular category as a distinct factor, especially given the small number of large stock movements driven by trade policy over the 21th century.

This problem is compounded when focusing primarily on large stock market movements. Such a restriction reduces sample sizes considerably and makes any automated approach more prone to issues of overfitting, especially when attempting to isolate a number of rare and distinct categories of events. As one example, one may attempt to gain granularity by increasing the number of dimensions to attempt to fit over (eg. moving from single words to 2-grams or n-grams in order to separate 'war' from 'trade war' or 'deficits' from 'trade deficits'), but decreasing the generalizability of the resultant classification system out-of-sample. While the automated system may perform well when automating the bifurcation of stock moves into two types of explanations, attempting to split the data into 10-20 categories that exhibit hugely different base rates tend to produce substantial Type 1 and Type 2 errors.

The issues arising from relatively small samples of events is amplified by the fact that the language employed by journalists and members of the financial industry have changed significantly over time. The choice of words that describe a large stock move caused by 'Corporate Earnings' or 'Trade & Exchange Rate Policy' can widely vary depending on whether the day in question was in 1910 or 2010. This is due both to changes in common phrases and terminology over time but also to the fact that the institutional framework of business, government, and financial markets has changed substantially in the past century. These changes span the creation of the Federal Reserve, the creation or end of different countries, the end of the gold standard, the rise and fall of new industries, and the broad innovations in financial reporting requirements and new trade agreements spanning the globe.

We use our Wall Street Journal codings as the training sample for a Naïve Bayes Classifier (see, for example, Russell and Norvig (2003)). To reduce overfitting, we follow the same procedure described above when constructing the category ranking. The main filters include removing stop words, words that appear in fewer than 5 articles, and words that appear in more than 70% of articles (ie. those with low signal-to-noise ratio). In-sample, the algorithm can fit nearly 100% of articles, but allowing this amount of flexibility may drive overfitting issues. To test for over-fitting, we measure the model's out of sample performance. For each day, we fit the Bayes Classifier on all other days and then pass that day's article into the classifier. To account for differences in base rates across categories, we restrict classification among those categories with a sufficiently large sample and similar base rates: Corporate Profits, Government Spending, Macroeconomic News & Outlook, Monetary Policy and Sovereign Military Actions. Although there are a significant number of jumps classified as Unknown, we omit this category, as it adds a noise to out of sample classifications. With this approach, we fit 63% of articles. On average, the Bayes Classifier works better out of sample than randomly picking categories from the unconditional distribution (which would achieve a match rate of 31%), but the fit is far from perfect.

Automated categorization is in part limited to the quality of the PDF files being converted to text. Earlier years (eg. pre-1940), in particular, suffer from poor image quality which results in less-than-perfect translation into machine-readable text. For this reason, we also perform our analysis with only data from 1984 to the present, the period in which we can obtain the text of the relevant article directly.

Restricting to the post-1984 sample slightly improves the fit, but this reveals a significant problem: because many of the categories are sparse, the model almost always guesses the modal category of 'Macroeconomic News & Outlook'. As discussed above, while it is possible to improve the out of sample fit by stemming words and trying to identify 'relevant' pieces of long articles (especially in the pre-World War II period), there is a limit to how good the out of sample fit can be with the 'bag-of-words' approach.

6. Volatility and GDP Following Stock Market Jumps

We have documented the fact that the categorical causes and geographic origins of stock market movements vary across countries and have changed substantially over time. We now turn to the question whether these categorical differences in what drives large stock market movements can predict future differences in financial or real variables.

6.1 Differences in Volatility by Jump Category

Looking first at financial markets, we find that, for a given size of stock market move, the reasons behind the move have systematically different implications for realized market volatility in the following days and weeks. Here we measure realized volatility over an *n* day horizon as the sum of squared returns on the CRSP Value-Weighted index over those *n* days. We use the uncentered second moment to avoid the difficulties inherent in measuring the mean stock return over a short horizon.

While all jump days lead to elevated levels of volatility, we test whether some types of jumps have more persistent effects than others, utilizing the following regression approach:

$$\mathbf{100}\sum_{i=1}^{n}\frac{r_{t+i}^{2}}{n} = a + b\left(r_t \times \mathbf{1}_{r_t>0}\right) + c\left(|r_t| \times \mathbf{1}_{r_t\leq 0}\right) +$$

$$d\left(r_{t-1}^{2}\right) + e\left(\sum_{i=1}^{5} r_{t-i}^{2}\right) + f\left(\sum_{i=1}^{22} r_{t-i}^{2}\right) +$$

$$g\ macro_t + h\ monetary_t + i\ other_t + Fixed\ Effects + e_t$$

r_t is the return on the CRSP value-weighted index. The left-hand side term is the average realized volatility over an n-day horizon. The first set of right-hand side variables are controls for the day's return, and allowing for an asymmetric effect of positive and negative returns on volatility (Black (1976)). The second set of RHS variables are 'HAR' controls to account for the effect of volatility over different horizons on future volatility (Corsi 2009). The last set of RHS variables represents our jump categories. For example, $macro_t$ will take the value 1 if all coders reading the article on that day classified the article as Macroeconomic News & Outlook, and will take the value 0 if no coders assigned the article Macro News. For days with disagreement between coders as to the primary category, the variable will take a value between zero and one.[12]

We find strong evidence that large stock jumps driven by macroeconomic news produce realized volatility substantially higher than those driven by monetary policy. We plot coefficients from this regression in Figure 7, looking at the 44 trading days (an approximate two month period) after a jump day. We hypothesize that some of these differences are driven by the fact that some types of events, such as a bad unemployment report, may *generate* uncertainty while others, such as monetary policy announcements about a rate change, may *resolve* uncertainty. These differences are economically significant, with volatility being 0.8 standard deviations above its mean 22 days after a negative monetary jump, less than the 2.3 standard deviations above the mean we observe after a negative macro jump.

22 Days After the Jump

Jump Type	Sd. Above Mean
Positive Monetary	0.553
Negative Monetary	0.826
Positive Macro	1.811
Negative Macro	2.293

The differences seen in Figure 7 may be driven in part by differences in the average sign and size of the different categories. We control for such differences by interacting jump days with indicators for positive or negative signed returns and plot the results in Appendix Figure A4, finding effects that are jointly significant for both types of jump days. Similarly, we test whether the difference between monetary and macroeconomic jump day follow-on volatility is driven by the state of the business cycle, adding interactions with indicators for NBER recessions. Both extensions retain significant differences between macroeconomic jump days and monetary policy jump days. The results of these are plotted in Appendix Figure A5. We find that while we do see higher volatility following macroeconomic jump days in both recessions and expansions, the difference is significantly larger during recessions.

Differences in post-jump volatility are not limited to jumps caused by macroeconomic news and monetary policy announcements. We see broader differences across policy-related jump days and non-policy-related jump days, as tested by interacting policy and non-policy indicators with the returns on the jump days. Table 5 displays results breaking down jump days

[12] Fixed effects include decade indicator variables, as well as a NBER recession indicator variable, though results are robust to year fixed effects instead of decade dummies.

along these lines, finding that, while both types of jump days increase volatility, jump days caused by policy reasons tend to increase volatility less than those caused by non-policy reasons.

6.2 Jump Clarity and Volatility

In addition to the differential effects of various categories of jump causes, we examine the in Table 6 for the clarity of a particular day's jump to impact future volatility using our clarity index. We show in column (1) for the overall clarity index, and in columns (2) to (5) for each of the sub-components that lower clarity is followed subsequently by higher volatility. This is consistent with Carlin, Longstaff and Matoba (2014) who find that increases in disagreement predict future realized volatility. We believe that clarity and disagreement may be related, as (1) one of the inputs is about differences in explanations for the jumps across newspapers (2) if confidence and/or ease of coding are low, different people reading the same article may have different interpretations.

These effects are persistent, as seen in Appendix Figure 6. Here we display a time series graph of coefficients from a regression of daily volatility following jumps, splitting into high-clarity and low-clarity (above and below median) jumps.[13] Given that low clarity jumps may be more likely when volatility is high, we include HAR controls as robustness.[14] We show that volatility is relatively higher after low clarity events than high clarity events.

6.3 Stock Jumps and GDP

Figure 8 plots the impulse response function from VAR with GDP, the number of negative macro jumps in a quarter, and number of positive macro jumps in a quarter. We include 12 lags of each variable, as selected by the AIC and BIC.[15] The figure displays an innovation of 1-standard deviation in the number of positive or negative jumps. We find that an increase in the

[13] Realized volatility is the sum of squared returns on the CRSP value-weighted index over the five days following the jump (does not include the jump day). The regression is telling us about partial effects, but we also want to understand the general relationship between clarity and volatility. Even without all the controls and fixed effects, there is a negative relationship between clarity and realized volatility over the next week.
[14] Addition controls include decade fixed effects and a NBER recession indicator variable. Results are robust to year fixed effects instead of decade dummies.
[15]GDP data is from FRED, so our sample here is restricted to 1947-2018. Standard errors feature the small-sample correction (with the result being robust to including cumulative within-quarter stock returns).

number of negative macro jumps are followed by an economic decline, while an increase in the number of positive macro jumps is followed by an expansion in GDP.

This pattern, however, is particular to jumps attributed to macro news. For other types of jump days, both positive and negative stock movements tend to depress GDP growth in the near term. Figure 9 shows the impulse response function for non-macroeconomic jumps. After both negative- and positive-signed jumps, we see a short-run decrease in output – consistent with volatility (even if it is associated with ‘positive’ jumps) being bad news (eg. Muir and Moriera (2017)).

7. Conclusion

We examine newspapers the day after major stock-market jumps to catalog the proximate cause, geographic source, and clarity of these events from 1900 in the US and 1980 (or earlier) in 13 other countries. We find three main results. First, the United States plays an outsized role in global stock-markets, accounting for 35% of jumps outside the US since 1980s, far above its 15% share of GDP. This matches other evidence on the dominance of the US in global finance. Second, the clarity of the cause of stock market jumps has been increasing since 1900, presumably because news and financial markets has become more transparent. Jump clarity predicts future volatility: doubling the clarity index of a jump reduces future volatility by 68%. Third, jumps caused by non-policy events (particularly macroeconomics news) lead to higher future stock-volatility, while jumps caused by policy events (particularly monetary policy) reduce future stock-volatility. This suggests while monetary policy surprises lead to stock-market jumps, they may reduce future volatility.

Bibliography

Barro, Robert J. "The stock market and investment." The Review of Financial Studies 3.1 (1990): 115-131.

Birz, Gene, and John R. Lott Jr. "The effect of macroeconomic news on stock returns: New evidence from newspaper coverage." *Journal of Banking & Finance* 35.11 (2011): 2791-2800.

Black, Fischer, "The Dividend Puzzle". The Journal of Portfolio Management Winter 1976, 2 (2) 5-8

Blei, David, Ng, Andrew, and Jordan, Michael. "Latent dirichlet allocation". The Journal of Machine Learning Research, Volume 3, 3/1/2003; Pages 993-1022

Bollerslev, Tim. "Generalized autoregressive conditional heteroskedasticity." Journal of econometrics 31.3 (1986): 307-327.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson. "ARCH models." Handbook of econometrics 4 (1994): 2959-3038.

Boudoukh, Jacob, et al. *Which news moves stock prices? a textual analysis*. No. w18725. National Bureau of Economic Research, 2013.

Boz, Emine, Gita Gopinath, and Mikkel Plagborg-Møller. *Global trade and the dollar*. No. w23988. National Bureau of Economic Research, 2017.

Brogaard, Jonathan, and Andrew Detzel. "The asset-pricing implications of government economic policy uncertainty." *Management Science* 61.1 (2015): 3-18.

Campbell, John Y., et al. "Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk." The Journal of Finance 56.1 (2001): 1-43.

Carlin, Bruce I., Longstaff, Francis A. and Matoba, Kyle, (2014), Disagreement and asset prices, Journal of Financial Economics, 114, issue 2, p. 226-238.

Corsi, Fulvio "A Simple Approximate Long-Memory Model of Realized Volatility" Journal of Financial Econometrics, Volume 7, Issue 2, 1 March 2009, Pages 174–196

Croux, Christophe, and Peter Reusens. "Do stock prices contain predictive power for the future economic activity? A Granger causality analysis in the frequency domain." *Journal of Macroeconomics* 35 (2013): 93-103.

Cutler, David M., James M. Poterba and Lawrence H. Summers. "What Moves Stock Prices?" The Journal of Portfolio Management Spring 1989, 15 (3) 4-12

Dougal, Casey, et al. "Journalists and the stock market." The Review of Financial Studies 25.3 (2012): 639-679.

Ehrmann, Michael, Marcel Fratzscher, and Roberto Rigobon. "Stocks, bonds, money markets and exchange rates: measuring international financial transmission." Journal of Applied Econometrics 26.6 (2011): 948-974.

Engelberg, Joseph E., and Christopher A. Parsons. "The causal impact of media in financial markets." The Journal of Finance 66.1 (2011): 67-97.

Engle, Robert F. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." Econometrica: Journal of the Econometric Society (1982): 987-1007.

Fama, Eugene F. "Stock returns, real activity, inflation, and money." The American economic review 71.4 (1981): 545-565.

Fischer, Stanley, and Robert C. Merton. "Macroeconomics and finance: The role of the stock market." (1984).

Fernandez-Perez, Adrian, Bart Frijns, and Alireza Tourani-Rad. "When no news is good news–The decrease in investor fear after the FOMC announcement." *Journal of Empirical Finance* 41 (2017): 187-199.

Fisher, Adlai J., Charles Martineau, and Jinfei Sheng. "Media attention, macroeconomic fundamentals, and the stock market." *University of British Columbia Working paper* (2017).

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. "On the relation between the expected value and the volatility of the nominal excess return on stocks." The journal of finance 48.5 (1993): 1779-1801.

Goldberg, Linda S., and Christian Grisse. *Time variation in asset price responses to macro announcements*. No. w19523. National Bureau of Economic Research, 2013.

Gopinath, Gita, and Jeremy C. Stein. *Banking, Trade, and the making of a Dominant Currency*. No. w24485. National Bureau of Economic Research, 2018.

Guo, Hui. "Stock market returns, volatility, and future output." *Review-Federal Reserve Bank of Saint Louis* 84.5 (2002): 75-84.

Gurkaynak, Refet; Brian Sack; Eric Swanson, 2005. "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," Macroeconomics 0504013, University Library of Munich, Germany.

Kelly, Bryan, Lubos Pastor, Pietro Veronesi “The price of political uncertainty: Theory and evidence from the option market”. Journal of Finance 71, 2417-2480, 2016.

Keynes, John Maynard, 1883-1946. The General Theory of Employment, Interest and Money. London :Macmillan, 1936. Print.

Li, Feng. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and economics*45.2-3 (2008): 221-247.

Lucca, David O., and Emanuel Moench. "The pre-FOMC announcementdrift." *The Journal of Finance* 70.1 (2015): 329-371.

Maggiori, Matteo, Brent Neiman, and Jesse Schreger. *International currencies and capital allocation*. No. w24673. National Bureau of Economic Research, 2018.

Manela, Asaf, and Alan Moreira. "News implied volatility and disaster concerns." *Journal of Financial Economics* 123.1 (2017): 137-162.

Mehl, Arnaud. "Large global volatility shocks, equity markets and globalisation: 1885-2011." (2013).

Muir and Moriera (2017) “Volatility-Managed Portfolios”. Journal of Finance. Vol 72, Issue 4, Pg 1611-1644

Niederhoffer, Victor. The Analysis of World Events and Stock Prices. The Journal of Business, 1971, vol. 44, issue 2, 193-219. Date: 1971

Obstfeld, Maurice. "Trilemmas and trade-offs: living with financial globalisation." (2015).

Pastor, Lubos, and Pietro Veronesi. "Uncertainty about government policy and stock prices." *The Journal of Finance* 67.4 (2012): 1219-1264.

Pastor, Lubos & Veronesi, Pietro, 2013. "Political uncertainty and risk premia," Journal of Financial Economics, Elsevier, vol. 110(3), pages 520-545.

Roll, Richard “R2” Journal of Finance, Volume 43, No. 2, July 1988

Russell, Stuart, Peter Norvig, “A Modern Approach” Artificial Intelligence. Prentice-Hall, Englewood Cliffs 25 (27), 79-80

Shiller, Robert, (1981), Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?, American Economic Review, 71, issue 3, p. 421-36.

Shiller, Robert J. "Narrative economics." *American Economic Review* 107.4 (2017): 967-1004.

Tetlock, Paul C. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance* 62.3 (2007): 1139-1168.

Appendix

A.1 Industry-level Excess Returns

As a final way to validate our jump categorization, we can measure the differential response of industry-grouped stock portfolios to jumps of different categories. In general, we would assume that these industries should be differentially sensitive to drivers of market movements of certain types. For instance, we would expect that banking stocks would respond more favorably than the average stock when favorable news about bank bailouts is released during the Global Financial Crisis.

To perform this test, we first obtain daily portfolio returns for 49 broad industry groupings. We utilize the detailed explanation provided by each coder in addition to the primary categorical classification to map to industry groupings. For each jump day, we define the variableTri_{it}. This variable is defined as equal to 1 if the jump description implies an amplified response of *i*, -1 if the jump description implies a dampened response of *i*, and 0 otherwise. Many jumps do not map readily to a single industry, and we sometimes assign two industries to a particular jump (eg. guns and aerospace). We end up with 115 jump X industry observations out of 339 jumps that span from 1960 to 2016.

We then can test for a relationship for a given industry or pooled across all industries, with the specification:

$$R_{it} = \sum_i \alpha_i + \sum_i \beta_i MR_t + \sum_i \delta_i Tri_{it} + \gamma Tri_{it} MR_t + \epsilon_t$$

where R_{it} is the daily return for industry portfolio *i* on day *t* and MR_t is the daily return on market portfolio on day *t*.

Appendix Table A4 displays the results of this analysis, with the results appearing to be in-line with our expectations – industries connected to particular categories of jumps see their returns substantially amplified (diminished) on positive (negative) jump days coded as that category.

A.2 Drivers of Monetary Policy Jumps

To better understand the triggers of our monetary policy category, we run variations of the following regression:

$$\begin{gathered} 100\ \times Monetary_{it} = \alpha + \beta_1\ 1_{\{FOMC_t==1\ OR\ FOMC_{t-1}==1\}} + \\ \beta_2\ RER6W_t + \beta_3\ RER6W_t \times 1_{\{FOMC_t==1\ OR\ FOMC_{t-1}==1\}} + \\ \beta_3\ ICR_t + ICR_t\ \times\ \beta_4\ 1_{\{FOMC_t==1\ OR\ FOMC_{t-1}==1\}} + \\ \beta_5\ PVOL_t + \beta_6\ PVOL_t\ \times\ 1_{\{FOMC_t==1\ OR\ FOMC_{t-1}==1\}} +\ \epsilon_t \end{gathered}$$

Where $1_{\{FOMC_t==1\ OR\ FOMC_{t-1}==1\}}$ is an indicator variable for a scheduled FOMC meeting at *t* or *t-1*. $RER6W_t$ is the average return on the CRSP value-weighted index over the 6 weeks preceding time *t*. ICR_t is total initial jobless claims over the past six weeks (excludes the week containing t), divided by BLS nonfarm payroll employment.[16] $PVOL_t$ is the sum of squared returns over the past six weeks, a measure of past realized volatility. The left-hand-side variable is our Monetary Policy jump category, which is multiplied by 100 to make the coefficients easier to interpret.

Relative to the previous validation results, we include real-side variables, in addition to scheduled FOMC announcements, as predictors of MP jumps. It's challenging to use real-side variables for this purpose, because (a) most real-side variables become available with a lag in real time, (b) they are subject to later revisions, further complicating the task of replicating the Fed's real-time information set, and (c) they are not issued at frequent intervals.

There is, however, one important real-side variable that suffers from none of these problems: initial claims for unemployment insurance benefits. The Fed and financial market participants follow this measure closely, especially when recession risk is high, because it's one of the best nonfinancial real-time early warning indicators of a downturn.

We run two versions of this regression: (1) Using all days in our sample (2) Using only jump days in our sample. The first specification tells us how (a) past average returns (b) past real-side data and (c) past average volatility affect the likelihood of Monetary Policy jumps. The second specification answers the same questions, but conditional on a jump occurring. We also run a version of the regression where we split the jump samples into days with positive and negative returns. Appendix Table A5 displays the results of this analysis. Overall, we find that MP jumps are more likely following low average returns and periods of high volatility.

16 We interpolate the monthly BLS data to weekly as follows: for day n in a 30-day month, we set the nonfarm payroll employment figure to (n/30) * nonfarm payroll employment in the current month + (30-n)/30* nonfarm payroll employment in the next month. We only use data from 1981 onward, 1981 is the first year with scheduled FOMC meetings.