Firm-level Risk Exposures and Stock Returns in the Wake of COVID-19

PRELIMINARY AND INCOMPLETE: DO NOT CITE OR CIRCULATE

Steven J. Davis[1], Stephen Hansen[2], and Cristhian Seminario-Amez[3]

[1]University of Chicago Booth School of Business and Senior Fellow, Hoover Institution

[2]Imperial College Business School and CEPR

[3]Kenneth C. Griffin Department of Economics, University of Chicago

September 14, 2020

Abstract

Firm-level stock returns differ enormously in reaction to COVID-19 news. We characterize these reactions using the *Risk Factors* discussions in pre-pandemic 10-K filings and two text-analytic approaches: expert-curated dictionaries and supervised machine learning (ML). Bad COVID-19 news lowers returns for firms with high exposures to travel, aircraft production and social distancing – directly and via downstream demand linkages – and raises them for firms with high exposures to healthcare policy, e-commerce and drug trials, among other effects. Monetary and fiscal policy responses to the pandemic strongly impact firm-level returns as well, but in ways quite distinct from pandemic news. Despite methodological differences, dictionary and ML approaches yield remarkably congruent return predictions. Importantly though, ML operates on a vastly larger feature space, yielding richer characterizations of risk exposures and greatly outperforming the dictionary approach in goodness-of-fit. By integrating elements of both approaches, we uncover new risk factors and sharpen our explanations for firm-level returns. To illustrate the broader utility of our methods, we also apply them to explain firm-level returns in reaction to the March 2020 Super Tuesday election results.

Acknowledgements: We thank the Becker Friedman Institute for Research in Economics for financial assistance. Davis also thanks the U.S. National Science Foundation (SES 1324257).

Hansen gratefully acknowledges financial support from the European Research Council (Consolidator Grant 864863).

1 Introduction

The economic disruption wrought by COVID-19 is unprecedented in modern times. Real GDP fell 11 percent in the United States and 15 percent in the Euro area in the first half of 2020, easily the largest drops since World War II. The International Monetary Fund forecasts the sharpest global output contraction on record in 2020.[1] Because its effects differ so greatly across sectors and firms, the COVID-19 shock is also likely to drive large-scale reallocation activity. Recent survey evidence (Barrero et al., 2020) and historical evidence (Davis and Haltiwanger, 1992) suggest that most of the reallocation response will involve shifts across firms within industries.

Indeed, firm-level stock returns differ enormously in reaction to COVID-19 news and policy responses. To highlight this point, Figure 1 plots daily market-level returns against the same-day interquartile range of firm-level returns on all trading days in 2019 – and on 20 "jump" days from 24 February to 27 March 2020 when the market rose or fell by at least 2.5%. Jump days show an extraordinary dispersion in firm-level returns. On 18 March, for instance, the 75th percentile stock had a one-day return advantage of 6.9 percentage points over the 25th percentile, more than 15 standard deviations greater (in 2019 units) than the average cross-firm IQR in 2019. Recent earnings announcements also underscore the asymmetric impact of COVID-19, with Amazon and Facebook reporting Q2 2020 revenue growth of 40% and 11%, respectively, both large upside surprises amidst a bleak earnings outlook for many firms.

We use the discussions of *Risk Factors* in pre-pandemic 10-K filings to characterize firm-level risk exposures, explain firm-level equity returns on jump dates, and interpret the drivers of those returns. The basic idea is simple: When the language firms use to describe their risks explains their stock price reactions to news about the pandemic (or other news), it reveals information about the channels through which the market expects the pandemic to affect their future earnings. We implement this idea in multiple ways. We focus on jump dates, because the news event that drove market reactions on

[1] See series GDPC1 (United States) and CLVMEURSCAB1GQEA19 (Euro area), both retrieved from FRED on 16 August 2020, and (IMF, 2020) for the global output forecast.

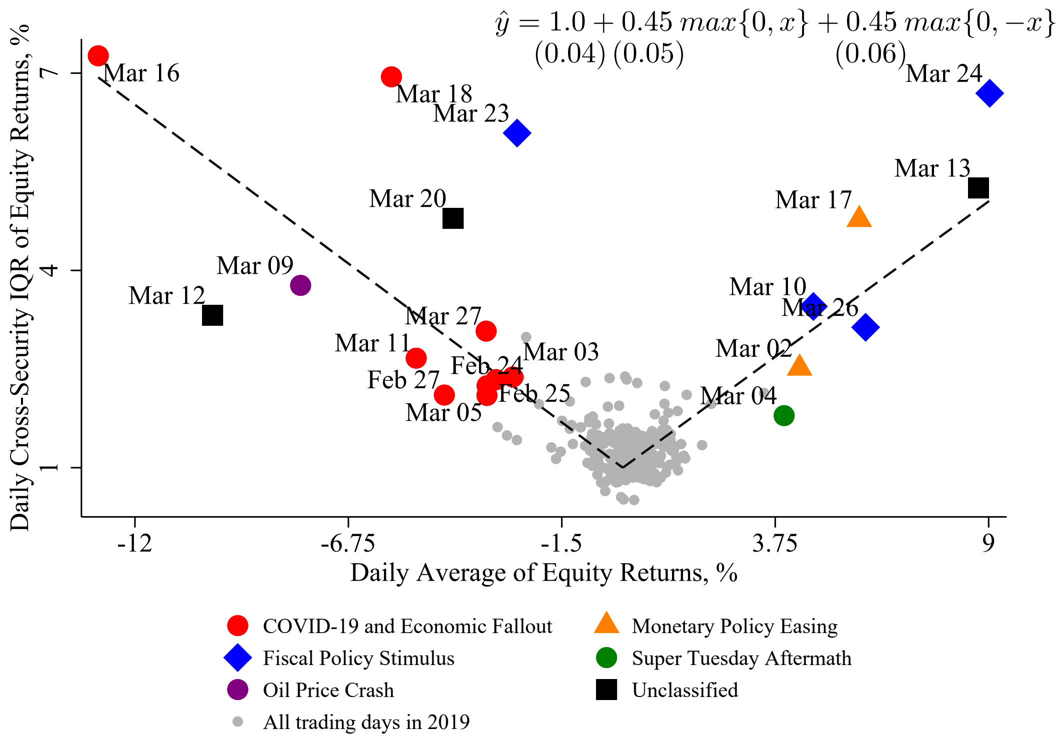


Figure 1: Value-Weighted Mean and Cross-Sectional IQR of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

We consider the value-weighted distribution of daily returns over 2,155 common stocks for all trading days in 2019 and all 20 jump days in 2020 through 27 March. The mean (s.d.) daily average return in 2019 is 0.12 (0.80) percent, and the mean (s.d.) of the daily IQR is 1.29 (0.36). The regression has 271 observations and an R-squared of 0.66, with standard errors in parentheses. A test of the null hypothesis that the two rays have equal slopes with opposite signs yields a p-value of 0.99. Jump classifications follow Baker et al. (2020a), who rely on human readings of next-day newspaper accounts.

those dates is usually apparent. In this regard, we rely on the classifications in Baker et al. (2020a), who consult next-day articles about the stock market jump in the *New York Times* and the *Wall Street Journal*. For the vast majority of COVID-era jump dates, the two newspapers advance a common explanation for the jump.

Our results show that the text in 10-K filings contains highly granular, quantifiable information about the forces that drive firm-level returns. For example, bad news about COVID-19 lowers returns for firms with high exposures to travel, aircraft production and social distancing – directly and via downstream demand linkages – and raises them for firms with high exposures to healthcare policy, e-commerce and drug trials, among other effects. We also find that the structure of firm-level return reactions differs systematically with the type of news that drove the market, as captured by the jump classifications. For example, on jump dates attributed to monetary policy easing, firm-level returns depend on exposures to inflation, interest rates, and real estate rather than exposures that matter in reaction to pandemic news.

As the use of text-as-data expands in economics and finance (Gentzkow et al., 2019a), it becomes ever more important to explore the strengths and weaknesses of different text-analytic methods. Under the widely-used dictionary approach (Tetlock, 2007; Loughran and Mcdonald, 2011; Baker et al., 2016), the researcher relies on expert-curated term sets to characterize and quantify the information content in relevant text documents. After extracting content measures, the researcher uses them to explain outcomes of interest. In our implementation of the dictionary approach, we use the term sets that Baker et al. (2019) apply to interpret aggregate stock market volatility.

A newer approach, growing in popularity, is supervised machine learning (ML). Under this approach, an algorithm selects the terms in a very large feature space that are most useful in explaining an outcome of interest. To implement the ML approach, we adopt the multinomial inverse regression (MNIR) method introduced by Taddy (2013) and recently applied in economics by Gentzkow et al. (2019b). We adopt MNIR because of its relative simplicity, its similarity to discrete-choice statistical models, and its successful application in other economic settings.

MNIR differs in two major respects from the dictionary approach. First, it considers all terms that appear in the discussions of *Risk Factors* as candidates for explaining returns. The set of all terms is an order of magnitude larger than the term sets encompassed by the curated dictionaries. Second, MNIR weights each term based on the strength of its association with the outcome of interest (firm-level returns, in our case). In contrast, dictionary approaches typically weight terms based on their frequency in the text documents of interest and perhaps in external libraries as well, as in some natural language processing methods.

Despite their differences, we find a remarkable congruence in predicted firm-level returns between the dictionary and ML approaches. Return predictions from one approach vary one-for-one, on average, with predictions from the other approach. In addition, the adjusted R^2 values in cross-sectional return regressions vary closely across jump dates under the two approaches. At the same time, MNIR achieves a uniformly higher adjusted R^2. This superior fit arises entirely because MNIR draws on a much larger feature space, as we show, rather than on other differences between the two approaches. By tapping a much larger feature space, MNIR captures many systematic aspects of the firm-level return structure that the dictionary approach misses.

However, it can be hard to obtain clear economic insights from raw ML results. To address this challenge, we proceed in steps. First, we identify terms that MNIR weights highly in explaining firm-level returns. These terms inform our selection of 'seeds' that form the basis for constructing term sets that define new risk exposure categories. In a second step, we build provisional term sets from the seeds based on similarity of linguistic context and MNIR weights. Third, we make judgmental adjustments to obtain a final version of new term sets. This process for constructing new term sets draws on both our estimated high-dimensional MNIR model and our domain expertise. Thus, it incorporates elements of both ML and expert-curated dictionary approaches. Finally, armed with our new term sets, we incorporate them into straightforward regression models that yield easily interpretable results.

When we apply this hybrid approach to pandemic-related jump dates, we obtain a much richer characterization of the forces that drive firm-level returns. It is how we uncover the role of exposures to social distancing restrictions, drug trials, and e-commerce. It is also how we uncover the role of upstream and downstream demand linkages. For example, downstream exposure to aircraft production predicts negative firm-level returns in reaction to bad COVID-19 news. The same news predicts positive returns for firms with high exposures to specific metals (e.g., tantalum and tungsten) that are critical inputs for semiconductors, lasers, and integrated circuits and for cloud computing and telecommunications. Acemoglu et al. (2016) show how downstream demand shocks operate in theory by propagating upstream through the production network.[2] Our results highlight the importance of the production network for understanding the effects of the COVID-19 shock. While evidence of COVID-19's impact on final consumer spending is now plentiful (Andersen et al., 2020; Baker et al., 2020b; Carvalho et al., 2020; Chetty et al., 2020; Cox et al., 2020; Surico et al., 2020), we provide evidence that its impact on intermediate goods suppliers is also important.

Dictionary methods and supervised ML are sometimes seen to occupy opposite ends of a methodological spectrum. Our study shows they are complements as much as substitutes. In this regard, our hybrid approach to uncovering risk factors and constructing term sets that capture them is useful more broadly. As an illustration, we use the hybrid approach to characterize firm-level return reactions to the 2020 Super Tuesday elections. The market rose 4% in reaction to these elections, widely regarded as a decisive victory for Joe Biden that greatly increased his prospects of becoming the Democratic presidential nominee. As before, we apply our hybrid approach to uncover risk factors, build term sets that capture them, and use the new term sets to explain firm-level returns. We find that Super Tuesday drove negative returns for firms with high exposure to mutual funds, private equity, oil, e-commerce, and lodging services; and positive returns for firms with high exposure to health insurance and property markets. Our hybrid approach lends itself to many other applications as well, including text-based analyses of what drives other outcomes.

An active literature considers how firm-level equity returns have responded to COVID-19. Here, we take note of a few studies that are particularly relevant to ours. Hassan et al. (2020) use natural language processing methods and human readings to quantify what firms say about COVID-19 and other infectious diseases in their earnings calls. They aggregate over their firm-level measures to show how sentiments related to COVID-19 vary by country and time.[3] They also show that equity returns are lower

[2]Other relevant theoretical analyses include Long and Plosser (1983), Atalay (2017), and Baqaee and Farhi (2019). See Carvalho and Tahbaz-Salehi (2019) for a review.

[3]Stephany et al. (2020) use the text in 10-K filings to track the evolution of COVID-related concerns at aggregate and industry levels.

in 2020 for firms that express greater concern about COVID-19 in their earnings calls. Our approach yields much more granular measures of business risk exposures, which we use to explain firm-level return reactions to several distinct types of news events. Our reliance on 10-K filings before the pandemic struck yields an ex ante characterization of risk exposures rather than an ex post one. Our approach is also likely to yield a fuller characterization of risk exposures, because firms face legal and financial liability for failing to disclose material risks in their regulatory filings (Mast et al., 2020). In contrast, earnings calls tend to focus on a limited set of salient concerns.

Papanikolaou and Schmidt (2020) and Pagano et al. (2020) use industry differences in employee ability to work from home (WFH) to explain firm-level returns (and other outcomes) in the wake of COVID-19. Their evidence fits well with our finding that bad COVID-19 news triggers positive return reactions at firms with high exposure to IT infrastructure, the demand for which rises with reliance on remote interactivity. However, our evidence about firm-level returns is largely distinct from the evidence in their study and others that use industry-level variation to characterize risk exposures. When we control for industry fixed effects, our text-based measures continue to have ample explanatory power for firm-level returns, underscoring how risk exposures and return reactions vary greatly even among firms in the same industry.

Laeven (2020) finds that social distancing measures adopted in response to the COVID-19 pandemic affect firms partly through input-output linkages. Ramelli and Wagner (2020) stress the role of upstream supply shocks due to disrupted exports from China, especially in the early stages of the pandemic. We focus on the period during which COVID-19 emerged as a global, rather than Chinese, health crisis and find an important role for many aspects of supply-chain linkages in driving firm-level abnormal return reactions to news about COVID-19.

A broader finance literature uses the *Risk Factors* to study equity returns, but few papers incorporate machine learning methods. Hanley and Hoberg (2019) and Lopez Lira (2019) use unsupervised learning approaches to group 10-K words into clusters that correlate with stock returns. Ke et al. (2019) propose a supervised learning framework for predicting stock returns using media text data, which they show outperforms standard sentiment dictionaries at return prediction. Our study is more focused on organizing terms into interpretable categories that inform our understanding of what drives firm-level return reactions to common shocks.

Our study proceeds as follows. Section 2 discusses our data sources, the dictionary approach, and MNIR. Section 3 presents results based on the dictionary approach. Section 4 implements the MNIR approach, establishes a close relationship between the return predictions generated by the two approaches, and shows that MNIR delivers better fitting models. We also show that a standard clustering algorithm applied to MNIR return predictions sorts jump dates into categories that align closely with the classification that Baker et al. (2020a) derive from next-day newspaper accounts of large market-level moves. It is both remarkable and reassuring that two entirely different methods, drawing on such dissimilar text sources, yield similar classifications of jump dates. Section 5 develops and implements our hybrid approach, first applying it to pandemic-related jump dates and then to Super Tuesday. Section 6 concludes.

2 Data and Empirical Methods

2.1 Firm-level returns and other financial measures

We consider daily returns for 2,155 equity securities on the 20 jump dates from 24 February to 27 March that Baker et al. (2020a) identify and classify. To compute returns, we obtain daily closing prices (PRCCD) of common equities traded on AMEX, NYSE and NASDAQ from the Compustat North America Security Daily file. We account for stock splits, dividends, etc. using the daily adjustment factor (AJEXDI) and the daily total return factor (TRFD) in the same Compustat file. We restrict attention to U.S.-incorporated firms with share prices quoted in U.S. Dollars. See Appendix A for more information about our sample.

Figure B.1 displays an analog to Figure 1 that considers the cross-firm standard deviation of daily returns in place of the IQR. Figures B.2 to B.4 display histograms of daily market-level returns, the IQR of firm-level returns, and the standard deviation of firm-level returns for trading days in 2019 alongside analogous statistics for the jump dates in 2020. These figures reinforce the chief message of Figure 1: The jump dates in our sample are extreme events with respect to both market-level returns and the dispersion of firm-level returns.

Our main outcomes of interest are firm-level abnormal returns constructed in the standard way. Specifically, we generate daily security-level abnormal returns for jump days as the difference between (i) a stock's actual return in excess of the risk free rate and (ii) its expected excess return per the Capital Asset Pricing Model (CAPM):

$$\text{AbnER}_{i,t} = log\left(\frac{p_{i,t}}{p_{i,t-1}}\right) - R_{f,t} - beta_i \times \left(R_{M,t} - R_{f,t}\right) \tag{1}$$

where p_{it} denotes the adjusted share price for stock i on day t, R_f denotes the four-week treasury bill rate (a proxy for the risk free rate), $beta_i$ is the stock's CAPM beta, and R_M is the value-weighted average market return. We estimate each stock's beta using an OLS regression of its daily excess return on the contemporaneous market-level excess return in the sample of all trading days in 2019.

Our statistical models for (abnormal) returns include two controls for financial characteristics. The first is a measure of the firm's equity market capitalization, Mcap_{it}, computed as shares outstanding (CSHOC) times closing price per share.[4] The second is firm leverage, computed as (long term debt (DLTT) + current liabilities (DLC)) divided by total assets (AT). We use the most recent data in the Compustat file for this purpose, yielding leverage values based on fiscal year 2019 (2018) for 89 (11) percent of firms. Appendix Table B.1 reports descriptive statistics for these and other variables used in our firm-level analyses.

[4] In a slight abuse of notation, we often refer to i as a firm whereas it actually indexes stocks. In the few cases where a firm has multiple stocks, we assign the same Mcap_{it} value to each one.

2.2 The *Risk Factors* text in 10-K files

Since 2006 (for fiscal year 2005), the Securities and Exchange Commission (SEC) has required the vast majority of publicly held firms to include a discussion of *Risk Factors* (*RF*) in Part 1a of their annual 10-K filings. The SEC advises that these discussions include any item that could impact future earnings. Investors can sue for compensation if the firm omits material information or risks (Mast et al., 2020). We use *RF* texts filed from 2010 to July 2016. (Machine-readable versions are available from EDGAR.) This choice of years mitigates the role of idiosyncratic language in a single filing and ensures that any relationship we find between the *RF* text and returns in 2020 reflects persistent risk exposures that long predate the arrival of COVID-19.

Appendix A describes how we pre-process the raw text files to obtain documents composed of words and phrases. After pre-processing, there are 18,911 unique terms that appear a total of 57 million times in our *RF* corpus. The large number of terms necessitates some form of dimension reduction, which we accomplish in two distinct ways: first, curated dictionaries that identify terms of interest and organize them into categories; second, Taddy's (2013, 2015) MNIR model, which operates on all terms.

2.3 Empirical approach 1: Curated dictionaries

We adopt the dictionaries of Baker et al. (2019), who expand on ones previously developed by Baker et al. (2016) and Davis (2017). One attraction of these dictionaries is their detail. They include 16 dictionaries that cover aspects of economic and financial conditions and another 20 that pertain to policy areas. Each one contains numerous terms that effectively define the dictionary's category. The construction of these dictionaries reflects the input and judgment of expert economists drawing on textbooks, newspaper articles, 10-K filings, and "their own knowledge of economic matters and input from other economists in seminars and personal communications." Baker et al. (2019) show that these dictionaries are useful for tracking and interpreting movements in stock market volatility, which is conceptually related to stock market jumps.

The dictionaries contain 430 terms that appear in our *RF* corpus, 244 after removing rare terms at the pre-processing stage.[5] These 244 terms appear nearly 1.4 million times, constituting 2.4% of the *RF* corpus. The *RF* texts for a given firm contain 28 distinct dictionary terms on average (standard deviation of 10) and 642 instances of dictionary terms (standard deviation of 620). To quantify a firm's exposure to a given risk category, we identify sentences in its *RF* texts that contain at least one term in the corresponding dictionary.[6] After computing the fraction of such sentences in each of the firm's *RF* texts, we calculate the average fraction. In this way, we obtain 36 firm-level exposure values, one for each category and its corresponding dictionary. Descriptive statistics for these firm-level exposure measures appear in table B.1.

2.4 Empirical approach 2: Multinomial inverse regression

MNIR treats the *RF* texts for each firm i as a *bag-of-words* represented by a V-dimensional vector $\mathbf{x}_i$ of terms or "features." $x_{i,v}$ is the count of term v for firm i, and $V = 18,911$ is the number of unique terms in our *RF* corpus. At the firm level, the average number of nonzero elements in $\mathbf{x}_i$ is 2,245, with a standard deviation of 891.

Many popular machine learning approaches to text analysis in economics and finance (e.g., latent Dirichlet allocation) represent documents in a latent space of "topics." These approaches reduce the dimensionality of text but can yield topics that lack clear relationships to the outcomes of interest. MNIR models the relationship between the terms in $\mathbf{x}_i$ and the outcomes of interest directly. The resulting statistical structure is similar to ones that arise in standard econometric models of discrete choice. Taddy's (2013) original MNIR model was inspired by an economics application (Gentzkow and Shapiro, 2010), and it has been further applied in Gentzkow et al. (2019b). These observations suggest that MNIR is a promising tool for the text-based analysis of firm-level

[5]Three Baker et al. (2019) dictionaries – foreign trade exposure, immigration policy, and government-sponsored enterprises – contain only rare terms. We drop these three categories in our implementation of the dictionary approach. Retaining them has little impact on the results.

[6]We handle plurals using the NLTK WordNet Lemmatizer. For example, "recession" and "recessions" are both captured by the "Broad Quantity Indicators" category reported in Appendix A.3.

returns and for exploring how machine learning methods can extend and complement the use of dictionary methods in economics and finance.

MNIR posits $\mathbf{x}_i \sim \mathrm{MN}(\mathbf{q}_i, N_i)$, where $\mathbf{q}_i$ is a multinomial V-dimensional probability vector and N_i is the total number of terms in firm i's *RF* texts (i.e., $N_i = \sum_v x_{i,v}$). The probability of feature v for firm i is

$$q_{i,v} = \frac{\exp(a_v + \mathbf{y}_i^T \mathbf{b}_v)}{\sum_v \exp(a_v + \mathbf{y}_i^T \mathbf{b}_v)}, \tag{2}$$

where $\mathbf{y}_i = (\mathrm{AbnER}_{i,t}, \mathbf{c}_i)$ contains firm-i abnormal returns on day t and firm controls $\mathbf{c}_i \in \mathbb{R}^P$. a_v is a parameter that controls for the baseline frequency of term v in the corpus, and $\mathbf{b}_v$ is a $P+1$ vector of coefficients that describe how firm observables map to the probability that term v appears in the *RF* texts.

Equation (2) describes a multinomial logistic regression over V categories, which we fit to 2,155 observations (per jump day), one per firm. The outcome being modeled is the probability that a particular term in V appears in a random draw from the firm's *RF* texts. The fitted model delivers 18,911 estimated probabilities for 2,155 firms. The non-standard aspect of (2) is the high dimensionality of V. So, we estimate (2) using Bayesian regularization methods with a Gamma-Laplace prior structure on the regression coefficients. (This estimation method is a more flexible form of the standard LASSO penalty, one that admits coefficient-specific penalization.) The selection of the prior trades off goodness-of-fit and model complexity via the maximization of an information criterion to avoid over-fitting (see Taddy, 2013 and Taddy, 2015 for details).

We seek to use *RF* text features to predict returns, while (2) models the inverse relationship of term probabilities given returns. To move from the estimated parameters of (2) to a forward regression with $\mathrm{AbnER}_{i,t}$ as the dependent variable, Taddy (2013) defines a *sufficient reduction projection* $z_i = \sum_v x_{i,v} b_{1,v}$ with the property $AbnER_{it} \perp \mathbf{x}_i \mid z_i, N_i, \mathbf{c}_i$. Thus, conditional on the scalar projection z_i, the high-dimensional raw data contain no extra predictive information for returns. This result does not specify the functional form for relating z_i to $AbnER_{it}$ in a forward regression, but it says we can model $AbnER_{it}$ as a function of z_i, N_i, $\mathbf{c}_i$, while disregarding $\mathbf{x}_i$.

2.5 Why two empirical approaches?

We adopt two distinct approaches to the analysis of firm-level equity returns for two reasons. First, we want to compare their strengths and weaknesses in a rich, concrete setting. Second, we want to explore whether and how empirical researchers can enrich their text-based analyses by combining elements of both approaches.

A clear advantage of the dictionary approach is it simplicity and transparency. Its implementation does not require the estimation of a first-stage statistical model, as in the inverse regression model (2). It relies instead on domain expertise, as codified in the dictionaries, to organize and quantify the text data and to use the resulting quantification to explain outcomes of interest.

A key advantage of MNIR (in common with all supervised learning models) is the ability to use all terms in the text corpus to explain the outcomes of interest. In our context, that means using the *RF* texts to explain systematic aspects of firm-level return reactions to pandemic-related news and other common shocks. Our MNIR model considers all 18,911 terms in our RF corpus, while the dictionary approach considers only 244 terms organized into 36 categories. As a result of its much larger feature space, MNIR can potentially capture aspects of the firm-level returns structure that the dictionary approach misses.

Two claims often arise in comparisons between dictionary and ML methods. First, that dictionary methods more readily yield results with clear interpretations. Second, that ML methods require less need for domain expertise or its costly codification. Each claim contains a kernel of truth, but the reality is more complex in our setting. In particular, dictionary methods often but not always yield easy-to-interpret results. We show how to use ML methods to sharpen the interpretations of dictionary-based results. Conversely, we also show how to use dictionary methods and domain expertise to interpret results that emerge from an MNIR implementation of the ML approach.

3 Results Based on the Dictionary Approach

To implement the dictionary approach, we fit regression models for daily firm-level returns via least-squares estimation. Our models have the following form:

$$\text{Abn}_{it} = \sum_{j=1}^{J} \beta_j \, \text{RExp}_i^j + \beta_{J+1} \, \text{Leverage}_i + \beta_{J+2} \, \log(\text{Mcap}_{it}) + \gamma_{s(i)} + \epsilon_{it}, \qquad (3)$$

where Abn_{it} is firm-i's abnormal return on day t, RExp_i^j is its exposure to risk category $j = 1, 2, ..., 36$, Leverage_i and Mcap_{it} are the financial controls defined earlier, and $\gamma_{s(i)}$ are NAICS2 fixed effects. Apart from fixed effects, all regressors enter (3) in standard deviation units. We fit (3) separately for each jump day or collection of same-type jump days according to the classification in Figure 1. When fitting to a collection of days, we use average values of Abn_{it} and $\log(\text{Mcap}_{it})$ for the days in question. The collection of fiscal policy jumps includes three days with a positive market return and one, 23 March, with a negative return in reaction to a delay in passing a fiscal relief bill. To account for this sign flip, we multiply firm-level abnormal returns on 23 March by -1 before averaging over days.

Table 1 reports estimates of (3), suppressing coefficients that are insignificant at the 10 percent level. Our simple model explains much of the (very large) abnormal return variation on jump days: Adjusted R^2 values range from 20% the day after Super Tuesday to 33% on pandemic fallout days and 36% for the March 9 Oil Price Crash. While not our focus, we observe that the financial controls are important return predictors. Market cap is a highly significant return predictor on jump days and, consistent with the evidence in Alfaro et al. (2020) and Ramelli and Wagner (2020), more leveraged firms perform worse in reaction to bad news about the pandemic.

Many of our dictionary-based exposure measures are also significant return predictors. As seen in Column (1), firms with high exposures to *inflation, credit indicators, taxes, entitlement programs, energy and environmental regulations,* and *transportation, infrastructure and utilities* react especially negatively to bad news about the pandemic

Jump Classification $\longrightarrow$ **Dependent Variable:** Abn_{it}	(1) COVID-19 and Its Fallout		(2) Monetary Policy Easing		(3) Fiscal Policy Stimulus		(4) Super Tuesday Election		(5) Oil Price Crash	
General Economic Categories										
Inflation	−0·21	(−2·5)	0·92	(4·7)			0·24	(3·5)	0·28	(1·9)
Interest Rates			0·78	(5·4)	0·31	(1·9)	0·13	(1·9)	−0·63	(−3·4)
Credit Indicators	−0·29	(−4·1)	−0·68	(−3·4)	−0·21	(−1·8)				
Labor Markets									0·34	(1·8)
Real Estate Markets			0·51	(2·2)	0·45	(2·8)	0·44	(2·8)		
Business Investment and Sentiment					0·31	(5·0)				
Consumer Spending and Sentiment			−0·36	(−1·9)						
Commodity Markets			−0·41	(−2·0)			−0·37	(−2·7)	−1·73	(−6·6)
Healthcare Matters			0·62	(2·0)						
Litigation Matters									0·34	(2·4)
Competition Matters			−0·37	(−1·8)						
Intellectual Property Matters	0·45	(6·2)			−0·61	(−6·9)	−0·29	(−2·9)	0·59	(3·0)
Policy-Related Categories										
Taxes	−0·28	(−2·1)			0·48	(3·4)				
Entitlement and Welfare Programs	−0·49	(−2·9)							−0·84	(−3·0)
Monetary Policy					−0·30	(−2·3)	−0·25	(−2·1)	−0·36	(−2·9)
Financial Regulation	0·12	(1·8)			0·29	(2·8)	−0·29	(−5·1)		
Competition Policy			0·32	(2·0)						
Intellectual Property Policy					0·14	(2·2)				
Energy and Environmental Regulation	−0·19	(−2·2)	−0·31	(−2·1)	0·21	(4·1)			−1·52	(−3·1)
Housing and Land Management					0·07	(1·8)				
Other Regulation			0·25	(3·2)	0·24	(3·6)				
Healthcare Policy	0·31	(1·9)							1·06	(5·8)
Transportation, Infrastructure, Utilities	−0·16	(−2·6)								
Elections and Political Governance							−0·10	(−2·9)	−0·36	(−2·0)
Financial Controls										
Log Market Cap	0·53	(7·3)	0·73	(3·0)	0·68	(4·5)	0·66	(4·8)	1·00	(4·4)
Leverage	−0·42	(−3·0)	−0·85	(−2·8)	0·40	(2·9)	−0·12	(−0·8)	−0·72	(−2·5)
Observations [Adjusted R^2]	2155	[0.329]	2155	[0.232]	2155	[0.285]	2155	[0.199]	2155	[0.361]

Table 1: Regression Results Based on Dictionary Approach

Jump days are grouped as follows: Column (1) February 24, 25, 27 and March 3, 5, 11, 16, 18, 27; (2) March 2, 17; (3) March 10, 23, 24, 26; (4) March 4; and (5) March 09. The dependent variable is the one-day percent abnormal equity return, or the corresponding security-level average across business days in columns 1 to 3. Each column considers the 36 dictionary-based risk exposure measures plus log market cap, leverage, and 2-digit NAICS fixed effects as controls. The t statistics reported in parentheses are computed using NAICS2 clustering. For presentation purposes, we omit coefficients of dictionary categories that are insignificant at the 0.1 level.

and its economic fallout. Firms with high exposures to *intellectual property* and *healthcare policy* perform relatively well in reaction to bad pandemic news. As reported in Table B.1, 21 firms (1% of our sample) have *intellectual property* exposures more than 3.5 standard deviations greater than the mean exposure, which implies a one-day positive abnormal return differential on pandemic fallout days of at least (0.45)(3.5)= 1.6 percentage points for these firms.[7] The *intellectual property* category is especially relevant for pharmaceutical firms, as its dictionary includes "patent" and "new drug application." Thus, the large, positive coefficient on *intellectual property* fits well with the view that bad news about the severity of the COVID-19 pandemic is relatively good for firms that own or develop healthcare-related intellectual property.

Looking across the columns in Table 1, the structure of firm-level return reactions differs systematically by jump type. For example, jumps attributed to monetary policy easing yield large positive return reactions at firms with high exposures to *interest rates* and *interest rates* but not to *intellectual property* or *transportation, infrastructure and utilities.* Jump days attributed to fiscal policy news generate the largest return reaction at firms with high exposure to the *tax* category. Firms exposed to tax-sensitive categories like real estate and business investment also outperform on fiscal policy jump days. However, the precise interpretation of some of these patterns is unclear. The *tax* category, for example, captures exposures to both high taxes and the potential for large tax credits (e.g., for R&D or investment). This example illustrates the interpretation challenges that can arise under the dictionary approach.

Several exposure measures play a role in driving firm-level return reactions to the Super Tuesday elections, including *real estate, commodity markets, intellectual property,* and *financial regulation.* These reactions reflect revised assessments of Job Biden's (and Bernie Sanders') prospects of becoming the Democratic Party's presidential nominee, how the general election campaign would play out in view of the expected nominee,

[7] Examples include Universal Display Corporation with an *intellectual property* exposure 5.7 standard deviations greater than the mean exposure, Editas Medicine, Inc. (5.6), Interdigital, Inc. (5.4), Dicerna Pharmaceuticals (5.1), Gilead Sciences (4.9), Kindred Biosciences (4.7), Biolife (4.5), Qualcomm (4.4) and Blueprint Medicines (4.0).

and the likelihood that Donald Trump would win re-election. The revision in assessments affect relative returns positively at firms exposed to *real estate,* for example, and negatively at firms exposed to *commodity markets,* e.g., oil and gas companies.

The Oil Price Crash came with a huge 7.9 percentage point drop in the overall stock market on 9 March. As reported in Column (5), firms with high exposure to *commodity markets* and *energy and environmental regulations* experienced especially large stock price drops. Consider a firm at the 99th exposure percentile for *commodity markets.* According to Table B.1, this firm's *commodity markets* exposure is 4.6 standard deviation units greater than the average firm's exposure. Thus, conditional on the other covariates in (3), the estimated model predicts that the 99th percentile firm has one-day negative abnormal return differential of $(-1.73)(4.6) = -8.0$ percentage points. This calculation illustrates a broader point: Some firms have extremely high exposures to one or a small number of risk categories. As a result, big shocks that pertain to particular exposure categories can drive very large firm-level return differentials.[8]

To summarize, our implementation of the dictionary approach yields an initial characterization of firm-level reactions to various market-moving news events in the wake of COVID-19. Some ambiguities arise when seeking to interpret the results, perhaps because the dictionaries were not specifically designed to characterize stock market behavior on our particualr jump days. Moreover, as we have stressed, the dictionary approach taps only a small fraction of the *RF* corpus.

[8]Given this, one might worry about our exclusion of dictionary terms that appear rarely in the *RF* corpus. While rare overall, these terms might capture important exposures at a few firms. To assess this concern, we ran jump-day regressions using exposure measures based on the full set of 430 terms in Baker et al. (2019) and compared them to the ones based on our more limited set of 244 terms. Figure B.5 plots the adjusted R^2 for each pair of regressions across days and shows they are nearly identical. Thus, nothing is lost in our application by dropping the rare dictionary terms.

4 Exploiting the Full *RF* Corpus

4.1 How much gain in fit from MNIR?

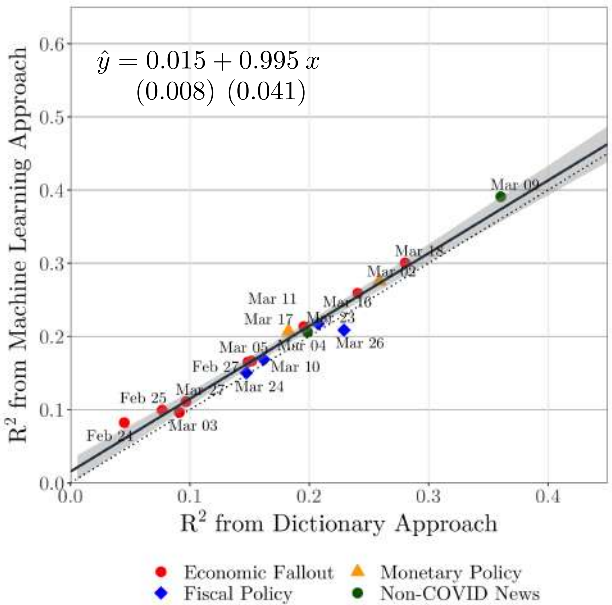
We estimate the inverse regression (2) separately for each jump day or collection of days with controls $\mathbf{c}_i = (\text{Leverage}_i, \log(\text{Mcap}_{it}), \gamma_{s(i)})$. Our forward regression is

$$Abn_{it} = \alpha_1\, z_i + \alpha_2\, N_i + \alpha_3\, Leverage_i + \alpha_4\, Log(Mcap_{it}) + \gamma_{s(i)} + \varepsilon_{it}, \tag{4}$$

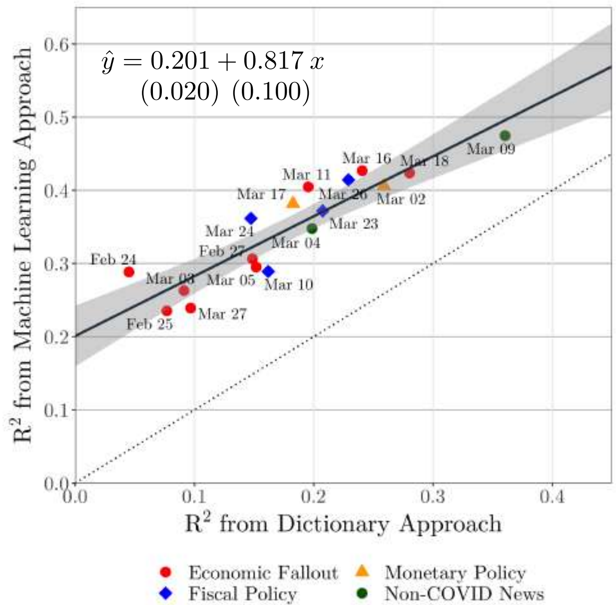
which we estimate by ordinary least squares. N_i is the number of terms in firm i's *RF* texts, and z_i is the sufficient reduction projection that summarizes the information in the inverse regression. We first examine the gain in fit achieved by MNIR relative to the dictionary approach. To do so, we separately estimate (3) and (4) for each of the 17 jump days covered by Table 1.

There are two reasons why MNIR might have greater explanatory power. First, (2) allows more flexibility in the relationship between returns and terms. Because (2) is fit with regularization, terms are selected or not based on the strength of their association with returns. Moreover, selected terms can have different regression coefficients. In contrast, the dictionary approach constrains all terms in the same category to have the same relationship to returns. There is, for instance, no down weighting of terms in a given dictionary that are less helpful in quantifying return-relevant exposures. Second, MNIR operates on a vastly larger feature space. Insofar as there is useful information about returns in the 18,667 terms ($= 18,911 - 244$) not classified in the dictionaries, we expect MNIR to achieve a better fit. To distinguish between these two possible reasons for a better fit, we also estimate MNIR using just the 244 dictionary terms. To summarize, we fit three sets of regressions and obtain three sets of adjusted R^2 values corresponding to the dictionary approach, MNIR based on the 244 dictionary terms, and MNIR based on all 18,911 terms.

Figure 2 displays the results. Figure 2a shows that the two approaches achieve essentially the same fit when using the same terms, and that this result holds across jump dates. That is, the greater flexibility of MNIR in how it relates returns to terms does not materially improve goodness of fit. Whatever interpretation value the dictionary approach offers by organizing terms into categories does not come at the expense of model fit – at least not in our application with our dictionaries.



(a) Restricted Feature Space in MNIR



(b) Full Feature Space in MNIR

Figure 2: R^2 achieved via Dictionary and Machine Learning Approaches

This figure plots the adjusted R^2 from regressions (3) and (4). The left panel displays results when using only the 244 dictionary terms as inputs into the MNIR model, while the right panel displays results when using all terms. Each plot presents a fitted regression line that relates the R^2 values (solid black line), along with 95% confidence intervals corresponding to the shaded region. The dashed line is the 45-degree line.

Figure 2b compares the dictionary approach to the MNIR implementation that uses all terms in the *RF* corpus. Here, we see a uniformly better fit for MNIR. The fit gain is large: The adjusted R^2 value is roughly 20 percentage points higher under MNIR on all jump dates. The implication is that, at least in our setting, the supervised machine learning approach more fully explains returns entirely because it considers a much larger set of terms than the dictionary approach.

Figure 2b carries a subtler message than supervised learning fits better. In particular, there is a close, approximately one-for-one relationship across jump dates between the R^2 values for the two approaches. The confidence interval for the slope coefficient relating the R^2 values contains 1. This result suggests that MNIR and the dictionaries draw on similar information to explain returns, since their ability to do so is highly correlated across jump dates. Next, we develop this point more fully.

4.2 Comparing predicted firm-level returns

Figure 3 plots fitted values for abnormal returns using the dictionary approach against the corresponding MNIR predictions (using all terms). To construct this figure, we fit abnormal return models separately for the 17 jump days to obtain 36,652 predictions under each approach. Remarkably, the regression line that relates the two sets of predicted outcomes is indistinguishable from the 45 degree line. In other words, the predicted firm-level return from the dictionary approach is an unbiased estimate of the MNIR prediction. This result also holds on each individual jump date.

One might worry that this one-for-one relationship is an artifact of including industry fixed effects under both approaches. To check this, we drop industry effects and refit models that focus on the text-based measures. Figure B.7 displays the results. In the pooled sample, we continue to find a near one-for-one relationship in fitted abnormal returns between the two approaches. There is also a close, near one-for-one relationship on most – but not all – individual jump days. For example, the regression slope is only 0.822 on 24 March. Thus, the striking result in Figure 3 is not an artifact of industry effects or the consequence of some other mechanical effect.

While predictions under the two approaches coincide on average, predicted return distributions differ in the higher moments. Figure 4 summarizes the predicted abnormal return distributions under several specifications. The dispersion of predicted returns is similar when including sector fixed effects alone or dictionary measures alone (including financial controls in both cases), as shown by comparing the second and third box plots in Figure 4. Using both industry effects and dictionary measures yields somewhat greater dispersion in predicted returns. Using MNIR yields a considerably

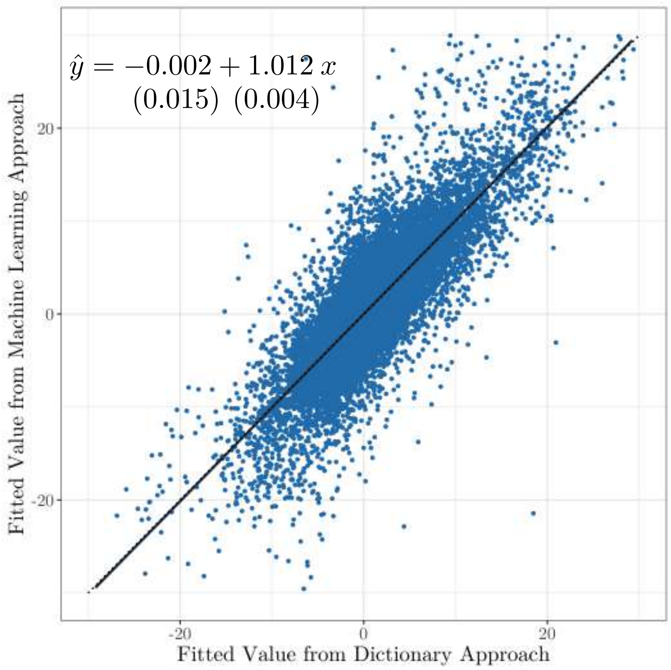


Figure 3: Fitted Values from Dictionary and Machine Learning Approaches
This figure plots fitted values from regressions (3) and (4), pooled across all 17 dates. It also presents a fitted regression line that relates the two (solid black line). Confidence bands are omitted from the graph due to their narrowness at standard levels. The dashed line is the 45 degree line.

more dispersed distribution of predicted returns, confirming that it captures additional return-relevant information in the RF corpus.

4.3 Text-based measures and narrow industry classifications

We now ask whether our text sources and methods capture information about returns beyond what is captured by narrow industry definitions. An affirmative answer provides evidence about firm heterogeneity and offers a means to characterize it. The question is also important in the context of the COVID-19 literature. Both Pagano et al. (2020) and Papanikolaou and Schmidt (2020) construct industry-level exposures (up to NAICS4 granularity) to restrictions on labor supply due to the inability of employees to work from home. They argue that differences in this source of exposure drive much of the firm-level variation in stock returns during 2020. By controlling for narrowly defined industries, we can assess whether text-based measures pick up information beyond the supply-side shocks that these studies highlight.

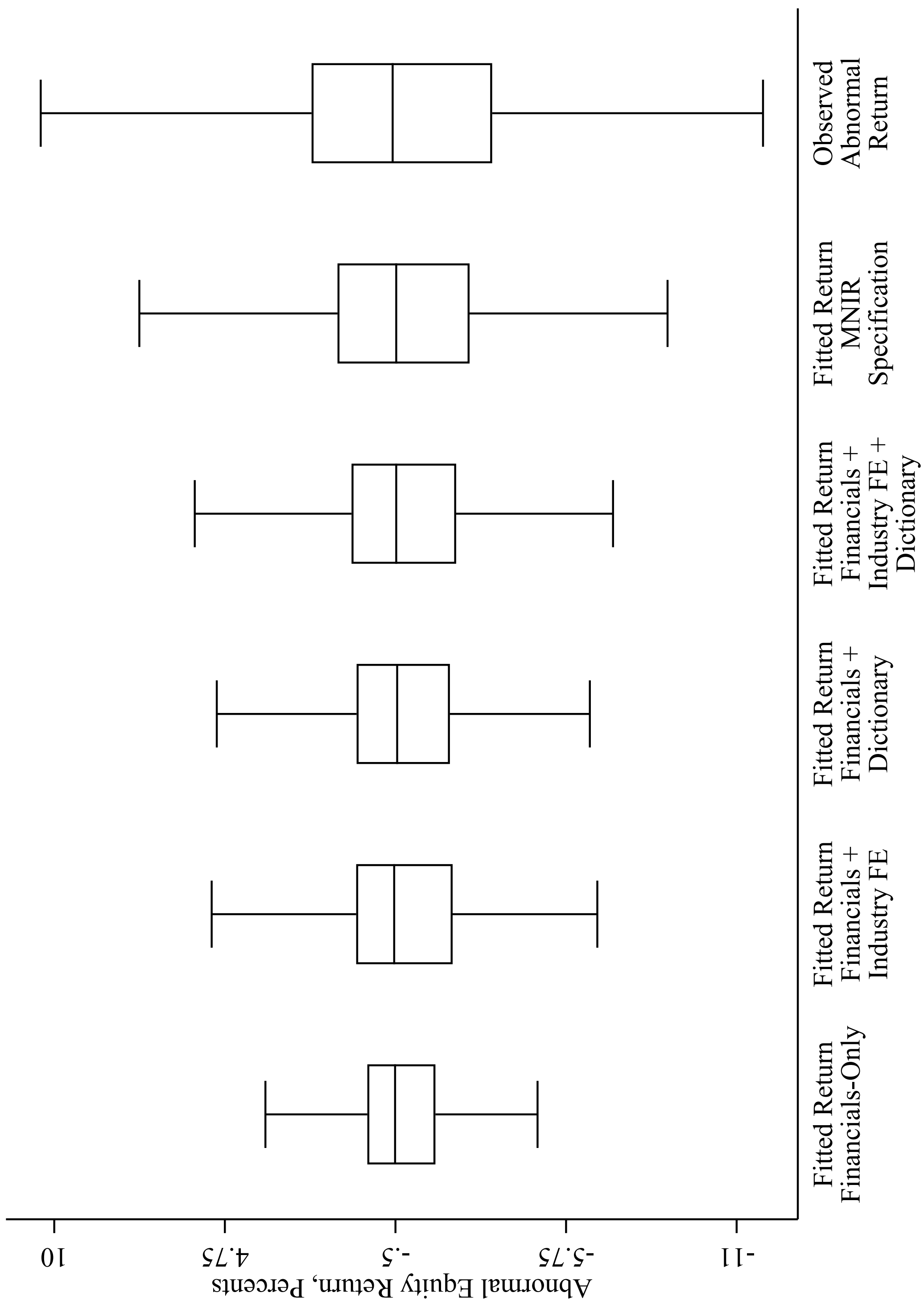
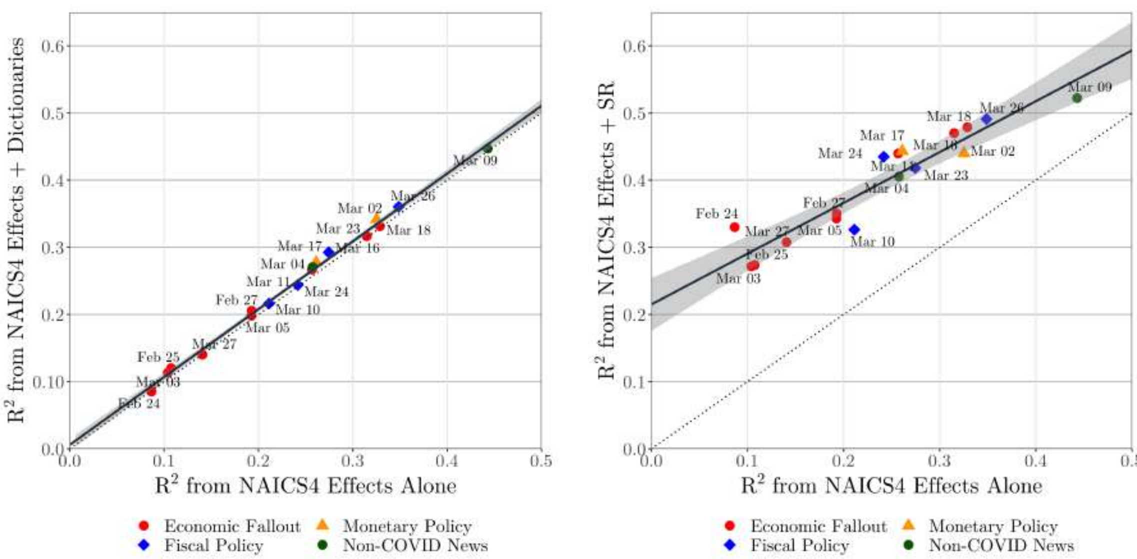


Figure 4: Range of Firm-Level (Fitted) Returns

This figure considers all 2,155 securities and 17 jump dates in our analysis sample. For each specification, we generate fitted abnormal returns at the security-day level and generate the corresponding boxplots. The upper (lower) whisker represents the largest (smallest) observation that is less (greater) than or equal to the third (first) quartile plus (minus) 1.5*IQR.

There are 216 unique NAICS4 industries in our baseline sample, from which we drop the 287 firms that lie in industry codes with fewer than five firms overall or with no available NAICS4 code. This leaves a subsample containing 1,868 firms distributed over 97 unique NAICS4 codes. For each jump day, we model abnormal returns as depending on firm financial controls and NAICS4 fixed effects, and we record the adjusted R^2. We then add the dictionary exposures and the sufficient reduction projection, and we again record the adjusted R^2. Figure 5 plots the results.



(a) Dictionary Method

(b) Machine Learning Method

Figure 5: Improvement in R^2 beyond Narrow Industry Codes

This figure plots adjusted R^2 values from regression models fit to firm-level returns on jump days. The horizontal scale shows the adjusted R^2 in models with NAICS4 fixed effects and firm-level financial controls. The vertical scale shows the adjusted R^2 values that result when adding dictionary-based measures (panel (a)) or the sufficient reduction projection estimated in (2) (panel (b)). Models are fit to the subsample of firms that lie in NAICS4 industries with five or more firms.

Figure 5a shows that adding dictionary measures does not shrink the residual variance in the return regressions, conditional on narrowly defined industry controls. That does not mean dictionary measures are uninformative about return drivers, but they don't improve model fit relative to detailed industry controls. In contrast, Figure 5b shows that including the sufficient reduction projection yields a large R^2 gain on every jump date.[9] In Section 5, we develop an approach to uncovering and interpreting the additional information about return drivers captured by MNIR.

4.4 Clustering jump days based on return structures

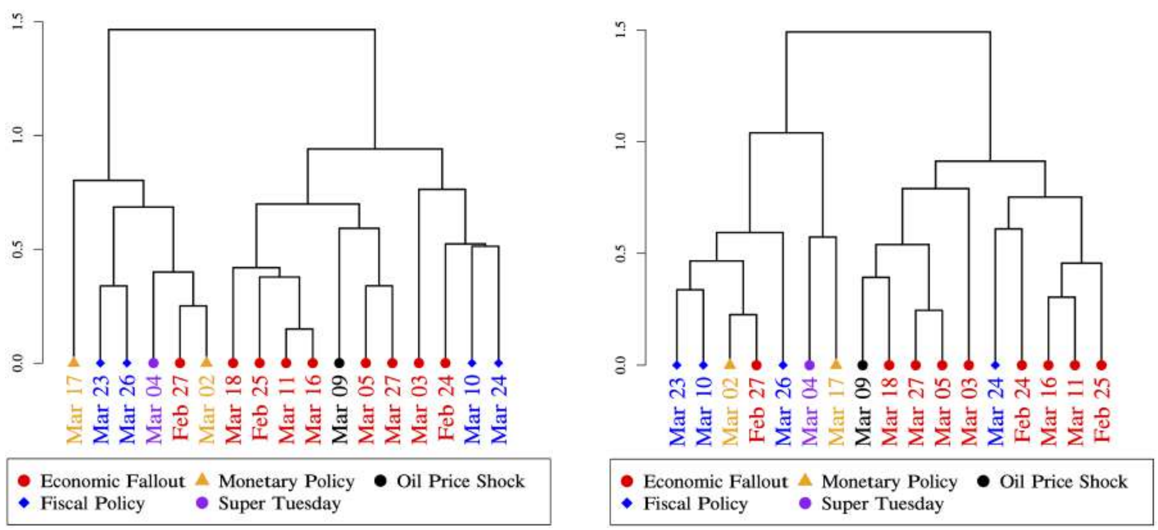
Recall that Baker et al. (2020a) use human readings of next-day newspaper accounts to classify *market-level* jumps as to reason (pandemic fallout, fiscal policy, etc.). We now apply an automated approach to classify jump days based on the *structure* of predicted firm-level abnormal returns. For this exercise, we measure returns in (2), (3), and (4) as $AbnER_{it} \times (\mathbf{5.5}/AvgER_t)$, where 5.5 is the average market-level move in our sample of 17 jump dates. This rescaling of returns ensures that our clustering reflects the structure of abnormal returns and *not* their overall magnitude or direction. We then fit our regression models separately by day and build a 17×17 matrix, where the (t_1, t_2) element is the correlation between the day-t_1 fitted values and the day-t_2 fitted values. Given this matrix, we apply a standard hierarchical clustering algorithm to group like days together, and display the results as dendograms in Figure 6.[10]

The dendograms reveal two interesting results. First, we obtain similar clusterings of jump days under the dictionary and ML approaches. This result reinforces our earlier conclusion that these two very different methodologies yield congruent return predictions. Second, the clustering that emergeS from our automated analysis of firm-level return *structures* based on *Risk Factors* texts in 10-K files is similar to the newspaper-based classification of *market-level* jumps in Baker et al. (2020a). The

[9]This result holds even though we estimate the inverse regression model (2) on a sample of 2,155 firms, which involves a potentially different relationship between terms and returns than the smaller sample of firms considered in Figure 5.

[10]The clustering algorithm works as follows. Start with a separate cluster for each jump day. In step 1, group the two days with the highest correlation of fitted values, yielding 16 clusters. Again compute the correlation between each cluster, using the minimum correlation value between all cross-cluster observations as the similarity metric when the cluster covers more than one day. This is known as 'complete link clustering'. Repeat this process until all observations lie in a single cluster. The height of the dendogram corresponds to the similarity level at which clusters are merged. See, for example, chapter 17 of Manning et al. (2009) for more details.

similarity is particularly evident for the MNIR approach. With few exceptions, the clustering algorithm groups the "fiscal" and "monetary policy" jumps (according to newspaper-based approach) into distinct blocks, as it does for the "pandemic fallout" jumps. Interestingly, the clustering algorithm groups Super Tuesday with other policy jumps, and it groups the oil price crash with other pandemic fallout days. These results support the idea that the Super Tuesday elections shifted expectations about future polices, while the oil price collapse shifted other aspects of the economic outlook. We conclude that identifiable categories of news events differ in how they interact with firm-level risk exposures to drive the structure of firm-level return reactions.



(a) Dictionary Method

(b) Machine Learning Method

Figure 6: Clustering of Days via Different Methods

This figure displays dendograms that represent clusterings of our 17 event days. In the left-hand panel, we cluster days based on the correlation matrix of the fitted values from (3) fit with $AbnER_{it} \times (\mathbf{5.5}/AvgER_t)$ as the dependent variable; in the right-hand panel, we cluster days based instead on the fitted values from (4) fit with the same dependent variable. In both cases, we use a complete-link hierarchical clustering algorithm, and color-code days in line with manually-assigned labels.

5 Uncovering and Interpreting Risk Factors

We now use our MNIR model to uncover and *interpret* the risk exposures that drive return reactions, a different undertaking than the usual prediction-oriented application of machine learning. The major challenge here is to organize the thousands of estimated model parameters in an insightful way. A first step in this direction is to exploit the same grouping of jump days as in Figure 1 and Section 3. From there, we fit MNIR models develop new risk exposure measures that yield insights into how shocks affect the structure of firm-level returns. Our first application is to pandemic fallout days.

5.1 Risk exposures for pandemic fallout days

When fit to pandemic fallout days, our MNIR model places positive weight on 9,948 terms and negative weight on 8,389 terms (574 terms are not selected into the model). Table B.2 displays terms with the largest positive and negative inverse regression coefficients on abnormal returns, given by $b_{1,v}$, the first element of $\mathbf{b}_v$ in (2). These terms suggest a more granular characterization of exposures than the baseline dictionaries. For example, while both "wheat" and "oil" appear in the dictionary for *Commodity Markets*, "wheat" is the term most associated with positive returns and "oil" is among those associated with negative returns. A natural interpretation is that pandemic news shifted their demands in opposite directions. Wheat is an input into basic food production, which remained stable as the pandemic unfolded – while oil supports the physical movement of goods and people, which fell dramatically. Other important terms include "games," "optics," "patents" and "clinical trials" as predictors of positive returns and "restaurants," "hotels," "airline industry" and "jet fuel" as predictors of negative returns.

Certain companies also feature prominently: "Intel" has a high association with positive returns, and "Delta," "Phillips" and "Boeing" have high associations with negative returns. Since we drop terms that appear in the *RF* texts of fewer than 25 firms, these findings say that the return reactions of many firms are strongly affected by their commercial connections to these companies.

We can inspect return drivers for individual firms as well. Table 2 lists the ten firms with the greatest residual shrinkage when adding the sufficient reduction projection to the regression model.[11] The table provides a short business description for each firm and lists its top five terms – calculated as $\hat{b}_{1,v}$ times the firm's tf-idf score for the corresponding term, $x_{i,v} \log(\frac{2134}{df_v})$, where $x_{i,v}$ is the count of term v in firm i's RF texts and df_v is the number of firms that use term v in their RF texts.[12] For example, top terms in predicting a positive return reaction to bad pandemic news for Netflix reflect its digital video services. The top terms for predicting a positive reaction for Novavax, which develops vaccines, include "vaccine" and "clinical trials." Marcus Corporation offers an example of a firm with exposures that create reinforcing negative return reactions, as it operates both hotels and cinemas.

5.2 Uncovering new risk exposures: A systematic approach

To construct risk exposures, we start with "seeds" drawn from (a) terms with large MNIR coefficients, $|\hat{b}_{1,v}|$; (b) terms with large tf-idf weighted MNIR coefficients, $|\hat{b}_{1,v}| x_v \log(\frac{2134}{df_v})$, where x_v is the count of term v in the RF corpus; and (c) terms with large values of $|\hat{b}_{1,v}| x_{i,v} \log(\frac{2134}{df_v})$ among firms with especially large fit improvements when adding the sufficient reduction projection to the regression (as in Table 2). We work with 45 seeds that reflect both positive and negative return reactions and that appear to cover the main exposures surfaced by our MNIR model fit to pandemic-related jump days. Table 3 reports the seeds and corresponding category names.

Next, we use the seeds to build sets comprised of related terms. Here, a typical NLP approach groups words based on semantic relatedness. We rely partly on that approach

[11] Figure B.8 shows that predicted returns for these firms move closer to actual returns as we go from a no-text model to one that includes dictionary-based measures to MNIR. Echoing results in Section 4, the two text-based approaches yield concordant results, but MNIR captures more information and yields better predictions.

[12] The $\log(\frac{2134}{df_v})$ expression down weights generic terms that are less useful in distinguishing among firms. 2134 is the number of unique firms in the data for which we have 10-K filings, so a term that appears in every firm's filings would get a weight of zero.

Company	Business description	Terms	tf-idf x MNIR coeff.
NOVAVAX INC ↗	Late-stage biotechnology company focused on the discovery, development and commercialization of vaccines to prevent serious infectious diseases.	vaccine	475·2
		influenza	297·5
		clinical trials	136·9
		candidates	99·4
		collaborators	96·1
NETFLIX INC ↗	World's leading internet television network with streaming memberships in over 190 countries.	dvd	943·2
		streaming	445·6
		subscribers	243·5
		titles	216·6
		studios	81·9
CEVA INC ↗	Leading licensor of wireless connectivity and smart sensing platforms.	ip	185·7
		irish	135·3
		royalty revenues	112·0
		royalty rates	95·5
		semiconductor	86·8
SURMODICS INC ↗	Leading provider of medical device and in vitro diagnostic technologies to the healthcare industry.	reagents	124·1
		incorporating	84·2
		licensees	42·7
		new applications	38·0
		coating	30·7
CADIZ INC ↗	Owner and developer of sustainable water and agricultural projects in California.	government appropriations	27·4
		railroad	24·9
		significant revenues	22·9
		value of options	13·6
		such plans	13·3
DOMINO'S PIZZA INC ↗	Multinational pizza restaurant chain with a large global network of franchise owners.	cheese	56·5
		quick	28·4
		foods	25·5
		pound	20·9
		interruption from earthquakes	16·8
RENEWABLE ENERGY GROUP INC ↘	North America's largest producer of advanced biofuels.	biodiesel	−2897·7
		biomass	−904·4
		feedstocks	−335·0
		gallons	−282·2
		soybean	−237·1
PLAINS ALL AMER PIPELNE -LP ↘	Provider of midstream energy infrastructure and logistics services for crude oil, natural gas liquids, natural gas and refined products.	crude	−1639·7
		ngl	−1499·2
		barrels per day	−1363·3
		pipeline	−815·5
		pipelines	−681·7
MARCUS CORP ↘	Owner and operator of real estate assets in the lodging and entertainment industries: movie theatres, hotels and resorts, a family entertainment center.	films	−216·8
		hotels	−215·2
		movie	−164·6
		film	−103·6
		patrons	−72·2
PBF ENERGY INC ↘	Petroleum refiner and supplier of unbranded transportation fuels, heating oil, petrochemical feedstocks, lubricants and other petroleum products.	refinery	−1122·3
		crude	−755·6
		refineries	−627·8
		feedstocks	−394·3
		refined products	−355·5

Table 2: Top Ten Firms with Greatest Fit Gains on Pandemic Fallout Days from MNIR Text Approach

This table lists the 10 firms with the greatest fit gain when adding the MNIR-based sufficient reduction production to a model with all of our non-text regressors. For each firm, we report the five most important terms in driving the gain in fit. We focus on terms with a positive (negative) MNIR coefficient if the firm-specific sufficient reduction projection is positive (negative), and we use tf-idf weighted scores to down weight generic terms.

and, to this end, deploy the popular word2vec model of word embeddings (Mikolov et al., 2013). This method inputs every sentence in our RF corpus and returns a vector representation of each of the 18,911 terms. We use these vectors to consolidate terms that are similar with respect to surrounding language in the corpus. This approach groups nearly identical terms like "pipeline" and "pipelines" and semantically similar terms like "cheese" and "foods." However, it neglects the fact that semantically related words can differ in the sign of their relationship to returns, as with "oil" and "wheat." Thus, we must account for both semantic similarity and relationship to returns. And, as in Table 2, we want to down weight generic terms that are less helpful in capturing exposures that differ among firms.

In view of these multiple considerations, we build new term sets by associating each seed with terms in the RF corpus that meet two criteria:

1. High specificity and same sign: Among terms v with an MNIR coefficient of the same sign as the one for the seed, we select those with $|\hat{b}_{1,v}| x_v \log(\frac{2134}{df_v}) > 200$.

2. High contextual similarity, as measured by cosine similarity of the embedding vectors. In practice, we require a term's embedding vector to have a cosine similarity greater than 0.4 with that of the seed.

Higher thresholds yield more suitable terms at the cost of excluding potentially relevant terms. Lower thresholds capture additional terms but at the cost of less suitability. We adjust the thresholds to strike a reasonable balance between these concerns. 8,513 seed-generated terms meet criterion 1. We then apply criterion 2 to obtain the set of semantically related terms that

Step 2 then searches within the set of terms from step 1 whose sign matches that of the seed for semantically related terms. In case the same term survives step 2 for more than one seed, we assign it to the seed for which its cosine similarity is highest. In total this approach selects 1,100 terms besides the 45 seeds.

Appendix C.1 lists all terms grouped with each seed under this approach. By and large, the groupings capture suitable terms that reflect coherent exposure categories,

Seed	Name	Retained Terms	Dropped Terms
advertisers	Advertizing	9	5
aircraft	Aircraft	10	10
biodiesel	Alternative Energy	10	17
card	Card Payments	25	0
clearing house	Clearing Houses	3	0
hotels	Commercial Property	18	3
gold	Currency Metals	2	5
display	Display Technology	16	13
pipelines	Energy Infrastructure	26	11
unrealized loss position	Financial Loss Management	15	0
yen	Foreign Exchange	5	0
franchisees	Franchising	13	0
gaming	Gambling	5	0
reinsurance	Insurance	23	0
surgeons	Medical Treatment	6	1
mortgage	Mortgages	44	0
oil	Oil and Gas	11	0
reit	REIT	29	0
homebuilding	Residential Construction	4	0
restaurants	Restaurants	3	13
retail	Retail	26	9
satellite	Satellite Communication	22	0
newspapers	Traditional Media	20	3
travel	Travel	11	6
workforces	Workforce	2	0

(a) Negative Exposures

Seed	Name	Retained Terms	Dropped Terms
bank	Banking	40	0
vessels	Containers	12	1
fdic	Deposits	20	0
preclinical	Drug Trials	43	0
ecommerce	Ecommerce	12	11
optics	Electronic Components	74	20
wheat	Foodstuffs	27	2
china	Foreign Countries	62	0
medicare	Health Insurance	35	0
investment funds	Investment Management	15	0
manufacturing	Manufacturing	35	5
steel	Metal Products	21	0
coal	Power Sources	13	0
tantalum	Raw Metals and Minerals	11	3
semiconductor	Semiconductor	15	5
software	Software Design	56	9
solutions	Software Services	65	8
freight	Transportation	21	0
games	Video Games	21	4
cloud	Web-Based Services	23	2

(b) Positive Exposures

Table 3: Targeted Exposures for Pandemic Fallout Dates

This table lists the 45 "seeds" that we use to construct MNIR-generated exposure categories and terms sets using the automated method described in the main text. The last two columns report how many automatically generated terms we retain and drop after judgmental pruning. We combine closely related exposure categories, as indicated by those with the same font color.

even when the seed is highly specific. For example, tantalum (a rare earth metal) is a key input into the manufacture of electronic circuits and equipment. The set seeded by "tantalum" contains other key inputs in the manufacturing supply chain for electronic equipment as well as "democratic republic of congo", a major site for rare earth mineral mining. For each seed we manually label the set of generated terms according to the type of exposure it captures, with labels displayed in Table 3.

Next we manually delete terms in each generated set that we judge do not convey sufficiently precise meaning, either because they are too generic or potentially refer to other concepts. This step relies on our own domain expertise and not on automation. Continuing with the above example, we delete the terms "adjoining countries", "requirements for companies", and "sheet" from the exposure associated with tantalum. After these deletions (fully described in appendix C.1) our targeted exposures contain 934 terms. Table 3 enumerates for each exposure the number of manual deletions we perform.

The final step in our construction is to generate firm-level exposures to each targeted risk factor j with associated word list $L(j)$ as $z_i^j = \sum_{v \in L(j)} x_{i,v} b_{1,v}$, which is a sub-component of the sufficient reduction projection based on the set of terms contained in $L(j)$. Table B.3 contains descriptive statistics for these exposures. Although quite specific, many targeted exposures are present in many firms' 10-K filings, which highlights their relevance in a variety of commercial contexts. To take one example, terms in "Web-Based Services" are present in over half the observations in the *RF* corpus, which suggests many firms are engaged at some level with this exposure across many industries.

Dependent Variable: Abn_{it}	(1) NAICS-2 Fixed Effects		(2) NAICS-2 Fixed Effects (comparable with 3)		(3) NAICS-4 Fixed Effects	
Targeted Exposures						
Advertizing	−0·09	(−2·4)	−0·10	(−2·2)	−0·12	(−3·0)
Alternative Energy	−0·10	(−6·8)	−0·09	(−8·7)	−0·05	(−1·9)
Card Payments	−0·14	(−3·3)	−0·12	(−3·2)	−0·17	(−4·8)
Clearing Houses	−0·10	(−9·7)				
Commercial Property					−0·15	(−2·3)
Currency Metals	−0·28	(−16·8)	−0·28	(−22·1)	−0·32	(−11·4)
Financial Loss Management	−0·23	(−11·5)	−0·24	(−12·8)	−0·29	(−3·5)
Foreign Exchange	−0·07	(−3·9)	−0·06	(−4·0)	−0·05	(−2·7)
Franchising	−0·10	(−1·8)	−0·12	(−3·2)	−0·15	(−2·2)
Gambling	−0·23	(−2·6)	−0·23	(−2·7)	−0·33	(−4·6)
Insurance	0·04	(2·1)	0·05	(2·4)		
Medical Treatment	−0·14	(−6·5)	−0·12	(−7·8)		
Mortgages	−0·11	(−3·3)	−0·13	(−5·6)		
REIT	−0·39	(−4·8)	−0·39	(−4·5)		
Residential Construction	−0·37	(−14·0)	−0·33	(−12·0)	−0·22	(−2·5)
Restaurants	−0·22	(−4·6)	−0·25	(−4·4)	−0·21	(−3·2)
Retail	−0·33	(−6·3)	−0·37	(−7·2)	−0·28	(−3·6)
Workforce	−0·19	(−3·1)	−0·20	(−2·9)	−0·20	(−3·3)
Oil and Gas + Energy Infr	−0·31	(−5·1)	−0·28	(−4·8)	−0·19	(−3·9)
Trad Media + Satellite Com	−0·09	(−2·4)	−0·09	(−2·3)	−0·11	(−2·9)
Travel + Aircraft	−0·24	(−2·7)	−0·25	(−2·9)		
Drug Trials	0·16	(11·4)	0·15	(10·7)	−0·04	(−2·7)
Ecommerce	0·15	(3·0)	0·15	(3·4)	0·14	(2·6)
Electronic Components	0·09	(4·1)	0·11	(4·2)	0·14	(3·6)
Foodstuffs	0·17	(4·3)	0·15	(4·9)	0·15	(4·8)
Foreign Countries	0·23	(2·7)	0·16	(1·8)		
Investment Management	0·22	(14·8)	0·22	(16·5)	0·21	(13·0)
Metal Products					−0·08	(−1·7)
Raw Metals and Minerals	0·29	(7·9)	0·28	(10·3)	0·26	(4·7)
Semiconductor					−0·07	(−2·0)
Video Games	0·12	(4·1)	0·10	(12·3)	0·11	(8·8)
Web-Based Services	0·22	(3·8)	0·20	(3·4)	0·21	(3·9)
Deposits + Banking	0·18	(5·4)	0·19	(5·1)	0·18	(4·0)
Financial Controls						
Log Market Cap	0·46	(4·4)	0·44	(4·1)	0·50	(6·2)
Leverage	−0·34	(−3·0)	−0·26	(−2·6)	−0·14	(−1·4)
Observations [Adjusted R^2]	2155	[0.410]	1868	[0.433]	1868	[0.470]

Table 4: Results for Targeted Exposures for Pandemic Fallout

Each column considers all targeted exposures for pandemic fallout dates (some of them are combined), except "Manufacturing". We also include log market cap and leverage. Additionally, columns 1 and 2 (3) consider 2-digit (4-digit) NAICS codes to introduce industry fixed effects and to cluster errors. For columns 2 and 3, we drop 4-digit NAICS codes with less than 5 companies. t stats are reported in parentheses; and, for presentation purposes, we omit the coefficients of regressors that are not significant at the 0.1 level. The pandemic fallout dates are Feb. 24, 25, 27 and March 03, 05, 11, 16, 18, 27. As a benchmark, note that estimating MNIR with all terms achieves an adjusted R^2 of 0.502 in the analogue forward regression.

5.3 Applying the new risk exposures

Column (1) of table 4 corresponds to the returns regression (3) but with the targeted exposures in place of the baseline dictionaries. This allows us to compare our targeted exposures in a comparable framework to the baseline dictionary exposures. Some exposures are strongly correlated, and in these cases we combine them into a single exposure by adding them together at the firm level. These combinations are color coded in table 3. We also do not include the "Manufacturing" exposure in the model due to its being generic as well as correlated with numerous more specific exposures, i.e. "Drug Trials" (correlation 0.56), "Foreign Countries" (correlation 0.48), and "Semiconductor" (correlation 0.45). We only report regression coefficients for exposures that are significant at the 10% level.

An initial observation is that there is a substantial improvement in fit from using the targeted exposures in place of the EMV dictionaries: the adjusted R^2 moves from 0.329 to 0.41. The targeted exposures also account for a substantial portion of the overall improvement in fit from estimating MNIR for the average firm-level abnormal return on pandemic fallout days using all terms, which produces an adjusted R^2 in the forward regression of 0.502.

The targeted exposures also provide a rich account of the varied impacts of bad pandemic news which aids in interpreting its impact. They capture the well-documented negative impact of exposure to consumer spending via "Retail" and "Card Payments", as well as the positive impact of exposure to online spending via "Ecommerce". Other exposures also suggest significant substitution patterns in non-durable consumption. "Foodstuffs" generates positive returns and "Restaurants" negative returns, which is consistent with a shift from market to home production of meals, an effect which is also present in observed consumer spending patterns in various countries. "Video Games" generates positive returns and "Gambling" and "Travel + Aircraft" both generate negative returns, which indicates a shift in recreation time.

Exposure to intermediate good and service provision linked to downstream demand shocks is also an important driver of returns. Exposure to energy via "Alternative Energy" and "Oil and Gas + Energy Infrastructure" is associated with negative returns, while the opposite is true for exposure to the technology supply chain via "Raw Metals and Minerals", "Electronic Components", and "Web-Based Services". The magnitudes of the estimated coefficients are comparable to those on final good consumption.

The impacts of different financial exposures also illustrate notable heterogeneity. "Mortgages" generates negative returns, part of a broader trend of low returns for property-related exposures also seen in "REIT" and "Residential Construction". Exposure to "Financial Loss Management" also leads to low returns. On the other hand, exposure to "Deposits + Banking" and "Investment Management" is positive, a plausible explanation for which is the increased savings rates seen among higher-income households during the pandemic.

Column (3) of table 4 introduces NAICS4 industry effects in place of NAICS2 effects. For comparison, column (2) reports results using NAICS2 effects on the same sample considered by column (3). While some of the exposures become insignificant, the majority do not (and some previously insignificant exposures become significant). The implication is that the targeted exposures are not a proxy for narrow industry membership but capture firm-level characteristics important for explaining the pandemic's impact on earnings.

Overall, then, we observe that the COVID-19 shock generates a wide array of positive and negative effects across firms. The richness of the text data in the *RF* corpus allows us to uncover dozens of separate effects that each play a role in explaining the overall structure of returns, including direct exposures to the demand shocks induced by social distancing as well as indirect exposures via supply chain linkages.

5.4 Targeted risk exposures for Super Tuesday

Dependent Variable: Abn_{it}	(1) NAICS-2 Fixed Effects		(2) NAICS-2 Fixed Effects (comparable with 3)		(3) NAICS-4 Fixed Effects	
Targeted Exposures						
Aircraft					−0·08	(−3·1)
Card Payments	−0·04	(−2·4)	−0·04	(−2·1)	−0·09	(−2·7)
Financial Contracting	−0·15	(−3·2)	−0·22	(−2·9)	−0·19	(−2·3)
Foodstuffs	−0·11	(−4·7)	−0·13	(−4·1)	−0·16	(−4·5)
Gambling	−0·20	(−7·4)	−0·18	(−6·2)	−0·16	(−2·8)
Hotels	−0·25	(−8·9)	−0·25	(−8·1)	−0·26	(−4·9)
Industrial Metals	−0·09	(−1·8)	−0·07	(−3·2)		
Motor Vehicles					−0·14	(−3·2)
Power Generation	−0·19	(−4·2)	−0·17	(−4·2)	−0·16	(−2·1)
Shipping	−0·21	(−4·8)				
Traditional Media	−0·15	(−8·0)	−0·14	(−8·9)		
Transportation	−0·08	(−3·9)	−0·07	(−3·2)		
Asset Mngmt + Gen Investment	−0·19	(−9·5)	−0·16	(−7·6)	−0·18	(−4·1)
Financial Regul + Banking	−0·18	(−7·5)	−0·18	(−5·4)	−0·11	(−3·4)
Fracking + Unconv Drilling	−0·19	(−2·0)				
Construction	0·22	(2·4)	0·28	(3·8)		
Drugs	0·13	(3·0)	0·13	(1·8)	0·26	(5·2)
Electronic Communication	0·28	(3·7)	0·29	(3·9)		
Foreign	0·08	(2·0)	0·12	(2·2)		
Franchising	0·11	(2·6)				
Government Contracting			0·18	(2·0)	0·17	(2·1)
Metals	0·16	(3·7)	0·20	(3·3)	0·23	(6·1)
Military	0·09	(2·6)				
Reinsurance	0·13	(8·7)	0·13	(5·5)		
REIT	0·43	(8·7)	0·42	(5·8)		
Rental Market	0·26	(3·1)	0·32	(7·6)	0·30	(9·7)
Utilities	0·18	(7·6)	0·19	(8·2)	0·16	(4·1)
Waste	0·16	(6·2)	0·13	(4·5)		
Health Ins + E-comm + Subsidies	0·20	(4·0)	0·17	(2·5)	0·20	(3·0)
Medical Infr + Gov Healthcare	0·30	(1·8)				
Financial Controls						
Log Market Cap	0·63	(4·8)	0·59	(4·1)	0·59	(4·5)
Leverage	−0·10	(−0·8)	−0·07	(−0·5)	−0·16	(−1·5)
Observations [Adjusted R^2]	2155	[0.242]	1868	[0.261]	1868	[0.308]

Table 5: Results for Targeted Exposures for the Super Tuesday Aftermath

Each column considers all targeted exposures for super tuesday (some of them are combined). We also include log market cap and leverage. Additionally, columns 1 and 2 (3) consider 2-digit (4-digit) NAICS codes to introduce industry fixed effects and to cluster errors. For columns 2 and 3, we drop 4-digit NAICS codes with less than 5 companies. t stats are reported in parentheses; and, for presentation purposes, we omit the coefficients of regressors that are not significant at the 0.1 level. As a benchmark, note that estimating MNIR with all terms achieves an adjusted R^2 of 0.349 in the analogue forward regression.

6 Discussion and Conclusion

We tap the *Risk Factors* discussions in 10-K filings to investigate how market-moving news affected the structure of firm-level equity returns in the wake of the COVID-19 pandemic. We find that *Risk Factors* texts helps interpret the enormous heterogeneity in firm-level responses that one observes on days with the arrival of news about the development of the pandemic. Some of the effects we find are in line with those found in other studies: there are large negative impacts on firms exposed to travel and out-of-home recreation, and positive impacts on firms exposed to e-commerce. Our environment also allows us to find knock-on effects of these direct impacts onto firms exposed to risks associated with the supply of intermediate goods. Again, these exposures are negative and positive depending on whether the downstream demand shock is negative or positive. The implication is that the full impact of the crisis stretches throughout the supply chain and beyond where the most obvious effects lie.

To arrive at these conclusions, we draw on two text analytic methods often portrayed as competitors, dictionary methods and supervised learning. When used separately, both have important limitations. The dictionaries we use are specifically designed to capture sources of equity market volatility, but are still broad enough to create ambiguities when interpreting their relationship to returns. This shows the difficulty of obtaining sufficiently narrow word lists to pin down precise effects of news on returns. On the other hand, while the supervised learning method generates a higher cross-sectional R^2 than the dictionaries do, the feature space it operates on is so vast that the estimated model alone is not immediately helpful for interpreting returns.

Only by combining machine learning with our own judgments do we make progress in interpreting the effects of pandemic news. We use our domain expertise both in picking the seed words that form the basis of our new exposures, and then again to reorganize word lists into coherent categories that are generated through an automated method in the first instance. It is this interplay between human input and automated methods that ultimately allows us to provide a clean characterization of risk exposures.

In other words, there are important complementary strengths of dictionary methods and machine learning that our paper shows can be powerful when used in tandem. This insight is the main methodological contribution of the paper, and we hope it stimulates more work in this direction in the future.

References

Acemoglu, D., Akcigit, U., and Kerr, W. (2016). Networks and the Macroeconomy: An Empirical Exploration. *NBER Macroeconomics Annual, University of Chicago Press*, 30(1):273–335.

Alfaro, L., Chari, A., Greenland, A. N., and Schott, P. K. (2020). Aggregate and Firm-Level Stock Returns During Pandemics, in Real Time. Working Paper 26950, National Bureau of Economic Research.

Andersen, A., Hansen, E. T., Johannesen, N., and Sheridan, A. (2020). Consumer Responses to the COVID-19 Crisis: Evidence from Bank Account Transaction Data. CEPR Discussion Paper 14809.

Atalay, E. (2017). How Important are Sectoral Shocks? *American Economic Journal: Macroeconomics*, 9(4):254–280.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., and Kost, K. J. (2019). Policy News and Stock Market Volatility. *National Bureau of Economic Research Working Paper Series*, (w25720).

Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., and Viratyosin, T. (2020a). The Unprecedented Stock Market Reaction to COVID-19. *The Review of Asset Pricing Studies, Forthcoming.*

Baker, S. R., Farrokhnia, R., Meyer, S., Pagel, M., and Yannelis, C. (2020b). How Does Household Spending Respond to an Epidemic? Consumption During the 2020 COVID-19 Pandemic. *The Review of Asset Pricing Studies, Forthcoming.*

Baqaee, D. and Farhi, E. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, 87, 4:1155–1203.

Barrero, J. M., Bloom, N., and Davis, S. J. (2020). COVID-19 Is Also a Reallocation Shock. *Brookings Papers on Economic Activity, Forthcoming.*

Carvalho, V. M., García, J. R., Ortiz, Á., Rodrigo, T., Mora, R., and Ruiz, P. (2020). Tracking the COVID-19 Crisis with High-Resolution Transaction Data. CEPR Discussion Paper 14642.

Carvalho, V. M. and Tahbaz-Salehi, A. (2019). Production Networks: A Primer. *Annual Review of Economics*, 11:635–663.

Chetty, R., Friedman, J. N., Hendren, N., and Stepner, M. (2020). How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data. Working Paper.

Cox, N., Ganong, P., Noel, P., Vavra, J., Wong, A., Farrell, D., and Greig, F. (2020). Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Saving Data. Becker Friedman Institute Working Paper 2020-82.

Davis, S. J. (2017). Regulatory Complexity and Policy Uncertainty: Headwinds of Our Own Making. Working Paper 2723980, Becker Friedman Institute Working Paper.

Davis, S. J. and Haltiwanger, J. (1992). Gross Job Creation, Gross Job Destruction, and Employment Reallocation. *The Quarterly Journal of Economics*, 107(3):819–863.

Gentzkow, M., Kelly, B., and Taddy, M. (2019a). Text as Data. *Journal of Economic Literature*, 57(3):535–574.

Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019b). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, 87(4):1307–1340.

Hanley, K. W. and Hoberg, G. (2019). Dynamic Interpretation of Emerging Risks in the Financial Sector. *The Review of Financial Studies*, 32(12):4543–4603.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2020). Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1. Working Paper 26971, National Bureau of Economic Research.

IMF (2020). World Economic Outlook Update. Technical report, International Monetary Fund.

Justeson, J. S. and Katz, S. M. (1995). Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, 1:9–27.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting Returns With Text Data. Working Paper 26186, National Bureau of Economic Research.

Laeven, L. (2020). Pandemics, Intermediate Goods, and Corporate Valuation. CEPR Discussion Paper 15022.

Long, J. B. and Plosser, C. I. (1983). Real Business Cycles. *Journal of Political Economy*, 91:39–69.

Lopez Lira, A. (2019). Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns. *SSRN Electronic Journal*.

Loughran, T. and Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.

Manning, C. D., Raghavan, P., and Shuetze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press.

Mast, J. T., Palmer, P. S., and Hickok, R. L. (2020). Operating in a Pandemic: Securities Litigation Risk and Navigating Disclosure Concerns.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Pagano, M., Wagner, C., and Zechner, J. (2020). Disaster Resilience and Asset Prices. *SSRN Electronic Journal*.

Papanikolaou, D. and Schmidt, L. D. (2020). Working Remotely and the Supply-side Impact of Covid-19. Working Paper 27330, National Bureau of Economic Research.

Ramelli, S. and Wagner, A. F. (2020). Feverish Stock Price Reactions to COVID-19. *The Review of Corporate Finance Studies*.

Stephany, F., Stoehr, N., Darius, P., Neuhauser, L., Teutloff, O., and Braesemann, F. (2020). The CoRisk-Index: A data-mining approach to identify industry-specific risk assessments related to COVID-19 in real time. *Covid Economics*, (41):129–146.

Surico, P., Kanzig, D., and Hacioglu, S. (2020). Consumption in the Time of COVID-19: Evidence from UK Transaction Data. CEPR Discussion Paper 14733.

Taddy, M. (2013). Multinomial Inverse Regression for Text Analysis. *Journal of the American Statistical Association*, 108(503):755–770.

Taddy, M. (2015). Distributed Multinomial Regression. *The Annals of Applied Statistics*, 9(3):1394–1414.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, LXII(3):30.

A Sample and Feature Space Construction

A.1 Sample of firms

The following are the details on how we construct our analysis sample:

- We link 3,154 firms (i) with at least one 10-K filing (with a non-empty Part 1a) from January 2010 to July 2016, and (ii) with equity return data for all business days between Feb 24, 2020 and March 27, 2020.

- We remove 19 firms with no leverage information.

- In order to compute abnormal returns, we first need to get estimates of stock-level betas. Hence, we keep stocks for which we have at least 125 daily return observations in 2019. We lose 28 firms in this step.

- We also drop small caps: either because they are in the first quartile of equity market value or because their share price is smaller than 5 dollars on February 21, 2020 (i.e. the last trading day before the stock market jump dates we consider in this paper). Dropped small caps account for 2.5 percent of total equity market value in the sample. In this step, we remove 968 firms.

- We discard 5 companies with no available NAICS code in our dataset. Finally, we keep only NAICS codes with at least 5 companies. We drop one firm in this last step.

- We end up with an analysis sample of 2,155 stocks for 2,133 companies.

A.2 Text preprocessing details

Our feature space construction begins with 10Ks from 2010 to 2016 scraped from EDGAR for 3,580 unique firms. Some of these firms are not part of our final sample, as explained above, we use them because they are potentially informative about the structure of language in the *Risk Factors* texts.

We first find and replace meaningful phrases in the 10-K corpus with a single term in the feature space. For example, ‘We owe additional income tax’ becomes ‘We owe additional income_tax’, where ‘income_tax’ is treated as an individual term. This ensures that the meaning conveyed by key phrases is retained in our analysis. These phrases come from multiple sources:

1. 433 phrases from the baseline dictionaries in Baker et al. (2019).

2. 3,803 phrases that correspond to named entities that appear more than 25 times in the corpus. We identify these entities with the named entity recognizer (NER) from the Stanford NLP group. The NER finds an additional 63 entities that also appear in the dictionaries, and so are redundant.

3. 9,649 additional multi-word expressions (MWE). To identify these, we first tag all words in the corpus using a part-of-speech tagger from the Stanford NLP group, and then tabulate tag patterns likely to correspond to meaningful sequences Justeson and Katz (1995). Our final set of MWE is the resulting trigrams that appear more than 150 times in the corpus, and bigrams that appear more than 500 times. This approach finds an additional 68 phrases also present in the dictionaries, and 265 phrases also present in named entities, and so are redundant.

We then follow standard steps to complete pre-processing:

- Lowercase all text (case-folding).

- Tokenize text by breaking it into individual terms. Continuing from the above example, the tokenized representation of ‘We owe additional income_tax’ would be the four-element list [‘we’, ‘owe’, ‘additional’, ‘income_tax’].

- Drop common words from a standard stopword list, e.g. ‘for’, ‘to’, etc.

- Drop any terms that appear in the *Risk Factors* text of fewer than 25 firms from 2010 to 2016.

A.3 Baseline Dictionary Categories and Terms

- Broad Quantity Indicators: {gdp, economic growth, depression, recession, economic crisis, industrial production}
- Inflation: {cpi, inflation, gold, silver}
- Interest Rates: {interest rates, yield curve}
- Credit Indicators: {bank loans, mortgage loans, credit spread, consumer credit, business credit}
- 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}[13]
- Real Estate Markets: {housing prices, home prices, homebuilding, homebuilders, housing starts, home sales, building permits, mortgages, residential construction, commercial construction, commercial real estate, real estate}
- Business Investment and Sentiment: {business investment, business confidence}
- Consumer Spending and Sentiment: {consumer spending, retail sales, consumer purchases, consumer confidence, consumer sentiment}
- Commodity Markets: {wheat, corn, sugar, cotton, beef, pork, petroleum, oil, coal, natural gas, biofuel, ethanol, steel, copper, zinc, tin, platinum, gold, metal, silver, aluminum, lead, commodity exchange, nymex, mercantile exchange, gas pipeline}
- Financial Crises: {financial crisis, financial crises}
- Exchange Rate: {exchange rate, currency devaluation}

[13]Terms in blue-font are not included in the 244-term dictionary currently considered by MNIR. We will add these in the next version of the draft.

- Healthcare Matters: {healthcare, health insurance, medicaid, medicare, affordable care act, medical malpractice, prescription drug, food and drug administration, fda, 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 law, federal trade commission, ftc, monopoly, hart scott rodino, european commission}
- Labor Disputes: {labor dispute, labor unrest, strike}
- Intellectual Property Matters: {patent, trademark, copyright, patent and trademark office, international trade commission, federal trade commission, ftc, intellectual property, hatch waxman, new drug application}
- Taxes: {taxes, tax, taxation, taxed, income tax, payroll tax, unemployment tax, sales tax, excise tax, value added tax, vat, carbon tax, corporate tax, business tax, accelerated depreciation, research and development tax credit, property tax, fiscal cliff, internal revenue service}
- Government Spending, Deficits and Debt: {government spending, government appropriations, defense spending, federal budget, government budget, debt ceiling, fiscal cliff, government shutdown, sovereign debt}
- Entitlement and Welfare Programs: {social security, disability insurance, medicaid, medicare, unemployment insurance, affordable housing}
- Monetary Policy: {monetary policy, money supply, open market operations, discount window, quantitative easing, central bank, federal reserve, the fed, european central bank}

- Financial Regulation: {financial reform, truth in lending, sarbanes oxley, dodd frank, tarp, troubled asset relief program, volcker rule, basel, capital requirement, stress test, deposit insurance, fdic, office of thrift supervision, ots, comptroller of the currency, occ, commodity futures trading commission, cftc, financial stability oversight council, securities and exchange commission, sec, bureau of consumer financial protection, consumer financial protection bureau, cfpb}

- Competition Policy: {competition law, federal trade commission, ftc, hart scott rodino, european commission}

- Intellectual Property Policy: {patent law, trademark law, copyright law, patent and trademark office, international trade commission}

- Labor Regulations: {department of labor, national labor relations board, minimum wage, workers compensation, occupational safety and health administration, osha, mine safety and health administration, at will employment, affirmative action, equal employment opportunity, erisa, pension benefit guaranty corporation, pbgc}

- Energy and Environmental Regulation: {energy policy, carbon tax, cap and trade, offshore drilling, pollution controls, environmental restrictions, clean air act, clean water act, environmental protection agency, epa, federal energy regulatory commission, ferc, endangered species, greenhouse gas regulation, climate change regulation, nuclear regulatory commission, pipeline and hazardous materials safety administration}

- Lawsuit and Tort Reform, Supreme Court Decisions: {supreme court}

- Housing and Land Management: {federal housing administration, department of housing and urban development, hud, bureau of land management, department of interior, zoning regulations, zoning laws, endangered species}

- Other Regulation: {consumer product safety commission, department of education, small business administration, federal communications commission, fcc, fish and wildlife service}

- Generic Regulation: {regulation, regulatory, regulate}

- National Security: {national security, war, military conflict, military action, terrorism, terror, defense spending, department of defense, department of homeland security, armed forces}

- Trade Policy: {tariff, dumping, world trade organization, north american free trade agreement, international trade commission}

- Healthcare Policy: {healthcare policy, health insurance, medicaid, medicare, affordable care act, national institutes of health}

- Food and Drug Policy: {food and drug administration, fda}

- Transportation, Infrastructure and Public Utilities: {department of transportation, national highway traffic safety administration, corps of engineers, federal aviation administration, faa, nasa, pipeline and hazardous materials safety administration}

- Elections and Political Governance: {presidential election}

- Agricultural Policy: {department of agriculture, usda}

B Additional Tables and Figures

B.1 Material for section 2

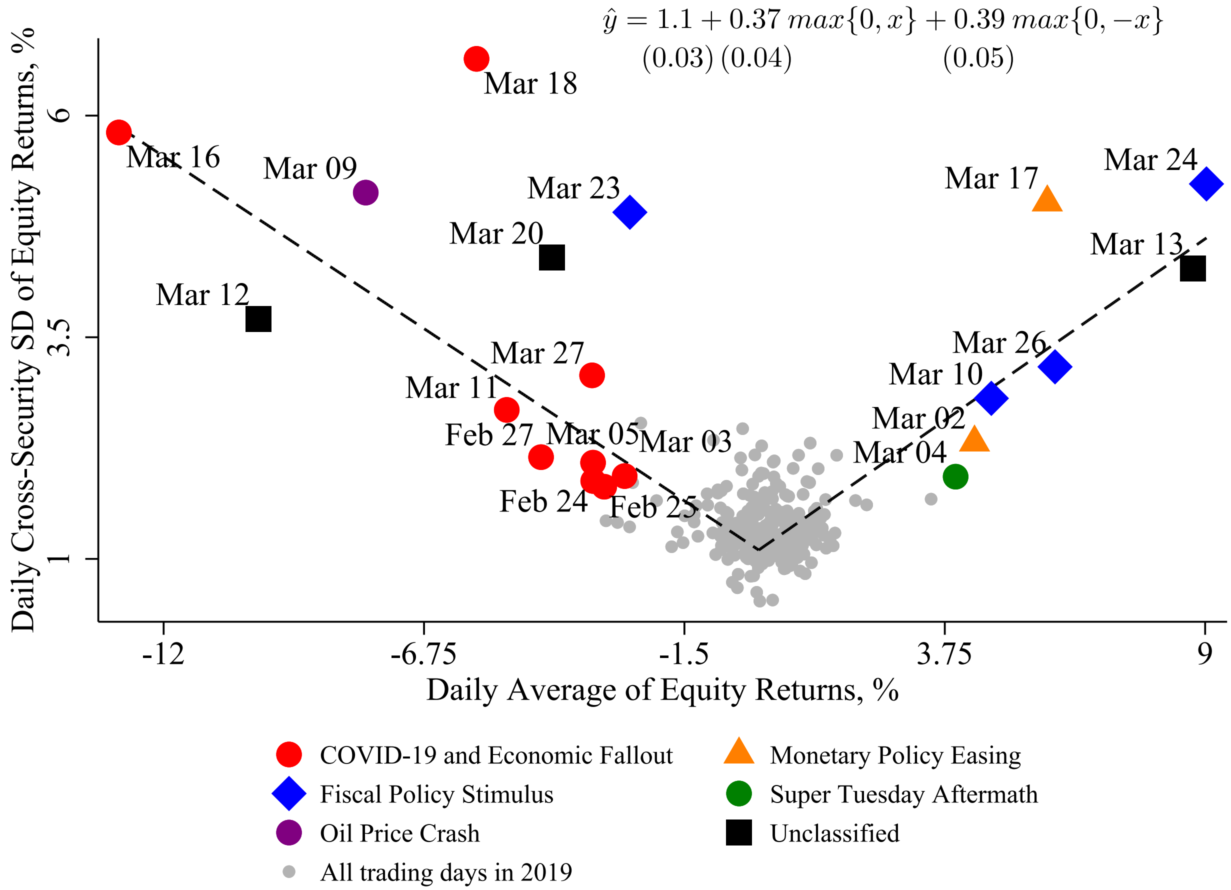


Figure B.1: Value-Weighted Mean and Cross-Sectional SD of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

We consider the value-weighted distribution of daily returns over 2,155 stocks for trading days in 2019 and jump days in 2020. The mean (s.d.) of the daily average return for trading days in 2019 is 0.12 (0.80) percent, and the mean (s.d.) of the daily SD is 1.34 (0.34). The regression has 271 observations and an R-squared of 0.61, with standard errors in parentheses. A test of the null hypothesis that the two rays have equal slopes with opposite signs yields a p-value of 0.66.

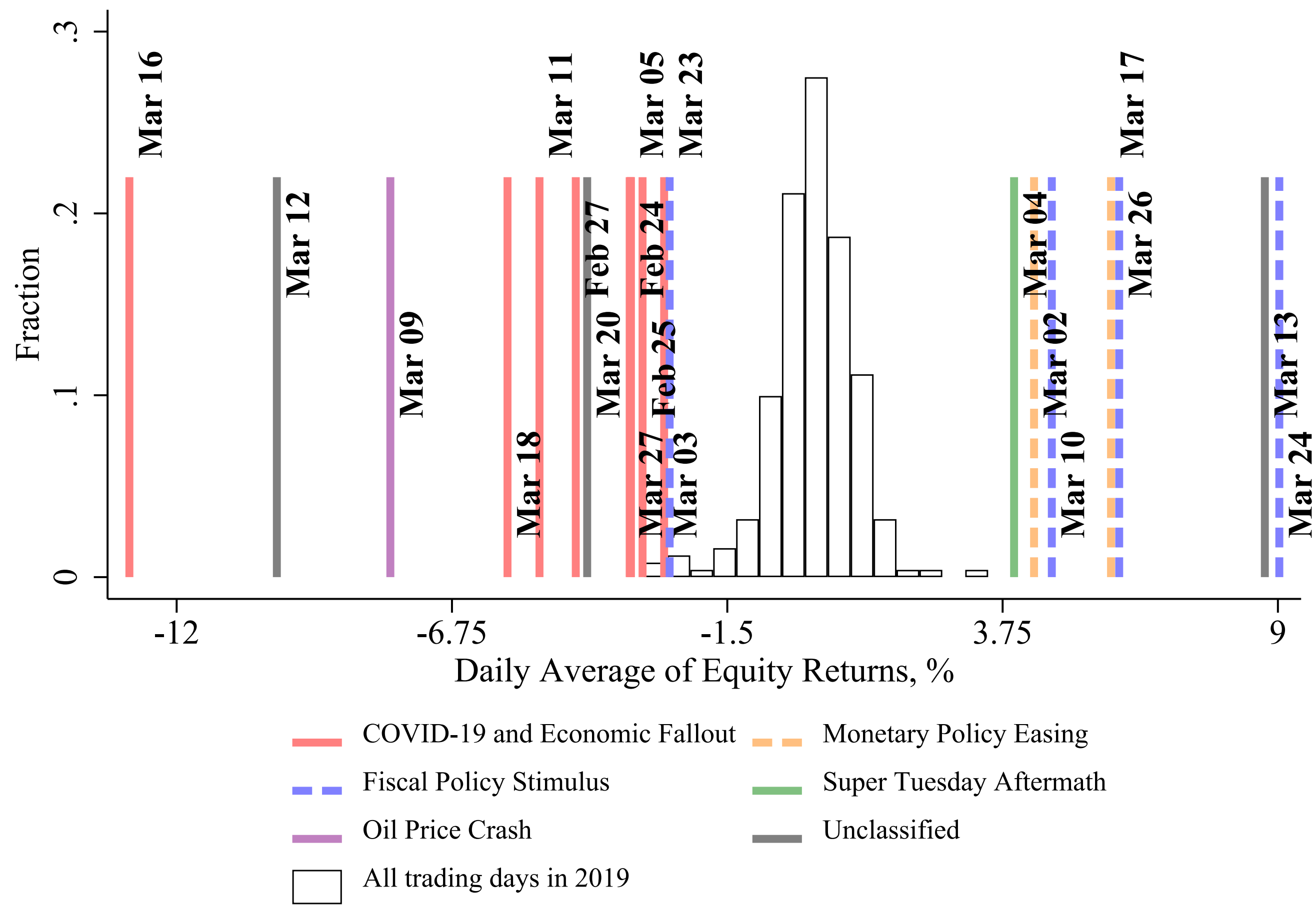


Figure B.2: Value-Weighted Mean of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

This Figure considers all 2,155 securities in our analysis sample. The daily cross-sectional average is value-weighted (i.e. by equity market value). The mean (s.d.) of the daily average return for trading days in 2019 is 0.12 (0.80) percent. The heights on the vertical axis refer to normalized values for the histogram of 2019 returns.

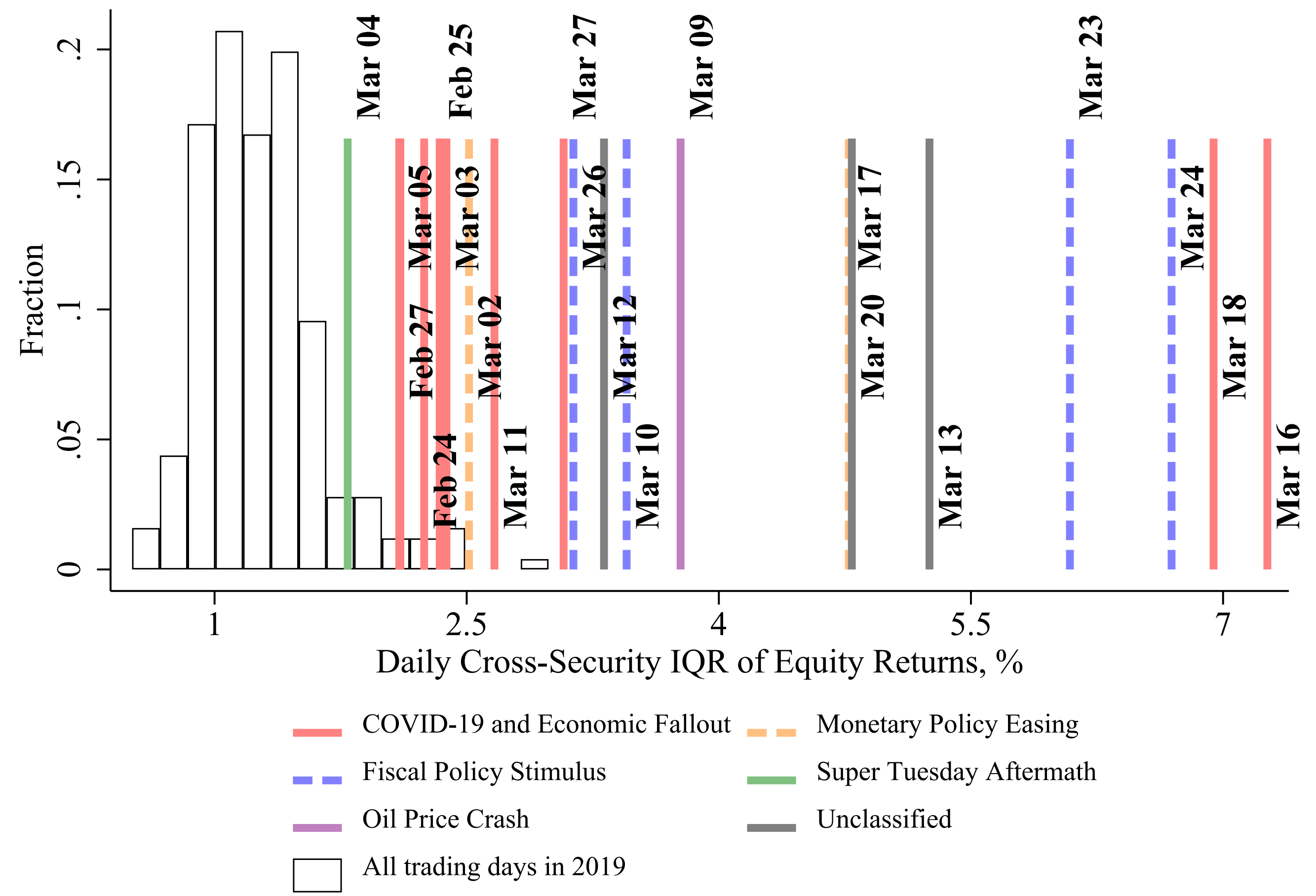


Figure B.3: Value-Weighted Cross-Sectional IQR of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

This Figure considers all 2,155 securities in our analysis sample. The daily cross-sectional IQR is value-weighted (i.e. by equity market value). The mean (s.d.) of the daily IQR for trading days in 2019 is 1.28 (0.36) percent. The heights on the vertical axis refer to normalized values for the histogram of 2019 returns.

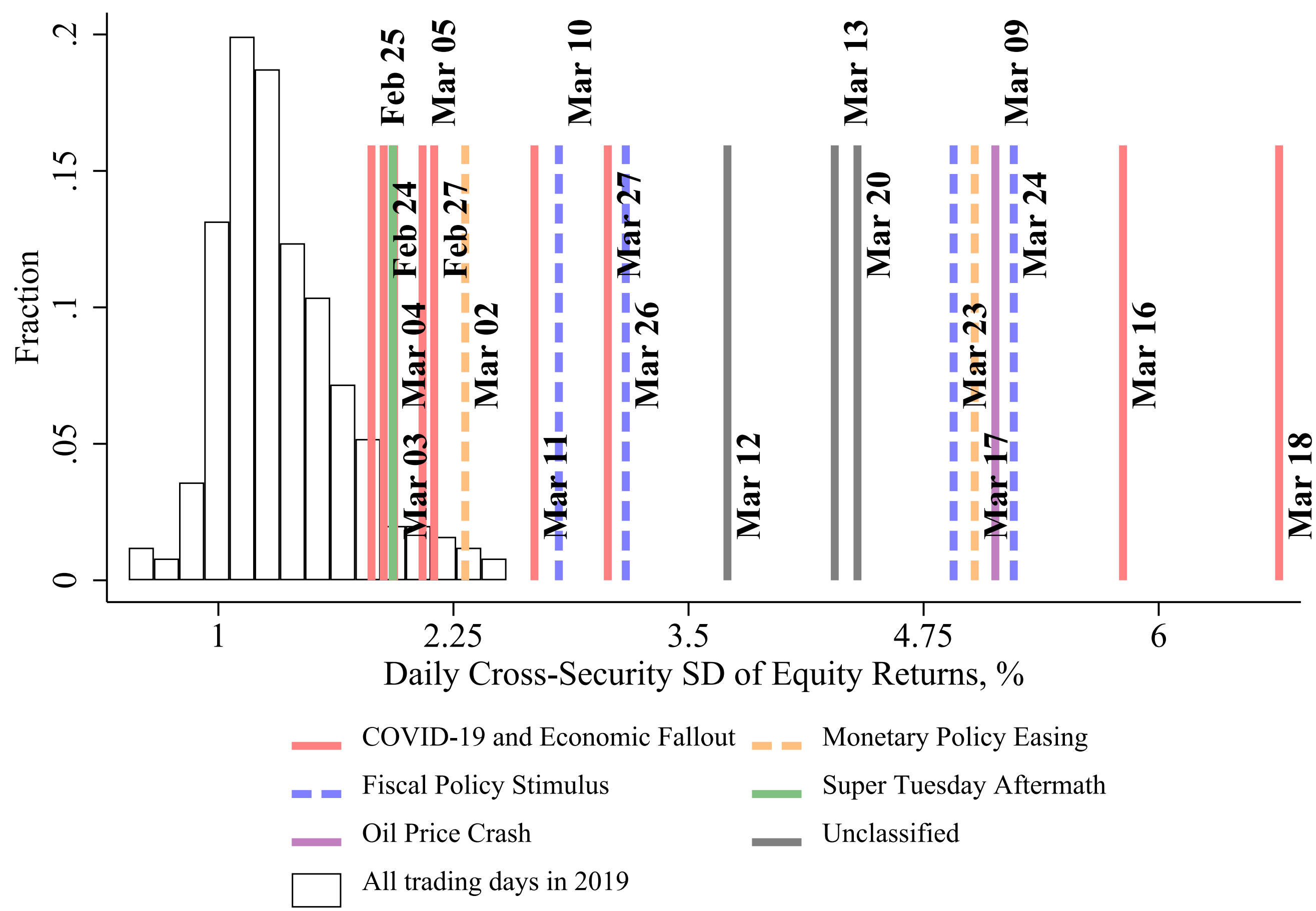


Figure B.4: Value-Weighted Cross-Sectional SD of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

This Figure considers all 2,155 securities in our analysis sample. The daily cross-sectional SD is value-weighted (i.e. by equity market value). The mean (s.d.) of the daily SD for trading days in 2019 is 1.34 (0.34) percent.

Variables	N	% > 0	Mean		SD		p1	p99
			All	> 0	All	> 0		
Abn. Returns and Financial Controls								
Percent Daily Abn. Return	36635	45.4	-1.1	3.7	7.1	4.3	-25.9	15.8
Log Market Cap	36635	100.0	21.5	21.5	1.7	1.7	18.6	25.9
Leverage	2155	96.3	0.3	0.3	0.3	0.3	0.0	1.1
General Economic Categories								
Broad Quantity Indicators	2155	60.6	0.3	0.6	0.6	0.6	0.0	2.6
Inflation	2155	53.3	0.3	0.6	0.7	0.9	0.0	2.9
Interest Rates	2155	80.3	1.7	2.1	2.3	2.3	0.0	10.0
Credit Indicators	2155	33.6	0.4	1.3	1.3	1.9	0.0	7.0
Labor Markets	2155	92.2	1.2	1.3	1.4	1.4	0.0	6.5
Real Estate Markets	2155	50.8	2.3	4.5	4.7	5.8	0.0	20.5
Business Investment and Sentiment	2155	7.6	0.0	0.3	0.1	0.2	0.0	0.6
Consumer Spending and Sentiment	2155	46.1	0.3	0.8	0.7	0.9	0.0	3.9
Commodity Markets	2155	96.1	3.0	3.1	5.3	5.4	0.0	27.4
Financial Crises	2155	22.2	0.1	0.3	0.2	0.3	0.0	0.9
Exchange Rate	2155	56.6	0.6	1.0	0.9	0.9	0.0	3.7
Healthcare Matters	2155	44.5	1.9	4.3	4.5	6.0	0.0	19.9
Litigation Matters	2155	94.4	2.1	2.2	1.7	1.7	0.0	7.6
Competition Matters	2155	32.0	0.1	0.4	0.3	0.5	0.0	1.6
Labor Disputes	2155	38.5	0.2	0.6	0.5	0.6	0.0	2.1
Intellectual Property Matters	2155	65.0	2.8	4.3	3.8	4.0	0.0	16.2
Policy-Related Categories								
Taxes	2155	95.5	3.4	3.6	3.4	3.4	0.0	14.4
Government Spending, Deficits, Debt	2155	32.1	0.1	0.4	0.3	0.5	0.0	1.8
Entitlement and Welfare Programs	2155	23.0	0.4	1.8	1.6	3.0	0.0	9.0
Monetary Policy	2155	26.2	0.3	1.3	1.0	1.6	0.0	4.7
Financial Regulation	2155	90.5	2.1	2.3	3.1	3.2	0.0	15.2
Competition Policy	2155	24.8	0.1	0.4	0.3	0.4	0.0	1.2
Intellectual Property Policy	2155	21.7	0.1	0.4	0.2	0.3	0.0	1.0
Labor Regulations	2155	34.7	0.2	0.6	0.7	1.0	0.0	2.6
Energy and Environmental Regulation	2155	27.1	0.5	1.8	1.5	2.3	0.0	7.7
Lawsuit and Tort Reform, Supreme Court Decisions	2155	16.4	0.0	0.2	0.1	0.3	0.0	0.6
Housing and Land Management	2155	14.6	0.1	0.4	0.3	0.8	0.0	0.8
Other Regulation	2155	12.1	0.2	1.8	1.4	3.7	0.0	8.2
Generic Regulation	2155	99.7	7.5	7.6	4.1	4.1	0.8	18.8
National Security	2155	74.2	0.5	0.7	0.7	0.7	0.0	3.3
Trade Policy	2155	36.8	0.2	0.6	0.5	0.6	0.0	2.0
Healthcare Policy	2155	30.6	0.6	1.8	2.1	3.5	0.0	10.3
Food and Drug Policy	2155	17.4	0.8	4.3	2.3	3.8	0.0	10.8
Transportation, Infrastructure, Utilities	2155	11.5	0.1	0.6	0.4	1.0	0.0	1.6
Elections and Political Governance	2155	1.3	0.0	0.1	0.0	0.1	0.0	0.0
Agricultural Policy	2155	4.1	0.0	0.6	0.2	0.8	0.0	0.7

Table B.1: Descriptive statistics

All statistics are unweighted and, except for log market cap and leverage, reported in percents. Sample sizes for returns and market cap refer to the 17 dates used in our regression models for returns.

B.2 Additional material for section 3

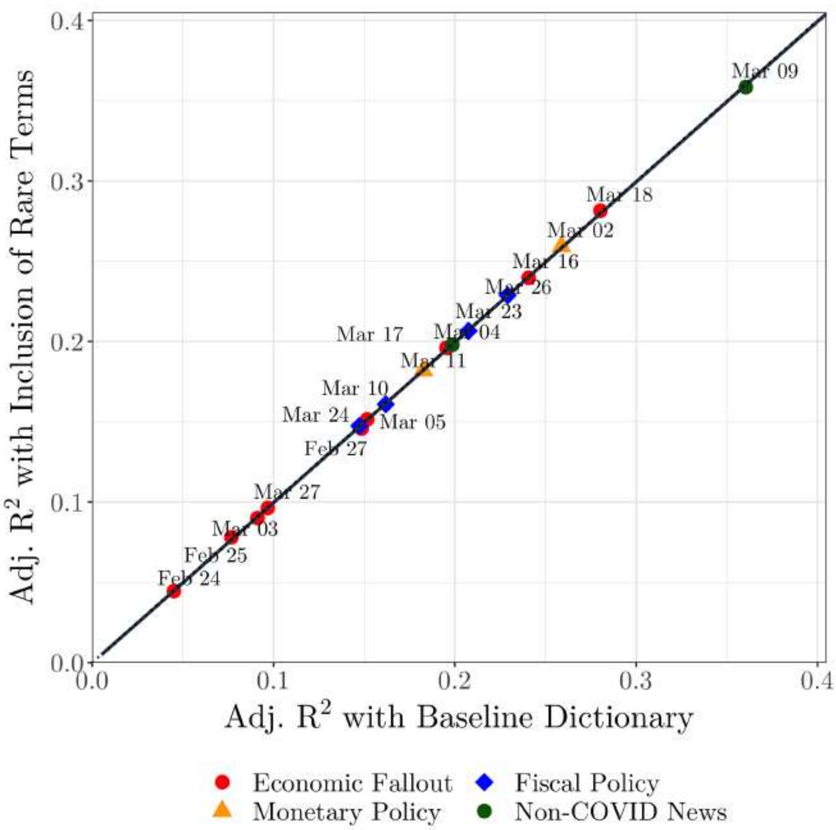
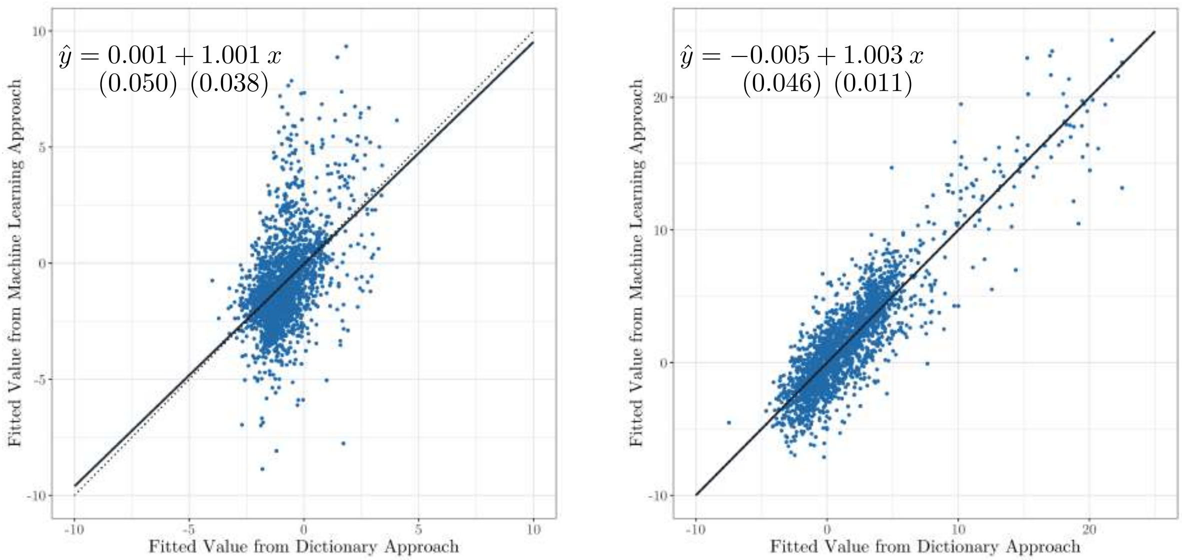


Figure B.5: Adjusted R^2 for Baseline and Complete Dictionaries

The horizontal axis displays the adjusted R^2 from 17 separate OLS regressions of abnormal returns on firm controls and the baseline exposures, one for each day that enters the event groupings. These baseline exposures are computed using the 244 terms in Baker et al. (2019) that appear in the preprocessed *RF* corpus. The vertical axis displays the adjusted R^2 from the same regressions but with the dictionary exposures computed using the 430 terms that appear in the original dictionaries. The inclusion or not of rare terms is inconsequential for goodness-of-fit.

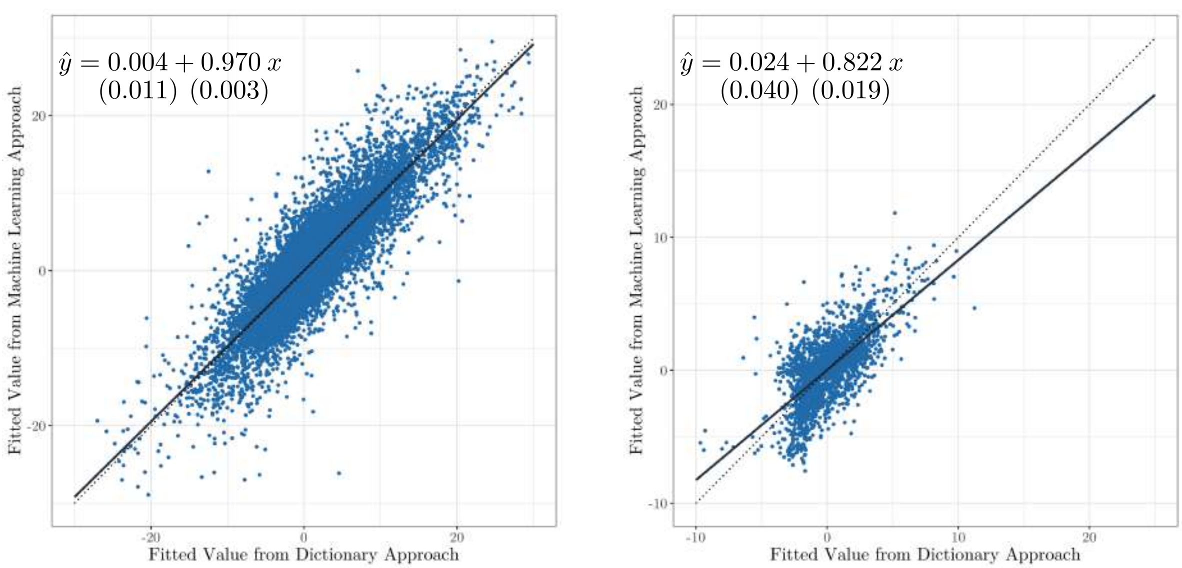
B.3 Additional material for section 4



(a) Day with Lowest Dictionary R^2 **(b)** Day with Highest Dictionary R^2

Figure B.6: Comparison of Fitted Values on 24 February and 9 March

This figure plots fitted values from (3) and (4) estimated on 24 February (left panel) and 9 March 9 (right panel). These days have the lowest and highest R^2 values, respectively, under the dictionary approach. The black solid lines are fitted regressions, and the dashed line is the 45 degree line. The scales of the x- and y-axes differ between the two panels.



(a) Pooled Days **(b)** March 24

Figure B.7: Comparing Fitted Values between Approaches without Sector Fixed Effects

This figure plots fitted values from regressions (3) and (4) without NAICS2-level fixed effects. The sufficient reduction project in (4) is built from coefficient estimates in the inverse regression model (2) that also does not include NAICS2 effects. The left panel plots the fitted values from both approaches across all days, and the right panel plots the fitted values for a day on which the fitted returns do not display a one-for-one relationship. The black solid lines are fitted regression lines, and the dashed line is the 45 degree line.

B.4 Additional material for section 5

B.4.1 Pandemic fallout day material

Pandemic Fallout Raw Coeffs	Pos	**wheat**	iceland	intel	acres of land	assays	salesforce	consoles	eye	optics	games
		ecommerce	private equity funds	alcohol	technology platform	game	santa clara	eus	virtual	tractors	purchased power
		graphics	titles	intercept	subscriptions	questionable	serial	new members	ecosystem	**health insurance**	cellular
	Neg	clearing house	establishments	patrons	atms	airline industry	delta	homebuyers	atm	slot	tribal
		homebuilding	visa	generations	**silver**	phillips	boeing	reuse	jet fuel	american express	corporate office
		spirit	cruise	capitalizations	master lease	jet	circulation	pos	unrealized loss position	native	negotiable
Pandemic Fallout Weighted Coeffs	Pos	product candidates	games	clinical trials	clients	**fda**	**health insurance**	drug candidates	**patent**	the corporation	**steel**
		fdic	solutions	patents	patients	our products	freight	clinical	cellular	game	alcohol
		drugs	collaborators	**intellectual property**	subscriptions	container	**wheat**	data center	platform	commercialize	carriers
	Neg	reit	gaming	hotels	crude	our manager	hotel	unitholders	restaurants	general partner	pipelines
		distributions	trs	restaurant	satellite	reinsurance	tenants	**fcc**	pipeline	**mortgage loans**	refinery
		oil	management agreement	homes	ngl	franchisees	aircraft	**homebuilding**	card	satellites	common units

Table B.2: Top and Bottom Terms on Pandemic Fallout Days

The top half of the table lists the 30 terms with the highest and lowest estimated coefficients in the MNIR model fit to pandemic fallout days, after screening out terms that appear in the *RF* texts of more than two-thirds of firms. The bottom half weights the MNIR coefficient by the corresponding term's tf-idf score, computed as $x_v \log(\frac{2134}{df_v})$ where x_v is the count of term v in the *RF* corpus and df_v is the number of firms that use term v in any of their *RF* texts. We then display the top and bottom 30 terms according to this weighted coefficient. Terms in bold also appear in the EMV dictionaries.

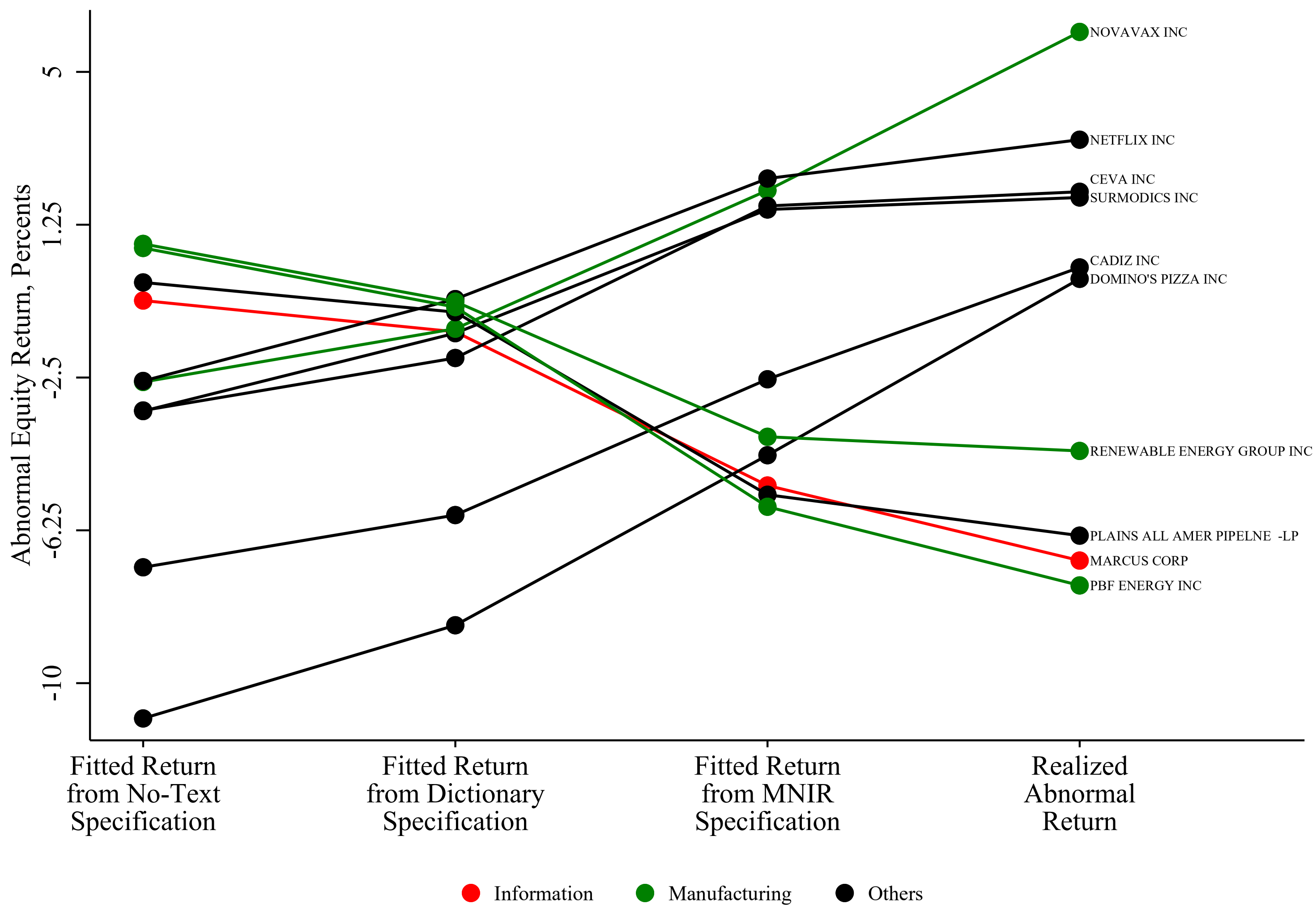


Figure B.8: Predicted and Actual Returns for Firms with Greatest MNIR Fit Gains on Pandemic Fallout Days

This figure considers ten firms with the greatest gain in fit by adding the sufficient reduction projection to a model fit to data for the 9 pandemic fallout jump dates listed in the notes to Table 1. We consider security-level averages over these jump dates and obtain predicted values for abnormal returns for each indicated specification. Dot colors show the firm's industry. For firms with multiple equity securities, we retain the one with the highest market cap on 21 February 2020.

Risk Exposure	N	% > 0	Mean All	Mean > 0	SD All	SD > 0	p1	p99
Advertizing	2155	39.3	282.9	719.8	1507.1	2338.5	0.0	6457.6
Aircraft	2155	12.3	498.8	4041.4	5023.5	13811.4	0.0	10360.5
Alternative Energy	2155	25.1	351.0	1400.8	3600.4	7094.4	0.0	4690.2
Card Payments	2155	40.6	844.4	2077.2	6997.3	10861.3	0.0	8899.6
Clearing Houses	2155	20.9	92.6	443.6	1273.5	2761.3	0.0	1220.5
Commercial Property	2155	77.2	1538.8	1994.1	7227.5	8172.6	0.0	25057.3
Currency Metals	2155	10.4	144.5	1389.8	2031.6	6174.9	0.0	1836.3
Display Technology	2155	38.8	406.0	1046.5	4362.2	6958.2	0.0	6035.2
Energy Infrastructure	2155	42.9	1434.4	3341.8	9783.0	14721.7	0.0	29597.1
Financial Loss Management	2155	81.9	761.0	929.7	2174.9	2371.2	0.0	12055.7
Foreign Exchange	2155	16.4	33.4	203.5	376.6	911.4	0.0	435.2
Franchising	2155	29.5	630.1	2138.5	3147.2	5515.5	0.0	12465.6
Gambling	2155	9.1	430.5	4709.3	4997.3	15943.5	0.0	6012.7
Insurance	2155	55.7	681.5	1222.9	4485.0	5953.5	0.0	16684.5
Medical Treatment	2155	16.1	103.2	641.2	846.5	2028.6	0.0	2385.7
Mortgages	2155	76.5	1785.7	2333.6	6784.8	7673.9	0.0	37103.0
Oil and Gas	2155	47.0	1226.3	2608.8	6513.0	9310.1	0.0	27892.1
REIT	2155	76.1	2353.0	3091.9	7200.9	8115.5	0.0	41279.2
Residential Construction	2155	22.5	230.8	1027.8	1488.6	3010.2	0.0	4249.8
Restaurants	2155	8.7	452.4	5213.9	5140.2	16763.7	0.0	14325.8
Retail	2155	60.7	424.8	699.9	1294.6	1603.0	0.0	5931.1
Satellite Communication	2155	36.4	942.0	2589.3	8006.2	13117.2	0.0	22547.5
Traditional Media	2155	49.5	541.7	1095.1	4026.1	5672.6	0.0	8409.4
Travel	2155	37.6	488.2	1297.2	2914.5	4640.9	0.0	8724.6
Workforce	2155	6.6	21.0	316.5	125.1	378.6	0.0	556.7
Banking	2155	82.4	1441.5	1749.1	5712.6	6250.0	0.0	29317.7
Containers	2155	24.8	414.4	1669.0	3590.1	7063.3	0.0	6487.6
Deposits	2155	19.4	447.8	2308.6	1943.6	3899.5	0.0	11453.8
Drug Trials	2155	67.7	2105.4	3109.7	8056.9	9632.0	0.0	39575.5
Ecommerce	2155	71.7	437.9	610.4	1620.6	1885.8	0.0	6109.7
Electronic Components	2155	73.0	833.8	1142.3	3368.0	3897.5	0.0	12615.3
Foodstuffs	2155	45.8	441.6	963.2	3551.3	5198.2	0.0	7587.8
Foreign Countries	2155	90.0	1607.8	1786.0	2999.9	3111.1	0.0	15099.5
Health Insurance	2155	71.0	1504.9	2121.0	7508.5	8841.2	0.0	21070.0
Investment Management	2155	30.7	440.3	1435.6	3266.0	5777.6	0.0	7765.1
Manufacturing	2155	85.0	1388.2	1633.8	2524.5	2664.5	0.0	11474.1
Metal Products	2155	51.1	518.3	1013.5	4799.9	6676.2	0.0	5493.3
Power Sources	2155	39.6	266.4	672.3	1643.5	2558.8	0.0	4655.1
Raw Metals and Minerals	2155	22.1	157.5	711.4	789.3	1557.1	0.0	2039.7
Semiconductor	2155	23.9	260.6	1092.4	1369.4	2638.9	0.0	7368.5
Software Design	2155	92.8	1092.4	1177.6	2475.5	2550.7	0.0	11332.8
Software Services	2155	95.8	2588.0	2702.1	4989.6	5068.1	0.0	24711.4
Transportation	2155	55.6	801.0	1439.6	4787.3	6347.2	0.0	12928.7
Video Games	2155	26.6	735.4	2765.7	9209.3	17713.0	0.0	11449.9
Web-Based Services	2155	67.9	909.0	1339.0	3441.8	4108.1	0.0	13521.4

Table B.3: Descriptive statistics, Targeted Exposures for Pandemic Fallout

All statistics are unweighted.

B.4.2 Super Tuesday material

Super Tuesday Raw Coeffs	ecommerce	utility operations	navy	health benefits	display	undeveloped land	aggregates	nonresidential	service territories	radar	Pos
	new members	shield	flex	landfills	health plans	subsidy	**health insurance**	customer care	**silver**	dec	
	medical costs	backstop	management programs	abbott	tactical	outreach	socio	**gold**	pharmacy	personal health information	
	sweep	hotel properties	behalf of clients	market mutual funds	casino	resorts	permian basin	vessel	lng	private equity funds	Neg
	millions of dollars	nfa	las vegas	mmcf	room rates	reservations	hotel rooms	lodging	drilling rigs	fractionation	
	wellhead	condensate	lincoln	vessels	sweet	bakken	morgan stanley	rigs	hotel	production companies	
Super Tuesday Weighted Coeffs	tenants	reit	**medicare**	**fda**	display	**health insurance**	tenant	**medicaid**	carriers	reinsurance	Pos
	healthcare	students	properties	ric	incentive fee	**coal**	patients	sba	franchisees	mining	
	drug candidates	portfolio	**real estate**	drugs	health care	cms	**department of education**	operating partnership	restaurants	electric	
	natural gas	**oil**	crude	hotels	hotel	clients	gaming	drilling	unitholders	pipelines	Neg
	the corporation	common units	pipeline	client	aircraft	vessels	**fdic**	midstream	partnership	offshore	
	general partner	**ferc**	casino	ngl	lng	refinery	hotel properties	**ethanol**	ngls	wells	

Table B.4: Top and Bottom Terms by MNIR Coefficient and TF-IDF Weighted MNIR Coefficient on Super Tuesday

The top half of this table lists the 30 terms with the highest and lowest estimated coefficient in the MNIR model on Super Tuesday, after screening out terms that appear in the *RF* texts for more than two-thirds of firms. The bottom half weights the MNIR coefficients by the corresponding term's tf-idf score, computed as $x_V \log(\frac{2134}{df_v})$ where x_v is the count of term v in the *RF* corpus and df_v is the number of firms that use term v in any of their *RF* texts. We then display the top and bottom 30 terms according to this weighted coefficient. Terms in bold also appear in the baseline dictionaries.

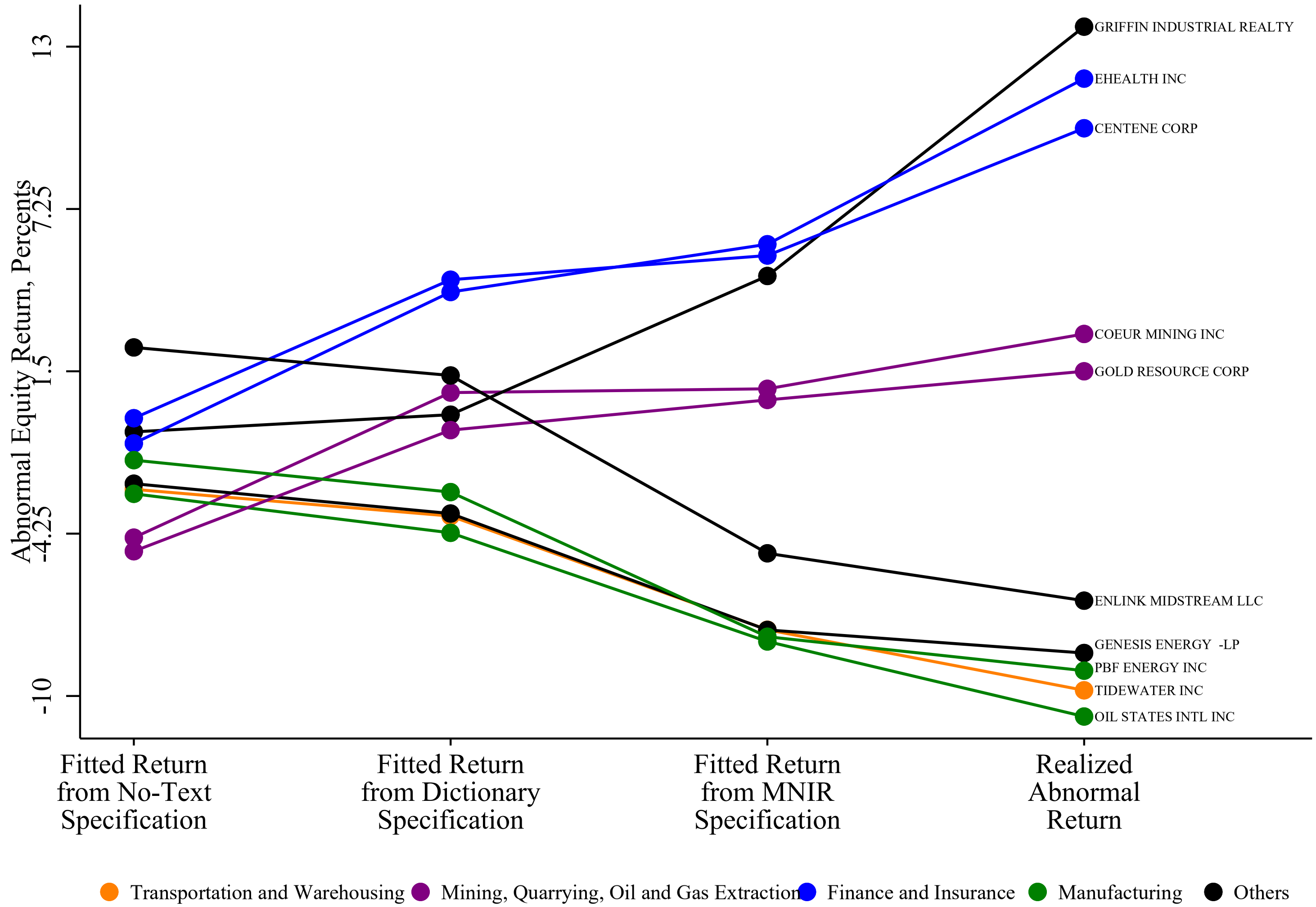


Figure B.9: Ten Firms with Largest Increase in Fit from No-Text to MNIR Specification for Super Tuesday

This Figure considers the jump date associated to the super tuesday aftermath. Based on each specification in the horizontal axis categories, we generate fitted abnormal returns at the security-level. Then, we keep only the securities with the highest market cap for each of the 2,133 firms on 21 Feb 2020, and plot the ten securities for which MNIR achieves the largest reduction in absolute residuals (compared to the no-text specification).

Company	Business Description	Terms	tf-idf x MNIR coeff.
GRIFFIN INDUSTRIAL REALTY ↗	Real estate business focused on developing, acquiring, managing and leasing industrial/warehouse properties.	undeveloped land connecticut square feet flex hartford	488·7 316·5 273·8 251·3 226·5
EHEALTH INC ↗	Online marketplace for health insurance and Medicare plans.	health insurance ecommerce carriers medicare carrier	3189·4 1177·5 946·9 718·0 372·7
CENTENE CORP ↗	Managed healthcare enterprise that serves as intermediary to government-sponsored and commercial healthcare programs, focusing on under-insured and uninsured individuals.	medicaid health plans health plan chip care	656·1 444·8 405·2 380·5 324·2
COEUR MINING INC ↗	Gold, silver, zinc, and lead producer.	silver gold mining mine ore	1007·7 599·5 551·4 316·7 157·7
GOLD RESOURCE CORP ↗	Producer of metal concentrates that contain gold, silver, copper, lead and zinc.	silver gold mine mining mexican	683·6 576·8 244·9 182·9 97·7
ENLINK MIDSTREAM LLC ↘	Provider of midstream energy services (e.g. fractionating, processing, transporting, storing, and selling) to the oil and gas industries.	midstream condensate partnership mmcf gathering	−1886·9 −658·8 −455·3 −388·9 −380·0
GENESIS ENERGY -LP ↘	Provider of midstream energy services (e.g. transportation, storage, and processing) to the oil and gas industry and producer of natural soda ash.	crude refinery pipelines oil unitholders	−695·3 −489·0 −387·9 −356·7 −343·9
PBF ENERGY INC ↘	Petroleum refiner and supplier of unbranded transportation fuels, heating oil, petrochemical feedstocks, lubricants and other petroleum products.	crude refinery refineries receivable agreement oil	−888·2 −768·9 −407·3 −394·7 −257·6
TIDEWATER INC ↘	Provides marine support and transportation services to the global offshore energy industry through the operation of a fleet of marine service vessels.	vessels vessel offshore deepwater crude	−2027·6 −1467·1 −722·8 −429·6 −300·0
OIL STATES INTL INC ↘	Provider of highly engineered oilfield products and services for the drilling, completion, subsea, production and infrastructure sectors of the oil and natural gas industry.	accommodations offshore oil drilling sands	−584·8 −576·1 −352·7 −348·3 −339·1

Table B.5: Ten Firms with Largest Increase in Fit
from No-Text to MNIR Specification for Super Tuesday Aftermath

This table lists the 10 firms with the greatest fit gain when adding the MNIR-based sufficient reduction production to a model with all of our non-text regressors. For each firm, we report the five most important terms in driving the gain in fit. We focus on terms with a positive (negative) MNIR coefficient if the firm-specific sufficient reduction projection is positive (negative), and we use tf-idf weighted scores to down weight generic terms.

Seed	Name	Retained Terms	Dropped Terms
aircraft	Aircraft	3	6
private equity funds	Asset Management	22	0
bank	Banking	35	16
pipelines	Energy Infrastructure	20	15
derivatives	Financial Contracting	34	0
fdic	Financial Regulation	23	0
wheat	Foodstuffs	11	14
hydraulic fracturing	Fracking	13	3
gaming	Gambling	6	0
investment funds	Generic Investment	10	0
hotels	Hotels	6	5
steel	Industrial Metals	17	8
vehicles	Motor Vehicles	12	8
oil	Oil and Gas	13	0
mastercard	Payment Systems	20	0
emissions	Pollution	40	0
electricity	Power Generation	34	1
vessels	Shipping	15	1
tariff	Tariffs	5	2
broadcast	Traditional Media	19	0
fleet	Transportation	9	1
travel	Travel	8	3
deepwater	Unconventional Drilling	21	22

(a) Negative Exposures

Seed	Name	Retained Terms	Dropped Terms
homebuilding	Construction	5	0
radar	Defense Technology	17	38
drugs	Drugs	34	13
ecommerce	Ecommerce	2	1
students	Education	9	9
mobile	Electronic Communication	35	10
mexican	Foreign	9	1
franchisees	Franchising	6	0
government contracts	Government Contracting	10	2
medicare	Government Healthcare	31	2
health insurance	Health Insurance	9	5
hospitals	Medical Infrastructure	20	5
gold	Metals	5	4
navy	Military	9	2
mining	Mining	11	0
reit	REIT	21	0
properties	Real Estate	22	0
reinsurance	Reinsurance	25	0
space	Rental Market	14	0
restaurants	Restaurants	2	8
subsidy	Subsidies	4	2
utility operations	Utilities	6	2
games	Video Games	6	1
landfills	Waste	5	5

(b) Positive Exposures

Table B.6: Targeted Exposures for Super Tuesday

This table enumerates our targeted risk factors for Super Tuesday. See notes in table 3.

Risk Exposure	N	%	Mean		SD		p1	p99
		> 0	All	> 0	All	> 0		
Aircraft	2155	6.2	146.6	2358.3	1738.2	6609.0	0.0	3716.1
Asset Management	2155	46.5	387.1	832.5	2142.6	3083.5	0.0	5416.6
Banking	2155	79.4	1166.6	1469.3	3943.0	4374.9	0.0	20950.8
Card Payments	2155	38.5	322.4	837.0	3176.0	5077.2	0.0	3451.2
Energy Infrastructure	2155	34.4	971.2	2820.7	7228.4	12110.3	0.0	21095.1
Financial Contracting	2155	81.0	848.1	1047.3	2305.8	2521.5	0.0	9460.2
Financial Regulation	2155	26.5	545.2	2057.8	2258.3	4019.3	0.0	12677.6
Foodstuffs	2155	22.6	167.5	741.2	1722.9	3568.0	0.0	2205.6
Fracking	2155	15.2	194.6	1278.8	1275.6	3054.0	0.0	5802.6
Gambling	2155	10.8	362.1	3349.4	4130.2	12179.1	0.0	5030.3
Generic Investment	2155	25.4	210.2	826.5	1437.8	2762.3	0.0	4179.3
Hotels	2155	11.8	684.8	5810.3	6394.7	17839.7	0.0	16468.0
Industrial Metals	2155	32.5	330.0	1014.3	2562.9	4417.8	0.0	5183.6
Motor Vehicles	2155	41.9	425.8	1017.3	2263.0	3411.9	0.0	7835.6
Oil and Gas	2155	45.9	2191.5	4775.2	11002.2	15860.5	0.0	50502.8
Pollution	2155	45.3	639.7	1411.1	2404.0	3415.5	0.0	12030.2
Power Generation	2155	87.1	1013.7	1163.9	4801.6	5128.0	0.0	21215.2
Shipping	2155	22.5	485.2	2160.4	4797.9	9951.6	0.0	9518.3
Tariffs	2155	40.4	85.7	212.1	569.5	881.0	0.0	1608.0
Traditional Media	2155	41.9	446.3	1064.0	4030.7	6172.2	0.0	7638.5
Transportation	2155	25.2	241.6	957.0	1883.3	3658.4	0.0	4980.9
Travel	2155	26.7	263.0	983.9	1552.1	2883.3	0.0	5673.8
Unconventional Drilling	2155	24.8	750.6	3029.0	5160.9	10036.2	0.0	20445.6
Construction	2155	23.5	153.1	650.9	841.6	1640.3	0.0	3261.4
Defense Technology	2155	43.5	229.1	526.5	1453.1	2167.4	0.0	3807.4
Drugs	2155	33.3	988.5	2966.8	4058.8	6604.0	0.0	20952.7
Ecommerce	2155	40.2	55.3	137.5	762.0	1197.0	0.0	524.2
Education	2155	19.8	205.4	1036.6	3071.0	6842.8	0.0	1180.4
Electronic Communication	2155	79.9	646.9	809.6	2772.9	3080.9	0.0	15197.0
Foreign	2155	30.7	135.1	439.9	894.8	1573.1	0.0	1953.0
Franchising	2155	17.4	526.9	3020.0	2314.3	4818.2	0.0	11579.0
Government Contracting	2155	46.1	155.1	336.3	647.6	921.2	0.0	2885.5
Government Healthcare	2155	61.4	1178.1	1917.5	5090.2	6384.8	0.0	18830.0
Health Insurance	2155	47.7	996.8	2089.7	5815.8	8285.9	0.0	12604.8
Medical Infrastructure	2155	35.3	598.6	1697.4	2685.8	4313.3	0.0	10471.9
Metals	2155	25.2	136.9	542.5	1687.3	3327.6	0.0	1661.7
Military	2155	43.7	220.8	505.6	1347.5	2004.2	0.0	3657.3
Mining	2155	33.6	450.4	1340.6	4327.4	7388.8	0.0	7619.3
Real Estate	2155	87.1	1724.9	1981.4	4321.3	4576.4	0.0	22134.3
Reinsurance	2155	72.9	684.3	938.7	3840.2	4471.4	0.0	18170.4
REIT	2155	50.4	975.0	1932.9	3242.1	4358.4	0.0	16733.3
Rental Market	2155	48.4	435.2	900.0	1739.4	2416.9	0.0	7045.4
Restaurants	2155	8.1	125.7	1556.2	1432.6	4828.6	0.0	4256.1
Subsidies	2155	33.7	71.9	213.5	287.3	463.6	0.0	973.6
Utilities	2155	24.4	81.0	332.4	611.9	1206.3	0.0	2088.7
Video Games	2155	11.8	114.5	971.8	1736.8	4984.6	0.0	938.3
Waste	2155	17.7	98.4	556.4	1084.6	2532.3	0.0	1311.4

Table B.7: Descriptive statistics, Targeted Exposures for Super Tuesday

All statistics are unweighted.

C Term Sets for Targeted Exposures

Here we list all terms associated with the targeted exposures for pandemic fallout dates and Super Tuesday. For each exposure, we first provide our chosen name followed by the set of terms representing the exposures in curly braces. The term marked by the asterisk is the seed term for building the set. Bold terms are also present in the EMV dictionaries. Our manual deletions are the terms with strike-through marks. Terms are ordered within sets according to their cosine similarity with the seed term in the embedding space.

C.1 Pandemic fallout days

Exposures associated with negative returns:

1. Advertizing: {advertisers*, advertiser, audience, audiences, ~~guests~~, advertising, ~~end customers~~, advertising revenue, ~~patrons~~, digital media, ~~subscriber~~, marketers, advertising expenditures, ~~buyers~~}

2. Aircraft: {aircraft*, ~~vehicles~~, commercial aircraft, boeing, flight, airlines, ~~trucks~~, ~~ships~~, ~~vehicle~~, ~~rig~~, **faa**, jet, ~~spare parts~~, flights, fly, ~~drilling rig~~, passenger, ~~replacement parts~~, ~~machines~~, ~~passengers~~}

3. Alternative Energy: {biodiesel*, **ethanol**, fuels, **corn**, biomass, ~~gasoline~~, ~~refined products~~, diesel, biofuels, ~~fuel~~, ~~feedstocks~~, ~~feedstock~~, ~~propane~~, ~~gallon~~, ~~refined~~, alternative fuel, ~~poultry~~, ~~jet fuel~~, ~~gallons~~, ~~refiners~~, ~~asphalt~~, ~~produced~~, ~~aggregates~~, ~~crude~~, alternative energy sources, ~~petrochemical~~, renewable}

4. Clearing Houses: {clearing house*, clearing, futures}

5. Commercial Property: {hotels*, hotel properties, hotel, properties, resorts, retail properties, property, such properties, shopping centers, ~~communities~~, commercial property, rooms, ~~homes~~, new properties, ~~management agreements~~, land parcels, such property, real properties, other properties, suites, management companies}

6. Currency Metals: {**gold***, **silver**, ~~concentrates~~, ~~ore~~, ~~sweet~~, ~~lumber~~, ~~el~~}

7. Display Technology: {display*, displays, format, digital, signage, displayed, screens, navigation, ads, ~~seat~~, ~~compact~~, interactive, radar, ~~machine~~, video, ~~signal~~, film, multimedia, cameras, ~~log~~, films, ~~pads~~, ~~meter~~, ~~filters~~, ~~crystal~~, ~~coupons~~, ~~wall~~, ~~ad~~, ~~turning~~}

8. Energy Infrastructure: {pipelines*, pipeline systems, pipeline, gathering systems, pipeline system, processing plants, storage tanks, processing facilities, terminals, storage facilities, gathering, refineries, ~~interstate~~, **gas pipeline**, terminal, ~~transportation systems~~, downstream, ~~intrastate~~, transmission facilities, gas processing, common carrier, ~~rail~~, ~~transportation~~, ~~plants~~, gas gathering, fractionation, ~~shippers~~, refinery, **ferc**, ~~dock~~, ~~gulf coast~~, ~~routes~~, wells, transmission system, midstream, ~~unloading~~, generation facilities}

9. Financial Loss Management: {unrealized loss position*, unrealized losses, fixed maturity securities, unrealized loss, fixed maturity, unrealized, investment portfolio, otti, fixed income securities, temporary impairments, loss position, market value, fair value, decline in value, portfolio}

10. Foreign Exchange: {yen*, canadian dollar, british pound sterling, rupee, dollar value}

11. Franchising: {franchisees*, franchisee, franchise, franchisors, franchised, franchise agreements, landlords, lessees, franchisor, franchise agreement, tenants, franchising, anchor tenants}

12. Gambling: {gaming*, casino, slot, horse, native}

13. Insurance: {reinsurance*, reinsurers, reinsurance agreements, reinsurance arrangements, ceded, reinsurance contracts, reinsured, reinsurer, commercial insurance, catastrophe, insurers, insurance policies, mortgage insurance, coverages,

insurer, captive, insurance policy, insureds, cost of reinsurance, casualty, statutory surplus, insurance company, insurance operations}

14. Medical Treatment: {surgeons*, hospitals, dentists, dental, clinics, pathology, ~~training programs~~}

15. Mortgages: {mortgage*, residential mortgage, **mortgages**, mortgage loan, commercial mortgage, certain mortgage, **mortgage loans**, other mortgage, residential mortgage loan, rmbs, loan, cmbs, mbs, abs, federal home loan mortgage corporation, **ginnie mae**, mortgage lending, **federal national mortgage association**, commercial mortgage loan, mortgage financing, other loans, subprime, securitized, first mortgage, such loans, first lien, agency securities, mortgage origination, securitization, mortgage market, originations, loan sales, origination, securitizations, asset, borrowers, mortgage banking, servicer, gse, backed, mortgaged, mortgage industry, **federal housing administration**, fha}

16. Oil and Gas: {**oil***, ngls, ngl, oils, liquids, **natural gas**, **petroleum**, hydrocarbon, hydrocarbons, marcellus shale, exploration}

17. Payment Systems: {card*, cards, credit card, visa, mastercard, debit, merchant, merchants, credit cards, cardholder, card issuers, card transactions, cardholders, atm, american express, electronic payment, interchange, payment services, pos, check, gift, interchange fees, pci, atms, point of sale}

18. REIT: {reit*, ric, reits, reit status, reit qualification, taxable reit subsidiary, taxable reit subsidiaries, trss, gross income test, trs, bdc, reit income, internal revenue, income test, reit distribution, partnership, income tests, taxable years, qualify, asset tests, hedge accounting treatment, gross income tests, gross income, reit gross income, investment company, **income tax**, distribution requirement, taxable year, spin}

19. Residential Construction: {**homebuilding***, **residential construction**, land

development, housing}

20. Restaurants: {restaurants*, restaurant, ~~stores~~, ~~retail stores~~, ~~dealerships~~, ~~guest~~, ~~retail locations~~, ~~customer traffic~~, ~~food products~~, ~~brands~~, dining, ~~locations~~, ~~openings~~, ~~concepts~~, ~~schools~~, ~~opened~~}

21. Retail: {retail*, ~~wholesale~~, outlet, **retail sales**, ~~foodservice~~, retailers, specialty stores, convenience stores, automotive, department stores, retail business, retailer, furniture, beauty, ~~home~~, retail outlets, retail operations, other retailers, ~~wholesale customers~~, new vehicle, shopping center, ~~residential customers~~, ~~food service~~, branded, club, casual, establishments, ~~oriented~~, cosmetics, building products, ~~commercial customers~~, upscale, retail space, recreational, ~~business services~~}

22. Satellite Communication: {satellite*, satellites, cable, band, broadband, frequencies, cable television, signals, gateway, carriage, wireless broadband, wireline, gps, microwave, data communications, programming, station, spectrum, broadcasters, **fcc**, transmitter, voip}

23. Traditional Media: {newspapers*, newspaper, television, circulation, movie, outlets, publications, radio, other media, print, advertising revenues, news, publishing, tv, broadcast, entertainment, pages, los angeles, stations, ~~outdoor~~, ~~clubs~~, hd, ~~households~~}

24. Travel: {travel*, air travel, business travel, travelers, ~~leisure~~, tourism, airline, ~~discretionary spending~~, vacation, airline industry, destinations, ~~**consumer spending**~~, ~~attendance~~, ~~disposable income~~, traveling, traffic, ~~recreation~~}

25. Workforce: {workforces*, labor force}

Exposures associated with positive returns:

1. Banking: {bank*, banks, bank subsidiary, state bank, savings bank, financial institution, bank subsidiaries, national bank, bank holding company, institution,

subsidiary bank, financial institutions, the corporation, **ots**, institutions, depository institution, national banks, bank holding companies, savings banks, banking, prudential, fhlb, banking institutions, savings institutions, community banks, financials, financial companies, depository, **federal home loan bank**, extensions of credit, bank regulators, chartered, wells fargo bank, federal bank, wells fargo, bhc act, bhca, corporations, bank of america, holding companies}

2. Containers: {vessels*, vessel, cargo, rigs, tank, fleets, drilling rigs, fleet, containers, trailers, ~~other equipment~~, engines, tractors}

3. Deposits: {**fdic***, fdics, **deposit insurance**, **occ**, insured institutions, frb, dif, insured depository institutions, special assessment, restoration plan, **comptroller of the currency**, assessment rate, assessment rates, reserve ratio, insurance assessments, federal banking regulators, federal banking agencies, loss sharing, loss share, federal banking agency}

4. Drug Trials: {preclinical*, nonclinical, preclinical studies, preclinical testing, preclinical development, clinical testing, clinical studies, clinical, clinical development programs, clinical trials, trials, toxicology, validation, clinical development, clinical data, development programs, confirmatory, trial results, clinical research, drug development, research and development, research programs, vivo, research, clinical trial, stage clinical trials, investigator, clinical study, drug candidates, clinical trial results, vitro, efficacy, product candidates, progress, commercialization activities, commercial use, collaborative, drug candidate, submission, antibody, compounds, inconclusive, investigational}

5. Ecommerce: {ecommerce*, e commerce, online, electronic commerce, ~~customer care~~, direct marketing, payment processing, amazon, ~~portals~~, ~~email~~, network, ~~catalogs~~, pc, ~~salesforce~~, ~~support systems~~, ~~offline~~, pcs, ~~yahoo~~, ~~portal~~, website, online services, ~~chat~~, ~~communications~~}

6. Electronic Components: {optics*, optical, sensor, ray, filter, graphics, high per-

formance, coating, electronic components, electronics, sensors, magnetic, chips, substrates, laser, micro, memory, analog, photovoltaic, fiber, coatings, thin, composites, logic, flash, chip, polymer, handheld, fibers, serial, surfaces, ir, lighting, industrial applications, boxes, glass, ~~hewlett packard~~, portable, ~~samsung~~, cables, electrical, transformers, appliances, audio, printers, intel, tech, ~~dell~~, assemblies, biomedical, appliance, data storage, ~~cisco~~, drives, valve, valves, peripheral, consumables, ~~sole supplier~~, ~~roof~~, stack, industrial, hvac, ~~powered~~, matrix, ~~plc~~, power systems, wired, modular, phones, ~~ltd~~, ~~disposable~~, ~~universal~~, libraries, chamber, embedded, catalyst, ~~microsoft corporation~~, reagents, ~~labs~~, batteries, ~~corp~~, plumbing, furnaces, ~~big~~, bio, ~~color~~, ~~biology~~, ~~strip~~, radiation, ~~sony~~, ~~diagnostics~~, finishing, graphic}

7. Foodstuffs: {**wheat***, grains, **sugar**, fruit, milk, grain, coffee, dairy, protein, proteins, sodium, powder, wine, packaging materials, crops, foods, fresh, agricultural products, synthetic, ~~intermediates~~, additives, enzymes, salt, ingredients, specialty, ~~trees~~, additive, organic, ingredient}

8. Foreign Countries: {china*, india, taiwan, chinese, south africa, asia, russia, beijing, shanghai, hong kong, asia pacific region, united arab emirates, countries, the philippines, korea, chinas, mexico, western europe, egypt, switzerland, overseas, latin america, unitedstates, united kingdom, europe, belgium, asian, germany, singapore, france, ukraine, indonesia, norway, finland, asia pacific, japan, certain countries, iceland, japanese, sweden, operations in mexico, operations in china, north america, peru, korean, australia, dubai, world, european, thailand, european union, industrialized, other countries, russian, england, many countries, worldwide, foreign countries, **central bank**, globally, german, chinese government}

9. Health Insurance: {**medicare***, **medicaid**, cms, payers, **prescription drug**, partd, health plans, physician, payors, reimbursement, **health insurance**, health care, **healthcare**, third party payers, hospital, health plan, payment system, hhs,

payer, clinical laboratory, third party payors, reimbursement levels, department of health and human services, payor, subsidy, prescription drugs, ppaca, mma, care organizations, coding, federal government, patients, private insurers, care programs, reimbursement policies}

10. Investment Management: {investment funds*, private equity funds, hedge funds, private equity fund, investment managers, private equity, limited partnerships, separate accounts, pooled, advisers, investment management, other investment, clo, investment advisers, asset managers}

11. Manufacturing: {manufacturing*, manufacture, product manufacturing, manufacturing process, manufacturing operations, manufacturing processes, manufacturing activities, production processes, manufacturing capabilities, commercial manufacturing, manufacturing facilities, production process, third party manufacturing, manufacturing equipment, assembly, wafer fabrication, contract manufacturers, third party manufacturers, ~~packing~~, contract manufacturing, product development, manufacturing capacity, commercial supply, manufacture of products, technical, new manufacturing, manufacturing facility, product components, production facilities, process technology, manufacturing services, ~~testing~~, commercial scale, contract manufacturer, volume production, finished products, manufacturers, ~~product design~~, ~~formulation~~, ~~materials~~}

12. Metal Products: {**steel***, **aluminum**, **metal**, **copper**, titanium, metals, stainless, pulp, plastics, resin, scrap, rubber, iron, rolled, raw materials, mill, mills, fabricated, raw material, diamond, hot}

13. Power Sources: {**coal***, electricity, ash, coke, steam, sand, power plants, power plant, electric power, energy sources, electric generating, water, tons}

14. Raw Metals and Minerals: {tantalum*, **tin**, tungsten, conflict minerals, democratic republic of congo, minerals, ~~adjoining countries~~, **zinc**, precious metals, such minerals, oxide, **platinum**, ~~requirements for companies~~, ~~sheet~~}

15. Semiconductor: {semiconductor*, semiconductors, silicon, semiconductor manufacturing, ic, semiconductor industry, semiconductor products, network equipment, consumer electronics, oems, ~~life sciences~~, technology industry, ~~customers products~~, ~~automotive industry~~, ~~life science~~, wafers, original equipment manufacturers, ~~industries~~, capital equipment, technology companies}

16. Software Design: {software*, software products, software applications, hardware, software systems, operating system, third party software, proprietary software, interfaces, interface, it infrastructure, architecture, other technology, computer hardware, operating systems, computer, software vendors, third party technology, hardware products, servers, new software, software development, proprietary technology, digital content, ~~designs~~, it systems, algorithms, ~~custom~~, ~~microsoft~~, data management, customization, analytic, ~~design~~, open source, malware, information systems, technology infrastructure, firewalls, open source software, ~~content~~, such technologies, bugs, communications systems, integrations, open source code, computers, compatibility, information management, proprietary, algorithm, source code, ~~laptops~~, technology systems, internal systems, customized, provisioning, computer systems, encryption, optimized, ~~designers~~, business processes, ~~ibm~~, proprietary technologies, ~~downloaded~~, undetected errors}

17. Software Services: {solutions*, solution, software solutions, technology solutions, platform, technology platform, communications services, service offerings, platforms, intelligent, analytics, tools, technologies, product offerings, edge, technology platforms, capabilities, modules, architectures, business solutions, functionality, devices, crm, innovative products, connectivity, new solutions, suite of products, automation, ecosystem, network services, new technologies, new services, ~~innovative~~, module, ~~features~~, management products, enterprise, ~~unified~~, functionalities, product line, next generation, scalability, professional services, applications, ~~touch~~, agile, new features, management system, new technology, testing services, service delivery, ~~other products~~, electronic devices, ~~new products~~,

wireless carriers, business model, enabled, seamless, ~~clients~~, enterprise customers, technical services, support services, new applications, new business models, integrated, lte, range of services, health information technology, diagnostic tests, ~~product lines~~, enhanced products, additional services, technical support services}

18. Transportation: {freight*, trucking, shipping, delivery services, ocean, carriers, shipping costs, other transportation, shipments, railroads, haul, fuel costs, railroad, inbound, transportation industry, ports, fuel surcharges, carrier, container, port, transit}

19. Video Games: {games*, game, titles, players, app, consoles, ~~movies~~, android, windows, player, mobile devices, streaming, facebook, studios, smartphones, music, handsets, smartphone, handset, console, subscribers, ~~download~~, ~~versions~~, ~~videos~~, mobile phones}

20. Web-Based Services: {cloud*, saas, cloud computing, web, hosted, server, internet, premise, ~~ip~~, ~~desktop~~, virtual, data center, networking, messaging, browser, mobility, wireless networks, hosting, subscription, network security, wireless, telephony, data centers, centric, bandwidth}

C.2 Super Tuesday

Exposures associated with negative returns:

1. Aircraft: {aircraft*, commercial aircraft, ~~railcars~~, ~~engine~~, boeing, ~~railcar~~, ~~spare parts~~, ~~equipment~~, ~~machines~~}

2. Asset Management: {private equity funds*, hedge funds, private equity, private equity fund, investment managers, mutual funds, limited partnerships, proprietary trading, pension funds, asset managers, certain investment, operating companies, clo, fixed income, institutional investors, buyout, institutional clients, alternative investment, investment banks, ventures, mutual fund, asset management}

3. Banking: {bank*, banks, bank subsidiary, state bank, financial institution, bank subsidiaries, banking subsidiaries, national bank, bank holding company, ~~institution~~, subsidiary bank, commercial bank, financial institutions, ~~the corporation~~, ~~trust company~~, ~~institutions~~, depository institution, national banks, bank holding companies, banking, ~~prudential~~, fhlb, banking institutions, ~~subsidiary~~, banking operations, financial services businesses, capital adequacy, financial group, community banks, ~~financials~~, ~~thrift~~, financial companies, supervisory, depository, lending, extensions of credit, bank regulators, ~~chartered~~, loans, regulator, federal bank, ~~wells fargo~~, bhc act, bhca, ~~corporations~~, ~~nonbank~~, brokered deposits, ~~holding companies~~, ~~s subsidiaries~~, ~~entity~~, ~~summit~~}

4. Energy Infrastructure: {pipelines*, pipeline systems, pipeline, gathering systems, pipeline system, processing plants, storage tanks, processing facilities, terminals, ~~gathering~~, refineries, ~~interstate~~, ~~transportation facilities~~, **gas pipeline**, terminal, ~~transportation systems~~, downstream, ~~intrastate~~, gas processing, common carrier, ~~rail~~, ~~transportation~~, ~~plants~~, gas gathering, ~~waterways~~, fractionation, ~~shippers~~, refinery, ~~production facilities~~, ~~leaks~~, transmission system, ~~transportation services~~, midstream, ~~unloading~~, ~~ruptures~~}

5. Financial Contracting: {derivatives*, derivative instruments, swaps, derivative, derivative transactions, swap, derivative contracts, financial instruments, hedges, futures contracts, derivative financial instruments, futures, foreign exchange contracts, commodity, hedging, hedging instruments, credit default, forward contracts, hedging activities, hedge accounting, hedge, otc, market risk, hedging arrangements, clearing, aoci, notional, **nymex**, hedged, trading activities, fair value measurements, cash collateral, **cftc**, counterparties}

6. Financial Regulation: {**fdic***, fdics, **deposit insurance**, **occ**, frb, dif, insured depository institution, fdia, insured depository institutions, insured deposits, restoration plan, **federal reserve**, assessment rate, assessment rates, reserve ratio, insurance assessments, federal banking regulators, federal banking agencies,

assessment base, loss sharing, **cfpb**, loss share, fsa}

7. Foodstuffs: {**wheat***, **sugar**, oils, **corn**, grain, proteins, fish, ~~powder~~, wine, ~~fabrics~~, ~~feedstocks~~, ~~blends~~, ~~sweet~~, ~~fibers~~, ~~synthetic~~, ~~intermediates~~, ~~precious metals~~, additives, ~~trees~~, ~~**tin**~~, ~~chips~~, additive, ~~apparel~~, organic, ~~produces~~}

8. Fracking: {hydraulic fracturing*, fracturing, hydraulic fracturing activities, hydraulic fracturing process, sdwa, water act, fracturing process, federal safe drinking, fluids, ~~stimulation~~, hydraulic, ~~lands~~, blm, ~~production activities~~, hydraulic fracturing practices, groundwater}

9. Gambling: {gaming*, casino, slot, las vegas, horse, native}

10. Generic Investment: {investment funds*, investment vehicles, separate accounts, asset classes, pooled, advisers, investment strategies, investment management, other investment, investment advisers}

11. Hotels: {hotels*, hotel properties, hotel, resorts, ~~communities~~, rooms, ~~stores~~, ~~franchisors~~, ~~management agreements~~, suites, ~~franchise~~}

12. Industrial Metals: {**steel***, **aluminum**, nickel, titanium, ~~paper~~, ~~plastics~~, scrap, alloys, ~~petrochemicals~~, concrete, iron, rolled, mill, composites, ~~fertilizer~~, ~~petrochemical~~, pipe, coatings, ~~raw material costs~~, diamond, silicon, ~~hot~~, **platinum**, ~~global demand~~, sand}

13. Motor Vehicles: {vehicles*, vehicle, cars, trucks, engines, new vehicles, car, ~~containers~~, motor vehicles, ~~batteries~~, ~~mounted~~, ~~battery~~, new vehicle, ~~appliances~~, motors, ~~furnaces~~, ~~heavy~~, ~~residual values~~, automotive, motor vehicle}

14. Oil and Gas: {**oil***, ngls, ngl, liquids, **natural gas**, **petroleum**, henry hub, liquefied, hydrocarbon, mcf, hydrocarbons, extraction, shales}

15. Payment Systems: {mastercard*, visa, card, merchants, cards, merchant, debit, cardholders, ach, payment card, payment cards, atm, card transactions, card-

holder, interchange, credit card, pci, processors, payment processing, interchange fees}

16. Pollution: {emissions*, ghg emissions, emission, greenhouse gas, ghgs, ghg, emissions of ghgs, air emissions, carbon dioxide, nox, carbon emissions, air pollutants, methane, ghg emission, emitted, emissions of greenhouse, emission standards, hazardous air pollutants, carbon, dioxide, nitrogen, fuel economy, emit, flaring, sulfur, **clean air act**, **cap and trade**, mact, caa, **epa**, stationary sources, epas, pollutant, nsps, psd, energy consumption, discharges, discharge of pollutants, tons per year, pollution}

17. Power Generation: {electricity*, electric power, power, energy, electricity generation, fuels, electrical power, propane, feedstock, spot market, energy sources, solar energy, renewable energy, refined products, hydro, coke, power generation, generation, **ethanol**, utility, hydroelectric, alternative energy sources, commodities, ~~output~~, generation facilities, renewables, forms of energy, gasoline, lng, renewable, solar panels, wholesale, heat, grid, alternative fuel}

18. Shipping: {vessels*, vessel, barges, cargo, rigs, tank, drilling rigs, rig, tanker, tanks, dock, barge, ~~other equipment~~, ports, crews, loading}

19. Tariffs: {**tariff***, tariffs, ~~**ferc**~~, ~~fercs~~, indexing, shipper, mechanism}

20. Traditional Media: {broadcast*, television, broadcasting, radio, broadcasters, programming, newspaper, stations, movie, **fcc**, fccs, station, newspapers, other media, studios, signals, audio, magazines, digital}

21. Transportation: {fleet*, fleets, truck, horsepower, container, ~~customer base~~, crew, van, miles, trains}

22. Travel: {travel*, air travel, travelers, leisure, tourism, vacations, vacation, destinations, ~~disposable income~~, ~~fears~~, ~~economic activity~~}

23. Unconventional Drilling: {deepwater*, gulf of mexico, shallow, offshore, marcellus shale, permian basin, ~~horizontal~~, bakken, drilling, wash, sands, basin, shale, ~~onshore~~, ~~depths~~, drilling rig, ~~seismic~~, ~~exploration~~, ~~exploratory~~, drilling activity, gas wells, gulf coast, ~~production operations~~, ~~frontier~~, ~~basins~~, ~~feet~~, ~~outer~~, unconventional, ~~wells~~, drilling operations, coastal, ~~north~~, ~~directional~~, deepwater horizon, ~~ocean~~, ~~mississippi river~~, ~~deep~~, ~~marine~~, ~~flowing~~, ~~drilled~~, ~~northern~~, ~~gulf~~, formations}

Exposures associated with positive returns:

1. Construction: {**homebuilding***, **commercial construction**, **residential construction**, land development, housing}

2. Defense Technology: {radar*, sensor, tactical, sensors, weapons, ~~lightweight~~, ~~electro~~, sensing, ~~adaptive~~, handheld, monitoring systems, command, instrumentation, ~~computerized~~, ~~console~~, ~~filtration~~, detection, signaling, ~~visual~~, ray, intelligence, ~~filter~~, ~~pads~~, ~~suppression~~, ~~id~~, ~~powered~~, ~~airborne~~, ~~peripheral~~, ~~imaging~~, ~~turbine~~, ~~interface~~, ~~conditioning~~, ~~hd~~, ~~micro~~, ~~ground~~, ~~optic~~, ~~vision~~, ~~backbone~~, ~~wall~~, ~~radiological~~, ~~cables~~, satellite, tracking, ~~dimensions~~, ~~air~~, ~~advanced~~, ~~workflow~~, ~~polymer~~, ~~architecture~~, ~~flight~~, intrusion, ~~infusion~~, ~~measurement~~, ~~controller~~, ~~radiation~~}

3. Drugs: {drugs*, drug, drug products, pharmaceutical products, drug candidates, new drugs, therapies, prescription drugs, vaccines, ~~approved products~~, ~~such products~~, pharmaceuticals, medical devices, treatments, ~~potential products~~, biosimilars, drug candidate, diagnostic products, biologics, medical device products, ~~compounds~~, ~~devices~~, pharmaceutical, ~~future products~~, new drug, inhibitors, medical products, therapy, antibodies, ~~marketed~~, inhaled, regimens, topical, intravenous, ~~device~~, ~~branded products~~, indications, ~~certain products~~, **fda**, controlled substances, ~~other indications~~, molecules, inhibitor, ~~ingredients~~, ~~label~~, other product candidate, prescription}

4. Ecommerce: {ecommerce*, ~~customer care~~, website}

5. Education: {students*, student, educational programs, ~~subscribers~~, ~~patients~~, ~~homebuyers~~, college, ~~consumers~~, student loans, courses, ~~individuals~~, ~~adults~~, applicants, colleges, ~~members~~, credentials, ~~users~~, ~~women~~}

6. Electronic Communication: {mobile*, mobile phone, apps, wireless, android, messaging, data communications, video, mobile applications, personal computers, ~~enabled~~, platforms, data services, handset, wireless communications, internet services, facebook, phones, ~~smart~~, voice, wireless networks, wireless carriers, communications services, download, broadband, handsets, apple, ~~portal~~, tablets, network, communications, ios, ~~centric~~, ~~networks~~, ~~entertainment~~, voip, lte, telecom, ~~pos~~, wireless services, ~~easy~~, internet access, operating systems, ~~monetization~~, ~~interoperable~~}

7. Foreign: {mexican*, swedish, peso, mexico, ~~railway~~, puerto rico, peru, franc, canadian dollar, operations in mexico}

8. Franchising: {franchisees*, franchisee, franchised, landlords, tenants, franchising}

9. Government Contracting: {government contracts*, subcontracts, government contracting, government contractor, fixed price contracts, government customers, ~~other contracts~~, government contractors, procurement, procurements, ~~contracting~~, government agencies}

10. Government Healthcare: {**medicare***, **medicaid**, cms, reimbursement rates, payment rates, ~~payers~~, inpatient, outpatient, part d, beneficiaries, **prescription drug**, partd, reimbursement, third party payers, aca, government programs, hhs, payer, care plans, third party payors, reimbursement levels, payor, ppaca, **affordable care act**, formularies, care organizations, independent payment advisory board, ~~coding~~, reductions in reimbursement, federal government, private insurers, care programs, reimbursement policies}

11. Health Insurance: {**health insurance***, health plans, health care, health benefits, health plan, health insurers, **healthcare**, ~~private insurance~~, ~~employers~~, ~~other insurance~~, ~~**workers compensation**~~, ~~employer~~, medical care, long term care}

12. Medical Infrastructure: {hospitals*, hospital, clinics, physicians, physician, clinicians, medical services, clinic, pharmacies, surgeons, nursing, ~~providers~~, care providers, ~~universities~~, relationships with physicians, ~~settings~~, ~~care~~, ~~admissions~~, nurses, health services, pharmacy, medical device manufacturers, medical, transplant, acute care}

13. Metals: {**gold***, **silver**, **copper**, metals, **metal**, ~~concentrates~~, ~~recycled~~, ~~pound~~, ~~lumber~~}

14. Military: {navy*, army, **department of defense**, dod, ~~installations~~, defense, military, prime contractor, prime contractors, ~~awarded~~, aerospace}

15. Mining: {mining*, mine, mining operations, mineral, mines, reclamation, **coal**, ore, underground, mined, land use}

16. REIT: {reit*, ric, reits, reit status, reit qualification, taxable reit subsidiary, taxable reit subsidiaries, bdc, irc, investment trust, income test, reit distribution, income tests, taxable years, qualify, asset tests, rics, hedge accounting treatment, gross income, distribution requirement, taxable year}

17. Real Estate: {properties*, property, such properties, certain properties, such property, real property, real properties, other properties, land, land parcels, office properties, commercial property, **real estate**, additional properties, undeveloped land, homes, lease, leases, apartment, property acquisitions, acres, lots}

18. Reinsurance: {reinsurance*, reinsurers, reinsurance coverage, ceded, reinsurance contracts, reinsured, property insurance, reinsurer, insurance subsidiary, commercial insurance, catastrophe, insurers, insurance policies, insurance subsidiaries,

coverages, insurer, casualty insurance, insurance coverage, such insurance, insureds, cost of reinsurance, casualty, statutory surplus, insurance company, insurance operations}

19. Rental Market: {space*, office space, retail space, vacant space, rentable, square feet, condominiums, let, buildings, leased, office buildings, vacancy rates, footage, vacant}

20. Restaurants: {restaurants*, restaurant, ~~shopping centers~~, ~~dealerships~~, ~~customer traffic~~, ~~foods~~, ~~food products~~, ~~clubs~~, ~~club~~, ~~convenience stores~~}

21. Subsidies: {subsidy*, subsidies, ~~veterans~~, grants, rebates, ~~eligibility~~}

22. Utilities: {utility operations*, utilities, electric utility, ~~electric~~, ~~es~~, distribution operations, service territories, electric transmission}

23. Video Games: {games*, game, titles, players, app, player, ~~new product offerings~~}

24. Waste: {landfills*, landfill, solid waste, ~~generating facilities~~, ~~beds~~, hazardous waste, wastewater, ~~ash~~, ~~water~~, ~~electric generating~~}