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

Steven J. Davis, Stephen Hansen, and Cristhian Seminario-Amez

Economics Working Paper 20118

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010

Revised February 3, 2021

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, traditional retail, aircraft production and energy supply—directly and via downstream demand linkages—and raises them for firrms with high exposures to healthcare policy, e-commerce, web services, drug trials and materials that feed into supply chains for semiconductors, cloud computing and telecommunications. Monetary and fisscal policy responses to the pandemic strongly impact firrm-level returns as well, but differently than 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 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.

Firm-level Risk Exposures and Stock Returns in the Wake of COVID-19
Steven J. Davis, Stephen Hansen, and Cristhian Seminario-Amez
Economics Working Paper 20118
Revised February 3, 2021
JEL Code: E44, G12, G14, G18
Keywords: Firm-level stock returns, COVID-19, text analysis, supervised machine learning, risk factors

Steven J. Davis
University of Chicago Booth
School of Business and
Senior Fellow, Hoover Institution

Stephen Hansen,
Imperial College Business School
and CEPR

Cristhian Seminario-Amez
Kenneth C. Griffin Department of
Economics,
University of Chicago

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, traditional retail, aircraft production and energy supply—directly and via downstream demand linkages—and raises them for firrms with high exposures to healthcare policy, e-commerce, web services, drug trials and materials that feed into supply chains for semiconductors, cloud computing and telecommunications. Monetary and fisscal policy responses to the pandemic strongly impact firrm-level returns as well, but differently than 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 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.

Acknowledgments:

We thank Alejandro Lopez Lira for helpful comments on an earlier draft, and seminar participants at BI Business School, Cornell, IESE, Swiss National Bank, and the University of East Anglia for their comments. 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.[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. 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 days, 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, for example, 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 days, because the news event that drove market reactions on those dates is usually apparent. In this regard, we rely on the classifications in Baker et al.

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

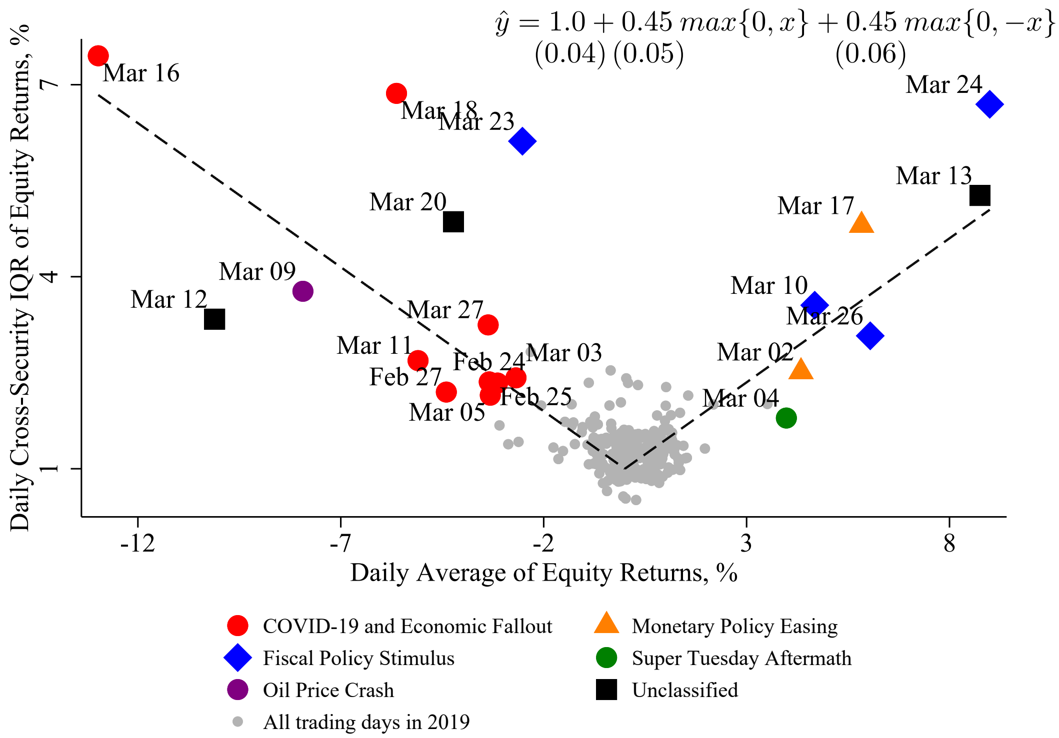


Figure 1: Value-Weighted Mean and Cross-Sectional IQR of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

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.) of the 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.

(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 days, 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, traditional retail and energy supply – directly and via downstream demand linkages – and raises them for firms with high exposures to healthcare policy, e-commerce, web services, drug trials and materials that feed into supply chains for semiconductors, cloud computing and telecommunications. 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 days 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 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 sources as well.

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 days under the two approaches. At the same time, MNIR achieves a uniformly higher adjusted R^2. For example, our MNIR model explains one-half of firm-level abnormal return variation on pandemic fallout days, as compared to one-third under the dictionary approach. As we show, the superior fit arises entirely because MNIR draws on a much larger feature space.

However, it is hard to obtain clear insights from raw MNIR results. To address this challenge, we proceed in steps. First, we identify seed terms that MNIR weights highly in explaining firm-level returns. In a second step, we use the seeds to automatically generate a set of related terms based on similarity of linguistic context and MNIR weights. Third, we prune the automatically generated lists to obtain term sets that define our new risk exposure categories. This process for developing exposure categories and associated term sets draws on both our high-dimensional MNIR model and our domain expertise. In this sense, it incorporates elements of both ML and expert-curated dictionary approaches. Armed with our new term sets, we quantify firm-level exposures to each category. Finally, we incorporate the firm-level exposure measures in straightforward regression models that yield easily interpretable results.

When we apply this hybrid approach, 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, e-commerce and more. It is also how we uncover the role of 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 for semiconductors, lasers, 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 role 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 show that its impact on input demands 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 associated term sets is useful more broadly. As a further illustration, we use the hybrid approach to characterize returns the day after 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 raised his chances of securing the Democratic nomination. We again apply our hybrid approach to uncover risk factors, build associated term sets, and use them to explain firm-level returns. We find that Super Tuesday drove negative returns for firms with high exposure to hotels, gambling, fracking, and financial management; and positive returns for firms with high exposure to healthcare, health insurance, REITs, property rentals, communications and construction.

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 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 distinct types of news events. Our reliance

[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.

on 10-K filings before the pandemic struck yields an ex ante characterization of risk exposures rather than an ex post one. Our approach also yields a fuller characterization of risk exposures, because firms face legal and financial liabilities 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 web services, the demand for which rises with reliance on remote interactivity. Our evidence about firm-level returns is largely distinct from the evidence in these two studies 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 an understanding of what drives firm-level return reactions to common shocks.

The next section 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 days into categories that align closely with the classification that Baker et al. (2020a) derive from next-day newspaper accounts of large market-level moves. Section 5 develops and implements our hybrid approach, first applying it to pandemic-related jump days and then to Super Tuesday. Section 6 shows that our text-based models of firm-level abnormal returns also have strong predictive content for corporate earnings surprises. Section 7 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 days 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 days in 2020. These figures reinforce the chief message of Figure 1: The jump days 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):

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

where $p_{i,t}$ 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.

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*

[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.

(*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

[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

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

MNIR posits $\mathbf{x}_i \sim \text{MN}(\mathbf{q}_i, N_i)$, where $\mathbf{q}_i$ is a multinomial V-dimensional probability

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.

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 $\mathbf{y}_i = (\text{Abn}_i, \mathbf{c}_i)$ contains firm-i abnormal returns on a given day or collection of days and firm controls $\mathbf{c}_i \in \mathbb{R}^P$. (We suppress time subscripts here.) 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 features, 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, 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 estimates of (2) to a forward regression with Abn_i as the dependent variable, we follow Taddy (2013) and define a *sufficient reduction projection* $z_i = \sum_v x_{i,v} b_{1,v}$ with the property $Abn_i \perp \mathbf{x}_i \mid z_i, N_i, \mathbf{c}_i$. 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 Abn_i in a forward regression, but it says we can model Abn_i 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 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:

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

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 the jump days attributed to fiscal policy.

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 market cap is a highly significant return predictor on jump days. Consistent with 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 and its economic fallout. Firms with high exposures to *intellectual property* and *health-*

Jump Classification ⟶ **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

The dependent variable is the daily firm-level (average) abnormal equity return on February 24, 25, 27 and March 3, 5, 11, 16, 18, 27 in Column (1); March 2, 17 in Column (2); March 10, 23, 24, 26 in Column (3); March 4 in Column (4); and March 9 in Column (5). Each column considers the 36 dictionary-based exposure measures plus log market cap, leverage, and 2-digit NAICS fixed effects as controls. The t statistics reported in parentheses are computed by clustering errors at the NAICS2 level. For presentation purposes, we omit coefficients on dictionary categories that are insignificant at the 0.1 level.

care 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 for firms with high exposures to *inflation* 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, and the likelihood that Donald Trump would win re-election. The revision in assess-

[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).

ments 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 a 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.

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 particular jump days. Moreover, as we have stressed, the dictionary approach taps only a small fraction of the *RF* corpus.

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 $\mathbf{c}_{it} = (\text{Leverage}_i, \log(\text{Mcap}_{it}), \gamma_{s(i)})$. Our forward regression is

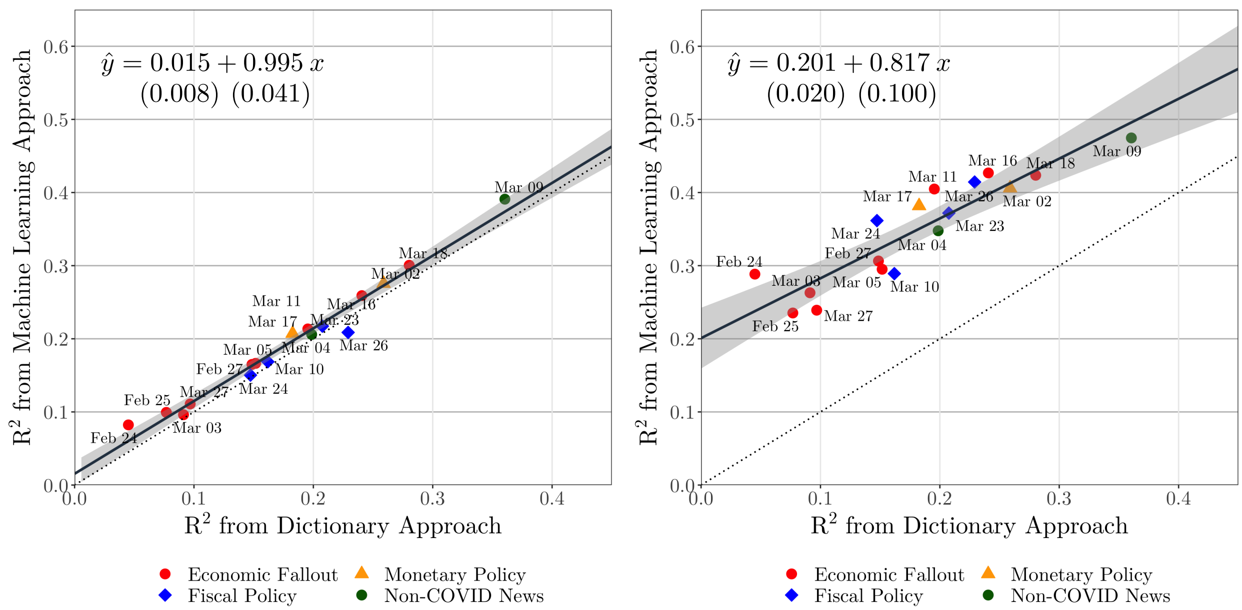
$$Abn_{it} = \alpha_1\, z_{it} + \alpha_2\, N_i + \alpha_3\, Leverage_i + \alpha_4\, Log(Mcap_{it}) + \gamma_{s(i)} + \varepsilon_{it}, \qquad (4)$$

which we estimate by ordinary least squares. N_i is the number of terms in firm i's RF texts, and z_{it} 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 (i) the dictionary approach, (ii) MNIR based on the 244 dictionary terms, and (iii) 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 days. 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.

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 days. The implication is that, at least in our setting, the supervised



(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.

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 days 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 days. 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 36,635 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 day.

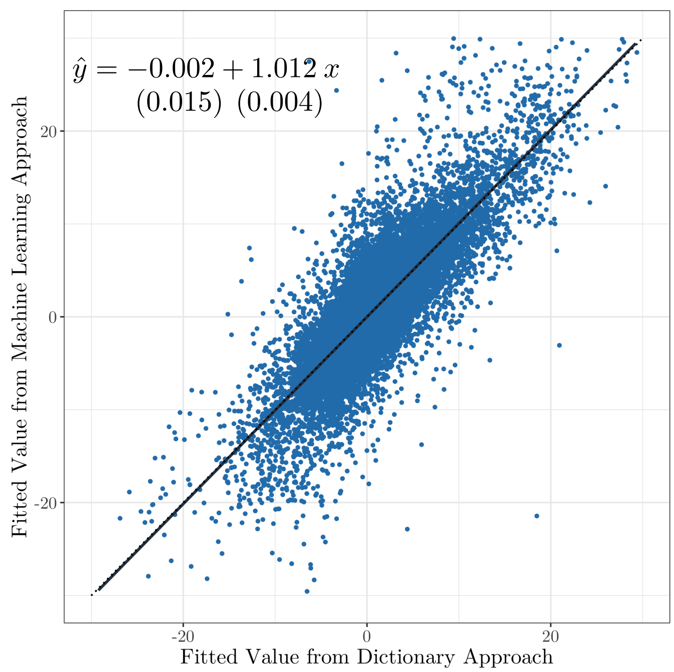


Figure 3: Fitted Values from Dictionary and Machine Learning Approaches
This figure plots fitted values from regressions (3) and (4), pooled across all 17 dates. It also presents a fitted regression line that relates the two (solid black line). Confidence bands are omitted from the graph due to their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

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 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.

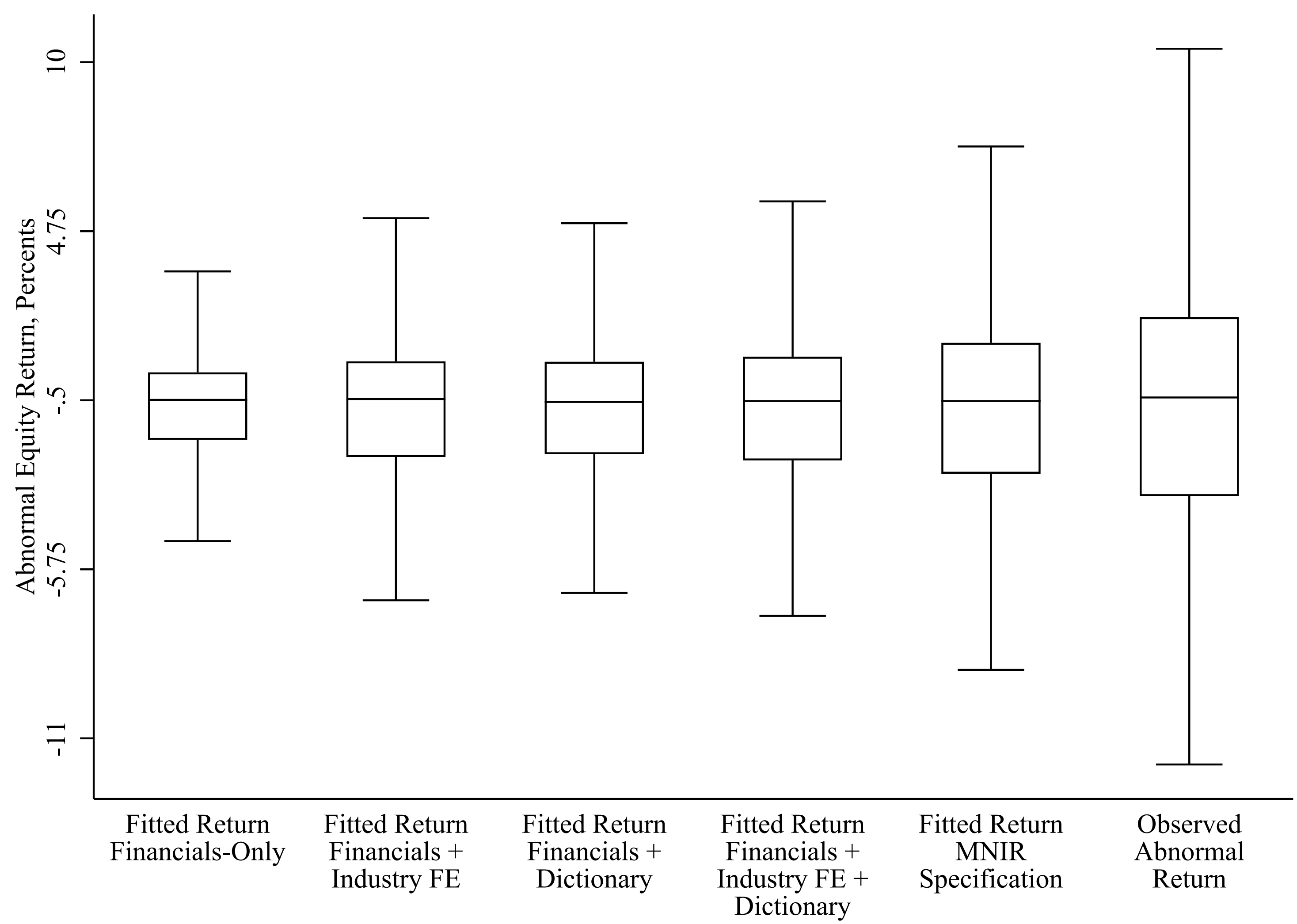
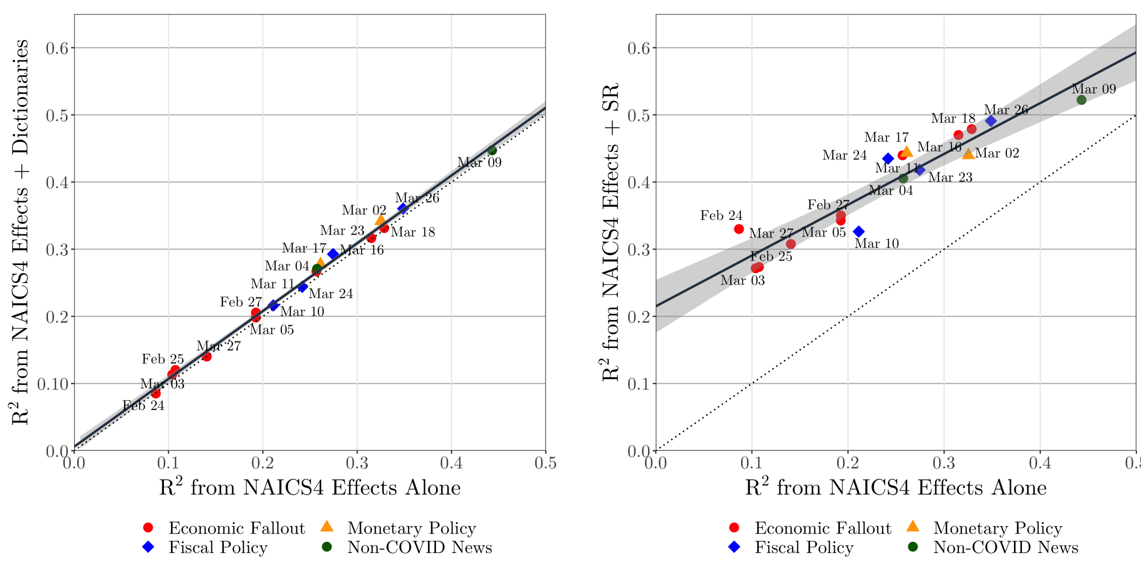


Figure 4: Range of Firm-Level (Fitted) Returns

This figure considers all 2,155 securities and 17 jump days 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, respectively, 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 day.[8] 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

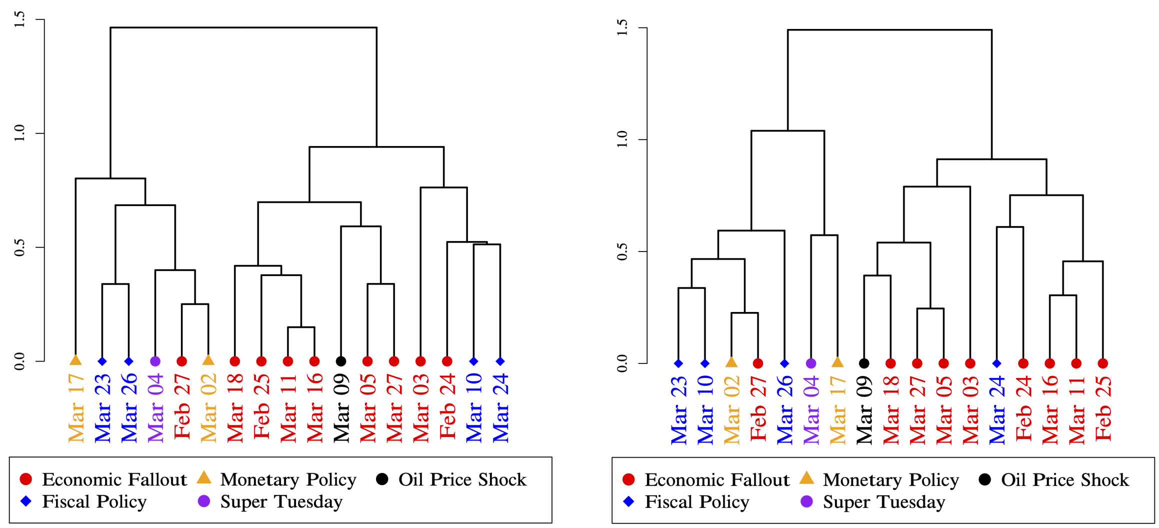
Recall that Baker et al. (2020a) use human readings of next-day newspaper accounts to classify *market-level* jumps as to reason (pandemic fallout, fiscal policy, etc.). We now apply an automated approach to classify jump days based on the *structure* of predicted firm-level abnormal returns. For this exercise, we measure returns in (2), (3), and (4) as $Abn_{it} \times (\mathbf{5.5}/AvgER_t)$, where 5.5 is the average market-level move in our sample of 17 jump days. 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.[9]

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

[8]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.

[9]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 $Abn_{it} \times (\mathbf{5.5}/AvgER_t)$ 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 grouping of jump days in Figure 1 and Section 3. From there, we use MNIR models to develop new 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. In contrast, 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.[10] 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, $x_{i,v}\log(\frac{2133}{df_v})$, 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.[11] 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 risk exposures: A systematic approach

To construct risk exposures, we start with "seeds" drawn from (a) terms with large MNIR coefficients, $|\hat{b}_{1,v}|$; and (b) terms with large tf-idf weighted MNIR coefficients, $|\hat{b}_{1,v}|x_v\log(\frac{2133}{df_v})$, where x_v is the count of term v in the RF corpus. 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. Specifically, we deploy the popular word2vec model of word embeddings to obtain a vector that, for each term $v \in \{1,\ldots,V\}$, characterizes the (average) semantic context of words in close proximity to v in our RF corpus.[12] We then use these vectors to group terms

[10] Figure B.8 shows that predicted returns for these firms align more closely with 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.

[11] The $\log(\frac{2133}{df_v})$ expression down weights generic terms that are less useful in distinguishing among firms. 2133 is the number of firms for which we have 10-K filings, so a term that appears in the 10-Ks of every firm gets zero weight.

[12] See Mikolov et al. (2013). In our application of word2vec, each word embedding is characterized

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 influenza clinical trials candidates collaborators	475·2 297·5 136·9 99·4 96·1
NETFLIX INC ↗	World's leading internet television network with streaming memberships in over 190 countries.	dvd streaming subscribers titles studios	943·2 445·6 243·5 216·6 81·9
CEVA INC ↗	Leading licensor of wireless connectivity and smart sensing platforms.	ip irish royalty revenues royalty rates semiconductor	185·7 135·3 112·0 95·5 86·8
SURMODICS INC ↗	Leading provider of medical device and in vitro diagnostic technologies to the healthcare industry.	reagents incorporating licensees new applications coating	124·1 84·2 42·7 38·0 30·7
CADIZ INC ↗	Owner and developer of sustainable water and agricultural projects in California.	government appropriations railroad significant revenues value of options such plans	27·4 24·9 22·9 13·6 13·3
DOMINO'S PIZZA INC ↗	Multinational pizza restaurant chain with a large global network of franchise owners.	cheese quick foods pound interruption from earthquakes	56·5 28·4 25·5 20·9 16·8
RENEWABLE ENERGY GROUP INC ↘	North America's largest producer of advanced biofuels.	biodiesel biomass feedstocks gallons soybean	−2897·7 −904·4 −335·0 −282·2 −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 ngl barrels per day pipeline pipelines	−1639·7 −1499·2 −1363·3 −815·5 −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 hotels movie film patrons	−216·8 −215·2 −164·6 −103·6 −72·2
PBF ENERGY INC ↘	Petroleum refiner and supplier of unbranded transportation fuels, heating oil, petrochemical feedstocks, lubricants and other petroleum products.	refinery crude refineries feedstocks refined products	−1122·3 −755·6 −627·8 −394·3 −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 five terms with the largest positive or negative MNIR coefficient in the firm-specific sufficient reduction projection when multiplied by tf-idf scores to down weight generic 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 $|\hat{b}_{1,v}| x_v \log(\frac{2133}{df_v}) > 200$.

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. Applying criterion 1 yields 8,513 seed-generated terms. Further applying criterion 2 leaves us with 1,100 terms in addition to the 45 seeds. When a term meets both criteria for multiple seeds, we assign it to the seed for which its cosine similarity is highest.

Appendix C.1 lists all terms grouped with each seed under this approach. By and large, the term sets reflect coherent exposure categories, 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

by a 200-element vector selected to maximize the probability of observing v in the corpus, conditional on a surrounding window of plus or minus five words. Our choice of 200 elements for the embedding vector and a proximity filter of +/- five words is fairly standard in the NLP literature. In practice, researchers evaluate the quality of embedding models based on downstream performance. As we will show, the exposure categories and term sets derived using our embedding models perform well in explaining and interpreting firm-level abnormal returns.

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

(a) Negative Exposures

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

(b) Positive Exposures

Table 3: MNIR-Generated Exposure Categories for Pandemic Fallout Dates

This table lists the 45 “seeds” that we use to construct MNIR-generated exposure categories and term sets. The last two columns report how many automatically generated terms we retain and drop after judgmental pruning. We combine closely related exposure categories in our downstream analysis, as indicated by those with the same font color.

inputs in the manufacturing supply chain for electronic components and "democratic republic of congo," a major source country for rare earth mineral mining.

Next, we manually prune the automatically generated term sets, deleting highly generic terms and those at high risk of referring to other concepts (as fully described in appendix C.1). Continuing with the tantalum example, we delete "adjoining countries," "requirements for companies," and "sheet." Table 3 reports the number of deleted and retained terms in each category. Lastly, we compute firm-level exposures to each category j and its associated term set, $L(j)$, as $z_i^j = \sum_{v \in L(j)} x_{i,v} |b_{1,v}|$, which captures the part of the sufficient reduction projection that derives from terms in $L(j)$. We combine categories with correlation higher than 0.7 across firms, which consolidates closely related categories such as "Software and Hardware Products" and "Software Services." We drop the "Manufacturing" category, because it generates firm-level exposure measures that are highly correlated with those generated by other, more focused categories. After combining categories and pruning, we have 934 terms in 38 categories.

Table B.3 reports descriptive statistics for the resulting firm-level exposures for each category j. Some categories contain terms that appear in the RF texts of relatively few firms (e.g., "Gambling," 9% of firms), whereas others appear in the RF texts of most firms (e.g., "Web-Based Services," 68%). Cross-firm exposure variation within categories is typically large.

5.3 Applying the new exposure measures

Table 4 reports abnormal return regressions on pandemic fallout days using the MNIR-generated exposure measures as explanatory variables. We fit (3) to the same data as in Column (1) of Table 1 except for using 38 MNIR-generated exposure measures rather than the 36 measures derived from expert-curated dictionaries. As in Table 1, we express explanatory variables in standard deviation units. A first key result is that the MNIR-generated exposures yield a large gain in fit: The adjusted R^2 increases from 0.33 in Column (1) of Table 1 to 0.41 in Column (1) of Table 4. In other words, our

MNIR-generated categories and measures outperform the dictionary approach, even when using about the same number of explanatory variables.

A second key result is that our MNIR-based measures capture much of the overall fit gain achieved by the full-blown MNIR implementation. When using the same firms and pandemic fallout days, the MNIR forward regression yields an adjusted R^2 value of 0.50. To be sure, that is a substantial fit gain relative to Column (1) in Table 4. By itself, however, the forward regression has little interpretation value. Embedded within the sufficient reduction projection that drives the strong fit of the forward regression are more than 18,000 terms that appear in the RF corpus and receive non-zero weight in the inverse regression. In contrast, Column (1) reports a regression based on 38 exposure measures and 934 terms. Moreover, each category involves a collection of related terms that, in most cases, offer ready interpretations.

Table 4 offers a rich account of how bad pandemic news drives the structure of firm-level abnormal returns. Exposures to "Traditional Retail" and "Card Payments" involve negative returns, while exposure to "Ecommerce" involves positive returns – in line with the pandemic-induced shift in consumer spending from traditional retail to online modes. Other results reflect major shifts in consumer spending across types of non-durable goods and services. For example, exposures to "Foodstuffs" and "Restaurants" generate positive and negative returns, respectively, in line with a shift from market to home production of meals. "Video Games" is associated with positive returns, while "Gambling" and "Aircraft + Travel" are associated with negative returns, reflecting a major shift across entertainment and recreation activities.

Downstream demand shocks are also key drivers of return reactions to pandemic news. For example, exposures to "Alternative Energy" and "Energy Infrastructure + Oil and Gas" involve negative returns, while the opposite holds for exposures to the technology supply chain as captured by "Raw Metals and Minerals," "Electronic Components and Devices," and "Web-Based Services." The magnitudes of the estimated coefficients on these exposure measures for intermediate goods and services are comparable to the ones for categories that capture spending on final consumption goods.

Dependent Variable: Abn_{it}	(1) NAICS-2 Fixed Effects		(2) NAICS-2 Fixed Effects		(3) NAICS-4 Fixed Effects	
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)
Financial 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)
Gold and Silver	−0·28	(−16·8)	−0·28	(−22·1)	−0·32	(−11·4)
Healthcare Providers	−0·14	(−6·5)	−0·12	(−7·8)		
Insurance	0·04	(2·1)	0·05	(2·4)		
Mortgages	−0·11	(−3·3)	−0·13	(−5·6)		
REITs	−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)
Traditional 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)
Aircraft + Travel	−0·24	(−2·7)	−0·25	(−2·9)		
Communications + Trad Media	−0·09	(−2·4)	−0·09	(−2·3)	−0·11	(−2·9)
Energy Infr + Oil and Gas	−0·31	(−5·1)	−0·28	(−4·8)	−0·19	(−3·9)
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 and Devices	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 Funds	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)
Semiconductors					−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)
Banking + Deposits	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: Regression Results Using MNIR-Generated Exposures, Pandemic Fallout Days

Each column considers 38 MNIR-generated exposure measures for pandemic fallout dates. 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 on exposures 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.

Financial exposure categories also also important. "Mortgages" generates negative returns, as do other property-related exposures like "REITs" and "Residential Construction." Exposure to "Financial Management" is also associated with negative return reactions. In contrast, exposures to "Banking + Deposits" and "Investment Funds" yield positive return reactions. A plausible explanation is that bad news about the pandemic drove greater precautionary savings and more concern about managing portfolio risks among higher-income households.

We can also produce a model-based explanation for any particular firm's return reactions by combining its exposure measures with the estimation results. Consider Plains All American Pipeline, which had an average daily abnormal return of -6.4% on pandemic fallout days as compared to -1.1% for the average firm. The Column (1) model in Table 4 predicts a daily abnormal return of -5.9% for Plains. The model attributes the company's sharply negative performance on pandemic fallout days to its heavy exposure to "Energy Infrastructure + Oil and Gas," "REITs," and "Alternative Energy" and its light exposure to categories associated with positive returns.

Column (3) in Table 4 includes NAICS4 industry effects in place of NAICS2 effects. For comparison, column (2) reports results using NAICS2 effects on the same sample as in column (3). Some exposure measures become insignificant, but most remain significant and a few previously insignificant ones become significant. The implication is that our MNIR-generated exposure measures are not simply a proxy for industry categories. Instead, they capture and explain important differences in firm-level return reactions within narrowly defined industries.

In sum, bad COVID-19 news generates a wide array of positive and negative return reactions across firms. The richness of the *RF* corpus enables us to uncover dozens of separate effects that play a role in driving the structure of return reactions, including effects that reflect demand shocks induced by social distancing and those that involve indirect exposures via supply-chain linkages.

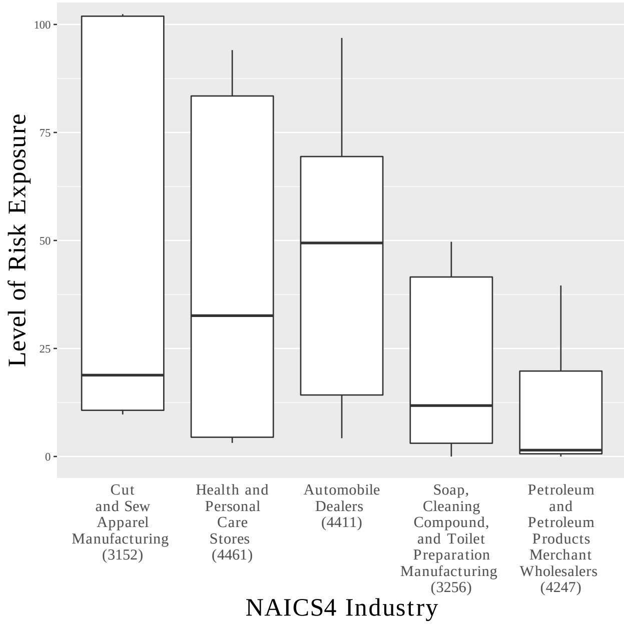
5.4 Risk exposures and industries

The strong explanatory power of our exposure measures in Table 4, even conditional on NAICS4 fixed effects, suggests that firm-level exposures vary considerably within and between industries. To develop this point, we examine the intra-industry distribution of firm-level exposures for four exposure categories, two with statistically significant positive coefficients in Table 4 and two with negative coefficients. We visualize these firm-level distributions as box plots in Figure 7.

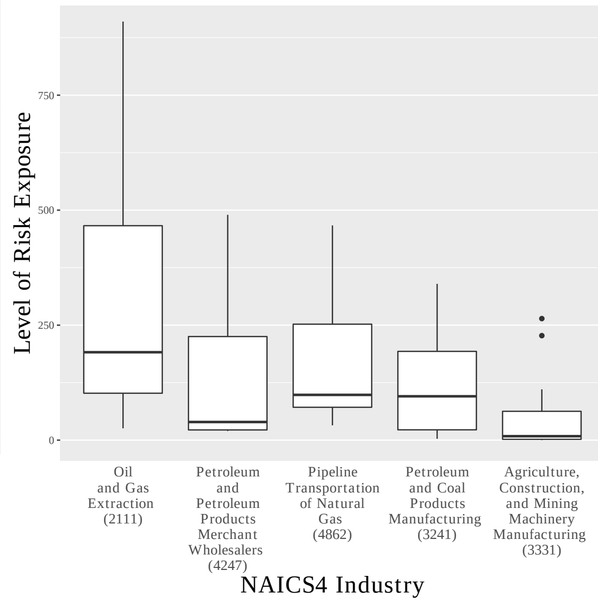
Two NAICS4 industries with high dispersion in exposures to "Traditional Retail" are, in fact, part of the retail sector: Health and Personal Care Stores and Automobile Dealers.[13] The other three industries with large dispersion in exposure to "Traditional Retail" sit upstream from the retail sector: Cut and Sew Apparel Manufacturing; Soap, Cleaning Compound, and Toilet Preparation Manufacturing; and Petroleum and Petroleum Products Merchant Wholesalers. This case illustrates how our exposure categories capture both direct and indirect effects of shocks to a given sector, and how greatly the firm-level exposures differ within industries. Figure 7b shows that exposure to the "Oil and Gas" category varies across and within industries that extract fossil fuels, transport fuels, use fuels as intermediate inputs, and sell fuels.

Four industries with high dispersion in exposures to "Foodstuffs" manufacture or process food and drinks, while the other industry comprises merchant wholesalers who sell these products to downstream grocery and food stores. Finally, a heterogeneous set of industries are exposed to Web-Based Services. These include hardware manufacturers that build physical cloud infrastructure, software publishers that provide code to run that infrastructure, and firms that use the web to provide services to downstream consumers (e.g., cable programming and telecommunications).

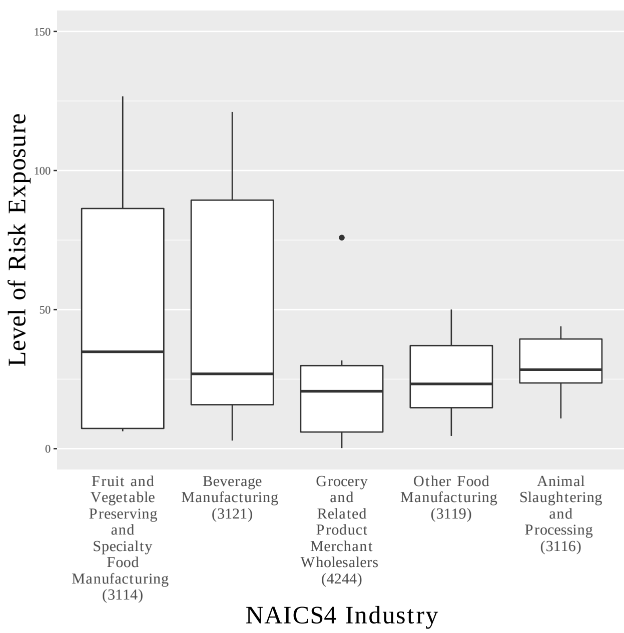
[13]The firm in Health and Personal Care Stores with minimum (maximum) exposure is Rite Aid Corporation (Ulta Salon, Cosmetics & Fragrance Inc.). The minimally and maximally exposed Automobile Dealers are, respectively, Rush Enterprises and Asbury Automotive Group.



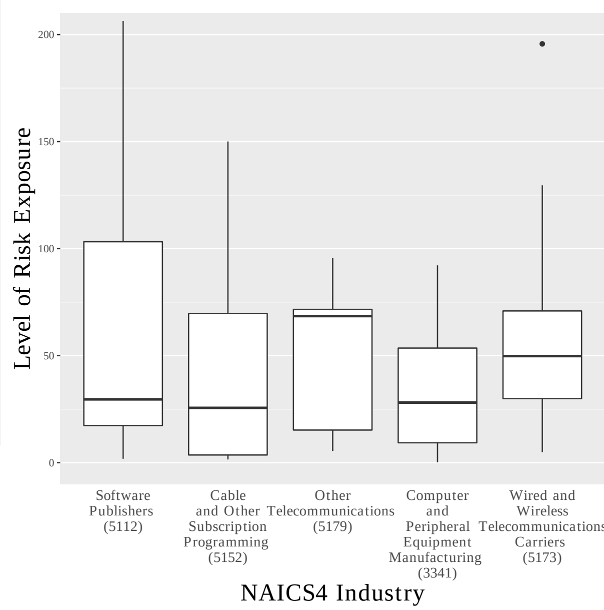
(a) ‘Traditional Retail’ Exposure Category



(b) ‘Oil and Gas’ Exposure Category



(c) ‘Foodstuffs’ Exposure Category



(d) ‘Web-Based Services’ Exposure Category

Figure 7: Intra-Industry Distributions of Firm-Level Risk Exposures

For a selection of four targeted exposures, we compute the risk exposure distribution across firms in absolute values within NAICS4 industries. Each figure shows box plots for the five industries for which the interquartile range of these distributions is highest. 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. Black dots represent outliers.

5.5 An Application to Super Tuesday

To illustrate its flexibility and power, we now apply our approach to explain and interpret firm-level equity returns in reaction to Super Tuesday. This application captures the impact of the attendant shift in expectations about future government policy due to changes in the likelihood of a Biden, Sanders or Trump Presidency, expectations about what policies Biden would pursue to secure the Democratic nomination, and expectations about the Trump campaign's policy agenda in response to the now-likely prospect of facing Biden in the general election contest.

Table B.4 reports the RF terms most associated with positive and negative return reactions to Super Tuesday. Terms related to hotels and gambling ("hotel properties," "casino," "las vegas") and fossil fuels ("natural gas," "oil," "permian basin") are among those highly associated with negative returns, while terms associated with real estate more broadly ("tenants," "reit," "undeveloped land") and healthcare ("health insurance," "medicare," "patients") are highly associated with positive returns. Recall that "reit" and "oil" are influential predictors of negative returns on pandemic fallout days. They remain highly influential in explaining reactions to Super Tuesday but in the opposite direction, illustrating how we capture distinct firm-level return reactions to different types of shocks. These results also illustrate the power of text to disentangle countervailing effects. Generic property terms predict positive return reactions to Super Tuesday, while terms relating to the hotels and gambling segment of the property market predict negative reactions. Perhaps not coincidentally, former President Trump has major business interests that involve hotels and gambling.

Table B.5 reports firms for which MNIR yields the largest fit gains relative to a no-text model. All five firms with the largest negative predicted reactions are in the oil and gas sector. Terms like "crude," "refinery," "offshore," and "drilling" are most responsible for driving these predictions. Two firms with large positive reactions facilitate government-sponsored healthcare (eHealth and Centene Corp), two are involved in mining (Coeur Mining Inc and Gold Resource Corp), and one is an industrial property developer (Griffin Industrial Realty). In each case, the terms associated with large positive returns relate naturally to the company's main business activity.

Next, we select 47 seeds (listed in Table B.6) and, as before, use them to construct term sets that define exposure categories. We obtain 666 automatically generated terms and prune 215 of them, as set forth in appendix C.2. Note that a given seed need not yield the same term set for Super Tuesday as for pandemic fallout days, because the estimated MNIR models differ. For example, the "ecommerce" seed yields 23 terms on pandemic fallout dates but just 3 on Super Tuesday (before manual pruning), because few terms semantically related to e-commerce are associated with negative returns on Super Tuesday. Nevertheless, several seeds yield similar term sets for Super Tuesday and pandemic fallout days, e.g., "travel," "reit," and "gaming."

We also uncover many distinct exposure categories for Super Tuesday. These include "Drilling Activity" and "Fracking," which describe specific fossil fuel extraction activities, in contrast to"Energy Infrastructure" and "Oil and Gas." Other examples include "Defense Technology" (military hardware), "Financial Instruments" (derivatives and hedging), "Government Healthcare" (Medicare, Medicaid, etc.) and "Waste Disposal." Another notable category is "Financial Regulation," a concept also present in the baseline dictionaries and correlated with Super Tuesday return reactions in Table 1. By and large, our MNIR-generated exposure categories for Super Tuesday add detail and nuance to concepts like healthcare, energy, finance, and real estate present in the baseline dictionaries. In contrast, our MNIR-generated exposure categories for pandemic fallout days are often quite distinctive and absent from the baseline dictionaries. Examples include categories for cloud computing, web services, and material inputs to the technology supply chain.

Dependent Variable: Abn_{it}	(1) NAICS-2 Fixed Effects		(2) NAICS-2 Fixed Effects		(3) NAICS-4 Fixed Effects	
Exposures						
Aircraft					−0·08	(−3·1)
Card Payments	−0·04	(−2·4)	−0·04	(−2·1)	−0·09	(−2·7)
Financial Instruments	−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 + Financial Mngmt	−0·19	(−9·5)	−0·16	(−7·6)	−0·18	(−4·1)
Banking + Financial Regul	−0·18	(−7·5)	−0·18	(−5·4)	−0·11	(−3·4)
Drilling Act + Fracking	−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)
Insurance	0·13	(8·7)	0·13	(5·5)		
Metals	0·16	(3·7)	0·20	(3·3)	0·23	(6·1)
Military	0·09	(2·6)				
REITs	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)		
Ecomm + Health Ins + Subsidies	0·20	(4·0)	0·17	(2·5)	0·20	(3·0)
Gov Healthcare + Healthcare Supp	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: Regression Results Using MNIR-Generated Exposures, Super Tuesday Aftermath

Each column considers 40 MNIR-generated exposure measures for Super Tuesday. 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 on exposures 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.

Table 5 presents the abnormal returns regression for Super Tuesday based on the new exposure categories. We again find a sizable fit gain compared to a model that relies on the baseline dictionaries: the adjusted R^2 rises from 0.199 in Column (4) of Table 1 to 0.242 in Column (1) of Table 5, roughly one-third of the gain from using the full-blown MNIR model (0.349). Many exposure categories remain statistically significant after controlling for NAICS4 effects, again showing that the RF texts contain information that helps explain intra-industry variation in abnormal returns.

6 Longer-Term Outcomes

We now investigate whether our text-based explanations for abnormal returns help predict firm-level surprises in corporate earnings in Q3 2020. Unlike abnormal returns, which can reflect revisions to cash flows and discount rates, earnings surprises pertain exclusively to new information about cash flows.

Because earnings forecasts formed in 2020 are potentially contaminated by knowledge of COVID-19 and its effects, we define surprises relative to forecasts made in 2019 Q4.[14] We obtain firm-level analyst forecasts of earnings per share (EPS) from I/B/E/S and express the earnings surprise for a given firm as a percentage of its pre-pandemic market value: 100(Realized Q3 2020 EPS minus average of analysts' forecast of Q3 2020 EPS)/(stock price per share on 15 November 2019).[15] We are able to construct this measure for 1,707 firms in our sample, and 1,507 after removing those in NAICS4 industries with fewer than three firms and extreme outliers in the surprise distribution.[16] The mean earnings surprise is -0.39 cents per dollar of share value, the standard deviation is 1.02, and the IQR is 0.7. As explanatory variables for the earnings surprises, we use the same NAICS4 and firm-level controls as before and the sufficient reduction projections from our MNIR models fit to (average) firm-level returns on the five classes of jump dates.[17] Projections for the oil shock date and the fiscal stimulus dates are

[14] The Wuhan Municipal Health Commission publicly reported an outbreak of pneumonia of unknown cause on 31 December 2019.

[15] Since we do not know the precise date on which analysts submit their forecasts, we use the middle date of Q4 2019 for scaling.

[16] Defined as surprise values greater than the upper quartile value plus five times the IQR and values lower than the lower quartile minus five times the IQR.

[17] Thus, we are *not* fitting MNIR models directly to the firm-level data on earnings surprises. While

	Dependent variable:				
	Earnings Surprise				
	(1)	(2)	(3)	(4)	(5)
SRP (Pandemic Fallout)		0.320**			0.313***
		(0.134)			(0.094)
SRP (Monetary Policy)			0.325***		0.122**
			(0.051)		(0.056)
SRP (Super Tuesday)				0.323***	0.279***
				(0.065)	(0.082)
Leverage	0.0002	0.025	0.008	−0.013	0.016
	(0.069)	(0.066)	(0.068)	(0.068)	(0.064)
Log Market Cap	0.106***	0.085**	0.106***	0.131***	0.107***
	(0.036)	(0.036)	(0.031)	(0.036)	(0.033)
Observations	1,507	1,507	1,507	1,507	1,507
NAICS4 Effects	Y	Y	Y	Y	Y
R^2	0.211	0.229	0.244	0.242	0.269
Adjusted R^2	0.143	0.163	0.179	0.176	0.204
Note:	*p<0.1; **p<0.05; ***p<0.01				

Table 6: Text-Based Models of Abnormal Returns Predict Earnings Surprises
Each column models firm-level 2020 Q3 earnings surprises as a function of NAICS4 effects, log market cap, and leverage. Columns (2)-(4) introduce the sufficient reduction projections (SRPs) from MNIR models fit to firm-level returns on pandemic fallout days, monetary policy days, and Super Tuesday, respectively. Each SRP variable is scaled to have unit standard deviation. Column (5) introduces all three SRPs. In all cases, we cluster at the NAICS4 level when computing standard errors, and drop industries with fewer than three observations.

highly correlated with the pandemic fallout projection, so we drop them.

Table 6 reports the results. Each projection has statistically significant and economically material effects on the earnings surprises. For example, according to Column (5), a unit standard deviation increase in the pandemic fallout projection value raises the earnings surprise by 0.313 percent of equity market value, conditional on the other explanatory variables, or nearly one half of the IQR of the surprise distribution. Considered jointly, the projections raise the adjusted R^2 value by 42 percent (to 0.204) relative to a specification with NAICS4 effects and firm-level controls. In short, our

that would be an interesting exercise, it would not speak to whether our MNIR models for abnormal returns help predict future earnings surprises.

text-based models for abnormal return reactions to common shocks in early 2020 also help predict earnings surprises later in the year.

As a final exercise, we ask whether the abnormal returns on jump days explained by our text-based models persist over longer horizons. Specifically, we regress cumulative firm-level returns from 14 February 2020 through various end dates on the projection values and NAICS4 effects. For end dates, we use 15 May, 30 June, 30 September, and 31 December 2020. Table B.8 reports the results. For the sake of brevity, we focus on the sufficient reduction projections obtained from our MNIR model fit to the nine pandemic fallout days listed in Table 1. Regardless of end date, we find that the pandemic fallout projection is highly significant and quantitatively important. Raising the projection value by one standard deviation involves an increase in cumulative returns of 0.43, 0.45, 0.36, and 0.37 standard deviations, respectively, on the four end dates. Hence, the firm-level abnormal return reactions to bad pandemic news in the early wake of the pandemic explained by our text-based MNIR models persist through at least the end of 2020.

Overall, these results for earnings surprises and cumulative returns allay concerns that our text-based exposure measures capture only short-lived effects of no consequence for the real economy.

7 Concluding Remarks

We exploit the *Risk Factors* texts in 10-K filings to investigate how market-moving shocks interact with prior risk exposures to drive the structure of firm-level equity returns. We focus on 17 trading days with large market-level moves from late February to the end of March 2020. These 17 days exhibit enormous dispersion in firm-level returns in reaction to news about COVID-19, monetary and fiscal policy, the Super Tuesday Democratic primary elections, and the oil price crash of 9 March.

Our text-based models explain up to half the variation in abnormal firm-level returns on these days. Bad pandemic news triggers large negative return reactions for firms with high exposures to travel, lodging, traditional retail, energy, aircraft production, residential construction, REITs and restaurants, among others. The same news triggers positive return reactions for firms with high exposures to drug trials, e-commerce, basic foodstuffs, web-based services, video games, financial management, and metals and minerals that feed into supply chains for semiconductors, electrical equipment and cloud computing. Many of these reactions reveal powerful effects of downstream demand shocks on upstream suppliers. Examples include aircraft and energy, harmed by the fall in travel demand, and positive return reactions for suppliers to firms that saw demand increases due to social distancing and the shift to working from home. Other market-moving news – such as monetary and fiscal policy actions or the Super Tuesday election outcome – generate quite different firm-level return reactions, which we also characterize using our text-based methods.

The pandemic-induced return reactions we uncover foretell future corporate earnings surprises, as we show. They also presage other shifts in the economy. Examples include major job losses in the traditional retail sector, employment gains at online shopping and delivery firms, a persistent collapse in air travel, job cuts in aircraft production, numerous bankruptcies among oil and gas companies, a collapse of advertizing revenue in print media, and surging demand for cloud computing.[18] Our results also square with evidence that COVID-19 accelerated ongoing shifts to digital services and remote interactions across a whole host of activities. Indeed, the share of new U.S. patent applications that advance technologies to support video conferencing, telecommuting, remote interactivity, and working from home more than doubled in the wake of the pandemic (Bloom et al., 2021). Managing the social and economic fallout of these shifts will present major policy changes for many years.

In terms of methodology, we draw on two text-analytic approaches often seen as alternatives: expert-curated dictionary methods and supervised machine learning. By combining elements of both, we obtain rich models that (a) fit better than models

[18]Barrero et al. (2020) discuss the first four developments mentioned here and Financial Times (2020abc) discuss the others.

based on expert-curated dictionaries, (b) uncover new, empirically relevant exposure categories missed by the curated dictionaries and, at the same time, (c) deliver interpretable patterns in the estimated structure of firm-level returns. This last feature pushes the supervised ML approach from prediction to interpretation.

Our hybrid approach starts with Taddy's (2013, 2015) MNIR implementation of supervised ML. We use a fitted MNIR model to identify influential "seed" terms. For each seed, we build term sets based on similarity of linguistic context in the *Risk Factors* texts and relationship to firm-level returns. These term sets effectively define exposure categories, which we apply to the *Risk Factors* texts to quantify firm-level exposures. The resulting firm-level measures then serve as explanatory variables in return regressions that yield readily interpretable results. While retaining elements of a conventional dictionary-based approach, our hybrid approach leans lightly on domain expertise. It completely sidesteps the laborious construction of expert-curated dictionaries. It also foregoes any reliance on external libraries, as in some NLP methods. As illustrated by our separate applications to pandemic fallout days and Super Tuesday, our hybrid approach is also flexible and adaptable. We hope it will prove useful in many other applications.

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.*

Bloom, N., Davis, S. J., and Zhestkova, Y. (2021). Covid-19 Shifted Patent Applications toward Technologies that Support Working from Home. *American Economic Association, Papers and Proceedings, 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.

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. *Harvard Law School Forum on Corporate Governance*.

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.

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 days 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 NAICS2 code in our dataset. Finally, we keep only NAICS2 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:[19] {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}[20]
- 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}

[19] This category corresponds to the "Other Financial Indicators" term set in Baker et al. (2019).

[20] 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

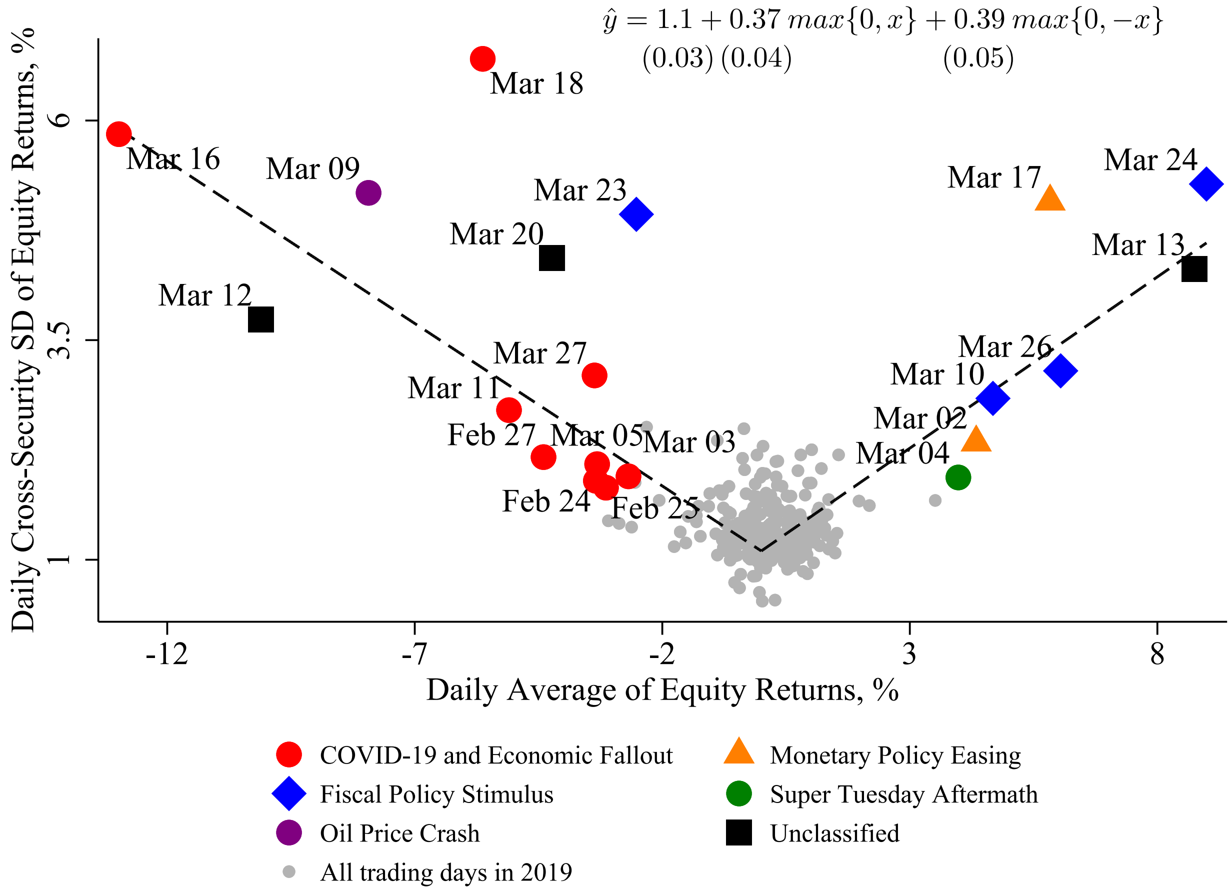


Figure B.1: Value-Weighted Mean and Cross-Sectional SD of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

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.

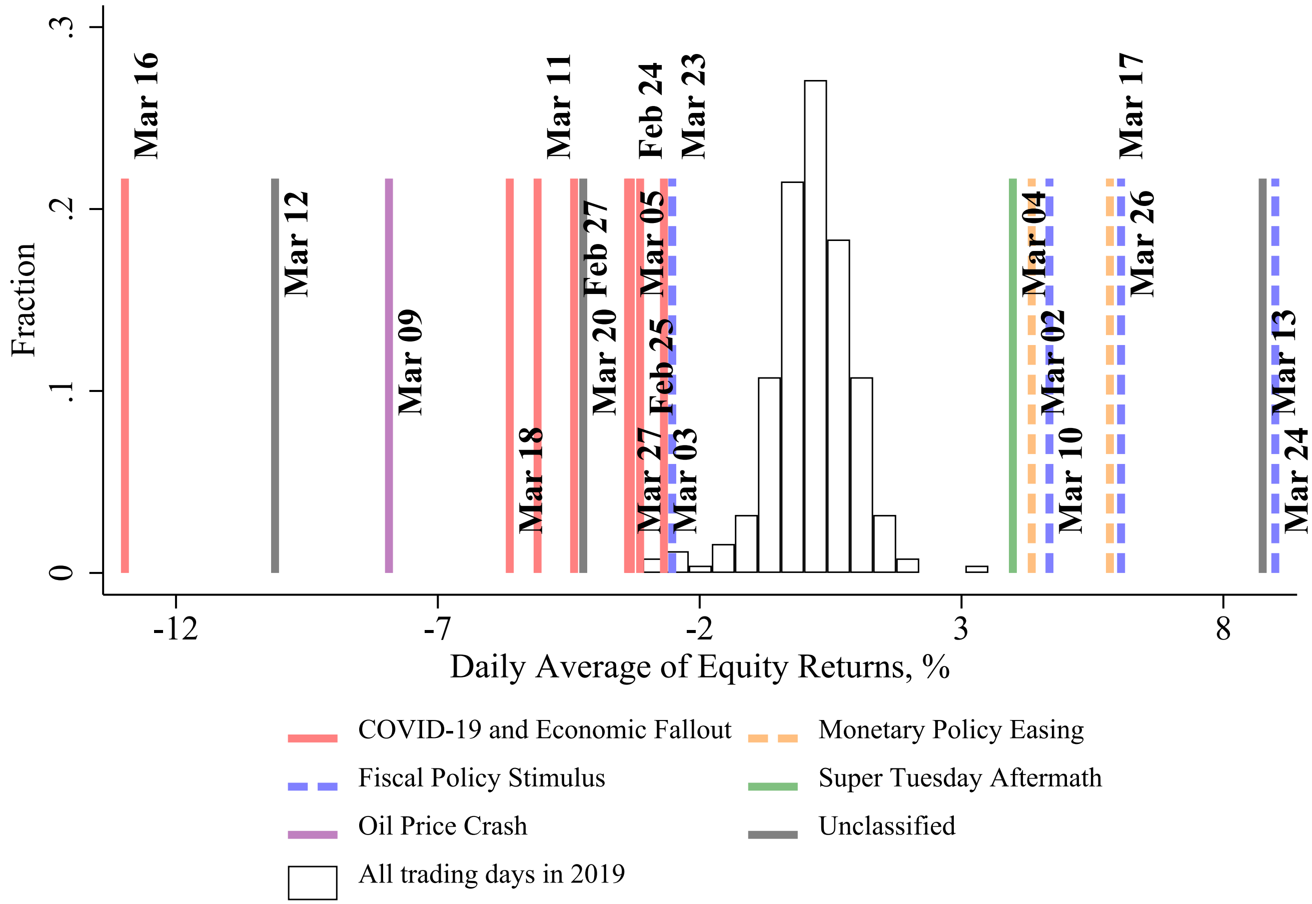


Figure B.2: Value-Weighted Mean of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

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.

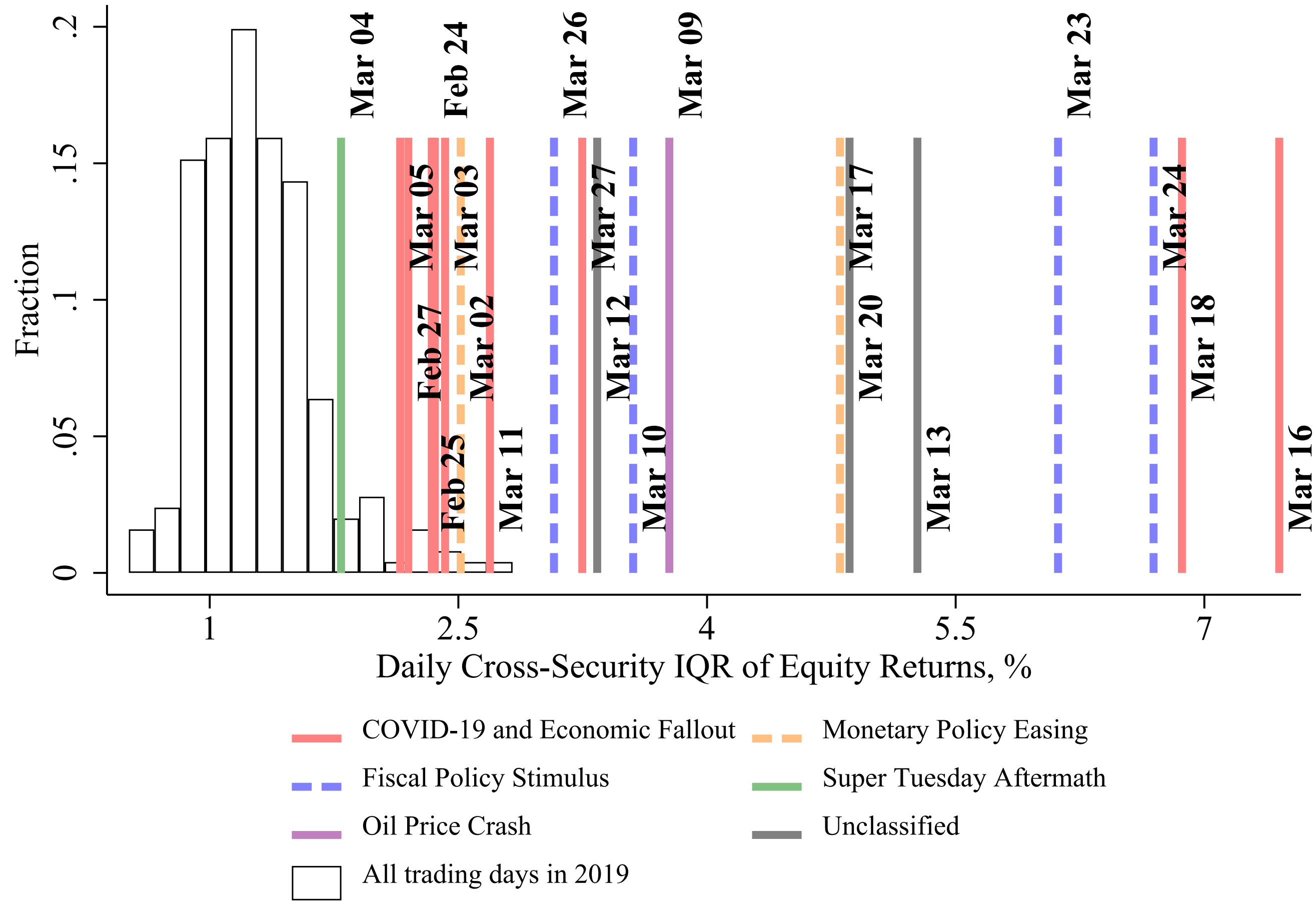


Figure B.3: Value-Weighted Cross-Sectional IQR of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

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.

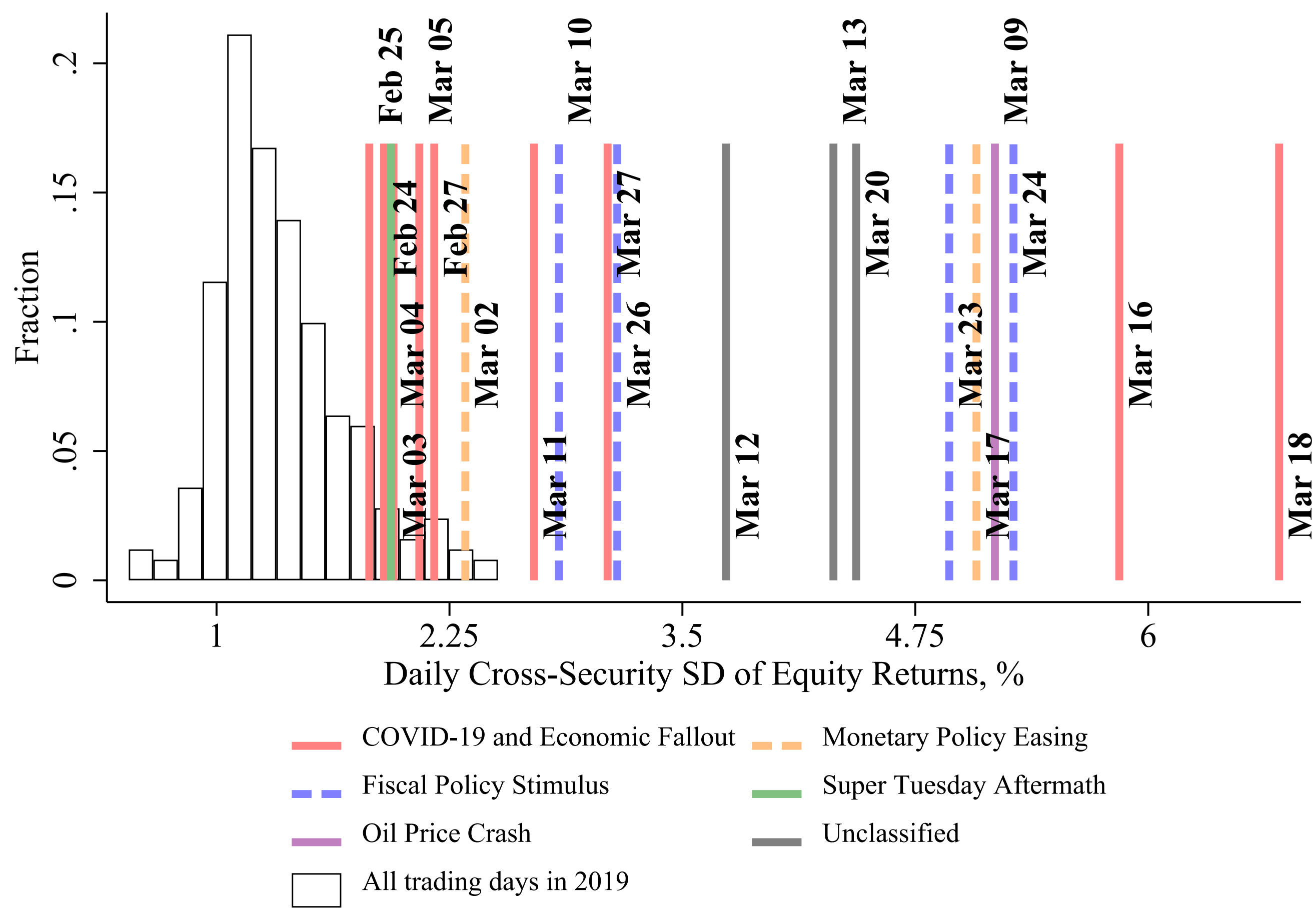


Figure B.4: Value-Weighted Cross-Sectional SD of U.S. Equity Returns, Daily for 2019 and for Large Daily Jumps in 2020

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

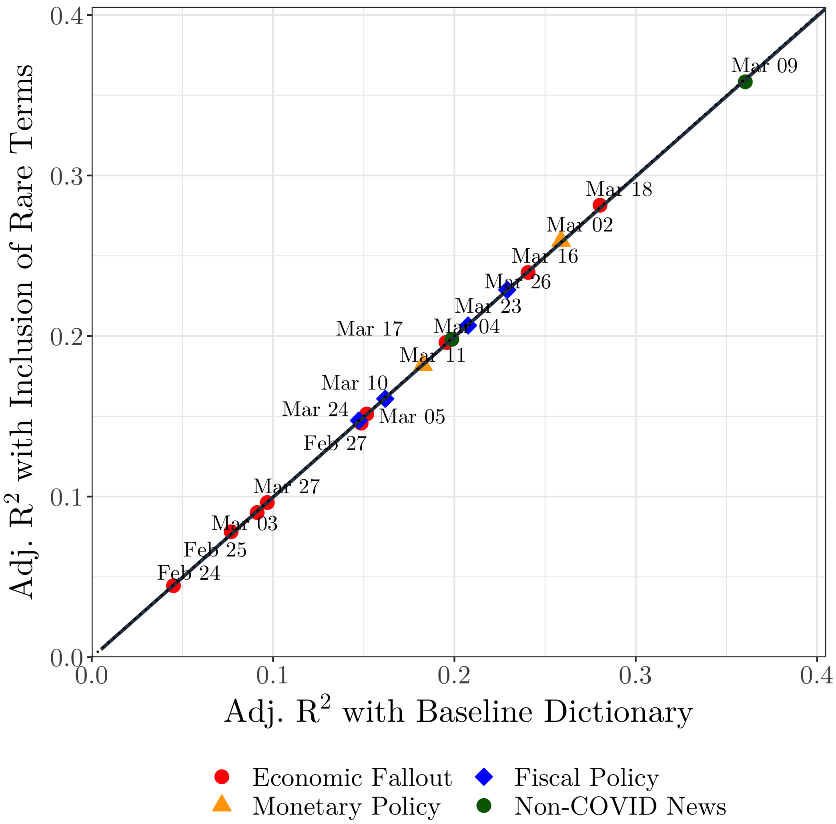
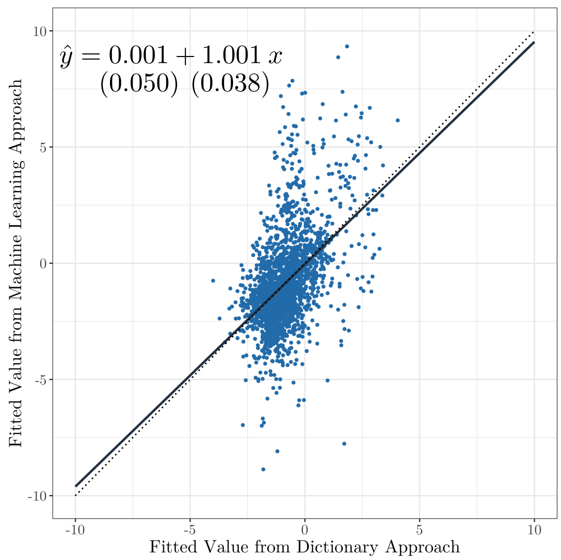


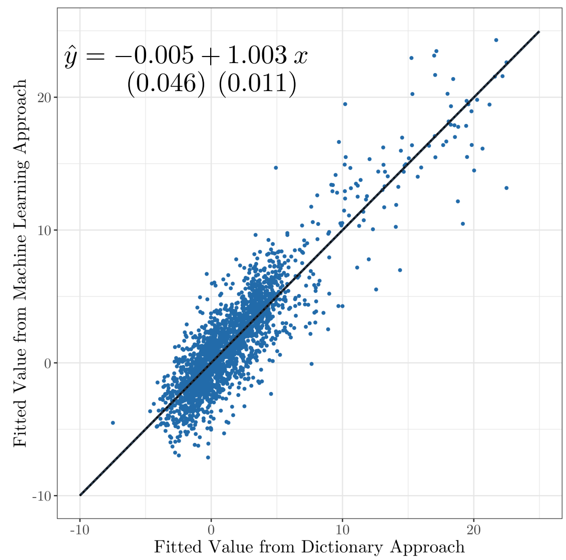
Figure B.5: Adjusted R^2 for Baseline and Complete Dictionaries

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



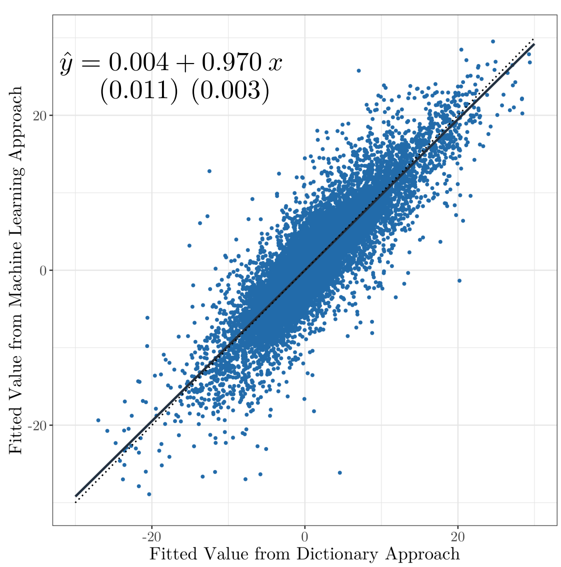
(a) Day with Lowest Dictionary R^2



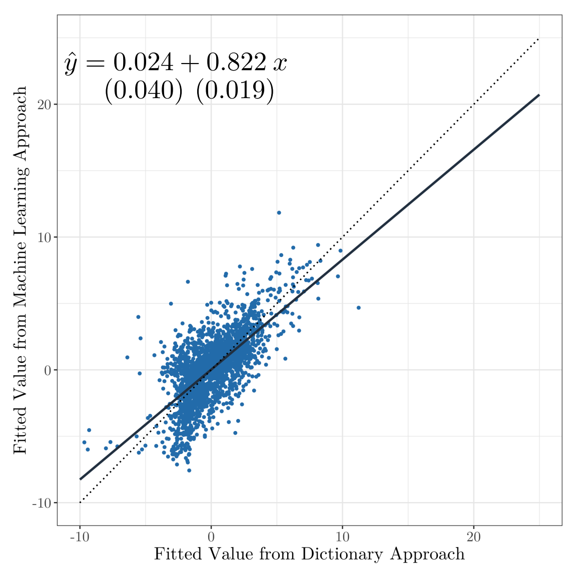
(b) Day with Highest Dictionary R^2

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

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 $x_v \log(\frac{2133}{df_v})$ 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.

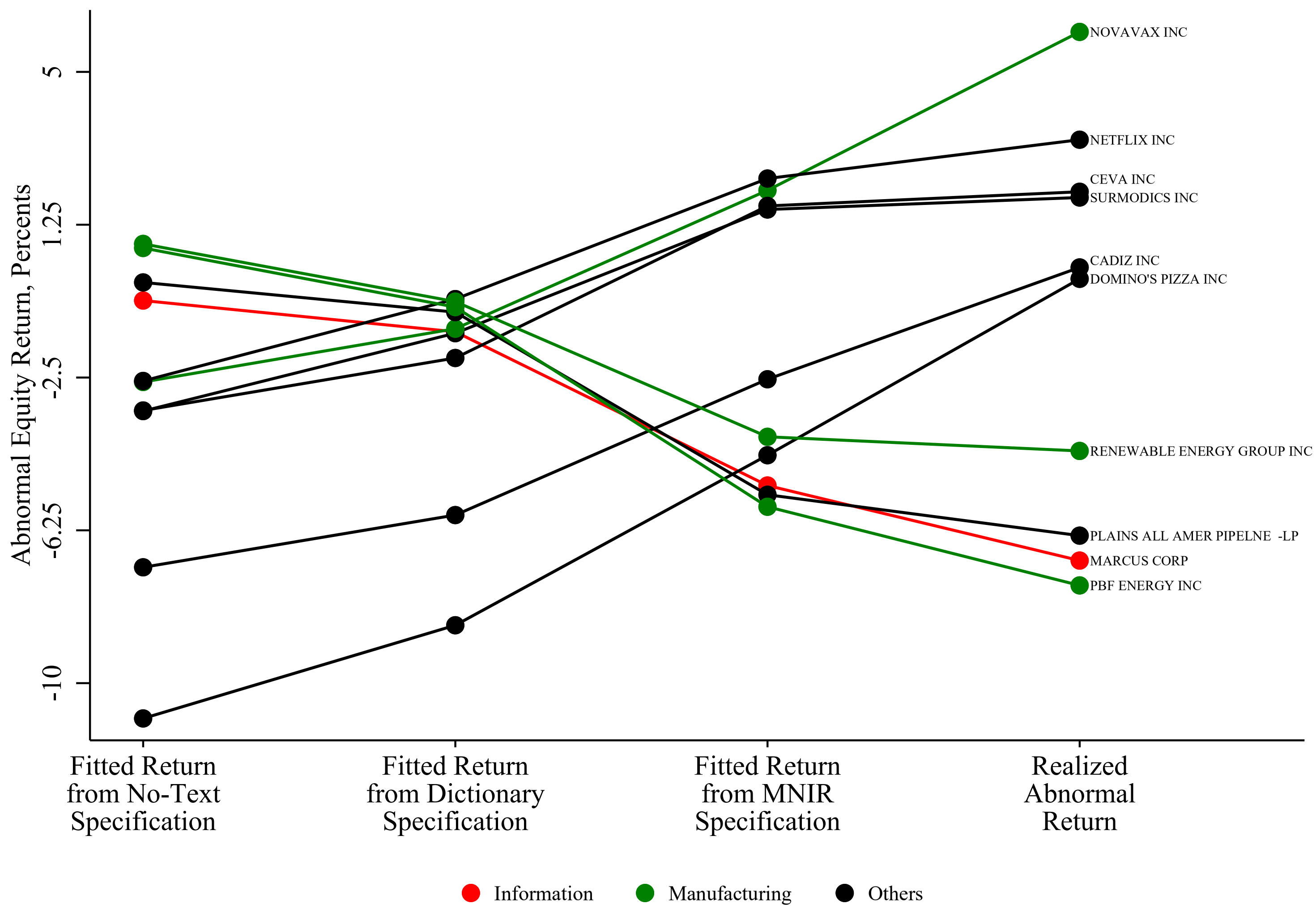


Figure B.8: Predicted and Actual Returns for Firms with Greatest MNIR Fit Gains on Pandemic Fallout Days

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 days listed in the notes to Table 1. We consider security-level averages over these jump days 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	%	Mean		SD		p1	p99
		> 0	**All**	> 0	**All**	> 0		
Advertizing	2155	39	283	720	1507	2338	0	6458
Alternative Energy	2155	25	351	1401	3600	7094	0	4690
Card Payments	2155	41	844	2077	6997	10861	0	8900
Clearing Houses	2155	21	93	444	1274	2761	0	1221
Commercial Property	2155	77	1539	1994	7227	8173	0	25057
Display Technology	2155	39	406	1046	4362	6958	0	6035
Financial Management	2155	82	761	930	2175	2371	0	12056
Foreign Exchange	2155	16	33	203	377	911	0	435
Franchising	2155	29	630	2139	3147	5516	0	12466
Gambling	2155	9	431	4709	4997	15944	0	6013
Gold and Silver	2155	10	144	1390	2032	6175	0	1836
Healthcare Providers	2155	16	103	641	847	2029	0	2386
Insurance	2155	56	682	1223	4485	5953	0	16685
Mortgages	2155	77	1786	2334	6785	7674	0	37103
REITs	2155	76	2353	3092	7201	8115	0	41279
Residential Construction	2155	22	231	1028	1489	3010	0	4250
Restaurants	2155	9	452	5214	5140	16764	0	14326
Traditional Retail	2155	61	425	700	1295	1603	0	5931
Workforce	2155	7	21	316	125	379	0	557
Aircraft + Travel	2155	42	987	2363	7350	11232	0	16330
Communications + Trad Media	2155	61	1484	2450	11191	14301	0	30317
Energy Infr + Oil and Gas	2155	64	2661	4152	15174	18794	0	67763
Drug Trials	2155	68	2105	3110	8057	9632	0	39575
Ecommerce	2155	72	438	610	1621	1886	0	6110
Electronic Components and Devices	2155	73	834	1142	3368	3898	0	12615
Foodstuffs	2155	46	442	963	3551	5198	0	7588
Foreign Countries	2155	90	1608	1786	3000	3111	0	15099
Health Insurance	2155	71	1505	2121	7508	8841	0	21070
Investment Funds	2155	31	440	1436	3266	5778	0	7765
Metal Products	2155	51	518	1014	4800	6676	0	5493
Power Generation	2155	40	266	672	1643	2559	0	4655
Raw Metals and Minerals	2155	22	157	711	789	1557	0	2040
Semiconductors	2155	24	261	1092	1369	2639	0	7368
Video Games	2155	27	735	2766	9209	17713	0	11450
Web-Based Services	2155	68	909	1339	3442	4108	0	13521
Banking + Deposits	2155	83	1889	2286	7184	7845	0	39324
Shipping Cont + Transportation	2155	60	1215	2021	7711	9863	0	18866
Sftw and Hrdw Prod + Sftw Serv	2155	99	3680	3727	6982	7014	0	33594

Table B.3: Descriptive statistics for MNIR-generated firm-level exposures to pandemic fallout news

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 $x_V \log(\frac{2133}{df_v})$ 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.

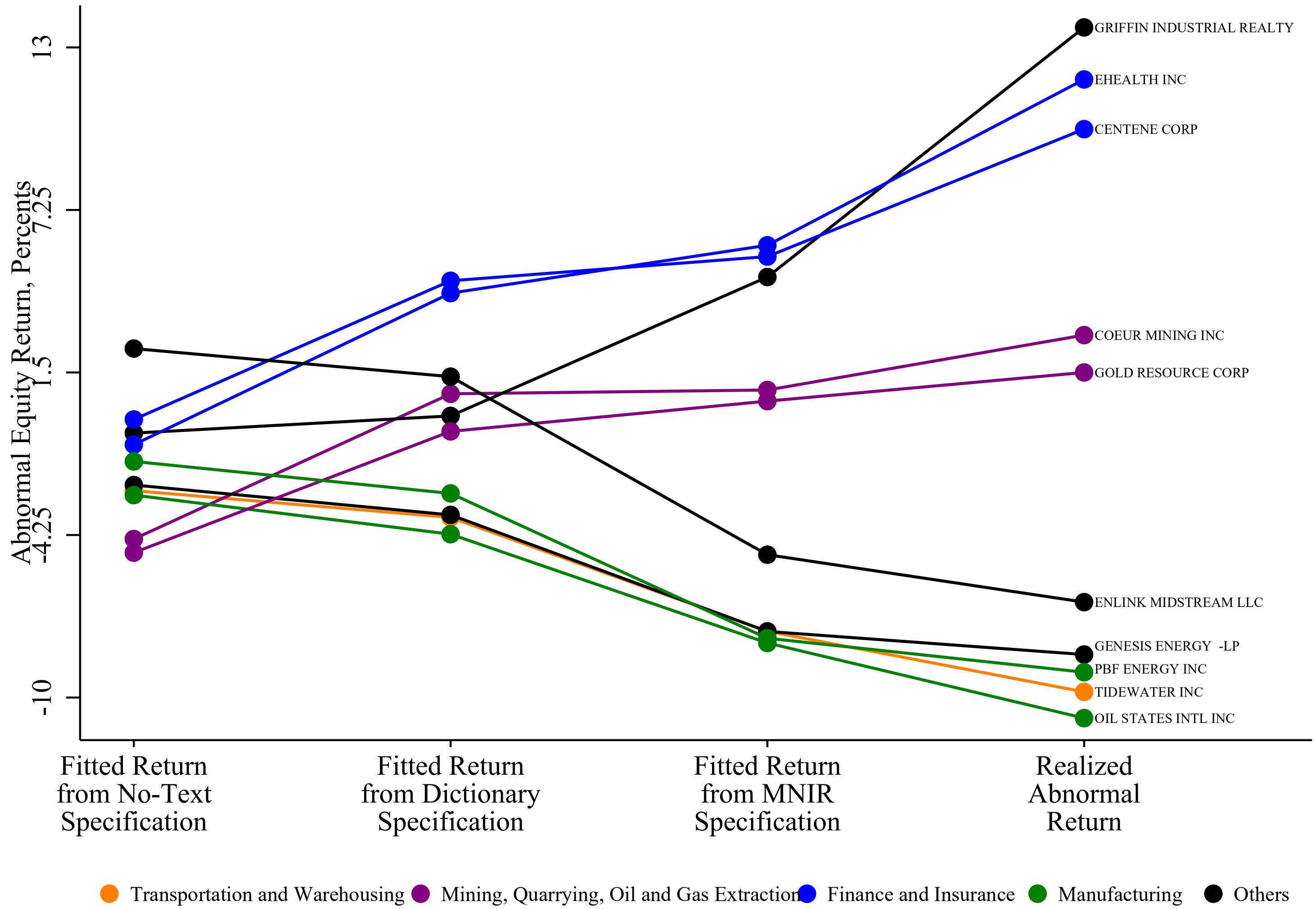


Figure B.9: Ten Firms with Largest Increase in Fit from No-Text to MNIR Specification for Super Tuesday

This figure considers ten firms with the greatest gain in fit by adding the sufficient reduction projection to a model fit to data for Super Tuesday. We 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.

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 five terms with the largest positive or negative MNIR coefficient in the firm-specific sufficient reduction projection when multiplied by tf-idf scores to down weight generic terms.

Seed	Name	Retained Terms	Dropped Terms
aircraft	Aircraft	3	6
mastercard	Card Payments	20	0
derivatives	Financial Instruments	34	0
wheat	Foodstuffs	11	14
gaming	Gambling	6	0
hotels	Hotels	6	5
steel	Industrial Metals	17	8
vehicles	Motor Vehicles	12	8
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
private equity funds	Asset Management	22	0
investment funds	Financial Management	10	0
bank	Banking	35	16
fdic	Financial Regulation	23	0
deepwater	Drilling Activity	21	22
hydraulic fracturing	Fracking	13	3
pipelines	Energy Infrastructure	20	15
oil	Oil and Gas	13	0

(a) Negative Exposures

Seed	Name	Retained Terms	Dropped Terms
homebuilding	Construction	5	0
radar	Defense Technology	17	38
drugs	Drugs	34	13
students	Education	9	9
mobile	Electronic Communication	35	10
mexican	Foreign	9	1
franchisees	Franchising	6	0
government contracts	Government Contracting	10	2
reinsurance	Insurance	25	0
gold	Metals	5	4
navy	Military	9	2
mining	Mining	11	0
reit	REITs	21	0
properties	Real Estate	22	0
space	Rental Market	14	0
restaurants	Restaurants	2	8
utility operations	Utilities	6	2
games	Video Games	6	1
landfills	Waste	5	5
ecommerce	Ecommerce	2	1
subsidy	Subsidies	4	2
health insurance	Health Insurance	9	5
medicare	Government Healthcare	31	2
hospitals	Healthcare Suppliers	20	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	147	2358	1738	6609	0	3716
Card Payments	2155	39	322	837	3176	5077	0	3451
Financial Instruments	2155	81	848	1047	2306	2521	0	9460
Foodstuffs	2155	23	168	741	1723	3568	0	2206
Gambling	2155	11	362	3349	4130	12179	0	5030
Hotels	2155	12	685	5810	6395	17840	0	16468
Industrial Metals	2155	33	330	1014	2563	4418	0	5184
Motor Vehicles	2155	42	426	1017	2263	3412	0	7836
Pollution	2155	45	640	1411	2404	3415	0	12030
Power Generation	2155	87	1014	1164	4802	5128	0	21215
Shipping	2155	22	485	2160	4798	9952	0	9518
Tariffs	2155	40	86	212	570	881	0	1608
Traditional Media	2155	42	446	1064	4031	6172	0	7639
Transportation	2155	25	242	957	1883	3658	0	4981
Travel	2155	27	263	984	1552	2883	0	5674
Asset Mngmt + Financial Mngmt	2155	52	597	1155	3401	4663	0	8793
Banking + Financial Regul	2155	80	1712	2150	6060	6722	0	32918
Drilling Act + Fracking	2155	30	945	3115	6118	10804	0	25243
Energy Infr + Oil and Gas	2155	58	3163	5439	16704	21625	0	80630
Construction	2155	24	153	651	842	1640	0	3261
Defense Technology	2155	44	229	526	1453	2167	0	3807
Drugs	2155	33	988	2967	4059	6604	0	20953
Education	2155	20	205	1037	3071	6843	0	1180
Electronic Communication	2155	80	647	810	2773	3081	0	15197
Foreign	2155	31	135	440	895	1573	0	1953
Franchising	2155	17	527	3020	2314	4818	0	11579
Government Contracting	2155	46	155	336	648	921	0	2885
Insurance	2155	73	684	939	3840	4471	0	18170
Metals	2155	25	137	542	1687	3328	0	1662
Military	2155	44	221	506	1348	2004	0	3657
Mining	2155	34	450	1341	4327	7389	0	7619
Real Estate	2155	87	1725	1981	4321	4576	0	22134
REITs	2155	50	975	1933	3242	4358	0	16733
Rental Market	2155	48	435	900	1739	2417	0	7045
Restaurants	2155	8	126	1556	1433	4829	0	4256
Utilities	2155	24	81	332	612	1206	0	2089
Video Games	2155	12	115	972	1737	4985	0	938
Waste	2155	18	98	556	1085	2532	0	1311
Ecomm + Health Ins + Subsidies	2155	73	1124	1533	6602	7670	0	12905
Gov Healthcare + Healthcare Supp	2155	69	1777	2573	7368	8752	0	28726

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

All statistics are unweighted.

B.5 Additional material for section 6

	Period for Computing Cumulative Returns:			
	2/14-5/15	2/14-6/30	2/14-9/30	2/14-12/31
	(1)	(2)	(3)	(4)
SRP (Pandemic Fallout)	14.908***	13.947***	14.470***	14.190***
	(1.824)	(1.626)	(2.012)	(2.589)
SRP (Monetary Policy)	4.718**	2.645	1.359	−3.023
	(2.084)	(1.764)	(2.101)	(2.348)
SRP (Super Tuesday)	5.702***	4.837***	7.250***	3.648**
	(1.217)	(1.220)	(1.655)	(1.815)
Leverage	−2.669*	−1.440	−1.092	−0.616
	(1.395)	(1.120)	(1.102)	(1.096)
Log Market Cap	7.402***	5.137***	7.444***	3.121***
	(0.886)	(0.750)	(1.005)	(0.953)
Observations	1,997	1,988	1,975	1,955
NAICS4 Effects	Y	Y	Y	Y
R^2	0.490	0.420	0.394	0.310
Adjusted R^2	0.449	0.374	0.346	0.254
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B.8: Cumulative Returns and the Sufficient Reduction Projections

Each column models cumulative returns as a function of the sufficient reduction projections (in standard deviation units) from MNIR models fit to (average) firm-level returns on pandemic fallout days, monetary policy stimulus days, and Super Tuesday. We also control for NAICS4 fixed effects, log market cap, and leverage. Columns (1) to (4) consider cumulative returns beginning from the pre-COVID stock market peak day of February 14 through various end dates: May 15, June 30, September 30, and December 31, respectively. In all cases, standard errors are computed using clustering at the NAICS4 level, and we drop industries with fewer than three observations.

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. 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}

3. Card Payments: {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}

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. 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~~}

7. Financial 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}

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

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

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

11. Gold and Silver: {**gold***, **silver**, ~~concentrates~~, ~~ore~~, ~~sweet~~, ~~lumber~~, ~~el~~}

12. Healthcare Providers: {surgeons*, hospitals, dentists, dental, clinics, pathology, ~~training programs~~}

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. 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}

15. REITs: {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}

16. Residential Construction: {**homebuilding***, **residential construction**, land development, housing}

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

18. Traditional 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~~}

19. Workforce: {workforces*, labor force}

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

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

22. Communications: {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. 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}

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

Exposures associated with positive returns:

1. 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}

2. 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~~}

3. Electronic Components and Devices: {optics*, optical, sensor, ray, filter, graphics, high performance, 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}

4. 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}

5. 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}

6. 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}

7. Investment Funds: {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}

8. 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 man-

ufacturers, 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~~}

9. 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}

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

11. 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~~}

12. Semiconductors: {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}

13. 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}

14. 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}

15. 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}

16. 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}

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

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. 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}

20. Software and Hardware Products: {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}

C.2 Super Tuesday

Exposures associated with negative returns:

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

2. Card Payments: {mastercard*, visa, card, merchants, cards, merchant, debit, cardholders, ach, payment card, payment cards, atm, card transactions, cardholder, interchange, credit card, pci, processors, payment processing, interchange fees}

3. Financial Instruments: {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}

4. 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~~}

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

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

7. 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}

8. 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}

9. 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}

10. 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}

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

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

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

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

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

16. 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}

17. Financial Management: {investment funds*, investment vehicles, separate accounts, asset classes, pooled, advisers, investment strategies, investment management, other investment, investment advisers}

18. 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~~}

19. 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}

20. Drilling Activity: {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}

21. Fracking: {hydraulic fracturing*, fracturing, hydraulic fracturing activities, hy-

draulic fracturing process, sdwa, water act, fracturing process, federal safe drinking, fluids, ~~stimulation~~, hydraulic, ~~lands~~, blm, ~~production activities~~, hydraulic fracturing practices, groundwater}

22. 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~~}

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

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

5. 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~~}

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

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

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

9. Insurance: {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}

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

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

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

13. REITs: {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}

14. 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}

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

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

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

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

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

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

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

22. 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}

23. 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}

24. Healthcare Suppliers: {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}