WORKING PAPER · NO. 2025-87

Macro Shocks and Firm-Level Response Heterogeneity

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

JUNE 2025

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Macro Shocks and Firm-Level Response Heterogeneity
Steven J. Davis, Stephen Hansen, and Cristhian Seminario-Amez
June 2025
JEL No. C55, E30, L20

ABSTRACT

Macro shocks produce high dispersion in firm-level equity returns, sales growth, and other outcomes. We show that this dispersion reflects observable differences in business characteristics. To do so, we combine firm-level returns on stock market ``jump'' days with text about business risks in prior 10-K filings to construct firm-specific shock exposures. Our exposure measures explain firm-level abnormal returns through interpretable variation in language. They also explain most of the increased dispersion in firm-level revenue growth after major shocks and much of the dispersion in employment growth, investment rates, and earnings surprises. Our evidence yields a novel interpretation for countercyclical dispersion, highlighting the key role of heterogeneous business characteristics in macro shock transmission.

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

Stephen Hansen
University College London
Department of Economics
stephen.hansen@ucl.ac.uk

Cristhian Seminario-Amez
Instituto Tecnológico
Autónomo de México (ITAM)
Department of Economics
cristhian.seminario@itam.mx

1 Introduction

Large shocks often disrupt economic activity in a highly uneven manner. As a recent example, real GDP fell 8% in the United States and 11% in the Euro area in 2020Q2, easily the largest drops since World War II.[1] Firm-level outcomes varied enormously in the wake of the pandemic (Barrero et al. 2020, Bartik et al. 2020, Barrero et al. 2021, Bloom et al. 2021, Papanikolaou and Schmidt 2022), echoing findings in a broader literature that documents – and seeks to explain – countercyclical dispersion in business performance.

One leading explanation for countercyclical dispersion is that recessions coincide with greater uncertainty, often modeled as an increase in the variance of idiosyncratic firm-specific shocks (Bloom 2009, Bloom et al. 2018, Senga 2018). Another is that firms are more reactive to idiosyncratic shocks of a given magnitude during downturns (Ilut et al. 2018, Berger and Vavra 2019). We propose a different explanation: Firms are heterogeneously exposed to the shocks that trigger downturns, and these exposures arise from observable, pre-existing business characteristics. Unlike the other two explanations, ours implies predictable patterns in firm-level outcomes, patterns that also differ predictably across macro shocks.

Consider two examples. The COVID shock had radically different effects on NetApp Inc and Scientific Games, two firms in the "Computer and Peripheral Equipment Manufacturing" industry. Their distinguishing business characteristic is the customer base. NetApp provides cloud services to various clients, while Scientific specializes in serving casinos. A second example involves Domino's Pizza and Ruth's Hospitality Group, both in the "Full-Service Restaurants" industry. Here, the distinguishing characteristic is suitability for a stay-at-home world. Pizza is the quintessential home delivery food, while Ruth's runs fine-dining steak restaurants. In these examples, dissimilar responses are predictable from the nature of the shock and observable business characteristics. Moreover, other macro shocks (e.g., an inflation surprise or a fiscal stimulus) yield predictably different response patterns.

To operationalize our explanation, we model firm-level responses to identified shocks as a function of text about business characteristics. We start with "jump" days that involve daily stock market moves greater than $|2.5\%|$. Following Baker et al. (2025b), we classify the news shock that triggers each jump, grouping the jumps and underlying shocks into well-defined categories. We then treat firm-level abnormal returns on jump days as reflective of how the shock alters the firm's outlook. To quantify a firm's exposure to jumps of a given type, we turn to text in the firm's prior 10-K filings. Before describing this step, it's helpful to consider the relationship between market-level and firm-level returns.

[1] These are quarter-to quarter changes, not annualized. See series GDPC1 (US) and CLVMEURSCAB1GQEA19 (Euro area), both retrieved from FRED on 6 October 2023.

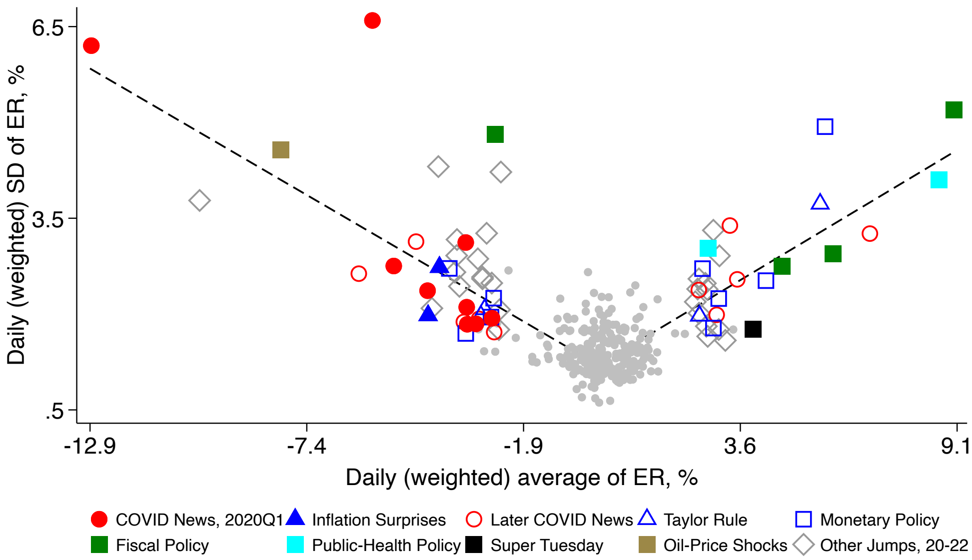


Figure 1: Mean and Dispersion of Daily U.S. Equity Returns, All Trading Days in 2019 and All Jump Days in 2020, 2021 and 2022. The chart shows the value-weighted mean daily return on the horizontal scale and the weighted standard deviation across firms on the vertical scale for about 2,000 firms, as detailed in Section 2. "Jump days" have absolute mean return greater than 2.5%. Gray dots show all trading days in 2019, a rather typical year.

Figure 1 plots daily market-level returns against the same-day standard deviation of firm-level returns on all trading days in 2019 (a typical year) and all jump dates in 2020, 2021, and 2022. We group jump days by the type of triggering news. Not surprisingly, many of the jumps during this period reflect news about COVID. Other jumps reflect inflation surprises, monetary policy, fiscal policy, election news (Super Tuesday), and more. Jump dates show an extraordinary dispersion in firm-level returns, consistent with large differences in firm-level shock exposures. The most extreme case is 18 March 2020, with a cross-firm standard deviation in one-day returns of 6.6 percentage points. That's about 15 standard deviations greater than the average cross-firm standard deviation in 2019.[2]

One could simply use firm-level returns on jump days as shock exposure measures. Cirelli and Gertler (2024) pursue that approach to build firm-level pandemic exposures, following our original paper. Similarly, Gürkaynak et al. (2022) use stock price reactions to monetary policy announcements to study firm-level transmission. While appealing in its simplicity, this approach raises two issues. First, returns may be a noisy measure of exposure. Second, by themselves, returns don't explain *why* exposures vary across firms and across macro shocks.

[2]Returning to our example, average daily abnormal returns across the nine "COVID News, 2020Q1" jumps depicted in Figure 1 were 4.34% for NetApp but -4.92% for Scientific. Similarly, they were -0.04% for Domino's but -9.28% for Ruth's. In contrast, these four firms had average daily abnormal returns of 0.91%, 2.93%, -1.51%, and 1.18%, respectively, across the two "Inflation Surprise" jumps in 2022, a very different cross-sectional pattern. Incidentally, Scientific rebranded as Light & Wonder in early 2022.

We address both issues by projecting abnormal returns onto observables that characterize firm-level shock exposures. Standard observables like industry and firm size are useful in this respect but capture only a small fraction of the relevant heterogeneity. To account for a much richer set of business characteristics, we turn to the *Risk Factors* (*RF*) section of pre-pandemic 10-K filings. Specifically, we build firm-level shock exposures for each jump date in the 2020-2022 period via multinomial inverse regression (Taddy 2013, 2015) applied to firm-level returns and *RF* text.[3] This method yields a scalar value that summarizes each firm's exposure to each jump. By grouping jumps that share a common trigger, we obtain firm-level exposures for each shock category listed in Figure 1. As we show, this approach successfully extracts the information in returns that is useful for predicting firm-level revenue growth, employment, investment, and earnings.

All shock categories listed in Figure 1 contribute to firm-level dispersion, but we subject COVID News, 2020Q1 and Inflation Surprises to a more detailed analysis. These two categories cover the largest negative jumps in 2020 and 2022, respectively, and yield quite distinct firm-level response pattens. To interpret these patterns, we propose a method for clustering the *RF* terms that capture the most important business characteristics for a given shock type. COVID clusters include terms for air travel, information technology, and clinical trials; inflation surprise clusters include terms for mortgages, retail, and fossil fuels. These exposure determinants cut differently across and within industries because of firm-specific differences in supply chains, technologies, customer bases, product characteristics, and more.

To examine how dissimilar shock exposures drive heterogeneity in real outcomes, we regress quarterly revenue growth on our text-based COVID and inflation exposures, each interacted with quarter fixed effects. COVID exposure generates revenue growth heterogeneity throughout 2020 and 2021 and, independently, inflation exposure generates heterogeneity during the early-2020 decline and mid-2022 rise in inflation expectations. These results reflect pre-existing business characteristics (and their interaction with macro shocks), because we build the exposures from *RF* text in pre-2020 10-K filings.

Remarkably, residual return variation on COVID and inflation jump dates is uncorrelated with revenue growth. Thus, our text-based exposures isolate the information in returns that helps predict future real outcomes, even though the residual component accounts for over 50% of return variation. Moreover, the word groups that more strongly drive abnormal returns also more strongly drive revenue growth. Finally, for annual changes in revenue, employment and investment growth, and for earnings surprises, COVID and inflation exposures have

[3] Other applications of MNIR include Gentzkow et al. (2019), who study political polarization in Congressional speech; Cage et al. (2021), who place party manifestos on a left-right scale; and García et al. (2023), who use earnings calls to identify words that correlate with returns in reaction to earnings announcements.

significant explanatory power, while residual returns have none.

Our framework lets us compare model-based predictions for growth-rate dispersion to its actual dispersion. The cross-sectional standard deviation (SD) of quarterly growth rates more than doubles in 2020Q2 relative to 2019. To quantify the contribution of macro shocks to this increase, we fit panel regression models to various shock exposures and compute the SD of the fitted values. A model with standard firm controls and no shock exposure explains very little. In contrast, adding the COVID and inflation exposures increases the fitted SD substantially. When we include exposure measures for all shock categories in Figure 1, we explain most of the SD rise from 2019 to 2020Q2 and more than 80% of the cross-firm growth rate variation in 2020Q2. In this sense, macro shocks are the main driver of cross-firm dispersion during the 2020 recession.

Are these findings specific to the pandemic era? The *RF* section of 10-K filings was first mandated in 2006 and took several years to reach maturity. So, we cannot reliably build text-based exposures to the macro shocks associated with the 2008-09 recession. Yet several facts suggest that shock-exposure heterogeneity is a major source of countercyclical dispersion more generally. First, since 1900, jumps occur three times more often during recessions than expansions. They also tend to be more extreme in recessions. Second, the relationship in Figure 1 between the magnitude of jump-date market returns and firm-level return dispersion holds since at least 2000. Third, while we cannot use text-based exposures over a long sample, we can use abnormal returns directly. When we do so for all jumps from 1995 through 2022 and repeat our panel regression analysis, we again find that response heterogeneity to macro shocks is a major source of countercyclical dispersion.

In summary, we provide empirical support for the relevance of heterogeneous exposures to macro shocks for countercyclical firm dispersion. These shocks shift the first moment of firm-level outcomes in highly dissimilar ways that reflect previously reported business characteristics. We also find that macro shocks differ in predictable ways in their patterns of firm-level response heterogeneity. Our empirical approach is suitable for the analysis of any macro shocks, provided that firm-level data on equity returns are available and can be matched to text about the business characteristics of individual firms. As such, many other applications are possible.

Previous studies of response heterogeneity to macro shocks include Ottonello and Winberry (2020) and Gürkaynak et al. (2022), both of which consider how firm-level financial conditions mediate monetary policy shocks. David et al. (2022) study how heterogeneous exposures to aggregate TFP shocks generate dispersion in capital's marginal product. These papers focus on a limited range of business characteristics and a limited set of macro shocks. We instead consider the full range of salient business characteristics, as reported in prior *RF* text, and many types of macro shocks. In this respect, our method is highly flexible and easily applied to other shocks and settings.

Many studies capture shocks by using asset price changes around policy announcements, inserting them into regression models to estimate effects on other variables. Examples include Kuttner (2001), Gürkaynak et al. (2005) and Ramey (2016). Our results suggest that raw asset price changes are noisy measures of a given shock's impact on individual firms. Thus, raw returns may yield an attenuation bias when estimating a shock's effects on other firm-level outcomes. Fortunately, policy announcements often come with text in the form of public statements (Hansen and McMahon 2016) or remarks by policymakers (Gorodnichenko et al. 2023). Projecting asset price moves onto these text sources and extracting a firm-specific shock exposure measure, as in our study, is one way to reduce the potential bias.

Hassan et al. (2019) pioneered the use of firm-level text data to build exposure measures for political risk. The basic method has been extended in many ways, including to firm-level pandemic exposures (Hassan et al. 2023). They use corporate earnings calls after the arrival of COVID-19 as their text source, and they focus on explicit mentions of the pandemic. In contrast, we build our COVID exposure measures by combining (i) stock market jumps attributed to COVID news in next-day newspaper accounts and (ii) pre-shock corporate filings that provide exhaustive descriptions of business risks. This is a direct measure of news shock exposure that can be computed immediately upon the arrival of the shock. We discuss the relationship between the two measures more below.

The paper proceeds as follows. Section 2 describes the data, and section 3 details the construction of our text-based shock exposures. Sections 4 and 5 establish their predictive content for abnormal returns and real outcomes, respectively. Section 6 discusses the role of dissimilar shock exposures in firm-level outcomes. Section 7 concludes.

2 Data

2.1 Risk-Factors text from 10-K reports

Since 2006 (for fiscal year 2005), the Securities and Exchange Commission (SEC) has required the vast majority of publicly held firms to include a discussion of *Risk Factors* (*RF*) in Part 1A of their annual 10-K filings. The SEC advises that these discussions include any item that could impact future earnings. Investors can sue for compensation if the firm omits material information or risks (Mast et al. 2020). The result is that *RF* text contains a wide variety of content related to firms' business models, trading partners, technology adoption, and markets that provides a rich accounting of firm heterogeneity that extends well beyond traditional numeric data.

We use *RF* texts filed from 2015 to 2019. 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 firm outcomes in later years reflects persistent firm attributes that long predate the arrival of macro shocks associated with COVID-19 and the subsequent inflationary episode. We obtain the near-universe of filings in this period by directly scraping EDGAR.[4] After removing a handful of companies with empty or short entries for *RF*, we obtain a sample of 7,259 total firms.

2.2 Firm-level returns and other financial measures

For each year in 2020, 2021, and 2022, we build a sample of firms' securities with available information in the Compustat daily securities dataset and that can be matched to a firm in the *RF* dataset. We restrict attention to U.S.-incorporated, non-financial firms with share prices quoted in U.S. Dollars. We also remove securities whose share price is sufficiently small. See Appendix A (Online) for more information about the sample construction. In total, our sample contains 1905 securities in 2020, 2028 securities in 2021, and 1818 securities in 2022. Overall there are 2,231 unique securities across all three yearly samples.[5]

To compute raw 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 construct daily abnormal returns in the standard way: the difference between (i) a stock's actual return in excess of the risk free rate and (ii) its expected excess return as per the Capital Asset Pricing Model (CAPM):

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

where $p_{i,t}$ denotes the adjusted share price for stock i on day t, $R_{f,t}$ denotes the four-week treasury bill rate (a proxy for the risk free rate), $beta_i$ is the stock's CAPM beta, and $R_{M,t}$ 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 the corresponding previous year (2019, 2020, or 2021).

[4]We obtain URLs for company filings via the API provided by https://sec-api.io/.

[5]Because the same firm can trade multiple securities, in a limited number of cases the same *RF* text is matched to multiple securities. The total number of firms in 2020, 2021, and 2022 is, respectively, 1876, 2002, and 1797. The number of unique firms in all three years is 2200.

In our statistical model of text and returns, we include two controls for financial characteristics. The first is a measure of the firm's equity market capitalization, computed as shares outstanding (CSHOC) times closing price per share. The second is firm leverage, computed as (long term debt (DLTT) + current liabilities (DLC)) divided by total assets (AT). To model returns for a jump date in year t, these values are taken from year $t-1$ Compustat files (or year $t-2$ in case of missing information).

2.3 Jump dates

Baker et al. (2025b) define a jump date as a day on which the aggregate US market return exceeds 2.5% in absolute value. They also provides a categorization of the type of macro news that triggers jump dates. The methodology uses human readings of next-day newspaper accounts to group jump dates into classes of related news. Baker et al. (2020) further refines the classification of jump dates in 2020 to account for the specific context of the pandemic. Jump dates and classifications are available through October 2020 at `https://www.stockmarketjumps.com/`.

Since our sample consists of jump dates from 2020 through 2022, we use an extended set of classifications produced by the same research team. There are 67 jump dates in total in our sample: 37 in 2020, 2 in 2021, and 28 in 2022. The jump dates and their classifications—which correspond to the groups in Figure 1—are presented in Table A.1. A full description of the data and methodology is available at `https://tinyurl.com/n5ptwsk2`. The "COVID News, 2020Q1" and "Later COVID News" categories both consist of dates driven by news on the economic fallout from COVID-19. Those in the former (latter) group consist of dates that fall in (after) the first quarter of 2020. We separate them to show how isolating jump dates early in the sample can explain subsequent firm dynamics. Dates in "Later COVID News" also provides out-of-sample dates we use to assess the performance of exposures derived from "COVID News, 2020Q1". The "Monetary Policy" and "Taylor Rule" categories are distinguished by whether the news relates to an unexpected policy action given economic conditions (Monetary Policy) or to the Fed's anticipated reaction to changes in economic conditions (Taylor Rule).[6]

The "Inflation Surprises" group is the only one that does not correspond to the categories in Baker et al. (2020) and Baker et al. (2025b). We form it since inflation was a large driver of macroeconomic dynamics in the wake of the pandemic. May 18, 2022 (-4.04% market return) and September 13, 2022 (-4.32%) are the two dates with the largest negative return among all categorized 2022 jump dates.[7] Moreover, next-day newspaper accounts

[6] See pages 64-71 in `https://www.stockmarketjumps.com/files/newguide.pdf` for more discussion.

[7] The categorized date with largest absolute return is November 10, 2022 (5.63%) which we use for out-

indicate an increasingly negative inflation outlook as the driver of the market return.[8] Note that, although both days' jumps are triggered by inflation news, the market reaction also incorporates knock-on effects arising from shifts in expected earnings and interest rates.

2.4 Real outcomes

Our primary real outcome measure is twelve-quarter revenue growth. We use twelve-quarter growth rates to guarantee that each post-pandemic quarter is compared with a pre-pandemic one, given our data runs through 2022Q4. To build the sample, we select firms that (i) have a complete record of quarterly revenue from 2015Q1 through 2022Q4 in Compustat, (ii) have a complete record of leverage (as computed above) and total assets (which we use as a size control) over the same horizon, and (iii) can be matched to our securities sample in 2020, 2021, and 2022. This yields a balanced sample of 992 firms.

For annual revenue and employment, we select firms with a complete record of outcomes and controls from 2007 through 2022. This yields a balanced sample of 703 firms for revenue and of 671 firms for employment. For annual investment, we follow the construction of Stein and Stone (2013) and obtain a balanced sample of 671 firms.

We additionally consider earnings surprises which we construct from I/B/E/S. 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 2019Q4.[9] 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 2020Q3 EPS minus average of analysts' forecasts of 2020Q3 EPS)/(stock price per share on 15 November 2019). 2020Q3 is the furthest horizon for which I/B/E/S provides forecasts formed in 2019Q4. We construct this measure for 1,302 firms that can be matched to our securities sample.

---

of-sample analysis below.

[8]The New York Times described the May 18, 2022 move as follows (bold type ours):

> A short reprieve for investors ended abruptly on Wednesday as stocks had their worst day yet in a series of already ugly drops after shrinking profits by major retailers **reignited Wall Street's fear of high inflation** (Marcos 2022).

while the Wall Street Journal described the September 13, 2022 move as:

> Stocks suffered their worst day in more than two years after **hotter-than-expected inflation data dashed investors' hopes** that cooling price pressures would prompt the Federal Reserve to moderate its campaign of interest-rate increases (Langley and Ostroff 2022).

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

3 Measuring Firm-Level Shock Exposures

The starting point for measuring firm-level exposures to macro shocks is firm-level equity price movements on jump dates. These reflect market beliefs on how a given news shock will impact future firm earnings. Figure 1 in the Introduction shows that there is vast dispersion in raw returns on jump dates. Figure C.1 (Online) repeats the same analysis but first strips returns of NAICS4 effects. While dispersion is mildly reduced, a large amount of firm heterogeneity remains, which further motivates the need for alternative data sources. The remainder of the analysis uses abnormal rather than raw returns. Let AbnRet_i^d be firm i's abnormal return on jump date d. In addition to date-level analysis, we also model the average abnormal return across the dates that make up the groups in Figure 1. Averaging abnormal returns across dates with related sources of macro news potentially strengthens the overall signal in prices.

Our view is that shock exposure arises from business characteristics difficult to measure with traditional data: product design, supply chains, technology adoption, and so forth. Our goal is to isolate variation in abnormal returns explained by this complex bundle to purge noise unrelated to fundamentals. To capture them, we use the Risk-Factors (RF) text of 10-K filings described in Section 2. We adopt a bag-of-words model in which each unique vocabulary term in the RF text is considered an independent feature. Appendix A (Online) details the preprocessing steps we undertake to generate features.

Let $c_{i,v}$ be the count of term v in the 10-K filings of firm i between 2015 and 2019, inclusive; $\mathbf{c}_i$ be the corresponding vector of term counts; $C_i = \sum_v c_{i,v}$ be the total number of terms in i's filings; and V the number of unique vocabulary terms. Since each year of returns data contains small differences in firm coverage, the set of 10-K filings that goes into the model is different for 2020, 2021, and 2022. In all cases, the total number of words is approximately 20 million; V is above 14,000; the average filing length is roughly 11,000 words; and the standard deviation in filing lengths is 7,000. The recent advent of large language models has introduced algorithms that represent text in more complex ways by allowing for individual words to interact with their context. Whether the bag-of-words model is nevertheless sufficient for extracting exposure-relevant information from our corpus is an empirical question. Our later results indicate this is the case.

3.1 Multinomial inverse regression and firm-level shock exposures

We use multinomial inverse regression (MNIR; Taddy 2013, 2015, Gentzkow et al. 2019) to estimate the cross-sectional relationship between abnormal returns on each jump date d and RF terms counts, while controlling for standard, non-text-based firm observables. MNIR posits that the count vector $\mathbf{c}_i$ is drawn from a multinomial distribution with probability vector $\mathbf{p}_i^d$. That is, each firm has a jump-date specific distribution over vocabulary terms. The parameterization of the vth element of the probability vector is

$$p_{i,v}^d = \frac{\exp(\eta_{i,v}^d)}{\sum_v \exp(\eta_{i,v}^d)} \quad \text{where} \quad \eta_{i,v}^d = \alpha_v^d + \beta_v^d \text{AbnRet}_i^d + \left(\boldsymbol{\gamma}_v^d\right)^T \mathbf{controls}_i^{y(d)} \tag{2}$$

where $\mathbf{controls}_i^{y(d)}$ contains NAICS2 sector, market capitalization, and leverage.[10] We also fit the model by replacing AbnRet_i^d with average abnormal returns for grouped dates. Taddy (2015) approximates MLE by fitting, for each term v, an independent Poisson regression of $c_{i,v}$ on AbnRet_i^d and $\mathbf{controls}_i^{y(d)}$.[11] The estimate $\hat{\beta}_v^d$ captures the relationship between returns AbnRet_i^d and pre-2020 usage of term v in *RF* text.

MNIR associates terms with abnormal returns, but does not immediately provide a firm-level exposure. To obtain such a measure, we use a *sufficient reduction projection* of the *RF* text (Cook 2007, Taddy 2015). Namely, firm i's exposure to the news shock on jump date d is $z_i^d = \sum_v \hat{\beta}_v^d \frac{c_{i,v}}{C_i}$. In words, z_i^d is a weighted average of all estimated MNIR coefficients on returns, where the weights correspond to the share of i's *RF* language made up of term v. Hence, the more firms use terms associated with positive (negative) returns on date d, the more positive (negative) their estimated exposure to date-d news will be. z_i^d is called a sufficient reduction projection because $\text{AbnRet}_i^d \perp \mathbf{c}_i \mid z_i^d, \mathbf{controls}_i^{y(d)}, C_i$. In words, all the information contained in the high-dimensional count vector $\mathbf{c}_i$ relevant for predicting AbnRet_i^d (conditional on controls) is summarized by the scalar value z_i^d. We can therefore use z_i^d as a low-dimensional representation of $\mathbf{c}_i$ that captures the association between pre-pandemic business practices as captured in *RF* text and jump-date returns.

3.2 Relationship among shock exposures

Our method yields a firm-level exposure to each of the nine distinct macro shocks in Figure 1. Any one of these can be isolated for further study, but for illustration purposes we select two. Table C.1 (Online) displays the correlation between exposures. The "COVID News, 2020Q1" and "Inflation Surprises" shocks have a low correlation with each other (-0.0913) while every other exposure has a correlation with at least one of these two that exceeds 0.6. Moreover, these two shocks represent the main sources of macro disruption in the 2020-

[10] The latter two controls are specific to the year y in which the jump date falls, hence the superscript $y(d)$.

[11] We apply LASSO regularization to regression coefficients other than the sector dummies. This is a conservative approach: we allow sector membership to explain term usage, and only include additional variables if they have sufficient, additional explanatory power.

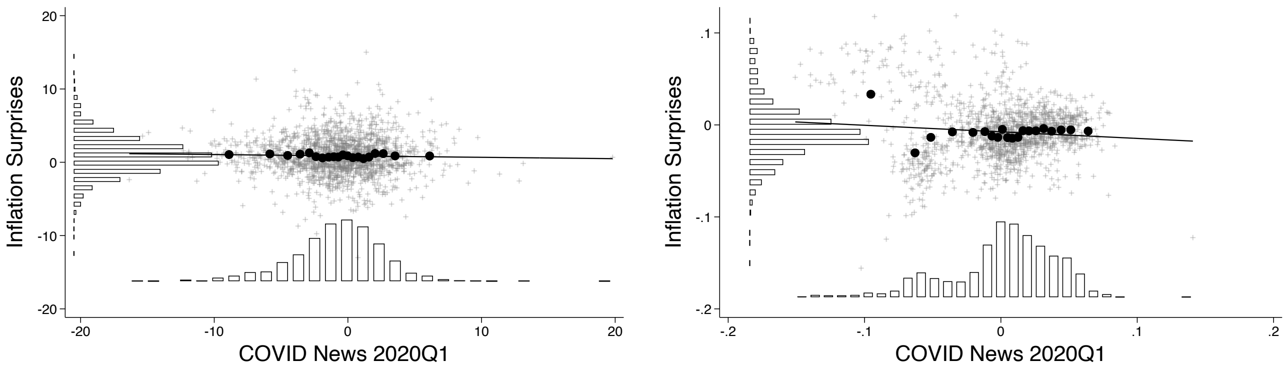
2022 period. We denote the average abnormal return on “COVID News, 2020Q1” dates as AbnRet_i^P and the associated text-based exposure as z_i^P. For brevity, we will hereafter refer to *pandemic* exposure for this shock. Similarly, let AbnRet_i^I and z_i^I denote returns and text-based exposures, respectively, for the ‘Inflation Surprises” shock. We will hereafter refer to *inflation* exposure for this shock. Our focus on these exposures does not imply other shocks do not generate firm-level dispersion in our sample. Indeed, in Section 6 we show other shocks have substantial independent effects on firm-level outcomes, and any of them could be further studied using the approach below.

Table A.2 lists terms most associated with positive and negative returns on pandemic jump dates. Among negative terms are those that relate to property, leisure, travel, and energy. Among positive terms are those that relate to remote services, drug trials, and technology. Table A.3 displays terms with negative and positive associations with inflation jump-date returns. Negative terms reflect language associated with property, debt, and healthcare. Positive terms reflect language associated with travel and energy. At first pass, then, the model gives a plausible account of key drivers of firm-level exposure. Since NAICS2 effects are in $\mathbf{controls}_i^{y(d)}$, these associations do not simply reflect sector membership. As we show below, specific exposures are widely distributed in the economy.

Sign of Terms in Inflation Regression Sign of Terms in Pandemic Regression	Negative	Zero	Positive
Negative	1701	1289	1833
Zero	1288	1220	1027
Positive	2243	1516	1713

Table 1: Signs of Return Coefficients in Pandemic/Inflation Inverse Regressions.
We collect the estimated β_v^P and β_v^I coefficients and cross-tabulate their signs. LASSO penalization generates a zero coefficient in some cases.

The influential terms associated with both shocks show they are expected to operate in a distinct manner. For example, mortgage-related terms appear in the negative inflation list but neither of the pandemic lists, while data-center-related terms are unique to the positive pandemic lists. In other cases, the sign on terms switches: strikingly, ‘hotel’ is the most negative pandemic term but the most positive inflation term. Similarly, ‘crude oil’ is among the most negative pandemic terms and most positive inflation terms. But such sign-flipping is not universal: ‘reit’ (an acronym for real estate investment trust) is among the most negative terms on both lists. To take a more systematic approach, Table 1 cross-tabulates the signs of the estimated β_v coefficients in the pandemic and inflation inverse regressions (because of LASSO penalization some coefficients may be zero). Clearly, there are systematic differences in the associations between terms and returns across the two shocks. Interestingly, more terms have opposite signs across the two sets of jump dates (1833 + 2243) than share a sign (1701 + 1713).



(a) Abnormal Returns

(b) Text Exposures

Figure 2: Distributions of Pandemic and Inflation Exposures
Left panel: scatterplot of AbnRet_i^P (x-axis) against AbnRet_i^I (y-axis) along with a fitted regression line. The histograms represent marginal distributions. Right panel: analogous figure for text exposures z_i^P and z_i^I.

The left panel of Figure 2 plots AbnRet_i^P and AbnRet_i^I, while the right panel plots z_i^P and z_i^I. In both cases, pandemic and inflation exposures are essentially uncorrelated, further emphasizing that the underlying shocks operate differently at the firm level. Exposures also exhibit generally smooth variation, meaning that heterogeneity arises at the level of the firm in a way that cannot be accounted for by a discrete clustering structure (such as would be provided by alternative sector definitions). The heterogeneity also suggests that firms do not adopt generic boilerplate language in *RF* text but instead highlight unique sources of risk—a claim further supported by our results below.

3.3 Components of pandemic and inflation exposure

A firm's exposure to a given macro shock can reflect many underlying characteristics. A key strength of our text-based approach—compared to using returns alone—is that it allows us to explore these more directly. However, the large number of individual terms in the MNIR model makes it challenging to interpret which specific forces are at play. A starting point to overcome this problem is to observe that various terms in Tables A.2 and A.3 appear closely related. For example, negative pandemic terms include 'airline', 'aircraft', and ‘airline.industry’, while negative inflation terms include ‘mortgage’ and ‘mortgage.loan’. This suggests there are sets of terms that reflect the drivers through which a shock operates.

To construct such sets, we propose a simple clustering algorithm that combines information on a term’s estimated MNIR coefficient and its semantic similarity to other terms as measured by a word embedding model. A word embedding model represents each vocabulary term v as a low-dimensional vector $\mathbf{e}_v \in \mathbb{R}^K$. The goal is for words that share similar meanings to have similar vectors, where similarity is typically computed using cosine similarity.[12] To estimate the embedding model, we use *RF* texts from all 7,259 firms for which we retrieve filings from 2015-2019. We use documents from all firms rather than those from just the returns sample in order to increase the available information for determining relationships among words.[13]

For a given jump date(s) d, we construct separate sets for positive and negative terms. The clustering algorithm proceeds in the following steps:[14]

1. **Rank Terms**. Let $\text{tf}_v = \sum_i c_{i,v}$ be the term-frequency for term v in the whole corpus. We rank terms based on $|\hat{\beta}_v^d| \times \text{tf}_v$, i.e. the average contribution of term v to the overall firm-level exposure z_i^d. Tables C.2 (Online) and C.3 (Online) rank terms under this alternative weighting scheme.

2. **Construct sets of *Seed* Terms**. Take the top 200 terms according to the preceding ranking. Place the top-ranked term in set L_1. Place the second-ranked term in the same set as the highest-scoring term if the cosine similarity of their word embeddings exceeds 0.7, else place it in a new set L_2. For each subsequent term, compute its cosine similarity with each existing set and assign to the most-similar set if the similarity exceeds 0.7, else place the term in a new set. The embedded representation of a multi-term set is formed by computing the average embedding vector over its constituent terms. After cycling through the first 200 terms, retain only sets containing at least two entries. Terms that form these sets are referred to as *seed* terms.

3. **Expand *Seed* sets**. Consider now the 1,000 terms with the highest absolute values for our ranking measure, excluding those already identified as *seed* terms in the previous

[12] The cosine similarity between vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ is $\mathbf{e}_i \cdot \mathbf{e}_j / ||\mathbf{e}_i||||\mathbf{e}_j||$.

[13] The specific model we use is the skipgram variant of Word2Vec (Mikolov et al. 2013). In our application of Word2Vec, each word embedding is characterized by a 100-element vector. These embeddings are trained to effectively predict which words surround each word in a window of plus or minus four words. See Ash and Hansen (2023) for more details on word embeddings models.

[14] Hanley and Hoberg (2019) and Li et al. (2021) also use seed words to initially populate lists that are then expanded by word embeddings. Methodologically, our construction can be viewed as an automated way of building dictionaries that are targeted at explaining firms’ reactions to particular shocks. Previous work by Baker et al. (2025a) introduced dictionaries for explaining stock market volatility from 10-K filings, but these are not tailored to the drivers that specific shocks induce.

stage. Assign each term to the *seed* set with the highest cosine similarity, provided it exceeds 0.7. Retain only sets that contain at least 5 terms.

Table A.4 displays output from the clustering procedure for the pandemic. There are 14 negative and positive groups, respectively, which we identify with an index value as well as the highest-ranking seed term. Appendix B (Online) prints all terms in each set. (We discuss the 'Coef' and 'T-Stat' columns in the next Section.) Table A.5 displays analogous information for inflation. There are a wide range of interpretable, specific forces through which the shocks operate. These include sets related to the terms that appear in Tables A.2 and A.3, e.g. a negative term set for the pandemic (inflation) shock containing travel-related (housing/debt) terms. But there are also many others. For example, we find negative pandemic exposure is driven by retailing, traditional media, and energy, among others, while positive exposure is driven by IT infrastructure, clinical trials, and intellectual property. For the inflation shock, negative drivers include real estate, debt, and retail sales, while positive drivers include energy, travel, and semiconductor manufacturing.[15]

These term sets allow one to isolate the part of the overall firm-level exposures arising from a given component. More concretely, for shock event $d \in \{P, I\}$ and for each term set j with associated term set L_j, we compute $z_{i,j}^d = \sum_{v \in L_j} |\hat{\beta}_v^d| \times \frac{c_{i,v}}{C_i}$, which is the analogue of the sufficient reduction projection but applied only to terms in L_j.

To better understand the sources of firm-level heterogeneity, Figure A.1a displays the distribution of a particular, negative pandemic exposure: number 11 from Table A.4 with top-ranked seed word 'airline' which we interpret as 'Travel' based on the full set of terms in Appendix B (Online). There is large cross- and within-industry exposure variation. As expected, the maximally-exposed firm is in the NAICS 4811 industry 'Scheduled Air Transportation'. Nevertheless, individual firms in other industries also have high exposure, in part from supply linkages. For example, Gogo Inc (5179, Other Telecommunications) is "the world's largest provider of broadband connectivity services for the business aviation market."[16] Sabre (5182, Computing Infrastructure Providers) is "a technology company that powers the global travel industry." The other firms highlighted in the Figure provide other instances of travel-linked businesses.

Figure A.1b shows the distribution of pandemic exposure 16 which includes terms 'solu-

[15]Since MNIR is a bag-of-words model, it does not distinguish the context in which particular words are used, and so we cannot distinguish between a term being positive for some firms and negative for others. For example, some firms might use the term 'oil' more frequently because they purchase oil as a production input. For these firms, an increase in energy prices would be negative. However, the MNIR estimates indicate that, on average, a higher use of the term 'oil' is perceived by markets as good news in the wake of the inflation events we study.

[16]The quoted text and those that follow in this section come from the Business section of the given firm's most recent 10-K filing.

tion', 'platform', 'cloud', and 'software' and which we interpret as 'IT Environment'. Again, there is wide variation within and across industries. Naturally, there are many exposed firms in software and computer industries. Beyond these, Cryoport (3391, Medical Equipment and Supplies Manufacturing) supports the movement of medical goods across temperature-controlled supply chains, "including apheresis collection and cryoprocessing, global logistics, technologically sophisticated packaging, biostorage and bio-services, informatics, and cryogenic systems." Mistras Group (5413, Architectural, Engineering, and Related Services) "enhances value for its customers by providing data driven solutions that digitalize the asset protection process and provide valuable insights to our customers that maximize uptime of the assets monitored."

Figure A.2 selects example negative (2 'Mortgages and Real Estate') and positive (19 'Oil and Gas') exposures arising from the inflation shock. As with the pandemic 'Travel' exposure, part of 'Mortgages and Real Estate' exposure arises from supply linkages. The maximally exposed firm is in the Residential Building Construction industry but other highly exposed firms lie in upstream manufacturing sectors (e.g. Building Material and Supplies Dealers; Other Wood Product Manufacturing). In the case of St. Joe, its NAICS industry is Traveler Accommodation but its business activities include (italics ours) "developing *residential*, hospitality, and commercial projects that meet growing market demands." Thus, its activities cut across traditional sectors and include a variety of real estate projects including housing. Finally, the 'Oil and Gas' exposure is not concentrated in any particular industry but encompasses many different firms involved in the extraction, transportation, and refinement of fossil fuels. As an illustration of the broad range of exposed activities, Heritage-Crystal Clean Inc (5622, 'Waste Treatment and Disposal') has an Oil Business segment which "consists of used oil collection activities, re-refining activities, oil filter removal and disposal services, and the sale of recycled fuel oil."

These case studies emphasize that our firm-level exposures reflect individual and highly specific firm characteristics. Each firm in our data is a bundle of specific exposure components that reflect its unique organization. In the remainder of the paper, we quantify how these account for the firm-level impact of aggregate shocks.

4 Shock Exposures and Asset Regressions

A long tradition in finance uses firm-level text data to explain asset behavior. For example, Kogan et al. (2009) and Boudoukh et al. (2019) use 10K filings and media articles, respectively, to predict volatility in individual stocks. More recently, Ke et al. (2020) and Lopez-Lira (2023) use media articles and RF text, respectively, to predict returns. These papers find that text contains information relevant for predicting firm-level stock performance, and that machine-learning models are useful for extracting this information.

Our goal here is different. First, we wish to assess *RF* text's ability to predict returns on the small subset of overall trading days with large aggregate market moves. Our view is that, on these jump dates, firm-level variation in abnormal returns is driven by a common news shock to which firms are heterogeneously exposed due to different business practices. On regular trading days, by contrast, idiosyncratic shocks presumably play a larger role, which may have a different statistical relationship with *RF* language. Second, we do not necessarily seek an algorithm or dataset that maximizes the ability to predict returns. Instead, we simply examine which part of total return variation arises from variation in the structure of firms as captured by our text exposures. We later examine to what extent this part, which in principle can be small or large, captures subsequent shifts in real outcomes across firms.

Our main finding is that *RF* text contributes a substantial amount to jump-date return predictability that is not accounted for by standard controls or by well-known existing measures of firm heterogeneity informed by text. At the same time, the majority of return variation remains unexplained.

4.1 Return predictability and Risk Factor language

MNIR models language as a function of firm observables, including abnormal returns. Instead, to assess return predictability, we need to formulate a "forward regression" of abnormal returns on language. Above we discussed that our text exposures satisfy the conditional independence relationship $\text{AbnRet}_i^d \perp \mathbf{c}_i \mid z_i^d, \mathbf{controls}_i^{y(d)}, C_i$. This implies that a forward regression need only model AbnRet_i^d as a function of $z_i^d, \mathbf{controls}_i^{y(d)}, C_i$ but does not specify the exact function. We take the simplest approach and choose a linear regression.

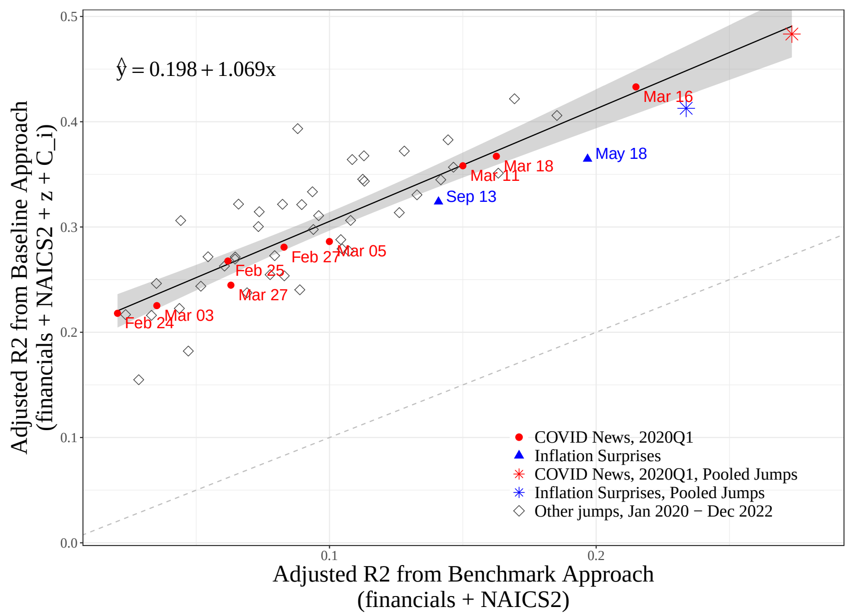
To assess the contribution of text to accounting for asset return variation, we estimate the following two models for each jump date in our sample:

$$\text{AbnRet}_i{}^d = \alpha_0^d + \left(\boldsymbol{\alpha}_1^d\right)^T \mathbf{controls}_i^{y(d)} + \epsilon_i{}^d \tag{3}$$

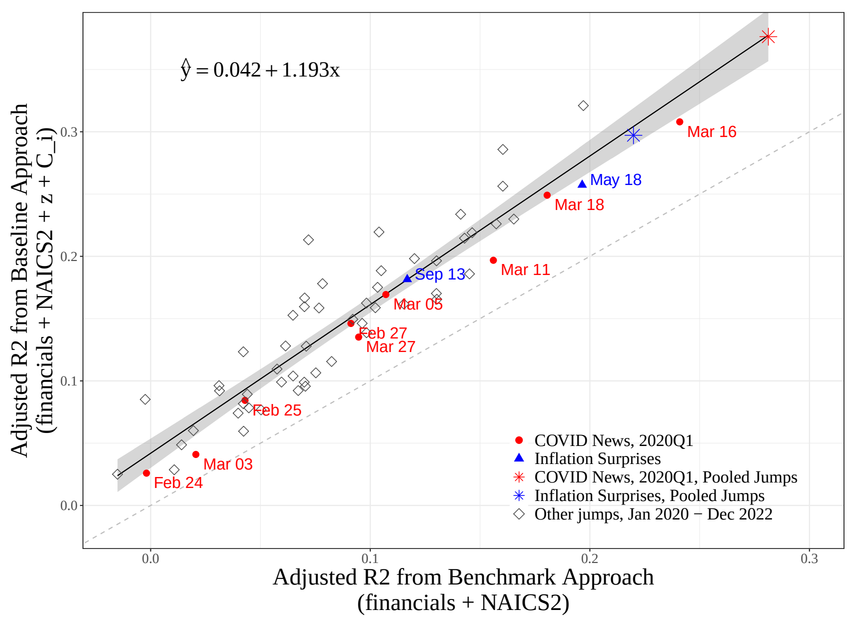
$$\text{AbnRet}_i{}^d = \gamma_0^d + \left(\boldsymbol{\gamma}_1^d\right)^T \mathbf{controls}_i^{y(d)} + \gamma_2^d z_i{}^d + \gamma_3^d C_i + \varepsilon_i{}^d \tag{4}$$

The vector of controls is the same as we use in the inverse regression model (2). (3) explains abnormal returns just with these controls. (4) additionally adds our corresponding text exposure $z_i{}^d$ and a control for the length of firm i's *RF* text (measured by word count C_i). We record the (adjusted) R^2 from both regression models and plot them in Figure 3a.

The top panel of Figure 3 shows a nearly uniform 20pp increase in R^2 from including the text exposures. We also include separate models for the average returns on pandemic and



(a) Within-Sample



(b) Out-of-Sample

Figure 3: Improvement in Goodness-of-Fit from Including Text in Asset Regressions. These figures report the adjusted R^2 from regressing abnormal returns from jump dates on controls for NAICS2 sector, leverage, and size (x-axis) and additionally including text-based shock exposures and controls for text length (y-axis). The top panel uses the entire sample, while the bottom panel reports the adjusted R^2 on held-out firm samples not used for the estimation of the MNIR coefficients.

inflation dates, respectively. Consistent with the intuition that averaging individual dates with common news reduces noise, the R^2 achieved by both models is higher for the average returns than for any of the individual dates that compose the average. Nevertheless, the increase in R^2 from including text is of similar magnitude as for individual jump dates: the R^2 from (3) for the pandemic (inflation) average is 0.27 (0.24) and from (4) is 0.48 (0.42).

To gauge out-of-sample goodness-of-fit, we examine model performance for each jump date on firms whose RF language does not inform the estimation of the text exposures. To begin, we randomly split the total sample of firms into ten groups, stratified by NAICS2 sector, to perform ten-fold cross validation. Then, for each jump date, and for each split of test data (one of the groups) and training data (the remaining groups), we (i) fit the inverse regression model (2) on the training data; (ii) use the estimated MNIR coefficients to form text exposures for each firm in the test data; and (iii) fit (3) and (4) on the test data. The bottom panel of Figure 3 plots the average R^2 across splits for each jump date. The addition of text increases R^2 for each jump date although, as expected, the gain is more muted than within sample—around 7pp on average. We again find that averaging returns across pandemic and inflation dates improves performance compared to all individual dates in each respective group. For the pandemic average, the inclusion of text shifts R^2 from 0.27 to 0.35. For the inflation average, the shift is from 0.23 to 0.30. Whether the reduction in the performance of the model with text is purely the result of spurious correlation is not clear. The relationship between terms and returns in the full sample may reflect highly specific business characteristics adopted by few firms. By their nature, these relationships would not generalize well out-of-sample.

	(1) Second Wave Fears (2020/06/11)	(2) Vaccine News (2020/05/18)	(3) Slower Fed Tightening (2022/11/10)
Exposure Variable	z_i^P	z_i^P	z_i^I
Coefficient	0.51***	-0.46***	-0.23***
	(0.032)	(0.060)	(0.026)
Observations	1,552	1,552	1,506
Adjusted R^2	0.349	0.339	0.116
Adjusted R^2 (only controls)	0.199	0.207	0.069

Table 2: Generalization of Text Exposures to Out-of-Sample Dates.
We fit the forward regression (4) on selected dates using z_i^P or, alternatively, z_i^I as the text exposure. The dates we select are not used to build the text exposures.

As a second out-of-sample test, we next consider dates—rather than firms—outside the training sample. The first two are part of the Later COVID News group, which allows us to test whether pandemic exposure from early in the pandemic as captured by z_i^P is informative for returns driven by related news later in time. The first such date is June 11, 2020, which has the largest negative market return (-6.07%) of any jump date after 2020Q1. Newspapers attribute this to the arrival of news that a second wave of COVID might develop. The second is May 18, 2020 (3.33%) when Moderna announced that early-stage clinical trials indicated the effectiveness of its COVID vaccine. We fit (4) for both dates, but use z_i^P as the text exposure which is built solely from dates in the COVID News, 2020Q1 category. Recall that firms with a higher z_i^P are those that perform relatively better, due to text-based fundamentals, in response to negative pandemic news. Since negative pandemic news arrived on June 11, the coefficient on z_i^P should be positive. By contrast, since the vaccine represents positive pandemic news, one would expect a negative coefficient on z_i^P for May 18. Table 2 shows both results hold. Moreover, the improvement in R^2 from including the pandemic exposure relative to only controls on these out-of-sample dates is substantial: 13pp on May 18, and 15pp on June 11.

For an out-of-sample date related to inflation, we choose November 10, 2022, when news arrived indicating a slowing path of interest rate hikes due to the easing of inflation pressures—the opposite of the news used to construct our inflation exposure. This has the highest absolute market return (5.63%) of any 2022 jump date and is part of the Taylor Rule group. As expected, Table 2 shows that our inflation exposure is significantly related to lower returns on this date. Recall that, by construction, higher (lower) inflation exposure reflects firms that—due to text-based fundamentals—perform better (worse) on days with news of rising inflation. A negative coefficient is therefore expected when the news signals easing inflation. The inclusion of the inflation exposure nearly doubles the R^2 relative to only controls.

Finally, note that in no case does model (4) achieve an R^2 above 0.5. That is, the majority of variation in jump-date returns remains unexplained even after including text exposures. This may be for several reasons. The bag-of-words model may be too simple to capture the relevant information in RF text for understanding how firms react to news shocks. Alternatively, RF text simply may not convey all the relevant information for how macro shocks transmit to firms. Finally, unexplained asset price variation may be unrelated to fundamentals, for example because of the presence of noise traders (Grossman and Stiglitz 1980, Milgrom and Stokey 1982). We return to this discussion in the following section.

4.2 Comparison to existing measures

Previous work has used 10-K text to capture firm-level characteristics that are difficult to measure with traditional data. One well-known measure is the text-based industry codes of Hoberg and Phillips (2016) that group firms according to the similarity of their 10-K product descriptions. To examine whether our shock exposures capture information beyond these, we replace NAICS2 fixed effects in (3) and (4) with text-based HP industry codes at the 50-category level of granularity. Figure C.2 (Online) compares adjusted R^2 and shows the increase in R^2 is comparable to the baseline setting. We conclude that the text exposures capture firm heterogeneity beyond the text-based HP industry codes.

Baker et al. (2025a) builds dictionaries to capture multiple dimensions of economic and policy risks, and applies them to firms' 10-K filings to account for sources of return volatility. As a further exercise, we include in $\mathbf{controls}_i^{y(d)}$ the firm-level counts of terms in RF texts from these dictionaries. Figure C.3 (Online) shows that the inclusion of these additional text-based controls has little impact on the increase in R^2 from including our text exposures.

One might expect our exposures to capture information beyond these existing measures because they are specific to particular types of shock. In contrast, the representation of firms in Hoberg and Phillips (2016) and Baker et al. (2025a) does not adjust to the nature of macro shocks to which firms are exposed.

4.3 Returns and specific exposures

To examine the effect of separate exposure components' impact on asset prices, we expand the forward regression model (4) as

$$\text{AbnRet}_i^d = \gamma_{0,j}^d + \left(\boldsymbol{\gamma}_{1,j}^d\right)^T \mathbf{controls}_i^{y(d)} + \gamma_{2,j}^d z_{i,j}^d + \gamma_{3,j}^d z_{i,-j}^d + \gamma_{4,j}^d C_i + \varepsilon_{i,j}^d \tag{5}$$

which we fit for all for $d \in \{P, I\}$ and j, where j denotes each specific component of our P and I exposures. $z_{i,j}^d$ is the analog of our exposure measure to macro shock d, but based solely on the subset of terms linked to its component j—as detailed in Section 3.3. $z_{i,-j}^d$ is the unexplained part of the overall exposure z_i^d from a regression onto $z_{i,j}^d$, and so controls for overall shock exposure orthogonal to component j. Tables A.4 and A.5 reports point estimates and standard errors of $\gamma_{2,j}^d$ for each of the 28 (30) pandemic (inflation) components. In all but one case, we find significant coefficients that have the expected sign. Thus, individual components, even when composed of relatively few terms, have an effect on returns that is independent from the variation in z_i^d arising from all other terms.

5 Shock Exposures and Real Outcomes

We now study how macroeconomic shocks transmit to firm outcomes in the real economy using alternative exposure measures. Our key finding is that—across a variety of outcomes—text exposures capture nearly all information in jump-date abnormal returns related to future firm performance beyond standard controls. This is despite text exposures' accounting for less than half of total variation in these returns. We find large heterogeneity induced by both pandemic and inflation shocks that is significantly understated using returns alone as an exposure measure.

5.1 Baseline regression model

Our main measure of firm-level real outcomes is quarterly revenue growth. Let $\text{rev}_{i,t}$ denote firm i revenue in quarter t. As discussed in Section 2 we construct a balanced sample of firms with observed revenue from 2015Q1 through 2022Q4. We compute the growth rate as

$$\Delta\text{rev}_{it} = \log(\text{rev}_{i,t}) - \log(\text{rev}_{i,t-12})$$

The use of twelve-quarter growth rates ensures that growth in each post-pandemic quarter in our sample is computed with respect to a pre-pandemic quarter. As we show in Section 6, the within-period standard deviation in this outcome is highly variable over time. For this reason, we standardize the outcome to have unit standard deviation and zero mean in each quarter to better compare the magnitudes of the coefficients in different quarters. We also trim the bottom and top 0.1% of observations according to Δrev_{it}.

Our main regression specification is

$$\Delta\text{rev}_{it} = I_i + I_{s(i),t} + \sum_{t=19Q1}^{22Q4} I_t \boldsymbol{\alpha}_t' \mathbf{e}_i + \sum_{t=18Q1}^{18Q4} I_t \boldsymbol{\beta}' \textbf{controls}_{it} + \sum_{t=19Q1}^{22Q4} I_t \boldsymbol{\beta}_t' \textbf{controls}_{it} + \epsilon_{it} \quad (6)$$

I_i is a firm fixed effect. $I_{s(i),t}$ is a NAICS2 $\times$ quarter fixed effect. $\mathbf{e}_i$ is a firm-level vector of shock exposure measures, which are held constant over time. $\textbf{controls}_{it}$ includes leverage and log assets in period $t-12$. Each continuous variable is standardized to have unit variance and zero mean by quarter. Because the coefficient on controls is constant during a baseline period (2018Q1 through 2018Q4), the $\boldsymbol{\beta}_t$ coefficients capture whether the association between controls and revenue growth shifts in quarter t relative to the baseline. Errors are clustered by period and by sector.

The primary coefficients of interest are $\boldsymbol{\alpha}_t$ which describe the relationship between shock exposures and real outcomes in period t. Since the model includes firm fixed effects, $\boldsymbol{\alpha}_t$

reflects whether an increase in shock exposure induces a deviation of a firm's period-t growth rate from its baseline growth rate from 2018Q1 through 2018Q4. This specification is similar to a standard local projection, although our shock exposures are not time-varying.

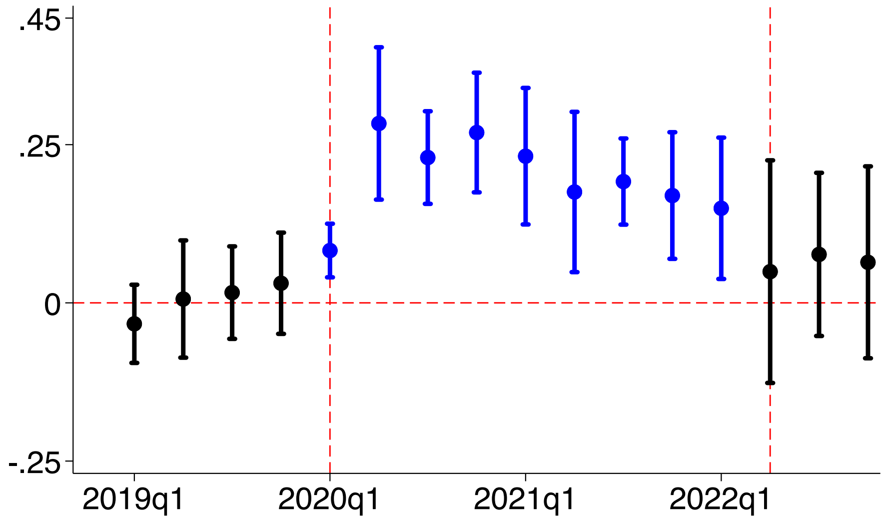
Our baseline exposure vector includes the pandemic and inflation exposures, as these capture business characteristics we conjecture should help determine how firm outcomes react to macro shocks. However, we also include abnormal return residuals since, as shown in Section 4, these have substantial variation that might also contain information relevant for real outcomes. More concretely,

$$\mathbf{e}_i = \left(z_i^P, \hat{\varepsilon}_i^P, z_i^I, \hat{\varepsilon}_i^I\right) \tag{EXP-TEXT}$$

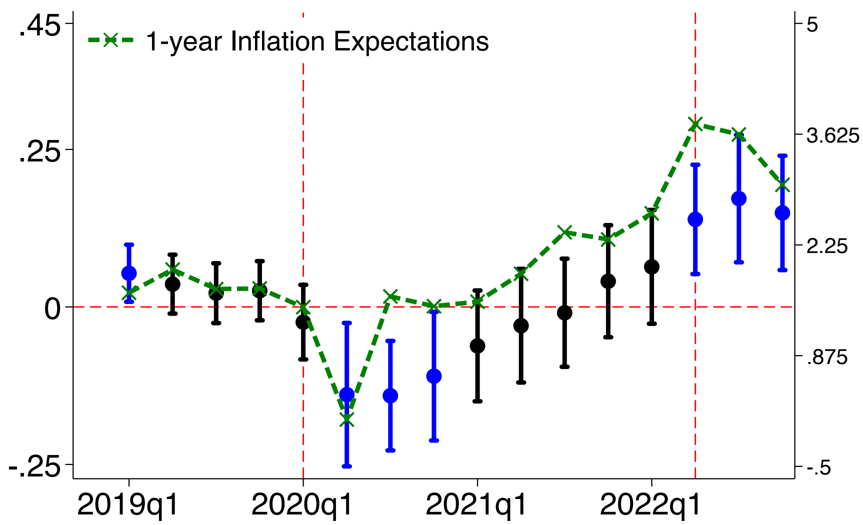
where $\hat{\varepsilon}_i^P$ and $\hat{\varepsilon}_i^I$ are the residuals from forward regression (4) estimated on pandemic and inflation abnormal returns, respectively.

Figure 4 displays the estimated effects of the text-based exposures. Differences in pandemic exposure are not associated with different pre-2020 growth trends as seen by the small and insignificant coefficients in 2019. Exposure becomes significant, albeit small, in 2020Q1 before growing large in 2020Q2, the quarter with highest pandemic-related disruption. The effect of a one-SD increase in pandemic exposure in 2020Q2 shifts growth rates by 0.3 SD. The interpretation is that firms that outperformed in the stock market on days with bad COVID news in 2020Q1 also saw relatively stronger revenue growth in 2020Q2. Pandemic exposure remains large and significant throughout 2020 and 2021. That is, beyond its negative and large aggregate impact, the pandemic had distributional effects across firms that persisted for nearly two years.

The inflation text exposure generates positive, significant effects from 2022Q2 onwards. While the magnitudes are not as sizeable as during the pandemic (which is not surprising given its historic magnitude) they are not trivial: a one-SD increase in inflation exposure generates a larger than 0.1-SD effect on revenue growth outcomes in each of the final three quarters in the sample. Interestingly, inflation exposure also generates negative effects in the three quarters following the arrival of COVID-19. To interpret these effects, Figure 4b superimposes one-year-ahead inflation expectations on the estimated coefficients. The estimated coefficients closely track the deviation of inflation expectations from their pre-pandemic level. Recall that our inflation exposure is built to explain movements in abnormal returns on 2022 dates during which inflation news arrived. But this construction will also proxy exposure to inflation news shocks throughout our sample. As such, the $\boldsymbol{\alpha}_t$ coefficients should be interpreted as the firm-level impact of inflation-related shocks that have arrived from 2019Q1 through period t. If our inflation exposure correctly captures this, firms that



(a) Effect of Pandemic Exposure



(b) Effect of Inflation Exposure

Figure 4: Effect of Shock Exposure (Text) on Quarterly Revenue Growth
We fit the panel regression model (6) using (EXP-TEXT) as the vector of exposures. This yields four coefficients in $\boldsymbol{\alpha}_t$ for each t. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals. The bottom figure overlays the log of one-year-ahead inflation expectations onto the estimated impact of text-based inflation exposure. Inflation expectations are measured using the Federal Reserve Bank of Cleveland's model available from FRED at `https://fred.stlouisfed.org/series/EXPINF1YR`.

benefit (in relative terms) from news suggesting rising inflation in 2022 should suffer from shocks pointing to reduced inflationary pressures in early 2020. Consistent with this, we see a negative coefficient on inflation exposure in 2020 when inflation expectations drop.

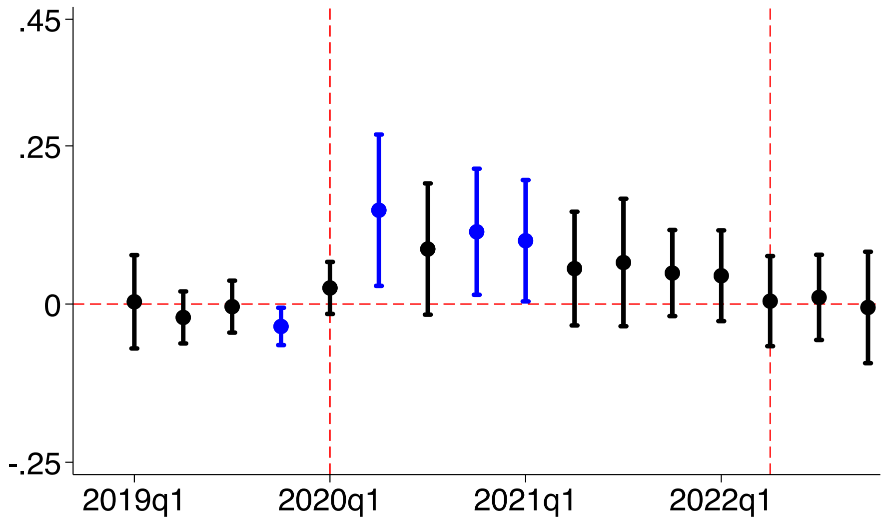
Figure A.3 instead displays the estimated effects of return residuals. Remarkably, they have no significant effect from 2020 onwards—even when they account for more than half of return variation. One interpretation is that returns combine signals of future firm outcomes and noise, and text exposures extract the former. The results also show there are limited gains to adopting a more complex text algorithm. It is reasonable to assume that a fine-tuned Transformer model (see Ash et al. 2025 for a recent review) could achieve higher R^2 in predictive regressions for asset movements. But this would simply better fit the residual which is uninformative about real outcomes. In fact, such an approach might lead to a *worse* text-based exposure insofar as it would incorporate more noise and less signal. A more complex model would also imply a loss of interpretability. In short, our results show the value of simpler text algorithms that use a basic bag-of-words representation.

5.1.1 Abnormal return exposures

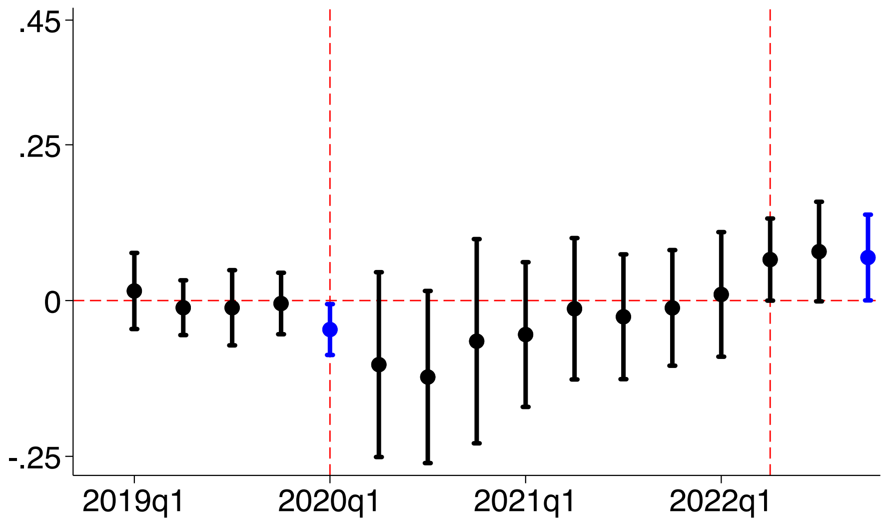
A large literature in macroeconomics uses asset prices movements during time windows where news arrives to build shock measures (Ramey 2016). The most common setting for the application of this method is to bond price movements around monetary policy announcements (Kuttner 2001, Gürkaynak et al. 2005), but more recently Cirelli and Gertler (2022) uses a measure closely related to AbnRet_i^P for pandemic exposure. As such, we re-estimate (6) with the exposure vector

$$\mathbf{e}_i = \left(\text{AbnRet}_i^P, \text{AbnRet}_i^I\right). \tag{EXP-RET}$$

Figure 5 shows the estimated $\boldsymbol{\alpha}_t$ coefficients. The patterns are qualitatively similar to those for the text exposures, but smaller in magnitude and less significant. This is again consistent with the view that abnormal returns combine signal and noise. Under this interpretation, including return exposures without projecting onto text attenuates coefficient estimates due to noise. These findings are potentially relevant for the wider literature that uses asset price movements to measure the impact of macro shocks. Projecting them onto a rich enough dataset that spans what should be fundamental drivers of return variation is a potentially promising method to overcome measurement error.



(a) Effect of Pandemic Exposure



(b) Effect of Inflation Exposure

Figure 5: Effect of Shock Exposure (Returns) on Quarterly Revenue Growth.
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (6) when using exposure measures (EXP-RET) derived from abnormal returns on pandemic and inflation jump dates. Point estimates displayed with 95% confidence intervals.

5.1.2 Early pandemic earnings call exposures

As discussed in the Introduction, Hassan et al. (2023) builds firm-level measures of COVID impact from 2020Q1 to 2022Q1 using mentions of the pandemic in earnings call transcripts. The 2020Q1 variables are constructed in the same quarter as the jump dates that inform our pandemic exposure z_i^P, so we can directly compare their ability to account for later real outcomes. To do so, we re-estimate the baseline regression using as exposures z_i^P, z_i^I, and the 2020Q1 exposure measures from earnings calls (positive sentiment, negative sentiment, risk, exposure). This reduces the sample of firms, but Figure C.4 (Online) shows the effects of our text exposures remain similar to those from the full sample. In contrast, Figure C.5 (Online) shows that neither COVID sentiment variable from earnings calls predicts future revenue growth. The same is true of the risk and exposure variables (results not reported). We conclude that the earnings-call-based exposures from 2020Q1 have limited forward-looking information for determining firm-level real outcomes.

5.2 Real effects of specific exposures

We hypothesize that text-based exposures explain real outcomes because they embody firm characteristics that affect market expectations about how firms will suffer or succeed due to macro shocks. In Section 3 we propose a method for describing the specific characteristics that drive the market reaction. Hence, a stricter test of our hypothesis is whether the exposure components that most strongly drive asset price responses also most strongly drive real-outcome heterogeneity. Recall that $z_{i,j}^d$ is exposure component j for firm i on jump date $d \in \{P, I\}$. We estimate (6) for each such component using as an exposure vector

$$\mathbf{e}_{i,j}^d = \left(z_{i,j}^d, z_{i,-j}^d, \hat{\varepsilon}_i^{\,d}, z_i^{-d}, \hat{\varepsilon}_i^{\,-d}\right) \tag{EXP-COMP-j}$$

$z_{i,-j}^d$ is the part of the total exposure z_i^d unexplained by a simple regression onto $z_{i,j}^d$. Note that in asset regression (5) we similarly include exposure component j and the residual of the total exposure after projecting onto $z_{i,j}^d$. $-d$ refers to the alternative in $\{P, I\}$.

Figure A.4 compares the strength of each component in the abnormal returns regressions with its strength in the real outcomes regression. For both pandemic and inflation exposures, we observe a positive relationship between the two. Moreover, there are no significant effects in the real outcomes model whose sign does not align with that of the returns model. Overall, then, we find evidence to support the view that text-based exposures explain real outcomes because they capture interpretable firm characteristics that interact with macro shocks to shift expected firm outcomes.

We can also use this exercise to isolate particularly strong exposure components. The pandemic exposure with the strongest negative (positive) effect on real outcomes is exposure 11 (exposure 16). These are the exposures whose distributions across firms are in Figure A.1 which we discuss in Section 3.3. Recall that exposure 11 reflects Travel and exposure 16 reflects IT Environment. The strongest negative and positive exposures on real outcomes for the inflation shock are 2 (Housing and Mortgages) and 19 (Oil and Gas), respectively, whose distributions are visualized in Figure A.2.

5.3 Other real outcomes

As discussed in Section 2, we also construct annual outcomes for revenue, employment, and investment. Denoting an outcome generically by y_{it} where t now references a year, we modify (6) slightly and estimate

$$\Delta y_{it} = I_i + I_{s(i),t} + \sum_{t=2018}^{2022} I_t \boldsymbol{\alpha}_t' \mathbf{e}_i + \sum_{t=2010}^{2017} I_t \boldsymbol{\beta}' \mathbf{controls}_{it} + \sum_{t=2018}^{2022} I_t \boldsymbol{\beta}_t' \mathbf{controls}_{it} + \epsilon_{it} \quad (7)$$

where $\Delta y_{it} = y_{it} - y_{i,t-3}$. The main difference with respect to (6) is we now use eight periods for a baseline (2010 through 2017) rather than four (2018Q1 through 2018Q4). Because annual outcomes are potentially noisier, doubling the number of baseline periods better estimates the pre-existing, firm-level growth trends. All annual outcomes are estimated with (EXP-TEXT) as the exposure vector.

Figure C.6 (Online) (C.7 (Online)) shows the effect of text (residual) exposure for annual revenue growth. The text exposures again capture the entire effect of post-2019 shock exposure embodied in jump-date abnormal returns. Figures C.8 (Online) and C.9 (Online) show analogous results for annual employment growth, and figures C.10 (Online) and C.11 (Online) for investment. Pandemic exposure is associated with significant employment responses in all post-pandemic years, and a significant investment response in 2020 and 2021. Inflation exposure is associated with a significant investment response in 2021. This provides evidence that macro shock exposure changes firm decisions, not just firm outcomes. For neither employment nor investment do residuals have significant effects.

Finally, we consider how the exposures explain earnings surprises. Table C.4 (Online) presents cross-sectional regression results. We again find that the text exposures correlate significantly with outcomes, and that residual exposures do not. Moreover, there is a notable R^2 improvement from introducing the text exposures instead of returns.

6 Shock Exposures and Cyclicality in Cross-Firm Dispersion

We now return to the central question of how much heterogeneous exposure to macro shocks can account for countercyclical increases in firm-level dispersion. The black curve in Figure 6 plots the weighted standard deviation of Δrev_{it} from 2019Q1 through 2022Q4, which displays a large increase during the 2020 recession. We then plot the weighted standard deviation in fitted growth rates from estimating (6) with different regressors. First, we consider the model without any exposure vector $\mathbf{e}_i$. Then we introduce, in succession, progressively enriched exposure vectors: (i) the pandemic exposure z_i^P; (ii) the inflation exposure z_i^I; (iii) exposures to the additional seven labeled shocks from Figure 1; (iv) exposures to shocks on each individual date that makes up the seven additional exposures; and (v) residualized returns for these individual exposure dates. Figure 6 shows the results. We include the individual dates that make up the seven additional exposures to account for potentially heterogeneity in triggering news within broad categories.

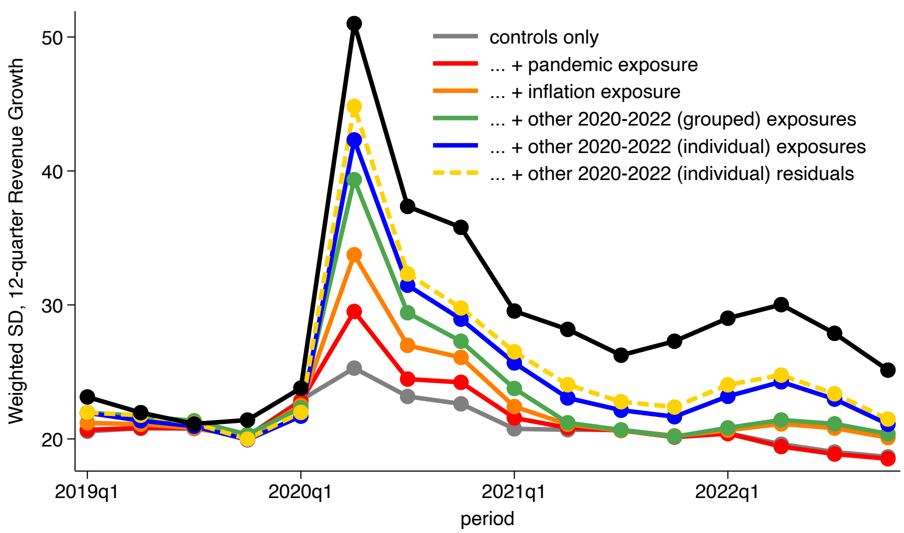


Figure 6: Dispersion in Firm-Level Revenue Growth

The black curve plots the standard deviation by quarter of the outcome variable for (6), i.e. twelve-quarter revenue growth, weighted by firm revenue in quarter $t-12$. The rest of the curves plot the weighted standard deviation in fitted values from alternative panel regression models that introduce variable in succession. The first model fits (6) without any exposure vector $\mathbf{e}_i$. The second model introduces z_i^P as an exposure meausure. The third further introduces z_i^I as an exposure. The fourth further introduces the other seven exposures in Figure 1. The fifth introduces all individual dates that make up these seven exposures separately. The final model also introduces abnormal return residuals computed from fitting (4) for these individual dates.

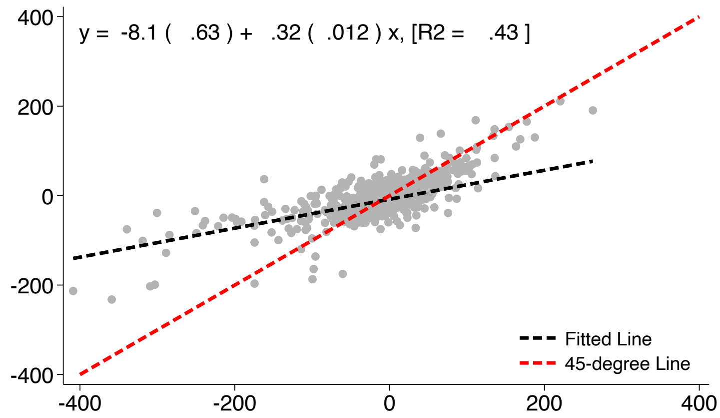
The model without text exposures has a limited ability to capture the enormous increase in dispersion in 2020Q2 despite controlling for time-specific sectoral shocks and heterogeneity related to time-varying firm size and leverage. In contrast, the model with text exposures greatly improves the ability to explain dispersion in firm-level revenue growth from 2020Q2 onwards. Interestingly, the contribution of the pandemic exposure to explaining dispersion beyond the non-text model is largely zero by 2021. In contrast, there are two episodes during which the inflation exposure contributes beyond the non-text model: during the initial quarters of the pandemic and during late 2022 when inflation expectations peak in the sample (see Figure 4b). The additional seven shock categories contribute a substantial amount of explanatory power for dispersion beyond the pandemic and inflation, showing that a variety of distinct macro shocks matter for dispersion. Finally, consistent with the results of Section 5, residual exposure is unimportant for explaining dispersion.

We next examine the fitted values for revenue growth from 20202Q2—the period with the largest drop in output during the pandemic recession—in more detail. Figure 7 compares actual vs predicted firm-level revenue growth for the model with no exposures and for model (iv) that includes pandemic and inflation exposures, as well as jump-date exposures for days that form other categories. The model with the rich set of exposures accurately recovers the full distribution of firm-level revenue growth in this quarter, achieving an R^2 of 0.82. In other words, heterogeneous exposure to macro shocks suffices to explain the bulk of firm dispersion at the onset of the recession. We emphasize again that the explanatory power in these regressions arises from RF texts that predate 2020 and jump-date return variation rather than idiosyncratic firm-level shocks realized during the recession.

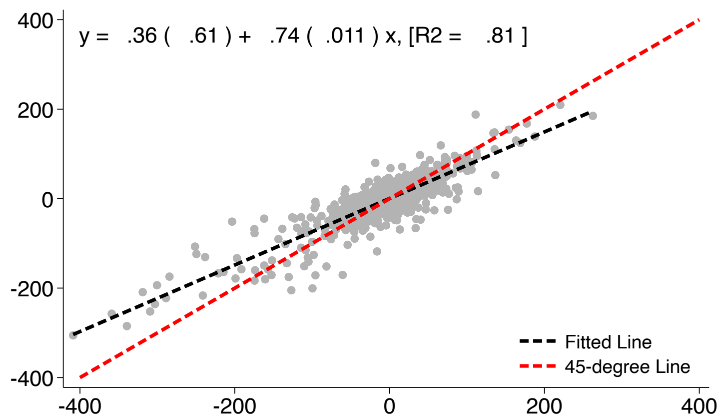
6.1 Historical evidence

These results establish that macro shock exposure drives much of the countercyclical dispersion from 2019-2022. A natural question is whether this result is unique to the sample period or more general. 10-K Risk Factors in their current form were only mandated from 2006. Many firms' filings were not fully developed by 2007 at onset of the Great Financial Crisis, the triggering event for the largest contraction prior to COVID-19. We can nevertheless explore whether the other data sources we use are consistent with countercyclical dispersion arising from macro shock exposure.

A first result is that jump date arrivals are themselves countercyclical. Figure C.12 (Online) plots the frequency of jump dates by month since 1900, with separate distributions for recessions and expansions. Jump dates are over three times more frequent in recessions. Moreover, Figure C.13 (Online) shows that—when they do arrive—jump dates tend to be



(a) Predictions from Controls Only



(b) Predictions from All Text Exposures

Figure 7: Actual (x-axis) vs. Predicted (y-axis) 2020Q2 Revenue Growth
The top figure plots the predicted values for 2020Q2 revenue growth from the panel regression (6) without any exposure vector $\mathbf{e}_i$. The bottom figure plots the predicted values when $\mathbf{e}_i$ consists of the pandemic exposure, the inflation exposure, and exposures from days that make up the seven additional classified categories in Figure 1.

larger in recessions than expansions.

Second, Figure C.14 (Online) replicates Figure 1 for all jump dates from 2000-2022. One continues to observe a positive relationship between the absolute size of market returns on jump dates and firm-level return dispersion, with a somewhat larger relationship for negative jumps. Hence, markets expect substantial firm-level heterogeneity in response to macro shocks for a much longer sample period.

Third, while we cannot build our text exposures over a long time period, we can use simple abnormal returns on jump dates as exposure measures. While these are noisy indicators of shock exposure, they are useful to reveal if the relationship between returns and real outcomes

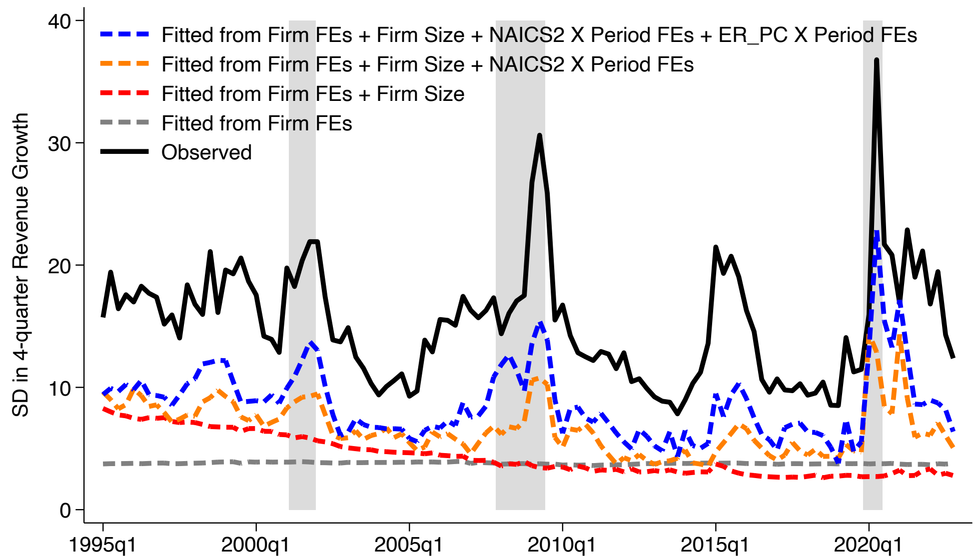


Figure 8: Dispersion in Firm-Level Revenue Growth since 1995

We define four-quarter revenue growth to be $\log(\text{rev}_{i,t}) - \log(\text{rev}_{i,t-4})$ and compute this value for a balanced sample of 950 firms from 1995Q1 through 2022Q4. The black curve plots the standard deviation in this variable by quarter, weighted by firm revenue in period $t-4$. The rest of the curves plot the weighted standard deviation in fitted values from alternative panel regression models that introduce variable in succession. The first model uses only firm fixed effects. The second model introduces firm size as measured by log assets interacted with quarter fixed effects. The third model introduces sector by quarter dummy variables. The fourth introduces loadings onto the first 10 principal components of jump-date returns over the sample period.

is systematically different during 2019-2022 versus prior recessions. Figure 8 shows dispersion in four-quarter revenue growth from 1995 through 2022 for a balanced panel of 950 firms whose quarterly revenue data is available over the entire period in COMPUSTAT. The black curve shows the standard deviation in growth rates. As expected, dispersion rises during recessions, marked by the shaded regions. The most recent, pandemic-induced recession shows particularly large dispersion.

As in Figure 6, we then introduce the dispersion in fitted values from models with additional variables. First, we regress growth rates on firm fixed effects which predicts little dispersion, meaning that individual firms experience substantially different growth rates over time. Next, we add firm size interacted with quarter fixed effects.[17] Time-series variation in firm size does not account for the cyclical pattern of dispersion. We then add NAICS2 by quarter fixed effects, which absorb sector- and time-specific shocks. This specification

[17]We do not include leverage controls because the number of firms would fall below 600 under the requirement of a balanced sample. Nevertheless, we have repeated the historical analysis on this smaller subsample and found similar patterns to those in Figure 8.

produces fitted values which rise somewhat during recessions.

Finally, we include jump-date returns over the sample period. Due to the large number of dates, incorporating them all jointly leads to a high-dimensional regression problem. We address this by performing a principal components decomposition over time and introducing an exposure vector $\mathbf{e}_i$ for each firm, which comprises its loadings on the first ten components. These components collectively explain 48.1% of the variation in jump-date returns.

In all three recessions, fitted dispersion rises with respect to the model with sector by time fixed effects after including information on jump-date returns. The maximum observed dispersion in the first recession is in 2001Q4 and takes a value 21.9. The dispersion predicted by the model with return exposures is 13.7. Thus, the ratio of explained to actual dispersion is 0.63. The equivalent ratio for 2009Q2—the quarter with maximum dispersion in the second recession—is 0.50 and for 2020Q2 is 0.62. In this respect, the pandemic recession is not unique in its relationship between jump-date returns and growth-rate dispersion. In contrast, the rise in growth-rate dispersion in 2015 which is not associated with a recession has a much more limited amount of explained dispersion from the returns model. Our return exposures are thus not mechanically linked to rising dispersion in the real economy.

The historical return data also allows us to ask whether firms with high exposure to macro shocks in one recession are highly exposed to macro shocks in others. In line with the NBER database on recession start and end dates, we compute three sets of average jump-date returns: (i) April 2001 through November 2001; (ii) January 2008 through June 2009; and (iii) March 2020 through April 2020. These capture, respectively, exposure to macro shocks during the dot-com recession, the Great Financial Crisis, and the pandemic recession. The correlation between (i) and (ii) is 0.0072, between (i) and (iii) is 0.1341, and between (ii) and (iii) is 0.1300. Hence, each recession involves a distinct set of macro shocks that, at least in terms of their effects across firms, differ substantially from those in other downturns.

7 Conclusion

We develop a method for measuring firm-level exposures to macro shocks. Our method works by combining abnormal returns on stock market jump dates with rich, firm-specific descriptions of business risk characteristics. We apply the method to outcomes from 2020 to 2022, a period characterized by great volatility in market-level returns and extraordinary dispersion in firm-level returns.

Our exposure measures explain firm-level abnormal returns through interpretable variation in language. They also explain most of the increased dispersion in firm-level revenue growth in the wake of the pandemic and much of the dispersion in employment growth, investment rates, and earnings surprises. More broadly, our evidence supports a novel interpretation of countercyclical dispersion, highlighting the role of heterogeneous business characteristics in macro shock transmission.

Our findings have potentially important implications for research and policy. Quantitative macro models with firm dynamics typically don't feature the mechanism for generating heterogeneity that we highlight. Showing its empirical relevance is an important first step towards better understanding how recessions, and macro shocks more generally, reshape economic activity. Because firm-level exposure patterns differ across macro shocks according to our results, the cross-sectional impact of recessions will also vary with the nature of the underlying macro shock. We show that the cross-sectional response pattern is predictable, conditional on the shock, and that response patterns differ in predictable ways across shocks. This type of information is useful in forecasting the effects of a given shock and, potentially, in designing targeted interventions to mute the effects of a shock or facilitate adjustments to it. More generally, our approach integrates text into a quantitative framework for shock identification and characterization, thereby helping move text analysis beyond description and towards its use in causal inference for macroeconomics.

References

Ash, E. and Hansen, S. (2023). Text Algorithms in Economics. *Annual Review of Economics*, 15(1):659–688.

Ash, E., Hansen, S., Marangon, C., and Muvdi, Y. (2025). Large language models in economics. In *Handbook of Economics and Language*, Palgrave Handbooks.

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., and Viratyosin, T. (2020). The Unprecedented Stock Market Reaction to COVID-19. *The Review of Asset Pricing Studies*, 10(4):742–758.

Baker, S. R., Bloom, N., Davis, S. J., and Kost, K. J. (2025a). Policy News and Stock Market Volatility. *Journal of Financial Economics*, forthcoming(w25720).

Baker, S. R., Bloom, N., Davis, S. J., and Sammon, M. C. (2025b). What Triggers Stock Market Jumps? Working Paper 28687, National Bureau of Economic Research.

Barrero, J. M., Bloom, N., and Davis, S. J. (2020). COVID-19 is Also a Reallocation Shock. *Brookings Papers on Economic Activity*, (Summer):329–371.

Barrero, J. M., Bloom, N., Davis, S. J., and Meyer, B. H. (2021). COVID-19 Is a Persistent Reallocation Shock. *AEA Papers and Proceedings*, 111:287–291.

Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M., and Stanton, C. (2020). The impact of COVID-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences*, 117(30):17656–17666.

Berger, D. and Vavra, J. (2019). Shocks versus Responsiveness: What Drives Time-Varying Dispersion? *Journal of Political Economy*, 127(5):2104–2142.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685.

Bloom, N., Fletcher, R. S., and Yeh, E. (2021). The Impact of COVID-19 on US Firms. Working Paper 28314, National Bureau of Economic Research.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2018). Really Uncertain Business Cycles. *Econometrica*, 86(3):1031–1065.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2019). Information, Trading, and Volatility: Evidence from Firm-Specific News. *The Review of Financial Studies*, 32(3):992–1033.

Cage, J., Le Pennec, C., and Mougin, E. (2021). Corporate Donations and Political Rhetoric: Evidence from a National Ban.

Cirelli, F. and Gertler, M. (2022). Economic Winners Versus Losers and the Unequal Pandemic Recession. Working Paper 29713, National Bureau of Economic Research.

Cirelli, F. and Gertler, M. (2024). Economic Winners versus Losers and the Unequal Pandemic Recession. *American Economic Journal: Macroeconomics*, forthcoming.

Cook, R. D. (2007). Fisher Lecture: Dimension Reduction in Regression. *Statistical Science*, 22(1):1–26.

David, J. M., Schmid, L., and Zeke, D. (2022). Risk-adjusted capital allocation and misallocation. *Journal of Financial Economics*, 145(3):684–705.

García, D., Hu, X., and Rohrer, M. (2023). The colour of finance words. *Journal of Financial Economics*, 147(3):525–549.

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

Gorodnichenko, Y., Pham, T., and Talavera, O. (2023). The Voice of Monetary Policy. *American Economic Review*, 113(2):548–584.

Grossman, S. J. and Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3):393–408.

Gürkaynak, R., Karasoy-Can, H. G., and Lee, S. S. (2022). Stock Market's Assessment of Monetary Policy Transmission: The Cash Flow Effect. *The Journal of Finance*, 77(4):2375–2421.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements. *International Journal of Central Banking*, 1(1).

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.

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133.

Hassan, T. A., Hollander, S., van Lent, L., Schwedeler, M., and Tahoun, A. (2023). Firm-Level Exposure to Epidemic Diseases: COVID-19, SARS, and H1N1. *The Review of Financial Studies*, 36(12):4919–4964.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects*. *The Quarterly Journal of Economics*, 134(4):2135–2202.

Hoberg, G. and Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies*, 23(10):3773–3811.

Hoberg, G. and Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Ilut, C., Kehrig, M., and Schneider, M. (2018). Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News. *Journal of Political Economy*, 126(5):2011–2071.

Justeson, J. S. and Katz, S. M. (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, 1(1):9–27.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2020). Predicting Returns with Text Data.

Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., and Smith, N. A. (2009). Predicting risk from financial reports with regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 272–280, USA. Association for Computational Linguistics.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

Langley, K. and Ostroff, C. (2022). Stocks Suffer Worst Day Since June 2020. *Wall Street Journal*.

Li, K., Mai, F., Shen, R., and Yan, X. (2021). Measuring Corporate Culture Using Machine Learning. *The Review of Financial Studies*, 34(7):3265–3315.

Lopez-Lira, A. (2023). Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns.

Marcos, C. M. (2022). Stocks resume their rout as falling profits reignite fears of inflation. *The New York Times*.

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

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

Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27.

Ottonello, P. and Winberry, T. (2020). Financial Heterogeneity and the Investment Channel of Monetary Policy. *Econometrica*, 88(6):2473–2502.

Papanikolaou, D. and Schmidt, L. D. W. (2022). Working Remotely and the Supply-Side Impact of COVID-19. *The Review of Asset Pricing Studies*, 12(1):53–111.

Ramey, V. A. (2016). Chapter 2 - Macroeconomic Shocks and Their Propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier.

Senga, T. (2018). A New Look at Uncertainty Shocks: Imperfect Information and Misallocation. Unpublished manuscript.

Stein, L. C. D. and Stone, E. (2013). The Effect of Uncertainty on Investment, Hiring, and R&D: Causal Evidence from Equity Options.

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

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

A Additional Figures and Tables

Category	Dates (MM/DD/YY)
COVID News, 2020Q1	02/24/20, 02/25/20, 02/27/20, 03/03/20, 03/05/20, 03/11/20, 03/16/20, 03/18/20, 03/27/20
Super Tuesday	03/04/20
Oil-Price Shocks	03/09/20
Public Health Policy	03/13/20, 04/17/20
Fiscal Policy	03/10/20, 03/23/20, 03/24/20, 03/26/20
Monetary Policy	03/02/20, 03/17/20, 04/22/22, 05/04/22, 05/05/22, 07/27/22, 08/26/22, 11/02/22, 11/30/22
Later COVID News	04/01/20, 04/06/20, 04/08/20, 04/14/20, 04/21/20, 05/18/20, 06/05/20, 06/11/20, 06/24/20, 10/28/20
Inflation Surprises	05/18/22, 09/13/22
Taylor Rule	05/27/22, 06/10/22, 10/07/22, 11/10/22
Other	03/12/20, 03/20/20, 03/30/20, 04/29/20, 05/01/20, 09/03/20, 09/08/20, 09/23/20, 02/25/21, 03/01/21, 03/07/22, 03/09/22, 03/16/22, 04/26/22, 04/29/22, 05/09/22, 05/13/22, 06/13/22, 06/16/22, 06/24/22, 07/19/22, 10/03/22, 10/04/22, 10/17/22, 12/15/22

Table A.1: 2020-2022 Jump Dates. Jump dates in 2020-2022 by source of triggering news. For all but the "Inflation Surprises" group, we use the categorizations of Baker et al. (2020) and Baker et al. (2025b). Full details on data and methodology are available at https://tinyurl.com/n5ptwsk2.

Bottom 30 Terms		Top 30 Terms	
Term	**Value**	**Term**	**Value**
hotel	-33549.80	game	19595.17
unitholder	-18019.33	product.candidate	18475.38
general.partner	-14038.83	client	17157.35
gaming	-11840.90	drug.candidate	12859.84
restaurant	-11830.94	clinical.trial	11963.56
reit	-10605.63	cellular	9029.67
tenant	-10177.66	subscription	7706.73
satellite	-9343.81	solution	7571.91
crude.oil	-9220.05	patient	6222.35
common.unit	-9087.59	student	6058.91
hotel.property	-8337.83	platform	5430.79
refinery	-7749.49	drug	5120.28
travel	-6931.05	collaborator	4641.59
franchisee	-6722.54	datum.center	4317.81
trs	-6657.09	celgene	4266.25
guest	-6097.13	fda	4185.00
airline	-5769.56	container	4182.65
our.common	-4922.40	commercialize	4106.13
refined.product	-4872.75	patent	3906.14
home	-4808.76	clinic	3552.93
biomass	-4800.37	player	3522.79
casino	-4785.70	laser	3399.12
partnership.agreement	-4501.55	data.center	3270.31
aircraft	-4479.71	clinical	3258.86
franchisor	-4350.14	china	3196.29
airline.industry	-4145.44	engine	3156.93
business.combination	-4041.02	cloud	3028.63
unit	-3951.97	chinese	2989.69
federal.income.tax	-3790.37	gene.therapy	2982.20
partnership	-3655.89	graphic	2976.21

Table A.2: Influential Terms for Pandemic Jumps (TF-IDF Weighting). This table shows the bottom and top terms from the inverse regression model for pandemic dates. The ranking of term v is constructed as follows. First we compute its corpus-wide term-frequency, inverse-document-frequency score. The term frequency of v is $\mathrm{tf}_v = \sum_i c_{i,v}$, the inverse document frequency of v is $\mathrm{idf}_v = \log\left(\frac{N}{\mathrm{df}_v}\right)$ where N is the number of firms in our sample and df_v is the number of firms that use term v in their RF texts. $\text{tf-idf}_v = \mathrm{tf}_v \times \mathrm{idf}_v$. Second, let $\hat{\beta}_v^P$ be the estimated coefficient on AbnRet_i^P in the inverse regression model. Terms are ranked based on $\text{tf-idf}_v \times \hat{\beta}_v^P$. The 'Value' column displays the value of this expression. As part of the preprocessing pipeline, we locate multi-word expressions and replace them with a single token. The table indicates these with a period separating the words in such phrases.

Bottom 30 Terms		Top 30 Terms	
Term	**Value**	**Term**	**Value**
tenant	-15810.95	hotel	31389.56
student	-9557.59	natural.gas	9899.49
operating.partnership	-7896.48	gaming	8074.73
reit	-7610.02	crude.oil	8016.95
real.estate	-7081.87	hotel.property	7193.75
homebuilding	-7078.13	aircraft	7111.44
the.company	-7038.95	solar	7065.24
product.candidate	-6580.21	pipeline	6868.64
home	-6515.21	oil	6852.39
client	-6057.50	ferc	6846.26
property	-5597.77	unitholder	6420.11
cellular	-5487.07	drilling	6027.62
fcc	-5200.31	ngl	5968.09
clinical.trial	-4671.19	fuel	5235.03
wireless	-4270.11	semiconductor	5142.39
spectrum	-4184.30	general.partner	5136.86
homebuyer	-3970.28	refined.product	5021.86
mortgage	-3870.20	lessee	4939.60
the.company	-3718.57	franchisor	4669.67
mortgage.loan	-3626.50	travel	4659.29
clinic	-3574.13	refinery	4409.03
patient	-3555.76	casino	4162.66
physician	-3555.10	ethanol	4058.72
rental.rate	-3502.23	business.combination	4012.34
driver	-3436.56	trust.account	3861.00
clinical.study	-3336.26	franchisee	3735.45
land	-3185.22	trs	3658.85
domain.name	-3182.90	vessel	3594.50
hospital	-3029.81	advertiser	3390.89
drug.candidate	-3002.49	common.unit	3280.60

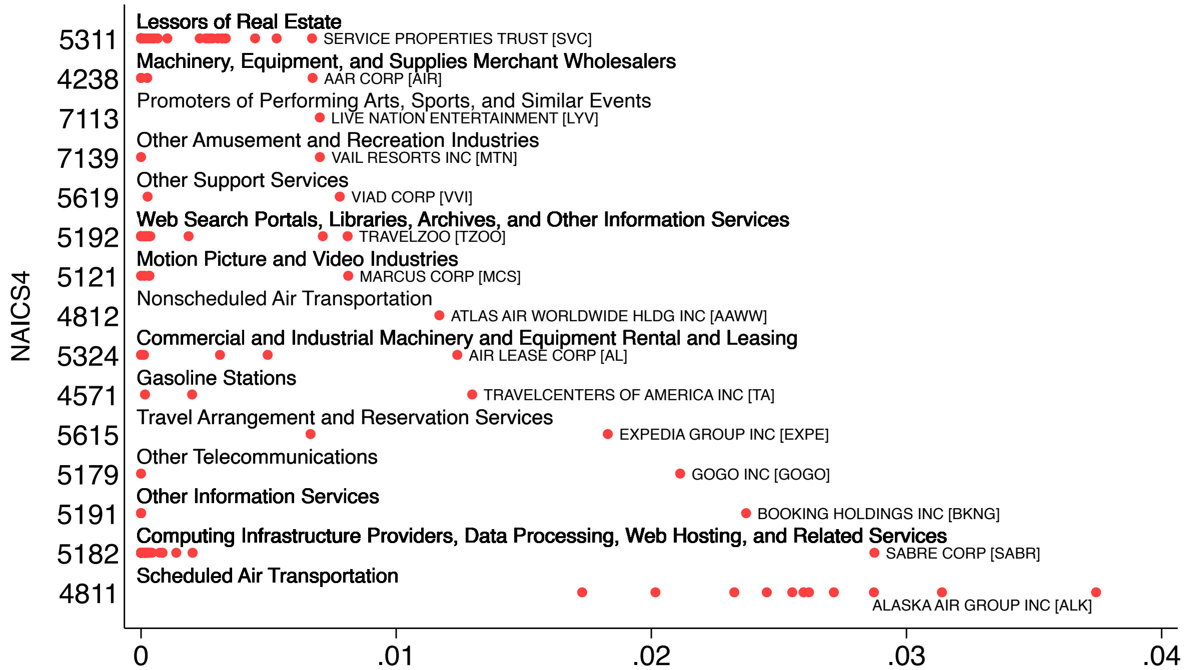
Table A.3: Influential Terms for Inflation Jumps (TF-IDF Weighting). This table shows the bottom and top terms from the inverse regression model for inflation dates. The ranking of term v is constructed as follows. First we compute its corpus-wide term-frequency, inverse-document-frequency score tf-idf_v. See notes for Table A.2 for explanation. Second, let $\hat{\beta}^I_v$ be the estimated coefficient on AbnRet^I_i in the inverse regression model. Terms are ranked based on $\text{tf-idf}_v \times \hat{\beta}^I_v$. The 'Value' column displays the value of this expression. As part of the preprocessing pipeline, we locate multi-word expressions and replace them with a single token. The table indicates these with a period separating the words in such phrases.

Index	First Seed Word	Direction	Count	Coef	T-Stat
1	hotel	neg	5	-0.21	-7.18
2	share	neg	7	-0.07	-4.57
3	unit	neg	5	-0.20	-5.61
4	tenant	neg	8	-0.28	-17.62
5	restaurant	neg	6	-0.20	-11.97
6	crude.oil	neg	15	-0.14	-6.98
7	business.combination	neg	5	-0.06	-2.73
8	satellite	neg	11	0.01	0.78
9	debt	neg	13	-0.20	-5.23
10	preferred.stock	neg	9	-0.14	-11.08
11	airline	neg	8	-0.19	-9.76
12	fuel	neg	6	-0.09	-3.74
13	retail	neg	6	-0.06	-2.39
14	indenture	neg	13		
15	product.candidate	pos	27	0.15	5.57
16	solution	pos	18	0.19	26.45
17	patent	pos	14		
18	game	pos	9	0.04	4.42
19	united.states	pos	11	0.26	6.30
20	fda	pos	7		
21	datum.center	pos	7	0.10	9.31
22	provider	pos	7	0.13	4.29
23	european.union	pos	6	0.17	6.99
24	collaboration	pos	6		
25	privacy	pos	5	0.13	5.96
26	new.customer	pos	5	0.15	5.33
27	dollar	pos	7	0.13	3.58
28	disease	pos	24		

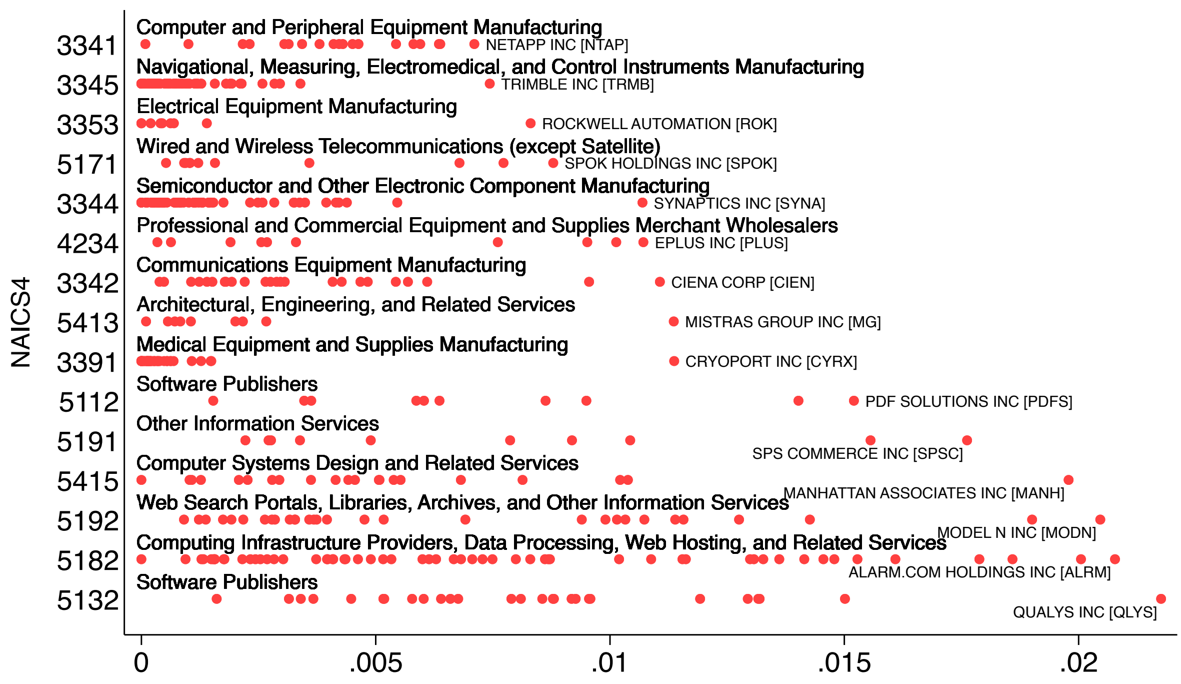
Table A.4: Pandemic Exposure Components. The Table displays information on the term sets that reflect separate components of pandemic exposure. The first column gives an integer index to each exposure. The second shows the highest-ranked seed word. The third records the direction of the exposure (*neg*ative or *pos*itive) and the fourth the total number of terms in the set. Appendix B (Online) prints all terms in each set. The fifth and sixth columns report the point estimates and T-statistics, respectively, of fitting 5. We combine exposures in the regression when the associated $z_{i,j}^{d}$ measures have a correlation above 0.5, which are denoted by common colors.

Index	First Seed Word	Direction	Count	Coef	T-Stat
1	tenant	neg	12	-0.03	-2.11
2	home	neg	7	-0.1	-6.23
3	product.candidate	neg	19		
4	patient	neg	8	-0.01	-1.7
5	mortgage	neg	6		
6	cellular	neg	13	-0.04	-5.59
7	retailer	neg	10	-0.15	-5.29
8	economic.condition	neg	6	-0.11	-5.79
9	internet	neg	14	-0.01	-1.42
10	senior.note	neg	17	-0.06	-3.97
11	regulatory.authorities	neg	6		
12	food	neg	7	-0.09	-3.1
13	supplier	neg	5	-0.05	-3.14
14	approval	neg	5		
15	information.technology.systems	neg	5	-0.09	-4.35
16	intangible.asset	neg	6	-0.06	-3.11
17	drug	neg	9		
18	cyber	neg	10	-0.07	-5.04
19	natural.gas	pos	13	0.12	4.3
20	common.stock	pos	5	0.04	1.48
21	fuel	pos	9	0.13	3.72
22	solution	pos	10	0.05	6.18
23	solar	pos	7	0.2	6.39
24	semiconductor	pos	10	0.13	11.05
25	refinery	pos	13	0.12	3.42
26	franchisor	pos	5	-0.0	-0.24
27	reserve	pos	7		
28	vessel	pos	5	0.03	1.3
29	ethanol	pos	7	0.06	3.95
30	airline	pos	7	0.11	5.22

Table A.5: Inflation Exposure Components. The Table displays information on the term sets that reflect separate components of inflation exposure. For an explanation of columns, see notes to Table A.4.

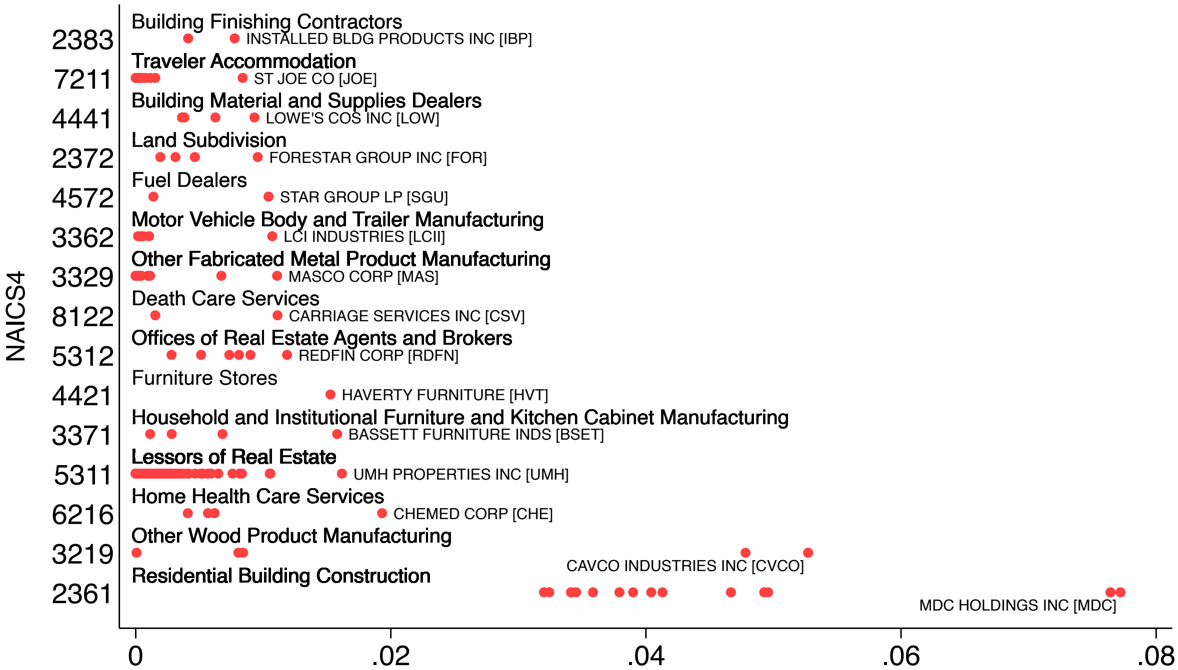


(a) Travel

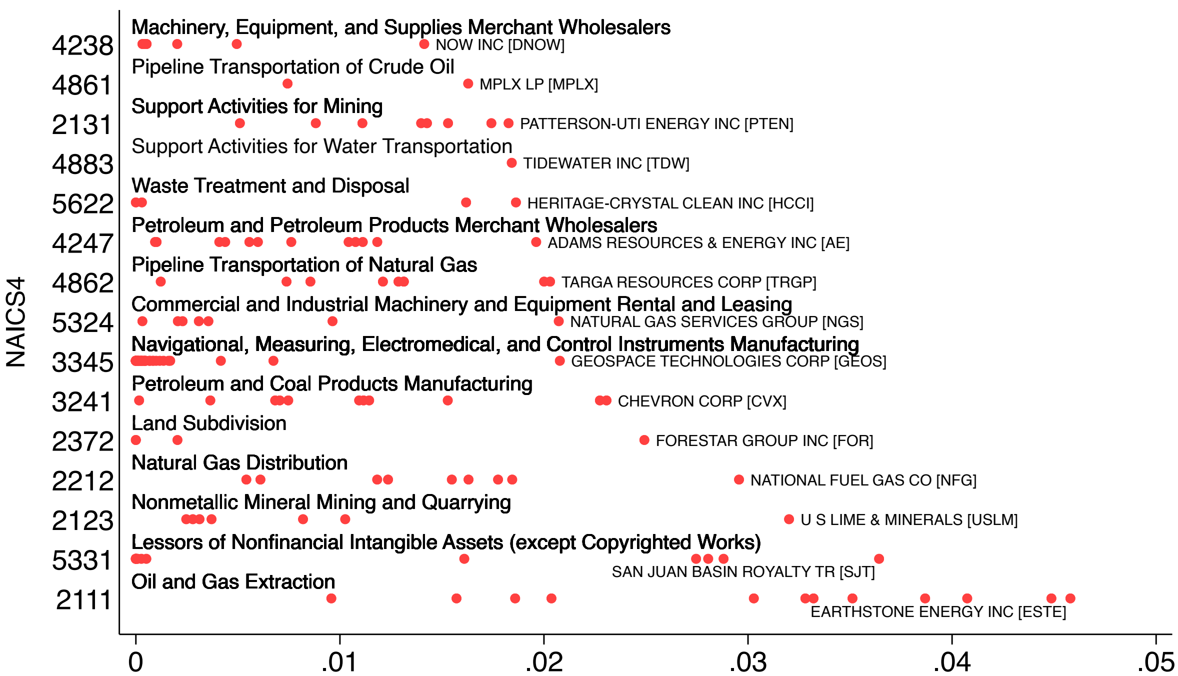


(b) Information Technology Environment

Figure A.1: Distributions of Specific Pandemic Exposures within NAICS4 Industry. For each specific exposure, we choose the top ten NAICS4 industries ranked according to maximally exposed firm. Each red dot represents a firm and the value of the given exposure is on the x axis.

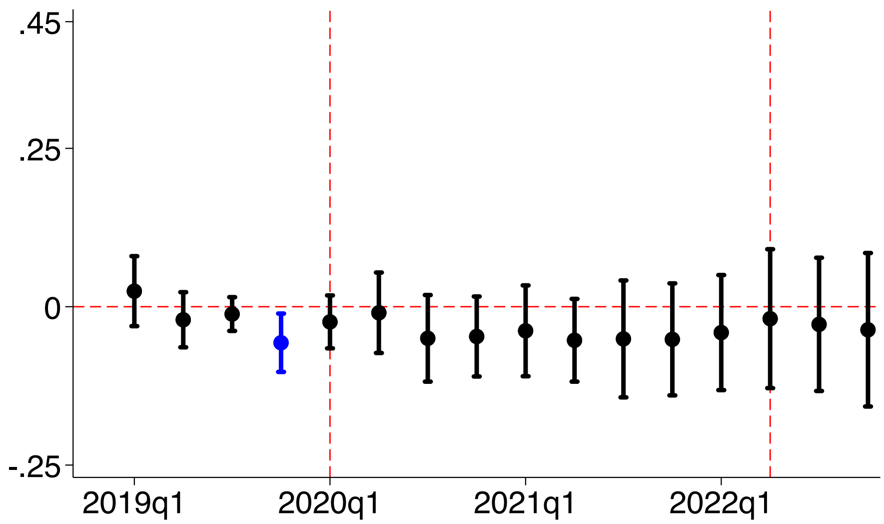


(a) Mortgages and Real Estate

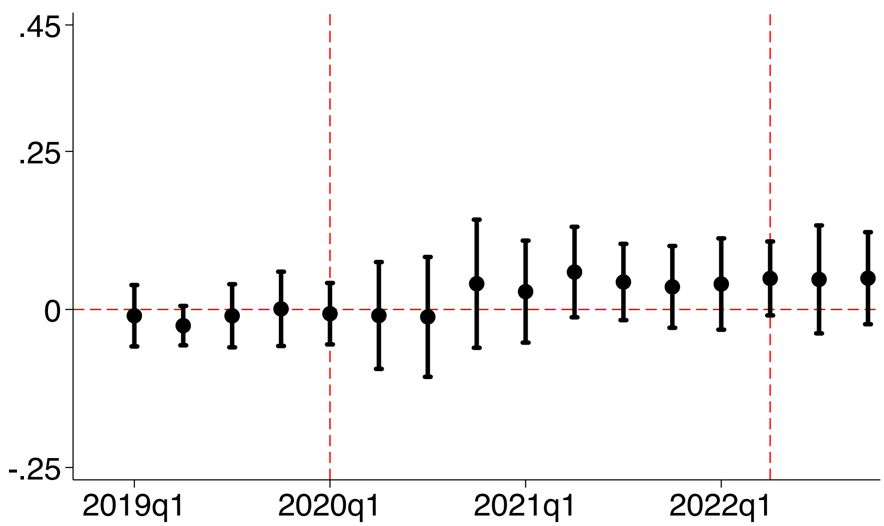


(b) Oil and Gas

Figure A.2: Distributions of Specific Inflation Exposures within NAICS4 Industry. For each specific exposure, we choose the top ten NAICS4 industries ranked according to maximally exposed firm. Each red dot represents a firm and the value of the given exposure is on the x axis.



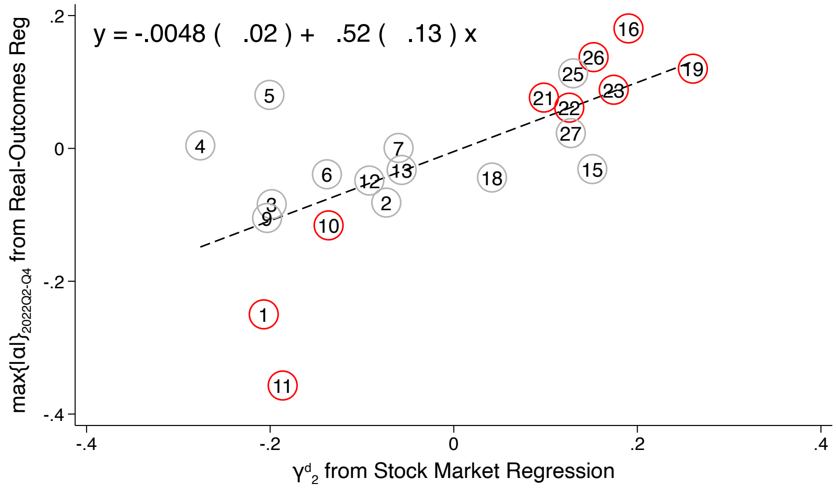
(a) Effect of Pandemic Exposure



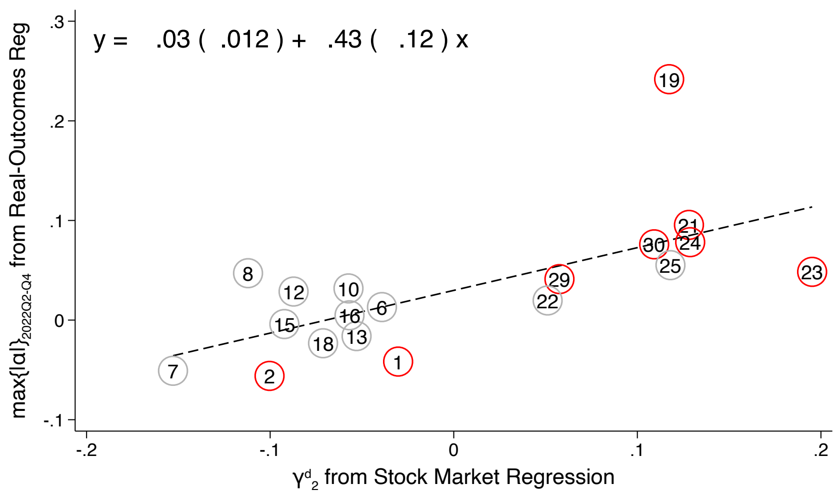
(b) Effect of Inflation Exposure

Figure A.3: Effect of Shock Exposure (Residual) on Quarterly Revenue Growth

We fit the panel regression model (6) using (EXP-TEXT) as the vector of exposures. This yields four coefficients in $\boldsymbol{\alpha}_t$ for each t. Top (bottom) panel displays coefficient on pandemic (inflation) residual exposure. Pandemic residual exposure is $\hat{\varepsilon}_i^P$ and inflation residual exposure is $\hat{\varepsilon}_i^I$. These capture variation in jump-date abnormal returns not captured by text exposures. Point estimates displayed with 95% confidence intervals.



(a) Specific Pandemic Exposures



(b) Specific Inflation Exposures

Figure A.4: Effect of Specific Exposures on Asset Prices vs Real Outcomes

This figure compares the effects of specific exposures on asset returns (x-axis) and on real outcomes (y-axis). The asset return effects are the same as reported in Tables A.4 and A.5, fifth column. The real pandemic effects are the estimated α_t coefficient for $t = 2020Q2$ when using the exposure vectors (EXP-COMP-j) for $d = P$ and all pandemic components j. The real inflation effects are the maximum estimated α_t coefficient for $t = 2022Q2, 2022Q3, 2022Q4$ when using the exposure vectors (EXP-COMP-j) for $d = I$ and all inflation components j. Each exposure is numbered using the Index column of Tables A.4 and A.5. Red circles indicate significance at the 95% level for real outcomes.

Online Appendix: Not for Publication

A (Online) Securities Sample and Feature Space Construction

A.1 (Online) Securities sample

The following are the details on how we construct our analysis sample: For example, for our jump dates in 2020:

- We link 2,858 firms (i) with at least one 10-K filing (with a non-empty Part 1a) from January 2015 to December 2019, and (ii) with equity return data for the 37 jump dates identified by Baker et al. (2020) for 2020.
- We remove 20 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 74 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 December 31, 2019. Dropped small caps account for 1.1 percent of total equity market value in the sample. In this step, we remove 894 firms.
- We discard 8 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 for our stock-market regressions of 1,905 stocks for 1,876 companies.

A.2 (Online) Text preprocessing steps

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 property tax' becomes 'We owe additional property_tax', where 'property_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. Phrases that correspond to named entities that appear more than 10 times in the corpus. We identify these entities with the named entity recognizer (NER) from the Stanford NLP group.

2. 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.

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', 'property_tax'].
- Drop common words from a standard stop word list, e.g. 'for', 'to', etc.
- Drop any terms that appear in the *Risk Factors* text of fewer than 25 firms from 2015 to 2019.

B (Online) Clusters of Influential Terms

B.1 (Online) Pandemic shock

Negative term sets

1. hotel, hotel property, property, real property, such property
2. share, common stock, class a, market price, common share, number of share, trading price
3. unit, our common, common unit, unitholder, common unitholder
4. tenant, lease, rent, master lease, lessee, ground lease, net lease, ground lease
5. restaurant, franchisee, guest, new restaurant, retail location, franchised
6. crude oil, refinery, refined product, feedstock, refining, natural gas, propane, ngl, transport, gas, crude, refined, ngls, terminal, petroleum product
7. business combination, initial, complete, consummate, consummation
8. satellite, station, programming, radio, broadcast, tv, cable, broadband, digital, interactive, television
9. debt, indebtedness, refinance, borrowing, repay, borrow, additional debt, debt obligation, additional indebtedness, mortgage debt, such debt, other debt, new debt
10. preferred stock, series, preferred unit, series a, convertible, preferred share, class b, voting right, share of series
11. airline, airport, travel, flight, passenger, ticket, route, air travel
12. fuel, diesel, jet, gasoline, motor, fuel oil
13. retail, retailer, channel, distribution channel, distribution partner, outlet
14. indenture, covenant, revolving credit facility, senior secured credit facilities, debt agreement, senior note, secured, event of default, credit facilities, subordinated note, debt instrument, other indebtedness, loan agreement

Positive term sets

15. product candidate, clinical trial, drug candidate, commercialize, collaborator, regulatory approval, trial, commercialization, clinical development, marketing approval, preclinical, study, clinical, preclinical study, efficacy, future collaborator, clinical study, future product candidate, companion diagnostic, new indication, such product candidate, commercial sale, investigational, ind, clinical testing, preclinical testing, nonclinical

16. solution, platform, cloud, software, functionality, hardware, module, service offering, web, saas, party software, computing, solutions and services, interface, networking, cloud computing, operating system, architecture

17. patent, intellectual property, intellectual property rights, infringe, patent application, trade secret, licensor, proprietary right, infringement, proprietary technology, issue patent, such patent, invention, unauthorized use

18. game, player, download, mobile device, smartphone, streaming, publisher, tablet, audio

19. united states, country, foreign jurisdiction, other country, internationally, certain country, international market, other jurisdiction, foreign market, abroad, foreign country

20. fda, regulatory authorities, ema, foreign regulatory authorities, regulatory authority, foreign authority, marketing authorization

21. datum center, data center, host, aw, party datum, amazon web services, server

22. provider, service provider, network, network operators, internet service, telecommunication, communication service

23. european union, eu, european, european commission, europe, eu member states

24. collaboration, celgene, sanofi, collaboration agreement, allergan, bayer

25. privacy, data protection, personal datum, datum privacy, personal information

26. new customer, customer base, exist customer, new client, client base

27. dollar, foreign currency, currency, euro, denominate, exchange rate, local currency

28. disease, cancer, cell, therapy, treatment, antibody, gene, genetic, novel, inhibitor, cell therapy, molecular, therapeutic, diagnostic, diagnostic test, biomarker, vector, refractory, oncology, immunotherapy, tissue, blood, protein, mutation

B.2 (Online) Inflation shock

Negative term sets

1. tenant, lease, rent, rental rate, space, rental revenue, vacant space, occupancy level, other tenant, lease term, tenant default, rental income

2. home, housing, new home, homebuyer, mortgage financing, mortgage interest, availability of mortgage

3. product candidate, clinical trial, clinical study, drug candidate, marketing approval, future collaborator, clinical development, clinical, future product candidate, other product candidate, nonclinical, commercialize, commercialization, preclinical, pivotal, such product candidate, collaboration partner, collaborative partner, lead product candidate
4. patient, physician, hospital, clinic, surgical, healthcare provider, surgery, healthcare professionals
5. mortgage, mortgage loan, loan, originate, commercial real estate, lending
6. cellular, wireless, broadband, cable, handset, networking, network operators, voice, internet service, wireline, telecommunication, telephone, communication service
7. retailer, retail, channel, outlet, mall, retail store, shopping center, shopping, grocery, retail sale
8. economic condition, market condition, economic uncertainty, consumer spending, general economic conditions, recessionary
9. internet, video, content, game, online, web, programming, movie, subscriber, popular, tv, studio, radio, sport
10. senior note, indebtedness, senior credit facility, indenture, credit facility, senior secured credit facility, borrowing, debt obligation, debt covenant, revolving credit facility, financial covenant, covenant, restrictive covenant, credit agreement, additional indebtedness, unsecured term loan, unsecured credit facility
11. regulatory authorities, ema, foreign regulatory authorities, regulatory authority, food and drug administration, fda
12. food, beverage, ingredient, food product, chicken, nutritional, nutrition
13. supplier, vendor, distributor, manufacturer, independent distributor
14. approval, clearance, regulatory approval, pma, regulatory clearance
15. information technology systems, information technology, information systems, computer system, technology infrastructure
16. intangible asset, lived, goodwill, carrying, impairment, impairment charge
17. drug, medicine, medical device, therapy, biosimilar product, injectable, medication, biologic, reference product
18. cyber, attack, security breach, information security, cyber attack, malware, threat, cyber incident, viruse, intrusion

Positive term sets

19. natural gas, oil, crude oil, ngl, gas, natural gas price, commodity price, ngls, hydrocarbon, natural gas production, natural gas liquid, price of oil, crude oil price
20. common stock, share, our stock, our common, common share
21. fuel, electricity, energy, coal, feedstock, power plant, electric, fired, fossil fuel
22. solution, platform, software, app, smartphone, digital, website, interactive, functionality, software solution
23. solar, renewable energy, energy efficiency, renewable, power generation, alternative energy source, renewable energy source
24. semiconductor, module, wafer, foundry, silicon, assembly, fabrication, process technology, production capacity, subsystem
25. refinery, refined product, terminal, tank, transport, pipeline, transportation, barge, pipeline system, petroleum product, crude, natural gas pipeline, rail
26. franchisor, franchisee, restaurant, franchise agreement, franchise
27. reserve, proved, natural gas reserve, reserve estimate, prove reserve, net cash flow, future production
28. vessel, rig, fleet, railcar, truck
29. ethanol, gasoline, renewable fuel, diesel, blend, gallon, refiner
30. airline, airport, travel, passenger, flight, route, ticket

C (Online) Additional Figures and Tables

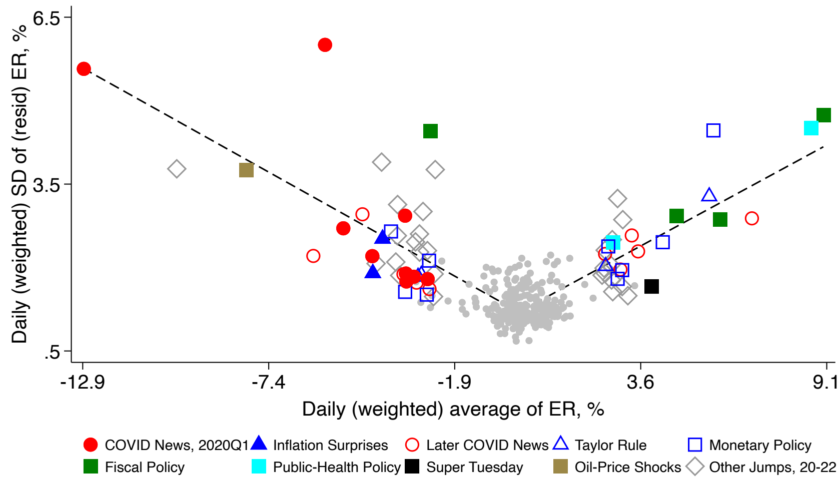


Figure C.1 (Online): Dispersion of Returns on Jump Dates (within sector). This plot is analogous to that in Figure 1, but with returns first stripped of NAICS4 fixed effects.

	COVID News, 2020Q1	Inflation Surprises	Fiscal Policy	Later COVID News	Monetary Policy	Taylor Rule	Oil-Price Shocks	Public Health Policy	Super Tuesday
COVID News, 2020Q1	1.0000								
Inflation Surprises	-0.0913	1.0000							
Fiscal Policy	0.8032	0.1340	1.0000						
Later COVID News	0.9261	0.0787	0.8517	1.0000					
Monetary Policy	-0.2665	0.6989	-0.0011	-0.2134	1.0000				
Taylor Rule	-0.2850	0.6001	-0.2705	-0.3004	0.6750	1.0000			
Oil-Price Shocks	0.7664	-0.2206	0.7723	0.7316	-0.2294	-0.4997	1.0000		
Public Health Policy	0.7637	0.0375	0.8742	0.8197	-0.0654	-0.4192	0.8221	1.0000	
Super Tuesday	0.1008	0.6691	0.2255	0.1187	0.7349	0.6673	-0.1245	0.0433	1.0000

Table C.1 (Online): Correlation Matrix of Shock Exposures. For each of the nine labeled groups of jump dates in Table A.1 we build a text-based exposure at the firm level. This table presents the correlation between these firm-level exposures.

Bottom 30 Terms		Top 30 Terms	
Term	**Value**	**Term**	**Value**
hotel	-10939.07	client	11362.10
distribution	-9131.11	technology	10634.90
property	-6332.94	product.candidate	9016.57
unitholder	-5852.97	harm	8536.23
share	-5444.45	solution	7445.12
unit	-5281.70	patent	6655.31
general.partner	-5213.07	clinical.trial	6413.66
tenant	-4854.90	platform	5850.42
travel	-4604.86	game	5429.08
reit	-4412.90	drug.candidate	4675.23
restaurant	-4126.08	development	4040.21
transaction	-3913.61	patient	3987.89
interest	-3883.07	intellectual.property	3976.48
gaming	-3695.86	united.states	3833.87
crude.oil	-3682.07	third	3558.54
business.combination	-3593.97	country	3431.84
federal.income.tax	-3482.58	subscription	3372.93
the.company	-3430.97	china	3250.98
the.company	-3260.88	equipment	3197.14
home	-3205.21	manufacture	3193.64
tax	-3120.75	drug	3173.92
satellite	-3069.60	application	3143.82
partnership	-2950.55	intellectual.property.rights	3058.14
our.common	-2944.03	commercialize	3023.66
common.unit	-2940.49	cellular	2977.45
qualify	-2885.72	datum	2853.03
cash	-2825.65	fda	2821.13
shareholder	-2797.74	manufacturing	2532.31
purpose	-2736.75	difficulty	2524.40
lease	-2682.22	achieve	2511.20

Table C.2 (Online): Influential Terms for Pandemic Jumps (TF Weighting). This table shows the bottom and top terms from the inverse regression model for pandemic dates. The ranking of term v is constructed as follows. First we compute its corpus-wide term-frequency score. The term frequency of v is $\text{tf}_v = \sum_i c_{i,v}$. Second, let $\hat{\beta}_v^P$ be the estimated coefficient on AbnRet_i^P in the inverse regression model. Terms are ranked based on $\text{tf}_v \times \hat{\beta}_v^P$. The 'Value' column displays the value of this expression.

Bottom 30 Terms		Top 30 Terms	
Term	**Value**	**Term**	**Value**
the.company	-15033.95	hotel	10327.32
property	-12048.25	project	7429.03
the.company	-12007.24	natural.gas	5723.75
tenant	-7339.45	technology	4831.76
real.estate	-5050.60	oil	4596.90
home	-4407.43	common.stock	4147.18
client	-3968.80	share	4000.50
member	-3518.39	pipeline	3971.35
product.candidate	-3164.41	fuel	3925.34
network	-3127.26	business.combination	3590.88
reit	-3054.17	crude.oil	3173.41
operating.partnership	-2667.43	travel	3136.16
store	-2478.23	partner	3010.75
clinical.trial	-2470.05	solution	2924.79
student	-2379.65	equipment	2924.17
patient	-2247.32	test	2885.83
physician	-2085.76	contract	2717.03
treatment	-2067.76	gaming	2508.12
land	-1992.26	initial	2504.09
community	-1911.12	platform	2350.78
interest.rate	-1844.90	drilling	2243.81
mortgage	-1828.89	partnership	2225.64
homebuilding	-1828.41	aircraft	2199.22
fcc	-1750.20	production	2161.28
brand	-1748.45	solar	2155.73
cellular	-1741.16	complete	2041.49
hospital	-1738.95	user	2039.45
consumer	-1720.85	operator	2034.10
lease	-1654.35	unitholder	2020.14
clinical.study	-1625.24	lessee	1927.92

Table C.3 (Online): Influential Terms for Inflation Jumps (TF Weighting). This table shows the bottom and top terms from the inverse regression model for inflation dates. The ranking of term v is constructed as follows. First we compute its corpus-wide term-frequency. See notes for Table C.2 (Online) for explanation. Second, let $\hat{\beta}_v^I$ be the estimated coefficient on AbnRet_i^I in the inverse regression model. Terms are ranked based on $\text{tf}_v \times \hat{\beta}_v^I$. The 'Value' column displays the value of this expression.

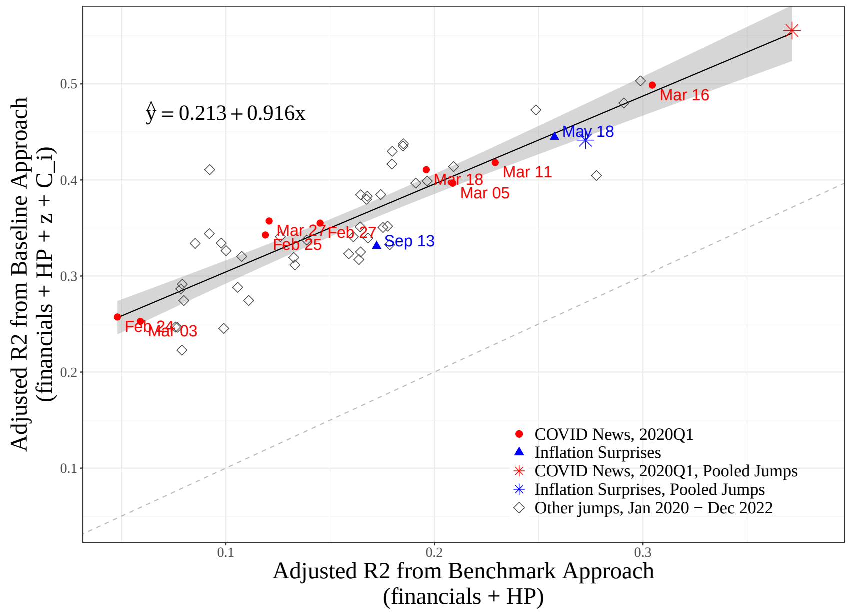


Figure C.2 (Online): This Figure is similar to Figure 3a but controls for the text-based industry industry instead of the NAICS2 industry. Text-based industry codes are developed in Hoberg and Phillips (2010) and Hoberg and Phillips (2016) are are based on similarity between 10K product descriptions.

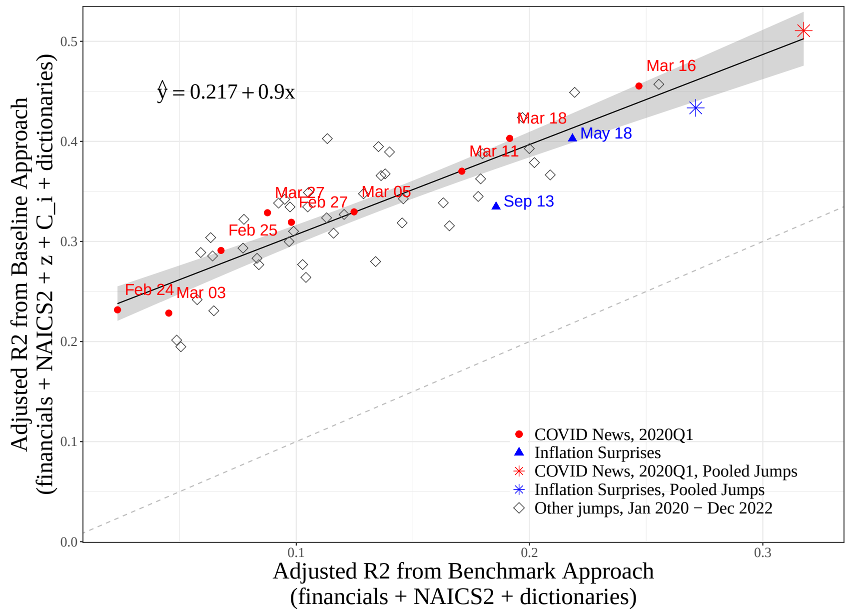
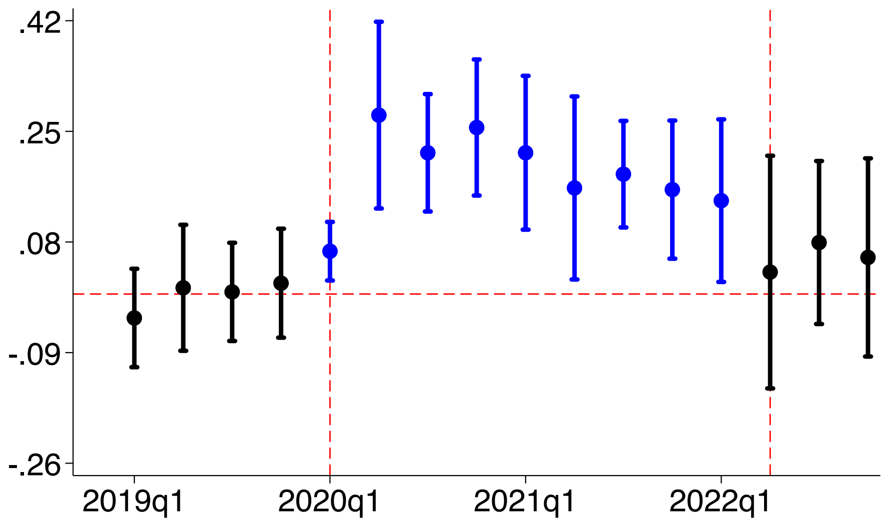
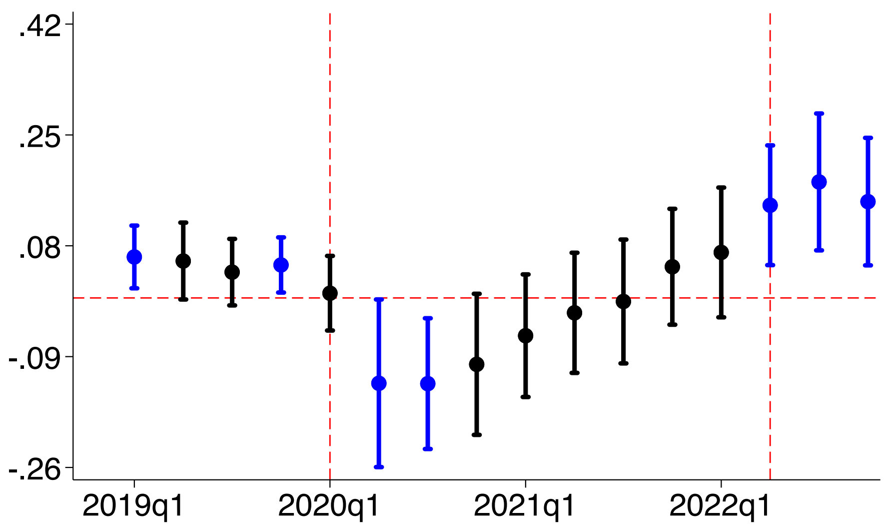


Figure C.3 (Online): These figures report the adjusted R^2 from regressing abnormal returns from jump dates on standard controls and the dictionary measures from Baker et al. (2025a) (x-axis) and additionally including text-based shock exposures (y-axis). The top panel uses the entire sample, while the bottom panel reports the adjusted R^2 on held-out firm samples not used for the estimation of the MNIR coefficients.

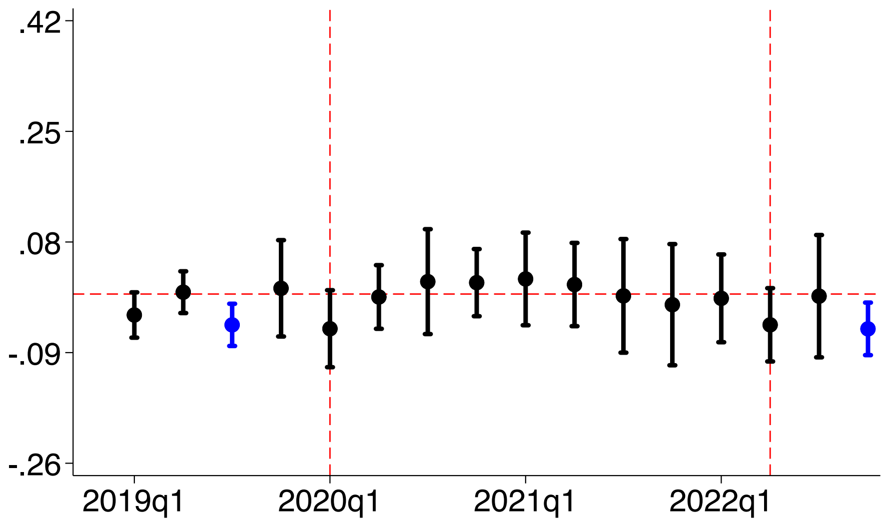


(a) Effect of Pandemic Exposure

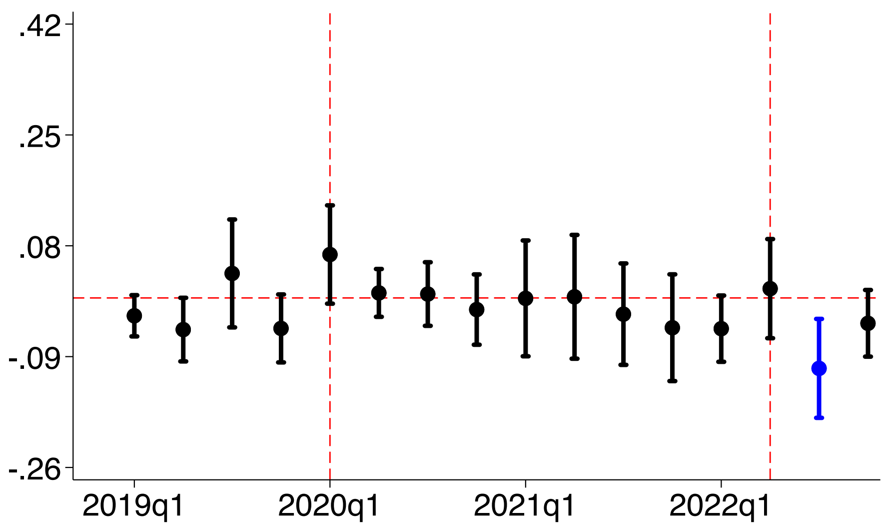


(b) Effect of Inflation Exposure

Figure C.4 (Online): Effect of Shock Exposure (Text) joint with Earnings Call Variables
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (6) when using pandemic and inflation exposures, as well as the four variables from 2020Q1 from Hassan et al. (2023) based on earnings calls. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals.



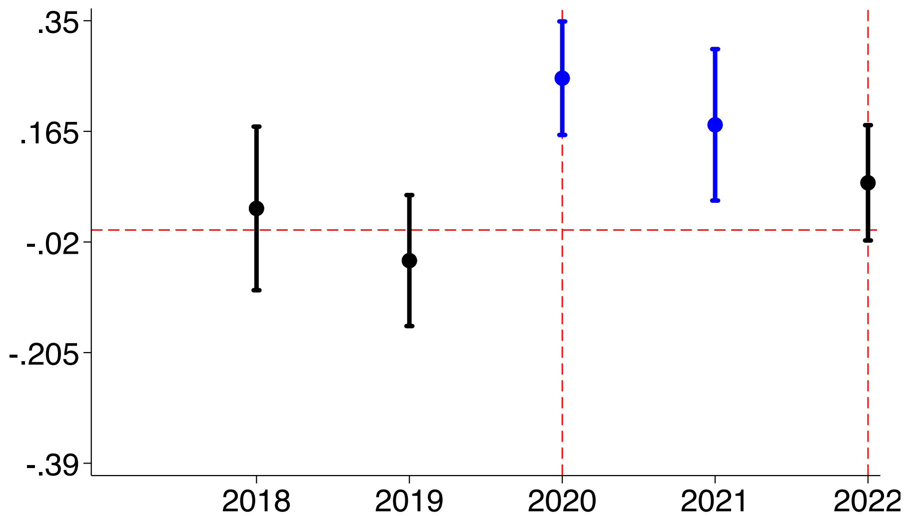
(a) Effect of Negative Sentiment



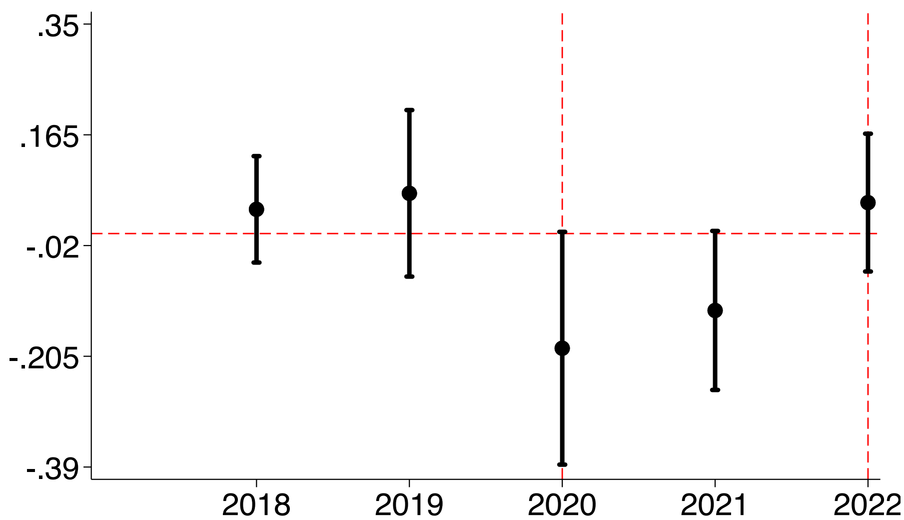
(b) Effect of Positive Sentiment

Figure C.5 (Online): Effect of Earnings Call Sentiment

This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (6) when using pandemic and inflation exposures, as well as the four variables from 2020Q1 from Hassan et al. (2023) based on earnings calls. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals.



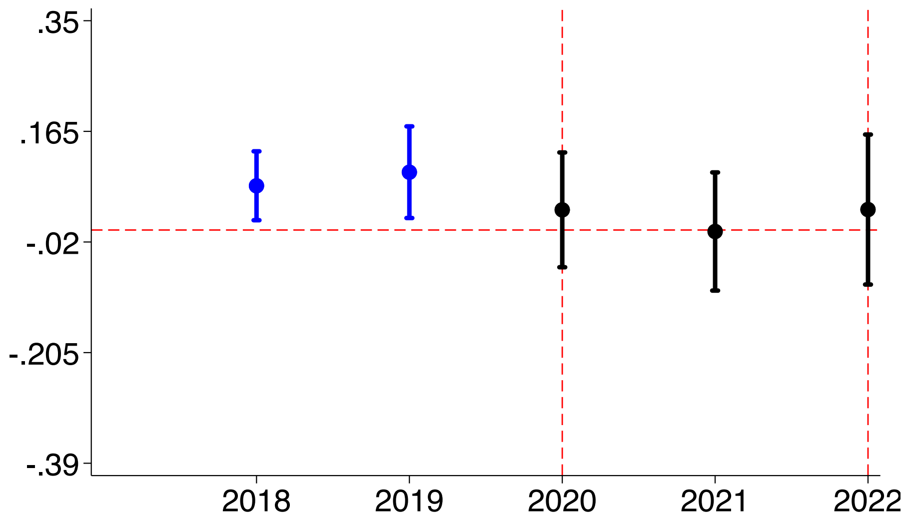
(a) Effect of Pandemic Exposure



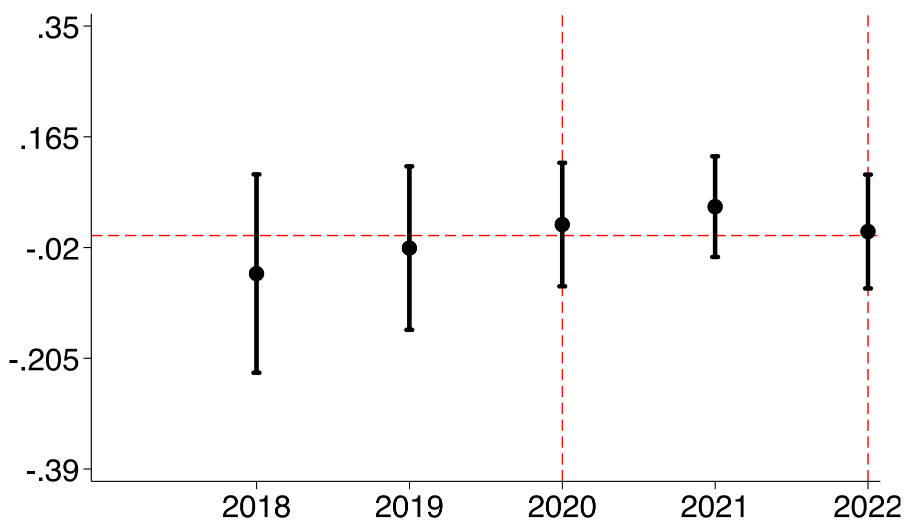
(b) Effect of Inflation Exposure

Figure C.6 (Online): Effect of Shock Exposure (Text) on Annual Revenue Growth

This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual revenue growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals.

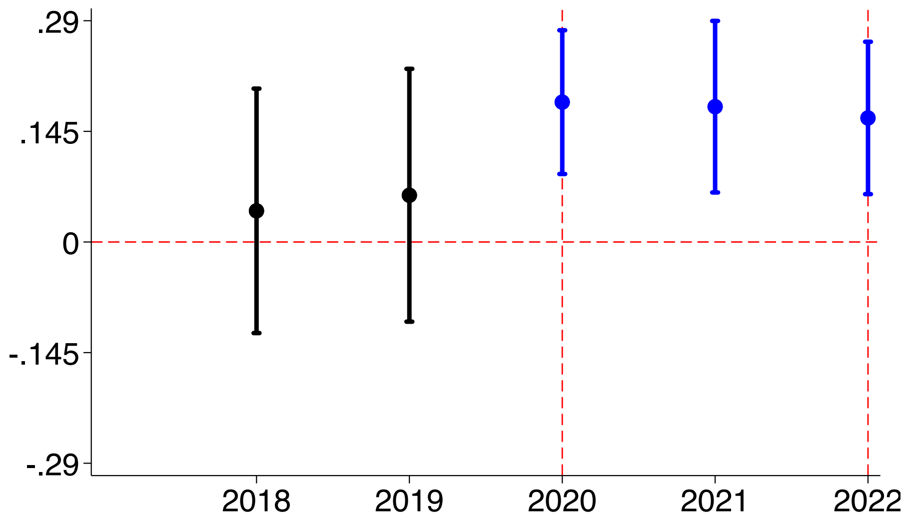


(a) Effect of Pandemic Exposure

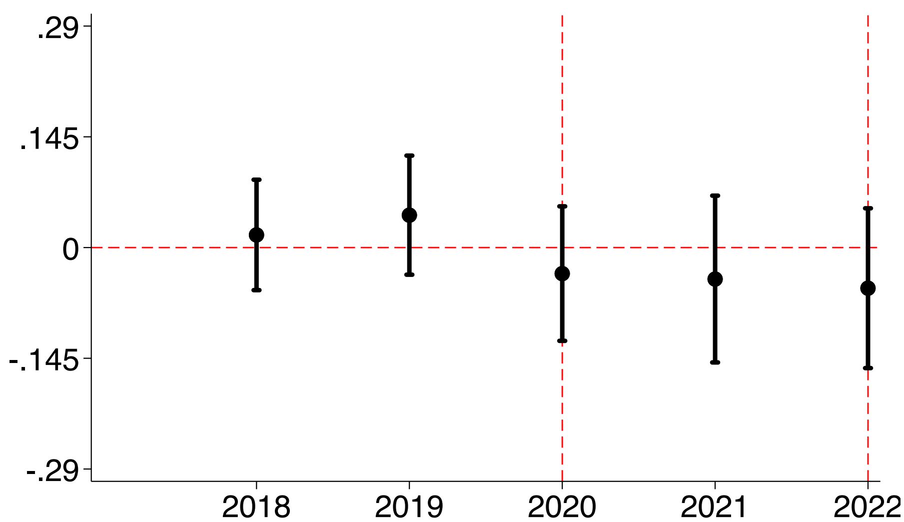


(b) Effect of Inflation Exposure

Figure C.7 (Online): Effect of Shock Exposure (Residual) on Annual Revenue Growth
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual revenue growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) residual exposure. Point estimates displayed with 95% confidence intervals.

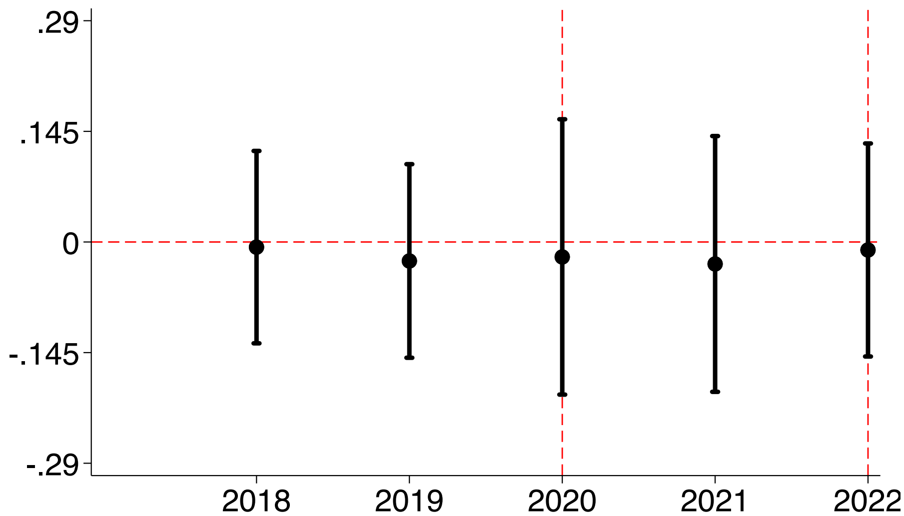


(a) Effect of Pandemic Exposure

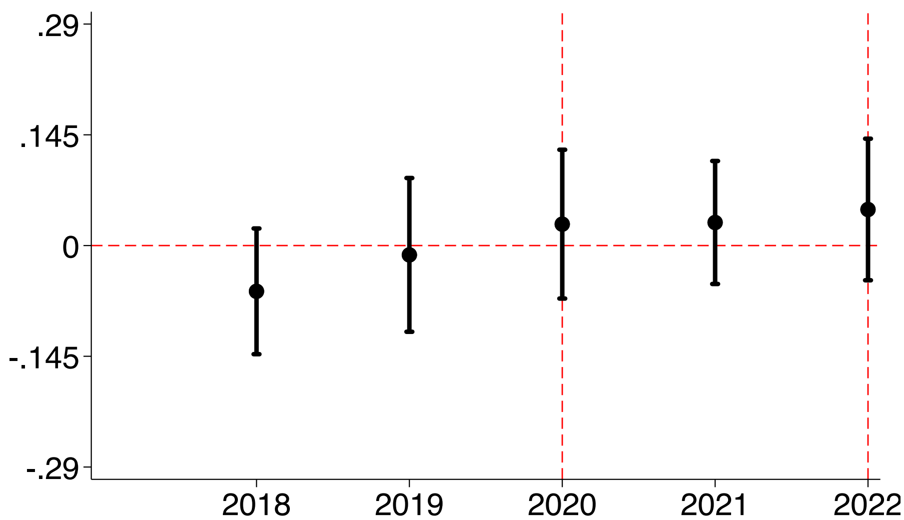


(b) Effect of Inflation Exposure

Figure C.8 (Online): Effect of Shock Exposure (Text) on Annual Employment Growth
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual employment growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals.

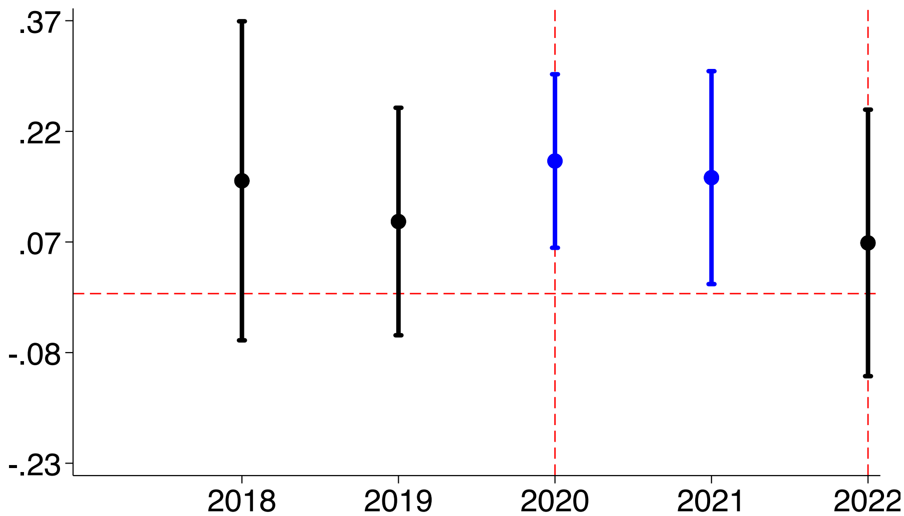


(a) Effect of Pandemic Exposure

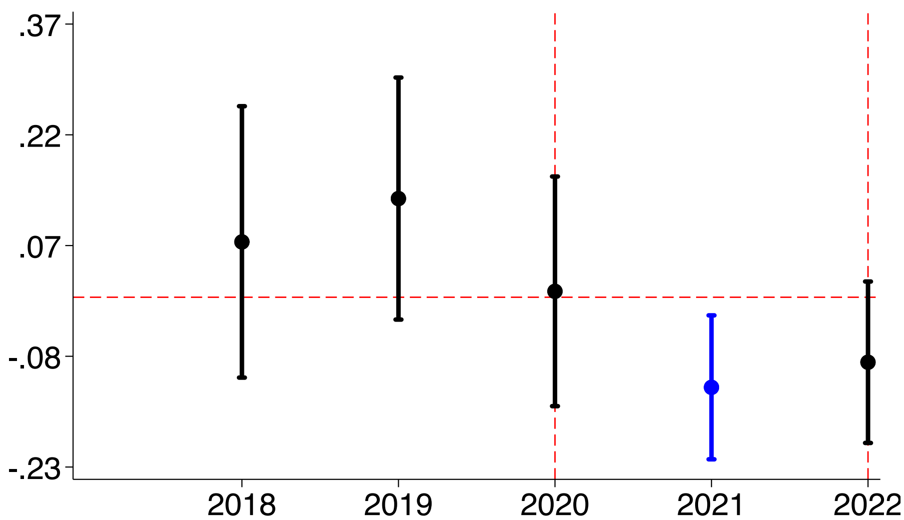


(b) Effect of Inflation Exposure

Figure C.9 (Online): Effect of Shock Exposure (Residual) on Annual Employment Growth
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual employment growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) residual exposure. Point estimates displayed with 95% confidence intervals.

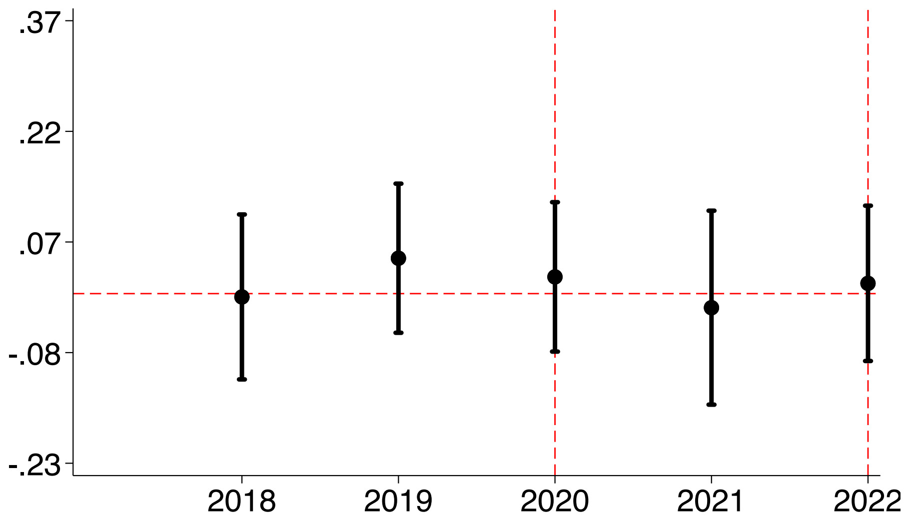


(a) Effect of Pandemic Exposure

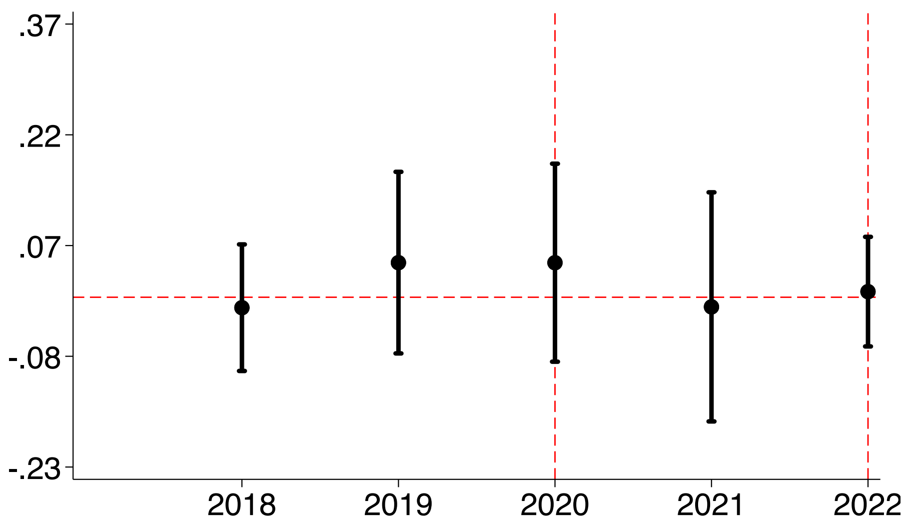


(b) Effect of Inflation Exposure

Figure C.10 (Online): Effect of Shock Exposure (Text) on Annual Investment Growth
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual investment growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) text exposure. Point estimates displayed with 95% confidence intervals.



(a) Effect of Pandemic Exposure



(b) Effect of Inflation Exposure

Figure C.11 (Online): Effect of Shock Exposure (Residual) on Annual Investment Growth
This figure displays the estimated $\boldsymbol{\alpha}_t$ coefficients from (7) when using (EXP-TEXT) as the vector of exposures and annual investment growth as an outcome measure. Top (bottom) panel displays coefficient on pandemic (inflation) residual exposure. Point estimates displayed with 95% confidence intervals.

VARIABLES	(1)	(2)	(3)	(4)	(5)
AbnRet, Econ		0.22***		0.21***	
		(0.065)		(0.060)	
AbnRet, HighInf				-0.09	
				(0.055)	
z-score, Econ			0.33***		0.28***
			(0.071)		(0.061)
AbnRet, Econ - resids			0.05		0.05
			(0.044)		(0.044)
z-score, HighInf					-0.17***
					(0.053)
AbnRet, HighInf - resids					0.01
					(0.035)
Observations	1,302	1,302	1,302	1,302	1,302
Adjusted R^2	0.054	0.086	0.116	0.091	0.137
Fixed Effects	NAICS2	NAICS2	NAICS2	NAICS2	NAICS2
Financials	Yes	Yes	Yes	Yes	Yes
Text Length	No	Yes	Yes	No	Yes

Table C.4 (Online): Effect of Exposures on 2020Q3 Earnings Surprises. The Table presents results of regressing 2020Q3 earnings per share surprises on firm-level covariates. The first column regresses on the same sector and financial controls used in asset regression (3) for $y = 2020$. The rest of the columns introduce measures of firm-level shock exposures. They also include a control for text length C_i. The second column includes AbnRet_i^P; the third column, z_i^P and $\hat{\varepsilon}_i^P$; the fourth column, AbnRet_i^P and AbnRet_i^I; and, the fifth column, z_i^P, z_i^I, $\hat{\varepsilon}_i^P$, and $\hat{\varepsilon}_i^I$.

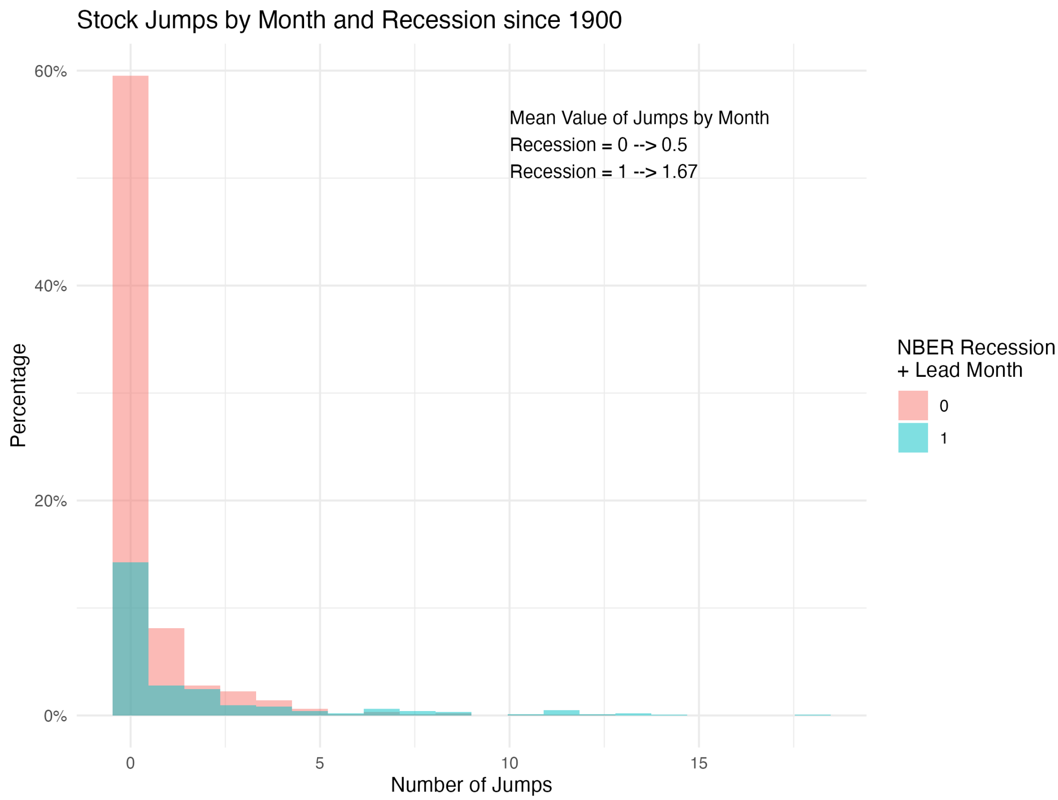


Figure C.12 (Online): Frequency of Jump Date Arrivals by Month

We tabulate the frequency of jump dates by month since 1990, and separately plot histograms according to whether months fall in recessionary vs expansionary periods. The recession classification comes from FRED series USREC. The jump date data comes from https://www.stockmarketjumps.com/. We include one month prior to the beginning of each recession in the definition of recession to account for jump dates' incorporating forward-looking news.

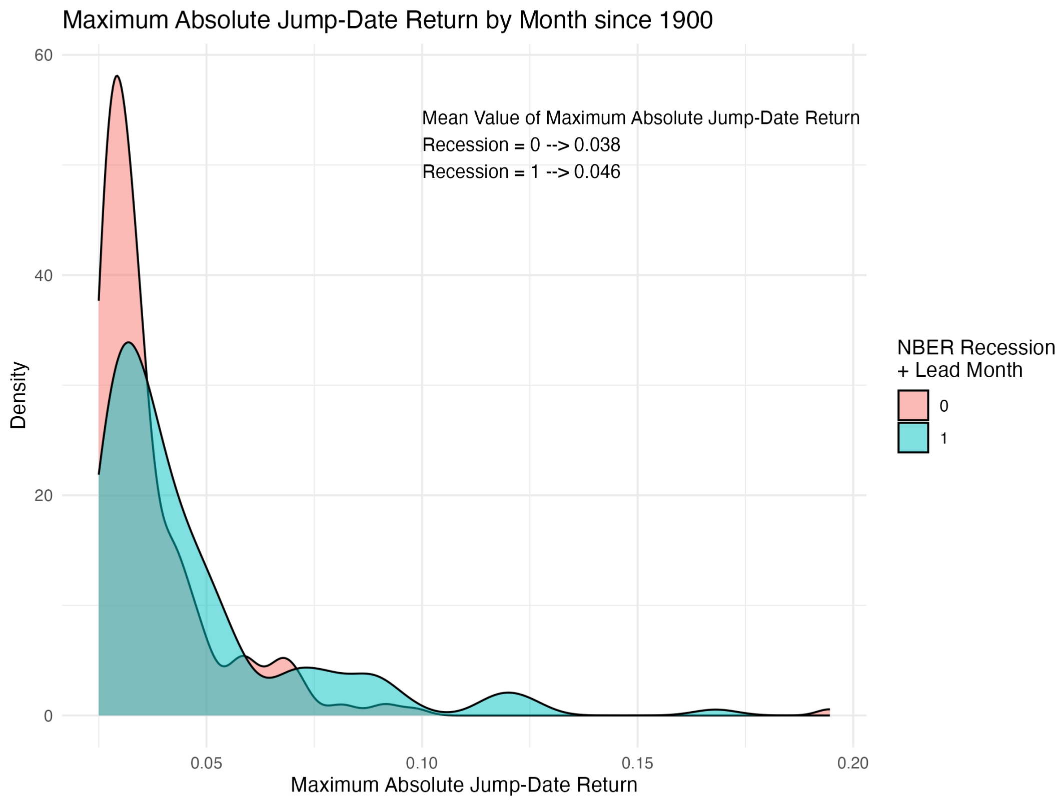


Figure C.13 (Online): Maximum Jump-Date Absolute Return by Month

We first consider all months with at least one jump date since 1900, then compute the maximum absolute value of the market return on such dates. We then plot the distributions of the maximum value according to whether months fall in recessionary vs expansionary periods. The recession classification comes from FRED series USREC. The jump date data comes from https://www.stockmarketjumps.com/. We include one month prior to the beginning of each recession in the definition of recession to account for jump dates' incorporating forward-looking news.

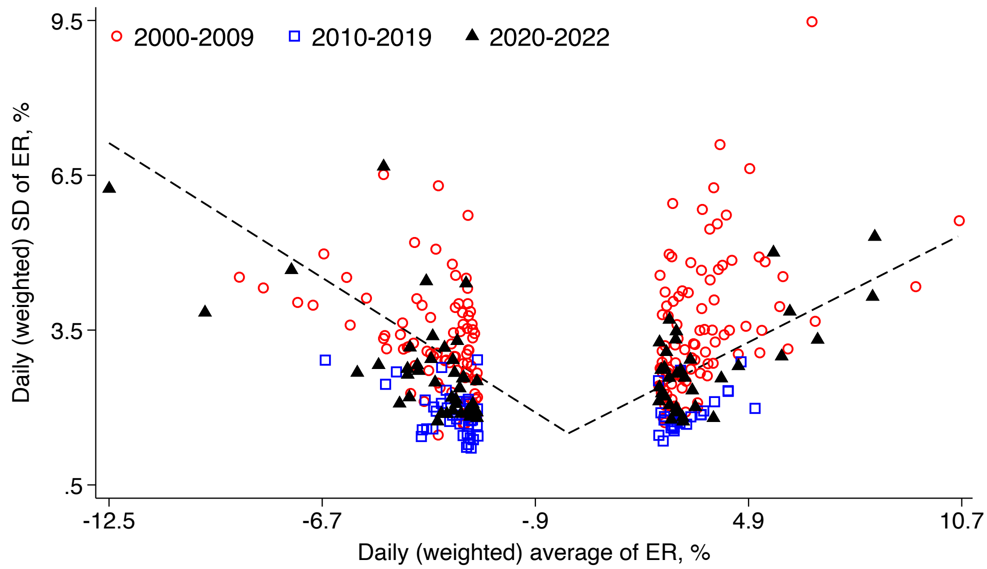


Figure C.14 (Online): Mean and SD of U.S. Equity Returns for all Jump Dates in 2000-2022. Each point in the scatterplot corresponds to a jump date. The x-axis value is the value-weighted mean return for a sample of equities detailed in Section 2. The y-axis value is the standard deviation of returns computed across firms.