The Composition of Firm Attention and Firm Risk☆

Preetesh Kantak[1], Alan Kwan[2]

Abstract

We examine how firm attention reflects dynamics in systematic exposure. To determine which activities define firms' risk profiles, we build a time-varying beta model that embeds idiosyncratic variation in operating leverage (DOL) and growth opportunities (GO). To this we add investor learning, which leads to specific asset pricing implications around information revealing events (IREs). We test whether variations in firm attention mimic the implications of the model using a proprietary dataset that captures high frequency employee attention to nearly 6,000 topics across 3,000,000 firms. We find that smaller firms with more GOs and higher DOL focus more on systematic versus investment related topics. Furthermore, positive innovations to this ratio lead not only to a drop in investment, but also predict negative shocks to asset prices and higher systematic risk around IREs. These findings map to model primitives, suggesting that variations in the composition of firm attention contain valuable information about variations in firm risk exposures.

Keywords: Information Choice, Learning, Private Information

JEL Classification: C55, D21, D22, D81, D83

[1]Indiana University, Kelley School of Business, pkantak@iu.edu, +1 812-855-4806

[2]Hong Kong University, Faculty of Business and Economics, apkwan@hku.hk, +852 9794-9483

Helpful comments and suggestions were provided by Riccardo Colacito, Zhi Da (discussant), Fotis Grigoris, Christian Heyerdahl-Larsen, Noah Stoffman, Charles Trzcinka, and Zhenyu Wang. We would also like to thank seminar participants at the 2021 Wabash River Conference at Notre Dame, and Indiana University-Bloomington. The authors further acknowledge the Indiana University Pervasive Technology Institute for providing HPC database, storage, and consulting resources that have contributed to the research results reported within this paper (Stewart, Welch, Plale, Fox, Pierce, and Sterling, 2017). Google Cloud Platform applications such BigQuery were used extensively in this research. All errors are our own.

1. Introduction

While many have focused on *investor* attention (see, inter alia, Ben-Rephael, Da, and Israelsen, 2017; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016), the literature has largely ignored *firm* attention and how it reflects variations in their own growth prospects and risk (see, e.g., Gondhi, 2017, for recent theoretical work). As the primary economic unit of production, measuring this relationship is an important addition to our understanding of the tradeoffs firms face. We propose a model that adds regime switching cash-flow growth and investor learning to a canonical framework of firms with both operating leverage (DOL) and real option (GO) considerations. We then test the degree to which the level and variation in firm attention across two sets of topics–those related to systematic risk and those related to within sector production–reflect these same considerations. The empirical work uses a novel dataset that quantifies attention to a broad set of topics by employees across more than 3,000,000 firms. In keeping with the model's implications, we find that smaller firms with greater GOs and DOL more intensely follow systematic versus sector topics. Additionally, dynamic shifts in attention from sector to systematic topics predict lower returns and higher covariances (β) with the market portfolio. This predictability is highly concentrated around earnings, suggesting investor learning about underlying cash-flow uncertainty during periods of information revelation.

Our model appeals to the time-varying beta literature (see, e.g., Berk, Green, and Naik, 1999; Carlson, Fisher, and Giammarino, 2004; Sagi and Seasholes, 2007; Babenko, Boguth, and Tserlukevich, 2016). Firm primitives reflect underlying business considerations–operating leverage and growth options–embedded in a continuous time model with a single priced risk factor. Systematic risk increases in poor economic states as the probability of a liquidity events rise and decrease in strong economic states as growth options are converted to assets-in-place. This simple model suggests two proxies associated with firm characteristics–the probability of a liquidity or systematic event and of exercising of growth options–that if observable would drastically improve the markets ability to evaluate a firm.

Our next motivation is to understand cross-sectional variation in firm behavior. We add idiosyncratic regime switching drift rates to firms in the economy. Market participants can learn about the regime a firm faces from cash-flow dynamics. Investors become more uncertain about the state of drift as cash-flows differ from their conditional expectations. This increases the variance of returns much more than implied by a model without learning. Furthermore, given that this information may not be immediately observable to investors, attention proxies may reflect asymmetric information on the part of the firm. This has important implications on the timing of asset pricing dynamics around information revealing events, something we exploit in our empirical analysis.

We use a proprietary dataset to test whether the composition of attention empirically captures these theoretical implications. The raw data is generated from employee reading of articles from a consortium of more than 4,000 publishers of online content (henceforth, the Consortium). Publishers provide information that allows the Consortium to link individuals' interactions with an article to their firm. Using a proprietary machine learning algorithm, they then predict the article's mix across topics. Interactions with a topic are aggregated across all employees within a firm at a daily frequency. This provides the Consortium with a measure of the intensity a firm is reading about a particular topic. The Consortium's primary purpose is to generate a signal of user purchase intent so that clients can more effectively direct sales, marketing and advertising dollars. As such, the topics themselves are geared towards identifying specific products or services, while changes in the topic interaction metrics reflect changes in a firm's purchase intent.

This business purpose makes the Consortium's dataset particular useful in testing our main set of hypothesis. Specifically, the model shows that firm β reflects the degree to which liquidity (risk) dominates growth option (opportunities) concerns within the firm. We thus hypothesize that firms with higher liquidity requirements will pay more attention to systematic versus firm-specific growth option related topics. Second, the model predicts that changes in β through time reflect changes to firm exposure to risk versus opportunity. Innovations to the composition of systematic versus growth option related reading should thus also reflect changes in this exposure. And finally, given the differing information sets between the firm and the marginal investor, innovations to the composition of reading should *predict* fluctuations in asset prices only during information revealing events. That is, when positive innovations to risk versus opportunity are revealed, one would expect (a) instantaneous returns to be lower, and (b) conditionally beta to increase. It is thus precisely during these times that what a firm is paying attention to will have the greatest predictability.

Our first empirical objective is to decompose the list of topics provided by the Consortium into those associated with firm-specific growth option (henceforth, sector topics) and liquidity considerations (henceforth, systematic topics). Given the large cross-section of topics captured by the Consortium, firms do not pay attention to every topic. For example, the 30th percentile firm in 2020 only pays attention to 75 of the 6,000+ topics available. If we focus exclusively on the publicly available firms in CRSP-COMPUSTAT (i.e., larger firms) the number of topics to which the 30th percentile firm pays increases to 1,945. We develop a methodology that exploits this heterogeneity to highly rank topics that occur with relatively high frequency within a sector (i.e., NAICS 3-digit level), but low frequency in the general population of firms. Visually, sector specific topics are closely related to the inputs required to produce the goods and services specific to a sector. We hypothesis that positive *abnormal* firm attention to topics important to its own sector is a strong indication of a positive firm-level growth shock that leads to an expansion of firm production.

Next we isolate from the remaining set of topics those related to liquidity. In the model, a firm's liquidity requirements are a function of either operating leverage (see Gilchrist, Schoenle, Sim, and Zakrajšek, 2017, for a similar interpretation) or external financing needs (see Whited, 2006, for microfoundations). These risks are reflected in higher covariances or beta with the market portfolio, which can be thought of as a firm-level measure of economic uncertainty. We thus start by constructing a reading list of articles for a fictitious industry that is interested in only economic uncertainty (see Baker, Bloom, and Davis, 2016). We then apply the same separation methodology above to identify those topics that have higher frequencies within this industry versus those in the general population of firms. We hypothesize that positive *abnormal* firm attention to these topics, regardless of sector, would be a strong indicator of a negative firm-level growth shock.

It's important to note, that liquidity and external financing considerations are important both when growth options are being exercised and economic uncertainty is high. That is, if tested without considering growth option, abnormal attention to liquidity related topics would have an ambiguous relationship with systematic risk of the firm. As such, it's critical that we aggregate reading in both categories into a single measure (i.e. systematic *versus* growth option related topics).[1]

We first see if our measure of firm attention has the intended interpretation. We do this by looking at the statistical relationship between the level of systematic-to-sector topic reading and firm characteristics during the Consortium sample period (2015-2019). Economically, there was little systematic variation in aggregate liquidity risk during this time. We can interpret any relationship as reflecting static differences in the presence of growth options or use of operating leverage between firms. We find that smaller firms with higher cash-flow betas, market-to-book and degrees of operating leverage spend more time reading about systematic versus sector reading. This maps directly to differences in model primitives.

Having validated our measures, we next test our model's dynamic implications. Industry specific consolidation or competitive pressures, for example, may create cross-sectional differences in cash flow growth expectations that are orthogonal to those seen market wide (see, e.g., David and Veronesi, 2013; David, 1997, for microfoundations of this behavior). Given the large cross section of firms in our database, by properly controlling for unobserved heterogeneity, we test how *innovations to* rather than *levels of* attention relate to changes in (within) firm behavior. Daily reading for each topic is aggregated to a weekly frequency (to remove seasonality), and then normalized and standardized to produce a score of unit normal distribution that captures the degree of abnormal reading versus a historical baseline. The methodology used to create a *spike* measure is similar to the reading measures in Baker, Bloom, and Davis (2016) and Ben-Rephael, Da, and Israelsen (2017). We find that within firm innovations in attention to systematic versus sector topics leads to a fall in SG&A and capital expenditure ratios. This provides further evidence of the underlying

[1] In fact, we see this empirically as well–the predictability of real economic decisions and asset prices are ambiguous when both variables are tested individually, but consistent across tests when included together or combined.

drivers of our model.

Armed with an empirical proxy for innovations in attention, we next turn to the model's asset pricing implications. Our first test is on returns around an information revealing event (i.e., earnings announcement). We find that abnormal reading in systematic versus sector topics predicts negative announcement day excess returns. This finding is robust to controls for firm characteristics normally associated with cross-sectional differences in excess returns–namely, the high-minus-low, small-minus-big, momentum and market beta risk factors. The intuition is straightforward: variations in drift rates have an idiosyncratic component such that on average the within firm drift rate is constant across the economy. By fixing each firm's exposure to sources of risk across our sample, statistically large changes in excess returns around these events reveal the information content embedded in our firm attention measures. Using this same intuition we also find that innovations to systematic versus sector topics predict high long-run risk-neutral variance and beta. These changes are also highly concentrated during the earnings announcement period, again suggesting strong learning behavior by investors in the market.

The remainder of the paper is organized as follows. In section 2, we present our model that motivates and clarifies the economic mechanisms involved in time-varying beta. In section 3 we describe the data and the construction of our empirical proxies in further detail. Section 5 reports the empirical results using our novel dataset and section 6 concludes.

2. A Conceptual Framework of the Firm

In this section, we provide a theoretical analysis of the relationship between simple, realistic firm characteristics and large, persistent changes in firm systematic risk (β). In section 5, we then test if measures derived from firm employee attention to topics related to these characteristics show similar relationships as seen in theory. We begin with a canonical model similar to that of Carlson, Fisher, and Giammarino (2004). Firms are exposed to systematic risk and have growth options and operating leverage. To this we add idiosyncratic variations in firm drift rates. This is meant to capture, in a partial equilibrium way, time varying industry, sector and product competition. This idiosyncratic variation, however, may be difficult for investors to fully decipher. Learning is then added, clarifying how and when idiosyncratic shocks will impact asset prices.

2.1. Setup of the Baseline Model

Each firm in the economy generates profits according to the instantaneous profit function

$$\Pi_{i,t} = \rho_i \cdot y_{i,t} - \rho_i \cdot c_i, \tag{1}$$

where ρ_i is the firm's sales sensitivity to systematic risk, y, and c_i is the fixed cost. The systematic state variable follows geometric brownian motion with constant drift μ_y and volatility σ_y

$$dy/y = \mu_y dt + \sigma_y dz_y.$$

All firms also have the opportunity to irreversibly expand production by γ for a fixed cost $\rho_i I_y$. Post option exercise the profit function becomes

$$\Pi_{i,t} = (1+\gamma)\rho_i \cdot y_{i,t} - (1+\gamma)\rho_i c_i.$$

The value function of the firm is given by $V(y_i) = E\int \Pi_{i,t} \cdot e^{-rt} dt$, where r is the constant discount rate. Ito's Lemma provides a differential equation that defines the evolution of the value function.

$$rV_i = \Pi_{i,t} + \mu_y \cdot y \frac{\partial V_i}{\partial y} + \frac{\sigma_y^2 y^2}{2} \cdot \frac{\partial^2 V_i}{\partial y^2} \tag{2}$$

We solve for $V_{i,t}$ by guess-and-verify, using the balanced budget requirement for external funding and the smooth pasting condition to estimate constants. The value function is defined as the sum of assets in place (V^{AY}), the expansion option (V^{GY}), and present value of fixed costs (V^F).

$$\begin{aligned} V_{i,t} &= \rho_i \left(V_{i,t}^{AY} + V_{i,t}^{GY} - V^F \right), \quad \text{where} \\ V_{i,t}^{AY} &= \frac{(1+\mathbb{I}\gamma)\, y}{r-\mu_y} \\ V_{i,t}^{GY} &= \frac{(1+\mathbb{I})\, \gamma y^*}{(r-\mu_y)\, b} \left(\frac{y}{y^*} \right)^b. \end{aligned} \tag{3}$$

where $b > 1$ is a constant. The threshold for option exercise, y^*, is defined as

$$y^* = \frac{b}{b-1} \cdot \frac{(r-\mu_y)\, I_y}{\gamma}. \tag{4}$$

This equation will become important when adding idiosyncratic regime switching drifts rates to the model. Higher drift rates, μ_y lower the threshold over which the firm will exercise the growth option. In addition, the slope of the value function with respect to the systematic state variable also changes. Using Ito's Lemma we can also estimate the systematic exposure of the firm and most importantly how it changes with the characteristics of the firm,

$$\beta_{i,t} = 1 + \rho_i (b-1) \frac{V_{i,t}^{GY}}{V_{i,t}} + \frac{V_{i,t}^{F}}{V_{i,t}}. \tag{5}$$

Right away, one can see that the beta is expected to increase as the firm size falls, fixed costs rise and growth options become a larger portion of firm value. In section 5 we first test this direct relationship using an empirically derived measure of asset β. Once validated we show that the level of systematic versus sector reading has the same relationship with size, sales β, DOL and the availability of growth options.

2.2. Regime Switching Drift Rates

We next add symmetric and idiosyncratic regime switching to cash flow drift rates at the firm level. This is meant to mimic shocks that impact firm valuations, but end up averaging out across the entire economy (e.g. intra industry competition).

$$dy/y = \mu_y(t)\,dt + \sigma_y dz_y.$$

$\mu_y(t)$ is a two-state Markov process with state space $\Theta = \{\mu_y^H, \mu_y^L\}$ where $\mu_y^H > \mu_y > \mu_y^L$. The transition probabilities of $\mu_y(t)$ over an instantaneous time interval is given by

$$\begin{bmatrix} (1-\lambda)\cdot dt & \lambda \cdot dt \\ \lambda \cdot dt & (1-\lambda)\cdot dt \end{bmatrix}$$

That is at any given moment of time half of firms will be in a high growth state and half in a low growth state. As such the average growth across the whole economy is still μ_y. This intuitive addition has the effect of creating even greater within firm heterogeneity in degrees of operating leverage and probabilities of growth option exercise by changing the *distance* the firm is from a liquidity event and the threshold over which the growth option will be exercised, respectively. As such, at the firm-level, underlying *idiosyncratic* shocks are also reflected in changes in *systematic* risk.

2.3. Investor Learning

So far firms (insiders) and investors (outsiders) have identical information sets through which to evaluate the firm, i.e.

$$\mathfrak{F}_t^{Inv} = \mathfrak{F}_t^{Firm}. \tag{6}$$

Under this assumption, especially in small samples, any relationship we find between regime changes and asset fluctuations could be miss identified (see Fama and French, 2004, for discussion). The equality of information sets between insiders and outsiders also belies anecdotal evidence of the demonstrated value that investors extract from formal or informal expert networks (see, e.g., Ahern, 2017). We address these shortcomings by adding investor learning to the model. We assume that investors do not fully know the state of firm growth at time t–that is they place some probability on the state either being high or low growth,

$$P\left(\alpha_i(t) = H \mid \mathfrak{F}_t^{Inv}\right) \neq 1 \text{ or } 0. \tag{7}$$

The asset pricing effects of learning from cash-flows would be most acute during information revealing events (henceforth, IREs) and when uncertainty of firm prospects is high. Empirically, learning confers a simple identification strategy. Insider's behavior (i.e., attention) will react immediately to idiosyncratic changes to prospects and risk. Assuming that investors are largely left in the dark after such a change, this is also when investors would be most uncertain about the state of cash flow growth. Investor reaction to news would thus be "large" when the state of growth was revealed. As a corollary, the predictability of variation in firm attention on asset prices and measures of risk to either changing liquidity needs or the execution of growth options would also be greatest immediately proceeding an IRE (e.g., after an earnings announcements or secondary stock offerings (see Hibbert, Kang, Kumar, and Mishra, 2020, for a similar identification strategy)).

3. Data and Variable Construction

Our proprietary data comes from a data provider (the Consortium) that specializes in analyzing content in internet articles published across thousands of media sites (members) to provide clients with actionable signals of *intent* to purchase specific business-to-business products and services.

Members include generalist publishers such as The Wall Street Journal, Forbes, and Bloomberg, as well as more niche providers of content such Hart (energy), Step Stone (private equity) and Quin Street (consumer products). In general, members span a wide array of industries and business activities and receive data analytic services from the Consortium for use of their raw data. Members feed information pinpointing the URL of online content and external IP addresses of the originating HTTP request ($\geq$ 15bn daily interactions with member content). In combination with cookie generated user profiles, the Consortium is able to use the IP addresses to associate a domain with a particular user. It is on this raw data that the Consortium deploys its NLP algorithm in order to better understand the content being read by firms (domains).

3.1. Daily Aggregates

The Consortium has developed a machine learning algorithm that is able to predict the topic composition of each article. Topics are word mixtures and associations that are *learned* using highly specific training corpora overlayed with exhaustive human verification. Generally topics come in two varieties: either they are (1) created for a client firm that sells a specific product or service (e.g., proteomics, circuit design, registered investment advisor), or (2) created to enhance the fit of the predictive algorithm to general variation in reading across users (e.g., political violence, vacation, best places to live). While the preponderance of topics have a business focus, a large proportion of actual reading occurs in the generalist bucket of topics.

The specific topics are of particular interest to the Consortium and their clients as their intertemporal variation provides a demonstratable signal of *intent* of the user to purchase the given product or service. All topic interactions, defined as the percentage a particular topic composes an article read by a user, is aggregated to a daily frequency across all users within a given domain. The rationale of this aggregation is simple. An article is an amalgamation of potentially many topics; the NLP algorithm deconstructs its content into a topic mix. The aggregation to a lower frequency (daily) of interactions across users (within a domain) utilizes the cross-sectional heterogeneity of article topic mixes to generate a more refined sense of a company's intent to purchase a particular good or service. This can be thought of in a simple probabilistic framework–i.e. the probability that the "sum" of interactions reflect true interest in a particular product or service can be enhanced when aggregated across N independent interactions,

$$P(\text{Intent to Purchase}|\text{Reading about Product or Service}) =$$
$$1-\prod_{i=1}^{N}(1-P_i(\text{Intent to Purchase}|\text{Reading about Product or Service}))$$

Of course this is a gross simplification given the assumption that both the unconditional probability of intent to purchase and reading about the product or service are always overlapping. This assumption is complicated by the fact that the probability of certain topics appearing within an article are highly interdependent. However, it should hold to some degree, especially as topic interactions are aggregated further in our main set of independent variables (see section 4). Finally, the Consortium applies other filters (sufficiency of topic interaction tests, bot filters, etc.) beyond the NLP algorithm to streamline the data to be as *intent* specific as possible.

3.2. Signal Extraction

A shortcoming of the daily measure is that the volume of articles and therefore topic interactions can vary across time and firms. First, intra week seasonality is substantial–e.g. weekend reading will be less than weekday reading. Second, the members of the Consortium change. Given that the preponderance of members are niche industry publications, this compositional shift can potentially distort the intertemporal behavior of any attention measure we create. Third, domains, which represent firms or business entities, have changing size in terms of number of employees represented in the data. And fourth, outages and latency issues may alter the volume of interactions documented. In many ways these issues are similar to those in the raw data used to create the Baker, Bloom, and Davis (2016) EPU measure; the Consortium uses a similar approach to construct a standardized measure of topic attention.

The aggregate daily score ($r^t_{j,d}$) for each topic j and domain d is first normalized by subtracting the mean domain topic interaction score across domains ($\hat{R}^t_{1:D}$) and dividing by its cross-sectional standard deviation ($\hat{\sigma}^t_{1:D}$).

$$\tilde{r}^t_{j,d} = \frac{r^t_{j,d} - \hat{R}^t_{j,1:D}}{\hat{\sigma}^t_{j,1:D}} \quad \text{where}$$
$$\hat{R}^t_{j,1:D} = \frac{1}{D}\sum_{i=1}^{D} r^t_{j,i} \quad \text{and}$$
$$\hat{\sigma}^t_{j,1:D} = \sqrt{\frac{1}{D-1}\sum_{i=1}^{D}\left(r^t_{j,d} - \hat{R}^t_{j,1:D}\right)^2}$$

This normalization mitigates changes in Consortium membership and algorithm, and the effects of outages and latency. It still leaves, however, the problem of comparing measures across domain-topics and changes in firm size. In addition, weekly seasonality could still be a persistent issue if attention varies across topic-days. This is addressed when constructing our attention measure for specific use cases next.

4. Topic Separation Methodology

Currently, the data represents more than 14mm domains and their interactions with over 6,000 topics (Consortium's taxonomy). As table 1 illustrates there is substantial cross-sectional and intertemporal heterogeneity in coverage across the dataset. First, the number of domains and topics represented has increased from 638k to 1.52m and 2.46k to 6.97k, respectively. These figures justify the steps taken by the Consortium to standardize scores through time. Second, the distribution of the number of topics paid attention to by individual firms is highly positively skewed. That is the median firm pays attention to only a few topics, while some firms pay attention to many, but still very far from *all* topics. If we focus only on COMPUSTAT/CRSP firms–the set of domains over which our main set of regressions are performed–we see that the skewness falls. This indicates that this skewness is largely driven by size.

The objective of this section is to develop an algorithm that separates the topics into those that represent the primitives of the model. In the model, results hinge on two firm characteristics: operating leverage and growth options. The *relative* importance of both drive key variations in firm systematic risk. Our hypothesis is that firm attention to the products and services related to liquidity events (due to a high degree of operating leverage) relative to those related to the expansion of core business activity will provide a high frequency proxy of changing exposure to each characteristic and thus capture key variation in firms' changing systematic risk.

4.1. Sector Topics

The heterogeneity across firms described above helps us separate core business related topics from those of general interest and those related to other businesses. Specifically, we first compute within a sector s for a given week t the number of domains that pay attention to topic j,

$$F_{j,t}^{s} = \sum_{d=1}^{D_s} \mathbb{I}_{\left(r_{j,d}^{t} > 0\right)},$$

where the indicator function $\mathbb{I}_{\left(r_{j,d}^{t} > 0\right)}$ equals 1 if and only if $r_{j,d}^{t} > 0$ and D_s is the total number of domains (associated companies) in sector s. Similarly, let $F_{j,t}$ denote the number of times topic j appears with non-zero score in the entire cross-section of domains. Equivalently,

$$F_{j,t} = \sum_{d=1}^{D} \mathbb{I}_{\left(r_{j,d}^{t} > 0\right)}.$$

We then denote a sector ranking or weighting for each topic j. This weight is obtained as

$$w_{j,t}^{s} = \frac{\tilde{w}_{j,t}^{s}}{\sum_{d=1}^{D_s} \tilde{w}_{j,t}^{s}} \quad \text{where}$$
$$\tilde{w}_{j,t}^{s} = \left(\frac{F_{j,t}^{s}}{F_{j,t}}\right)^{\alpha} \cdot F_{j,t}^{s}.$$

The goal of the algorithm is to penalize topics that are being paid attention to by a relative large number of domains in the full cross-section of domains, but only by a small number of domains within a particular sector (NAICS 3-digit industry). A simple expositional example illustrate the approach. Assume there are 100 firms of which 10 are in the crop production sector (NAICS 111). Two topics within the taxonomy– Vacations and Tillage– are of obvious general and sector specific interest. As such all 100 firms pay attention to *vacations*, but only the 10 firms pay attention to tillage. Applying an α of 1, $\tilde{w}_{vacation,t}^{crop} = (10/100)^{1} \cdot 10 = 1$ and $\tilde{w}_{tillage,t}^{crop} = (10/10)^{1} \cdot 10 = 10$. The algorithm thus tilts the weight or ranking of topics towards those (products and services) that are most associated with either inputs or outputs of the particular sector or industry. As the ranking is dependent on α, rather than constraining the algorithm to a particular parameter value, we use the

Newton-Raphson method to solve for the parameter using the following constraint,

$$\left(\sum_{j=1}^{N} w_{j,t}^{s}\right) - q\% = 0.$$

That is, we choose α such that the sum of $w_{j,t}^{s}$, sorted in descending order, over the first N topics within a sector composes q percent of the weight. For our main set of results we use $N = 100$ and $q = 90$, but results are robust to reasonable values of N and q. By reasonable we mean they are bounded far enough from q being 100% and N representing *all* topics in the taxonomy. Figure 1 are word clouds representing the topics with the largest 100 weights within various sectors. The differential sizes of the topic name represent their relative rank. As one can see, the algorithm does a good job of isolating very sector specific topics, which look like products or services important to the sector in question. This is despite some of the topics not being a large share of either the COMPUSTAT/CRSP or Full sample. Table 2 is a list of the top 30 industries by either employment share or firm count share. As one would expect, the full sample does a better job of capturing the full scope of economic activity in the economy versus the exclusively public firm sample of CRSP/COMPUSTAT.

4.2. Systematic Topics

Liquidity related or systematic topics on the other hand will be of a general interest variety. By this we mean all firms with non-zero exposure to the market will be concerned about the effects of these shocks on their bottom line and by extension will pay some attention to related topics. Of the remaining non-sector specific topics, we attempt to distinguish between general "nuisance" topics, e.g. "vacations", and general "systematic" topics, e.g. "liquidity risk", by creating a fictitious sector which pays attention to *only* macroeconomic uncertainty.

We create a macroeconomic corpus of ≥2,500 articles. The corpus includes articles from 6 publishers from 2005-2019: the Wall Street Journal (WSJ), Economist (segmented by 6 section tabs), Financial Times (FT), Federal Reserve Beige Books (segmented in 12 regions), Federal Reserve Notes, and Bank of International Settlement publications. For the WSJ, Economist and FT we follow Baker, Bloom, and Davis (2016) in searching for all articles with the terms *economic*, *uncertainty*, and either *Congress*, *legislation*, *regulation*, or *White House*. In many ways the level of reading of these topics can be thought of as reflecting firm exposure to economic and policy uncertainty.

The Consortium's proprietary algorithm was then deployed onto the articles, generating the predictions of the percentage that each article is of a particular topic. This was aggregated up to each publication-type (22 when including segments). The separation methodology described above (where $F^s_{j,t} = F^\beta_{j,t}$) then used to capture topic saliency to our macroeconomic industry. Figure 2 is the word cloud representing the topics with the largest 100 weights within the fictitious macroeconomic industry. As one can see, liquidity, macroeconomic uncertainty and macro oriented institution related topics are well represented. The basic idea is that if a firm is paying relatively more attention to these topics it must be that their concern or exposure to systematic risks have increased.

4.3. Attention Factors

We then aggregate topics at the firm level across these two topic buckets: systematic and sector. As our interests are in both the level and change in systematic versus sector reading, we construct measures that reflect extensive and intensive margins of attention, respectively. In both cases it's important to note, given that all topics are not represented within each domain, that 250 systematic topics are available for nearly all domain-week observations with our COMPUSTAT/CRSP sample. Furthermore, as noted above, normalized scores are not comparable through time or across domain-topics. To address this, we compute the precentile of the sum of current topic scores (extensive margin) and current topic score versus some historical baseline (intensive margin). This then allows us to compare scores across time and firms.

For extensive attention, we first sum standardized scores ($\tilde{r}^t_{j,d}$) across the top 250 systematic and sector-specific topics available within a given domain-week, which removes weekly seasonality. This domain sum is then compared with the cross section of sum of scores across the COM-

PUSTAT/CRSP sample within a given week for each topic cluster type.

$$F_R\left(\bar{r}^t_{j,d}\right) = \int_{-\infty}^{\bar{r}^t_{j,d}} f_R(r)\,dr = \frac{C_{r \le \bar{r}^t_{j,d}} + \frac{C_{\bar{r}^t_{j,d}}}{2}}{N}, \tag{8}$$

where j is defined as either sum of systematic or sector reading scores, C represents the number of observations in the cross-section that are below a given firm's sum of score ($\bar{r}^t_{j,d}$). The extensive attention measure is then simply the logarithm of $F_R\left(\bar{r}^t_{j,d}\right)/F_R\left(\bar{r}^t_{j,d}\right)$. Note that this construction implicitly controls for any size effect. For example, large firms will have both large numerators and denominators, and, as such, we are measuring relative not absolute reading of systematic versus sector reading.

For the intensive measure, on the other hand, we are interested in within-firm variation of attention. The baseline measure is therefore scores across a historical time interval rather than a cross-section of firm scores. We first compute a comparison of the *current interval* (i.e., rolling 21 day average of normalized scores) to *baseline interval* (84 days prior) of the same domain-topic score. Equation 8 is used again to estimate the empirical cumulative distribution function, where $\bar{r}^t_{j,d}$ is now the standardized score ($\tilde{r}^t_{j,d}$) and C represents the number of score observations in the baseline interval that are below the current interval score. This weekly domain-topic measure captures *positive* innovations (or anomalies) to domain attention to a particular topic. Unfortunately, in the time series, due its rolling nature, the score also demonstrates a higher probability of $\leq$50 score in the future as the current interval moves into the baseline interval. Our intensive attention measure therefore focuses exclusive on right tail events,

$$A^{\beta}_{k,t} - A^{s}_{k,t} = \left(\sum_j \left(\frac{1}{250}\right) \cdot \mathbb{I}_{(x_{k,j,t} \ge \bar{x})}\right) - \left(\sum_j \left(\frac{1}{250}\right) \cdot \mathbb{I}_{(x_{k,j,t} \ge \bar{x})}\right).$$

That is we subtract the average number of spiking sector scores from spiking systematic scores in a given week. For our baseline results $\bar{x} = 80$. This measure is computed across the top 250 sector and systematic topic scores for each domain-week. In the next section we show that our measures, which relate to the *composition* of attention and its changes, provide important information about the level and time variation in firm risk.

5. Empirical Results

We first show that the cross-section of firm characteristics relates to the extensive margin of attention with similarly directionality as an empirically derived measure of asset β. This exercise can be thought of as validating our attention measure as a proxy for a firm's systematic risk exposure. We then relate intertemporal changes in reading (intensive margin) to changes in firm behavior (real outcomes). Finally, allowing for the possibility of information frictions between investors and insiders (employees within the firm), we test whether and when our attention measure predicts returns and innovations to variances and βs.

5.1. *Measure Validation*

The predictions from the canonical model can be visualized in equation 5: asset β should be inversely related to firm size ($V_{k,t}$), while directly related to DOL ($V^F_{k,t}/V_{k,t}$), availability of growth options ($V^{GY}_{k,t}/V_{k,t}$) and firm sales sensitivity to aggregate shocks (ρ_k). We use empirical proxies to test for these relationships.

Our theoretical model ignores leverage, focusing on asset β. We follow Doshi, Jacobs, Kumar, and Rabinovitch (2019) in delevering within quarter return data by using the total liabilities available in COMPUSTAT, $r^A_{k,t} = r_{k,t} \times \left(1 - L_{k,t}/\left(L_{k,t} + P_{k,t} \times \text{Shr Out}_{k,t}\right)\right)$. We estimate asset β (β^A) using daily delevered returns regressed onto daily delevered market returns.

$$r^A_{k,t} = \alpha_k + \widehat{\beta}^A_k \cdot r^A_{m,t} + \varepsilon_{k,t}$$

Market returns, $r^A_{m,t}$, are estimated by asset weighting (mark-to-market) delevered returns for our universe of stocks. Sales sensitivities are similarly derived from regressing sales growth for firm k onto aggregate sales, $\Delta Sales_{k,t} = \widehat{\rho}_k \Delta Sales_{agg,t} + \varepsilon_{k,t}$. This is estimated using quarterly sales data on a 10yr rolling basis. Growth options are proxied by the log of Tobin's Q,

$$Q_t = \frac{P_{k,t} \times \text{ShrOut}_{k,t} + \text{Prefs}_{k,t} + \text{Curr Liab}_{k,t} - \text{Curr Assets}_{k,t} + \text{LT Debt}_{k,t}}{\text{Assets}_{k,t}}. \tag{9}$$

Market capitalization of asset is proxied as the log of the numerator of equation 9. Finally, following the work of Gilchrist, Schoenle, Sim, and Zakrajšek (2017), we proxy for operating leverage by the log of the liquidity ratio which is defined as cash and cash equivalents from the balance sheet divided by total assets. The intuition of this proxy follows the finding of Lins, Servaes, and Tufano (2010)–non-operating cash holdings are largely driven by a desire of firms to hedge against future negative cash flow shocks. As the liquidity of firms with high degrees of operating leverage (DOL) is more likely to be impacted by these types shocks, one would conjecture that high DOL firms will also hold relatively more liquidity in the cross-section.

We regress $\beta^A_{k,t}$ onto each of these proxies,

$$\widehat{\beta}^A_{k,t} = b_1 \cdot \widehat{\beta}^s_{k,t} + b_2 \cdot \mathrm{Q}_{k,t} + b_3 \cdot \mathrm{Capitalization}_{k,t} + b_4 \cdot \mathrm{Liquidity}_{k,t} + \varepsilon_{k,t}. \tag{10}$$

Results are presented in table 3 panel A. It's important to note that both independent and dependent variables demonstrate a great deal of persistence. We therefore cluster standard errors by both firm and time. Unsurprisingly, higher sales sensitivity translates to higher β^A. The liquidity ratio as well is positively related to β^A. Viewed through the lens of cash and cash equivalents being held as a hedge against potential negative shocks, this relationship is very intuitive.

Tobin's Q is also positively related to β^A. This is in keeping with the model findings. Valuations of firms with near or in the money growth options have implicit leverage–their value includes exposure to yet to be exercised growth options while their physical asset base is relatively small. This induces higher asset weighted exposure to systematic risks.

Unfortunately, capitalization is positively, not negatively–as suggested by our model–related to β^A. This, however, seems to be entirely related to observations in the lowest decile of size. In figure 3 panel A, we split size into three buckets–the lowest 10th precentile, the 10-55th percentile and the 55-100th percentile–and rerun the regression specification.

$$\widehat{\beta}^A_{k,t} = b_1 \cdot \widehat{\beta}^s_{k,t} + b_2 \cdot \mathrm{Q}_{k,t} + \sum_{j=1}^{3} b_{3,j} \cdot \mathbb{I}^{cap}_{k,t} + b_4 \cdot \mathrm{Liquidity}_{k,t} + \varepsilon_{k,t}, \tag{11}$$

where $\mathbb{I}^{cap}_{k,t}$ is a dummy variable that takes a one if the capitalization of firm k is in one of the three buckets on week t. As noted by the simple p-values from Wald Statistics comparing coefficients between buckets within the figure, the sensitivity of β^A to size actually falls by a statistically significant amount from the 10-55th percentile to 55-100th percentile buckets. As β^A is a generated quantity, we believe that the non-linear, hump shaped profile exhibited in 3 panel A may relate to estimation error.

In the fifth column of table 3 panel A, we add all four independent variables to the same regression. As noted in Lins, Servaes, and Tufano (2010), the liquidity ratio while driven primarily by hedging needs, is also related to the need for easy access to expansion capital. This may confound the interpretation of the liquidity ratio as a measure of operating leverage; in this specification we place the residual from regressing the log liquidity ratio onto $\rho_{k,t}$, Tobin's Q and capitalization as our DOL proxy.

5.2. *Attention and Firm Characteristics*

In this section, we show that the set of cross-sectional firm characteristics described above relate to the level of macro versus sector reading with signs similar to those of regression 10. We regress the read ratio onto our proxy for sales β, availability of growth options, size and DOL:

$$\log\left(F_R\left(\sum r^t_{\beta,d}\right)/F_R\left(\sum r^t_{s,d}\right)\right) = b_1 \cdot \beta^s_{k,t} + b_2 \cdot Q_{k,t} + b_3 \cdot \text{Capitalization}_{k,t} + b_4 \cdot \text{Liquidity}_{k,t} + \varepsilon_{k,t}. \tag{12}$$

Note that $F_R\left(\sum r^t_{j,d}\right)$, defined by equation 8, is the percentile of a firm's total reading in either sector or systematic topics versus the distribution of all CRSP-COMPUSTAT firms on a given day. Given the findings from our theoretical model, we would expect b_1, b_2, b_3 and b_4 to be positive, positive, negative and positive, respectively. Results are presented in table 3 panel B. Similar to β^A, the read ratio demonstrates a significant amount of intra-firm time series persistence. We thus cluster standard errors by both firm and time.

All firm characteristics correlate with the read-ratio as expected, including size. In figure 3 panel A, we perform the same size split as represented by regression 11. A dummy variable is created that is a one if firm k's size falls into one of three buckets–the lowest 10th percentile, the 10-55th percentile and the 55-100th percentile–during a given week. *P*-values from Wald statistics testing whether $b_{3,j}$ are statistically different between buckets, shows that the relationship is now strongly monotonically decreasing. As the data comes from directly observing what firm employees are reading, we believe our proxy reduces the risk of estimation error effecting statistical inference.

We conduct one more piece of cross-sectional analysis to verify that the read-ratio is partially related to DOL. In the series of regressions presented so far, the liquidity ratio is our DOL proxy, which, as noted, is at best an inferred proxy. We therefore run the classic degree of operating leverage regression,

$$\Delta \log EBITDA_{k,t} = a_{0,k} + \sum_{j=1}^{3} a_j \cdot \mathbb{I}_{k,j}^{RR} \cdot \Delta \log Sales_{k,t} + \varepsilon_{k,t}. \tag{13}$$

We first average the read-ratio for firms across our sample (2015-2020) and then place each firm into terciles based on the average level of systematic versus section reading. We then regress, on annual data, the change in EBITDA (operating income plus non-cash depreciation) versus change in sales. Due to seasonality the regression has less clear results (e.g. degrees of operating are on average less than one across the sample) when using quarterly data. Following Eisfeldt and Papanikolaou (2013) and others, we are allowing for heterogeneity in DOL by firm characteristics– in our case differences in the average read-ratio. In figure 4 we present the three a_j from equation 13 and test if they are different from one another. Given that our terciles are statitic through time, we can test the DOL over longer samples. In figure 4 panel A and 4 panel B we run the sample over the regression from 2014 to today and 2015 to today, respectively. Higher average read-ratio firms have a statistically significant 25% higher DOL.

Finally, as noted there is a significant amount of persistence in both the dependent and independent variables in both sets of regressions. Furthermore, the relationship between β^A and firm characteristics in the model is largely cross-sectional. States of growth are also highly persistent and our data only covers roughly 5yrs. We should therefore expect, with the right sets of fixed effects, that the relationships highlighted in table 3 should substantially attenuate. In table 4 we sequentially add both time and firm fixed effects to model 5 from table 3. The first column of each dependent variable (β^A and the read-ratio), is the original regressions with time and firm clustered standard errors, and no fixed effects. In the second column we add time fixed effects. This seems to have little effect on the results. In the third column we add firm-level fixed effects. The relationships disappear or greatly diminish. This is a strong indication that the construction of the independent variables and the relationships we see to are almost entirely cross-sectional.

5.3. *Attention and Firm Dynamics*

In the previous section we show that cross-sectional differences in the level of systematic versus sector reading reflect differences in exposure to risk across firms. In our theoretical framework negative *shocks* to trend growth rates lead to an increase in systematic risk and a drop in investments within firms. We would therefore also expect that intertemporal changes in systematic versus sector reading would reflect within firm changes in exposure to risk.

We test for this possibility by regressing cumulative investments at different horizons h onto our intensive measure of attention

$$I_{k,t+h} = \beta \cdot \sum_{w=1}^{W_t} \left(A^{\beta}_{k,w,t} - A^{s}_{k,w,t} \right) + \text{controls}_{k,t} + \varepsilon_{k,t}. \tag{14}$$

$A^{\beta}_{k,w,t} - A^{s}_{k,w,t}$ is the average number of systematic versus sector topics spiking versus a historical average at any given week w and quarter t (see section 4 for further details). The intuition is straight forward–as firms face a negative shocks to cash-flows their focus changes from efforts to exercise growth options towards those to shore up liquidity due to fixed costs. Intertemporal changes in the composition of attention should therefore tell us something about changes in investments. As income statement information is only available on a quarterly basis, we average $A^{\beta}_{k,w,t} - A^{s}_{k,w,t}$ over the quarter and test for both contemporaneous and long-run predictability in investment behavior.

To capture within firm effects all regressions include firm fixed effects. Our theoretical framework also makes the distinction that variations in trend growth have both a systematic and idiosyncratic component. While this is critical when we test for the asset pricing implications of attention in section 5.4, our dynamic investment regressions also require some version of time fixed effects. Furthermore, we must control for heterogeneity in firm investment seasonality. We add NAICS 3-digit by quarter fixed effects to control for both aspects.

Our two main independent variables, $I_{k,t+h}$ in equation 14, are capital expenditures (CapEx) and selling, general and administrative expenses (SG&A). We normalize both by historical balance sheet items–one quarter lagged assets, and implied organizational capital for CapEx and SG&A,

respectively. This is common in the literature and done to minimize the impact of time-varying firm size on investment flows (see, inter alia, Welch and Wessels, 2000). We follow Eisfeldt and Papanikolaou (2013) in constructing a stock of organizational capital, $O_{i,t}$. Specifically, we assume a law of motion for any quarter t of

$$O_{k,t} = (1-\delta_O)\cdot O_{k,t-1} + \text{SG\&A}_{k,t}.$$

δ_O is depreciation of organization capital. $\text{SG\&A}_{i,t}$ is available in COMPUSTAT. We estimate organization capital at time t based on an initial stock which we assume to be an infinite backwards sum from the first period SG&A is available, i.e. $O_0 = \text{SG\&A}_{k,1}/(g+\delta_O)$. We use parameter estimates for δ_O and g of 30% and 10% annually. g is the average growth rate of SG&A across our sample. It is important note that the depreciation rate we use is larger than the 15% used in Eisfeldt and Papanikolaou (2013). Recent work by the BEA suggests that rates of R&D depreciation are consistently greater than 25% across most industries (see Li and Hall, 2020). These parameters are adjusted to our quarterly data. SG&A is then divided by $O_{k,t-1}$; this not only maintains consistency with the idea of normalizing flows by a measure of investment stock, but also makes the variable more stable when estimating our vector autoregression (VAR) below.

Our regression results are presented in table 5 for five different horizons h: contemporaneous and then cumulative investments 1-4 quarters ahead. All variables are winsorized at the 1% level. Economic magnitudes are difficult to compare between regressions using the stated coefficient values. Quarterly CapEx/Assets has a standard deviation of around 0.75% and average $A^{\beta}_{k,t} - A^{s}_{k,t}$ has a standard deviation of around 1.2%. This implies that a $+1\sigma$ spike systematic versus sector reading predicts a cumulative 0.06σ drop in $CapEx_t/Asset_{t-1}$ over one year. Similarly normalized SG&A has a demeaned standard deviation of around 2.5%, which implies a cumulative 0.08σ drop in SG&A over one year from a $+1\sigma$ spike in systematic versus sector reading.

These magnitudes may seem small; a VAR, however, would provide a stronger interpretation of the long-term, cumulative effects to investments from a shock to attention. We run the following panel VAR specification with $J = 4$, which removes seasonal variation in the investment variables.

$$Y_{k,t+1} = \mu_k + \mu_t + \sum_{j=1}^{J} \Phi_j \cdot Y_{k,t-j} + \sum \cdot u_{k,t+1} \quad \text{where} \tag{15}$$

$$Y_{k,t} = \begin{bmatrix} A^{\beta}_{k,t} - A^{s}_{k,t} & I_{k,t} \end{bmatrix}$$

Table 5 indicates both a contemporaneous and predictive relationship between shocks to attention and investments. This is due to differing frequencies of our data; attention variables are available weekly whereas investment data is quarterly. In order to analyze the dynamic impact of a random disturbance on the system of variables we utilize a Cholesky orthogonalization of our VAR. Given the variable ordering implied in equation 15, the way in which we aggregate attention over a quarter is critical. Our approach is to use a Barlett weighted average over the *forward* N weeks,

$$A^{\beta}_{k,t-1} - A^{s}_{k,t-1} = \sum_{n=1}^{N} \frac{\omega(n) \cdot \left(A^{\beta}_{k,n,t} - A^{s}_{k,n,t}\right)}{\sum \omega(n)} \quad \text{where}$$

$$\omega(n) = \begin{cases} \frac{2n}{N}, & \text{if } 0 \leq n \leq \frac{N}{2} \\ 2 - \frac{2n}{N}, & \text{otherwise.} \end{cases}$$

The aggregate attention measure for quarter $t-1$ thus actually contains data from quarter t. We set N equal to weeks, which is approximately 1 quarter of attention data. Given that the past quarter attention proxy is using data from the current quarter–where attention 6-7 weeks before the end of quarter is weighted most heavily–we perserve the possibility that investments can respond to a *pseudo*-contemporaneous structural shock to attention. Following Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bover (1995), we utilize forward demeaning and cross-sectional means to control for firm and time fixed effects, respectively.

The impulse response functions (IRFs) are presented in figure 5. First a simple Granger Causality test verifies the statistically plausible direction of causality–changes in attention Granger cause changes in investment behavior. Second, the cumulative fall in investments is similar in magnitude for both CapEx/Assets and SG&A/Organiational Capital, although given that SG&A is a larger portion on average of Organizational Capital than CapEx is of Assets, it's fall in percentage terms is less pronounced. Cumulatively the orthogonalized IRFs predict a fall of more than 0.3σ in

CapEx/Assets and 0.15σ for SG&A over 10 quarters. The IRFs capture the intuition that while the initial shock to investment is small, it is long lasting, accumulating to a much lower baseline over time.

Finally, we test if the primitive shocks we claim induce changes in attention are reasonably represented in the data. In our theoretical framework idiosyncratic shifts in trend cash-flow or sales growth shift the risk profiles of firms. Granger causality tests of the VAR represented by equation 15 when investments are substitute with quarterly sales growth suggest this exact direction of causality.[2] That is, whereas changes in attention seem to *predict* changes in investment, changes in sales growth Granger cause changes in attention. This suggests flipping the ordering when estimating our orthogonalized IRF,

$$Y_{k,t} = \begin{bmatrix} \Delta \text{Sales}_{k,t} & A^{\beta}_{k,t} - A^{s}_{k,t} \end{bmatrix}. \tag{16}$$

We present the IRF in figure 6. It is clear that positive shocks to sales growth induce a fall in attention. However, the rebound of attention spike is also quick. Given the construction of our attention variable (21-day average versus previous 84-day baseline), one would expect a reversal over 2-3 quarters as the current data moves into the historical. Nonetheless, a standard deviation shock to sales induces a cumulative drop in our attention variable of roughly 0.1σ over 5 quarters.

5.4. *Asset Pricing Implications*

Our final empirical work tests the premise that shifts in firm-level attention provide a novel signal of firm state variables to investors. As noted in section 2, our analysis thus far does not assume a separation in information sets between investors and the firm. There are two reasons to include this separation. First, a world with no separation belies the tremendous efforts that investors have historically, and assuming rationally, placed in learning about firm prospects before others. Second, statistical identification of any relationship between changes in attention and asset prices would be difficult given the short sample size of our attention proxies. Separation allows for the possibility of

[2]Running regression 14 with sales growth as the dependent variables, assumes that changes in macro versus sector reading *predict* changes in sales growth. This produces directionally correct, but statistically inconclusive, results.

discrete points in time, such us earnings announcements, where information is revealed to investors. This provides a natural experiment for us to test our theoretical implications by comparing asset prices at times before and after such points of revelation.

One implication of our model is that negative idiosyncratic shocks to trend growth lead to a fall in asset prices. If we layer on investor learning, while we would expect an immediate change in the composition of attention by the firm, investors reaction to the news would be delayed. In fact, asset prices, which are determined by the marginal investor, would react to changes in the firm's composition of attention specifically around periods of time when this information is revealed.

Our first set of return analysis is portfolio sorts that test the medium-term predictability of excess returns around our information revealing event, earnings announcements. We first stack all firms on their earnings date for a given quarter. Firms do not all have the same earnings announcement date or fiscal year end. This process therefore *staggers* the implementation of any plausible trading strategy within a quarter. We refer to this new timeline as the trading calendar where $t = 30$ is the date of earnings announcement. We rank each firm for a given trading quarter into deciles based on their $t = 25$ (i.e., five trading days prior to earnings announcement) spike in systematic versus sector reading, $A^{\beta}_{k,25} - A^{s}_{k,25}$. We analyze the evolution of daily returns for a quarter around the trading calendar earnings date. Under the assumption that announcements are evenly spaced $t = 0$ and $t = 60$ are trading dates equidistant from the previous versus current and current versus future quarter announcement, respectively. We then average the return of all stocks within a decile across all trading windows of which we have roughly 20 (5yrs, 4 qtrs/year). The tenth decile stocks (relatively high systematic reading spikes) is considered short while the first decile stocks (relative high sector reading spikes) are considered long. Standard errors are estimated on each trading day across all stock-quarters. Cumulative returns and standard errors are then estimated assuming intertemporal independence, which is common in the event-study literature.

We present this analysis in figure 7. We have approximately 500 stocks in each portfolio for each trading quarter. The aggregation step assumes that with so many stocks the portfolios would have minimal residual idiosyncratic risk. In addition, while we control for this more rigorously in a regression framework below, it is unlikely that the composition of portfolios would be that different from one another because the portfolio sorting variable, $A^{\beta}_{k,t} - A^{s}_{k,t}$, is already normalized and standardized by within time and firm level topic reading. The standard errors allow us to test whether across trading time the excess return spread between first and tenth decile (factor) portfolios are statistically different than zero. As one sees in figure 7, the factor return is not statistically different than zero before portfolio formation (dashed line), but becomes consistently and statistically positive within 15 days or three trading weeks after. This confirms that there is little pre-trend activity although there seems to be some "leakage" of information before actual announcement. The spread, although not statistically significant, begins to turn higher a week or two before earnings announcement. The spread is also driven by both components of the portfolio; included in the graph are the average returns over market of both the high systematic and high sector reading portfolios. The bifurcation of returns around earnings is clear in both the equal and value weighted portfolios.

We next formalize this non-parametric analysis with regressions, which allow us to test for predictability of returns around earnings while controlling for the usual asset pricing factors,

$$r_{k,t+1} - r_{f,t+1} = E_t\left[r_{k,t+1} - r_{f,t+1}\right] + \sum_{b=1}^{B} \beta_b \cdot \mathbb{I}(b) \cdot \left(A^{\beta}_{k,t} - A^{s}_{k,t}\right) + \varepsilon_{k,t+1}. \tag{17}$$

We aggregate returns over the risk-free rate to a weekly panel, matching the frequency of our attention variables. We further split the predictive sensitivities into B buckets. In our main specification there are three: three weeks before earnings ($[-3,0)$), three weeks after ($[0,+3)$) and all other times. Given our theoretical results, we would expect little predictability during normal times. Information on idiosyncratic variations in trend growth rates within a firm are unlikely to filter to the marginal investor unless formally revealed by the firm. As such, we should see, conditional on macro reading spiking versus sector reading, a lower return in the post-earnings versus pre-earnings bucket. Our null hypothesis therefore is that $\beta_{[-3,0)} \leq \beta_{[0,+3]}$–i.e., that spikes to systematic versus sector reading predict higher returns post- versus pre-earnings.

Table 6 panel A presents the results of regression 17. In columns 1-4 we limit the sample to those stocks that have listed options. This is done in order to maintain the same sample of firms as those in our variance and risk neutral β regressions below. Given that smaller stocks tend not to have listed options this analysis can be thought of as approaching the value weighted non-parametric analysis above. In column 1 we do not include indicators for intra-quarter time periods. It is clear that over the full sample lagged attention spikes have little predictive power. In column 2 we control for the market risk factor. While both pre- and post-earnings period indicators do not individually show predictive power, our null hypothesis that spikes to systematic versus sector reading predict higher returns post-earnings is rejected at the 10% level. This continues to be true after adding SMB, HML, and UMD risk factors. In column 5 we show that results, while quantitatively little changed, have become statistically weak using the full sample and all return factor controls. As noted in figure 7, there seems to be *leakage* of information into returns in the week prior to earnings. In table 6 panel B we include this week in the post-earnings indicator and test if $\beta_{[-3,-1)} \leq \beta_{[-1,+3]}$ using regression 17. With this additional week, results are stronger across the board. In particular, the regression specification in column 5, which includes the full sample and all risk factors, statistically rejects our null hypothesis. Overall, results are quantitatively similar to those presented in figure 7. The standard deviation of weekly attention spikes is around 1.6%, which implies that a one-standard-deviation spike predicts $\sim$20pbs higher returns across the three weeks post earnings announcement; a long-short portfolio would thus produce approximately $\sim$40bps. This return could be enhanced if portfolio formation took place a week before.

The second and perhaps more direct implication of our model is that negative idiosyncratic shocks to trend growth lead to higher firm-level risk. Unfortunately, shifts in high frequency volatility are difficult to identify using returns as it requires a lengthy window for estimation. We therefore turn to options data, where forward looking (risk neutral) measure of volatility risk are easily available. We follow Bakshi, Kapadia, and Madan (2003) in constructing a firm-level measure of expected variance using options data on any given day t,

$$V = \int_{S(t)}^{\infty} \frac{2\left[1-\ln(K/S(t))\right]}{K^2} \cdot C(t,\tau;K)\,dK + \int_{-\infty}^{S(t)} \frac{2\left[1+\ln(K/S(t))\right]}{K^2} \cdot P(t,\tau;K)\,dK. \tag{18}$$

K is the strike of either call ($C(t,\tau;K)$) or put ($P(t,\tau;K)$ options of expiry date $t+\tau$ and $S(t)$ is the current (spot) stock price. Equation 18 captures the idea that investors' exposure to a squared return contract is a function of the probability weighted expected return squared across *all* possible share price values. One can then back out these values from an infinite string of options in the positive (calls) and negative (puts) return domains–i.e. the volatility surface. Following Buss and Vilkov (2012), we discretize equation 18 over moneyness or K/S values from 0.33 to 3 by increments of 0.01 and use the volatility surface files from Option Metrics to construct a daily estimate of variance contract strikes across all stocks.

We then test whether our measure of attention spikes predict next period variation in risk neutral variance.

$$\left(\sigma_{k,t+1}^{Q}\right)^2 = \omega_i + \beta_Q \cdot \left(\sigma_{k,t}^{Q}\right)^2 + \sum_{b=1}^{B} \beta_b \cdot \mathbb{I}(b) \cdot \left(A_{k,t}^{\beta} - A_{k,t}^{s}\right) + \varepsilon_{k,t+1}. \tag{19}$$

This regression specification can be thought of as a panel GARCH where attention spikes are a *proxy* for the square of random return residuals. Another benefit of using risk-neutral measure of forward variance is that we can use various expiries, which will give us some sense of the persistence of shifts in trend growth. Due to the underlying mean reverting (i.e., towards α) process, if long dated options show a great deal of predictability around a revealing event, it means that the information revealed changes investors long-run expectation of firm prospects.

Table 7 panel A presents the results of regression 19 on forward variance of 1 month, quarter, semi-annual, and annual frequencies. In column 1 we look at predictability over the whole sample, which is marginally positive. In columns 2 we split the quarter into three buckets: 3 weeks before and after earnings, and all other weeks. There are a few points of note. First, the majority of predictability of attention spikes occurs in the period around earnings not outside. This is intuitive–spikes in systematic versus sector reading would be related to a spike in operating leverage risk. Second, as one would expect the degree of predictability decays as the forward expiry of expected variance increases. This is in keeping with intuition of mean reversion in our model. Third, focusing exclusively on the 1 year forward variance regression in column 5, the increase in *long-run* variance from a one-standard deviation increase in lagged attention is quite small. Weekly systematic versus sector spikes has a standard deviation of 1.6%, while individual firm demeaned standard deviation of 1 year variance is around 0.10. This implies that a one-standard-deviation higher attention spike score predicts only a roughly 0.01σ higher variance over the next year.

This perceived small effect can be driven by two potential issues. First, the post earnings indicator is the average effect of predictability over three weeks. As we will shortly show the predictability is highly concentrated in the first week; as we're averaging over three weeks this could be fractioning out the effect. More importantly, however, we should distinguish between predictability of the components of variance risk. The findings from table 7 panel A can be interpreted as specific to total risk–that priced (systematic) and that not priced (idiosyncratic). As the regressions we are running are done on a firm-panel basis, idiosyncratic risk may overwhelm any predictability of the sort that matters for pricing. This is also important in the context of our theoretical motivation; in the model while the underlying shock to cash-flows that we are interested in are *idiosyncratic*, their effect is to change the *systematic* risk exposure of the firm. This is similar in spirit to the compositional effects of idiosyncratic shocks rigorously analyzed in Babenko, Boguth, and Tserlukevich (2016).

Systematic risk, henceforth β_k, is the covariance between the returns of stock k and the market m divided by the variance of market. The covariance term can be further split into the variance of k and all index components j, as well as the correlation between k and j.

$$\beta^{Q}_{km,t} = \frac{\sigma^{Q}_{km,t}}{\left(\sigma^{Q}_{m,t}\right)^{2}} = \frac{\sigma^{Q}_{k,t} \cdot \sum_{j=1}^{N} w_j \sigma^{Q}_{j,t} \rho^{Q}_{ij,t}}{\left(\sigma^{Q}_{m,t}\right)^{2}} \tag{20}$$

We have estimates of $\sigma^{Q}_{j,t}$ from equation 18, but no estimates for $\rho^{Q}_{ij,t}$. We follow Buss and Vilkov (2012) in parametrically estimating this measure. Specifically, we assume that the relationship between risk-neutral and physical correlation is

$$\rho^{Q}_{ij,t} = \rho^{P}_{ij,t} - \alpha_t \cdot \left(1 - \rho^{P}_{ij,t}\right). \tag{21}$$

This equation allows for differences in pairwise correlation between stocks, but only an average correlation premium across index constituents. Buss and Vilkov (2012) highlights various regularity conditions that must be satisfied for equation 21 to work, which were verified for our sample and specification. We estimate physical correlation ($\rho^{P}_{ij,t}$) as the 250-day moving average correlation between stock i and j. Given equations 20 and 21 α_t can be estimated as,

$$\alpha_t = -\frac{\left(\sigma^Q_{m,t}\right)^2 - \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma^Q_{i,t} \sigma^Q_{j,t} \rho^P_{ij,t}}{\sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma^Q_{i,t} \sigma^Q_{j,t} \left(1 - \rho^P_{ij,t}\right)}.$$

It is important to note that in our case versus that of Buss and Vilkov (2012) N is not the same as the constituents of the index; this is because the cross-section of stocks in our sample will be different than the constituents of any index with available options. Our next order is to decide our proxy for the true market portfolio. Given our huge cross-section of stocks, we use the broadest available index, the Russell 3000, versus the traditionally used S&P500. $\left(\sigma^Q_{m,t}\right)^2$ is estimated using the volatility surface files for the Russell 3000 from Option Metrics and equation 18, and w_i and w_j are estimated using market capitalizations *within* our sample of stocks. $\beta^Q_{k,t}$ is then estimated using equation 20 using the estimate of $\rho^Q_{ij,t}$ from equation 21.

We then run a similar GARCH-type regression now focused on testing whether our measure of attention spikes predict next period variations in risk-netural β,

$$\beta^Q_{k,t+1} = a_0 + a_\beta \cdot \beta^Q_{k,t} + \sum_{b=1}^{B} a_b \cdot \mathbb{I}(b) \cdot \left(A^\beta_{k,t} - A^s_{k,t}\right) + \varepsilon_{k,t+1}. \tag{22}$$

Table 7 panel B provides the results from this regression. Mirroring the regressions represented in table 7 panel A we regress future risk-neutral beta onto current beta and our measure of attention spikes at a forward beta of 1 month, quarter, semi-annual, and annual frequencies. In the first column we look at predictability over the whole sample. Starting in column 2 we look at various measure of beta derived from different option durations. Like the volatility regressions there is a decay in predictability as option duration increases, although the decay seems to be substantially less steep; this leads to a still economically large effect at more long-run estimates of β. Using 365 day option derived measure of β the individual firm demeaned standard deviation is 0.18. This implies that a one-standard-deviation higher attention spike predicts a 0.02σ higher risk-neutral beta over the next year. As noted above, shifts in variance are highly concentrated in the first week or two of earnings; we would assume that something similar would be there for estimates of β. Given that the coefficients on our indicators represents the *average* predictability for the three weeks post earnings, this would imply that this figure represents a lower-bound. We can more directly test by expanding the indicators to cover all weeks within a quarter around the earnings announcement date. That is, we can test when around an earnings announcement the predictability of our attention measure is greatest. Figure 8 plots the interaction coefficient values (i.e., a_b in equation 22 where $b \in \{0, ..., 11\}$) around earnings. a_0 is absorbed into our regression constant and thus not included in the plot, while $b = 6$ is the week of the earnings announcement. It is clear both from an economic and statistical perspective that our attention spike measure as greatest predictability during earnings week and the week after. It is also clear that the averaging across the three weeks reduces the economic significance of predictability. Using the weekly coefficient values a one-standard-deviation higher attention measure predicts a 0.04σ higher risk-neutral beta over the next year.

6. Conclusion

This paper analyzes the dynamic relationship between firm risk and opportunity and firm attention. As the primary economic unit of production, understanding the theoretical underpinnings and then measuring this relationship is an important contribution to the literature. We further these objectives by first developing a time-varying β model that includes regime switching idiosyncratic cash-flow growth and investor learning. Firms are exposed to both operating leverage and the availability of growth options. This produces two points of liquidity constraints for the firm: (1) due to negative idiosyncratic shocks to trend growth the probability that firms need to raise capital to cover fixed costs increases, and (2) due to positive shocks to trend growth the probability of needing capital to expand production increases. This generates idiosyncratic variation in systematic exposure, which is similar in spirit to Babenko, Boguth, and Tserlukevich (2016) although the underlying mechanism is different. Furthermore the execution of growth options reduces systematic risk as the asset base increases. The basic idea is that systematic exposure increases at both extremes–positive and negative shocks to cash-flow growth–but only reduces when positive shocks translate into the expansion of production. Investor learning then generates the specific prediction that asset pricing reactions to shifts in trend growth and thus discount rates will not be immediate, but concentrated around information revealing events.

This theoretical framework suggests that shifts in focus of the firm between liquidity and business expansion concerns would provide tremendous insight into the state of a firm's systematic risk. We test this premise using a novel dataset that quantifies attention to a broad set of topics by employees across more than 3,000,000 firms. We develop an algorithm that separates topics into those related to liquidity and macro economic uncertainty and those related to the expansion of products and services either consumed or produced by specific firms. Using these two sets of topics we first find that, in keeping with our theoretical motivation, smaller firms with greater growth options and degrees of operating leverage more intensely follow systematic versus sector topics. Also dynamic shifts in attention from sector to systematic topics predict lower returns, and higher total risk and covariances with the market portfolio. Finally, this predictability is highly concentrated around earnings announcements, suggesting concentrated investor learning about underlying state of firm cash-flow growth during periods of information revelation.

References

Ahern, Kenneth R, 2017, Information networks: Evidence from illegal insider trading tips, *Journal of Financial Economics* 125, 26–47.

Arellano, Manuel, and Olympia Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29–51.

Babenko, Ilona, Oliver Boguth, and Yuri Tserlukevich, 2016, Idiosyncratic cash flows and systematic risk, *The Journal of Finance* 71, 425–456.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, 2016, Measuring economic policy uncertainty, *The Quarterly Journal of Economics* 131, 1593–1636.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *The Review of Financial Studies* 16, 101–143.

Ben-Rephael, Azi, Zhi Da, and Ryan D Israelsen, 2017, It depends on where you search: Institutional investor attention and underreaction to news, *The Review of Financial Studies* 30, 3009–3047.

Berk, Jonathan B, Richard C Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *The Journal of Finance* 54, 1553–1607.

Buss, Adrian, and Grigory Vilkov, 2012, Measuring equity risk with option-implied correlations, *The Review of Financial Studies* 25, 3113–3140.

Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2004, Corporate investment and asset price dynamics: Implications for the cross-section of returns, *The Journal of Finance* 59, 2577–2603.

David, Alexander, 1997, Fluctuating confidence in stock markets: Implications for returns and volatility, *Journal of Financial and Quantitative Analysis* 32, 427–462.

———, and Pietro Veronesi, 2013, What ties return volatilities to price valuations and fundamentals?, *Journal of Political Economy* 121, 682–746.

Doshi, Hitesh, Kris Jacobs, Praveen Kumar, and Ramon Rabinovitch, 2019, Leverage and the cross-section of equity returns, *The Journal of Finance* 74, 1431–1471.

Eisfeldt, Andrea L, and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance* 68, 1365–1406.

Fama, Eugene F, and Kenneth R French, 2004, The capital asset pricing model: Theory and evidence, *Journal of Economic Perspectives* 18, 25–46.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajšek, 2017, Inflation dynamics during the financial crisis, *American Economic Review* 107, 785–823.

Gondhi, Naveen, 2017, Rational inattention, misallocation, and asset prices, *Unpublished working paper. INSEAD*.

Hibbert, Ann Marie, Qiang Kang, Alok Kumar, and Suchi Mishra, 2020, Heterogeneous beliefs and return volatility around seasoned equity offerings, *Journal of Financial Economics*.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S Rosen, 1988, Estimating vector autoregressions with panel data, *Econometrica* pp. 1371–1395.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Li, Wendy CY, and Bronwyn H Hall, 2020, Depreciation of business r&d capital, *Review of Income and Wealth* 66, 161–180.

Lins, Karl V, Henri Servaes, and Peter Tufano, 2010, What drives corporate liquidity? an international survey of cash holdings and lines of credit, *Journal of Financial Economics* 98, 160–176.

Sagi, Jacob S, and Mark S Seasholes, 2007, Firm-specific attributes and the cross-section of momentum, *Journal of Financial Economics* 84, 389–434.

Stewart, Craig A, Von Welch, Beth Plale, Geoffrey Fox, Marlon Pierce, and Thomas Sterling, 2017, Indiana university pervasive technology institute, https://pti.iu.edu/.

Welch, Ivo, and David Wessels, 2000, The cross-sectional determinants of corporate capital expenditures: A multinational comparison, *Schmalenbach Business Review* 52, 103–136.

Whited, Toni M, 2006, External finance constraints and the intertemporal pattern of intermittent investment, *Journal of Financial Economics* 81, 467–502.

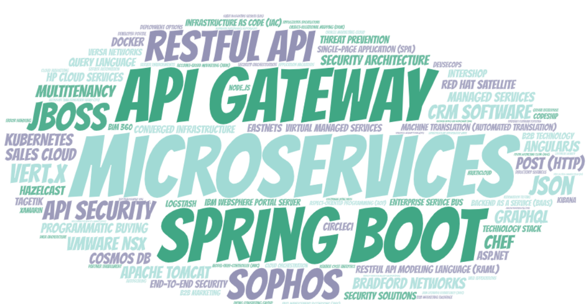
Figure 1. Industry Topic Clouds

In this figure we present word clouds of the top 100 topics associated with a particular 3-digit NAICS codes. Sizes of illustrated topic names roughly correspond to the topic ranking. Rankings are derived using the algorithm described in section 4.

panel A. Chemical Manufacturing

panel B. Credit Intermediation and Related Activites

panel C. Professional, Scientific and Technical Services



panel D. Publishing (except Internet)

panel E. Fabricated Metal Manufacturing

panel F. Oil and Gas Extraction

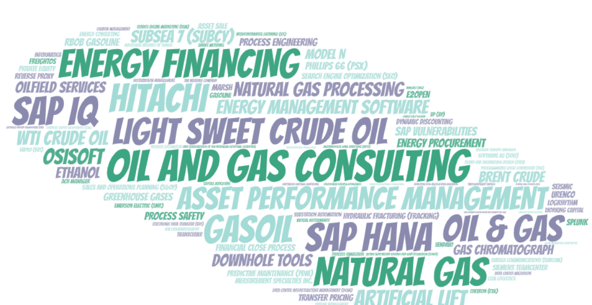
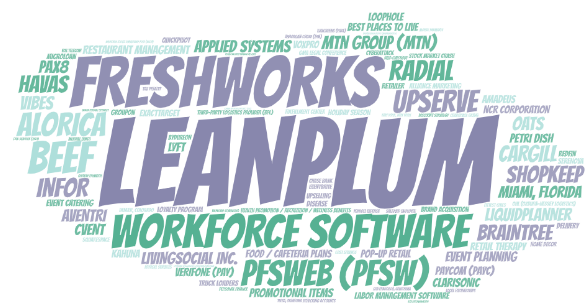


Figure 1. Industry Topic Clouds

(continued)

panel G. Insurance Carriers and Related Activities

panel H. Food Services and Drinking Places



panel I. Educational Services

panel J. Real Estate

panel K. Health and Personal Care Stores

panel L. Ambulatory Health Care Services

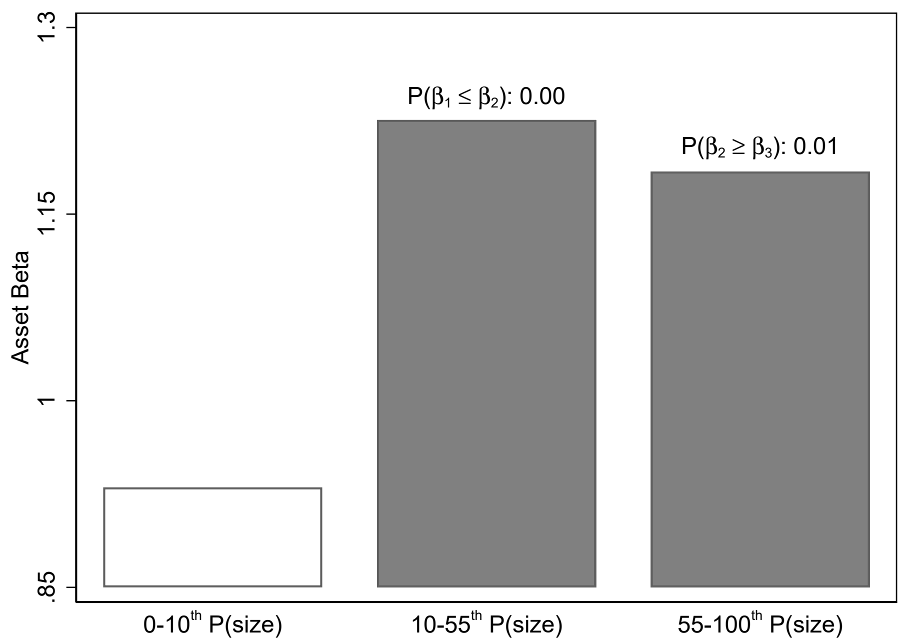
Figure 2. Systematic Topic Cloud

In this figure we present the word cloud of the top 100 topics associated with our fictitious *macro* industry. Sizes of illustrated topic names roughly correspond to the topic rankings. The derivation of the macro industry reading list and topic ranking is described in section 4.

Figure 3. Cross-sectional Sensitivity by Size

In this figure we present the sensitivities of β^A (panel A) and the read-ratio (panel B) to size as presented in regression 11. β^A is a derived measure of firm asset β as formulated by Doshi, Jacobs, Kumar, and Rabinovitch (2019). The read-ratio is defined as the log percentile of firm k's normalized reading score across systematic topics minus its log percentile of normalized reading score across sector topics. The dependent variable (size) is split into three buckets: less than 10th, 10-55th and 55-100th percentiles. Standard errors clustered by firm and date. The difference in sensitivities between buckets are tested using a Wald Statistic; p-values are presented within the graphs.

panel A. β^A Versus Portfolios by Size



panel B. Read-ratio Versus Portfolios by Size

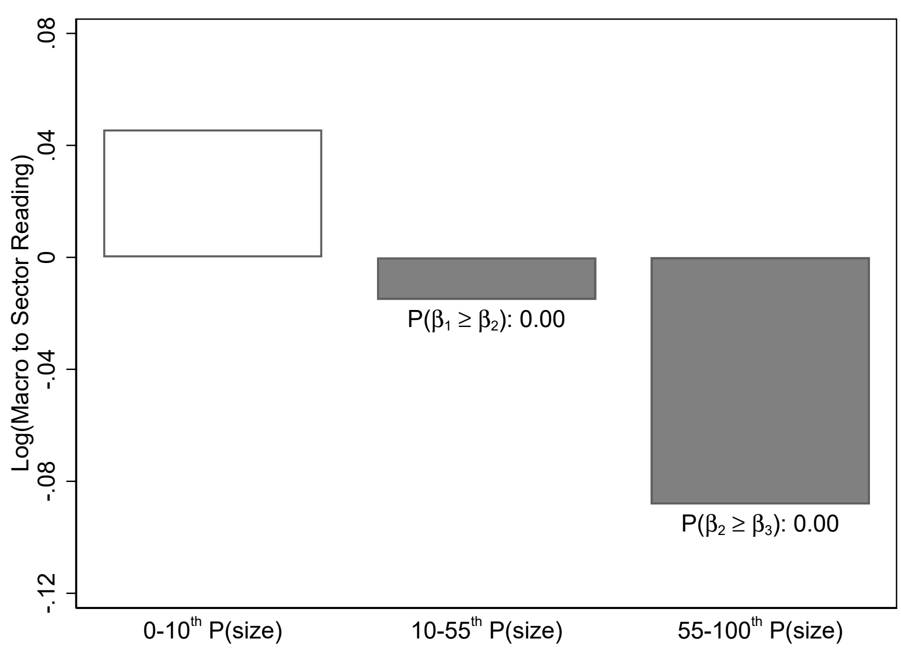
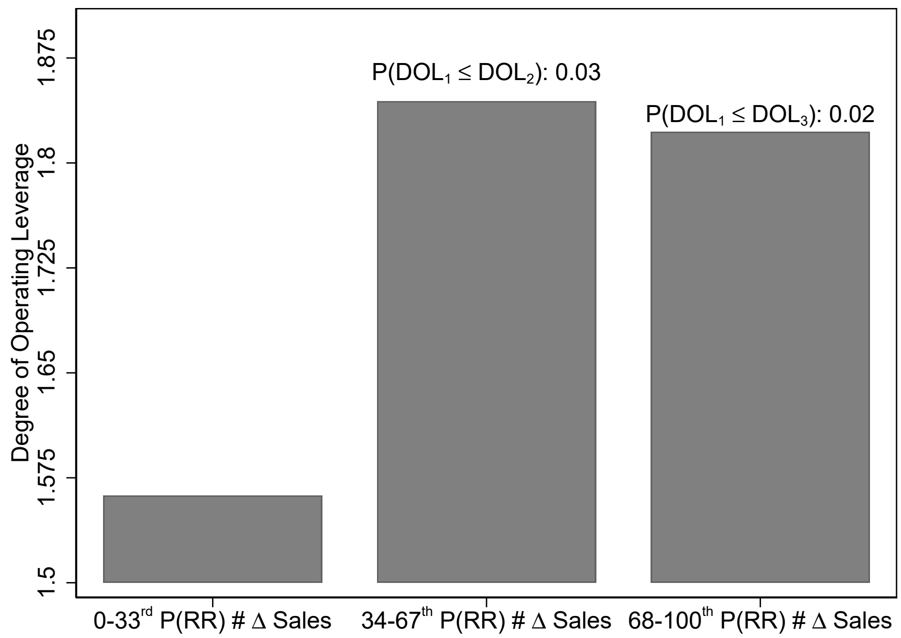


Figure 4. Degree of Operating Leverage Regressions

In this figure represents the interaction coefficients from regression 13. The dependent variable is annual $\Delta EBITDA$ which is defined as growth in operating cash flow (operating include plus non-cash depreciation); the independent is annual sales growth. The sensitivity of $\Delta EBITDA$ to $\Delta Sales$ is then allowed to different between firms of different systematic to sector reading. Firms are split into terciles based on the average read-ratio over the full sample (2015-2020). Panel A and panel B presents the sensitivities of terciles from 2014-today and 2015-today, respectively. Wald statistics are computed comparing the second and third terciles to the first; the statistics p-value is estimated using firm and time clustered standard errors.

panel A. DOL Coefficients $\geq$ 2014 Sample



panel B. DOL Coefficients $\geq$ 2015 Sample

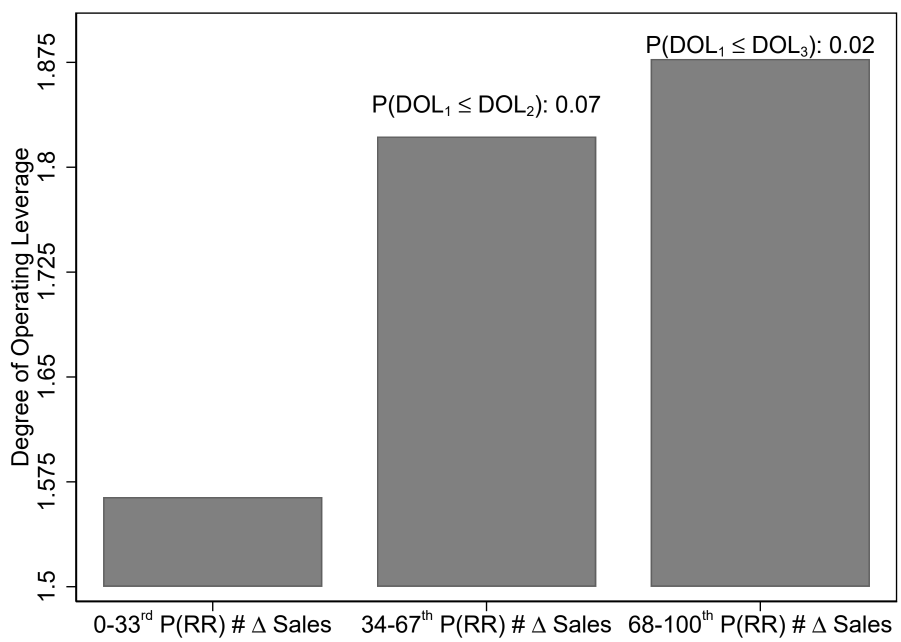
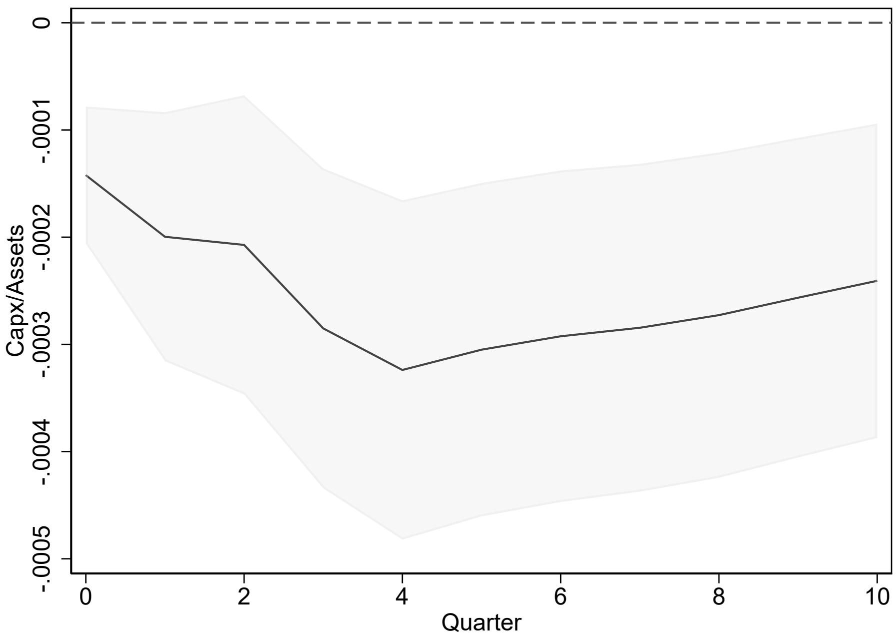


Figure 5. Investments Impulse Response Functions

This figure presents the Cholesky orthogonalized impulse response functions of our normalized investment variables to a one-standard-deviation shock to attention. The underlying vector autoregression and variable stacking is described by 15. Confidence intervals are robust to heteroskedasticity. Time and firm fixed effects are estimated by forward and cross-sectionally demeaning variables following Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bover (1995).

panel A. Capital Expenditures



panel B. Selling, General and Administrative Expenses

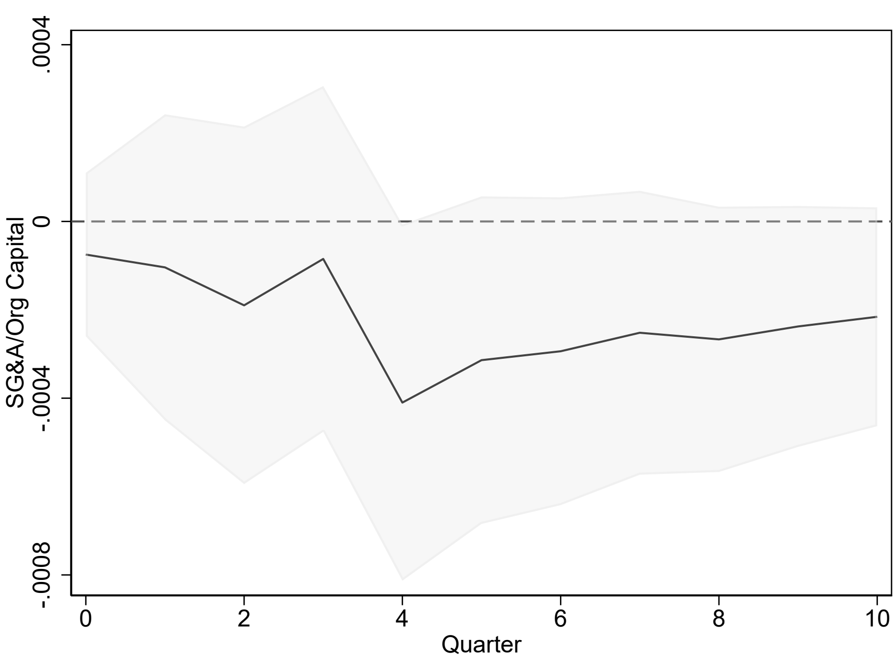


Figure 6. $A^{\beta}_{k,t} - A^{s}_{k,t}$ Impulse Response Function

This figure presents the Cholesky orthogonalized impulse response functions of quarterly aggregated attention from a one-standard-deviation shock to quarterly sales growth. The underlying vector autoregression is described by 15; however, now variable stacking is given by equation 16. Time and firm fixed effects are estimated by forward and cross-sectionally demeaning variables following Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bover (1995).

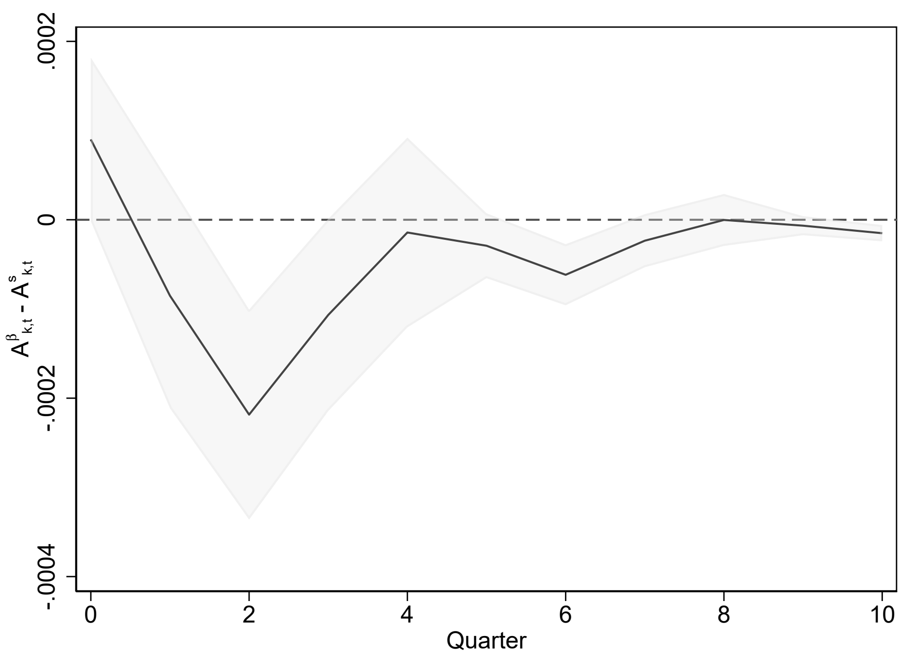
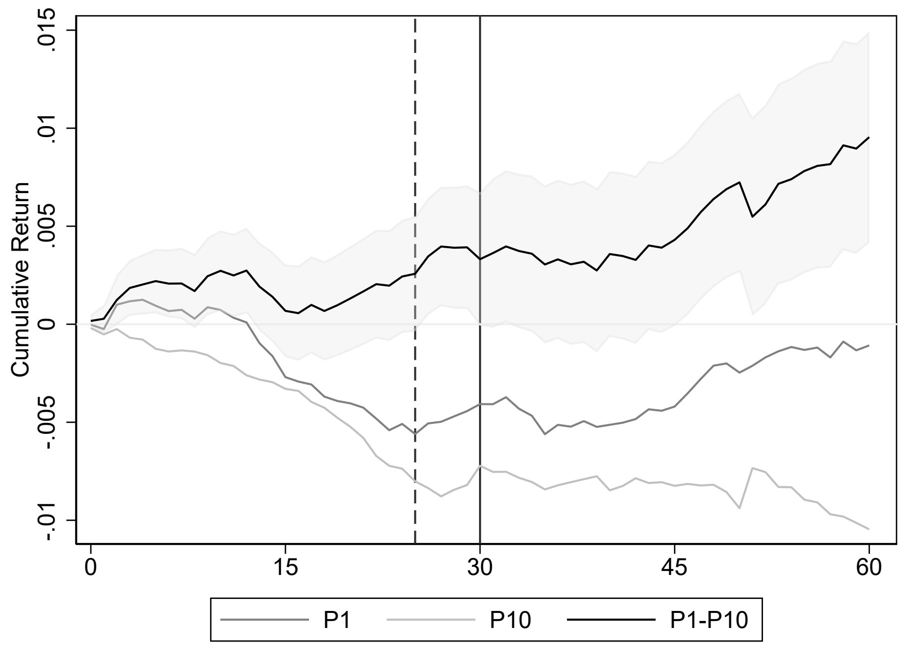


Figure 7. Portfolio Returns

This figure show cumulative returns of a long-short strategy identified using our intensive attention measure. All firms within a quarter are placed in deciles according to intensive margin score five days before earnings announcement (dashed line). Days within a quarter are indexed as being 30 trading days before (0 to 29 days) and 30 days after (31 to 61 days) announcement date (solid line). Returns are averaged on a given trading day across all stocks and quarters. The total long-short portfolio is assumed to be of net zero investment. Return standard errors are estimated using the cross-section of stock returns. Cumulative return standard errors are then estimated assuming intertemporal independence. Also graphed are the average first and tenth decile portfolio returns in excess of market returns.

panel A. Equal Weighted Returns



panel B. Value Weighted Returns

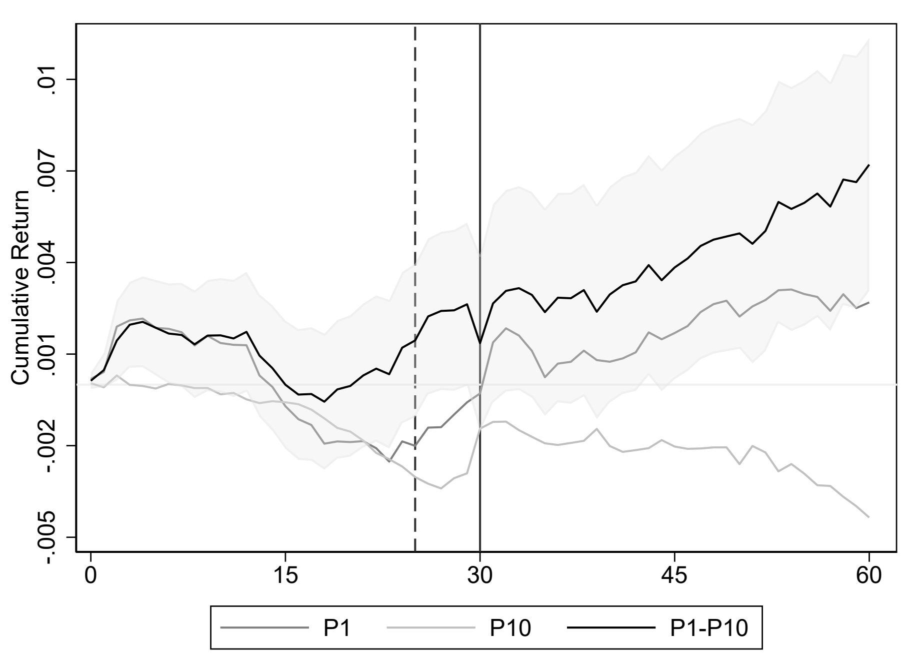


Figure 8. Risk Neutral β Predictability

In this figure we present the predictability coefficient values of regression 22 where indicators are given for each within a quarter. Wald statistics are then used to tests for shifts in predictability of our intensive attention measure by week. *p*-values are presented within the figure. Earnings announcement date is indicated as week 6. Week 1 through 5 are the 5 weeks before earnings announcement and weeks 7-11 are the 5 weeks after. Regression 22 otherwise uses the same specification as in column 5 of table 7 panel B.

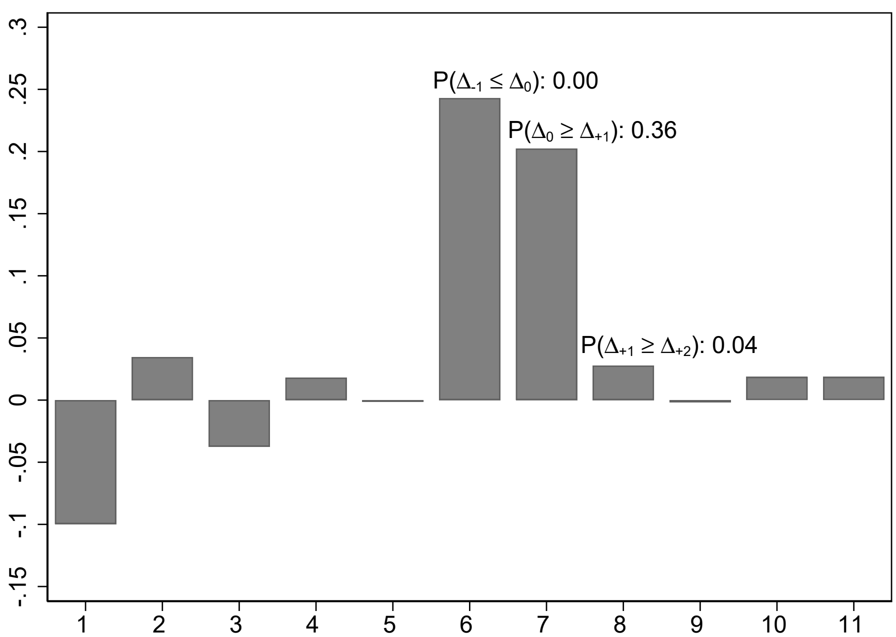


Table 1. Dataset Summary Statistics

This table reports summary statistics of our raw data. "No Topics" and "No Domains" are the number of unique topics or firms paid attention to across the entire database for a given year. We then illustrate the representativeness of attention by presenting the 10th, 30th, median, 70th and 90th percentile of topics per domain across the years. This is done on the full dataset and limited to the firms in the merged CRSP-COMPUSTAT file.

	2015	2016	2017	2018	2019	2020
No Topics	2459	2998	3740	5434	6766	6968
No Domains (Firms)	638086	760677	1814507	2225211	2252270	1521256
Topics per Domain						
Full Sample:						
p10	24	21	58	54	34	19
p30	98	95	276	297	196	75
p50	215	219	627	766	542	188
p70	394	410	1172	1525	1175	426
p90	822	847	2264	2968	2781	1305
CRSP-COMPUSTAT Sample:						
p10	168	167	1165	1674	1431	502
p30	615	677	2645	3729	3945	1945
p50	1095	1236	3361	4639	5592	3822
p70	1567	1823	3643	5139	6489	5737
p90	2040	2431	3716	5389	6730	6828

Table 2. Top 30 Represented Industries

This table breaks down the top 30 industries within different subsamples of our dataset. Our separation methodology (see section 4) utilizes the full sample of firms whereas our empirical tests, due to data availability, only utilize the CRSP-COMPUSTAT sample. Using a coarse measure of employment available in our dataset, the table includes the fraction of total employment and total number of firms that each top 30 NAICS 3-digit industry represents for the various subsamples.

COMPUSTAT Sample			Full Sample		
Industry Name	Emp.	Firms	Industry Name	Emp.	Firms
Professional, Scientific, and Technical Services	0.077	0.225	Professional, Scientific, and Technical Services	0.132	0.225
Chemical Manufacturing	0.071	0.037	Educational Services	0.064	0.037
Credit Intermediation and Related Activities	0.066	0.014	Credit Intermediation and Related Activities	0.047	0.014
Publishing Industries (except Internet)	0.057	0.012	Chemical Manufacturing	0.041	0.012
Fabricated Metal Product Manufacturing	0.054	0.036	Administrative and Support Services	0.04	0.036
Miscellaneous Manufacturing	0.053	0.031	Publishing Industries (except Internet)	0.038	0.031
Transportation Equipment Manufacturing	0.046	0.024	Motor Vehicle and Parts Dealers	0.037	0.024
Utilities	0.039	0.011	Food Manufacturing	0.032	0.011
Insurance Carriers and Related Activities	0.038	0.009	Transportation Equipment Manufacturing	0.027	0.009
Electrical Equipment, Appliance, and Component Manufacturing	0.037	0.02	Insurance Carriers and Related Activities	0.026	0.02
Food Manufacturing	0.034	0.016	Machinery Manufacturing	0.025	0.016
Machinery Manufacturing	0.033	0.017	Computer and Electronic Product Manufacturing	0.025	0.017
Food Services and Drinking Places	0.033	0.025	Merchant Wholesalers, Durable Goods	0.024	0.025
Computer and Electronic Product Manufacturing	0.023	0.068	Religious, Grantmaking, Civic, and Similar Organizations	0.023	0.068
Accommodation	0.021	0.024	Food Services and Drinking Places	0.022	0.024
Truck Transportation	0.02	0.013	Miscellaneous Manufacturing	0.022	0.013
Merchant Wholesalers, Durable Goods	0.019	0.011	Fabricated Metal Product Manufacturing	0.02	0.011
Administrative and Support Services	0.019	0.005	Electrical Equipment, Appliance, and Component Manufacturing	0.019	0.005
Beverage and Tobacco Product Manufacturing	0.018	0.012	Accommodation	0.019	0.012
Oil and Gas Extraction	0.017	0.013	Miscellaneous Store Retailers	0.018	0.013
Motor Vehicle and Parts Dealers	0.017	0.005	Utilities	0.017	0.005
Mining (except Oil and Gas)	0.016	0.027	Real Estate	0.016	0.027
Health and Personal Care Stores	0.016	0.05	Ambulatory Health Care Services	0.016	0.05
Telecommunications	0.015	0.006	Mining (except Oil and Gas)	0.015	0.006
Real Estate	0.012	0.013	Health and Personal Care Stores	0.014	0.013
Repair and Maintenance	0.012	0.008	Telecommunications	0.011	0.008
Air Transportation	0.011	0.036	Specialty Trade Contractors	0.011	0.036
Heavy and Civil Engineering Construction	0.01	0.004	Truck Transportation	0.011	0.004
Amusement, Gambling, and Recreation Industries	0.009	0.015	Securities, Commodity Contracts, and Other Financial Investments Activities	0.01	0.015
Paper Manufacturing	0.008	0.002	Air Transportation	0.009	0.002

Table 3. Cross-sectional Validation Tests

These tables present the results of cross-sectional tests from section 5. The parameterized tests are represented by regressions 10 (panel A) and 12 (panel B). In 3 panel A, a derived measure of firm asset β is regressed onto β^{Sales}, Tobin's Q, market capitalization and the liquidity ratio. $\widehat{\beta}^{Sales}$ is derived from regressing quarterly firm sales growth onto aggregate sales growth onto aggregate quarterly sales growth. Tobin's Q is defined by formula 9. Capitalization is the market value of assets (equity plus liabilities). Following Gilchrist, Schoenle, Sim, and Zakrajšek (2017), the liquidity ratio is used as a proxy for the firm degree of operating leverage. In 3 panel B the dependent variables changes to the read-ratio. The read-ratio is defined as the log percentile of firm k's normalized reading score across systematic topics minus its log percentile of normalized reading score across sector topics. Reported t-statistics are clustered by firm and time. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

panel A. β^A Regressions

	(1)	(2)	(3)	(4)	(5)
β^{Sales}	0.0062***				0.0069***
	[7.54]				[8.20]
Tobin's Q		0.0307*			0.0878***
		[1.87]			[4.37]
Capitalization			0.0224***		0.0165***
			[5.62]		[4.25]
Liquidity Ratio				0.0141**	0.0381***
				[2.17]	[5.53]
Debt-to-Equity					+
Observations	33,778	33,899	33,899	34,007	33,362
R^2	0.0143	0.0012	0.0112	0.0020	0.0478

panel B. Read-ratio Regressions

	(1)	(2)	(3)	(4)	(5)
β^{Sales}	0.0018***				0.0013***
	[3.85]				[2.81]
Tobin's Q		0.0271**			0.0279**
		[2.33]			[2.32]
Capitalization			−0.0214***		−0.0216***
			[−6.52]		[−6.36]
Liquidity Ratio				0.0329***	0.0216***
				[6.34]	[3.48]
Debt-to-Equity					+
Observations	39,903	39,923	39,923	40,441	38,980
R^2	0.0014	0.0008	0.0082	0.0093	0.0148

Table 4. FEs for Cross-sectional Regressions

This table presents the addition of fixed effects to the regression specification presented in column 5 of tables 3 panel A and 3 panel B. We begin with the original specification, including a debt-to-equity control (columns 1 and 4). To that we add time fixed effects (columns 2 and 5), and, finally, both time and firm fixed effects (columns 3 and 6). Reported t-statistics are clustered by firm and time. The independent variable is either our derived measure of asset β (columns 1-3) or our proxy for systematic to sector reading (columns 4-6). Reported t-statistics are clustered by firm and time. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Asset Beta			Read Ratio		
β^{Sales}	0.0069***	0.0066***	0.0003	0.0013***	0.0011**	0.0028**
	[8.20]	[7.95]	[0.37]	[2.81]	[2.46]	[2.61]
Tobin's Q	0.0878***	0.0864***	−0.0146	0.0279**	0.0286**	0.0122
	[4.37]	[4.32]	[−1.05]	[2.32]	[2.38]	[0.71]
Capitalization	0.0165***	0.0172***	0.0267***	−0.0216***	−0.0211***	−0.0145
	[4.25]	[4.43]	[2.85]	[−6.36]	[−6.15]	[−1.21]
Liquidity Ratio	0.0381***	0.0376***	0.0021	0.0216***	0.0220***	0.0001
	[5.53]	[5.51]	[0.85]	[3.48]	[3.52]	[0.02]
Debt-to-Equity	+	+	+	+	+	+
Time FE		+	+		+	+
Firm FE			+			+
Observations	33,362	33,362	33,224	38,980	38,980	38,801
R^2	0.0478	0.0588	0.9317	0.0148	0.0227	0.4281

Table 5. Investment Predictive Regressions

This table presents the results of regression 14. In panel A the dependent variable is Capital Expenditures (CapEx) normalized by lagged assets. In panel B the dependent variable is SG&A normalized by lagged organizational capital. We follow Eisfeldt and Papanikolaou (2013) in estimating organizational capital although the rate of depreciation assumption is higher at 25%. This is in line iwth recent work done at the BEA (see Li and Hall, 2020). As CapEx and SG&A are only available quarterly the independent variable, $A^{\beta}_{k,t} - A^{s}_{k,t}$, which is available weekly, is averaged over the quarter. For horizons greater than contemporaneous (i.e. $h \geq 0$), investment flow is accumulated over the horizon and then normalized by the stock at quarter $t-1$. Reported t-statistics are clustered by firm and time. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

panel A. Capital Expenditures

	h = 0	h = 1	h = 2	h = 3	h = 4
$A^{\beta}_{k,t} - A^{s}_{k,t}$	-0.0082^{*}	-0.0147^{*}	-0.0270^{**}	-0.0385^{**}	-0.0446^{***}
	[−1.81]	[−2.02]	[−2.22]	[−2.89]	[−3.36]
Time × NAICS FE	+	+	+	+	+
Firm FE	+	+	+	+	+
Observations	48,821	44,182	39,899	35,985	32,341
R^2	0.6906	0.7469	0.7852	0.8132	0.8377

panel B. Selling, General and Administrative Expenses

	h = 0	h = 1	h = 2	h = 3	h = 4
$A^{\beta}_{k,t} - A^{s}_{k,t}$	-0.0154	-0.0501^{*}	-0.0919^{***}	-0.1673^{***}	-0.2106^{***}
	[−1.03]	[−2.03]	[−2.90]	[−3.12]	[−3.04]
Time × NAICS FE	+	+	+	+	+
Firm FE	+	+	+	+	+
Observations	44,620	40,043	36,056	32,460	29,151
R^2	0.6738	0.7375	0.7753	0.8031	0.8307

Table 6. Return Predictive Regressions

This table presents the results from regression 17. Expected rates of return are subtracted from next period's excess returns and then regressed onto our measure of attention. Expected rates of return are estimated using market returns, SMB, HML and UMD return factors. Which factors used are highlighted in the column header. As we are interested in changes in predictability around earnings announcement dates, we split attention into those periods 3 weeks before and 3 weeks after earnings. We then test, using a Wald Statistic, whether predictability changes. p-values are presented in the tables. Panel A uses announcement date as a firm dissecting point for testing changes in predictability; panel B includes the week before earnings in the post-earnings indicator. Reported t-statistics are clustered by firm and time. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

panel A. Contemporaneous to Earnings Indicator

	None	$r_{m,t} - r_{f,t}$	$+ SMB_t, HML_t$	$+ UMD_t$	Full, All
$(A^{\beta}_{k,t} - A^{s}_{k,t})$	0.0077				
	[0.65]				
$\mathbb{I}_{\text{other}} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0094	0.0104	0.0127	0.0069
		[0.62]	[0.71]	[0.85]	[0.46]
$\mathbb{I}_{[-3,0)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0234	0.0247	0.0254	0.0221
		[1.21]	[1.28]	[1.29]	[1.11]
$\mathbb{I}_{[0,+3)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		−0.0182	−0.0095	−0.0081	−0.0036
		[−0.94]	[−0.51]	[−0.42]	[−0.17]
Time FE	+	+	+	+	+
Firm FE	+	+	+	+	+
$(\mathbb{I}_{[0,+3)} \times \cdot) - (\mathbb{I}_{[-3,0)} \times \cdot)$		-0.0417	-0.0342	-0.0335	-0.0257
$P(\mathbb{I}_{[-3,0)} \times \cdot < \mathbb{I}_{[0,+3)} \times \cdot)$		0.0299	0.0608	0.0689	0.1342
Observations	611,494	611,494	611,494	611,494	916,839
R^2	0.1818	0.0310	0.0111	0.0097	0.0091

panel B. One Week Lag to Earnings Indicator

	None	$r_{m,t} - r_{f,t}$	$+ SMB_t, HML_t$	$+ UMD_t$	Full, All
$(A^{\beta}_{k,t} - A^{s}_{k,t})$	0.0077				
	[0.65]				
$\mathbb{I}_{\text{other}} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0030	0.0070	0.0094	0.0090
		[0.22]	[0.53]	[0.69]	[0.66]
$\mathbb{I}_{[-3,-1)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0485**	0.0414*	0.0429*	0.0383
		[2.02]	[1.74]	[1.78]	[1.63]
$\mathbb{I}_{[-1,+3)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		−0.0198	−0.0142	−0.0157	−0.0264
		[−0.77]	[−0.55]	[−0.60]	[−0.99]
Time FE	+	+	+	+	+
Firm FE	+	+	+	+	+
$(\mathbb{I}_{[-1,+3)} \times \cdot) - (\mathbb{I}_{[-3,-1)} \times \cdot)$		-0.0683	-0.0557	-0.0586	-0.0647
$P(\mathbb{I}_{[-3,-1)} \times \cdot < \mathbb{I}_{[-1,+3)} \times \cdot)$		0.0217	0.0509	0.0437	0.0292
Observations	611,494	611,494	611,494	611,494	916,839
R^2	0.1818	0.0310	0.0111	0.0097	0.0091

Table 7. Variance and β Predictive Regressions

This table presents the results of regression 19 in panel A and regression 22 in panel B. Risk neutral variance and β are estimated using the methodologies of Bakshi, Kapadia, and Madan (2003) and Buss and Vilkov (2012), respectively. Both independent variables are computed using options of various durations: 1 month, 1 quarter, 1 semi-annual period and annual as described in the table columns. We split the period around earnings announcement to test if predictability changes around earnings. p-value from the Wald Statistics are presented. Reported t-statistics are clustered by firm and time. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

panel A. Risk Neutral Variance

	30 day	30 day	90 day	182 day	365 day
$\left(\sigma^{Q}_{k,t}\right)^2$	0.4716***	0.4715***	0.6097***	0.6416***	0.6324***
	[39.10]	[39.10]	[45.79]	[49.21]	[52.01]
$(A^{\beta}_{k,t} - A^{s}_{k,t})$	0.0355				
	[1.40]				
$\mathbb{I}_{\text{other}} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0218	0.0275	0.0236	0.0182
		[0.71]	[1.38]	[1.52]	[1.26]
$\mathbb{I}_{[-3,0)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		−0.0894*	0.0031	0.0090	0.0068
		[−1.90]	[0.15]	[0.45]	[0.40]
$\mathbb{I}_{[0,+3)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.2279***	0.0831***	0.0581***	0.0490***
		[4.57]	[3.58]	[3.04]	[2.69]
Time FE	+	+	+	+	+
Firm FE	+	+	+	+	+
$P(\mathbb{I}_{[-3,0)} \times \cdot > \mathbb{I}_{[0,+3)} \times \cdot)$		0.0000	0.0027	0.0254	0.0417
Observations	609,609	609,609	609,609	609,609	609,609
R^2	0.6512	0.6512	0.8054	0.8434	0.8498

panel B. Risk Neutral Beta

	30 day	30 day	90 day	182 day	365 day
$\beta^{Q}_{k,t}$	0.5575***	0.5575***	0.7124***	0.7601***	0.7838***
	[58.02]	[58.02]	[80.25]	[88.87]	[94.29]
$(A^{\beta}_{k,t} - A^{s}_{k,t})$	0.0978				
	[1.58]				
$\mathbb{I}_{\text{other}} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.0998	0.0389	0.0165	−0.0021
		[1.49]	[0.97]	[0.46]	[−0.06]
$\mathbb{I}_{[-3,0)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		−0.2181*	−0.0755	−0.0224	−0.0084
		[−1.71]	[−1.37]	[−0.46]	[−0.19]
$\mathbb{I}_{[0,+3)} \times (A^{\beta}_{k,t} - A^{s}_{k,t})$		0.4687***	0.2115***	0.1753***	0.1559***
		[3.91]	[3.26]	[3.26]	[3.08]
Time FE	+	+	+	+	+
Firm FE	+	+	+	+	+
$P(\mathbb{I}_{[-3,0)} \times \cdot > \mathbb{I}_{[0,+3)} \times \cdot)$		0.0002	0.0006	0.0042	0.0111
Observations	609,609	609,609	609,609	609,609	609,609
R^2	0.6364	0.6364	0.8013	0.8409	0.8419