Stock Price Comovement and the Consumption of Qualitative Information

*Travis Box**

ABSTRACT

This paper provides a new empirical strategy for testing models of information choice based on observing the type of information that is consumed and incorporated into asset prices. Consistent with the predictions of the information-driven comovement hypothesis (Veldkamp 2006a), I find that market-wide correlations are higher when many investors consume qualitative information about firms whose payoffs covary strongly with many others. Furthermore, as aggregate correlation falls, so does the demand for these high covariance signals. Thus, my findings imply that investor information consumption choices are shaped by a market for information.

JEL classification: C33, C53, D83, G00, G11, G12, G14

Keywords: Comovement, Dynamic Panel Estimation, Textual Analysis, Correlation

The process by which investors make information consumption choices is poorly understood, but critical to the functioning of financial markets. These consumption decisions are necessary because a single equity investor cannot keep pace with the combined volume of press releases, regulatory filings and news reports from more than just a few firms. Responsibility for all remaining

---

* tbox@bus.olemiss.edu, School of Business Administration, University of Mississippi, 340 Holman, P.O. Box 1848, University, MS 38677-1848, Office: (662) 915-2553, Fax: (662) 915-5821. I thank the University of Mississippi for their research support. I am grateful for the helpful comments of Eric Kelley, Andrew Lynch, Ryan Davis, Robert Van Ness, Geert Bekaert and Kenneth Singleton. I thank Paul Tetlock from Columbia University and Richard Brown and Maciek Pomalecki from Thomson Reuters Machine Readable News for data considerations. All mistakes in this article are my own.

content is often delegated to professional analysts and fund managers who transform this text volume into buy or sell recommendations and place bets on our behalf.

While these agents may possess the talent and training to manange this type of content accurately, they are also subject to the same basic information consumption constraints faced by individual investors. To broaden the appeal of their limited output, equity analysts often choose to focus their efforts on companies whose payoffs are most illustrative of the broader market (Hameed, et al. 2015). While this sort of concentration should maximize the aggregate rewards from their research (Veldkamp 2006a), clustering also ensures that many of the similarities between firm-pairs that lack analyst coverage will go sometimes unnoticed. Unfortunately, these gaps cannot be eliminated by simply redistributing analyst assignments uniformly across the market. Even if individual analysts were allocated to non-overlapping subsets of companies, important qualitative similarities between many firm-pairs would remain unobserved. While a diligent analyst would recognize qualitative simlairities across firms in their own portfolio, they would be unable to observe similarities between the companies that they follow and those followed by other analysts. Furthermore, no combination of research reports from different analysts would not be sufficient to reconstruct a comprehensive record of all similiarites between their respective portfolios.

In practice, these similarities are overlooked because qualitative information cannot be easily transferred and consumed by another investor. Liberti & Petersen (2017) describe how hard information, which is often recorded quantitatively, and soft information, which is often communicated as text, can be applied to financial market decisions. When quantitative information is collected by one person and transmitted to another, both people know exactly the same thing. Therefore, it is possible to delegate the collection of data to someone other than the decision maker. With qualitative information, however, the individual consuming the data may not know which parts are relevant or useful until long after they begin collecting the data. Thus, qualitative information must be collected in person by the decision maker. It is precisely this characteristic of qualitative information that I exploit to identify fluctuations in the type and quantity of information consumed by equity market investors.

My analysis is based on changes in the relation between firm-pair stock return correlation and the similarity of their qualitative information. More often than not, the stock prices of companies that are similar to each other tend to move in the same direction. The field of finance has identified a variety of individual characteristics that, when shared across firms, might predict comovement in their equity returns. Many of the characteristics that have been shown to explain comovement, such as as firm beta (Ledoit and Wolf 2003), size (Pindyck and Rotemberg 1993), book-to-market (Bekaert, Hodrick and Zhang 2009), momentum (Asness, Moskowitz and Pedersen 2013) and industry ((Campbell, et al. 2001), (Irvine and Pontiff 2009), and (Brandt, et al. 2010)), are easily measured and widely disseminated.

Other comovement predictors, such as the textual similarity in newswire text (Box 2017) and product market descriptions (Hoberg and Phillips (2010a) and (2010b)), are based on qualitative characteristics that are difficult to categorize effectively. Therefore, the qualitative similarities described by newswire text or product market descriptions will only affect stock return correlation if a nontrivial subset of investors are aware of the information connecting both companies. When a firm announces earnings that do meet expectations, the stock prices of their industry peers often fall contemporaneously because investors understand that the announcement may have implications for all the companies engaged in that type of business. Thus, return correlation between the peer firms is high because their stock prices frequently respond to the same information in real time. When investors choose to forgo the consumption of qualitative information related to a particular firm, the similarities between that company and all others in the market are nit recognized. Here, a disappointing earnings announcement may still have implications for qualitatively similar firms, yet the stock prices of those firms do not respond contemporaneously, and return correlations are low.

Veldkamp (2006a) presents a theoretical model of information choice based on the observation that information has a high fixed cost of production and a low cost of replication. Competition between information producers lowers the cost of high-demand content and encourages investors to consume the same information that others are purchasing. According to her information-driven comovement hypothesis, stock prices comove excessively when investors value firms based on a common subset of information. The model predicts that aggregate comovement will be highest when investors cluster their information consumption on firms whose payoffs covary strongly with many other companies.

By measuring changes in the relation between firm-pair stock return correlation and the similarity of their qualitative information, I am able to determine which types of information investors consume and incorporate into stock prices. Therefore, I can test whether investors focus their consumption on the type of content that leads to comovement. If qualitative similarities between a pair of firms predict how their equity payoffs covary, individual firms with higher average measures of textual similarity in their newswire text or product market descriptions should also have higher average payoff covariances. Consistent with the predictions of the information-driven comovement hypothesis, I find that market-wide correlations are higher when many investors consume qualitative information about firms whose payoffs covary strongly with many others. Furthermore, as aggregate correlation falls, so does the demand for these high covariance signals. Thus, my findings imply that investor information consumption choices are shaped by a market for information.

I also test two other predictions of the information-driven comovement hypothesis by examining how information consumption varies across firms and time. First, as the value of an investment rises, it comprises a larger share of the average investor's portfolio, and information about the investment becomes more valuable. Therefore, information consumption should increase as firm values grow larger. Next, the marginal benefit of consuming additional information rises as security payoffs become less predictable. Thus, demand for asset-specific information should also increase during times of uncertainty. In support of both predictions, I find that investor consumption of firm-specific information expands with market values and payoff volatility.

Theories of investor information choice have been unable to achieve broad acceptance because they are difficult to analyze without reliable quantitative measures describing investor information sets. Certain implications of the information-driven comovement hypothesis have been tested previously by examining changes in the production of information (Brockman, Liebenberg, & Schutte (2010) and Hameed, Morck, Shen, & Yeung (2015)). However, my paper is the first to demonstrate empirically that the consumption of information is determined by firm-specific characteristics and ambient market conditions.

I. Conceptual underpinnings

Grossman and Stiglitz (1980) build a rational expectations equilibrium model of information consumption where investors can choose to pay a fixed price and observe a signal about the future payoff of a single risky asset. As more investors learn the information, the signal becomes more easily inferred from the asset's price, and the benefit from observing the signal begins to fall. When the model is extended to multiple risky assets, a strategic substitutability emerges. Because investors prefer to buy low-demand assets that have lower prices, they also prefer to learn about assets that others know less about (Veldkamp 2011). Therefore, otherwise identical investors may choose to observe signals about different assets.

Veldkamp (2006a) replaces Grossman and Stiglitz's (1980) fixed information price with an information market. Her information-driven comovement hypothesis is motivated by the observation that information is fundamentally distinct from other goods because of its high fixed cost of production and near-zero cost of replication. This information production technology, coupled with free entry in the information market, creates a strategic complementarity that works through the market price for information. The lower price of high-demand content makes investors want to purchase the same information that others are purchasing. If investors buy mostly the same signals, and the signals they buy have high covariance with the other assets' payoffs, price comovement is strong.[1] Veldkamp (2006a) likens the cost of discovering a signal with that of hiring a journalist to find primary sources of information. Once discovered, the information can be distributed to other traders at zero marginal cost. Thus, primary sources of information are not reflected in asset prices until investors bear the cost of discovery, either by reading and evaluating the content themselves or by hiring an analyst or journalist to conduct the analysis on their behalf.

[1] The connection between information consumption and stock price comovement is not unique to the Veldkamp (2006a) model. Motivated by psychological evidence that attention is a scarce cognitive resource, Peng & Xiong (2006) model how investors allocate limited attention in an effort to reduce portfolio uncertainty. They propose that limited investor attention leads to category-learning behavior, whereby investors process more market- and sector-wide information than firm-specific information.

The text flowing across a financial newswire and appearing in the annual report should approximate the universe of available primary sources. I examine the degree to which certain types of newswire content are consumed and incorporated into asset prices. Recent empirical evidence suggests that aggregate consumption of comparable primary sources might be lower than expected. Loughran and McDonald (2015) examine download requests from the SEC's EDGAR server log. They find that an average publicly-traded firm has its 10-K requested only 27 times on the day of and the day following the filing date, and, for firms in the smallest three size quintiles, the average number of daily requests falls to five. The breadth of information processed by analysts and journalists appears to be similarly limited. Hameed, et al. (2015) report that almost one third of the listed firms in their sample lack analyst coverage, and Fang & Peress (2009) observe that only 75% of NYSE stocks and 42% of NASDAQ stocks are featured in newspaper articles. These findings support an assumption that is vital to my evaluation of the information-driven comovement hypothesis. Specifically, I assume that investors must choose which types and quantities of information to consume because it is not economical to process all the qualitative information appearing in the market.

According to the information-driven comovement hypothesis, a signal must contain information about the value of many assets, and it must be observed by many investors for it to produce comovement. Thus, aggregate comovement will be high relative to the covariance of underlying fundamentals when investors cluster their information consumption on firms whose payoffs covary strongly with others. Hameed, et al. (2015) examine whether profit-motivated information producers cluster on firms that have high payoff covariances. In their analysis, the payoff covariance between two firms is approximated by the historical correlation in their accounting profits. They provide evidence that equity analysts disproportionately follow firms whose fundamentals are good predictors of many other companies. Using each firm's overall level of qualitative similarity as an alternative measure of payoff covariance, I also find that analysts coordinate on firms whose signals contain information about the value of many others. To test whether payoff covariance motivates other types of information producers, I perform a similar analysis on the output of journalists. However, I find no evidence that these other profit-motivated producers also base their coverage decisions on a firm's average level of qualitative similarity.

Beyond just observing the types of information available, a more direct analysis of the information-driven comovement hypothesis requires knowledge of investors' information consumption choices. Figure 1 illustrates the mechanism through which these choices impact stock price comovement. Each period t, primary sources of information arrive in the market containing signals φ_{it} about the future payoffs of each firm i. These payoffs and, therefore the information signals about these payoffs, are correlated across companies. The degree to which information about firm i is qualitatively similar to information about firm j is described by π_{ij}, and the comovement between their stock prices is denoted ρ_{ij}. Individual investors choose which signals to discover at the beginning of each period, and λ_i represent the fraction of investors that demand information about firm i. When $\lambda_i, \lambda_j = 0$, signals φ_{it+1} and φ_{jt+1} about firms i and j go undetected by the market and have no impact on their stock price comovement. Therefore, signal correlation π_{ijt} has no relation with ρ_{ijt+1} unless firm-specific information consumption, λ_i or λ_j, is positive. As demand for information about either firm i or j strengthens, so does the relation between price comovement and qualitative similarity.

Without loss of generality, assume that the market consists of three firms A, B and C. If investors only choose to consume information about firm A, such that $\lambda_A > 0$ and $\lambda_B, \lambda_C = 0$, then signals φ_{At+1} about firm A will be incorporated into the stock prices of firms A, B and C. If qualitative information related to firms A and B has been similar in the past, $\pi_{ABt} > 0$, investors will bid up the stock price of firm B after observing positive signals about the payoffs of A. Even though the signals φ_{Bt+1} related to firm B are not incorporated into the price of either company's stock, their comovement during period $t + 1$ will be positively related to the correlation of their payoff signals π_{ABt} during period t. The comovement between firms B and C, however, will have no direct relation with their signal correlation covariance π_{BCt}. Instead, each company's stock price will fluctuate with signals φ_{At+1} about firm A, and the comovement between B and C will be determined by the payoff signal correlations π_{ABt} and π_{ACt}.[2]

Figure 1 demonstrates that the relation between qualitative similarity π_{ij} and stock price comovement ρ_{ij} depends on investor information consumption choices λ_i and λ_j. By measuring the strength of the relation between qualitative similarity and future comovement across firm-pairs, I can determine which types of information investors choose to consume. Therefore, I can test whether investors focus their consumption on the type of signals that lead to comovement.

According to the information-driven comovement hypothesis, investors only cluster their information consumption on firms whose payoffs covary strongly with others when aggregate information consumption is low. To illustrate this prediction, the previous framework is augmented

---

[2] In this scenario $\rho_{ABt+1} = f(\lambda_A \times |\varphi_{At+1}|, \lambda_A\pi_{ABt})$, $\rho_{ACt+1} = f(\lambda_A \times |\varphi_{At+1}|, \lambda_A\pi_{ACt})$ and $\rho_{BCt+1} = f(\lambda_A \times |\varphi_{At+1}|, \lambda_A\pi_{ABt}, \lambda_A\pi_{ACt})$.

with an information market characterized by a high fixed cost of production and near-zero cost of replication. Investors minimize the total variance of their portfolios by choosing which signals to purchase. Let Λ, equal to the sum of λ_i across all firms i, represent the total amount of information consumed by investors, and let $\overline{\rho_t}$, defined as the average of all pairwise return correlations ρ_{ijt} in each period t, describe the aggregate level of price comovement.

If the payoff signals of firm A are correlated with the signals of firms B and C, $\pi_{ABt}, \pi_{ACt} > 0$, but the payoffs of firms B and C are uncorrelated with each other, $\pi_{BCt} = 0$, then only signals about firm A can reduce uncertainty about the payoffs of all three assets. Thus, when aggregate information demand Λ is sufficiently low, investors will coordinate on the signal that has the highest covariance with the payoffs of the other two firms. This situation describes a strategic complementarity in information acquisition. Market-wide comovement $\overline{\rho_t}$ is high because the values of two firms, B and C, are being inferred based on common information. Now suppose that information demand Λ begins to rise. If it becomes optimal for investors to pay for two signals, they can now eliminate the most uncertainty by observing both low covariance signals, B and C, and inferring the value of the high covariance stock, A. This situation describes a strategic substitutability in information acquisition. Now, only one price is being determined based on inference, and market-wide comovement $\overline{\rho_t}$ is low.

The previous example illustrates how the aggregate level of information consumption Λ determines whether investors coordinate on high covariance signals. Without controlling for market states that determine the overall demand for information, I find that investors typically consume less qualitative information about firms whose payoffs covary strongly with most other assets. This result implies that the market's aggregate level of information consumption is high enough to support a strategic substitutability in information acquisition most of the time. However, I also find that coordination on high covariance signals becomes more common as market-wide return correlations $\overline{\rho_t}$ increase. Together, these results imply that comovement rises when many investors observe a limited number of high covariance signals, but that demand for low covariance signals is higher on average. Thus, complementarity leads to comovement, but substitutability typically prevails.

Two other implications of the theory provide a basis for testing the information-driven comovement hypothesis. In the model, the value of a signal is determined by its ability to reduce total payoff variance, where total payoff variance depends on risk and the value of the asset at risk. With regards to risk, asset-specific information becomes more valuable as security payoffs become less predictable.[3] Likewise, demand for asset-specific information increases whenever the asset comprises a larger share of the average investor's portfolio. In support of these predictions, I find that the production and consumption of information about a firm positively relates to its daily stock return volatility and market capitalization.

These same predictions also apply to aggregate information consumption. In times of uncertainty, the marginal benefit of observing additional signals rises, causing market-wide information consumption to increase. Similarly, when the total value of an asset rises, investors must hold that additional asset value for the asset market to clear. Therefore, aggregate demand for

[3] The investor attention models developed by Peng & Xiong (2006), Mondria (2010) and Kacperczyk, Van Nieuwerburgh, & Veldkamp (2016) suggest a similar relation between payoff variance and information processing.

information should increase when many assets are highly valued.[4] With regards to these broad market conditions, I find that firm-specific information consumption increases with total stock market volatility and cumulative market returns.

II. Sample description and qualitative similarity measures

The firm universe for this study consists of all domestic common stocks trading on the NYSE, NASDAQ and Amex exchanges with CRSP share codes 10 or 11. I calculate the NYSE price and size decile breakboints each six-month period from January 2003 to December 2013 based on the price and shares outstanding for the final trading day of the previous interval. Firms falling in the smallest price or size decile for a particular time period are removed from the sample where the average lowest breakpoints across all intervals are \$7.89 and \$259 million, respectively. The resulting sample contains an average of 1,982 firms at the beginning of each period with 2,740 unique firms appearing in at least one interval.

Figure 1 introduced the variable π_{ij} to describe how payoff signals are correlated across companies. For my analysis, the qualitative similarity of payoff signals is based on either the textual similarity in newswire text (Box 2017) or product descriptions (Hoberg and Phillips (2010a), and

[4] In models where incomplete information is motivated by limited attention, as opposed to costly information, aggregate information consumption is usually determined by a fixed processing capacity. Andrei & Hasler (2015) model the relation between attention to news, return volatility, and risk premia, but they avoid providing a theoretical foundation for fluctuating attention. Andrei & Hasler (2016) investigate a costly attention allocation decision. But, with just one risky asset their model is silent on comovement.

(2010b)). The newswire text comes from the Thomson Reuters NewsScope Archive, a historical database of *Reuters News* and select third party content. The Archive is derived from the Reuters Integrated Data Network (IDN) newswire feed and consists of the message stream which communicates text to client workstations. My approach to calculating the textual similarity of newswire text is identical to the process described in Box (2017). First, I calculate the cosine similarity, $\widetilde{WireSim}_{ijt}^{all}$, between the firm vectors i and j in the term-document matrix[5] for period t constructed from all text appearing on the Reuters Integrated Data Network. Next, firms with some relevant text are classified into deciles based on total word counts for each period in the sample. The variable $\overline{WireSim}_{ijt}^{all}$ represents the average document similarity between firms appearing in the same word count deciles as i and j during period t. Finally, the qualitative similarity of newswire text, $WireSim_{ijt}^{all}$ is calculated by subtracting $\overline{WireSim}_{ijt}^{all}$ from $\widetilde{WireSim}_{ijt}^{all}$.

The Thomson Reuters NewsScope Archive also describes the attribution, or source, of each story. There are a total of 12 attributions contributing relevant takes, however, only *Reuters News* consists primarily of content produced by journalists. Other attributions, such as *Business Wire* or

[5] The term-document matrix is a mathematical representation of the frequency of terms that occur in a collection of documents. The intuition behind this methodology is as follows: if the frequency of words used in the takes about different firms is similar, then the qualitative information contained in those stories is also similar. As an example, if the takes about two firms use words like "interest," "debt," and "default," it may be the case that both firms are having some difficulty accessing capital. Even if these firms are in entirely different industries and have entirely different market capitalizations, a newswire subscriber might expect some covariance in their future payoffs relative to firms whose newswire text does not mention these words.

PR Newswire, are more likely to contain content generated by the firms themselves in the form of press releases and legal disclosures. While the consumers of newswire content are likely to base investment decisions on takes produced by the companies or by the journalists, I give special attention to content generated by *Reuters News*. The qualitative similarity of newswire text written by journalist is represented by $WireSim_{ijt}^{rtrs}$

From the online Hoberg-Phillips Industry Classification Library, the variable $HobSim_{ijt}$ is the yearly firm-by-firm pairwise similarity score calculated by parsing the product descriptions of company 10-Ks, then forming word vectors for each firm to compute continuous measures of product similarity. Their variable is very similar to unadjusted document similarity $\widetilde{WireSim}_{ijt}^{all}$ discussed above. Unofortunately, Hoberg and Phillips (2015c) only make their measure publicly available for firms having pairwise similarities that are above a certain threshold.[6]

III. Empirical analysis

The subsequent analysis will attempt to answer two economic questions. First, do information producers focus their efforts on firms whose payoffs covary most strongly with other companies? If journalists and analysts process more information about firms qualitative information is similar to most other companies, this would imply that information producers provide the type of signals capable of generating comovement. Second, can information consumption choices help us

[6] From the Hoberg-Phillips Industry Classification Library (Hoberg and Phillips 2015c), "the TNIC-3 classification data only records firms having pairwise similarities with a given firm *i* that are above a threshold as required based on the coarseness of the three digit SIC classification. The level of coarseness of TNIC-3 thus matches that of three digit SIC codes, as both classifications result in the same number of firm-pairs being deemed related."

understand the origins of comovement? If investors cluster their information demand on a few signals that predict the values of many companies, price comovement will be high relative to the covariance of underlying fundamentals.

A. Information production

My analysis begins with an examination of information production. By studying the output of analysts and journalists, I investigate whether profit-motivated information producers focus their efforts on firms whose payoffs covary most strongly with others. Fang & Peress (2009) find that journalists cluster their coverage on large firms, but they do not test whether payoff covariance is a determinant of media following. Using correlation in historical accounting profits to measure total payoff covariance, Hameed, et al. (2015) provide evidence that equity analysts disproportionately follow firms whose fundamentals are good predictors of other companies'.

I propose two alternative proxies for total payoff covariance based on each firm's average level of newswire similarity or total product similarity. If the firm-specific signals between two companies are correlated, their payoffs will covary. When signals related to one firm are correlated with the signals of many different companies, that firm's total payoff covariance will be high. To determine whether information about one firm contains information about the payoffs of many others, I calculate firm i's average newswire similarity with all other firms j:

$$\overline{WireSim_{it}^{all}} = \frac{1}{N-1}\sum_{j\neq i} WireSim_{ijt}^{all} \tag{1}$$

where N is the number of firms with some positive volume of text appearing on the IDN during period t. Box (2017) shows that newswire similarity can predict how the future equity payoffs of two firms are correlated. Following Hoberg and Phillips (2016), we take the sum of pairwise similarity scores for each supplier firm in our sample to determine its overall degree of product similarity, $\sum HobSim_{kt}$. High values for this summation indicate that firms sell similar products to many other companies.

The information-driven comovement hypothesis also predicts that asset-specific information becomes more valuable when the security's payoffs become less predictable or when the security comprises a larger share of the average investor's portfolio. To measure average portfolio share, market capitalizations are calculated on the final trading day of each 6-month span, and every firm i is included in a NYSE size decile $SizeDec_{it}$ for the following period t. Payoff predictability is approximated by the firm's daily stock return standard deviation σ_{it}.

The level of information production is approximated by word count and analyst following. $WrdCnt_{it}^{rtrs}$ is the total number of words written about the firm and distributed by *Reuters News* during the 6-month span t, and $WrdCnt_{it}^{firm}$ is the total number of words contributed by all other attributions. Thus, the former applies to content produced by journalists, while the latter measures content generated by the companies themselves. The number of unique analysts with an earnings prediction recorded in the I/B/E/S database during period t is represented by $AnaNum_{it}$. With a median of 83 and an average of 554 total words, the summary statitistics reported in Table I reaffirm that the bulk of media coverage is focused on a very small number of companies. Analysts, on the other hand, follow a much broader universe of firms. In a typical 6-month span, 83% of the companies in my sample have an analyst earnings prediction, but only 62% have a positive quantity of text produced by *Reuters News*.

In addition to analyzing the determinants of information production, I am also interested in whether the availability of firm-specific information reduces comovement. In the simple case with three assets, where all investors observe signals related to asset A, Veldkamp's (2006a) model predicts that there will be no comovement between assets A and B, or assets A and C, in excess of their payoff covariance. Conversely, comovement will be high between assets B and C because investors must make correlated inferences about their values. Thus, higher volumes of firm-specific information consumption should be inversely related to that firm's average level of comovement. Presumably, profit-motivated information producers, like analysts and journalists, attempt to generate the type of content that investors ultimately purchase and consume. Therefore, the volume of their firm-specific output should be inversely related to each firm's average level of comovement. The Pearson correlation ρ_{ijt} between the daily stock returns of firms i and j describes their pairwise stock price comovement, and $\overline{\rho_{it}}$ is calculated by averaging ρ_{ijt} over all firms $j \neq i$.

The information-driven comovement hypothesis predicts that profit-motivated information producers focus their efforts on larger and more volatile firms and, given sufficiently low levels of aggregate information consumption Λ, companies whose payoffs covary most strongly with others. These predictions motivate the following model:

$$\begin{aligned} Dep_{it+1} = \beta_0 + \beta_1 SizeDec_{it} + \beta_2 \sigma_{it} + \beta_3 WireDum_{it}^{all} + \beta_4 \overline{WireSim_{it}^{all}} \\ + \beta_5 \sum HobSim_{kt} + \beta_6 \overline{\rho_{it}} + \sum_{k=7}^{8} \beta_k Other_{kit} \\ + \sum_{k=9}^{K} \beta_k Control_{kit} + \alpha_{t+1} + \varepsilon_{it+1} \end{aligned} \qquad (2)$$

where α_{t+1} is a fixed effect for each 6-month span. The variable Dep_{it+1} will be some measure of information production, $WrdCnt_{it+1}^{firm}$, $WrdCnt_{it+1}^{rtrs}$ or $AnaNum_{it+1}$, depending on the specification. Equation (2) suggests that content producers determine their coverage during period $t+1$ after observing individual firm characteristics during period t. The binary variable $WireDum_{it}^{all}$ indicates whether the firm has some positive volume of text appearing on the IDN.

This variable is necessary to differentiate when contemporaneous average newswire similarity, $\overline{WireSim_{it}^{all}}$, is 0 because the firm's qualitative information not excessively similar or dissimilar to most other firms, or because it was never mentioned on the newswire. The information-driven comovement hypothesis predicts that the coefficients β_1 and β_2 should be positive when the dependent variable is either measure of profit-motivated information production, $WrdCnt_{it+1}^{rtrs}$ or $AnaNum_{it+1}$. The coefficients β_4 and β_5 will also be positive if there is a strategic complementarity in content generation. Controlling for the firm's average level of price comovement $\overline{\rho_{it}}$ ensures that the relation between information production and total payoff covariance is independent of realized average return correlation. To determine whether different types of information producers influence each other, contemporaneous observations of each of the other two production measures, $Other_{kit}$, are also included in each specification. A description for all other included controls, $Control_{kit}$, is provided in Panel B of Table A-1.

The distributions of all three information production variables are described in Figure 2. Any summation of word count or analyst following is obviously bound below by 0, but Figure 2 demonstrates that a large portion of the pooled sample is also clustered at this bound for each variable. Moreover, even when information production is positive, realized values are still confined to a discrete set of integers. The simplest framework for analyzing counted data is the Poisson regression model (Cameron and Trivedi 2013),[7,8] however, an important limitation of the Poisson distribution is that the conditional variance is assumed to equal the conditional mean. According to Table I, this assumption might be inappropriate for my analysis because the unconditional variance of each information production variable is much larger than its sample mean.

A negative binomial distribution should be specified in cases where the variances derived from the data are higher than their conditional means (Gardner, Mulvey and Shaw 1995). Unlike the Poisson distribution, which is fully characterized by one parameter, the negative binomial distribution is a function of both its mean and a measure of overdispersion. Adapting Equation (2) to this framework gives:

$$
\begin{aligned}
&Dep_{it+1} \sim \text{Poisson}(\mu_{it+1}) \\
&\mu_{it+1} = \beta_0 + \beta_1 SizeDec_{it} + \beta_2 \sigma_{it} + \beta_3 WireDum_{it}^{all} + \beta_4 \overline{WireSim_{it}^{all}} \\
&\qquad + \beta_5 \sum HobSim_{kt} + \beta_6 \overline{\rho_{it}} + \sum_{k=7}^{8} \beta_k Other_{kit} \\
&\qquad + \sum_{k=9}^{K} \beta_k Control_{kit} + \alpha_{t+1} + \nu_{it+1} \\
&e^{\nu_{it+1}} \sim Gamma\left({}^{1}\!/_{disp_{it+1}}, disp_{it+1}\right)
\end{aligned}
\qquad (3)
$$

[7] Ordinary least squares estimation of Equation (2) assumes that the regression errors ε_{it+1} follow a normal distribution. This assumption is not appropriate when the left-hand side variables are limited to nonnegative integer values.

[8] A common approach to modeling data that is not normally distributed is to transform the variables, usually by taking their natural logarithm. Section IA-III in the internet appendix highlights some of the issues with this technique when the non-normally distributed dependent variable is a counted measure.

Equation (3) stipulates that the number of words written about, and the number of analysts following, a firm i during period $t+1$ is a negative binomial random variable with mean μ_{it+1} and dispersion parameter $disp_{it+1}$.[9]

Word counts and analyst following are observed over time, so my analysis must account for the correlation between repeated measures of information production related to the same firm. Companies that are covered by analysts and the financial press, during period t are also likely to be covered during period $t+1$. The generalized estimating equations approach introduced by Liang & Zeger (1986) specifies how the average of a response variable, $\bar{\mu}$, adjusts to changes in the independent variables while allowing for correlation between repeated measurements on the same individual over time. Parameters from this method of estimation have a population average interpretation. For every unit increase in an independent variable across the population, generalized estimating equations reveal how much the average response $\bar{\mu}$ would change (Ballinger 2004).[10]

[9] When the overdispersion parameter is 0, the negative binomial distribution becomes the Poisson distribution. Equation (2) is estimated with a Poisson and a negative binomial regression on the pooled sample of observations. For all three information production variables, a likelihood ratio test strongly rejects strongly rejects the null hypothesis that the overdispersion parameter is 0.

[10] The generalized estimating equations model specifies only the conditional mean μ_{it+1} and treats the correlation structure as a nuisance parameter (Gardiner, Luo, & Roman (2009) and Hardin & Hilbe (2013)). The algebraic form of the correlation structure is specified by the researcher through a working correlation matrix whose parameters are estimated by the method of moments. When the mean response is correctly specified, consistent parameter estimates will be derived even if the algebraic form of the correlation structure is misspecified. However, some loss of efficiency could result if the specified working correlation matrix is far from the true correlation. I estimate Equation (3) assuming an autoregressive correlation structure for each measure of information production.[10] Pan (2001) proposed a model-selection method for generalized estimating equations known as the quasi-likelihood information criterion. The

My analysis of average comovement is summarized by the following regression model:

$$\begin{aligned}\overline{\rho_{\iota t+1}} = \beta_0 + \beta_1 SizeDec_{it} + \beta_2 \sigma_{it} + \beta_3 WireDum_{it}^{all} + \beta_4 \overline{WireSim_{it}^{all}} \\ + \beta_5 \sum HobSim_{kt} + \beta_6 \frac{WrdCnt_{it}^{firm}}{1{,}000} + \beta_7 \frac{WrdCnt_{it}^{rtrs}}{1{,}000} \\ + \beta_8 AnaNum_{it} + \beta_9 \overline{\rho_{\iota t}} + \sum_{k=10}^{K} \beta_k Control_{kit} + \alpha_{t+1} + Ind_i \\ + \varepsilon_{it+1}\end{aligned} \quad (4)$$

To account for varying levels of average correlation between industries, every firm in the sample is assigned to one of the 49 industry portfolios as defined on Kenneth French's website. Ind_i is a fixed effect describing industry affiliation. Both word count variables, $WrdCnt_{it}^{firm}$ and $WrdCnt_{it}^{rtrs}$, are rescaled to ease the presentation of results. The coefficients β_7 and β_8 will be negative if the availability of firm-specific information produced by journalists and analysts reduces stock price comovement. The results from estimating Equations (3) and (4) are reported in Table II. The standard errors for the generalized estimating equations specifications are clustered by firm, and the standard errors for the ordinary least squares specification are clustered by firm and time using the Cameron, Gelbach and Miller (2011) multi-way clustering procedure.

If firm-generated content is often related to required disclosures, then output volume, $WrdCnt_{it}^{firm}$, is not determined by a market for information. Table II confirms that future firm-generated text volume is not positively associated with stock return volatility or either proxy for

---

specification of a negative binomial distribution with an autoregressive correlation structure is supported by this criterion.

total payoff covariance, $\overline{WireSim_{it}^{all}}$ or $\sum HobSim_{kt}$. Increasing market capitalization, however, does appear to raise firm-generated output. Companies that move into a higher size decile during period t subsequently increase their self-generated word count by 5.4%.[11] It is not possible to determine from Table II whether larger firms produce more content because of higher investor information demand or more strenuous disclosure requirements. An increase in contemporaneous analyst following does predict future firm-generated volume, but the economic impact is small. There is no similar relation between contemporaneous journalist output and future firm-generated text volume.

Consistent with the predictions of the information-driven comovement hypothesis, Table II shows that analysts coordinate on firms whose average newswire similarity, $\overline{WireSim_{it}^{all}}$, is high. However, I find evidence that journalist-produced text volume is negatively influenced by total payoff covariance. Thus, there is a strategic complementarity in information produced by analysts, but a strategic substitutability in information distributed by *Reuters News*. My second proxy for total payoff covariance based on product similarity, $\sum HobSim_{kt}$, is not positively related to the future output of analysts or journalists. While the future cash flows should be highly correlated between firms that make similar products, $HobSim_{kt}$ only captures one dimension of pairwise payoff covariance. Firm-pairs with higher levels newswire similarity, however, are likely to be

[11] For a one-unit change in the predictor variable, the difference in the logs of expected counts of the dependent variable is expected to change by the respective regression coefficient. For the coefficient on $SizeDec_{it}$, $e^{0.0524} = 1.0538$.

qualitatively similar in a variety of ways. Thus, $\overline{WireSim_{it}^{all}}$ represents a more robust measure of total payoff covariance.

Table II also demonstrates that contemporaneous average price comovement $\overline{\rho_{it}}$ has only a modest impact on analyst following and does not contribute positively to future text volume. Thus, average newswire similarity, $\overline{WireSim_{it}^{all}}$, is a better predictor of analyst information production than realized comovement, $\overline{\rho_{it}}$. I also find that future analyst following and journalist coverage increase with firm size, but only journalists are influenced positively by contemporaneous volatility. While journalists and analyst should both be motivated to focus their efforts on generating the most profitable content, their methods for creating value seem to diverge. Overall, I find that analyst following is concentrated on firms whose fundamentals are good predictors of other companies', whereas journalists focus on recent volatility.

The positive and nearly significant coefficients on $AnaNum_{it}$ and scaled $WrdCnt_{it+1}^{firm}$ in the second column provide evidence that future journalist coverage is positively influenced by contemporaneous firm-generated text volume and analyst following. When a company increases its own output by 1,000 words, or when one additional analyst begins to follow a firm, journalist-produced text volume increases by 3.0% and 1.0%, respectively. Table II also reveals, however, that the level of analyst following is not similarly related to either measure of word count.

Journalists are portrayed as information producers in the Veldkamp (2006a) model, but the positive association with contemporaneous firm-generated output implies that *Reuters News* may function more like an echo for primary sources and other profit-motivated producers. This result is consistent with the findings of Ahern & Sosyura (2014), who show that firms originate and disseminate information through the financial media.[12] Their conclusions are based on an even narrower classification of journalist-produced content. Publications like *The Wall Street Journal*, *The New York Times* and *The Washington Post* are described as media sources in their study, but *Reuters News*, *Dow Jones News Service* and *Business Wire* are lumped together as "firm-originated news." While the *Business Wire* stories included in my sample are clearly firm-generated, those from *Reuters News* have journalist bylines. Still, Ahern & Sosyura (2014) justify their classification by arguing that newswire stories provide little analysis. If content from *Reuters News* is at least somewhat "firm-originated," the market for information will play a smaller role in determining their coverage decisions.

For the final column in Table II, the relation between contemporaneous analyst following and future comovement is consistent with the predictions of the information-driven comovement hypothesis. The coefficient on $AnaNum_{it}$ is negative and significant implying that a firm's average level of comovement with all other firms in the market, $\overline{\rho_{it+1}}$, is inversely related to the amount of information produced by analysts. Thus, future comovement is highest when analyst following is low and investors are most likely to be making correlated inferences about a particular firm's value. The availability of relevant firm- and journalist-produced content, however, does not reduce a particular company's average level of stock price comovement with all other firms. Therefore, the production of information, by firms or journalists, may not reflect investor information consumption.

[12] The Pew Research Center (2011) analyzed several major storylines reported on television, radio, newspaper or online outlets and found that only 14% originated with reporters.

Overall, the results in Table II imply that firm-generated newswire content may be viewed as a primary information source, whereas analyst following may be determined by a market for information. My analysis of information production, however, does not consider aggregate changes in information demand, Λ. In Section III.B, I examine how aggregate information consumption responds to market conditions, and I study whether the type of information that investors choose to consume differs across market states.

B. Information-driven price comovement

Box (2017) shows that the similarity of qualitative information can predict how the future equity payoffs of two firms are correlated. Section III.A establishes that some profit-motivated information producers focus their efforts on firms whose payoff signals covary most strongly with other companies. The remaining analysis investigates whether or not investors also cluster their information demand on the types of signals that cause stock price comovement to be high relative to the covariance of underlying fundamentals. First, I analyze how aggregate information consumption changes with market conditions. Second, I examine how investors choose which types of information to consume. Finally, I study whether the type of information consumed differs across market states.

Figure 1 demonstrates that the degree to which information about two firms is qualitatively similar, π_{ij}, and stock price comovement, ρ_{ij}, depends on investor information consumption choices λ_i and λ_j. When investors choose to disregard the newswire content related to a particular firm, new information about that firm cannot affect the valuation of others. As the consumption of signals related to that firm increases, however, investors begin to include those signals into their appraisals of other firms. Thus, for a particular firm-pair, the relation between contemporaneous signal correlation and future stock price comovement should increase with the quantity of information that investors choose to consume about each firm.

B.1. Market conditions and information consumption

According to Figure 1, stock return correlations ρ_{ij} will follow the correlation of payoff signals π_{ij} more closely as total information consumption Λ increases. When the value of an asset rises, investors must hold that additional asset value in order for the asset market to clear. Therefore, there will be more aggregate demand for information about high-value assets when many assets are highly valued. Thus, the relation between qualitative similarity and future price comovement should become stronger as aggregate market levels rise. Changing asset values will be measured by the total return of the CRSP Value Weighted Index, R_t^{Mkt}, during period t. Panel A of Figure 3 portrays the level and return of the CRSP Market Weighted Index over the entire sample period. The market loses and regains half of its value during this span, providing ample opportunity to examine how information consumption responds to market-wide stock returns.

I use the daily return standard deviation σ_t^{Mkt} of the CRSP Market Weighted Index during period t to gauge the importance of asset-relevant information when equity payoffs become less predictable. In times of uncertainty, the marginal benefit of observing additional signals rises, causing market-wide information consumption Λ to increase. Thus, the covariance of payoff signals, π_{ij}, will be a better predictor of stock return correlation ρ_{ij} when market-wide uncertainty σ_t^{Mkt} is high. In untabulated results, an alternative measure of payoff predictability, the Chicago Board Options Exchange Market Volatility Index (VIX), is also included as an interaction variable and the inferences are unchanged. Panel B of Figure 3 shows that both measures of uncertainty are highly correlated throughout the sample.

According to the information-driven comovement hypothesis, price comovement will be highest when investors are making correlated inferences about the values of many assets. As demand for asset-specific information Λ increases, however, the pairwise return correlations ρ_{ij} between firms should approach the covariance of their payoff signals π_{ij}. Thus, the relation between qualitative similarity and future price comovement should vary inversely with aggregate return correlation. The variable $\overline{\rho_t}$, defined as the sample average of all pairwise return correlations ρ_{ijt} in a given period t, is used to capture the aggregate level of price comovement. According to Panel C of Figure 3, average return correlation rose as high as 61.8% in the third quarter of 2011.

Most of the subsequent analysis will center on the following basic regression model:

$$\begin{aligned}\rho_{ijt+1} = \beta_0 &+ \beta_1 \max_{k\in i,j} Size_{kt} + \beta_2 \max_{k\in i,j} \sigma_{kt} + \beta_3 \max_{k\in i,j} \overline{WireSim_{kt}^{all}} \\ &+ \beta_4 \max_{k\in i,j} \sum HobSim_{kt} + \beta_5 \pi_{ijt} + \beta_6 \left(\pi_{ijt} \times R_t^{Mkt}\right) \\ &+ \beta_7 \left(\pi_{ijt} \times \sigma_t^{Mkt}\right) + \beta_8 \left(\pi_{ijt} \times \overline{\rho_t}\right) + \sum_{k=9}^{K} \beta_k Control_{kijt} + \alpha_{t+1} \\ &+ \gamma_{i\wedge j} + \delta_{i\vee j} + +\varepsilon_{ijt+1}\end{aligned} \tag{5}$$

where α_{t+1} is a time series fixed effect, $\gamma_{i\wedge j}$ is a panel effect for a unique pair of firms i and j, and $\delta_{i\vee j}$ is a panel effect for each individual firm i or j. The first four variables in Equation (5) account for cross-sectional differences in average correlations based on individual firm characteristics. First, the market capitalizations of individual firms are calculated on the final trading day of period $t-1$, and the variable $\max_{k\in i,j} Size_{kt}$ represents the maximum market value between the firms i and j. Next, the daily return standard deviation is calculated for each firm over all of the trading days in period t, and $\max_{k\in i,j} \sigma_{kt}$ is the maximum of the two standard deviations. Finally, controls based on both measures of total payoff covariance introduced in Section III.A are included to account for firms k with higher average levels of newswire similarity, $\max_{k\in i,j}\overline{WireSim_{kt}^{all}}$, and total product similarity, $\max_{k\in i,j}\sum HobSim_{kt}$. The qualitative similarity of payoff signals, π_{ijt}, is either the textual similarity of all content appearing on the Reuters IDN, $WireSim_{ijt}^{all}$, the textual similarity of content contributed by *Reuters News*, $WireSim_{ijt}^{rtrs}$, or the Hoberg and Phillips (2010a), and (2010b) product similarity score, $HobSim_{ijt}$.

As written, the disturbances estimated from Equation (5) contain some unfavorable structure. Equation (5) attempts to measure the change in future return correlation that would result from a hypothetical change in contemporaneous qualitative similarity. It is possible that contemporaneous changes in qualitative similarity are responses to changes in return correlation earlier in the same period. Therefore, the specification should also account for the current period's, and possibly even earlier periods', observations of pairwise return correlation.

Next, all the estimated return correlations have a value bound between -1 and 1, but the error term ε_{ijt+1} is assumed to be distributed over a range of $-\infty$ to ∞. To improve the accuracy of the coefficient standard errors, the Fisher transformation is applied to the correlation estimates:

$$z_{ijt} = \frac{1}{2}\ln\frac{1+\rho_{ijt}}{1-\rho_{ijt}} \tag{6}$$

The variable z_{ijt} is an approximate variance-stabilizing transformation for ρ_{ijt} when the stock returns of firms i and j follow a bivariate normal distribution. The transformed pairwise return correlation z_{ijt+1} at time $t+1$ for firms i and j is also related to the transformed return correlation of the same firm-pair at all other points in time due to persistent firm-specific characteristics. The cross-sectional disturbances are also likely to have structure induced by firm-specific relations.

Taken together, these concerns motivate the following model with transformed and lagged dependent variables and time series fixed effects α_{t+1}:

$$
\begin{aligned}
z_{ijt+1} = \sum_{s=0}^{S} \phi_s z_{ijt-s} + \beta_0 + \beta_1 \max_{k \in i,j} Size_{kt} + \beta_2 \max_{k \in i,j} \sigma_{kt} \\
+ \beta_3 \max_{k \in i,j} \overline{WireSim_{kt}^{all}} + \beta_4 \max_{k \in i,j} \sum HobSim_{kt} + \beta_5 \pi_{ijt} \\
+ \beta_6 \left(\pi_{ijt} \times R_t^{Mkt}\right) + \beta_7 \left(\pi_{ijt} \times \sigma_t^{Mkt}\right) + \beta_8 \left(\pi_{ijt} \times \bar{\rho}_t\right) \\
+ \sum_{k=9}^{K} \beta_k Control_{kijt} + \alpha_{t+1} + \gamma_{i \wedge j} + \delta_{i \vee j} + + \varepsilon_{ijt+1}
\end{aligned} \tag{7}
$$

Unfortunately, OLS estimation of Equation (7) would still be biased and inconsistent. Therefore, I proceed with the dynamic panel estimator (henceforth DPE) proposed by Arellano and Bover (1995) and Blundell and Bond (1998).

The relation between aggregate information consumption and market conditions is analyzed in Table III. $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$ are binary variables indicating that both firms had some positive volume of text during period t appearing on the Reuters IDN or generated by *Reuters News*, respectively. Hoberg and Phillips (2015c) only make their product score publicly available for firms having pairwise similarities that are above a certain threshold.[13] The binary variable $HobDum_{ijt}$ is set to 1 if both firms i and j are above this minimum level. The newswire binanry

[13] From the Hoberg-Phillips Industry Classification Library (Hoberg and Phillips 2015c), "the TNIC-3 classification data only records firms having pairwise similarities with a given firm *i* that are above a threshold as required based on the coarseness of the three digit SIC classification. The level of coarseness of TNIC-3 thus matches that of three digit SIC codes, as both classifications result in the same number of firm-pairs being deemed related."

variables, $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$, are necessary to differentiate when qualitative similarity is 0 because information about the two firms was unrelated, or because one of the firms did not have a positive text volume during the period. Conversely, the product similarity binary variable, $HobDum_{ijt}$, is necessary because product similarity scores are only available for firm-pairs whose products are somewhat similar. Thus, $HobSim_{ijt}$ is a censored measurement of the actual textual simalirity between 10-K product descriptions. To mitigate the impact of this censored variable, I perform some of my analysis on a subsample of firms whose product similarity scores are above the minimum threshold ($HobDum_{ijt} = 1$) at least once during in my sample.

All of the systematic and alternative controls introduced in Box (2017) are included in every specification. A description of these variables is also provided Table A-1. The significance of the qualitative similarity measure calculated from all text appearing on the Reuters IDN $WireSim_{ijt}^{all}$ is not diminished with the inclusion of the interacted variables.

When firm documents are constructed from text combined across all attributions, the relation between contemporaneous newswire similarity, $WireSim_{ijt}^{all}$, and future return correlation ρ_{ijt+1} becomes stronger as market values rise and aggregate payoff uncertainty σ_t^{Mkt} increases. Thus, the degree to which the signals contained in primary sources of information are incorporated into asset prices is consistent with the predictions of the information-driven comovement hypothesis; more investors are willing to bear the cost of information discovery as the variance of their total payoff increases. The lack of systematic variation in document similarity observed in Box (2017), lessens the possibility that these results stem from market-wide changes in linguistic commonality. Table III also demonstrates that the relation between $WireSim_{ijt}^{all}$ and future comovement ρ_{ijt+1} weakens when aggregate return correlation $\overline{\rho_t}$ increases. Thus, total firm-specific information consumption Λ is low during periods where market-wide comovement is high.

B.2. Firm characteristics and information consumption

The market-level analysis demonstrates how the level of total information consumption Λ changes with the cumulative returns, aggregate volatility and average comovement of the equity market. While these aggregate consumption changes are consistent with the information-driven comovement hypothesis, Table III does not address why investors choose to consume specific pieces of information. The subsequent analysis examines whether firm-specific information consumption λ_i increases as security i's payoffs become less predictable, the stock comprises a larger share of the average investor's portfolio, or signals about the firm contain more information relevant to the valuation of others.

If firm i is larger and more volatile than firm j, investors will consume more information about the former than the latter because signals about firm i can reduce more total payoff variance. Furthermore, as firm i's size and standard deviation increase, the fraction of investors that demand information about firm i will also rise. Thus, according to Figure 1, the price comovement observed between firms i and j should move closer to the covariance of their underlying fundamentals as stock-specific information consumption λ_i grows. Therefore, the relation between qualitative similarity and future return correlation should be strongest when $\max_{k \in i,j} Size_{kt}$ and $\max_{k \in i,j} \sigma_{kt}$ are large.

According to the information-driven comovement hypothesis, a signal must contain information about the value of many assets and must be observed by many investors in order for it to produce comovement. To gauge whether signals about a particular firm contain information about the value of many other companies, I rely on the proxies for total payoff covariance introduced in Section III.A. If there is a strategic complementarity in information acquisition, investors will consume more information about firms with higher aggregate signal correlation. Similar to my strategy for examining how λ_i responds to changes in individual firm size and volatility, the variables $\max_{k\in i,j} \overline{WireSim_{kt}^{all}}$ and $\max_{k\in i,j} \sum HobSim_{kt}$ represent the maximum average payoff covariance of firms i and j. If the relation between signal correlation and future return correlation is stronger when $\max_{k\in i,j} \overline{WireSim_{kt}^{all}}$ or $\max_{k\in i,j} \sum HobSim_{kt}$ is large, then investors coordinate on the types signals that are predicted to generate comovement in the information-driven comovement hypothesis.

Table IV shows how the consumption of qualitative information relates to the value, risk and average payoff covariance of individual firms. Untabulated in each specification are five lags of the systematic variables introduced in Box (2017) and all 14 of the alternative controls included in Table III. For every interacted variable, the multiplier and multiplicand are also included individually as regressors. Inferences from the untabulated variables are the same as in previous tables. Once again, the significance of the qualitative similarity measure calculated from all text appearing on the Reuters IDN $WireSim_{ijt}^{all}$ is not diminished with the inclusion of the interacted variables.

As expected, the coefficient on $WireSim_{ijt}^{all} \times \max_{k\in i,j} Size_{kt}$ is positive and significant, implying that information consumption λ_i is increasing with firm size. Thus, when it is not economical to process all of the content appearing in the IDN feed, investors focus their resources on the subset information that can be used to evaluate the most asset value. The coefficient on $WireSim_{ijt}^{rtrs} \times \max_{k\in i,j} Size_{kt}$ is also positive, but not significant. With regards to payoff uncertainty, the situation is somewhat reversed. Table II shows that journalists focus their production on recent volatility, and, according to Table IV, investors consume more information generated by *Reuters News* when the content relates to firms with high daily return standard deviations. However, information consumption from all primary sources is not similarly influenced by firm volatility.

The results in Table IV are not consistent with a strategic complementarity in information acquisition. The coefficients on $WireSim_{ijt}^{all} \times \max_{k \in i,j} \overline{WireSim_{kt}^{all}}$ and $WireSim_{ijt}^{rtrs} \times \max_{k \in i,j} \overline{WireSim_{kt}^{all}}$ are negative in all specifications. This implies that investors consume less qualitative information about firms whose payoffs covary strongly with most other companies. When investors can eliminate the most uncertainty by observing low covariance signals and inferring the values of the high covariance firms, there is a strategic substitutability in information acquisition. Overall, this finding suggests that, on average, investors do not cluster their information demand on the types of signals that can cause stock price comovement to be high relative to the covariance of underlying fundamentals.

B.3. Firm characteristics, market conditions and information consumption

Direct empirical tests of Veldkamp's (2006a) information-driven comovement hypothesis are complicated by aggregate changes in information consumption. In the model, investors only coordinate on high covariance signals when aggregate information consumption is sufficiently low. As information consumption begins to rise, however, signal demand can spill over into other assets, and a strategic substitutability in information acquisition begins to appear. Thus, whether or not investors coordinate on high covariance signals depends on the aggregate level of information consumption.

Without controlling for market conditions that could influence the overall demand for information, Table IV shows that investors consume less qualitative information about firms whose payoffs have higher average covariances. However, Table III reveals that the aggregate level of information consumption varies with market-wide average comovement, cumulative returns and volatility. Table VI examines whether or not these same market conditions influence how investors choose which types of information to consume. If aggregate information consumption Λ recedes when market returns R_t^{Mkt} are negative, aggregate return volatilities σ_t^{Mkt} are low and average return correlation $\overline{\rho_t}$ is high, then these same conditions should encourage investors to coordinate on a limited number of high covariance signals.

Once again, the multiplier and multiplicand are included individually as regressors for every interacted variable. Therefore, all of the interaction terms appearing in Table III and Table IV are included in the specifications reported in Table VI. Inferences from the untabulated variables are the same as in previous tables. The significance of the qualitative similarity measure calculated from all text appearing on the Reuters IDN $WireSim_{ijt}^{all}$ is not diminished by the inclusion of the additional interacted variables.

When firm documents are constructed from text combined across all attributions, the negative and significant coefficients on $WireSim_{ijt}^{all} \times \max_{k\in i,j} \overline{WireSim_{kt}^{all}} \times R_t^{Mkt}$ and $WireSim_{ijt}^{all} \times \max_{k\in i,j} \overline{WireSim_{kt}^{all}} \times \sigma_t^{Mkt}$ imply that investor coordination on high covariance signals becomes more common when market values are falling and aggregate volatility is low. According to the information-driven comovement hypothesis, these conditions make it less economical to read and evaluate primary sources of information.

Table III reveals that firm-specific information consumption is low when market-wide comovement is high. The positive and significant coefficient on $WireSim_{ijt}^{all} \times \max_{k \in i,j} \overline{WireSim_{kt}^{all}} \times \overline{\rho_t}$ in Table VI implies that coordination on high covariance signals becomes more common as market-wide return correlations $\overline{\rho_t}$ increase. Thus, episodes of high average comovement coincide with an increased consumption of information λ_i about a limited number of firms i with high average payoff covariances $\max_{k \in i,j} \overline{WireSim_{kt}^{all}}$. Consistent with the information-driven comovement hypothesis, I find that market-wide correlations are higher when many investors consume qualitative information about firms whose payoffs covary strongly with many other companies. Likewise, as aggregate correlation falls, so does the demand for these high covariance signals.

Overall, the results in Table IV and Table VI imply that comovement rises when many investors observe a limited number of high covariance signals, but also that demand for low covariance signals is higher on average. Thus, complementarity leads to comovement, but substitutability typically prevails.

IV. Closing remarks

The process by which investors choose the type and quantity of information to consume is poorly understood, but critical to the functioning of financial markets. This paper provides a new empirical strategy for testing models of information choice based on observing the type of information that is consumed and incorporated into asset prices. Consistent with a theoretical model presented by Veldkamp (2006a), I find that price comovement is high relative to the covariance of underlying fundamentals when investors cluster their information demand on just a few firms whose payoffs covary strongly with many other companies. However, as the breadth of information consumption increases, I find that stock return correlations move closer to their fundamental covariances. Overall, my findings imply that investor information consumption choices are influenced by a market for information.

References

Ahern, Kenneth R., and Denis Sosyura. 2014. "Who Writes the News? Corporate Press Releases during Merger Negotiations." *The Journal of Finance Vol. LXIX, No. 1* 241-291.

Andrei, Daniel, and Michael Hasler. 2016. "Dynamic Attention Behavior Under Return Predictability." *Working paper.*

Andrei, Daniel, and Michael Hasler. 2015. "Investor Attention and Stock Market Volatility." *The Review of Financial Studies / v 28 n 1* 33-72.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics 68 (1)* 29-51.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *The Journal of Finance Vol. LXVIII, No. 3* 929–985.

Ballinger, Gary A. 2004. "Using Generalized Estimating Equations for Longitudinal Data Analysis." *Organizational Research Methods vol. 7 no. 2* 127-150.

Bekaert, Geert, Robert J. Hodrick, and Xiaoyan Zhang. 2009. "International Stock Return Comovements." *The Journal of Finance Vol. LXIV, No. 6* 2591-2627.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics 87 (1)* 115-143.

Box, Travis. 2017. "Qualitative Similarity and Stock Price Comovement." *Working paper.*

Brandt, Michael W., Alon Brav, John R. Graham, and Alok Kumar. 2010. "The Idiosyncratic Volatility Puzzle: Time Trend of Specualtive Episode." *The Review of Financial Studies / v 23 n 2* 865-899.

Brockman, Paul, Ivonne Liebenberg, and Maria Schutte. 2010. "Comovement, information production, and the business cycle." *Journal of Financial Economics 97* 107-129.

Cameron, A. Colin, and Pravin K. Trivedi. 2013. *Regression Analysis of Count Data, Second Edition.* New York: Cambridge University Press.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multi-way Clustering." *Journal of Business and Economic Statistics 29(2)* 238-249.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *The Journal of Finance Vol. LVI, No. 1* 1-43.

Fang, Lily, and Joel Peress. 2009. "Media Coverage and the Cross-section of Stock Returns." *The Journal of Finance Vol. LXIV, No. 5* 2023-2052.

Gardiner, Joseph C, Zhehui Luo, and Lee Anne Roman. 2009. "Fixed effects, random effects and GEE: what are the differences?" *Statistics in Medicine Volume 28, Issue 2* 181–360.

Gardner, William, Edward P Mulvey, and Esther C. Shaw. 1995. "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models." *Psychological Bulletin, Vol 118(3)* 392-404.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The Amercian Economic Review, Vol. 70, No. 3* 393-408.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen, and Bernard Yeung. 2015. "Information, Analysts, and Stock Return Comovement." *The Review of Financial Studies / v 28 n 11* 3153-3187.

Hardin, James W., and Joseph M. Hilbe. 2013. *Generalized Estimating Equations, Second Edition.* Boca Raton, FL: CRC Press.

Hoberg, Gerard, and Gordon Phillips. 2010b. "Dynamic Text-Based Industries and Endogenous Product Differentiation." *NBER Working Papers 15991, National Bureau of Economic Research, Inc.*

Hoberg, Gerard, and Gordon Phillips. 2015c. *Hoberg-Phillips Industry Classification Library.* Accessed 11 6, 2015. http://cwis.usc.edu/projects/industrydata/industryclass.htm.

Hoberg, Gerard, and Gordon Phillips. 2010a. "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *The Review of Financial Studies / v 23 n 10* 3773-3811.

Irvine, Paul J., and Jeffrey Pontiff. 2009. "Idiosyncratic Return Volatility, Cash Flows, and Product Markets." *The Review of Financial Studies / v 22 n 3* 1149-1177.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A Rational Theory of Mutual Funds' Attention Allocation." *Econometrica 84 (2)* 571-626.

Ledoit, Olivier, and Michael Wolf. 2003. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance 10* 603-621.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika Vol. 73* 13-22.

Liberti, José María, and Mitchell A. Petersen. 2017. "Information: Hard and Soft." *Working paper.*

Loughran, Tim, and Bill McDonald. 2015. "Information Decay and Financial Disclosures." *Working paper.*

Mondria, Jordi. 2010. "Portfolio choice, attention allocation, and price comovement." *Journal of Economic Theory 145* 1837-1864.

Pan, Wei. 2001. "Akaike's Information Criterion in Generalized Estimating Equations." *Biometrics Vol. 57, No. 1* 120-125.

Peng, Lin, and Wei Xiong. 2006. "Investor attention, overconfidence and category learning." *Journal of Financial Economics 80* 563-602.

Pew Research Center. 2011. *How News Happens.* January 11.

Pindyck, Robert S., and Julio Rotemberg. 1993. "The Comovement of Stock Prices." *The Quarterly Journal of Economics, Vol. 108, No. 4* 1073-1104.

Veldkamp, Laura. 2011. "Information Choice with Sunstitutability in Actions." In *Information Choice in Macroeconomics and Finance*, by Laura Veldkamp, 83-138. Princeton, New Jersey: Princeton University Press.

Veldkamp, Laura. 2006a. "Information Markets and the Comovement of Asset Prices." *Review of Economic Studies 73* 823-845.

Table I
Summary statistics for production regressions

This table presents summary statistics for the variables appearing in Equation (4). $WrdCnt_{it+1}^{rtrs}$ is the total number of words written about firm i and distributed by *Reuters News* during each 6-month period t, and $WrdCnt_{it}^{firm}$ is the total number of words contributed by all other attributions. $AnaNum_{it}$ is the number of unique analysts with an earnings prediction recorded in the I/B/E/S database during period t. $\overline{\rho_{it}}$ is calculated by averaging ρ_{ijt}, the Pearson correlation in the daily stock returns of firms i and j, over all firms j. Similarly, $\overline{WireSim_{it}^{all}}$ is firm i's average newswire similarity $WireSim_{ijt}^{all}$ over all firms j. σ_{it} is firm i's daily stock return standard deviation during period t.

	Mean	Std Dev	P1	P5	P50	P95	P99
$WrdCnt_{it}^{firm}$	4,048.31	6,117.14	0	0	2,457	13,003	25,409
$WrdCnt_{it}^{rtrs}$	554.15	2,116.46	0	0	83	2,133	7,872
$AnaNum_{it}$	9.53	8.43	0	0	8	26	36
$\overline{WireSim_{it}^{all}}$	0.00	0.02	-0.07	-0.04	0.00	0.03	0.06
$\overline{\rho_{it}}$	0.29	0.13	0.05	0.11	0.27	0.53	0.67
σ_{it}	2.74	1.69	0.82	1.08	2.33	5.80	9.04

Table II
Information production and firm characteristics

This table reports the estimation of Equations (3) and (4). $WrdCnt_{it+1}^{rtrs}$ is the total number of words written about firm i and distributed by *Reuters News* during each 6-month period t, and $WrdCnt_{it}^{firm}$ is the total number of words contributed by all other attributions. $AnaNum_{it}$ is the number of unique analysts with an earnings prediction recorded in the I/B/E/S database during period t. $\overline{\rho_{it}}$ is calculated by averaging ρ_{ijt}, the Pearson correlation in the daily stock returns of firms i and j, over all firms j. Similarly, $\overline{WireSim_{it}^{all}}$ is firm i's average qualitative similarity $WireSim_{it}^{all}$ over all firms j. $WireDum_{it}^{all}$ is a binary variable set to 1 whenever firm i has any positive number of words appearing on the Reuters Integrated Data Network during period t. $SizeDec_{it}$ is firm i's NYSE decile based on market value from the last trading day of period $t-1$, and σ_{it} is firm i's daily stock return standard deviation during period t. A description for all other included variable calculations is provided in Panel B of Table A-1. A generalized estimating equations approach, specified with a negative binomial distribution and an autoregressive correlation structure, is used when the dependent variable measures future information production, either $WrdCnt_{it+1}^{firm}$, $WrdCnt_{it+1}^{rtrs}$ or $AnaNum_{it+1}$. Ordinary least squares is used when the dependent variable measures future average comovement, $\overline{\rho_{it+1}}$. The t-statistics (reported in parenthesis) in the information production specifications are calculated from standard errors clustered by firm, and t-statistics in the comovement specification are derived from standard errors clustered by firm and time using the Cameron, Gelbach and Miller (2011) multi-way clustering procedure. * and ** represent significance at the 5% and 1% level, respectively.

Table II—Continued

	Generalized Estimating Equations—Negative Binomial Distribution			Ordinary Least Squares
	$WrdCnt_{it+1}^{firm}$	$WrdCnt_{it+1}^{rtrs}$	$AnaNum_{it+1}$	$\overline{\rho_{it+1}}$
$BetaDec_{it}$	0.0154** (5.849)	0.0277** (4.290)	0.00730** (6.532)	0.0180** (4.072)
$Bk/MktDec_{it}$	-0.0108** (-3.523)	0.0425** (5.383)	-0.00299* (-2.175)	0.0111** (4.286)
$MomDec_{it}$	0.00301* (2.546)	0.0195** (5.146)	-0.00158** (-3.582)	-0.000288 (-0.0970)
$AmiDec_{it}$	-0.0586** (-9.683)	-0.187** (-9.950)	-0.0596** (-20.95)	0.0101* (2.015)
$PrcDec_{it}$	-0.0322** (-8.331)	-0.0573** (-5.983)	-0.00437** (-2.706)	0.00700** (2.839)
$InstDec_{it}$	0.00185 (0.552)	-0.0168 (-1.893)	0.0214** (12.62)	0.000468 (0.302)
$SP500_{it}$	0.255** (7.822)	0.295** (5.300)	0.0990** (5.005)	0.0511** (2.601)
$SizeDec_{it}$	0.0524** (8.336)	0.198** (11.73)	0.0363** (14.43)	0.0167* (1.987)
σ_{it}	-0.0679** (-11.15)	0.131** (5.868)	-0.0200** (-7.899)	-0.0168 (-0.517)
$WireDum_{it}^{all}$	0.762** (30.05)	0.537** (7.578)	0.0337** (4.249)	0.0185 (1.408)
$\overline{WireSim_{it}^{all}}$	-0.919** (-6.838)	-2.634** (-4.877)	0.162** (3.289)	0.449** (3.214)
$\sum HobSim_{it}$	-0.00622** (-2.989)	-0.0341** (-5.908)	0.00124 (1.192)	0.00940** (4.943)
$WrdCnt_{it}^{firm}/1{,}000$		0.0299** (12.00)	0.000527 (1.028)	0.000348 (0.617)
$WrdCnt_{it}^{rtrs}/1{,}000$	0.00550 (1.148)		0.000245 (0.241)	-0.00158 (-0.855)
$AnaNum_{it}$	0.0137** (8.858)	0.00955 (1.945)		-0.00276** (-3.788)
$\overline{\rho_{it}}$	-0.00634 (-1.707)	-0.0354** (-2.905)	0.00883** (6.137)	0.398** (18.41)
Working Correlation Matrix	AR(1)	AR(1)	AR(1)	
Time Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	No	No	No	Yes
R-squared				0.778
Dispersion	2.012	4.370	0.837	
Observations	40,155	40,155	40,155	40,155

Table III
Market conditions and information consumption

The dependent variable in all specifications is the Fisher transformation z_{ijt+1} of the Pearson correlation ρ_{ijt+1} calculated from the daily returns of firms i and j in excess of the risk free rate for each 6-month period $t+1$. The binary variables $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$ are set to 1 whenever both firms have some positive number of total words transmitted across the Reuters Integrated Data Network and *Reuters News*, respectively. Qualitative similarity measures $WireSim_{ijt}^{all}$ and $WireSim_{ijt}^{rtrs}$ are defined in Box (2017). $HobDum_{ijt}$ is a binary variable set to 1 if both firms i and j are members of the same TNIC-3 industry, as defined in the online Hoberg-Phillips Industry Classification Library. $HobSim_{ijt}$ is the yearly firm-by-firm pairwise product similarity. Firm *i*'s and *j*'s average qualitative similarity is calculated for each period t, and $\max_{k\in i,j} \overline{WireSim_{kt}^{all}}$ is the standardized maximum average qualitative similarity between both firms. Firm *i*'s and *j*'s average product similarity score is calculated for each period t, and $\max_{k\in i,j} \sum HobSim_{kt}$ is the standardized maximum score between both firms. Similarly, $\max_{k\in i,j} Size_{kt}$ and $\max_{k\in i,j} \sigma_{kt}$ are the standardized maximum market value and daily return standard deviation between the firms. The market condition variables R_t^{Mkt}, σ_t^{Mkt}, and $\bar{\rho}_t$, defined in Figure 3, are standardized with a mean of 0 and a standard deviation of unity. A description for all other included variable calculations is provided in Table A-1. Results are generated using the approach described in Arellano and Bover (1995) and Blundell and Bond (1998) with bias-corrected robust variance-covariance estimates of the model parameters. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis. All of the independent variables are used as predetermined instruments in the dynamic panel estimation. "Systematic lags" refers to the total number of lags included in each specification for the variables z_{ijt}, $BetaDum_{ijt}$, $BetaCorr_{ijt}$, $SizeDum_{ijt}$, $SizeCorr_{ijt}$, $Bk/MktDum_{ijt}$, $Bk/MktCorr_{ijt}$, $MomDum_{ijt}$, $MomCorr_{ijt}$, $IndDum_{ijt}$ and $IndCorr_{ijt}$.

Table III—Continued (Control Variables)

	Sampled from all eligible firms						Sampled from firms sharing TNIC-3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_{ijt}	0.196** (123.7)	0.198** (125.0)	0.195** (122.9)	0.196** (123.5)	0.196** (124.0)	0.198** (125.2)	0.259** (159.0)	0.261** (159.3)
$BetaDum_{ijt}$	0.0263** (12.05)	0.0248** (11.36)	0.0263** (12.07)	0.0263** (12.04)	0.0262** (12.02)	0.0253** (11.65)	-0.0138** (-4.713)	-0.0136** (-4.633)
$BetaCorr_{ijt}$	0.0281** (11.64)	0.0262** (10.87)	0.0283** (11.75)	0.0282** (11.68)	0.0284** (11.78)	0.0277** (11.53)	-0.0213** (-6.527)	-0.0206** (-6.302)
$SizeDum_{ijt}$	0.0716** (7.502)	0.0690** (7.226)	0.0753** (7.883)	0.0740** (7.743)	0.0756** (7.915)	0.0765** (8.030)	0.178** (15.01)	0.199** (16.89)
$SizeCorr_{ijt}$	0.0653** (6.601)	0.0625** (6.312)	0.0690** (6.971)	0.0677** (6.839)	0.0694** (7.008)	0.0704** (7.130)	0.179** (14.58)	0.200** (16.39)
$Bk/MktDum_{ijt}$	0.109** (26.37)	0.110** (26.46)	0.110** (26.46)	0.111** (26.80)	0.109** (26.35)	0.111** (26.77)	0.0659** (12.73)	0.0800** (15.47)
$Bk/MktCorr_{ijt}$	0.119** (25.96)	0.119** (26.04)	0.119** (26.08)	0.121** (26.43)	0.119** (25.99)	0.120** (26.38)	0.0702** (12.31)	0.0858** (15.07)
$MomDum_{ijt}$	0.0444** (22.74)	0.0434** (22.16)	0.0445** (22.79)	0.0443** (22.65)	0.0451** (23.07)	0.0446** (22.87)	0.0708** (27.00)	0.0694** (26.45)
$MomCorr_{ijt}$	0.0453** (21.06)	0.0440** (20.41)	0.0454** (21.11)	0.0450** (20.94)	0.0461** (21.45)	0.0456** (21.27)	0.0734** (25.45)	0.0716** (24.81)
$IndDum_{ijt}$	0.0687** (11.87)	0.0771** (13.27)	0.0698** (12.04)	0.0718** (12.38)	0.0576** (9.788)	0.0509** (8.676)	0.00355 (1.045)	0.0187** (5.453)
$IndCorr_{ijt}$	-0.0662** (-33.47)	-0.0690** (-35.10)	-0.0657** (-33.19)	-0.0667** (-33.75)	-0.0649** (-32.87)	-0.0594** (-30.13)	-0.0478** (-15.37)	-0.0296** (-9.160)
ρ_{ijt}^{1mo}	0.0185** (20.86)	0.0186** (20.97)	0.0186** (20.92)	0.0186** (21.00)	0.0184** (20.75)	0.0184** (20.78)	0.0498** (43.65)	0.0485** (42.59)
ρ_{ijt}^{2mo}	0.0250** (18.81)	0.0253** (18.98)	0.0248** (18.66)	0.0250** (18.79)	0.0249** (18.73)	0.0252** (18.91)	0.00932** (5.548)	0.0142** (8.421)
$AnaDum_{ijt}$	-0.00176 (-0.220)	0.000415 (0.0515)	0.000782 (0.0974)	0.00170 (0.211)	-0.00196 (-0.245)	-0.00207 (-0.258)	-0.0424** (-4.223)	-0.0138 (-1.379)
$AnaCorr_{ijt}$	-0.00343 (-0.408)	-0.00127 (-0.151)	-0.000935 (-0.111)	-1.49e-05 (-0.00177)	-0.00367 (-0.437)	-0.00367 (-0.438)	-0.0449** (-4.279)	-0.0152 (-1.449)
$InstDum_{ijt}$	0.0216** (4.422)	0.0218** (4.460)	0.0233** (4.767)	0.0238** (4.872)	0.0224** (4.607)	0.0209** (4.321)	0.0402** (6.847)	0.0413** (7.064)
$InstCorr_{ijt}$	0.0224** (4.312)	0.0226** (4.347)	0.0241** (4.644)	0.0247** (4.759)	0.0232** (4.483)	0.0217** (4.221)	0.0395** (6.268)	0.0405** (6.444)
$AmiDum_{ijt}$	0.115** (12.24)	0.112** (11.94)	0.114** (12.19)	0.114** (12.15)	0.112** (11.99)	0.113** (12.05)	0.0737** (6.341)	0.0860** (7.409)
$AmiCorr_{ijt}$	0.117** (12.19)	0.115** (11.89)	0.117** (12.13)	0.117** (12.10)	0.115** (11.94)	0.115** (12.00)	0.0707** (5.924)	0.0834** (6.993)
$PrcDum_{ijt}$	0.0155** (4.217)	0.0145** (3.935)	0.0168** (4.584)	0.0168** (4.579)	0.0168** (4.567)	0.0159** (4.323)	-0.0697** (-14.68)	-0.0774** (-16.33)
$PrcCorr_{ijt}$	0.0138** (3.469)	0.0127** (3.187)	0.0152** (3.834)	0.0152** (3.834)	0.0152** (3.821)	0.0142** (3.584)	-0.0804** (-15.67)	-0.0888** (-17.33)
$SP500_{ijt}$	0.00408 (1.724)	0.00791** (3.390)	0.00650** (2.743)	0.00849** (3.599)	0.00379 (1.601)	0.00358 (1.520)	0.0294** (11.80)	0.0264** (10.71)

Table III—Continued (Interest Variables)

	Sampled from all eligible firms						Sampled from firms sharing TNIC-3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\max_{k\in i,j} Size_{kt}$	0.00672**	0.00620**	0.00677**	0.00656**	0.00687**	0.00694**	0.00837**	0.00810**
	(18.82)	(17.73)	(18.83)	(18.44)	(19.15)	(19.43)	(18.73)	(18.57)
$\max_{k\in i,j} \sigma_{kt}$	-2.52e-05	-0.000158	8.01e-05	-4.21e-05	8.81e-05	0.000349	-0.0113**	-0.0107**
	(-0.0694)	(-0.436)	(0.221)	(-0.116)	(0.242)	(0.964)	(-28.14)	(-26.70)
$\max_{k\in i,j} \overline{WireSim_{kt}^{all}}$	-0.000226	-0.000265	0.000803**	0.000834**	0.000761**	0.000783**	0.00217**	0.00208**
	(-1.066)	(-1.244)	(4.413)	(4.584)	(4.179)	(4.303)	(9.506)	(9.126)
$\max_{k\in i,j} \sum HobSim_{kt}$	0.0104**	0.0100**	0.0104**	0.0102**	0.00980**	0.0108**	0.0253**	0.0276**
	(38.12)	(36.63)	(37.88)	(37.28)	(34.05)	(40.35)	(50.60)	(65.62)
$WireDum_{ijt}^{all}$	0.0112**	0.0128**						
	(11.42)	(13.14)						
$WireSim_{ijt}^{all}$	0.0565**	0.0890**						
	(9.924)	(17.92)						
$WireDum_{ijt}^{rtrs}$			-0.000974*	-0.000921*				
			(-2.321)	(-2.196)				
$WireSim_{ijt}^{rtrs}$			0.00999**	0.0271**				
			(3.293)	(8.745)				
$HobDum_{ijt}$					0.0159**	0.00849	0.00318**	0.00338**
					(3.631)	(1.959)	(3.596)	(3.837)
$HobSim_{ijt}$					0.221**	-0.202**	0.262**	0.109**
					(2.844)	(-3.565)	(15.03)	(12.35)
$WireSim_{ijt}^{all} \times R_t^{Mkt}$		0.00911*						
		(1.963)						
$WireSim_{ijt}^{all} \times \sigma_t^{Mkt}$		0.0179*						
		(2.512)						
$WireSim_{ijt}^{all} \times \bar{\rho}_t$		-0.0182**						
		(-3.417)						
$WireSim_{ijt}^{rtrs} \times R_t^{Mkt}$				-0.00339				
				(-0.612)				
$WireSim_{ijt}^{rtrs} \times \sigma_t^{Mkt}$				0.0227**				
				(2.614)				
$WireSim_{ijt}^{rtrs} \times \bar{\rho}_t$				0.0158*				
				(2.295)				
$HobSim_{ijt} \times R_t^{Mkt}$						-0.252**		-0.165**
						(-13.86)		(-31.80)
$HobSim_{ijt} \times \sigma_t^{Mkt}$						-0.353**		-0.294**
						(-11.85)		(-38.02)
$HobSim_{ijt} \times \bar{\rho}_t$						0.147**		0.221**
						(6.391)		(38.05)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-pair Panel Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Systematic Lags	5	5	5	5	5	5	5	5
AR(2) Test	0.527	1.289	0.462	0.544	0.217	-0.170	-1.104	-1.566
Observations	1,364,711		1,364,711		1,364,711		1,156,033	

Table IV

Firm characteristics and information consumption

The dependent variable in all specifications is the Fisher transformation z_{ijt+1} of the Pearson correlation ρ_{ijt+1} calculated from the daily returns of firms i and j in excess of the risk free rate for each 6-month period $t+1$. The binary variables $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$ are set to 1 whenever both firms have some positive number of total words transmitted across the Reuters Integrated Data Network and *Reuters News*, respectively. Qualitative similarity measures $WireSim_{ijt}^{all}$ and $WireSim_{ijt}^{rtrs}$ are defined in Box (2017). $HobDum_{ijt}$ is a binary variable set to 1 if both firms i and j are members of the same TNIC-3 industry, as defined in the online Hoberg-Phillips Industry Classification Library. $HobSim_{ijt}$ is the yearly firm-by-firm pairwise product similarity. Firm i's and j's average qualitative similarity is calculated for each period t, and $\max_{k \in i,j} \overline{WireSim_{kt}^{all}}$ is the standardized maximum average qualitative similarity between both firms. Firm i's and j's average product similarity score is calculated for each period t, and $\max_{k \in i,j} \sum HobSim_{kt}$ is the standardized maximum score between both firms. Similarly, $\max_{k \in i,j} Size_{kt}$ and $\max_{k \in i,j} \sigma_{kt}$ are the standardized maximum market value and daily return standard deviation between the firms. A description for all other included variable calculations is provided in Table A-1. Results are generated using the approach described in Arellano and Bover (1995) and Blundell and Bond (1998) with bias-corrected robust variance-covariance estimates of the model parameters. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis. All of the independent variables are used as predetermined instruments in the dynamic panel estimation. "Systematic lags" refers to the total number of lags included in each specification for the variables z_{ijt}, $BetaDum_{ijt}$, $BetaCorr_{ijt}$, $SizeDum_{ijt}$, $SizeCorr_{ijt}$, $Bk/MktDum_{ijt}$, $Bk/MktCorr_{ijt}$, $MomDum_{ijt}$, $MomCorr_{ijt}$, $IndDum_{ijt}$ and $IndCorr_{ijt}$. "Alternative Controls" refers to the inclusion of $AnaDum_{ijt}$, $AnaCorr_{ijt}$, $AmiDum_{ijt}$, $AmiCorr_{ijt}$, $SP500_{ijt}$, $SPVal_{ijt}$, $SPGrw_{ijt}$, $PrcDum_{ijt}$, $PrcCorr_{ijt}$, $InstDum_{ijt}$, $InstCorr_{ijt}$, MSA_{ijt}, ρ_{ijt}^{1mo} and ρ_{ijt}^{2mo} as untabulated controls.

Table IV—Continued

	Sampled from all eligible firms				Sampled from firms sharing TNIC-3	
$\max_{k\in i,j} Size_{kt}$	0.00672**	0.00725**	0.00677**	0.00659**	0.00837**	0.00671**
	(18.82)	(19.53)	(18.83)	(18.64)	(18.73)	(14.30)
$\max_{k\in i,j} \sigma_{kt}$	-2.52e-05	-7.71e-05	8.01e-05	-1.74e-05	-0.0113**	-0.00922**
	(-0.0694)	(-0.212)	(0.221)	(-0.0479)	(-28.14)	(-21.33)
$\max_{k\in i,j} \overline{WireSim^{all}_{kt}}$	-0.000226	-2.62e-05	0.000803**	0.000827**	0.00217**	0.000884**
	(-1.066)	(-0.123)	(4.413)	(4.546)	(9.506)	(3.377)
$\max_{k\in i,j} \sum HobSim_{kt}$	0.0104**	0.0102**	0.0104**	0.0103**	0.0253**	0.0292**
	(38.12)	(37.24)	(37.88)	(37.69)	(50.60)	(62.16)
$WireDum^{all}_{ijt}$	0.0112**	0.0122**				
	(11.42)	(12.42)				
$WireSim^{all}_{ijt}$	0.0565**	0.0818**				
	(9.924)	(14.43)				
$WireDum^{rtrs}_{ijt}$			-0.000974*	-0.000947*		
			(-2.321)	(-2.258)		
$WireSim^{rtrs}_{ijt}$			0.00999**	0.0256**		
			(3.293)	(7.019)		
$HobDum_{ijt}$					0.00318**	0.00313**
					(3.596)	(3.451)
$HobSim_{ijt}$					0.262**	0.103**
					(15.03)	(6.220)
$WireSim^{all}_{ijt} \times \max_{k\in i,j} Size_{kt}$		0.0403**				
		(6.016)				
$WireSim^{all}_{ijt} \times \max_{k\in i,j} \sigma_{kt}$		0.0109*				
		(2.346)				
$WireSim^{all}_{ijt} \times \max_{k\in i,j} \overline{WireSim^{all}_{kt}}$		-0.0184**				
		(-6.799)				
$WireSim^{all}_{ijt} \times \max_{k\in i,j} \sum HobSim_{kt}$		0.00846				
		(1.750)				
$WireSim^{rtrs}_{ijt} \times \max_{k\in i,j} Size_{kt}$				0.00392		
				(0.682)		
$WireSim^{rtrs}_{ijt} \times \max_{k\in i,j} \sigma_{kt}$				0.0229**		
				(5.414)		
$WireSim^{rtrs}_{ijt} \times \max_{k\in i,j} \overline{WireSim^{all}_{kt}}$				-0.00228		
				(-0.911)		
$WireSim^{rtrs}_{ijt} \times \max_{k\in i,j} \sum HobSim_{kt}$				-0.00812**		
				(-2.841)		
$HobSim_{ijt} \times \max_{k\in i,j} Size_{kt}$						0.0205**
						(3.053)
$HobSim_{ijt} \times \max_{k\in i,j} \sigma_{kt}$						-0.0534**
						(-12.50)
$HobSim_{ijt} \times \max_{k\in i,j} \overline{WireSim^{all}_{kt}}$						0.0398**
						(7.917)
$HobSim_{ijt} \times \max_{k\in i,j} \sum HobSim_{kt}$						-0.00545
						(-0.615)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm-pair Panel Effects	Yes	Yes	Yes	Yes	Yes	Yes
Alternative Controls	Yes	Yes	Yes	Yes	Yes	Yes
Systematic Lags	5	5	5	5	5	5
AR(2) Test	0.527	0.812	0.462	0.432	-1.104	-0.374
Observations	1,364,711		1,364,711		1,156,033	

Table V
Analyst following and information consumption

The dependent variable in all specifications is the Fisher transformation z_{ijt+1} of the Pearson correlation ρ_{ijt+1} calculated from the daily returns of firms i and j in excess of the risk free rate for each 6-month period $t+1$. The binary variables $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$ are set to 1 whenever both firms have some positive number of total words transmitted across the Reuters Integrated Data Network and *Reuters News*, respectively. Qualitative similarity measures $WireSim_{ijt}^{all}$ and $WireSim_{ijt}^{rtrs}$ are defined in Box (2017). $HobDum_{ijt}$ is a binary variable set to 1 if both firms i and j are members of the same TNIC-3 industry, as defined in the online Hoberg-Phillips Industry Classification Library. $HobSim_{ijt}$ is the yearly firm-by-firm pairwise product similarity. $EPSSim_{ijt}$ is equal to $N_{ijt}^{an}/\sqrt{N_{it}^{an}N_{jt}^{an}}$ where N_{ij}^{an} is the number of analysts following both firms i and j in a period t, and N_{it}^{an} and N_{jt}^{an} are the number of analysts following firms i and j respectively. Firm i's and j's average qualitative similarity is calculated for each period t, and $\max_{k\in i,j}\overline{WireSim_{kt}^{all}}$ is the standardized maximum average qualitative similarity between both firms. Firm i's and j's average product similarity score is calculated for each period t, and $\max_{k\in i,j}\sum HobSim_{kt}$ is the standardized maximum score between both firms. Similarly, $\max_{k\in i,j} Size_{kt}$ and $\max_{k\in i,j}\sigma_{kt}$ are the standardized maximum market value and daily return standard deviation between the firms. A description for all other included variable calculations is provided in Table A-1. Results are generated using the approach described in Arellano and Bover (1995) and Blundell and Bond (1998) with bias-corrected robust variance-covariance estimates of the model parameters. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis. All of the independent variables are used as predetermined instruments in the dynamic panel estimation. "Systematic lags" refers to the total number of lags included in each specification for the variables z_{ijt}, $BetaDum_{ijt}$, $BetaCorr_{ijt}$, $SizeDum_{ijt}$, $SizeCorr_{ijt}$, $Bk/MktDum_{ijt}$, $Bk/MktCorr_{ijt}$, $MomDum_{ijt}$, $MomCorr_{ijt}$, $IndDum_{ijt}$ and $IndCorr_{ijt}$. "Alternative Controls" refers to the inclusion of $AnaDum_{ijt}$, $AnaCorr_{ijt}$, $AmiDum_{ijt}$, $AmiCorr_{ijt}$, $SP500_{ijt}$, $SPVal_{ijt}$, $SPGrw_{ijt}$, $PrcDum_{ijt}$, $PrcCorr_{ijt}$, $InstDum_{ijt}$, $InstCorr_{ijt}$, MSA_{ijt}, ρ_{ijt}^{1mo} and ρ_{ijt}^{2mo} as untabulated controls.

Table V—Continued

	Sampled from all eligible firms			Sampled from firms sharing TNIC-3			Sampled from firms without product market linkages	
$\max_{k\in i,j} Size_{kt}$	0.00673**	0.00676**	0.00687**	0.00821**	0.00817**	0.00845**	0.00628**	0.00632**
	(18.84)	(18.83)	(19.16)	(18.39)	(18.35)	(18.98)	(18.08)	(18.03)
$\max_{k\in i,j} \sigma_{kt}$	-2.13e-05	9.22e-05	8.05e-05	-0.0108**	-0.0109**	-0.0109**	0.000567	0.000645
	(-0.0586)	(0.254)	(0.222)	(-27.19)	(-27.19)	(-27.41)	(1.561)	(1.775)
$\max_{k\in i,j} \overline{WireSim_{kt}^{all}}$	-0.000224	0.000809**	0.000768**	0.00171**	0.00207**	0.00216**	8.45e-05	0.000909**
	(-1.059)	(4.446)	(4.218)	(6.741)	(9.072)	(9.457)	(0.402)	(5.013)
$\max_{k\in i,j} \sum HobSim_{kt}$	0.0104**	0.0104**	0.00983**	0.0352**	0.0354**	0.0296**	0.0112**	0.0112**
	(37.93)	(37.73)	(34.18)	(82.50)	(82.04)	(56.29)	(41.72)	(41.53)
$EPSSim_{ijt}$	-0.0263	-0.0238	-0.0222	0.119**	0.127**	0.106**	-0.0204	-0.0141
	(-1.584)	(-1.434)	(-1.309)	(21.35)	(23.56)	(16.73)	(-1.008)	(-0.637)
$WireDum_{ijt}^{all}$	0.0113**			0.00708**			0.00923**	
	(11.49)			(6.522)			(9.525)	
$WireSim_{ijt}^{all}$	0.0564**			0.0181**			0.0454**	
	(9.869)			(2.802)			(7.943)	
$WireDum_{ijt}^{rtrs}$		-0.000977*			0.000401			-0.00106*
		(-2.327)			(0.778)			(-2.525)
$WireSim_{ijt}^{rtrs}$		0.00930**			0.0106**			-0.00147
		(3.042)			(2.994)			(-0.485)
$HobDum_{ijt}$			0.0150**			0.00322**		
			(3.463)			(3.641)		
$HobSim_{ijt}$			0.220**			0.216**		
			(2.710)			(11.35)		
$WireSim_{ijt}^{all} \times EPSSim_{ijt}$	0.0292			0.0730*			0.145	
	(0.260)			(2.180)			(0.668)	
$WireSim_{ijt}^{rtrs} \times EPSSim_{ijt}$		0.0929			0.0266			0.287*
		(1.109)			(1.238)			(2.006)
$HobSim_{ijt} \times EPSSim_{ijt}$			-0.315			0.339**		
			(-1.209)			(4.643)		
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-pair Panel Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Alternative Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Systematic Lags	5	5	5	5	5	5	5	5
AR(2) Test	0.548	0.479	0.268	0.317	-0.260	-1.528	-0.637	-0.663
Observations		1,364,711			1,156,033		1,359,859	

Table VI

Firm characteristics, market conditions and information consumption

The dependent variable in all specifications is the Fisher transformation z_{ijt+1} of the Pearson correlation ρ_{ijt+1} calculated from the daily returns of firms i and j in excess of the risk free rate for each 6-month period $t+1$. The binary variables $WireDum_{ijt}^{all}$ and $WireDum_{ijt}^{rtrs}$ are set to 1 whenever both firms have some positive number of total words transmitted across the Reuters Integrated Data Network and *Reuters News*, respectively. Qualitative similarity measures $WireSim_{ijt}^{all}$ and $WireSim_{ijt}^{rtrs}$ are defined in Box (2017). $HobDum_{ijt}$ is a binary variable set to 1 if both firms i and j are members of the same TNIC-3 industry, as defined in the online Hoberg-Phillips Industry Classification Library. $HobSim_{ijt}$ is the yearly firm-by-firm pairwise product similarity. The market condition variables R_t^{Mkt}, σ_t^{Mkt}, and $\overline{\rho_t}$, defined in Figure 3, are standardized with a mean of 0 and a standard deviation of unity. Firm i's and j's average qualitative similarity is calculated for each period t, and $\max_{k\in i,j}\overline{WireSim_{kt}^{all}}$ is the standardized maximum average qualitative similarity between both firms. Firm i's and j's average product similarity score is calculated for each period t, and $\max_{k\in i,j}\sum HobSim_{kt}$ is the standardized maximum score between both firms. A description for all other included variable calculations is provided in Table A-1. Results are generated using the approach described in Arellano and Bover (1995) and Blundell and Bond (1998) with bias-corrected robust variance-covariance estimates of the model parameters. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis. All of the independent variables are used as predetermined instruments in the dynamic panel estimation. "Systematic lags" refers to the total number of lags included in each specification for the variables z_{ijt}, $BetaDum_{ijt}$, $BetaCorr_{ijt}$, $SizeDum_{ijt}$, $SizeCorr_{ijt}$, $Bk/MktDum_{ijt}$, $Bk/MktCorr_{ijt}$, $MomDum_{ijt}$, $MomCorr_{ijt}$, $IndDum_{ijt}$ and $IndCorr_{ijt}$. "Alternative Controls" refers to the inclusion of $AnaDum_{ijt}$, $AnaCorr_{ijt}$, $AmiDum_{ijt}$, $AmiCorr_{ijt}$, $SP500_{ijt}$, $SPVal_{ijt}$, $SPGrw_{ijt}$, $PrcDum_{ijt}$, $PrcCorr_{ijt}$, $InstDum_{ijt}$, $InstCorr_{ijt}$, MSA_{ijt}, ρ_{ijt}^{1mo} and ρ_{ijt}^{2mo} as untabulated controls. " Market Condition Interactions " refers to the inclusion of all interactions introduced in Table III. "Firm Characteristic Interactions" refers to the inclusion of all interactions introduced in Table IV.

Table VI—Continued

Panel A: Market and $\max_{k\in i,j}\overline{WireSim_{kt}^{all}}$ interactions			
	Sampled from all eligible firms		Sampled from firms sharing TNIC-3
$WireDum_{ijt}^{all}$	0.0172** (17.99)		
$WireSim_{ijt}^{all}$	0.0840** (15.98)		
$WireDum_{ijt}^{rtrs}$		-0.000773 (-1.840)	
$WireSim_{ijt}^{rtrs}$		0.0239** (6.815)	
$HobDum_{ijt}$			0.00284** (3.244)
$HobSim_{ijt}$			0.0473** (5.291)
$\max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times R_t^{Mkt}$	0.000785** (3.412)	0.000849** (4.389)	-8.97e-05 (-0.323)
$\max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \sigma_t^{Mkt}$	0.000595 (1.637)	0.000706* (2.291)	0.00206** (4.652)
$\max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \overline{\rho}_t$	0.000153 (0.566)	-5.22e-05 (-0.223)	0.000607 (1.805)
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times R_t^{Mkt}$	-0.0157** (-4.198)		
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \sigma_t^{Mkt}$	-0.0359** (-5.634)		
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \overline{\rho}_t$	0.0309** (6.570)		
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times R_t^{Mkt}$		0.0220** (3.628)	
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \sigma_t^{Mkt}$		0.0320** (2.930)	
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \overline{\rho}_t$		-0.0199* (-2.210)	
$HobSim_{ijt} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times R_t^{Mkt}$			-0.0335** (-6.564)
$HobSim_{ijt} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \sigma_t^{Mkt}$			-0.119** (-12.98)
$HobSim_{ijt} \times \max_{k\in i,j}\overline{WireSim_{kt}^{all}} \times \overline{\rho}_t$			0.0567** (8.652)
Time Fixed Effects	Yes	Yes	Yes
Firm-pair Panel Effects	Yes	Yes	Yes
Alternative Controls	Yes	Yes	Yes
Firm Characteristic Interactions	Yes	Yes	Yes
Market Condition Interactions	Yes	Yes	Yes
Systematic Lags	5	5	5
AR(2) Test	1.048	0.901	-0.196
Observations	1,364,711		1,156,033

Table VI—Continued

Panel B: Market and $\max_{k\in i,j}\sum HobSim_{kt}$ interactions			
	Sampled from all eligible firms		Sampled from firms sharing TNIC-3
$WireDum_{ijt}^{all}$	0.0143** (14.78)		
$WireSim_{ijt}^{all}$	0.107** (20.38)		
$WireDum_{ijt}^{rtrs}$		-0.000497 (-1.189)	
$WireSim_{ijt}^{rtrs}$		0.0234** (6.759)	
$HobDum_{ijt}$			0.00296** (3.234)
$HobSim_{ijt}$			0.0945** (7.712)
$\max_{k\in i,j}\sum HobSim_{kt} \times R_t^{Mkt}$	-0.00632** (-31.60)	-0.00652** (-32.85)	-0.0128** (-36.24)
$\max_{k\in i,j}\sum HobSim_{kt} \times \sigma_t^{Mkt}$	-0.0115** (-38.78)	-0.0117** (-39.72)	-0.0198** (-37.97)
$\max_{k\in i,j}\sum HobSim_{kt} \times \bar{\rho}_t$	0.00575** (25.18)	0.00580** (25.56)	0.0114** (29.92)
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\sum HobSim_{kt} \times R_t^{Mkt}$	-0.0237** (-4.641)		
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\sum HobSim_{kt} \times \sigma_t^{Mkt}$	-0.0435** (-5.738)		
$WireSim_{ijt}^{all} \times \max_{k\in i,j}\sum HobSim_{kt} \times \bar{\rho}_t$	0.0287** (5.115)		
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\sum HobSim_{kt} \times R_t^{Mkt}$		-0.00347 (-0.472)	
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\sum HobSim_{kt} \times \sigma_t^{Mkt}$		-0.0385** (-3.336)	
$WireSim_{ijt}^{rtrs} \times \max_{k\in i,j}\sum HobSim_{kt} \times \bar{\rho}_t$		0.0310** (3.129)	
$HobSim_{ijt} \times \max_{k\in i,j}\sum HobSim_{kt} \times R_t^{Mkt}$			-0.0271** (-5.048)
$HobSim_{ijt} \times \max_{k\in i,j}\sum HobSim_{kt} \times \sigma_t^{Mkt}$			-0.103** (-12.17)
$HobSim_{ijt} \times \max_{k\in i,j}\sum HobSim_{kt} \times \bar{\rho}_t$			0.123** (20.02)
Time Fixed Effects	Yes	Yes	Yes
Firm-pair Panel Effects	Yes	Yes	Yes
Alternative Controls	Yes	Yes	Yes
Firm Characteristic Interactions	Yes	Yes	Yes
Market Condition Interactions	Yes	Yes	Yes
Systematic Lags	5	5	5
AR(2) Test	-0.562	-1.210	-0.475
Observations	1,364,711		1,156,033

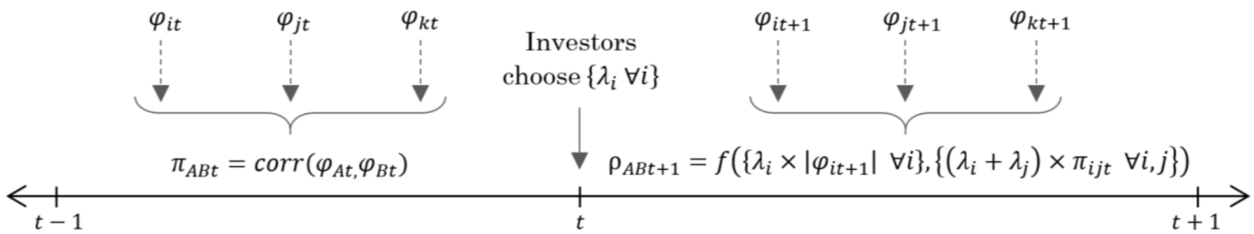
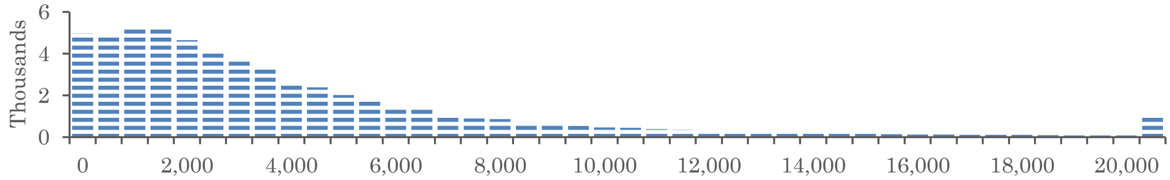
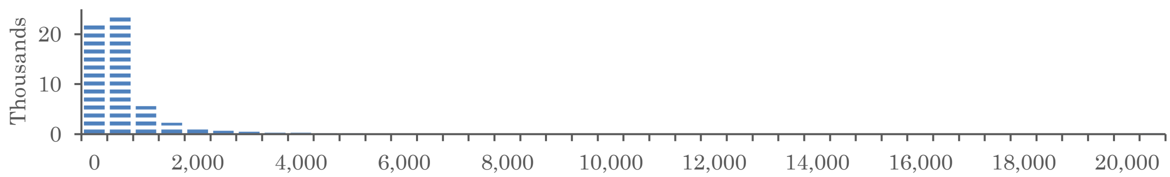


Figure 1. Timeline describing information consumption and price comovement

Panel A: Distribution of $WrdCnt_{it}^{firm}$



Panel B: Distribution of $WrdCnt_{it}^{rtrs}$



Panel C: Distribution of $AnaNum_{it}$

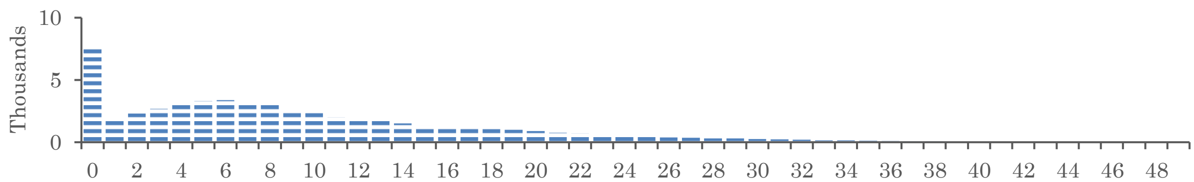
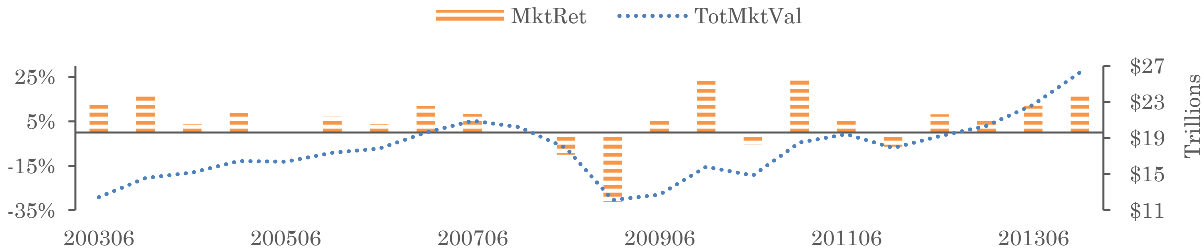


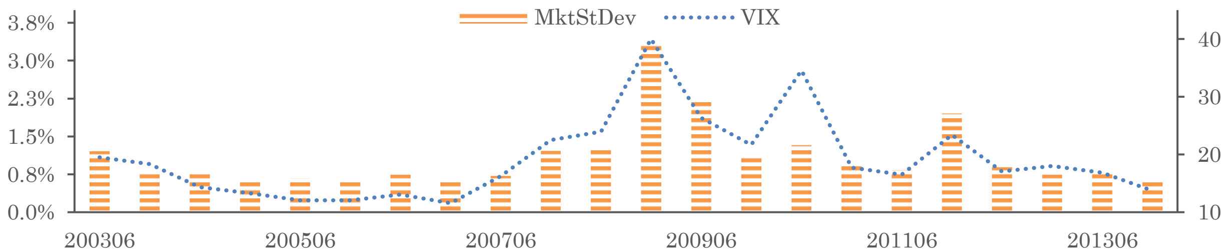
Figure 2. Production variable histograms

Panel A illustrates the pooled distribution of $WrdCnt_{it}^{firm}$, or the total number of words written about firm i and distributed by all attributions other than *Reuters News*. Panel B describes $WrdCnt_{it}^{rtrs}$, or the total number of words written about firm i and distributed by *Reuters News*. Panel C represents the distribution of $AnaNum_{it}$, or the number of unique analysts with an earnings prediction recorded in the I/B/E/S database during period t.

Panel A: Market value and 6-month cumulative return



Panel B: 6-month market daily return standard deviation and VIX level



Panel C: Average pairwise return correlation

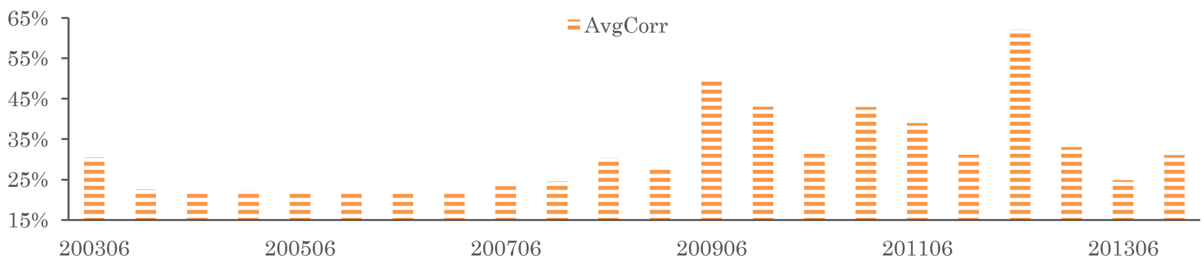


Figure 3. Market-wide financial variables 2003-2013

Panel A illustrates the closing aggregate market level $TotMktVal_t$ (right axis) from the last trading day of period t and the cumulative return R_t^{Mkt} (left axis) of the CRSP Market Weighted Index over period t. Panel B depicts the daily return standard deviation σ_t^{Mkt} (left axis) of the CRSP Market Weighted Index during period t and the Chicago Board of Options Exchange Market Volatility Index VIX_t (right axis) closing value on the last trading day of period t. Panel C represents the $\bar{\rho}_t$, or the sample average of all pairwise return correlations ρ_{ijt} during period t.

A. Supplementary descriptors

Table A-1

Regression variable definitions

Variable	Definition
	Table A-1 Panel A: First appearing in Table I
$WrdCnt_{it}^{firm}$	Total number of words written about firm i and distributed by all attributions other than *Reuters News* during period t.
$WrdCnt_{it}^{rtrs}$	Total number of words written about firm i and distributed by *Reuters News* during period t.
$AnaNum_{it}$	The number of unique analysts with an earnings prediction recorded in the I/B/E/S database during period t.
$\overline{WireSim_{it}^{all}}$	Firm i's average qualitative similarity with all other firms j, $\overline{WireSim_{it}^{all}} = \frac{1}{N-1}\sum_{j\neq i} WireSim_{it}^{all}$, where N is the number of firms with some positive volume of text appearing on the IDN during period t.
$\sum HobSim_{it}$	Overall degree of product similarity calculated as the sum of pairwise similarity scores for each supplier firm in our sample.
$\overline{\rho_{it}}$	Pearson correlation ρ_{ijt} between the daily stock returns of firms i and j averaged over all firms $j \neq i$.
σ_{it}	Firm i's daily stock return standard deviation σ_{it}.
	Table A-1 Panel B: First appearing in Table II
$BetaDec_{it}$	Firm i's NYSE decile based on daily market model beta calculated over two years ending on the last day of period t.
$Bk/MktDec_{it}$	Firm i's NYSE decile based on book-to-market from the most recent quarterly report before the beginning period t.
$MomDec_{it}$	Firm i's NYSE decile based on total return over the previous $t-12$ to $t-2$ months.
$AmiDec_{it}$	Firm i's NYSE decile based on daily Amihud ratio calculated over two years ending on the last day of period t.
$PrcDec_{it}$	Firm i's NYSE decile based on closing price on the last trading day of period $t-1$.
$InstDec_{it}$	Firm i's NYSE decile based on level of institutional holdings during period t.
$SP500_{it}$	Binary variable set to 1 if firm i is a member of the S&P 500 Index on the last trading day of period t.
$SizeDec_{it}$	Firm i's NYSE decile based on market value from the last trading day of period $t-1$.
$WireDum_{it}^{all}$	Binary variable has a value of 1 whenever firm i has some positive number of total words appearing on the Reuters Integrated Data Network during period t.
	Table A-1 Panel C: First appearing in Box (2017)
ρ_{ijt}	Pearson daily return correlation between firms i and j during period t.
z_{ijt}	Fisher transformation of Pearson return correlation. Equal to $\frac{1}{2}\ln\frac{1+\rho_{ijt}}{1-\rho_{ijt}}$.
$BetaDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on daily market model beta calculated over two years ending on the last day of period t.

$BetaCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on daily market model beta calculated over two years ending on the last day of period t. $BetaCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$SizeDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on market value from the last trading day of period $t-1$.
$SizeCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on market value from the last trading day of period $t-1$. $SizeCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$Bk/MktDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on book-to-market from the most recent quarterly report before the beginning period t.
$Bk/MktCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on book-to-market from the most recent quarterly report before the beginning period t. $Bk/MktCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$MomDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on total return over the previous $t-12$ to $t-2$ months.
$MomCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on total return over the previous $t-12$ to $t-2$ months. $MomCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$IndDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same 49-industry portfolio, as defined on Kenneth French's website.
$IndCorr_{ijt}$	Each firm in the sample is assigned to one the 49 industry portfolios, as defined on Kenneth French's website. $IndCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
ρ_{ijt}^{1mo}	Pearson daily return correlation between firms i and j during the last month of period t.
ρ_{ijt}^{2mo}	Pearson daily return correlation between firms i and j during the last two months of period t.
$AnaDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on the number of unique analyst releasing an earnings forecast during period t.
$AnaCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on the number of unique analyst releasing an earnings forecast during period t. $AnaCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$InstDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on level of institutional holdings during period t.
$InstCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on level of institutional holdings during period t. $InstCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$AmiDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on daily Amihud ratio calculated over two years ending on the last day of period t.
$AmiCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on daily Amihud ratio calculated over two years ending on the last day of period t. $AnaCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$PrcDum_{ijt}$	Binary variable set to 1 if both firms i and j are members of the same NYSE decile portfolio based on closing price on the last trading day of period $t-1$.
$PrcCorr_{ijt}$	Each firm in the sample is assigned to NYSE decile portfolios based on closing price on the last trading day of period $t-1$. $PrcCorr_{ijt}$ is the daily return correlation between the portfolios containing firms i and j during period t.
$SP500_{ijt}$	Binary variable set to 1 if both firms i and j are members of the S&P 500 Index on the last trading day of period t.

$\max\limits_{k\in i,j} Size_{kt}$	Standardized maximum market value between firms i and j on the last trading day of period $t-1$.
$\max\limits_{k\in i,j} \sigma_{kt}$	For each period t, the daily return standard deviation is calculated for each firm i and j. $\max\limits_{k\in i,j} \sigma_{kt}$ is the standardized maximum standard deviation between both firms.
$WireDum_{ijt}^{all}$	Binary variable has a value of 1 whenever both firms have some positive number of total words appearing on the Reuters Integrated Data Network.
$\widetilde{WireSim}_{ijt}^{all}$	Document similarity variable is the cosine similarity between the firm vectors i and j in the term-document matrix for period t constructed from text appearing on the Reuters Integrated Data Network.
$\overline{WireSim_{ijt}^{all}}$	For each period in the sample, firms with some relevant text are classified into deciles based on total word counts. The variable $\overline{WireSim_{ijt}^{all}}$ represents the average document similarity between firms appearing in the same word count deciles as i and j during period t. The variable is constructed from all attributions appearing on the Reuters Integrated Data Network.
$WireSim_{ijt}^{all}$	Qualitative similarity variable is calculated by subtracting $\overline{WireSim_{ijt}^{all}}$ from $\widetilde{WireSim}_{ijt}^{all}$.
$WireDum_{ijt}^{rtrs}$	Binary variable has a value of 1 whenever both firms have some positive number of total words appearing originating from *Reuters News*.
$\widetilde{WireSim}_{ijt}^{rtrs}$	Document similarity variable is the cosine similarity between the firm vectors i and j in the term-document matrix for period t constructed from text generated by *Reuters News*.
$\overline{WireSim_{ijt}^{rtrs}}$	For each period in the sample, firms with some relevant text are classified into deciles based on total word counts. The variable $\overline{WireSim_{ijt}^{rtrs}}$ represents the average document similarity between firms appearing in the same word count deciles as i and j during period t. The variable is constructed from text generated by *Reuters News*.
$WireSim_{ijt}^{rtrs}$	Qualitative similarity variable is calculated by subtracting $\overline{WireSim_{ijt}^{rtrs}}$ from $\widetilde{WireSim}_{ijt}^{rtrs}$.
R_t^{Mkt}	Standardized cumulative return of the CRSP Market Weighted Index over period t.
σ_t^{Mkt}	Standardized daily return standard deviation of the CRSP Market Weighted Index during period t.
$\bar{\rho}_t$	Standardized sample average of all pairwise return correlations ρ_{ijt} in a given period t.