Information Flows among Rivals and Corporate Investment*

Darren Bernard†
London Business School
Regents Park
London, United Kingdom NW1 4SA
dbernard@london.edu

Terrence Blackburne
Oregon State University
Austin Hall #430
2751 SW Jefferson Way
Corvallis, OR 97331
terrence.blackburne@oregonstate.edu

Jacob Thornock
Brigham Young University
Marriott School of Business
556 Tanner Building
Provo, UT 84602
thornock@byu.edu

ABSTRACT: Using a novel pairwise measure of firms' acquisition of rivals' disclosures, we show that investment opportunities drive interfirm information flows. We find that these flows predict subsequent mergers and acquisitions as well as how and how much firms invest, relative to rivals. Moreover, firms' use of rivals' information often hinges on the similarities of their products. Our results suggest that rivals' public information, far from being unusable, helps facilitate investment and product decisions, including acquisitions and product differentiation strategies. The findings also support a learning mechanism that could partly underlie the emerging literature on peer investment effects.
JEL classification: D83, D85, G30, G34, L22
Keywords: Information acquisition, information flows, rivalry, investment, M&A

---

* We thank Bill Schwert (editor) and Gerard Hoberg (referee), workshop participants at the 2015 UBCOW Conference, Boston College, Brigham Young University, MIT, University of California at Berkeley, University of Chicago, University of Colorado at Boulder, University of Rochester, Washington State University, and Wharton, and Dave Burgstahler, Ran Duchin, John Gallemore, Jarrad Harford, Jonathan Karpoff, Anya Kleymenova, Cathy Schrand, and Lakshmanan Shivakumar for helpful comments on earlier drafts of the paper. We also extend our gratitude to Samantha Leck for excellent research assistance, to Slavi Marinov for computational assistance, to Matthew Denes for sharing the Compustat-Federal Procurement Data System linking table, and to Gerard Hoberg and Gordon Phillips for providing data related to their text-based industry measures. We appreciate financial support from the London Business School, Oregon State University, the Foster School of Business, and the Marriott School of Business.
† Corresponding author. Tel.: +44 7966 908114.

You know, when I really wanted to understand in depth what a company was doing, Amazon or Apple, I'd get their 10-K and read it. It's wonky, it's this, it's that, but it's the greatest depth you're going to get, and it's accurate. – Steve Ballmer, former Microsoft CEO, April 2017

1. Introduction

Good management requires good information. In this vein, the firm's information plays a central role in models of corporate investment. However, tying investment decisions to the firm's information is empirically challenging because information acquisition is mostly unobservable. Evidence on how the firm acquires information, how good that information is, and how it affects investment decisions is limited.

This paper examines the links between the firm's information acquisition and corporate investment. We develop a novel measure that captures firms' acquisition of information about rival firms. As the quotation from Steve Ballmer suggests, public disclosures can be a key source of information about rivals, as these disclosures reveal rivals' strategic plans, accounting performance, material contracts, product information, plans for capital expenditures (Capex), and other details. We therefore focus on firms' acquisition of information about their rivals in public disclosures and the effects this has on investment decisions.

Our measure is derived from server logs of the Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database, which captures every click or download of public firms' SEC filings. These data resemble those used in recent research (e.g., Ljungqvist and Qian, 2016; Drake et al., 2015; Lee et al., 2015) but with an important difference—our data identify *both* the entity acquiring the information and the entity whose information is acquired. Thus we construct an extensive panel, *firm-pair* data set of each searching firm's acquisition of rival firms' SEC filings.

This measure of interfirm information flows overcomes several limitations in prior research. First, many measures of information acquisition assume that agents take actions to become informed based solely on the availability of information (e.g., Veldkamp, 2011). Rather than assume demand based on supply, our new measure directly captures the acquisition of information *by rivals*. Such search indicates direct employee actions to gather information, which likely affects managers' views directly (as managers ask employees to collect this information) or indirectly (as this information "rises to the top"). In this way, our approach allows us to focus on managers' rather than investors' information acquisition (e.g., Da et al., 2011) as well as assess the type of information acquired (e.g., Badertscher et al., 2013).

Second, other measures typically do not identify either the firm acquiring the information or the firm whose information is acquired. Our measure is at the firm-pair level, which better reflects the intuition that the actions and position of specific rivals influence many investment decisions (e.g., Gilbert and Lieberman, 1987). For instance, Apple's entry into the wearable devices product space with its Apple Watch depended not only on conditions facing the broader tech industry but also on the investment opportunities and behavior of specific rivals in the space, such as FitBit. Our measure captures instances such as Apple's acquisition of FitBit's filings, providing a more direct view into firms' use of information than would be possible using coarser firm- or industry-based measures. As a result, we can examine how pairwise information flows relate to, and predict, pairwise interactions in investment decisions.

We begin by examining how a firm's acquisition of public information depends on its investment opportunities *and* those of its rivals. On one hand, models of observational learning suggest that information flows should be greater when the firm or its rivals have significant investment opportunities whose values are sensitive to managers' decisions about how to exploit them (e.g., Bikhchandani et al., 1998). Managers uncertain about optimal investment can acquire information about other firms to build precision—for example, to predict the success of a potential product introduction. On the other hand, firms' rich set of private information related to investment decisions could make public information about rivals redundant. This critique is particularly applicable to public information in regulatory filings, which often contain a substantial amount of boilerplate (e.g., Dyer et al., 2017; Kravet and Muslu, 2013; Hanley and Hoberg, 2010). Given that this information is compliance-driven, largely backwards-looking, and easily accessible, opportunities to exploit it could be limited.[1]

We model firm-pair information flows as a function of a variety of economic fundamentals, both firm-specific and specific to the firm-pair. The results suggest that the searching firm's average acquisition of rivals' filings increases with its own investment opportunities, as proxied by its market-to-book ratio. We similarly find that the acquisition of a given rival's information increases with the magnitude of that rival's investment opportunities. These results illustrate both firm- and rival-specific drivers of information flows—firms appear to learn from public filings in response to their own as well as specific rivals' investment opportunities.

To ensure the identification of these effects, we employ two quasi-natural experiments in unrelated settings. The first is the U.S. government budget crisis and sequestration cuts of 2011–2013. The crisis reduced current and expected product demand for government contractors, which caused a substantial decline in these firms' investment opportunities. Using a difference-in-differences design, we find that, during the crisis and ensuing imposition of sequestration cuts, information flows decreased more for firms more highly dependent on government contracts than for other public firms.

[1] For example, many studies question the value of this information, even to investors (e.g., Li and Ramesh, 2009; Stice, 1991).

The second quasi-natural experiment exploits changes in import tariffs. Fresard (2010) shows that reductions of industry-level import tariffs significantly reduce the investment opportunities of firms subject to the reductions, via an increase in foreign competition. We find that, in the two years following tariff reductions, the acquisition of rivals' information declines more for firms exposed to tariff reductions. Overall, these results provide consistent evidence that investment opportunities, for both the firm and its rivals, drive the acquisition of rivals' public information.

We next explore *how* rivals' public information is used for the firm's external and internal investment decisions. We report three central findings. First, information flows relate strongly to mergers and acquisitions (M&A) activity. Exploiting the time-specificity of the data to examine pairwise search in event time, we find that information acquisition builds prior to M&A, spikes around announcements, and remains elevated for several months. We also examine the predictive power of pairwise information flows for M&A, which we find is substantial: a one standard deviation increase in information flows corresponds to a greater than 50% increase in the odds of a pairwise acquisition the following year. Further, the evidence suggests an externality of public firm information for private companies: public acquirers collect information on the public peers of future private targets, an effect strongest when the acquirer and target are in different industries—that is, when information asymmetries are high. These results underscore the use of public information for vetting targets and facilitating due diligence and highlight the importance of this information for investments subject to high uncertainty, such as acquisitions that are plausibly differentiating.

Second, we find evidence that pairwise information flows predict pairwise investment levels. Firms appear to learn about rivals' capital and research and development (R&D) investment levels and mimic them in subsequent years, consistent with the intuition that public information lets managers better evaluate investment levels necessary to remain competitive. We also find that the predictive power of information flows for future R&D mimicking is greater when the product similarity of the firms is lower—again, when information asymmetries are high. Together, these results point to corporate learning as a plausible mechanism for peer effects in the investment literature (e.g., Bustamante and Fresard, 2018; Roychowdhury et al., 2018).

Third and finally, we find that acquired information facilitates product differentiation strategies. Although the use of rivals' information to set investment levels could give rise to similar project selection, firms have strong incentives to differentiate their products within a market to gain pricing power and reduce direct competition (e.g., Hoberg and Phillips, 2016; Lieberman and Asaba, 2006). For example, Amazon could use Netflix's disclosures to benchmark spending on streaming content while also evaluating product strategies (e.g., spending on sports content) that creates separation in the product space. Consistent with this, we find that greater firm-pair information flows correspond to subsequent product differentiation *relative to the searched firm*. We also find that this effect is somewhat reversed for firm-pairs that are more dissimilar; information flows between dissimilar firms is more predictive of product mimicking than differentiation. These results suggest that firms use rivals' public information in part to improve product positioning.

Our paper contributes along several dimensions. First, we contribute to the literature on corporate investment. Uncertainty undermines investment (e.g., Guiso and Parigi, 1999). Yet the actions of firms to address information deficits are not well understood. Our findings support a role for public information acquisition in facilitating capital and product investment decisions. The results suggest an economically significant sensitivity of information acquisition to the searching firm's and the searched firm's investment opportunities. We also provide the first direct evidence on relations between *firm-pair* information flows and corporate outcomes, results that highlight the role of public information acquisition in shaping interactions among rivals.

Second, we contribute to the literature on peer effects. Prior work finds that corporate decisions are often a function of the actions of peer firms, but the mechanism for these effects is unclear (e.g., Roychowdhury et al., 2018; Leary and Roberts, 2014; Delong and Deyoung, 2007). Our evidence points to a plausible learning mechanism, as firms appear to acquire information about rivals to facilitate their own investment decisions. This evidence builds on research that provides indirect evidence of learning. For example, Badertscher et al. (2013) show that investment sensitivities of private firms increase in the presence of public peers, and Durnev and Mangen (2009) provide evidence that restatements convey information about the investment projects of restating firms' competitors. Our findings also complement the growing literature that studies how managers learn from the prices of peer firms (e.g., Dessaint et al., 2019; Foucault and Fresard, 2014; Bond et al., 2012). Our evidence suggests corporate learning from peers' public disclosures is also important, plausibly because information in these disclosures is often richer and less noisy than stock prices.

Finally, the study contributes to the body of work on the decision-usefulness of firms' public disclosures. As a public good, information in these disclosures is often maligned as boilerplate and prepared to satisfy regulators more than investors or other users. Our evidence highlights the decision-usefulness of this public information specifically for investment, especially when information asymmetries are high. Firms appear to free ride on rivals' information production, suggesting a byproduct of regulatory requirements for transparency is to facilitate both internal and external investments, such as acquisitions of private companies.

2. Data

2.1. Data collection and sample construction

We measure information acquisition using novel data that capture firms' search activities on the SEC EDGAR repository (e.g., Drake et al., 2015; Drake et al., 2016; Lee et al., 2015). SEC filings contain substantial amounts of firm information, including information about product costs, operational decisions, forecasts, segment disclosures, material contracts, and management discussion of risks and performance, among many other items. As a result, our measure reflects a broad scope of incentives to acquire information (e.g., Li et al., 2013).[2] SEC filings are also audited, easy to obtain, and highly standardized—the structure, basic content, and periodicity of many of these filings are essentially fixed. By directly examining the acquisition of public firms' standardized, mandatory filings, we minimize concerns that differences in the endogenous supply of disclosures or variation in the costs of acquiring and processing them could explain the results.

We start with the log files of EDGAR servers, which host all SEC filings. To measure *firm-to-firm* search, we separate activity of the searching firm, i, from that of the firm being searched for, j. We identify firm i by its IP address and firm j by its Central Index Key (CIK), both of which are recorded in the logs. We match firm j's CIK to its GVKEY in Compustat using the SEC Analytics match table.

Identifying the IP address for firm i requires several steps. We begin by matching firm i's IP address from the server logs to its owner's name, using header files from the American Registry of Internet Numbers (ARIN). To this end, we employ a proprietary query program that scrapes

[2] Of course, our analyses examine the acquisition of SEC filings *specifically via EDGAR*, and some proportion of information acquisition likely occurs through alternative channels. However, few other sources are free, easily accessible, complete, and offer relatively anonymous access. For example, firms tend to monitor traffic to investor relations pages of their websites (Hodge and Pronk, 2006), so rivals are more likely to access filings via the SEC website to the extent they wish to mask information search from disclosing firms.

all IP addresses and their corresponding ownership data (known as WHOIS data) from ARIN. These data identify the owner of each IP address, as of February 2014. We then identify, by manual checks, the searching IP owners' names that correspond to public companies and retrieve these firms' CIKs (and GVKEYs, using the SEC Analytics match table) to identify an initial set of searching firms.[3]

Next, we augment the WHOIS data, using historical IP address ownership records (known as WHOWAS data) for the initial set of searching firms. The use of WHOWAS data is necessary, because the WHOIS data represent IP address ownership at a single point in time (as of 2014 in this case) and IP address ownership can vary over time. Thus, with the full time-series of ownership for each IP address, we identify and remove IP-searching firm matches for those years that the owner was different from that in February 2014.[4] This additional step ensures we identify IP ownership with a high degree of fidelity over the full sample period, which eliminates the possibility of attributing EDGAR search activity to the wrong firm. The resulting sample, which we call the "verified sample," consists of searching firms that have downloaded disclosure filings from EDGAR for which we can identify IP addresses using WHOIS and WHOWAS data and for which Compustat includes requisite data for the analyses.

The sample covers the period 2004–2015, which we choose based on data availability for our primary measures. Because we aim to better understand information flows *among rivals*, for most analyses, the data are restricted to observations of firm-pairs for which firm *i* (the searching firm) and firm *j* (the searched firm) compete in similar product markets, based on the Hoberg and

[3] We originally identified a set of firms in 2014, using EDGAR data of firms' search for years 2004–2011. We subsequently acquired more recent EDGAR data, which we appended to the initial data set to extend the sample period through 2015 for the firms we originally identified.

[4] We cannot form an initial set of searching firms based on WHOWAS data, as they are not obtainable in machine-readable form. To incorporate WHOWAS, we send a separate request by email to ARIN for every IP address we identify based on WHOIS for the initial set of searching firms.

Phillips (2016) text-based network industry classification schema ("TNIC3" industry). The TNIC3 industries are established based on the similarity of mandated product descriptions provided in firms' 10-K filings and are designed to be as coarse as three-digit Standard Industrial Classification (SIC) codes; thus the TNIC3 industries are time-varying, firm-specific, and fairly broad, ensuring they capture even minor competitors. Further, Hoberg and Phillips (2016) provide evidence that this classification system more accurately identifies a firm's actual competitors than fixed schema, such as SIC and North American Industry Classification System (NAICS) codes, and generally does not capture upstream or downstream firms.

2.2. Description of the final sample and data validation

Our final verified sample for the firm-pair analyses consists of 252,370 *i*,*j*,*t* triplets over the sample period. This sample has complete coverage of the search activities on EDGAR from a given IP address, *conditional* on the IP address being included in the sample. In other words, if the sample includes a given IP address for a searching firm *i*, we can observe the full scope of search from that IP address for all other public firms *j* with filings on EDGAR.[5] While this characteristic of the data limits the scope of any potential selection issues, we caveat that the data likely do not include all public firms that search on EDGAR. One reason for this is that our selection criteria exclude IP addresses through which firms do not search on EDGAR above a minimum threshold—this is necessary to make the hand-matching of IP owner name to CIK feasible (see the Online Appendix). Another reason is that the SEC anonymizes the last octet of every IP address, so we cannot attribute an IP address to a specific company, unless it owns the entire final octet. These factors likely account for the reduced sample coverage, relative to

[5] Thus we cannot observe information acquisition by U.S. firms for most foreign rivals (e.g., Apple for Samsung), as these firms are generally not required to provide regulatory filings to the SEC.

Compustat; as we show in Panel A of Appendix B, our sample contains 579 unique searching firms after imposing all data requirements.[6]

Given the incomplete sample coverage, we provide descriptive evidence of the sample composition and identify characteristics that differ from the broader population of firms. First, we compare the distribution of our sample with the Compustat universe by Fama-French 30 industry classification to examine whether some unanticipated selection issue disproportionately excludes firms in certain industries. As we show in Panel B of Appendix B, the sample has broad representation across industries. The industry proportions of our sample, relative to those in Compustat, are visually similar but are frequently statistically different, as shown in the far-right column. Across the 30 industries, three are proportionally more than 2.5% different from Compustat: utilities, healthcare, and banking.[7] Still, the univariate correlation between the two columns in Panel B of Appendix B is roughly 0.90, indicating similarity in the industry proportions. We conclude that the industry distribution in the verified sample generally mirrors the industry distribution of the Compustat universe but acknowledge that some industries are disproportionately represented in our verified sample.

Second, we compare a variety of characteristics of our sample firms to those of the broader population of firms in Compustat. The results, tabulated in Panel C of Appendix B, show that the firms for which we have identified IP addresses are substantially larger (in terms of total assets

[6] Panel A of Appendix B also shows the number of sample firms by year. Loughran and McDonald (2017) find an increasing trend in EDGAR usage over time. Our data are largely consistent with this trend through 2011. As noted above, the sample years 2012–2015 represent augmented data for those firms available in 2011. The declining sample coverage after 2011 appears to relate to IP ownership changes, rather than survivorship issues per se. To this point, the decline in coverage after 2011 also occurs for Standard & Poor's (S&P) 500 firms, which are large and stable. Further, untabulated results show that firms that fall out of the sample are broadly similar to those that remain on dimensions such as size and growth.

[7] The largest proportional difference between the verified sample and Compustat is in banking and insurance: whereas 13% of our sample firms are in banking or insurance, roughly 25% of Compustat firms are. This sample underrepresentation could be due to financial institutions' use of other means to acquire information contained in filings, such as Bloomberg terminals, or because financial institutions are more likely to scrape filings using bots, which our selection criteria exclude (see the Online Appendix).

and market value) and more profitable than the typical firm in Compustat. These differences are expected, given our selection criteria and the SEC's anonymization of the last octet of IP addresses—smaller firms have fewer IP addresses and are likely to search less in total. Still, our results should be interpreted with these characteristics of the sample in mind.

Notwithstanding these comparisons, our sample could differ from the broader Compustat universe in less observable (and less innocuous) ways. For example, the limited sample coverage, due to requirements for hand-matching IP addresses, could reflect some kind of unknown systematic selection bias. We conduct additional tests, detailed in Section 3.3 below, to address these concerns.

3. Investment opportunities and information flows

3.1. Variable measurement and descriptive statistics

Our sample comprises the Cartesian product of ordered pairs of rival firms, searching firm *i* and searched firm *j*. Thus we model information flows among rivals using a *firm-pair* design. This design provides a degree of fineness unique to the literature; we can account for a variety of economic fundamentals shared by the two rival firms in the pair as well as characteristics specific to the broader product space and specific to each firm.

Our first goal is to examine the effects of the firm's *and* its rivals' investment opportunities on information acquisition. Our firm-pair measure of interfirm information acquisition, *Information acquisition*$_{i,j,t}$, is equal to the number of firm *j* SEC filings downloaded from the EDGAR database by firm *i* in year *t* (but excluding search by the firm for its own filings).[8] By using a firm-pair measure as the primary dependent variable, we are modeling *i*'s search for a

[8] This measure captures the acquisition of the full spectrum of SEC filings, as we expect most SEC forms are useful to rivals. For example, while the 10-K arguably contains the stalest information, it also contains the greatest depth of quantitative and qualitative information, nearly all of which is audited. In contrast, the 8-K is narrow in scope but also timely and often includes details about important company events, such as earnings releases.

given rival *j*, as distinct from *i*'s *total* search for all rivals. To capture each firm's investment opportunities, we use its market-to-book assets ratio, as used extensively in prior research (e.g., Jung et al., 1996), defined as total assets plus the market value of equity less the book value of equity, all scaled by total assets (e.g., Nini et al., 2009).[9] We employ market-to-book as a measure of investment opportunities, because it is a forward-looking measure that captures prevailing market views on a company's prospects. Further, it is suitable for a pairwise, cross-sectional research design, which allows us to separately identify how variation in *i*'s and in *j*'s investment opportunities relates to *i*'s average acquisition of rivals' filings.

Because information flows are generally unobservable, an important element of our contribution is to provide initial evidence of other fundamental economic factors that relate to information flows between rivals. Firm-pair factors include measures of product similarity and return correlation, which account for shared economic exposures of the two firms—for instance, to common revenue streams, capital market pressures, or input markets (de Bodt et al., 2018; Hoberg and Phillips, 2010a; Lieberman and Asaba, 2006). To capture compliance uncertainty, the firm-pair factors also include an indicator variable equal to one if the firms share the same auditor during the year (e.g., Hanley and Hoberg, 2010). In addition, we include a measure of the distance between the headquarters of the firms in the pair to capture shared exposures to local economic risks and regulations.

Finally, we examine a number of characteristics of each searching firm *i* and searched firm *j*. For symmetry, we include proxies for profitability, leverage, size (total assets), firm age, and sales growth for both firm *i* and firm *j*. We also include measures of key characteristics of the

[9] Measuring investment opportunities is empirically challenging (e.g., Frank and Goyal, 2003). Our approach to deal with this issue is threefold. First, we employ market-to-book assets (see Table 2), which is one of the most frequently used measures of investment opportunities. Second, we employ two quasi-experimental shocks to investment opportunities (see Tables 3 and 4). Finally, in untabulated analyses, we employ investment opportunities proxied by Tobin's *Q*, as do Erickson and Whited (2012), and find results generally consistent with those in Table 2.

searching firm's product market—namely, measures of product market instability and market concentration (Hoberg et al., 2014). In addition, for the searched *j* firms, we include measures of industry leadership, information supply (i.e., number of filings), and financial distress. See Appendix A for full variable definitions.

Table 1 presents summary statistics for the variables. Panel A shows that *Information acquisition*$_{i,j,t}$ is right skewed, with a median of zero, a mean of approximately one, and a standard deviation of approximately five. These values are low by construction, because the pairwise measure covers the entire product market and TNIC3 industries are fairly broad, meaning that most firms within the industry are only minor rivals.[10] Panel B presents selected univariate correlations. (The Online Appendix presents a complete correlation table.) The correlations suggest a low multicollinearity risk and are generally intuitive. For instance, information flows are strongly positively correlated with firm-pair product similarities and negatively correlated to the geographic distance between the firms in the pair.

3.2. Results—the influence of investment opportunities on information flows

We begin with tests of the effects of the firm's and its rivals' investment opportunities on information acquisition. The results are presented in Table 2. As the dependent variable (*Information acquisition*$_{i,j,t}$) is a count variable and a likelihood ratio test suggests overdispersion, we use the fixed effects negative binomial model of Hausman et al. (1984) (Cameron and Trivedi, 2013).[11] In each panel of Table 2, Columns 1, 3, and 5 present the coefficients from estimating the model, while Columns 2, 4, and 6 present the estimated

[10] Still, search is predictably concentrated *within*, rather than outside, the firm's TNIC3 industry—see the Online Appendix for an illustration of intra- vs inter-industry information flows.

[11] Because the negative binomial model is a conditional model (i.e., not a true fixed effects estimator) and is potentially subject to the incidental parameters problem, we also estimate a Poisson model for Table 2, and the results are similar (see the Online Appendix).

incidence rate ratios.[12] Columns 1 and 2 omit year fixed effects and firm-pair effects, Columns 3 and 4 include only year fixed effects, and Columns 5 and 6 include year fixed effects and firm-pair effects. Year fixed effects account for potential time-series variation in information acquisition, perhaps due to changing use of the EDGAR database or macroeconomic effects. Firm-pair effects control for time-invariant sources of unobservable heterogeneity unique to each *i*, *j* pair, though inclusion of these effects reduces the sample size, as there are many firm-pairs for which search is nil for all years. To ease interpretation, we normalize all nonbinary independent variables to be mean zero and have unit standard deviation. Standard errors are clustered by firm-pair.

Panel A reports the results for the full sample period 2004–2015 using our verified sample.[13] We find that the coefficient on the searching firm's market-to-book is consistently positive and statistically significant across the specifications, consistent with information acquisition being sensitive to the firm's investment opportunity set. When we identify the result strictly on variation within each firm-pair (i.e., when we include firm-pair effects), the incidence rate ratio suggests a one standard deviation increase in market-to-book corresponds to about a 5% increase in information acquisition, on average, of filings of each firm *j* in firm *i*'s product space. We similarly find evidence that *rivals*' investment opportunities influence information acquisition. The coefficient on firm *j* market-to-book is positive and significant in all specifications; again focusing on Columns 5 and 6, we find that a one standard deviation increase in firm *j*'s market-to-book corresponds to a roughly 3% increase in *i*'s acquisition of its information. All together,

[12] The incidence rate ratio (IRR) can be interpreted like an odds ratio. For example, an IRR of 1.2 implies that the predicted value of the dependent variable is 1.2 times greater, given a one unit increase in the independent variable, holding constant other regressors.

[13] Our results are robust to clustering by searching firm or limiting the sample period to 2004–2011 (see Section 2).

the results support the premise that investment opportunities are associated with information flows.

Other economic factors also relate to pairwise information flows. We find that the rival pair's product similarities and return correlation have highly significant effects, consistent with the intuition that shared capital market outcomes and common economic exposures increase information flows between the firms. We similarly find that firm-pairs that share the same auditor access more of one another's filings, as do firm-pairs that are more geographically proximate. These findings are consistent with prior work that suggests audit firms convey unique compliance pressures (Dunn and Mayhew, 2004) and the intuition that geographic proximity captures shared local economic factors and breeds familiarity.

Firm-specific attributes with clear effects on information flows include firm *i*'s product market fluidity (Hoberg et al., 2014), which has a consistently negative and statistically significant coefficient across the specifications. This finding runs counter to the intuition that uncertainty induces information acquisition but is supported by theory that product market instability can make it more difficult to extract decision-relevant information from rivals' public filings (e.g., Moscarini, 2004). The effect of firm *i*'s market concentration is positive and highly significant, consistent with the hypothesis of Hoberg and Phillips (2010b) that firm-specific information is more informative and less costly to acquire in more concentrated industries. We also find that search for firm *j* is greater if the firm is an industry leader, an effect that is incremental to the positive effect of firm *j*'s size. These effects suggest firm *i* searches more for firm *j* as *j* becomes a larger and more dominant rival, indicating an outsized role for industry leaders in reducing uncertainty (e.g., Gilbert and Lieberman, 1987).[14]

---

[14] The mean variance inflation factor (VIF) for our analyses is 1.37, with a maximum VIF of 2.61 for $Size_{j,t}$. All other VIFs are less than 2, which indicates that multicollinearity likely does not affect our inferences.

3.3. Alternative sample construction

As discussed in Section 2, the verified sample is somewhat incomplete as a result of the need to hand-match IP addresses. To address concerns about this incompleteness, we create an alternative sample called the "predicted sample" that is based on a predictive self-search methodology, as discussed in greater detail in the Online Appendix.[15] The intuition for this sample is that users often search for the company's own filings more than for other firms' filings. For example, in untabulated results, we find that users from Apple's IP addresses tend to search for Apple's own SEC filings at about ten times the rate they search for the next-most searched firm's (Alphabet's) filings. Thus, by using the IP that disproportionately searches for a given company as a proxy for the company itself, we can identify the search patterns of many additional firms. This approach identifies 1,279,692 firm-pair observations, about five times more than the verified sample.

Table 2 Panel B presents the results of the analyses based on this predicted sample. We find that the results are similar in sign and magnitude to the effects we show in Panel A. For example, the coefficients on firm *i*'s market-to-book and on firm *j*'s market-to-book remain consistently positive and significant across the specifications. The overall consistency of these results with those in Panel A suggests the influence of unknown potential biases related to the construction of the verified sample is minimal.[16]

[15] We thank the referee for proposing this novel approach.

[16] Another potential issue is that some firms present in the early part of the sample period but missing in the latter part are entirely omitted from the verified sample. To be clear, familiar forms of survivorship, such as those related to delisting or most forms of bankruptcy, are minimized because they do not necessarily cause a reassignment of IP addresses. Still, to the extent some firms are missing, the results based on our verified sample could be subject to survivorship concerns. Thus, as an additional robustness test, the Online Appendix presents the results using the subsample of firms that are members of the S&P 500. These firms are larger and more stable and therefore arguably less susceptible to the kinds of events that generally underlie survivorship concerns than the average firm in the sample. Our findings for Table 2 are similar when we use this subsample as well.

3.4. Quasi-natural experiments: 2011 federal government budget crisis and tariff reductions

Measuring the investment opportunities of the firm as well as those of its rivals is empirically challenging and subject to some measurement error, as noted elsewhere (e.g., Erickson and Whited, 2012). Thus, we employ two shocks to isolate plausibly exogenous variation in the firm's and its rivals' investment opportunities: the U.S. federal government budget crisis, beginning in 2011, and staggered import tariff reductions. Unlike other shocks to investment opportunities used in the literature (e.g., state-level tax rate changes), these shocks are symmetric in the sense that each has the same directional effect on the investment opportunities of the firm *and* its rivals. The shocks are also unrelated to each other, which adds to the tests' discriminant validity; each shock creates substantial and plausibly exogenous variation in investment opportunities, but other causes and consequences of the shocks have little in common. Similar results in two dissimilar contexts point to investment opportunities as an important mechanism that drives information flows.

The U.S. budget crisis and subsequent sequestration began after the Republican Party re-gained control of Congress in early 2011 and used the threat of default as leverage in negotiations with President Obama to secure passage of the Budget Control Act ("BCA"). The BCA led to both a substantial immediate cut in government spending and reduced expectations about growth in government spending, which acted as an unexpected shock to current and future expected demand for industries that generate a significant portion of their revenue from U.S. government contracts.[17] While the changes in fiscal policy had potential economy-wide implications, the effect on a given firm's investment opportunities is most direct for companies that count the U.S. government as a major customer. Therefore we can use this shock to employ

[17] For example, while the FY 2010 budget proposal and the amounts appropriated were similar ($10 billion less than requested), in FY 2011, appropriated amounts were $115 billion less than the amount originally requested. We provide further detail on the passage of the BCA and the events surrounding it in the Online Appendix.

a difference-in-differences design on the basis that reductions in product demand for contractors reduced their and their rivals' investment opportunities.[18]

To identify industries most affected by this shock, we use the linking table developed by Brogaard et al. (2019) to match contracts from the Federal Procurement Data System with firms in Compustat. We measure the percentage of firms' revenue derived from government contracts in 2010 and classify firms for which U.S. government contracts accounted for more than 5% of their revenue as firms for which *High contract*$_i$, an indicator variable, equals one. We then interact an indicator variable for years 2011–2013, *BCA*$_t$, with *High contract*$_i$ to identify the differential effect of the shock for government contractors on interfirm information acquisition relative to other firms. Thus we estimate a difference-in-differences regression of *Information acquisition*$_{i,j,t}$ on *High contract*$_i$ and *BCA*$_t$×*High contract*$_i$, using 2008–2010 as the pre-period and including a large set of economic factors as controls (as in Table 2).

The results of this analysis are presented in Table 3.[19] We find that the coefficient on the interaction of *BCA*$_t$ with *High contract*$_i$ is negative and statistically significant. The reduction in current and future demand for government contractors is associated with a reduction in information acquisition among contractors of approximately 59%, a large effect consistent with the magnitude of the immediate and potential long-run effects of budget sequestration.[20]

Next, we use plausibly exogenous reductions in import tariffs. Fresard (2010) provides evidence that these declines in tariff rates are followed by substantial increases in import

[18] Press reports during the period highlighted the investment cuts likely to follow sequestration. For example, see https://www.cio.com/article/2394658/government/sequestration-threatens-tech-firms--dod-contractors-and-national-security.html.

[19] In Table 3 and those that follow, we tabulate all pairwise (*i*,*j*) control variables (as distinct from those that are searching firm *i*- or searched firm *j*- specific), given that the research design emphasizes pairwise search activity of one firm for a rival firm's SEC filings. The Online Appendix presents Tables 3–8 with full tabulation of controls.

[20] This result is robust to multiple design adjustments. For example, the results are similar in economic and statistical significance when we exclude 2008 (the financial crisis) or limit the sample period to 2010–2011.

penetration and reductions in domestic firms' market-to-book ratios.[21] Fresard (2010) also shows these tariff reductions are not clustered in a specific period, are not systematically related to several dimensions of affected firms' ex ante financing and performance characteristics (such as profitability), and are not fully anticipated by equity markets. Thus tariff cuts represent plausibly exogenous reductions in investment opportunities for the domestic firm and its domestic rivals, via an increase in foreign competition.[22]

We collect product-level import data from the United States International Trade Commission (USITC) for the period 2004–2014 at the four-digit SIC industry level, similar to that compiled for earlier periods by Feenstra (1996) and Feenstra et al. (2002).[23] As in prior studies, these data are limited to manufacturing firms (four-digit SIC codes between 2000 and 3999). We calculate the ad valorem tariff rates for each industry-year as the duties collected by U.S. Customs divided by the free-on-board value of imports. We follow Fresard (2010) and create an indicator variable, *Tariff cut*$_{i,t}$, which equals one if firm *i* experienced a negative change in tariff rates that is three times the median change and not followed by an equivalent increase in the following two years. To examine the effect of the tariff reductions on information flows among rivals, we regress the measure of interfirm search on *Tariff cut*$_{i,t}$ and the firm-pair and firm-specific variables described in Section 3.1.

The results are presented in Table 4. We find a statistically significant reduction in the level of search for rivals' filings following unanticipated tariff reductions—the coefficient on *Tariff cut*$_{i,t}$ is negative and significant. The economic magnitude of the effect is substantial: the

[21] Fresard's (2010) findings are consistent with similar results in other studies, including those of Bernard et al. (2006) and Tybout (2003). Huang et al. (2017) further validate that the tariff cuts are associated with increases in imports and declines in market-to-book of domestic firms during our sample period.

[22] Because few foreign firms file with the SEC, our tests cannot speak to changes in search by domestic firms for foreign rivals after the tariff shock. We have no prediction of whether such search for foreign rivals would increase or decrease: while the domestic firm's investment opportunities are lower following a tariff cut, foreign rivals' investment opportunities are greater, so the net effect on search is unclear.

[23] We end the sample in 2014 due to limitations on tariff import data.

incidence rate ratio corresponding to the coefficient on *Tariff cut*$_{i,t}$ is roughly 0.87, suggesting a 13% decline in the average acquisition of rivals' filings.[24]

Overall, the findings from these two shocks reinforce our interpretation that public information flows are highly sensitive to the firm's and its rivals' investment opportunities.

4. Information flows and subsequent investment decisions

In this section, we examine the predictive power of public information flows for investment decisions. Information acquisition facilitates active learning, so the data can be applied to study not only factors associated with information flows ex ante, but also the relation between these flows and ex post firm decisions. Thus we can provide evidence on *how* the acquired information is used, assuming the information's role in the searching firm's corporate strategies is subsequently revealed in its investment activities, such as by changes in product positioning.

Given the unique pairwise structure of the data, we focus our tests on pairwise external and internal investment outcomes. We separately examine external and internal investment decisions, because corporate information problems are different for these decisions. To examine external investments, we focus on the role of information flows in executing M&A. M&A are vital to implement the strategic direction of the firm but entail substantial information asymmetries between acquirers and targets (e.g., Hoberg and Phillips, 2010a; Betton et al., 2008). Consistent with the literature showing uncertainty inhibits investment (e.g., Guiso and Parigi, 1999), these asymmetries imply a strong incentive for firms to vet targets ahead of deals. Issues of adverse selection are less severe for internal investments; instead, a principal source of uncertainty stems from interactions with rivals. Internal investments in product and service development require simultaneously evaluating rivals' investments to ensure the firm's actions advance its positioning

[24] These results are robust to instead using a negative binomial model with industry and year fixed effects, which yields a difference-in-differences design similar to that used by Fresard (2010) and Balakrishnan and Cohen (2014).

with consumers. In this vein, to examine internal investments, we focus on the role of information flows in setting capital and R&D levels and selecting product investments.

4.1. Facilitating external investments: Mergers and acquisitions

Following the intuition that public information can help resolve uncertainties inherent in M&A, Rossi and Volpin (2004) show that the quality of disclosures is positively related to the percentage of a country's traded companies that are targeted for acquisitions. They interpret the evidence as suggesting that "good disclosure is a necessary condition for identifying potential targets" (p. 283). Our analyses extend this basic intuition in three ways. First, we directly measure information flows between the acquirer and target in event time surrounding M&A announcements. Because M&A are discrete investment events, these event-study analyses help identify *when* information flows help resolve M&A uncertainty and mitigate concerns that information acquisition relates only indirectly (or spuriously) to investment. Second, we examine the product similarity of the acquirer and target as a source of heterogeneity in the predictive power of information flows for acquisitions. Public information is plausibly more important for differentiating or diversifying acquisitions, as information asymmetries are likely higher the more dissimilar the acquirer is relative to the target. Third, we test a potential externality of public firm presence for private firms—the use of public rivals' disclosures ahead of acquisitions of private targets, whose information is publicly inaccessible ex ante.

We identify each firm's U.S. acquisition targets using Securities Data Company (SDC) Platinum, limiting our analyses to acquisitions announced between January 1, 2004 and December 31, 2015. We link acquirer and target CUSIP to acquirer and target CIK using the Wharton Research Data Services (WRDS) SEC Analytics CUSIP-CIK link table and identify acquirers' GVKEY using the WRDS SEC Analytics GVKEY-CIK link table. In all cases, we exclude share repurchases and minority stake acquisitions but retain deals even if they are eventually withdrawn or result in less than 100% ownership by the acquirer.

4.1.1. Information flows in event-time around M&A announcements

To begin, we examine abnormal information flows between firm-pairs in the months before and after M&A transactions between the firms. Fig. 1 visually depicts abnormal information acquisition by public acquirers for public targets in event time, plotting abnormal information acquisition by acquiring firm *i* for SEC filings of target firm *j* by month, relative to the month it is publicly announced that *i* will acquire *j* (month = 0). We measure abnormal information acquisition as the count of firm *i*'s monthly EDGAR downloads of the target firm's filings less the count of firm *i*'s monthly downloads of a propensity-score matched control firm that is not an acquisition target of firm *i*.[25] We construct this measure relative to nontarget firms to isolate the incremental effect of search due to the M&A itself, as firms plausibly search for and vet other potential targets ahead of an acquisition.

Fig. 1 shows that abnormal search builds in the months leading to acquisition announcements, particularly within the three months immediately prior. The most salient feature of the figure is the sharp peak in acquirer-target information acquisition in the month of the announcement, which corresponds to roughly five to six times more search activity than the average for the first six months in the figure (i.e., months -12 to -7). In addition, abnormal information acquisition remains elevated for several months after the acquisition announcement. Thus acquirer-target information acquisition appears to be important in multiple stages of the

[25] We calculate the propensity score by conducting acquiring firm *i*-year specific logistic regressions of potential targets *j*, controlling for size, market-to-book, leverage, return on assets (ROA), plant, property, and equipment (PP&E), logged cash holdings, firm age, a blockholder indicator, and industry indicators of each firm *j* (Eckbo, 2014). The control firm chosen is the non-acquired firm *j* in the year prior to the acquisition that has the propensity score closest to the acquired firm *j*.

M&A process—both for identifying potential targets as well as during due diligence, as more employees become involved to execute the acquisition.

4.1.2. Predicting public firm M&A with public information flows

Next, we empirically examine the predictive power of pairwise information flows for subsequent M&A decisions. That is, we address the question: can information flows between two firms predict the subsequent merger of those firms? Predicting the targets of M&A transactions is empirically challenging (Routledge et al., 2016; Betton et al., 2008), and on top of that, very little research predicts deals at the level of acquirer-target pairs. We illustrate the usefulness of our measure in this context by regressing $Acquisition_{i,j,t+1}$, an indicator variable equal to one if public firm i acquires public firm j in year t+1, on $Information\ acquisition_{i,j,t}$ and other firm and firm-pair characteristics that could predict an acquisition by i of j. Thus, in this analysis, we use search activity in calendar year t to predict acquisitions announced in calendar year t+1.

Panel A of Table 5 presents the prediction results of the logit model. The coefficient on $Information\ acquisition_{i,j,t}$ is positive and highly significant, suggesting that greater acquisition of rival firm j information by firm i in t is associated with an increase in the probability that i acquires j in year t+1. The coefficient is relatively stable across all three specifications and the effect is economically significant: the odds ratios indicate a one standard deviation increase in information acquisition relates to a 55%–61% increase in the odds i acquires j in year t+1.

Fig. 2 illustrates this predictive power based on a Receiver Operating Characteristics (ROC) curve analysis, which evaluates the performance of the prediction model based on its ability to accurately classify acquisitions as occurring or not occurring based on a given set of regressors. When $Information\ acquisition_{i,j,t}$ is the sole explanatory variable in the prediction model, the area under the curve is about 0.73 (untabulated), which is conventionally considered acceptable discrimination (Hosmer and Lemeshow, 2000). When the full model is used, the area under the curve increases to more than 0.89, which represents excellent discrimination. When we compare the discriminatory power of all the regressors, we find that *Information acquisition*$_{i,j,t}$ provides more discriminatory power to predict firm-pair acquisitions than any other variable (untabulated). In sum, the results show that information flows are highly predictive of M&A activity, with discriminatory power that is economically meaningful.

We also examine the predictive ability of information flows based on variation in the similarity of the acquirer and target—that is, based on the level of information asymmetries between the firms. In Panel B of Table 5, we expand the regression by interacting information acquisition with firm-pair product similarity. The results show that information flows are a weaker predictor of a subsequent acquisition when the product similarity of the pair is higher. We interpret this as evidence that public information is especially important to alleviate information asymmetries ahead of acquisitions of more dissimilar targets. That is, the more dissimilar the target—and thus the more uncertainty acquirers likely have regarding deal structure, integration risks, etc.—the more useful firm information appears to be for facilitating the purchase.

4.1.3. Predicting private firm M&A with public information flows

Finally, we examine a potential externality of public firms' disclosures, which can act as an alternative source of information about private targets. Unlike for public targets, information for U.S. private firms is publicly unavailable. As a result, potential acquirers without access to private accounts must use alternative information sources to identify and vet potential targets. One possibility is that acquirers use the disclosures of other public firms that are close rivals of potential private targets, perhaps to evaluate industry sales trends or recent innovations in the product space. If so, the implication would be that information acquisition of public rival *j*'s disclosures in year *t* would predict public firm *i*'s acquisition of a private firm in *j*'s product space in year *t*+1, consistent with public firm presence creating an informational externality that extends to private firms.

To test this prediction, we regress *Private acquisition*$_{i,j,t+1}$, an indicator variable equal to one if public firm *i* acquires a private firm in the same two-digit SIC code as public firm *j* in year *t*+1, on *Information acquisition*$_{i,j,t}$ and other predictors. We rely on SIC codes for this analysis because text-based similarity measures are unavailable for private firms. We present the results in Table 6. In Panel A, we find that the acquisition of *j*'s information has strong predictive power for *i*'s acquisition of a private target in *j*'s product space, and Panel B shows that this predictive power is substantially stronger when the acquirer and the target do not share the same four-digit SIC—that is, when information asymmetries are higher. We interpret this finding as evidence of an externality of public firm information, which can act as a partial substitute for information that is typically unavailable for private companies.

Overall, the findings illustrate firms' use of public firm information to facilitate external investments, suggesting that requiring transparency via public disclosure requirements could aid in M&A. The cross-sectional tests also illustrate an important interaction with product market characteristics and the target's information environment—information flows better predict acquisitions when information asymmetries are high, such as when an acquisition is differentiating or when little public information exists for the target.

4.2. Facilitating internal investments: Capex levels, R&D levels, and product differentiation

Firms advance their competitive position via continuous investment.[26] However, both the optimal level of investment and the optimal selection of projects are unknown to the manager. The firm's private information set is certainly critical to resolving these uncertainties, but peer information is also plausibly important. For example, Kroger might acquire Walmart's disclosures to gauge its investments in distribution or IT systems, or Delta might acquire Alaska Airlines' disclosures to better gauge capacity growth plans on the West coast. Consistent with this intuition, Durnev and Mangen (2009) and Badertscher et al. (2013) find that peers' disclosures provide information useful for the firm's capital investment decisions.

Yet theory and empirical work leave it unclear how firms use rivals' information for simultaneously setting investment levels and selecting projects. A relatively straightforward view is that information signals from rivals' disclosures converge prior beliefs about potential investments (e.g., Lieberman and Asaba, 2006; Devenow and Welch, 1996), which can lead the firm to invest more similarly to rivals. This view has some support from Bustamante and Fresard (2018), who find that firms increase investment in response to increases in the investment of product market peers. They interpret their evidence consistent with a learning perspective, whereby public information about peers creates endogenous complementarity in investment decisions. Models of endogenous product differentiation yield similar predictions. For example, in Hellwig and Veldkamp (2009), agents' information acquisition incentives inherit the same strategic motives as the agents' actions, so in a game with strategic substitutability, agents prefer

[26] For example, Shaked and Sutton (1987) explain that industries become more concentrated "because the possibility exists, primarily through incurring additional fixed costs, of shifting the technological frontier constantly forward towards more sophisticated products." As Ellickson (2007) notes of the supermarket industry, "[f]irms that fail to match the quality increases of their rivals cannot survive."

to acquire information others do not have. The empirical implication is that firms acquire less *public* information about the rivals from which they are trying to differentiate.

Other theory and empirical work suggest alternatives. For example, a number of models predict growth-deterring or pre-emptive investment (e.g., Gilbert and Lieberman, 1987; Fudenberg and Tirole, 1983). If firms use rivals' information largely to learn about these capacity commitments, greater information acquisition could predict less similar investment levels. Similarly, greater information acquisition could predict less similar product choices. Product differentiation strategies are risky, in part because competitors move simultaneously; rivals' position in the product space is not static. Thus a strong incentive to acquire information about a rival is to ensure product investments create the desired separation.

We test these competing incentives with a design akin to our prior analyses. Retaining the pairwise structure of the data, we examine the link between pairwise information flows and changes in pairwise product similarity and pairwise investment similarity. Using measures of pairwise similarity allows us to speak to firm *i*'s investment decisions, relative to the rival *j*—for example, how *i* changes its products, relative to *j*'s products—as a function of *i*'s acquisition of *j*'s information. For investment similarity, we focus on changes in the pairwise similarity of capital and R&D expenditures. We measure the change in similarity of capital expenditures (R&D expenditures) as negative one multiplied by the change in the absolute difference in firm *i*'s and its rival firm *j*'s annual capital expenditures scaled by lagged fixed assets (R&D expenses scaled by total sales) between year *t* and *t*+1.[27] We measure encroachment, the change between year *t* and *t*+1 in product similarity, using the Hoberg and Phillips (2010a) similarity score, as

[27] The measurement of capital expenditures scaled by fixed assets and R&D expenses scaled by revenues is consistent with variable measurement in prior research, including Hennessy et al. (2007) and Frank and Shen (2016), among others. When R&D expense is missing in Compustat, we set it equal to the industry-year mean, where industry is measured at the two-digit SIC code level, following Koh and Reeb (2015).

above. We winsorize the dependent variables at the first and 99th percentiles to reduce the influence of outliers and again include firm-pair fixed effects, so variation in information acquisition is relative to other years for the same firm *j*, not to other firms in the cross-section.

Table 7 reports the results for changes in pairwise investment similarity, and Table 8 presents the results for changes in pairwise product similarities. In both tables, we include specifications interacting information flows with firm-pair product similarities, following the intuition above that firms' use of rivals' information plausibly differs depending on pairwise product positioning. In Table 7 Panel A, we find a positive association between pairwise information flows and the ex post change in investment similarities, consistent with acquired information being used to help set investment levels. The results in Table 7 Panel B show that information flows predict future R&D mimicking more strongly when the product similarity of the firms is lower—again, when information asymmetries are high.[28]

In Table 8 Panel A, we find that greater pairwise information flows predict *lower* levels of subsequent product similarities between the rival-pair, evidence consistent with the use of public information to facilitate product differentiation. Table 8 Panel B suggests that this ex post differentiation, relative to *j*, is stronger the more similar *i* is to *j* ex ante.[29] Together, these results suggest that information flows help facilitate forms of both mimicking and differentiation; on average, the acquisition of a rival's public information leads to more similar investment levels but more dissimilar product choices, consistent with firms seeking information to avoid under- or over-investment but also to create separation from rivals' product offerings.

---

[28] Our measure of R&D similarity is subject to a high degree of kurtosis, even after winsorization. Still, we use this measure in the tabulated results to avoid additional data adjustments, as even extreme changes in R&D could reflect learning. However, in untabulated robustness tests, we find that the results in Table 7 Panel A for R&D similarity are somewhat sensitive to the presence of the extreme values in the measure of R&D similarity; the results in Table 7 Panel B (i.e., the interaction term) remain consistent, regardless of additional steps to remove extreme values.

[29] An equivalent interpretation is that acquired information appears to be used relatively more for product mimicking when the firms are ex ante dissimilar.

The findings suggest an important nuance to the common interpretation of peer effects. Consistent with prior work, our results support interfirm learning as a mechanism for more similar investment levels among rivals (e.g., Bustamante and Fresard, 2018). However, this effect does not imply more homogeneous choices in all investment decisions. The firm can benchmark investment levels relative to rivals but still seek to differentiate on product or service quality, variety, or other dimensions best suited to its strengths and competitive positioning. This approach reduces the risk the firm falls behind on process and product improvements but also ensures brand distinction, which can improve customer loyalty, pricing power, etc. (e.g., Aaker, 2011). The broader implication is that peer effects are not easily reduced to a single dimension, and pairwise similarity in the product market appears to be an important conditioning variable for how firms use rivals' public information.

5. Conclusion

Interfirm information flows play an important role in investment theory but are rarely examined empirically because they are generally unobservable. We develop a direct measure of these information flows at the *firm-pair* level. This measure allows us to construct an asymmetric directed network of information acquisition that sidesteps many of the problems inherent in other measures of corporate learning, such as those based on information supply endogenous to investors' demand (e.g., Veldkamp, 2006). The results suggest information flows among public rivals are closely related to investment decisions. Investment opportunities drive information flows, and firms appear to use acquired information in multiple contexts, including in M&A, setting capital investment and R&D levels, and selecting differentiating product investments. Our results illustrate the important role of public information acquisition in shaping firm-pair interactions, contributing to the empirical work on corporate investment and peer effects.

Our novel data should be useful in future research. One promising avenue for future work would be to further explore the underlying reasons certain characteristics of the product space, such as product market fluidity, are systematically associated with differences in information acquisition among rivals. Better understanding the mechanisms for these relations is important in part to illuminate costs to disclosing firms of mandated filings, as the acquisition of competitor information sometimes benefits both the firm obtaining the information and the one disclosing it (e.g., Smith, 1981). Future work could also examine the specific content of regulatory filings to provide further evidence on the type of information for which competitors search and explore how this information acquisition relates to subsequent investment behavior.

Because our data have the dimensionality to measure active learning within corporate rivalries, they also allow for more direct tests of empirical predictions in a number of other topics. Examples include the literature on information cascades (e.g., Anderson and Holt, 1997), monitoring (e.g., Shroff et al., 2014), and antitrust topics, such as collusion (e.g., Porter, 2005; Green and Porter, 1984). The data could also be extended to capture the acquisition of public firm filings by other important audiences, such as private firms, auditors, nonprofits, regulators, etc. (e.g., Bozanic et al., 2017; Drake et al., 2019). Such research could help to reveal the roles of public information flows in resolving uncertainties and aiding in corporate (and noncorporate) decision-making.

Future work also has the opportunity to make improvements to the data. One limitation of our study is that sample coverage is limited. Future work could progressively relax the selection criteria we impose to build out the sample of searching firm IP addresses. Relatedly, there could be more refined methods that yield a more comprehensive and more accurate sample based on predictive self-search. There could also be alternative methods that help address the possibilities that firms could mask their IP addresses or that some proportion of search behavior is due to extraneous reasons, such as employees' day trading. Further, future work could identify methods to better capture the searches of smaller firms, which would help to guarantee the generalizability of findings.

Appendix A

Variable definitions

This table presents details on the definition and computations of all variables in the paper. *i* indexes searching firms, *j* indexes searched-for firms, and *t* indexes years. Compustat data codes are in bold.

Dependent variables	
$Information\ acquisition_{i,j,t}$	The total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* in year *t*.
$Encroachment_{i,j,t\ to\ t+k}$	$Product\ similarity_{i,j,t+k}$-$Product\ similarity_{i,j,t}$.
$\Delta Capex\ similarity_{i,j,t+1}$	-1*($\lvert Capex_{i,t+1}-Capex_{j,t+1}\rvert$ - $\lvert Capex_{i,t}-Capex_{j,t}\rvert$), where $Capex_{i,t}$ equals firm *i*'s net capital expenditures in year *t* scaled by lagged total fixed assets, $(\boldsymbol{capxv}_{i,t}-\boldsymbol{sppe}_{i,t})/\boldsymbol{ppent}_{i,t-1}$.
$\Delta R\&D\ similarity_{i,j,t+1}$	-1*($\lvert R\&D_{i,t+1}-R\&D_{j,t+1}\rvert$ - $\lvert R\&D_{i,t}-R\&D_{j,t}\rvert$), where $R\&D_{i,t}$ equals firm *i*'s research and development expenses in year *t* scaled by sales, $\boldsymbol{xrd}_{i,t}/\boldsymbol{sale}_{i,t}$. Missing values of $\boldsymbol{xrd}_{i,t}$ are set to the industry-year mean, where industry is measured at the two-digit SIC code level, following Koh and Reeb (2015).
$Public\ acquisition_{i,j,t+1}$	An indicator variable equal to one if public firm *i* acquires an ownership stake of more than 50% in public firm *j* in year *t*+1 (SDC).
$Private\ acquisition_{i,j,t+1}$	An indicator variable equal to one if public firm *i* acquires an ownership stake of more than 50% in a non-public firm in the same two-digit SIC as public firm *j* in year *t*+1 (SDC).
Firm-pair regressors	
$Product\ similarity_{i,j,t}$	The cosine similarity between firm *i*'s and firm *j*'s product word vectors during year *t* (Hoberg and Phillips, 2010a).
$Return\ corr_{i,j,t}$	The correlation of daily stock returns for firm *i* and firm *j* during year *t* (The Center for Research in Security Prices (CRSP)).
$Same\ auditor_{i,j,t}$	An indicator variable equal to one if firm *i* and firm *j* share the same auditor during year *t* (Audit Analytics).
$Log(Distance)_{i,j,t}$	The natural log of one plus the distance in kilometers between the headquarters of firm *i* and firm *j*.
Firm-level regressors	
$Product\ market\ fluidity_{i,t}$	The cosine similarity between firm *i*'s product word vector and the aggregate change vector of rivals' product words (Hoberg et al., 2014).
$TNIC\ HHI_{i,t}$	The Herfindahl-Hirschman sum of squared market shares based on the Hoberg Phillips TNIC3 industry classification (Hoberg et al., 2014).
$Size_{i,t}$	The natural log of firm *i*'s total assets, log($\boldsymbol{at}_{i,t}$).
$Market\text{-}to\text{-}book\ ratio_{i,t}$	Market-to-book assets ratio of firm *i*, $(\boldsymbol{at}_{i,t}+\boldsymbol{prcc\_f}_{i,t}*\boldsymbol{csho}_{i,t}-\boldsymbol{ceq}_{i,t}-\boldsymbol{txdb}_{i,t})/\boldsymbol{at}_{i,t}$.
$Leverage_{i,t}$	Book leverage of firm *i*, $(\boldsymbol{dlc}_{i,t}+\boldsymbol{dltt}_{i,t})/\boldsymbol{at}_{i,t}$.
$ROA_{i,t}$	Return-on-assets of firm *i*, $\boldsymbol{ib}_{i,t}/\boldsymbol{at}_{i,t}$.
$Sales\ growth_{i,t}$	Sales growth of firm *i*, $(\boldsymbol{sale}_{i,t}-\boldsymbol{sale}_{i,t-1})/\boldsymbol{sale}_{i,t-1}$.
$Firm\ age_{i,t}$	The number of years firm *i* has been included in the Compustat database.
$Total\ product\ similarity_{i,t}$	The sum of pairwise similarities between firm *i* and its competitors within a given year (Hoberg and Phillips, 2016).
$Log(Cash_{i,t})$	The natural log of firm *i*'s cash and investments scaled by total assets, log($\boldsymbol{ch}_{i,t}/\boldsymbol{at}_{i,t}$).
$PPE_{i,t}$	Firm *i*'s net plant, property, and equipment, scaled by total assets $\boldsymbol{ppent}_{i,t}/\boldsymbol{at}_{i,t}$.
$Log(Filings)_{j,t}$	The natural log of one plus the total number of filings posted on EDGAR by firm *j* in year *t*.
$Leader_{j,t}$	An indicator variable equal to one if firm *j* has the largest volume of sales for firm *i*'s TNIC3 industry (Hoberg and Phillips, 2016) in year *t*.

$Distress_{j,t}$	An indicator variable equal to one if firm *j*'s Altman's *Z* score, $(3.3*\boldsymbol{pi}_{j,t}+\boldsymbol{sale}_{j,t}+1.4*\boldsymbol{re}_{j,t}+1.2*(\boldsymbol{act}_{j,t}-\boldsymbol{lct}_{j,t}))/\boldsymbol{at}_{j,t}$, is in the bottom 10th percentile of the Compustat universe in year *t*.
$Blockholder_{j,t}$	An indicator variable equal to one if an individual shareholder owns more than 5% of firm *j*'s shares outstanding in year *t*.
$Industry\ M\&A_{j,t\text{-}1}$	An indicator variable equal to one if there were mergers and acquisitions of publicly listed firms within firm *j*'s Fama-French 48 industry classification in year *t*-1.
$Tariff\ cut_{i,t}$	An indicator variable equal to one if the negative change in the ad valorem tariff rates for firm *i*'s industry in year *t* is at least three times the median change and not followed by an equivalent increase in the following two years, as in Fresard (2010).
BCA_t	An indicator variable equal to one for the years 2011–2013.
$High\ contract_i$	An indicator variable equal to one if U.S. government contracts accounted for more than 5% of firm *i*'s revenue in 2010 (Federal Procurement Data System).

Appendix B

Sample composition

The table presents details on the sample composition, over time and across industries. Across all panels, the "verified sample" consists of 3,931 sample firm-years (579 unique firms) for which we have identified search activity on EDGAR for the sample period (2004–2015). "Compustat" consists of all firms in the Hoberg and Phillips universe that have total assets greater than $1 million and financial data available on Compustat (52,527 firm-years). Panel A presents the number of unique searching firms by year and in total. Panel B provides the average proportion of firms over the sample period (2004–2015) that is in the Fama-French 30 industry classification for both the verified sample and the Compustat sample. Panel C provides descriptive statistics for both the verified sample and the Compustat sample. *Total assets* is defined as ***at***$_{i,t}$, *Market value of equity* is defined as ***prcc_f***$_{i,t}$****csho***$_{i,t}$, and *Sales turnover* is defined as ***sale***$_{i,t}$***/at***$_{i,t}$ (Compustat variable names are bold and in italics). All variables are winsorized at the first and 99th percentiles. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

Panel A: Coverage of unique searching firms

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
Verified sample	244	273	318	372	428	450	440	386	314	277	232	197	579
Of S&P 500	107	110	115	129	145	153	150	135	111	97	87	77	196

Panel B: Industry composition relative to Compustat

	Proportion		
Industry (Fama-French 30)	Verified sample	Compustat	*z*-stat for difference
Food Products	0.030	0.018	*5.33****
Beer & Liquor	0.007	0.002	*7.36****
Tobacco Products	0.008	0.001	*8.93****
Recreation	0.021	0.017	*2.03***
Printing & Publishing	0.007	0.007	*0.23*
Consumer Goods	0.016	0.009	*4.33****
Apparel	0.013	0.010	*1.41*
Healthcare, Medical Equipment, Pharma Products	0.088	0.127	*7.17****
Chemicals	0.019	0.017	*0.87*
Textiles	0.000	0.002	*-2.29***
Construction and Construction Materials	0.012	0.019	*-3.24****
Steel Works, etc.	0.010	0.009	*0.63*
Fabricated Products and Machinery	0.033	0.026	*2.67****
Electrical Equipment	0.008	0.014	*-3.13****
Automobiles and Trucks	0.016	0.012	*2.05***
Aircraft, Ships, and Railroad Equipment	0.009	0.006	*2.07***
Precious Metals, Non-Metallic, and Metal Mining	0.005	0.007	*-1.17*
Coal	0.000	0.002	*-3.66****
Petroleum and Natural Gas	0.028	0.043	*-4.42****
Utilities	0.053	0.022	*12.00****
Communication	0.033	0.026	*2.70****
Personal and Business Services	0.112	0.113	*-0.24*
Business Equipment	0.118	0.100	*3.46****
Business Supplies and Shipping Containers	0.019	0.010	*5.70****
Transportation	0.047	0.023	*9.47****

Wholesale	0.049	0.026	*8.75****
Retail	0.066	0.043	*6.87****
Restaurants, Hotels, and Motels	0.021	0.015	*2.84****
Banking, Insurance, Real Estate, Trading	0.130	0.245	*-16.36****
Other	0.024	0.031	*-2.27***

Panel C: Summary statistics relative to Compustat

		Verified sample	Compustat	*t*-stat for difference
Total assets ($m)	Mean	12,464	4,287	*17.12****
	Median	2,331	622	
	Std	29,693	13,545	
Market value of equity ($m)	Mean	11,066	2,816	*18.65****
	Median	2,155	430	
	Std	27,635	8,124	
Market-to-book	Mean	1.82	1.92	*-5.17****
	Median	1.45	1.36	
	Std	1.14	1.56	
Leverage	Mean	0.24	0.22	*7.56****
	Median	0.21	0.15	
	Std	0.21	0.23	
Return-on-assets	Mean	0.02	-0.04	*29.14****
	Median	0.04	0.01	
	Std	0.12	0.24	
Sales turnover	Mean	1.02	0.80	*16.56****
	Median	0.82	0.60	
	Std	0.80	0.79	

References

Aaker, D.A., 2011. Brand Relevance: Making Competitors Irrelevant. Jossey-Bass, San Francisco.

Anderson, L.R., Holt, C.A., 1997. Information cascades in the laboratory. American Economic Review 87, 847–862.

Badertscher, B., Shroff, N., White, H.D., 2013. Externalities of public firm presence: Evidence from private firms' investment decisions. Journal of Financial Economics 109, 682–706.

Balakrishnan, K., Cohen, D.A., 2014. Competition and financial accounting misreporting. Unpublished working paper, London Business School and University of Texas at Dallas.

Bernard, A.B., Jensen, B.J., Schott, P.K., 2006. Trade costs, firms, and productivity. Journal of Monetary Economics 53, 917–937.

Betton, S., Eckbo, B.E., Thorburn, K.S., 2008. Corporate takeovers. In: Eckbo, B.E. (Ed.), Handbook of Empirical Corporate Finance (vol. 2). Elsevier, North Holland, 291–430.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1998. Learning from the behavior of others: conformity, fads, and informational cascades. Journal of Economic Perspectives 12, 151–170.

de Bodt, E.J., Eckbo, B.E., Roll, R.W., 2018. Corporate rivalry and return comovement. Unpublished working paper, University of Lille, Dartmouth College, and California Institute of Technology.

Bond, P., Edmans, A., Goldstein, I., 2012. The real effects of financial markets. Annual Review of Financial Economics 4, 339–360.

Bozanic, Z., Hoopes, J.L., Thornock, J.R., Williams, B.M., 2017. IRS attention. Journal of Accounting Research 55, 79–114.

Brogaard, J., Denes, M., Duchin, R., 2019. Political influence and the renegotiation of government contracts. Unpublished working Paper, University of Utah, Carnegie Mellon University, and University of Washington.

Bustamante, M.C., Fresard, L., 2018. Does firm investment respond to peers' investment? Unpublished working paper, University of Maryland and University of Lugano, Swiss Finance Institute.

Cameron, C., Trivedi, P., 2013. Regression Analysis of Count Data, second ed. Cambridge University Press, New York.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. Journal of Finance 66, 1461–1499.

Delong, G., Deyoung, R., 2007. Learning by observing: Information spillovers in the execution and valuation of commercial bank M&As. Journal of Finance 62, 181–216.

Dessaint, O., Foucault, T., Fresard, L., Matray, A., 2019. Noisy stock prices and corporate investment. Review of Financial Studies 32, 2625-2672.

Devenow, A., Welch, I., 1996. Rational herding in financial economics. European Economic Review 40, 603–615.

Drake, M., Roulstone, D., Thornock, J., 2015. The determinants and consequences of information acquisition via EDGAR. Contemporary Accounting Research 32, 1128–1161.

Drake, M., Roulstone, D., Thornock, J., 2016. The usefulness of historical accounting reports. Journal of Accounting and Economics 61, 448–464.

Drake, M., Lamoreaux, P., Quinn, P., Thornock, J., 2019. Auditor benchmarking of client disclosures. Review of Accounting Studies 24, 393–425.

Dunn, K.A., Mayhew, B.W., 2004. Audit firm industry specialization and client disclosure quality. Review of Accounting Studies 9, 35–58.

Durnev, A., Mangen, C., 2009. Corporate investments: Learning from restatements. Journal of Accounting Research 47, 679–720.

Dyer, T., Lang, M., Stice-Lawrence, L., 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. Journal of Accounting and Economics 64, 221–245.

Eckbo, B.E., 2014. Corporate takeovers and economic efficiency. Annual Review of Financial Economics 6, 51–74.

Ellickson, P.B., 2007. Does Sutton apply to supermarkets? RAND Journal of Economics 38, 43–59.

Erickson, T., Whited, T.M., 2012. Treating measurement error in Tobin's q. Review of Financial Studies 25, 1286–1329.

Feenstra, R.C., 1996. U.S. imports, 1972–1994, data and concordances. NBER Working Paper No. 5515.

Feenstra, R.C., Romalis, J., Schott, P.K., 2002. U.S. imports, exports, and tariff data, 1989 to 2001. NBER Working Paper No. 9387.

Foucault, T., Fresard, L., 2014. Learning from peers' stock prices and corporate investment. Journal of Financial Economics 111, 554–577.

Frank, M.Z., Goyal, V.K., 2003. Testing the pecking order theory of capital structure. Journal of Financial Economics 67, 217–248.

Frank, M. Z., Shen, T., 2016. Investment and the weighted average cost of capital. Journal of Financial Economics 119, 300–315.
Fresard, L., 2010. Financial strength and product market behavior: The real effects of corporate cash holdings. Journal of Finance 65, 1097–1122.
Fudenberg, D., Tirole, J., 1983. Capital as a commitment: Strategic investment to deter mobility. Journal of Economic Theory 31, 227–250.
Gilbert, R.J., Lieberman, M., 1987. Investment and coordination in oligopolistic industries. RAND Journal of Economics 18, 17–33.
Green, E.J., Porter, R.H., 1984. Noncooperative collusion under imperfect price information. Econometrica 52, 87–100.
Guiso, L., Parigi, G., 1999. Investment and demand uncertainty. Quarterly Journal of Economics 114, 185–227.
Hanley, K., Hoberg, G., 2010. The information content of IPO prospectuses. Review of Financial Studies 23, 2821–2864.
Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52, 909–938.
Hellwig, C., Veldkamp, L., 2009. Knowing what others know: Coordination motives in information acquisition. Review of Economic Studies 76, 223–251.
Hennessy, C., Levy, A., Whited, T., 2007. Testing Q theory with financing frictions. Journal of Financial Economics 83, 691–717.
Hoberg, G., Phillips, G., 2010a. Product market synergies and competition in mergers and acquisitions: A text-based analysis. Review of Financial Studies 23, 3773–3811.
Hoberg, G., Phillips, G., 2010b. Real and financial industry booms and busts. Journal of Finance 65, 45–86.
Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. Journal of Political Economy 124, 1423–1465.
Hoberg, G., Phillips, G., Prabhala, N., 2014. Product market threats, payouts, and financial flexibility. Journal of Finance 69, 293–324.
Hodge, F., Pronk, M., 2006. The impact of expertise and investment familiarity on investors' use of online financial report information. Journal of Accounting, Auditing & Finance 21, 267–292.
Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, second ed. John Wiley & Sons, Inc., New York.
Huang, Y., Jennings, R., Yu, Y., 2017. Product market competition and managerial disclosure of earnings forecasts: Evidence from import tariff rate reductions. The Accounting Review 92, 185–207.
Jung, K., Kim, Y., Stulz, R.M., 1996. Timing, investment opportunities, managerial discretion, and the security issue decision. Journal of Financial Economics 42, 159–185.
Koh, P., Reeb, D., 2015. Missing R&D. Journal of Accounting and Economics 60, 73–94.
Kravet, T.D., Muslu, V., 2013. Textual risk disclosures and investors' risk perceptions. Review of Accounting Studies 18, 1088–1122.
Leary, M.T., Roberts, M.R., 2014. Do peer firms affect corporate financial policy? Journal of Finance 69, 139–178.
Lee, C., Ma, P., Wang, C., 2015. Search-based peer firms: Aggregating investor perceptions through Internet co-searches. Journal of Financial Economics 116, 410–431.
Li, E.X., Ramesh, K., 2009. Market reaction surrounding the filing of periodic SEC reports. The Accounting Review 84, 1171–1208.
Li, F., Lundholm, R., Minnis, M., 2013. A measure of competition based on 10-K filings. Journal of Accounting Research 51, 399–436.
Lieberman, M.B., Asaba, S., 2006. Why do firms imitate each other? The Academy of Management Review 31, 366–385.
Ljungqvist, A., Qian, W., 2016. How constraining are limits to arbitrage? Review of Financial Studies 29, 1975–2028.
Loughran, T., McDonald, B., 2017. The use of EDGAR filings by investors. Journal of Behavioral Finance 18, 231–248.
Moscarini, G., 2004. Limited information capacity as a source of inertia. Journal of Economic Dynamics & Control 28, 2003–2035.
Nini, G., Smith, D.C., Sufi, A., 2009. Creditor control rights and firm investment policy. Journal of Financial Economics 92, 400–420.
Porter, R., 2005. Detecting collusion. Review of Industrial Organization 26, 147–167.
Rossi, S., Volpin, P.F., 2004. Cross-country determinants of mergers and acquisitions. Journal of Financial Economics 74, 277–304.

Routledge, B.R., Sacchetto, S., Smith, N.A., 2016. Predicting merger targets and acquirers from text. Unpublished working paper, Carnegie Mellon University, University of Navarra, and University of Washington.
Roychowdhury, S., Shroff, N., Verdi, R.S., 2018. The effects of financial reporting and disclosure on corporate investment: A review. Unpublished working paper, Boston College and Massachusetts Institute of Technology.
Shaked, A., Sutton, J., 1987. Product differentiation and industrial structure. Journal of Industrial Economics 36, 131–146.
Shroff, N., Verdi, R.S., Yu, G., 2014. Information environment and the investment decisions of multinational corporations. The Accounting Review 89, 759–790.
Smith, R.L., 1981. Efficiency gains from strategic investment. Journal of Industrial Economics 30, 1–23.
Stice, J.D., 1991. Using financial and market information to identify pre-engagement factors associated with lawsuits against auditors. The Accounting Review 66, 516–533.
Tybout, J., 2003. Plant- and Firm-Level Evidence on “New” Trade Theories. In: E.K. Choi and J. Harrigan (Eds.), Handbook of International Economics (vol. 1). Oxford, Basil-Blackwell, 388–415.
Veldkamp, L., 2011. Information Choice in Macroeconomics and Finance. Princeton University Press, Princeton.
Veldkamp, L., 2006. Media frenzies in markets for financial information. American Economic Review 96, 577–601.

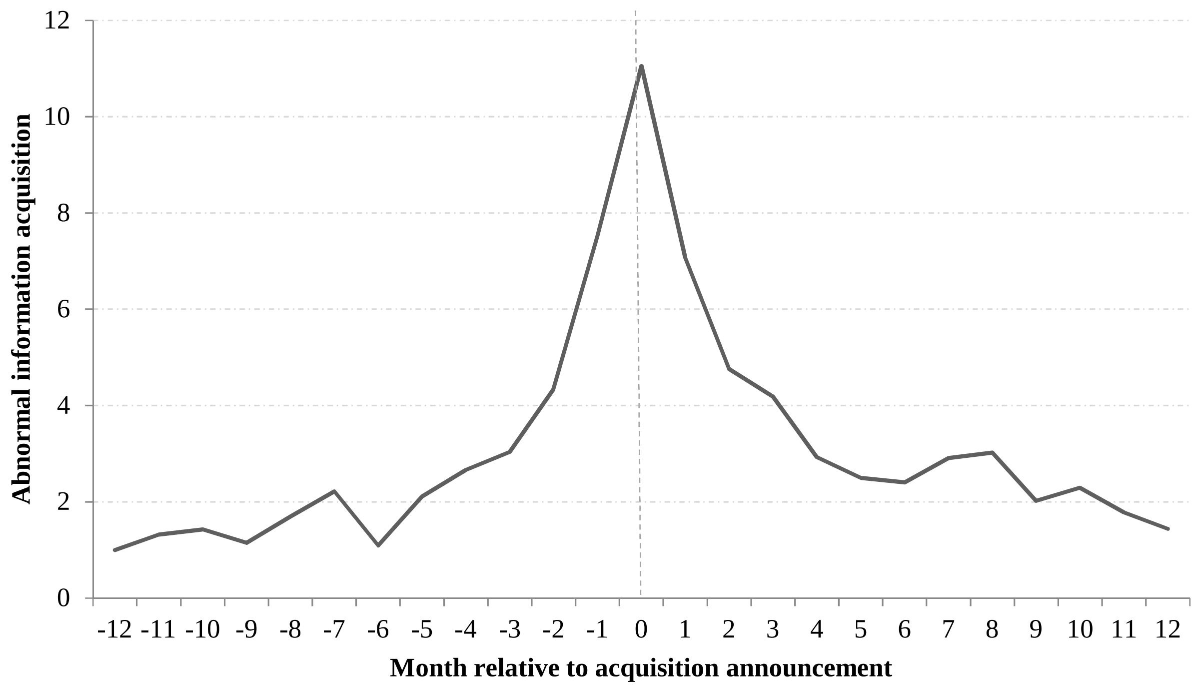

Fig. 1. Abnormal information acquisition around M&A announcements. This figure plots average abnormal information acquisition by public firm *i* for public target firm *j*'s EDGAR filings in monthly event time relative to the announcement of firm *i*'s acquisition of firm *j*. We measure abnormal information acquisition as the number of acquiring firm *i*'s monthly downloads of the target firm *j*'s EDGAR filings less the count of firm *i*'s monthly downloads of a propensity-score matched, non-target control firm, *j*. We calculate the propensity score by conducting acquiring firm *i*-year specific logistic regressions of potential targets *j*, controlling for size, market-to-book, leverage, ROA, PP&E, logged cash holdings, firm age, a blockholder indicator, and industry indicators of each firm *j* (Eckbo, 2014). The control firm chosen is the non-acquired firm *j* in the year prior to the acquisition that has the propensity score closest to the acquired firm *j*. Data on M&A transactions are retrieved from SDC Platinum and purged of share repurchases, acquisitions of non-public targets, and minority stake acquisitions. See Appendix A for variable definitions and Appendix B for sample characteristics.

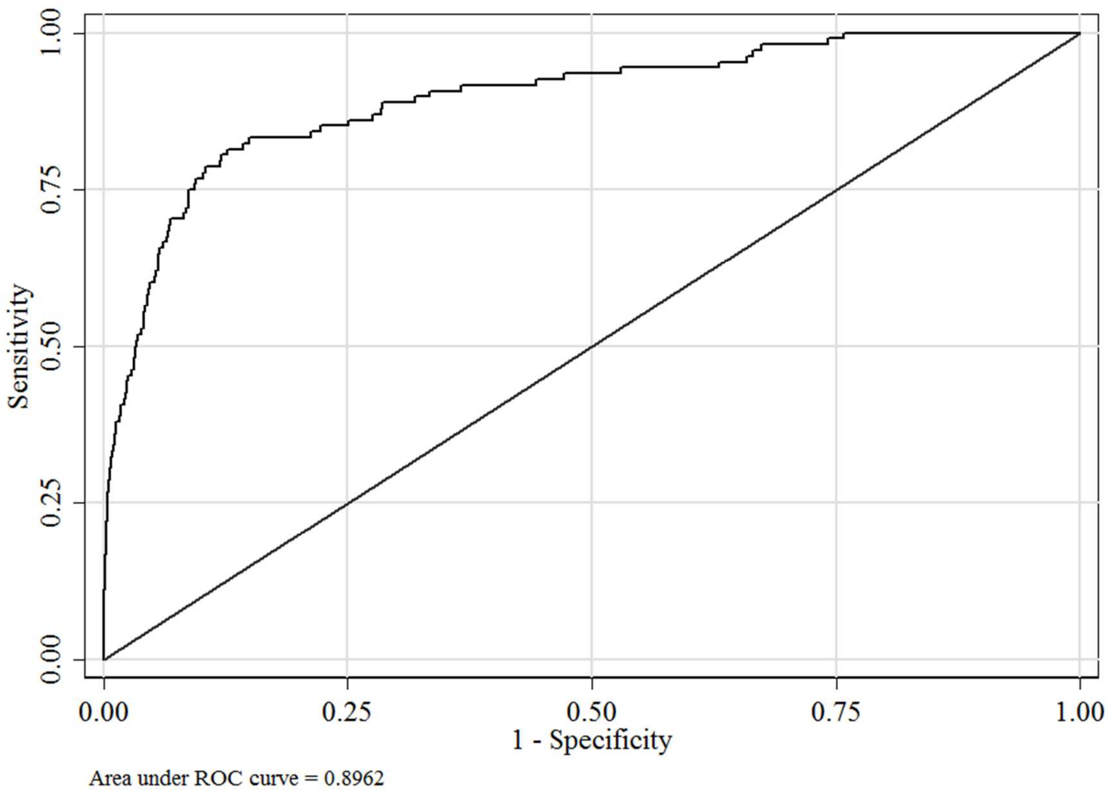

Fig. 2. Predicting future corporate acquisitions using pairwise information flows. This figure plots the Receiver Operating Characteristic (ROC) Curve for predicting the future acquisition of target firm *j* by acquiring firm *i* based on information flows between the *i*,*j* firm-pair. The ROC curve derives from the logit model presented in Table 5 Panel A Column 3.

Table 1
Sample descriptive statistics

This table provides descriptive statistics and correlations for each variable in the verified sample. The sample covers the period 2004 to 2015 and includes 252,370 *i,j,t* firm-pair-years. Dependent variables and firm-pair regressors are at the *i,j,t* level, where *i* indexes searching firms, *j* indexes searched firms, and *t* indexes years. Searching-firm regressors are at the *i,t* level; searched-firm regressors are at the *j,t* level. Panel A presents summary statistics for all variables in the verified sample. Panel B presents correlations among selected variables; see the Online Appendix for the full correlation table. Spearman's rank correlations are reported below the diagonal and Pearson correlations above the diagonal. Data are restricted to each searching firm's TNIC3 industry, based on the Hoberg and Phillips (2016) text-based network industry classification schema. All continuous variables are winsorized at the first and 99th percentiles. See Appendix A for variable definitions.

Panel A: Summary statistics

	mean	sd	p10	p25	p50	p75	p90
Dependent variables							
Information acquisition$_{i,j,t}$	1.083	4.626	0	0	0	0	1
Encroachment$_{i,j,t \text{ to } t+1}$	-0.009	0.031	-0.039	-0.017	-0.005	0.006	0.020
ΔCapex similarity$_{i,j,t+1}$	0.016	0.374	-0.234	-0.060	0.000	0.081	0.274
ΔR&D similarity$_{i,j,t+1}$	-0.166	5.416	-0.214	-0.007	0.000	0.005	0.181
Acquisition$_{i,j,t+1}$	0.0004	0.021	0	0	0	0	0
Private acquisition$_{i,j,t+1}$	0.140	0.347	0	0	0	0	1
Firm-pair regressors							
Product similarity$_{i,j,t}$	0.049	0.047	0.004	0.013	0.033	0.073	0.120
Return corr$_{i,j,t}$	0.293	0.220	0.0230	0.109	0.267	0.454	0.612
Same auditor$_{i,j,t}$	0.207	0.405	0	0	0	0	1
Log(Distance)$_{i,j,t}$	6.430	1.541	4.478	5.971	6.815	7.560	7.848
Searching-firm regressors							
Product market fluidity$_{i,t}$	8.746	3.792	4.242	5.921	8.104	11.18	14.19
TNIC HHI$_{i,t}$	0.114	0.081	0.041	0.067	0.092	0.133	0.207
Size$_{i,t}$	8.123	2.090	5.349	6.821	7.954	9.610	10.92
Market-to-book ratio$_{i,t}$	1.830	1.367	0.969	1.039	1.320	2.054	3.227
Leverage$_{i,t}$	0.218	0.205	0.000	0.056	0.173	0.323	0.478
ROA$_{i,t}$	-0.002	0.158	-0.143	-0.003	0.022	0.064	0.117
Sales growth$_{i,t}$	0.114	0.385	-0.171	-0.044	0.057	0.180	0.380
Firm age$_{i,t}$	26.60	17.02	10	13	20	40	56
Total product similarity$_{i,t}$	14.09	16.80	1.857	2.960	5.305	21.28	46.09
Log(Cash)$_{i,t}$	-2.961	1.443	-5.028	-3.865	-2.746	-1.829	-1.253
PPE$_{i,t}$	0.201	0.246	0.007	0.018	0.084	0.298	0.644
Searched-firm regressors							
Size$_{j,t}$	6.806	2.058	4.089	5.430	6.779	8.145	9.498
Market-to-book ratio$_{j,t}$	1.864	1.479	0.950	1.014	1.268	2.066	3.553
Leverage$_{j,t}$	0.202	0.215	0.000	0.0213	0.137	0.315	0.497
ROA$_{j,t}$	-0.054	0.252	-0.307	-0.026	0.010	0.051	0.103
Sales growth$_{j,t}$	0.179	0.671	-0.188	-0.043	0.064	0.212	0.490
Firm age$_{j,t}$	19.78	14.25	7	10	15	24	44
Log(Filings)$_{j,t}$	3.855	0.572	3.135	3.466	3.850	4.234	4.605
Leader$_{j,t}$	0.015	0.122	0	0	0	0	0
Distress$_{j,t}$	0.022	0.145	0	0	0	0	0
Log(Cash)$_{j,t}$	-2.983	1.651	-5.095	-4.093	-2.857	-1.657	-0.947
PPE$_{j,t}$	0.200	0.259	0.006	0.017	0.069	0.293	0.673
Blockholder$_{j,t}$	0.812	0.391	0	1	1	1	1
Industry M&A$_{j,t-1}$	0.904	0.295	1	1	1	1	1

Panel B: Correlation table (selected variables)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) *Information acquisition*$_{i,j,t}$	1	-0.025	-0.005	0.002	0.047	0.016	0.083	0.181	0.054	-0.092	0.007	0.002
(2) *Encroachment*$_{i,j,t \text{ to } t+1}$	-0.010	1	-0.000	-0.005	-0.017	0.039	-0.296	0.011	-0.001	0.013	-0.005	0.014
(3) *ΔCapex similarity*$_{i,j,t+1}$	-0.007	-0.001	1	0.006	0.000	-0.016	-0.010	-0.005	0.001	-0.004	0.007	-0.008
(4) *ΔR&D similarity*$_{i,j,t+1}$	0.005	0.016	0.004	1	0.002	0.007	0.003	-0.004	0.001	-0.001	-0.065	-0.025
(5) *Acquisition*$_{i,j,t+1}$	0.026	0.001	-0.002	0.004	1	0.045	0.008	0.007	0.004	-0.005	-0.001	0.001
(6) *Private acquisition*$_{i,j,t+1}$	0.021	0.046	-0.023	0.015	0.024	1	0.057	-0.015	0.024	-0.010	-0.002	0.044
(7) *Product similarity*$_{i,j,t}$	0.075	-0.156	0.002	0.010	0.009	0.069	1	0.029	0.031	-0.012	-0.067	-0.061
(8) *Return corr*$_{i,j,t}$	0.216	-0.016	-0.000	-0.023	0.009	-0.007	0.025	1	0.082	-0.538	-0.135	-0.125
(9) *Same auditor*$_{i,j,t}$	0.070	0.002	0.003	0.002	0.004	0.022	0.029	0.088	1	-0.010	0.031	0.036
(10) *Log(Distance)*$_{i,j,t}$	-0.097	0.014	-0.005	-0.005	-0.003	-0.009	-0.032	-0.067	0.008	1	-0.019	-0.025
(11) *Market-to-book ratio*$_{i,t}$	0.040	-0.012	-0.022	0.014	0.002	0.027	-0.178	-0.112	0.045	0.051	1	0.321
(12) *Market-to-book ratio*$_{j,t}$	0.064	0.003	-0.018	0.012	0.002	0.050	-0.147	-0.038	0.066	0.028	0.476	1

Table 2

The associations between information flows among rivals and economic fundamentals

This table presents the results of negative binomial regressions of the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t* (*Information acquisition*$_{i,j,t}$) on firm-pair characteristics, searching firm *i* characteristics, and searched-for firm *j* characteristics. For each panel, Columns 1 and 3 present the coefficients from estimating the model, while Columns 2 and 4 present the estimated incidence rate ratios. The coefficients, standard errors, and incidence rate ratios in Columns 5 and 6 are estimated using the fixed effects negative binomial model of Hausman et al. (1984). Data are restricted to each searching firm's TNIC3 industry, based on the Hoberg and Phillips (2016) text-based network industry classification schema. Panel A presents the results using our verified sample and Panel B presents the results using a predicted sample of IP addresses based on self-search patterns. The construction of the verified and predicted samples is outlined in Section 2 and the Online Appendix. All continuous variables are winsorized at the first and 99th percentiles. We normalize all nonbinary independent variables to have a mean of zero and standard deviation of one and cluster standard errors by firm-pair. See Appendix A for variable definitions. The sample period covers 2004–2015. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Panel A: Verified sample

Variables	(1) Coeff	(2) IRR	(3) Coeff	(4) IRR	(5) Coeff	(6) IRR
Market-to-book ratio$_{i,t}$	0.131***	1.140	0.132***	1.141	0.053***	1.054
	(0.014)		(0.015)		(0.009)	
Market-to-book ratio$_{j,t}$	0.264***	1.302	0.225***	1.252	0.029***	1.029
	(0.017)		(0.017)		(0.010)	
Product similarity$_{i,j,t}$	0.439***	1.551	0.444***	1.560	0.138***	1.148
	(0.016)		(0.017)		(0.009)	
Return corr$_{i,j,t}$	0.491***	1.634	0.489***	1.630	0.085***	1.088
	(0.017)		(0.019)		(0.010)	
Same auditor$_{i,j,t}$	0.308***	1.360	0.297***	1.346	0.103***	1.108
	(0.034)		(0.034)		(0.020)	
Log(Distance)$_{i,j,t}$	-0.272***	0.762	-0.266***	0.767	-0.086***	0.917
	(0.013)		(0.013)		(0.008)	
Product market fluidity$_{i,t}$	-0.336***	0.715	-0.406***	0.667	-0.136***	0.873
	(0.016)		(0.018)		(0.010)	
TNIC HHI$_{i,t}$	0.189***	1.208	0.188***	1.206	0.044***	1.045
	(0.015)		(0.016)		(0.008)	
Size$_{i,t}$	-0.278***	0.758	-0.242***	0.785	-0.250***	0.779
	(0.028)		(0.029)		(0.016)	
Leverage$_{i,t}$	0.114***	1.121	0.110***	1.116	0.106***	1.112
	(0.018)		(0.018)		(0.011)	
ROA$_{i,t}$	0.110***	1.117	0.067***	1.069	0.014	1.014
	(0.016)		(0.016)		(0.009)	
Sales growth$_{i,t}$	0.003	1.003	0.048***	1.049	0.006	1.006
	(0.010)		(0.011)		(0.007)	
Firm age$_{i,t}$	0.152***	1.164	0.147***	1.158	0.150***	1.162
	(0.019)		(0.020)		(0.012)	
Size$_{j,t}$	0.484***	1.622	0.385***	1.469	0.280***	1.324
	(0.026)		(0.026)		(0.015)	
Leverage$_{j,t}$	0.127***	1.135	0.127***	1.135	0.006	1.006
	(0.017)		(0.017)		(0.010)	
ROA$_{j,t}$	-0.046**	0.955	-0.048**	0.953	-0.004	0.996
	(0.023)		(0.023)		(0.012)	
Sales growth$_{j,t}$	-0.016	0.984	-0.019	0.981	-0.023***	0.977
	(0.013)		(0.012)		(0.007)	
Firm age$_{j,t}$	0.106***	1.112	0.089***	1.093	-0.012	0.988
	(0.016)		(0.016)		(0.010)	

Log(Filings)$_{j,t}$	0.154***	1.166	0.310***	1.364	0.040***	1.041
	(0.016)		(0.019)		(0.009)	
Leader$_{j,t}$	0.386***	1.471	0.419***	1.521	0.101***	1.106
	(0.062)		(0.064)		(0.037)	
Distress$_{j,t}$	-0.248*	0.780	-0.267**	0.766	-0.044	0.957
	(0.128)		(0.120)		(0.083)	
Observations	252,370		252,370		74,917	
Firm-pair effects	No		No		Yes	
Year FE	No		Yes		Yes	

Panel B: Predicted IP address sample

Variables	(1) Coeff	(2) IRR	(3) Coeff	(4) IRR	(5) Coeff	(6) IRR
Market-to-book ratio$_{i,t}$	0.203***	1.226	0.216***	1.241	0.111***	1.118
	(0.008)		(0.008)		(0.006)	
Market-to-book ratio$_{j,t}$	0.272***	1.313	0.261***	1.298	0.075***	1.078
	(0.008)		(0.008)		(0.005)	
Product similarity$_{i,j,t}$	0.259***	1.295	0.276***	1.318	0.106***	1.111
	(0.008)		(0.008)		(0.006)	
Return corr$_{i,j,t}$	0.487***	1.627	0.449***	1.567	0.142***	1.152
	(0.007)		(0.008)		(0.005)	
Same auditor$_{i,j,t}$	0.118***	1.125	0.135***	1.144	0.047***	1.048
	(0.016)		(0.016)		(0.011)	
Log(Distance)$_{i,j,t}$	-0.247***	0.781	-0.245***	0.783	-0.120***	0.887
	(0.006)		(0.006)		(0.005)	
Product market fluidity$_{i,t}$	-0.176***	0.838	-0.244***	0.784	-0.076***	0.927
	(0.008)		(0.009)		(0.006)	
TNIC HHI$_{i,t}$	0.131***	1.139	0.110***	1.116	0.054***	1.056
	(0.007)		(0.007)		(0.005)	
Size$_{i,t}$	0.242***	1.274	0.277***	1.319	0.190***	1.209
	(0.010)		(0.010)		(0.009)	
Leverage$_{i,t}$	0.100***	1.106	0.106***	1.112	0.086***	1.090
	(0.007)		(0.007)		(0.006)	
ROA$_{i,t}$	0.006	1.006	-0.007	0.993	0.048***	1.049
	(0.008)		(0.009)		(0.007)	
Sales growth$_{i,t}$	0.056***	1.057	0.066***	1.068	0.001	1.001
	(0.007)		(0.007)		(0.005)	
Firm age$_{i,t}$	0.121***	1.129	0.096***	1.101	0.097***	1.102
	(0.007)		(0.008)		(0.006)	
Size$_{j,t}$	0.350***	1.418	0.282***	1.326	0.199***	1.220
	(0.010)		(0.011)		(0.008)	
Leverage$_{j,t}$	0.151***	1.163	0.155***	1.168	0.053***	1.054
	(0.007)		(0.007)		(0.005)	
ROA$_{j,t}$	-0.052***	0.949	-0.028***	0.972	-0.002	0.998
	(0.009)		(0.009)		(0.007)	
Sales growth$_{j,t}$	0.025***	1.026	0.029***	1.029	-0.015***	0.985
	(0.006)		(0.006)		(0.004)	
Firm age$_{j,t}$	0.132***	1.141	0.115***	1.122	0.095***	1.100
	(0.007)		(0.007)		(0.005)	
Log(Filings)$_{j,t}$	0.249***	1.283	0.390***	1.477	0.123***	1.131
	(0.007)		(0.008)		(0.005)	
Leader$_{j,t}$	0.756***	2.130	0.802***	2.231	0.169***	1.184
	(0.033)		(0.034)		(0.022)	
Distress$_{j,t}$	-0.119*	0.888	-0.042	0.959	0.065	1.067
	(0.067)		(0.069)		(0.045)	
Observations	1,279,692		1,279,692		240,141	
Firm-pair effects	No		No		Yes	
Year FE	No		Yes		Yes	

Table 3

Information flows among rivals and shocks to investment opportunities—the U.S. federal government budget crisis

This table presents the results from estimating difference-in-differences regressions of the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t* ($Information\ acquisition_{i,j,t}$) on a shock to investment opportunities for government contractors during a governmental budget crisis ($BCA_t \times High\ contract_i$). We use unexpected reductions in U.S. government spending in 2011–2013 due to the passage of the Budget Control Act of 2011 as a shock to investment opportunities, via a reduction in product demand for government contractors. We identify firms that are subject to the shock as those firms for which U.S. government contracts accounted for 5% or more of total revenue in 2010 ($High\ contract_i = 1$). We then create an indicator variable for the calendar years 2011–2013 ($BCA_t = 1$) and interact it with $High\ contract_i$. Column 1 presents the coefficients from estimating the model, while Column 2 presents the estimated incidence rate ratios. Data are restricted to each searching firm's TNIC3 industry, based on the Hoberg and Phillips (2016) text-based network industry classification schema. The model includes firm-pair characteristics, searching firm *i* characteristics, and searched-for firm *j* characteristics and uses the fixed effects negative binomial regression model of Hausman et al. (1984). All continuous variables are winsorized at the first and 99th percentiles. We normalize all nonbinary independent variables to have a mean of zero and standard deviation of one and cluster standard errors by firm-pair. Some controls are untabulated for brevity and consist of the following variables: $Market\text{-}to\text{-}book\ ratio_{i,t}$, $Market\text{-}to\text{-}book\ ratio_{j,t}$, $Product\ market\ fluidity_{i,t}$, $TNIC\ HHI_{i,t}$, $Size_{i,t}$, $Leverage_{i,t}$, $ROA_{i,t}$, $Sales\ growth_{i,t}$, $Firm\ age_{i,t}$, $Size_{j,t}$, $Leverage_{j,t}$, $ROA_{j,t}$, $Sales\ growth_{j,t}$, $Firm\ age_{j,t}$, $Log(Filings)_{j,t}$, $Leader_{j,t}$, and $Distress_{j,t}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. The sample period covers 2008–2013. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Variables	(1) Coeff	(2) IRR
$BCA_t \times High\ contract_i$	-0.888***	0.412
	(0.098)	
$High\ contract_i$	0.389***	1.475
	(0.101)	
$Product\ similarity_{i,j,t}$	0.162***	1.175
	(0.013)	
$Return\ corr_{i,j,t}$	0.082***	1.085
	(0.015)	
$Same\ auditor_{i,j,t}$	0.062**	1.064
	(0.029)	
$Log(Distance)_{i,j,t}$	-0.047***	0.954
	(0.011)	
Observations	37,543	
Controls included	Yes	
Firm-pair FE	Yes	
Year FE	Yes	

Table 4

Information flows among rivals and shocks to investment opportunities—industry-level import tariff cuts

This table presents the results from estimating regressions of the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t* ($Information\ acquisition_{i,j,t}$) on a shock to investment opportunities based on tariff cuts ($Tariff\ cut_{i,t}$). We use plausibly exogenous changes in tariff rates as a shock to investment opportunities, via an increase in foreign competition. The indicator variable, $Tariff\ cut_{i,t}$, is constructed as follows: we first collect product-level import data from the United States International Trade Commission (USITC) for the period 2004–2014 at the four-digit SIC industry level, similar to that compiled for earlier periods by Feenstra (1996) and Feenstra et al. (2002). We then calculate the ad valorem tariff rates for each industry-year as the duties collected by U.S. Customs divided by the free-on-board value of imports. We follow Fresard (2010) and measure unexpected tariff cuts as a negative change in tariff rates that is at least three times the median change and not followed by an equivalent increase in the following two years. Column 1 presents the coefficients from estimating the model, while Column 2 presents the estimated incidence rate ratios. Data are restricted to each searching firm's TNIC3 industry, based on the Hoberg and Phillips (2016) text-based network industry classification schema. As in prior studies, these data are limited to manufacturing firms (four-digit SIC codes between 2000 and 3999). The model includes firm-pair characteristics, searching firm *i* characteristics, and searched-for firm *j* characteristics and uses the fixed effects negative binomial regression model of Hausman et al. (1984). All continuous variables are winsorized at the first and 99th percentiles. We normalize all nonbinary independent variables to have a mean of zero and standard deviation of one. Standard errors are clustered by firm-pair. Some controls are untabulated for brevity and consist of the following variables: $Market\text{-}to\text{-}book\ ratio_{i,t}$, $Market\text{-}to\text{-}book\ ratio_{j,t}$, $Product\ market\ fluidity_{i,t}$, $TNIC\ HHI_{i,t}$, $Size_{i,t}$, $Leverage_{i,t}$, $ROA_{i,t}$, $Sales\ growth_{i,t}$, $Firm\ age_{i,t}$, $Size_{j,t}$, $Leverage_{j,t}$, $ROA_{j,t}$, $Sales\ growth_{j,t}$, $Firm\ age_{j,t}$, $Log(Filings)_{j,t}$, $Leader_{j,t}$, and $Distress_{j,t}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. The sample period covers 2004–2014. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Variables	(1) Coeff	(2) IRR
$Tariff\ cut_{i,t}$	-0.139***	0.870
	(0.038)	
$Product\ similarity_{i,j,t}$	0.174***	1.190
	(0.016)	
$Return\ corr_{i,j,t}$	0.056***	1.058
	(0.017)	
$Same\ auditor_{i,j,t}$	0.085**	1.089
	(0.036)	
$Log(Distance)_{i,j,t}$	-0.158***	0.854
	(0.015)	
Observations	24,425	
Controls included	Yes	
Firm-pair effects	Yes	
Year FE	Yes	

Table 5
The predictive ability of pairwise information flows for future acquisitions

This table presents the results from regressions of future M&A activity between *i*,*j* firm-pairs ($Acquisition_{i,j,t+1}$) on information flows between the rival-pair ($Information\ acquisition_{i,j,t}$), measured as the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t*. Panel A presents the results of logit models examining factors associated with future acquisitions of public firms *j* by firms *i*. The dependent variable is an indicator variable set to one if firm *i* acquires public firm *j* in year *t*+1. Panel B examines cross-sectional variation in the associations between information acquisition and future acquisitions based on the product similarity of firms *i* and *j*. To isolate information flows among rival firms, observations must fall within acquiring firm *i*'s product space [text-based industry classifications, as in Hoberg and Phillips (2016)]. All continuous variables are winsorized at the first and 99th percentiles. In both panels, we normalize all nonbinary independent variables to have a mean of zero and standard deviation of one. Standard errors are clustered by firm-pair. Some controls are untabulated for brevity and consist of the following variables: $Market\text{-}to\text{-}book\ ratio_{i,t}$, $Market\text{-}to\text{-}book\ ratio_{j,t}$, $Product\ market\ fluidity_{i,t}$, $TNIC\ HHI_{i,t}$, $Size_{i,t}$, $Leverage_{i,t}$, $ROA_{i,t}$, $Firm\ age_{i,t}$, $Total\ product\ similarity_{i,t}$, $Log(Cash)_{i,t}$, $PPE_{i,t}$, $Size_{j,t}$, $Leverage_{j,t}$, $ROA_{j,t}$, $Firm\ age_{j,t}$, $Log(Cash)_{i,t}$, $PPE_{j,t}$, $Blockholder_{j,t,}$ and $Industry\ M\&A_{j,t\text{-}1}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. The sample period covers 2004–2015. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Panel A: Information flows between the rival-pair and subsequent pairwise acquisitions

	(1)		(2)		(3)	
	$Acquisition_{i,j,t+1}$		$Acquisition_{i,j,t+1}$		$Acquisition_{i,j,t+1}$	
Variables	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio
$Information\ acquisition_{i,j,t}$	0.479***	1.614	0.441***	1.554	0.463***	1.589
	(0.028)		(0.043)		(0.045)	
$Product\ similarity_{i,j,t}$			0.555***	1.742	0.544***	1.724
			(0.099)		(0.100)	
$Return\ corr_{i,j,t}$			0.163	1.177	0.303**	1.354
			(0.116)		(0.122)	
$Log(Distance)_{i,j,t}$			-0.059	0.942	-0.056	0.946
			(0.083)		(0.084)	
Observations	252,370		252,370		252,370	
Controls included	No		Yes		Yes	
Firm-pair FE	No		No		No	
Year FE	No		No		Yes	

Panel B: Information flows and subsequent pairwise acquisitions—product similarity heterogeneity

	(1)		(2)	
	$Acquisition_{i,j,t+1}$		$Acquisition_{i,j,t+1}$	
Variables	Coeff	Odds ratio	Coeff	Odds ratio
$Information\ acquisition_{i,j,t}$	0.468***	1.598	0.492***	1.635
	(0.040)		(0.042)	
$Information\ acquisition_{i,j,t} \times Product\ similarity_{i,j,t}$	-0.063**	0.939	-0.066**	0.936
	(0.027)		(0.027)	
$Product\ similarity_{i,j,t}$	0.757***	2.131	0.756***	2.129
	(0.112)		(0.112)	
$Return\ corr_{i,j,t}$	0.158	1.171	0.301**	1.351
	(0.114)		(0.120)	
$Log(Distance)_{i,j,t}$	-0.062	0.940	-0.057	0.945
	(0.082)		(0.083)	
Observations	252,370		252,370	
Controls included	Yes		Yes	
Firm-pair FE	No		No	
Year FE	No		Yes	

Table 6

The predictive ability of public firm information flows for future acquisitions of private targets

This table presents the results from regressions of future private-firm acquisitions (*Private acquisition*$_{i,j,t+1}$) on public rival-pair information flows (*Information acquisition*$_{i,j,t}$), measured as the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t*. Panel A presents the results of logit models examining factors associated with future acquisitions of private firms by firms *i*, where the acquired firm has the same two-digit SIC code as firm *j*. The dependent variable is an indicator variable set to one if firm *i* acquires a private firm in the same two-digit SIC code as firm *j* in year *t*+1. Panel B examines cross-sectional variation in the associations between information acquisition and future acquisitions of private firms based on whether the acquired firm is in the same four-digit SIC code as both firms *i* and *j*. To isolate information flows among *i*,*j* rival-pairs, observations must fall within acquiring firm *i*'s product space [text-based industry classifications, as in Hoberg and Phillips (2016)]. All continuous variables are winsorized at the first and 99th percentiles. In both panels, we normalize all nonbinary independent variables to have a mean of zero and standard deviation of one. Standard errors are clustered by firm-pair. Some controls are untabulated for brevity and consist of the following variables: *Market-to-book ratio*$_{i,t}$, *Market-to-book ratio*$_{j,t}$, *Product market fluidity*$_{i,t}$, *TNIC HHI*$_{i,t}$, *Size*$_{i,t}$, *Leverage*$_{i,t}$, *ROA*$_{i,t}$, *Firm age*$_{i,t}$, *Total product similarity*$_{i,t}$, *Log(Cash)*$_{i,t}$, *PPE*$_{i,t}$, *Size*$_{j,t}$, *Leverage*$_{j,t}$, *ROA*$_{j,t}$, *Firm age*$_{j,t}$, *Log(Cash)*$_{i,t}$, *PPE*$_{j,t}$, *Blockholder*$_{j,t,}$, and *Industry M&A*$_{j,t-1}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. The sample period covers 2004–2015. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Panel A: Information flows for public firms and acquisitions of private firms

	(1)		(2)		(3)	
	Private acquisition$_{i,j,t+1}$		*Private acquisition*$_{i,j,t+1}$		*Private acquisition*$_{i,j,t+1}$	
Variables	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio
Information acquisition$_{i,j,t}$	0.042***	1.043	0.065***	1.067	0.075***	1.078
	(0.006)		(0.006)		(0.007)	
Product similarity$_{i,j,t}$			0.194***	1.214	0.203***	1.225
			(0.009)		(0.009)	
Return corr$_{i,j,t}$			-0.042***	0.959	0.029***	1.030
			(0.008)		(0.009)	
Log(Distance)$_{i,j,t}$			-0.011	0.990	-0.009	0.991
			(0.007)		(0.007)	
Observations	252,370		252,370		252,370	
Controls included	No		Yes		Yes	
Firm-pair FE	No		No		No	
Year FE	No		No		Yes	

Panel B: Acquisitions of private firms—industry heterogeneity

Variables	(1) *Private acquisition*$_{i,j,t+1}$ Coeff	Odds ratio	(2) *Private acquisition*$_{i,j,t+1}$ Coeff	Odds ratio
Information acquisition$_{i,j,t}$	0.083***	1.086	0.092***	1.096
	(0.010)		(0.010)	
Information acquisition$_{i,j,t}$ × *Same SIC*$_{i,j,t}$	-0.063***	0.939	-0.061***	0.941
	(0.013)		(0.013)	
Same SIC$_{i,j,t}$	0.753***	2.123	0.704***	2.021
	(0.017)		(0.018)	
Product similarity$_{i,j,t}$	0.082***	1.085	0.097***	1.102
	(0.009)		(0.009)	
Return corr$_{i,j,t}$	-0.091***	0.913	-0.022**	0.978
	(0.008)		(0.009)	
Log(Distance)$_{i,j,t}$	-0.015**	0.985	-0.013*	0.987
	(0.007)		(0.007)	
Observations	252,370		252,370	
Controls included	Yes		Yes	
Firm-pair FE	No		No	
Year FE	No		Yes	

Table 7
Information flows and subsequent rival-pair investment mimicking

This table presents the results from regressions of changes in investment similarities between *i,j* firm-pairs (Δ*Capex similarity*$_{i,j,t+1}$ or Δ*R&D similarity*$_{i,j,t+1}$) on information flows between the rival-pair (*Information acquisition*$_{i,j,t}$), measured as the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t*. To capture changes in investment similarities, we examine firm-pair level changes in capital expenditures (scaled by lagged fixed assets) and changes in research and development expenses (scaled by sales). Panel A presents the results of Ordinary Least Squares (OLS) regressions examining factors associated with the change in similarity between firm *i* and firm *j's* capital expenditures and between firm *i* and firm *j's* research and development expenses. Panel B examines cross-sectional variation in the associations between information acquisition and investment similarities based on the product similarity of firms *i* and *j*. To isolate information flows among *i,j* rival-pairs, observations must fall within firm *i*'s product space [text-based industry classifications, as in Hoberg and Phillips (2016)]. All continuous variables are winsorized at the first and 99th percentiles. We normalize all nonbinary independent variables to have a mean of zero and standard deviation of one. Standard errors are clustered by firm-pair. Some controls are untabulated for brevity and consist of the following variables: *Market-to-book ratio*$_{i,t}$, *Market-to-book ratio*$_{j,t}$, *Product market fluidity*$_{i,t}$, *TNIC HHI*$_{i,t}$, *Size*$_{i,t}$, *Leverage*$_{i,t}$, *ROA*$_{i,t}$, *Sales growth*$_{i,t}$, *Firm age*$_{i,t}$, *Size*$_{j,t}$, *Leverage*$_{j,t}$, *ROA*$_{j,t}$, *Sales growth*$_{j,t}$, *Firm age*$_{j,t}$, *Log(Filings)*$_{j,t}$, *Leader*$_{j,t}$, and *Distress*$_{j,t}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Panel A: Pairwise information flows and pairwise investment similarity

Variables	(1) Δ*Capex similarity*$_{i,j,t+1}$	(2) Δ*R&D similarity*$_{i,j,t+1}$
Information acquisition$_{i,j,t}$	0.006**	0.005*
	(0.003)	(0.003)
Product similarity$_{i,j,t}$	-0.017***	0.022***
	(0.005)	(0.005)
Return corr$_{i,j,t}$	0.032***	0.001
	(0.006)	(0.005)
Same auditor$_{i,j,t}$	0.068***	-0.058***
	(0.014)	(0.012)
Log(Distance)$_{i,j,t}$	0.024*	0.021
	(0.013)	(0.014)
Observations	204,555	218,208
R-squared	0.293	0.373
Controls included	Yes	Yes
Firm-pair FE	Yes	Yes
Year FE	Yes	Yes

Panel B: Pairwise information flows and pairwise investment similarity—product similarity heterogeneity

Variables	(1) Δ*Capex similarity*$_{i,j,t+1}$	(2) Δ*R&D similarity*$_{i,j,t+1}$
Information acquisition$_{i,j,t}$	0.006**	0.009***
	(0.003)	(0.003)
Information acquisition$_{i,j,t}$ × *Product similarity*$_{i,j,t}$	0.000	-0.011***
	(0.002)	(0.003)
Product similarity$_{i,j,t}$	-0.017***	0.023***
	(0.005)	(0.005)
Return corr$_{i,j,t}$	0.032***	0.001
	(0.006)	(0.005)
Same auditor$_{i,j,t}$	0.068***	-0.058***
	(0.014)	(0.012)
Log(Distance)$_{i,j,t}$	0.024*	0.021**
	(0.013)	(0.014)
Observations	204,555	218,208
R-squared	0.293	0.373
Controls included	Yes	Yes
Firm-pair FE	Yes	Yes
Year FE	Yes	Yes

Table 8

Information flows and subsequent rival-pair product similarities

This table presents the results from regressions of subsequent changes in product similarities between *i,j* rival-pairs ($Encroachment_{i,j,t\ to\ t+k}$) on pairwise information flows between the rivals ($Information\ acquisition_{i,j,t}$), measured as the total number of firm *j* filings downloaded from the SEC's EDGAR database by firm *i* during year *t*. Panel A presents the results of OLS regressions of ex post product encroachment on pairwise information flows, where product encroachment is defined as the change in similarity of firm *i*'s products with rival firm *j*'s products, measured over three future periods (years $t+k$, where k = 1, 2, or 3). Panel B examines cross-sectional variation in the associations between information acquisition and encroachment based on the product similarity of firms *i* and *j*. To isolate information flows among *i,j* rival-pairs, observations must fall within firm *i*'s product space [text-based industry classifications, as in Hoberg and Phillips (2016)]. All continuous variables are winsorized at the first and 99th percentiles. We normalize all nonbinary independent variables to have a mean of zero and standard deviation of one. Standard errors are clustered by firm-pair. Some controls are untabulated for brevity and consist of the following variables: $Market\text{-}to\text{-}book\ ratio_{i,t}$, $Market\text{-}to\text{-}book\ ratio_{j,t}$, $Product\ market\ fluidity_{i,t}$, $TNIC\ HHI_{i,t}$, $Size_{i,t}$, $Leverage_{i,t}$, $ROA_{i,t}$, $Sales\ growth_{i,t}$, $Firm\ age_{i,t}$, $Total\ product\ similarity_{i,t}$, $Size_{j,t}$, $Leverage_{j,t}$, $ROA_{j,t}$, $Sales\ growth_{j,t}$, $Firm\ age_{j,t}$, $Log(Filings)_{j,t}$, $Leader_{j,t}$, and $Distress_{j,t}$. See Appendix A for variable definitions and the Online Appendix for the tabulation of all control variables. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (two-tailed).

Panel A: Ex post product encroachment

Variables	(1) $Encroachment_{i,j,t\ to\ t+1}$	(2) $Encroachment_{i,j,t\ to\ t+2}$	(3) $Encroachment_{i,j,t\ to\ t+3}$
$Information\ acquisition_{i,j,t}$	-0.008**	-0.011***	-0.017***
	(0.003)	(0.003)	(0.003)
$Product\ similarity_{i,j,t}$	-1.046***	-1.141***	-1.078***
	(0.008)	(0.008)	(0.007)
$Return\ corr_{i,j,t}$	0.045***	0.047***	0.011**
	(0.005)	(0.005)	(0.004)
$Same\ auditor_{i,j,t}$	0.000	0.049***	-0.018
	(0.013)	(0.013)	(0.013)
$Log(Distance)_{i,j,t}$	0.002	0.009	0.017*
	(0.011)	(0.009)	(0.009)
Observations	240,169	224,311	208,036
R-squared	0.443	0.588	0.679
Controls included	Yes	Yes	Yes
Pair FE	Yes	Yes	Yes

Panel B: Ex post product encroachment—product similarity heterogeneity

Variables	(1) $Encroachment_{i,j,t\ to\ t+1}$	(2) $Encroachment_{i,j,t\ to\ t+2}$	(3) $Encroachment_{i,j,t\ to\ t+3}$
$Information\ acquisition_{i,j,t}$	-0.005*	-0.007**	-0.011***
	(0.003)	(0.003)	(0.003)
$Information\ acquisition_{i,j,t} \times Product\ similarity_{i,j,t}$	-0.007**	-0.010***	-0.015***
	(0.004)	(0.004)	(0.004)
$Product\ similarity_{i,j,t}$	-1.045***	-1.139***	-1.077***
	(0.008)	(0.008)	(0.007)
$Return\ corr_{i,j,t}$	0.045***	0.047***	0.011**
	(0.005)	(0.005)	(0.004)
$Same\ auditor_{i,j,t}$	0.000	0.049***	-0.018
	(0.013)	(0.013)	(0.013)
$Log(Distance)_{i,j,t}$	0.002	0.009	0.017*
	(0.011)	(0.009)	(0.009)
Observations	240,169	224,311	208,036
R-squared	0.443	0.588	0.680
Controls included	Yes	Yes	Yes
Pair FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes