The real effects of financial markets on scientific disclosure: Evidence from a quasi-natural experiment*

Stefano H. Baruffaldi[†a], Markus Simeth[‡b], and David Wehrheim[§c]

[a]University of Bath & Max Planck Institute for Innovation and Competition
[b]Copenhagen Business School
[c]IESE Business School

This version: *December 28, 2020*

ABSTRACT

While innovation disclosure is essential for cumulative knowledge production and economic growth, evidence on firm incentives to disclose innovation outcomes is lacking. We examine the role of financial markets in firms' decisions to disseminate scientific research results. We employ a quasi-natural experiment that exploits plausibly exogenous variation in analyst coverage, resulting in higher information asymmetries. We find that firms respond by a quick and enduring increase in scientific publications. We also show that disclosure decisions are shaped by financial constraints and managerial incentives. We discuss important implications, such as potential crowding out effects between transparency initiatives and socially desirable innovation disclosure.

KEYWORDS: information asymmetries, disclosure, corporate science, financial analysts, proprietary costs, innovation, appropriation, signaling

JEL Classification: O32, O34, G30

---

*We thank Benjamin Balsmeier, Iván Blanco, Bruno Cassiman, Daniel Streitz, Thomas Rønde as well as the workshop participants at the Barcelona GSE Summer Forum (2019), the University of Luxembourg (2019), Ecole Polytechnique Fédérale de Lausanne (2020), ZEW Mannheim (2020), and seminar participants at CUNEF (2020), ESADE Entrepreneurship Institute (2020), and IESE Business School (2020) for their helpful comments and suggestions. All errors are our own. The authors acknowledge financial support from the Independent Research Fund Denmark (Project 0133-00119B) and the Spanish Ministry of Science, Innovation and Universities (Project PGC2018-094418-B-I00). Authors are listed in alphabetic order.

†University of Bath, School of Management, United Kingdom; University of Bath and Max Planck Institute for Innovation and Competition, Munich, Germany; s.h.baruffaldi@bath.ac.uk.

‡Copenhagen Business School, Department of Strategy and Innovation, Denmark; ms.si@cbs.dk.

§IESE Business School, Strategic Management Department, Spain; dwehrheim@iese.edu.

1. Introduction

Does a firm's financial information environment affect its innovation disclosure practices? Theoretically, a reduction in asymmetric information improves a firm's access to financing, which can lead to an increase in innovation activities, because such activities are particularly likely to be sensitive to financing constraints (Arrow, 1962; Hall and Lerner, 2010). On the other hand, an enhanced information environment makes it more difficult for firms to conceal their innovative activities from the marketplace, and the costs associated with information leakage to competitors are particularly severe with regard to the disclosure of innovative ideas. If these competitive costs are high enough, then a richer information environment can make managers more reluctant to disclose extensive information about innovation (Bhattacharya and Ritter, 1983; Anton and Yao, 2004). While prior work has provided valuable insights on the relationship between information asymmetry and financial disclosure decisions (Balakrishnan, Billings, Kelly, and Ljungqvist, 2014; Irani and Oesch, 2013), we lack evidence in contexts where disclosure is costly to the firm.

In this paper, we analyze the link between asymmetric information and the decision to voluntarily disclose research outcomes in the form of scientific publications. The requirements for academic novelty and the peer review process make publications a truthful and credible signal to investors (Dasgupta and David, 1994). Consistent with this view, qualitative evidence in the context of pharmaceutical companies suggests that "[g]etting research published in a peer-reviewed journal is like winning a stamp of approval from its most influential audience. It's an automatic validation unmatched by any other medium" (Polidoro and Theeke, 2012, p. 1137). At the same time, it is well established that scientific disclosure is accompanied by serious appropriability problems. Indeed, recent literature examines the private returns from scientific disclosure, and while these studies show a positive association between scientific publication stocks and firm value, they also document substantial heterogeneity in this relationship, confirming the trade-off in disclosure decisions (Simeth and Cincera, 2016; Arora, Belenzon, and Sheer, 2020).

Understanding when and why firms disclose scientific research outcomes is of high interest to policy makers and scholarly researchers given the critical role of knowledge disclosure in endogenous growth models (e.g., Romer, 1990; Lucas and Moll, 2014) and the broad shift from government-funded research to private funding of research.[1] Figure 1 plots basic and applied research expenditures by U.S. firms between 1990 and 2015 (Panel A), and their output as measured by the total number of scientific publications and patents (Panel B). Scientific publication outputs are substantial, albeit growth has been moderate and stagnated in recent years, especially relative to the increasing basic and applied research as well as the increase in patenting. These diverging trends have revived the debate on the determinants of firms' contributions to the scientific knowledge stock (Arora, Belenzon, and Sheer, 2020; Bloom, Jones, Van Reenen, and Webb, 2020).

[Figure 1 here]

In light of the theoretical tension, empirical investigation is necessary. Analyzing the relationship between proxies for information asymmetry and corporate disclosure is challenging due to potential endogeneity.[2] To overcome this hurdle, we employ a quasi-experimental design relying on reduced analyst coverage resulting from plausibly exogenous broker house mergers and closures. Broker mergers and closures lead to termination of analyst employments, and consequently, coverage decreases for those firms previously covered by the analysts. This approach was originally proposed by Hong and Kacperczyk (2010) to test whether such declines are associated with increases in earnings forecast bias. Consistent with reduced information production, the authors document an increase in the forecast error variance among analysts who continue to cover the stock. Kelly and Ljungqvist (2012) use asset pricing tests and several proxies for information asymmetry based on measures of market reaction to earnings announcements and also show that reduced coverage causes an increase in information asymmetry.

Our identification strategy uses 43 broker merger and closure events staggered between the years 2000 and 2010. It accommodates large, publicly traded U.S. firms with at least one year of positive R&D expenditures and at least one patent during our sample period. Associated with these events are 750 firms that were either covered by both merging houses or the closing house in the year prior to the event (our treatment sample). Using a difference-in-differences (DiD)

[1] R&D performed in the United States has been traditionally funded in part by the business sector and in part by the federal government, with nearly identical shares in the 1980s. The business sector's share has been increasing and reached almost 70% by 2017. In absolute terms, the business sector invested over $380 billion in 2017 and is by far the largest performer of U.S. R&D. The above data were sourced from the National Center for Science and Engineering Statistics, available at https://ncses.nsf.gov/pubs/nsb20203.

[2] For instance, overconfident managers could engage in greater risk taking, which manifests in higher levels of asymmetric information. Simultaneously, those managers could invest less in information production, leading to more limited disclosure. Our relationship of interest could also suffer from reverse causality: Firms making more voluntary disclosures may naturally be easier to evaluate(e.g., they could be larger and followed by more analysts), resulting in lower information asymmetry.

approach, we compare the changes in the scientific research outputs of the treated firms relative to a control group of observationally similar firms that were unaffected by the merger/closure, thus identifying the causal change in corporate scientific publication outputs resulting from the loss of coverage.

Using this empirical design, we provide causal evidence that a firm's capital market information environment influences its scientific disclosure policies. We find a prompt and persistent increase in scientific publications following an exogenous reduction in coverage: losing one analyst causes a substantial 12% increase in scientific publication over a 3-year period compared to firms without any decrease in analyst coverage. We further document that this treatment effect extends to the significance and quality of scientific research outputs (as measured by journal impact factor and future citations). Our main estimates include firm, event, and year fixed effects to account for time-invariant unobservable factors particular to an event or a firm that may influence scientific outputs, as well as time trends.

We conduct several tests to confirm the validity and robustness of these results. First, we provide evidence in support of the parallel assumption trends underlying our DiD approach: the publication behaviors of our treatment and control firms diverge only after the reduction in coverage. We conduct additional tests to examine whether our results remain robust to using a DiD matching estimator and broker closures compared to broker mergers, and whether they are driven by clustering in time of broker disappearances. We find that our results continue to hold in each of these tests. We also address serial correlation problems by collapsing the time series information into pre-treatment and post-treatment periods and experimenting with various clustering schemes. Moreover, consistent results emerge when we consider an alternative data source for scientific publications as well as variations in pre-treatment and post-treatment contrasts.

To assess the extent to which managers face trade-offs in choosing optimal disclosure levels, first, we examine how the effect of coverage terminations on firm scientific disclosure behavior depends on the need to access external financing. If scientific disclosure is costly, then there is little reason to believe that managers are willing to increase publication levels when the costs associated with the loss of coverage are irrelevant to the firm. Derrien and Kecskés (2013) show that the decrease in analyst coverage is not relevant to firms that have sufficient internal capital to finance their investments. On the other hand, firms with a greater need to access capital markets suffer to a greater extent from an exogenous increase in asymmetric information and therefore have a greater need to convince investors of their quality and growth potential. Our tests show that the effect of coverage terminations on scientific publications is concentrated among firms with greater ex ante financial constraints and those with better ex ante growth opportunities.

Second, we document how our treatment effect varies according to managerial incentives.[3] Using predictions from the takeover exposure model of Stein (1988), we argue that some managers are more concerned about their careers and therefore care more about short-term stock prices, potentially at some cost. Other managers are less concerned and therefore avoid disclosure costs of signaling to the market that the long-run outlook of the firm is promising. Our tests show that the effect of coverage on scientific disclosure is concentrated among firms run by managers with ex ante greater career concerns (measured by anti-takeover provisions or the executive's age), consistent with managerial signaling motives. In addition, we observe no shift in scientific publications among treated firms where managerial wealth is more closely tied to firm wealth; this is, the managers are more concerned about the long-run competitive costs of scientific disclosure. Overall, our results imply that managers face substantial cost-benefit trade-offs in their scientific disclosure decisions; disclosure occurs only when the benefits from disclosure outweigh the costs.

Finally, an increase in asymmetric information is not the only theory that can predict a rise in scientific publication due to a reduction in analyst coverage. Other explanations could be based on the role of analysts as effective monitoring agents and theories in which analysts assume a dark side that induces short-term performance pressures because they ignore long-term investments. Both are potentially important, and we discuss their empirical relevance. The monitoring hypothesis would predict a complementarity between monitoring by analysts and traditional governance mechanisms. However, we observe the opposite in our data and therefore conclude that the monitoring role of analysts does not appear to be a mechanism that links coverage terminations and scientific publications. The second class of theories (the dark side) implies changes in innovation outcomes across the board. We document that R&D investments and patent outcomes remain unaffected by the loss in coverage. Moreover, we do not find that the reduction in coverage causes a change in the propensity to use scientific knowledge

[3] Beyer, Cohen, Lys, and Walther (2010, p. 305) review the literature on information problems and corporate disclosure practices and conclude that managerial incentives are important because it is the manager and not the "firm" that decides on disclosure levels. In their own words, "the costs and benefits of disclosure that explain disclosure decisions reflect management's utility and disutility from making a disclosure." The authors also emphasize that two potentially important factors that affect management's utility from disclosure are corporate governance and managerial compensation, both of which we analyze.

in the firm's downstream invention.

The remainder of this paper is organized as follows. Section 2 relates our contribution to the existing literature. Section 3 presents the data, and Section 4 discusses our identification strategy. The empirical results are presented in Section 5. We first provide our baseline results, then discuss the extensive robustness tests, and finally report evidence on the underlying economic mechanism. Section 6 concludes this paper.

2. Related literature

Our paper contributes to several strands of literature. First, this work relates to the finance literature that explores how capital markets affect corporate R&D outcomes. Earlier studies examined, for instance, the effects from institutional ownership (Aghion, Van Reenen, and Zingales, 2013), stock liquidity (Fang, Tian, and Tice, 2014), public offering decisions (Bernstein, 2015), financial derivatives (Blanco and Wehrheim, 2017), or hedge fund activism (Brav, Jiang, Ma, and Tian, 2018). We add to this literature by offering an alternative perspective on how specific features of the capital market can affect observable innovation outcomes. Our findings show that the capital market consequences for patent- and non-patent based measures of innovation can be different. Therefore caution is needed when interpreting results based on a single innovation outcome (see also the discussion in Kerr and Nanda, 2015). In particular, our evidence suggests that the literature attributing capital markets a role in shaping the incentives of firms to innovate may in part reflect differences in disclosure dynamics.[4] It is important to distinguish between these explanations, because they differ in terms of their implications for welfare and policy.

We add to the literature that study firms' motives to disclose scientific research outcomes. Recent evidence by Arora, Belenzon, and Sheer (2020) documenting that scientific publications are associated with knowledge spillovers has fueled a long-standing debate over how firms benefit from publishing activities. Rosenberg (1990) and Hicks (1995) argue that scientific publications enable firms to access scientific communities and to transfer knowledge from these communities to the firm, thereby allowing them to more rapidly reach the frontier and exploit first-mover

[4]In their conclusions, Fang, Tian, and Tice (2014, p. 2123) acknowledge the disclosure channel as an alternative interpretation for how stock illiquidity might translate into more innovation outcomes and highlight "this issue as an interesting and important area for future research, particularly if measures of non-patentable innovation and longer-term investment outputs become available." Our paper contributes in the methodological context by providing such a measure.

advantages.[5] Another motivation relates to the role of granting publication opportunities to corporate researchers in order to attract and retain the best PhD graduates and researchers (Henderson and Cockburn, 1994; Stern, 2004). Scientific publication may also help for the adoption of science-based products, such as for advertising the effectiveness of drugs to doctors and hospitals (Azoulay, 2002; Polidoro and Theeke, 2012). That close ties with universities enable small biotech firms to signal the quality of the firm to investors has been recognized in the literature (Hicks, 1995; Audretsch and Stephan, 1996). However, there is a dearth of systematic evidence on scientific disclosure for capital market-related reasons. We analyze the revealed preferences for publishing by managers due to an expected shock in the firm's financial information environment. Our results imply that the market is not a sideshow, but rather exerts a powerful effect on scientific disclosure strategies.

We also contribute to the accounting literature on voluntary disclosure. Theoretical models have identified several important frictions that prevent full disclosure, including proprietary costs related to informing rivals (e.g., Verrecchia, 1983; Darrough and Stoughton, 1990). The empirical literature, however, faces challenges in testing the predictions of these models due to a lack of reliable measures (see survey by Leuz and Wysocki, 2016). We consider a disclosure type that is credible and where proprietary costs are significant, namely, firms' trade-offs between publishing their discoveries in scientific outlets and keeping them secret. In terms of the cost, our study is perhaps most closely related to that of Guo, Lev, and Zhou (2004), who examines the impact of various competitive cost proxies on the extent of product-related information disclosed by biotech initial public offerings (IPOs) in their prospectuses. Their analysis implies three main determinants of disclosure: the stage of product development, availability of patent protection, and venture capital backing. This is consistent with our view that upstream research activities (not downstream development) are associated with substantial competitive concerns and that managers may therefore be reluctant to disclose this information.[6]

We also build on the literature that uses broker house mergers or closures as exogenous shocks to study the effects of information asymmetry on corporate policies. Derrien and Kecskés (2013) show that capital and acquisition expenditures as well as R&D decline in response to

[5]See Simeth and Raffo (2013) and Baruffaldi and Poege (2020) for empirical evidence.

[6]Prior work in this context has largely focused on patents as innovation disclosure vehicle, particularly for start-ups (e.g., Brown and Martinsson, 2019; Farre-Mensa et al., 2020; Saidi and Žaldokas, 2020; Glaeser et al., 2020). Nevertheless, as intellectual property rights to exclude others, they predominantly serve as legal safeguards in product markets. Hence, those studies face the challenge that innovation disclosure through patents might have limited proprietary costs. Moreover, limited proprietary costs are reinforced by the legal language of patent documents, enabling inventors to obscure critical details (Hall and Harhoff, 2012).

the shocks, as do financing and cash holdings. Irani and Oesch (2013) document that firms increase their earnings management through discretionary accruals in response to such shocks. Chen, Harford, and Lin (2015) show that after exogenous brokerage exits, managers engage in expropriate behavior to a greater extent. They conclude that analysts are important for governance because they monitor other costly forms of managerial behavior. In the innovation context, He and Tian (2013) show that reductions in coverage cause an increase in patent and patent citation counts. The authors suggest that analysts, by imposing short-term pressure on managers, exacerbate managerial myopia and impede corporate innovation. However, this interpretation has been questioned by Clarke, Dass, and Patel (2015) and Guo, Pérez-Castrillo, and Toldrà-Simats (2019), who find no link with overall patent and patent citations.

Our finding that treated firms engage more strongly in scientific disclosure also relates to the study by Balakrishnan, Billings, Kelly, and Ljungqvist (2014). The authors use the same experimental setting to examine the effects on voluntary disclosure and stock liquidity. They show that managers mitigate the loss of liquidity and public information associated with the reduction in analyst coverage by increasing the levels and timeliness of earnings guidance. We provide insights into their channel by focusing on settings where managers face substantial cost-benefit trade-offs in their disclosure decisions. We also show that despite the potential loss of liquidity due to the loss of coverage, managers at firms with strong internal governance mechanisms in place refrain from making such disclosures when proprietary costs are significant. Hence, such responses could be due to managers who care about reducing information asymmetries for reasons other than liquidity.

Our results on the role of managerial behavior for scientific disclosure are consistent with studies on managerial incentives to shape market perceptions and short-term stock prices. On the theoretical side, Stein (1996) shows that in inefficient markets, a manager who is concerned about the stock price exploits investors' imperfect rationality by catering to time-varying investor sentiment. Empirical studies, such as Baker and Wurgler (2004) and Lou (2014), confirm that corporate financing and investment policies are at least in part motivated by short-term stock price considerations. In addition, past work finds that attention-grabbing events, such as important news, unusual trading volumes, and extreme stock returns, temporarily boost the returns of the stocks by generating more buy orders than sell orders (e.g., Gervais, Kaniel, and Mingelgrin, 2001; Barber and Odean, 2008). We contribute to this body of the literature by providing novel evidence that a firm's choice of disclosure schedules, such as those pertaining to scientific research, is also partially motivated by the desire to increase investors' awareness of that firm and that managers with a short horizon are willing to sacrifice long-term value in order to generate such awareness.

3. Data

3.1. Sample selection

The baseline sample includes information on U.S. public firms for the period between 1997 and 2014.[7] We collect scientific publication information from Elsevier's Scopus database and analyst coverage data from the Institutional Brokers' Estimate System (I/B/E/S) database. To calculate the control variables and the variables used in additional specifications, we add balance sheet and stock price data from the Centre for Research in Security Prices (CRSP)/Compustat Merged Database, data on institutional ownership from Thomson's CDA/Spectrum 13F Holdings database, bid-ask spread estimates from Farshid Abdi's website (http://www.farshidabdi.net/data/), text-based financial constraint measures from Gerard Hoberg's website (http://faculty.marshall.usc.edu/Gerard-Hoberg/), corporate governance data from the IRRC Governance database, and CEO age and compensation data from the Compustat Executive Compensation File. Finally, we obtain patent information from the Duke Innovation & Scientific Enterprises Research Network (DISCERN).[8]

3.2. Variable construction

When constructing the sample of firm-year observations, we restrict our attention to observations with positive assets, sales, and equity in the CRSP/Compustat Merged Database annual file. We eliminate financial and utility firms [firms with Standard Industrial Classification (SIC) codes 4900-4999 and 6000-6999], and firms that are not headquartered in the U.S. based on their current headquarters location. We then merge these data with the remaining databases. As in Arora, Belenzon, and Sheer (2020), we require a firm to have at least one patent and at least one year of positive R&D expense during our sample period (1997-2014). After restricting

[7] Our identification strategy employs broker disappearances between 2000 and 2010, and we examine the three "fiscal" years before and after each event. Following the literature (e.g., He and Tian, 2013), we also construct a 12-month "disappearance period" symmetrically around the identified events to avoid overlaps in years -1 and $+1$. For these reasons, the baseline sample covers the period between 1997 and 2014.

[8] DISCERN builds and improves on the historical National Bureau of Economic Researchpatent database, by extending the dynamic Compustat-USPTO patent data match until 2015 and implementing several methodological improvements. We use Version 5, which is available at https://doi.org/10.5281/zenodo.3976774.

our sample to observations with non-missing information on control variables, we are left with a baseline sample of 20,342 firm-year observations on 2,491 firms in an unbalanced panel. Table 1 defines all the variables used in this study and lists their sources.

3.2.1. Corporate scientific publications

We obtain scientific publication data from Scopus, which contains detailed information on approximately 71 million records from peer-reviewed journals, trade publications, book series, and conference proceedings from 1969 to 2020, and all citations made by these publications since 1970 (over 1.7 billion).[9] Each publication further includes information on the title of the publication, the title of the journal, authors, and an affiliation field with the name and address of the publishing institute or the firm. We focus on "Articles" and "Conference Proceedings" from the list of document types as the most relevant outlets for novel scientific results. To identify scientific publications of firms, we standardize the names in the affiliation field and match these names to all historical company names from our sample of CRSP/Compustat Merged Database firms (for a detailed description on our extensive matching process, see Internet Appendix, Section A.1).

The main corporate publication variable used in this paper is the firm's total number of scientific publications in peer-reviewed journals and conference proceedings in a given year (e.g., as in Simeth and Cincera, 2016; Arora, Belenzon, and Sheer, 2020). Following the literature, we set the publication counts to zero for firms without available publication information in the Scopus database and then use the natural logarithm of the publication counts, PUB, as the main publication measure in our analysis. To avoid losing firm-year observations with zero publications, we add one to the actual values when calculating the natural logarithm. Since we are interested in the decision to disclose scientific research and the average delays from submission to publication in the natural sciences and engineering as well as biomedical research tend to be less than one year (see, e.g., Björk and Solomon, 2013), our preferred specification relates the coverage shock in the current year to scientific publications over the same period.

[9] We prefer Scopus over Web of Science (WoS) database because it includes a more expanded spectrum of journals and, in particular, conference proceedings, which are important final outlets for research in computer sciences and engineering (e.g., Mongeon and Paul-Hus, 2016). However, our results are insensitive to the choice of the data repository for the publication data. As we demonstrate in the robustness tests, our results remain robust to using scientific publication data from WoS.

3.2.2. *Analyst coverage and control variables used in baseline specification*

We obtain analyst information from the I/B/E/S database. We compute analyst coverage, COV, as the average of the number of analysts over 12 months providing current-fiscal-year earnings forecasts; LN_COV is defined as the natural logarithm of (one plus) the average analyst coverage. When selecting control variables, we follow the innovation-related finance literature (e.g., Fang, Tian, and Tice, 2014), which presents a standard vector of firm and industry characteristics. All those variables are computed for firm i over its fiscal year t. The usual variables are firm size, LN_TA, which is the natural logarithm of total assets; $RDTA$, which denotes the R&D expenditures scaled by total assets; LN_AGE, which is the natural logarithm of the number of years the firm is listed on CRSP/Compustat; $PPETA$, which is net property, plant, and equipment scaled by total assets; $CAPEXTA$ or capital expenditures scaled by total assets; profitability, ROA, which is measured by return on assets; leverage, LEV, which is measured by total debt-to-total assets; growth opportunities, Q, which is measured by Tobin's Q; financial constraints, $DELAYCON$, which is sourced from Hoberg and Maksimovic (2015) and captures the risk of delaying a firm's investment due to overall liquidity issues; institutional ownership $INSTOWN$, measured as the fraction of a firm's stock owned by institutional investors; stock illiquidity, $ILLIQ$, the natural logarithm of estimates of bid-ask spreads; product market competition, $HINDEX$, measured by the Herfindahl index based on annual sales; and $HINDEX^2$ or the squared Herfindahl index.

[Table 1 here]

3.3. *Summary statistics*

Table 2 provides the summary statistics for the baseline sample. To minimize the effect of outliers, we winsorize all variables at the top and bottom 1% of each variable's distribution and focus our discussion on medians. The corporate publication output is highly skewed: the mean number of scientific publications is 18, but the median is only 1. The raw measure of analyst coverage has a mean value of 6.569 and a median value of 4.083, which is comparable to previous studies. Regarding other variables, a median firm in our sample has a book value of assets of \$310 million, return on assets of 9.8%, Tobin's Q of 1.7, R&D-to-assets ratio of 6%, property, plant, and equipment scaled by total assets of 14.1%, capital expenditure ratio of 2.9%, and total debt-to-total assets of 10.5%. Fifteen years have elapsed since average firm's inclusion in the CRSP/Compustat Merged Database, and 58.3% are institutional investors.

[Table 2 here]

Table 3 displays the number and fraction of firms with and without scientific publications by industry. In our sample, firms with publications are spread broadly across all industries. Using the Fama–French 12-industry classification obtained from Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), we show that all 10 industries included in our analysis have firms with nonzero publications during our sample period and the fraction of firms with nonzero publications ranges from 44% to 85%.

[Table 3 here]

4. Identification strategy

To account for endogeneity concerns when analyzing the effect of analyst coverage on firms' scientific publication behavior, we adopt broker house mergers and closures that generate exogenous variation in firms' analyst coverages. The validity of the experiment relies on the assumption that broker mergers and closures lead to a loss of analyst coverage for a given firm. We provide evidence for the relevance condition later by showing that our treated firms experience a decrease in analyst coverage of roughly one analyst. The identification assumption that is central to a causal interpretation of our findings is that such coverage terminations are uncorrelated with time-variant unobservable firm characteristics. The most straightforward reason for this assumption to hold is that such termination decisions are not made by the analyst (i.e. they are driven by the broker house discontinuations) and hence are likely to be independent of firm policies and other confounding factors. In addition, authors such as Hong and Kacperczyk (2010), Kelly and Ljungqvist (2012), and Derrien and Kecskés (2013) document that such shocks are plausibly exogenous.

We focus on broker disappearances between 2000 and 2010, and examine the effect of exogenous shocks to analyst coverage for the treated firms over the three fiscal years before $[-3, -1]$ and the three years after $[+1, +3]$ the brokerage merger/closure date.[10] We consider post-Regulation Fair Disclosure (Reg FD) broker disappearances to avoid complications related to

[10]The choice of a three-year window before and after the coverage loss follows the related literature. However, our results are similar irrespective of our using a two-, four-, or five-year window, as can be seen in the Internet Appendix, specifically Table A4.

the implementation of this important regulatory change for analysts (as in Balakrishnan, Billings, Kelly, and Ljungqvist, 2014). We stop with brokers that disappeared in 2010 to avoid truncation problems in some outcome measures during the last years of our sample. Our treatment sample is a combination of firms affected by the brokerage merger/closure events from Kelly and Ljungqvist (2012), who cover the period between 2000 and 2008, and firms affected by events from Fich, Juergens, and Officer (2018), who provide data on brokerage closures and mergers that occurred between 2008 and 2014.

To identify firms whose coverage levels are affected by the merger/closure events, we follow the approach put forth in He and Tian (2013) and Billett, Garfinkel, and Yu (2017), among others. First, for each event, we define the event period as the six months around the broker disappearance month; this accounts for the fact that some mergers span several days or even a couple of months. Next, we retrieve all firms covered by the event brokers in the 12 months before the event period $[-15, -3)$ as well as the analysts working for them. We assume that an analyst covers a firm if there is at least one earnings estimate in the I/B/E/S Detail History file for that firm in the pre-event period. Similarly, we assume that an analyst disappears if there is no earnings estimate by them in the I/B/E/S records in the 12 months after the event period $(+3, +15]$. For brokerage closures, we retain firms for which the analyst disappears from I/B/E/S in the post-event period; using those analysts that do not issue any earnings estimate during that period ensures that analysts who transition to other broker houses do not continue to cover those firms. For broker mergers, we retain firms that are covered by both the acquirer and the target broker before the merger period and for which one of their analysts disappears; this ensures that the resulting loss in coverage is indeed due to the brokerage merger. Further, we drop firms from the sample that are no longer covered by the acquirer in the period after the event; the reason for this restriction is that such terminations could be endogenous because the acquiring broker has chosen to stop covering the firm for reasons that are not observable to us.

For each event, we then continue to construct a control sample of unaffected firms, that is, firms with analysts following in the pre-event period but which are not covered by both merging brokers or the closing broker. Following the other studies that relate broker disappearance to firm-level outcomes, we use a seven-year event window consisting of three years before $[-3, -1]$ and three years after $[+1, +3]$ the merger/closure date. We also consider two additional points. First, because our variables are measured on an annual basis, we must avoid overlaps in the pre- and post-event periods. To do so, we use the last fiscal year that ends in the pre-event period

$[-15, -3)$ as year -1 and the first fiscal year that starts in the post-event period $(+3, +15]$ as year $+1$. Second, it is possible that firm-year observations overlap across events. To address this, we restrict the control sample to firms that are not included in any treatment cohort during the relevant pretreatment or posttreatment time window. This design implies that some firms serve as treatment firms in one period and control firms in another, although never within the seven-year window surrounding treatment.

To implement the DiD identification strategy, one remaining concern is that the treatment and control samples differ in observable dimensions, which could affect the estimate on the coverage loss. Table 4 presents the summary statistics for both the treatment and control samples in the year prior to the broker house merger/closure events. It is apparent that the treated firms are larger and have more coverage that the control firms. This occurs for two reasons. First, treated firms affected by broker house mergers are covered by at least two broker houses in the period before the coverage shock. Second, treated firms receive analyst coverage from larger broker houses, and larger brokers tend to cover larger firms (as discussed in Hong and Kacperczyk, 2010). We account for such differences using two approaches. First, our basic approach is to incorporate firm fixed effects and control variables into the DiD regression framework. Second, we implement a DiD matching estimator.

[Table 4 here]

Our final quasi-natural experiment sample compromises 750 treated firms corresponding to 43 broker disappearances between 2000 and 2010. Of these, 18 result from broker closures and 25 from broker mergers. Figure 2 (Panel A) depicts the distribution of broker disappearances by calendar year. As in previous studies that examine broker mergers and closures (e.g., Derrien and Kecskés, 2013), we observe some clustering of broker disappearances in 2000 and 2001. Panel B depicts the distribution of treated firms by calendar year. It shows that treated firms are even stronger clustered in time: 378 firms are treated between 2000 and 2002, and the other 372, between 2003 and 2010. Moreover, a small number of broker disappearances accounts for a larger number of treated firms. For instance, the shutdown of the U.S. equities operations by Dutch bank ABN Amro accounts for 14% of treated firms, and the top 10 broker disappearances, for 73%.[11] Our DiD approach rules out the possibility that time series effects explain our results.

[11] Perhaps this is a good opportunity to briefly revisit the identification assumption, as Kelly and Ljungqvist (2012, p. 1384) explicitly discuss the case of ABN Amro. Specifically, the authors quote board member, who clarified that the decision to shut down the U.S. equities and investment banking operations was related to

However, in the robustness tests, we provide separate estimates for the small number of broker disappearances that account for a large number of treated firms.

[Figure 2 here]

5. Empirical results

5.1. OLS estimations

We begin our empirical analysis by examining the association between analyst coverage and corporate scientific publications using naïve ordinary least squares (OLS) regression models. Specifically, we estimate

$$PUB_{i,t} = \alpha + \beta LN\_COV_{i,t} + \gamma Z_{i,t} + \delta_t + \lambda_i + \varepsilon_{i,t} \tag{1}$$

where i indexes firms and t indexes time. The dependent variable, PUB, is the natural logarithm of (one plus) the total number of scientific publications for firm i. Since the publishing process in most scientific disciplines takes less than one year, and in light of our interest to measure disclosure, we examine the effect of a firm's analyst coverage on its scientific publications in the current year. The analyst coverage measure, LN_COV, is measured for firm i over its fiscal year t as the logarithmic transformation of analyst coverage. The vector $Z_{i,t}$ contains firm and industry characteristics that could affect a firm's publication behavior. δ_t and λ_i correspond to year and firm fixed effects, respectively; standard errors are robust to heteroskedasticity and are clustered at the firm level.

Table 5 presents the results. The first three columns regress the number of scientific publications only on the coverage variable. We start without any fixed effects. The next two columns gradually add firm and year fixed effects. The final three columns gradually add the set of control variables. Under the hypothesis that asymmetric information encourages managers to improve disclosure about firm's research activities (because they are concerned that the increase in the cost of capital increases the costs of doing research), we expect the coefficient estimate on LN_COV to be negative (because the number of analysts following is negatively correlated

competitive pressure and strategic reasons rather than characteristics of the stocks their analysts covered: "We're withdrawing from businesses in which we're strategically ill-positioned and cannot create a sustainable profit stream, whether the market turns around or not."

with information asymmetry).

However, the coefficient on LN_COV is positive and significant across all the columns of Table 5; hence, higher analyst coverage is associated with more scientific publications. This is inconsistent with our baseline hypothesis and could occur due the endogenous relationship between analyst coverage and scientific disclosure. For instance, it is possible that larger increases in scientific publications are only seen among larger firms because those firms are less concerned about proprietary costs (e.g., Graham, Harvey, and Rajgopal, 2005). At the same time, it is well established that those firms are followed by more analysts. Alternatively, perhaps voluntary disclosure and analyst coverage are complements, as Lang and Lundholm (1996) propose, and this also extends to scientific disclosure. These panel regressions are difficult to interpret due to the well-known difficulties in controlling for endogeneity when estimating the relationship between asymmetric information in capital markets and corporate policies

[Table 5 here]

5.2. DiD estimations: Baseline

In the previous section, we show a positive correlation between analyst coverage and scientific publications, even after controlling for other factors that could drive firm publication behavior, and fixed effects. To examine how firms react to a plausibly exogenous change in their information environment, we now turn to the DiD approach. As mentioned above, Kelly and Ljungqvist (2012) and others demonstrate that broker mergers/closures are plausibly exogenous, and thus, as long as we account for observable differences between treated and control firms, it is unlikely that variations in firms' characteristics drive either the coverage shock or our treatment outcome. To account for firm-level differences, our basic approach is to incorporate control variables into the DiD regression framework. Specifically, we estimate

$$\begin{aligned} PUB_{i,e,t} = \alpha + \beta_1 Post_{e,t} + \beta_2 Treated_{i,e} + \beta_3 Post_{e,t} \times Treated_{i,e} \\ + \gamma Z_{i,t} + \delta_t + \lambda_i + \theta_e + \varepsilon_{i,t} \end{aligned} \tag{2}$$

where i, e, and t indexes the firm, event, and time, respectively. Our main dependent variable remains PUB: the natural logarithm of (one plus) firm i's total number of scientific publications in a given year. The variable $Post_{e,t}$ is a dummy that equals one if a firm-year observation is in the post-event period for event e and zero otherwise. $Treated_{i,e}$ is a dummy that equals one if firm i is part of the treatment sample for event e and zero if it is part of the control sample. The coefficient of interest is β_3, which corresponds to the DiD effect, namely, the impact of the loss of an analyst due to a broker merger/closure on changes in scientific publications of the treated firms relative to the control firms. The variables δ_t, λ_i, and θ_e correspond to year, firm, and event fixed effects, respectively; as before, we cluster standard errors at the firm level.[12] This specification allows us to easily incorporate control variables that account for other potential sources of systematic differences across treated and control firms that are not picked up by fixed effects (i.e., time-varying firm- and industry-specific determinants). To this end, $Z_{i,t}$ is the set of control variables listed above.

We first evaluate whether our DiD approach meets the key idea of the experiment: on average, treated firms should lose about one analyst relative to control firms in the period after a broker house event. We test this by estimating Eq. (2) with the dependent variable replaced by analyst coverage, and without including fixed effects or control variables. Given our experimental setup, we should observe a DiD coefficient of roughly minus one. Indeed, the first column of Table 6 suggests that this is the case: the DiD coefficient is -1.334 and significant at the 1% level. In terms of magnitude and significance, this outcome is consistent with the other studies that use a similar research design (e.g., Hong and Kacperczyk, 2010; Derrien and Kecskés, 2013; Guo, Pérez-Castrillo, and Toldrà-Simats, 2019).

We also examine the evolution of analyst coverage during the three years before and after the coverage stock. Panel A of Figure 3 presents the results. The median difference in coverage between the treatment sample and the control group is relatively stable before the brokerage house disappearance $[-3, -1]$ and decreases by about one analyst between years -1 and $+1$. Given this drop in analyst coverage for our treatment sample, we also assess whether we observe increased asymmetric information in our data. The typical proxy used in the related literature is analysts' forecast dispersion. We report the results in column 2 of Table 6. The estimated DiD coefficient in the forecast dispersion equation is 0.010 and statistically significant. Hence, the loss of coverage causes asymmetric information to increase.

Next, we examine how this increase in asymmetric information translates into the scientific publication behavior of the firms. The remaining columns of Table 6 present the results. We

[12] In the section of robustness tests, we experiment with clustering standard errors at the event or the firm-event level. However, clustering at the firm level tends to produce the largest, and thus most conservative, standard errors. Thus, we report these throughout the paper. We also show that our results remain robust when we collapse the time series information into two effective periods (before and after the event) which explicitly takes into account the effective sample size.

begin with estimating Eq. (2) without any fixed effects or control variables. The next three regressions gradually add firm, year, and, finally, event fixed effects. The last three columns includes these as well as we gradually add the set of control variables. Across all the relevant columns of Table 6, we observe that the DiD coefficient is positive (ranging between 0.092 and 0.133) and both economically and statistically significant. For example, the point estimate of 0.115 in column 8 implies that an exogenous loss of one analyst following a firm causes it to increase scientific publications by about 12% over a three-year period compared to firms without any decrease in analyst coverage. Our interpretation of this finding is that firms react to an increase in asymmetric information by increasing scientific disclosure in an attempt to actively signal their quality. This rationale is consistent with economic models of financial signaling through the disclosure of proprietary information (e.g., Bhattacharya and Ritter, 1983). In such models, when managers face informational asymmetries between the firm and capital markets, they will have a greater incentive to communicate private information to the capital market, but the only way to do so is through the disclosure of information that is also useful to rival firms.

The step-wise inclusion of fixed effects and controls merits some discussion. Column 3 presents a regression without any fixed effects or control variables. The DiD coefficient is positive with a marginal effect of 0.092. In column 4, we include firm fixed effects. Conditioning on firm fixed effects slightly increases the DiD coefficient (from 0.092 to 0.133), suggesting that, if anything, time-invariant differences across firms bias the DiD estimate downward. Additionally including year fixed effects and event fixed effects (columns 5 and 6) does not change this picture. This provides reassurance that the estimated impact of the coverage shock on scientific publications is not driven by systematic differences in disclosure behavior across firms, years, or events.

The last three columns gradually add the standard set of time-varying firm and industry characteristics employed by existing studies on firm innovation. Omitting these characteristics from the regressions could lead to biases as well. However, the results shown in Table 6 indicate that this is unlikely, as the DiD coefficients change only marginally. Thus, the inclusion of additional control variables has a limited impact on the estimated treatment effect, suggesting that our coverage shock is plausibly exogenous and that the increase in scientific publications is not the result of omitted variable bias.

[Table 6 here]

We also inspect the evolution of the scientific publication behavior of the firm surrounding a decrease in analyst coverage. Panel B of Figure 3 presents the differences in corporate publications between the treatment and control firms in the three years before and after the loss of an analyst. As in the formal regression-based tests discussed above, we observe a similar pattern of a causal increase in scientific publications due to asymmetric information. The median difference in scientific publications is constant over the years leading up to the coverage shock $[-3, -1]$, following which the line starts to increase between years -1 and $+1$. Hence, scientific publications react immediately to coverage terminations. The fact that firms can respond without delay supports the interpretation that firms disclose knowledge strategically. The graph also supports the parallel trends assumptions underlying our DiD approach. Moreover, formal statistical tests confirm this pattern. We report these complementary tests in the Internet Appendix, specifically column 1 of Table A4. The parallel trend assumption requires similar trends in scientific publication variables during the pre-event period for both the treatment and control samples, whereas it does not imply that scientific publication levels be the same across the treatment and control samples during the pre- and post-event regimes, because these differences are cancelled out in the estimation process. As can been seen, it does not appear that changes in corporate scientific publications are part of long-term trends before analyst coverage decreases.

[Figure 3 here]

5.3. DiD estimations: Additional scientific publication measures

We now turn to several additional publications metrics to analyze whether the shocks to information asymmetries also affect the composition of scientific publications. Conceptually, if managers' scientific disclosure responses to coverage terminations are motivated by an attempt to influence market perceptions, we should observe that our results also extend to the quality of scientific publications. The fine-grained vertical differentiation of journals quality (by ranking and prestige), in particular, constitutes a rather transparent template to discern publications by their quality. An advantage of impact-factors is that it is not an ex-post indicator of quality such as citations. Publications in top journals likely constitute stronger signals and means of disclosure, relative to alternative documents, such as other publications or patents. Extant literature shows that firms are indeed sensitive to the reputation effects of their publications, and tend to favour publications of better quality (Hicks, 1995; Baruffaldi and Poege, 2020).

To assess the relevance of quality, we create five different variables. For the first variable, we follow common practice by weighting the number of publications in a given year by the Journal Impact Factor (JIF) and, also, by the number of citations received in subsequent years. Alternatively, we rank all the academic journals by the JIF, select the bottom and top 50% in each year, and count the number of publications in those journals per firm and year. The final dimension of interest involves linkages with university-based scientists. Audretsch and Stephan (1996) argue that such links lend more prestige to a firm, and Stephan (1994) shows that capital markets do recognize such prestige. We measure it by the fraction of the number of publications that are co-authored with university-based scientists to the total number of publications in a given year.[13]

The results from estimating Eq. (2) on this set of dependent variables are presented in Table 7. Column 1 shows the regression with cite-weighted publications, and column 2 reports the results with regard to JIF-weighted publications. The DiD coefficient for cite-weighted publications is 0.186 and that for JIF-weighted publications is 0.223. Both remain significant at the 1% level. Hence, for the average treated firm, we see an increase in publications between 19% and 22% when quality is taken into account, which is larger than the 12% obtained in our baseline estimation. We observe a similar pattern when separately counting the raw publications in the bottom and top 50% of journals ranked by the JIF each year. In column 3, where JIF is high, the DiD coefficient is large, whereas in column 4, where JIF is low, the DiD coefficient is smaller (0.173 versus 0.032). Finally, in column 5, we use the ratio of publications with academic institutions as the dependent variable. Consistent with the reputational implications of having ties with university-based scientists, we obtain a positive and statistically significant DiD coefficient. Overall, we observe stronger treatment effects when accounting for publication quality and/or affiliations with universities. This finding is consistent with our interpretation, as one would expect that the capital market is particularly responsive to scientific disclosure when the scientific value of the published output is high.

[Table 7 here]

5.4. DiD estimations: Robustness tests

Before analyzing in greater detail the underlying economic mechanisms, we first conduct a number of tests to examine the validity of our quasi-experiment and the robustness of the

[13] Note that survey evidence by Blumenthal, Causino, Campbell, and Louis (1996) documents that most agreements with universities include specific clauses about what and when university-based scientists may or may not publish.

baseline treatment effect. We first show that our results remain similar when using a matching estimator (Section 5.4.1). Next, we investigate the validity of our research design by testing for differences across specific broker house events (Section 5.4.2). Finally, we consider alternative pre- and post-period contrasts, and perform several additional tests to show that our baseline results are robust in terms of magnitudes and statistical significance (Section 5.4.3).

5.4.1. DiD Matching estimator

The identifying assumption of our quasi-natural experiments is that changes in the scientific publications behavior of treated firms across the merger/closure events are not due to any factor aside from the merge/closure leading to a drop in analyst coverage. Our basic approach to account for differences in the characteristics between treated and control firms is to incorporate control variables into the DiD specification. The evidence from Table 6 seems consistent with this assumption, as both the magnitude and the statistical significance of the estimated treatment effects change only slightly after including the vector of firm characteristics. To further augment the confidence in our findings, we now examine the robustness of our results when we enforce common support and covariate balance in the pre-event period using coarsened exact matching (CEM) (Iacus, King, and Porro, 2012).

Our choice of matching variables follows other studies that employ broker merger/closure events as quasi-natural experiments (e.g., Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012; Derrien and Kecskés, 2013). Specifically, we match treated and control firms based on the joint distribution of firm size, growth opportunities, analyst coverage, past returns, two-digit SIC industry code dummies, and event dummies. We also include a publication growth variable (i.e., the growth in the number of publications from years -3 to -1), as this parameter has been used in related innovation studies (e.g., He and Tian, 2013). In an attempt to achieve the highest comparability before the treatment, we take the values of those variables in the year before each broker house merger/closure (year -1) as matching criteria. To eliminate any statistically significant differences between the treated and control firms, we conduct CEM with the condition to differentiate firms according to 10 categories in analyst coverage, 5 categories in firm size, and 3 categories each in past returns, growth opportunities, and publication growth.

Table 8 presents the results. The first five columns display the summary statistics for the treated and control firms after the matching procedure. The summary statistics indicate that, for the variables we match on, the differences between both groups are small in terms of economic magnitudes and statistically insignificant. The last column of Table 8 presents the treatment effects of the coverage shock using the CEM-balanced sample: the DiD coefficient remains significant with a marginal effect very similar to that in Table 6. This is another piece of evidence suggesting that our baseline DiD estimate is unlikely to be explained by heterogeneity between the treatment and control firms. Specifically, we show that observable differences that could explain selection into the treatment, such as firm size, analyst coverage, growth opportunities, or past stock return do not drive our results. Our estimates remain strikingly similar, which provides further evidence that our natural experiments are indeed exogenous.[14]

[Table 8 here]

5.4.2. Merger/closure characteristics

Next, we show that our baseline results are not driven by broker disappearances in specific years and by the small number of mergers/closures that cause a large number of treated firms. This is important because Figure 2 shows that treated firms are strongly clustered in time and a substantial number of broker disappearances overlaps with the dot-com bubble as well as the financial crisis of 2008. We also confirm that our results are not driven by either broker mergers or broker closures alone, and our results remain robust to the exclusion of broker disappearances that are not included in the list provided by Kelly and Ljungqvist (2012). To this end, we perform four different analyses.

We tabulate the results in the Internet Appendix, specifically Table A1. First, we estimate Eq. (2) only for the group of brokers that disappeared in 2000, 2001, 2002, 2008, and 2009, contrasting these results with the group of brokers that disappeared in the other years. The DiD coefficient remains positive and significant in all sub-periods. Second, we repeat our analysis separately for the group of top 10 brokers ranked by the number of firms that lose an analyst, which collectively account for 73% of the treated firms. The DiD estimate is positive and significant in both subsamples, suggesting that our full sample is not driven by a small number of brokers that account for a large number of coverage terminations. Third, we repeat our analysis for mergers and closures separately, and find that our results are similar across both groups. In the fourth test, we drop broker disappearances that are not included in the list in Kelly and Ljungqvist (2012). We find even stronger results: firms affected by brokerage merger/closure

[14]We obtain results almost identical to the ones shown in Table 8 if we match by publication levels instead of publication growth rates and/or include stock return volatility as an additional matching criterion.

events from this list experience a 14% larger increase on average in scientific publications in the three years after the coverage shock than firms without any decrease in coverage.

5.4.3. Additional robustness tests

We conduct a rich set of additional tests to verify the robustness of our baseline results in terms of magnitudes and statistical significance. The results are tabulated in the Internet Appendix (see Tables A2-A4). First, we rerun our baseline DiD specification on a balanced panel to check for attrition bias. This attempt delivers results similar to those of the unbalanced panel. Our results are also robust to including industry-by-year fixed effects, which allow us to control for time-varying opportunities across different industries. We collect firm-year scientific publication information from Scopus. As discussed elsewhere, Scopus offers a larger coverage in natural science and engineering as well as biomedical research than WoS. However, Scopus may suffer from other limitations, as discussed in Mongeon and Paul-Hus (2016). Therefore, we replicate our DiD tests using data from WoS and show that our results remain robust. We test whether our results suffer from serial correlation issues. To this end, we experiment with clustering standard errors either at the event level or the event-firm level. However, using different clustering schemes makes little difference. If anything, by clustering standard errors at the firm level throughout the paper, we overstate the standard deviation of the DiD estimator. We also repeat our DiD analysis by collapsing the firm-year observations by broker into a pre and post period. This experiment also yields very similar results.

Second, our baseline specification uses a three-year window before and after the broker merger/closure with a 12-month disappearance period, defined as the 6 months symmetrically around the event date. We do so to account for the fact than many broker mergers and closures span a longer time, which makes it difficult to recognize the specific event date. To explore the sensitivity of our results to this choice, we first move the pretreatment interval further backward from the event year by either one or two years, while keeping constant the posttreatment window. The DiD estimate in both cases remains robust, with a slightly larger economic impact compared to the baseline. This suggests that our findings are unlikely to be driven by timing effects in the years leading up to broker mergers and closures. We also move the posttreatment window further outward by either one and two years, while keeping constant the pretreatment window. Varying the posttreatment window as well produces remarkably stable coefficient estimates. Finally, we explore the sensitivity of the publication results to the selection of the three-year measurement window $[-3, +3]$. We show that our results remain robust if we use a two-year $[-2, +2]$, four-year $[-4, +4]$ or five-year $[-5, +5]$ window.

5.5. *Heterogeneity analysis*

In this section, we explore heterogeneity in the treatment effect by examining relevant subsamples. Our baseline DiD results suggest that managers consider scientific disclosure as a means to mitigate the adverse consequences of an exogenous increase in information asymmetry. While this line of thought is consistent with existing studies on financial reporting, in that increased disclosure lowers the cost of capital and increases firm value (e.g., Irani and Oesch, 2013; Balakrishnan, Billings, Kelly, and Ljungqvist, 2014), we stand out from other studies in that we consider a case where proprietary costs are significant. If this is the case, the treatment effect should depend on the costs associated with the loss of coverage. Specifically, it should only affect scientific publication behavior for firms that suffer to a greater extent from exogenous shocks to their analyst coverage. Put differently, if scientific disclosure is costly, there are few reasons to believe that managers are willing to voluntarily sacrifice long-term firm value if the benefits of disclosure are small relative to the size of the coverage shock they suffer.

5.5.1. *Conditioning on financial constraints*

We consider two dimensions that have been identified in the related literature as key determinants of shaping the costs associated with the loss of coverage. First, we investigate if the disclosure behavior of firms experiencing a coverage shock varies with their access to external financing. Kelly and Ljungqvist (2012) suggest that decreases in analyst coverage (due to broker merger/closure) increases information asymmetry as well as the cost of capital. Specifically, the authors show that coverage shocks have first-order adverse effects on liquidity and firm value. Derrien and Kecskés (2013) show that this leads to a reduction in external financing and investment, especially for firms that are financially constrained. For financially unconstrained firms, the decrease in analyst coverage is largely irrelevant, as they have sufficient internal capital to finance their operations. Therefore, firms that are financially constrained suffer to a greater extent from an exogenous reduction in analyst coverage.

To test this hypothesis, we create subsamples of firms based on whether they are financially constrained in the year prior to the event. The extant literature offers numerous proxies for financial constraints. We start with the text-based measures developed by Hoberg and Maksimovic (2015). Specifically, we employ the delaycon and equitydelaycon measures. We consider unconstrained (constrained) firms to represent those with scores below (above) the sample median. Billett, Garfinkel, and Yu (2017) use two indirect proxies to classify firms as unconstrained: having a credit rating or paying dividends. We use both. The last proxy we consider is firm age. Younger firms tend to be characterized by a higher degree of informational asymmetry and high growth opportunities. Older and more established firms tend to have a lower degree of informational asymmetry with outside investors and fewer growth opportunities. We classify firms in the bottom (top) of the age distribution as constrained (unconstrained).

Table 9 presents the DiD results for the five subsamples. For all the financial constraints measures, we find consistent evidence that the increase in scientific publications is concentrated among firms that are financially constrained. In columns 1 and 2, we begin with the delaycon measure. As column 1 shows, the estimated marginal effect of the coverage shock on scientific publications for firms that are financially unconstrained is small in magnitude (0.047) and statistically indistinguishable from zero. On the other hand, the estimated treatment effect is positive and significant in the case of financially constrained firms. The DiD coefficient is also larger in magnitude (0.177) than the corresponding average treatment effect for the full sample (see Table 6, specifically column 9). In columns 3 and 4, we observe a very similar pattern when we use the equitydelaycon measure, which additionally captures firms that indicate plans to issue equity to address their liquidity challenges: financially unconstrained firms show an insignificant treatment effect relative to control firms, while constraint firms show a strong treatment effect of roughly 22%. Columns 5-10 report results for the more conventional constraints measures. Consistent with our findings so far, the estimated treatment effect on corporate publications is positive and statistically significant, and only present for treated firms without credit ratings, non-dividend payers, and younger firms.

Overall, these results show that the coverage loss does not lead to an adjustment in scientific publications among the financially unconstrained subset of the treated firms. This finding is consistent with the view that scientific disclosure is costly. By revealed preference, managers that do not increase scientific publication levels view the costs of scientific disclosure as greater than the benefits. The fact that only those firms that suffer more from the shock increase scientific publications indicates that the costs of scientific disclosure are substantial. The literature on disclosure corroborates this interpretation. Specifically, survey evidence by Graham, Harvey, and Rajgopal (2005) shows that a substantial fraction of managers views proprietary costs as a significant barrier to more disclosure because they "do not want to explicitly reveal sensitive proprietary information on a platter to competitors" (p. 62). Simultaneously, managers of young companies consider increased disclosure as a more important instrument to correct for information problems than managers of older companies. Interestingly, however, the motivation for these managers is not so much a desire to relax financing constraints than to reduce the risk of job loss. This naturally leads to our second cross-sectional analysis.

[Table 9 here]

5.5.2. Conditioning on managerial incentives

In this section, we aim to understand the extent to which changes in firm disclosure behavior in the treatment group are motivated by managerial incentives. According to the theories of Narayanan (1985), Stein (1988), and Bebchuk and Stole (1993), asymmetric information and managers' concern for their short-term reputation or stock prices can lead to decisions that yield short-term gains at the expense of the long-term interest of the shareholders. The most basic empirical implication of those models is that the magnitude of distortions will depend on the extent to which managers are concerned about impressing the market. Under the view that scientific disclosure has proprietary costs, there is no strong motivation for managers to increase publication levels to ensure that their stock is never undervalued. Stein (1988) suggests that signaling behavior becomes important when there is a great exposure to hostile takeovers for the firm. Since increased information asymmetry leads to lower share prices, the exogenous loss of analyst coverage can increase the probability of a hostile takeover attempt, and maximizing this price then becomes a pressing concern for managers. Consequently, we might reasonably expect to see that managers who are concerned about their careers are more willing to incur disclosure costs in an attempt to signal to the market that the long-run outlook of the firm is good, thereby increasing its current valuation.

To assess the role of managerial incentives in our settings, we follow the governance literature and condition upon several proxies for managerial entrenchment, CEO age, and CEO equity-based compensation. The first proxy for managerial entrenchment is the governance index introduced by Gompers, Ishii, and Metrick (2003), which is built upon 24 anti-takeover provisions. As argued in the literature, anti-takeover provisions insulate (or entrench) management from the disciplining effects of shareholders and the market for corporate control. The second proxy is the entrenchment index proposed by Bebchuk, Cohen, and Ferrell (2009). The authors argue that 6 of the 24 provisions are the most effective to entrench incumbent managers and are not influenced by the noise produced by the other provisions. Next, we consider whether the CEO is also the chairman of the board of directors. One function of the chairman of the board is to oversee the process of hiring, compensating, and firing the CEO. Hence, when the chairman of the board is the CEO herself, the CEO is more entrenched. We also examine another dimension: the chief executive officer's (CEO's) age. Prendergast and Stole (1996) predict that younger managers who wish to establish a reputation for quickly learning take greater risks to signal superior ability. In particular, younger managers overweight their personal beliefs and exaggerate their behavior to appear talented. Differences in how much managerial wealth is tied to the value of a firm's assets is another test. We measure it using the change in the sensitivity of pay to the stock price measure of Core and Guay (2002).

Table 10 presents the DiD results obtained from the split sample regressions. Columns 1 through 6 report the results for low and high managerial entrenchment; columns 7 and 8, for the young and the old CEOs; and columns 9 and 10, for low and high pay-performance sensitivity. Across all the columns of Table 10, we obtain evidence that managerial incentives matter for firms' disclosure responses to coverage terminations. In columns 1 and 2, we classify firm-year observations based on the median governance index in the year prior to the event. We see that the point estimate of the treatment effect is larger in magnitude (0.163) and statistically significant when CEOs are less entrenched. Moreover, the estimated DiD coefficient is larger than the average treatment effect reported in Table 6. In contrast, when managers are insulated against the reputational risk from a hit to their market value (due to broker merger/closure), the estimated treatment effect is small (0.043) and statistically insignificant. In columns 3 and 4, we partition the sample based on the median value of the entrenchment index in the year prior to the relevant event. We find results almost identical to those of the governance index, with the statistical significance of the DiD coefficients being of the same order of magnitude. In columns 4 and 5, we consider CEO duality as an alternative proxy for managerial entrenchment. We find qualitatively similar results, as the partial effect of increased information asymmetry on scientific disclosure is only present among firms where the CEO does not play the role of the chairman.

In columns 7 and 8, we examine whether firms managed by younger CEOs behave differently to the loss of coverage than firms run by older CEOs. The ages of the young (old) CEOs are below (above) the median. We find that the cross-sectional effect is concentrated among firms run by young CEOs. For this group, the estimated DiD coefficient is positive and statistically significant. Other things being equal, firms with CEOs aged between 39 and 56 years show treatment effects of roughly 16%. In contrast, firms with CEOs aged 57 years and above show treatment effects that are close to zero. A potential concern with the previous result is that older CEOs tend to have higher pay-performance sensitivity in their compensation contracts to account for fewer career concerns, as documented by Gibbons and Murphy (1992). At the same time, a greater pay-performance sensitivity may induce risk aversion and deter activities that are costly in the long run. Consistent with the latter, Edmans, Gabaix, Sadzik, and Sannikov (2012) show that increasing pay-performance sensitivity offsets short-termism/myopia by ensuring that the CEO loses more dollars in the future than she gains today by inflating current stock prices. The final two columns of Table 10 report the results of separating our sample into firms managed by CEOs with low and high pay-performance sensitivity based on the sample median prior to the event. Our results suggest that compensation incentives also matter. We observe that the treatment group with low pay-performance sensitivity increases the number of scientific publications following the coverage shock. We do not observe any such behavior among firms with high sensitivity of pay to stock price. Hence, such managers view the competitive costs of committing to disclosure as outweighing the capital market benefits.[15]

Overall, these findings indicate that managerial career concerns and compensation incentives are important empirical determinants that link coverage shocks and firm scientific disclosure behavior. We find that disclosure responses appear only in settings where managers are less entrenched and subject to more pressure imposed by outside shareholders. We also find that younger managers and those with compensation incentives less closely tied to long-term firm value show strong responses. In contrast, we observe no adjustment behavior in settings where the costs of the coverage shocks to the manager are lower or the disclosure costs are substantial. All of the above findings appear consistent with the implications of a signaling interpretation.

Of course, other factors could also shape the costs associated with the loss of coverage and the cost-benefit analysis from scientific disclosure. However, those tend to generate rather ambiguous predictions. For example, Derrien and Kecskés (2013) suggest that smaller firms have more information asymmetry than larger firms, and hence suffer from the disappearance of an analyst.

[15] Our findings remain unaltered if we replace the sensitivity of pay to the stock price measure of Core and Guay (2002) with the scaled wealth-performance sensitivity (WPS) measure of Edmans, Gabaix, and Landier (2009). In a specification identical to columns 9 and 10 of Table 10, when WPS is low, the DiD coefficient is large positive, positive, and significant (we obtain a coefficient of 0.124 with a standard error of 0.054), whereas when WPS is high, the coefficient is smaller and insignificant (we arrive at a coefficient of 0.060 with a standard error of 0.049).

However, Graham, Harvey, and Rajgopal (2005) show that it is precisely these firms that are much more restrictive when disclosure involves sensitive information that might harm the firm's competitive position. Another example is product market competition. While Chen, Harford, and Lin (2015) find that the loss of an analyst matters more in less competitive industries, revealing company secrets to competitors is probably less of an issue for firms operating in such industries. Understanding where and how these differences operate provides a valuable direction for future theoretical and empirical research.

[Table 10 here]

5.6. Discussion of alternative perspectives

In principle, exogenous reductions in analyst coverage due to broker mergers and closures may increase scientific publications through mechanisms other than a rise in information asymmetry. In this section, we discuss two alternative perspectives on the role of analysts in capital markets (other than producing information that matters to investors): the role of analysts as effective monitors (Section 5.6.1) and their role in pressuring managers to meet short-term performance benchmarks (Section 5.6.2). We also provide additional evidence in support of our suggested mechanism by showing that coverage terminations have no impact on patenting activities or the use of scientific knowledge in firms' downstream inventions. We conclude that the information view best describes the relationship between coverage terminations and increased scientific disclosure.

5.6.1. Analyst coverage and external monitoring

A large strand of the literature reports on the monitoring role of analysts in mitigating managerial extraction of private benefits from outside shareholders. Under this view, a drop in analyst coverage implies that managers face lower detection probabilities and are therefore more likely to engage in value-destroying activities for private benefits. Chen, Harford, and Lin (2015), for example, show that when analyst coverage decreases due to exogenous broker closures and mergers, shareholders value their internal cash holding less, CEOs receive higher excess compensation, and they are more likely to engage in value-destroying acquisitions. This explanation requires that managers, unless closely monitored, will always act in self-serving and opportunistic ways. In contrast, under the information setting, such behavior does not occur unless the manager perceives that their interests are clearly and directly at stake.

To evaluate the monitoring role of analyst coverage, we can revisit the findings from Table 10. As Irani and Oesch (2013) show, a governance complementarity exists between analyst coverage and takeover pressure in reducing managerial misbehavior. Hence, the key implication of this hypothesis is that the loss of an analyst should be felt most sharply when managers are more entrenched. However, we observe the opposite in our data: contrary to the monitoring story (and consistent with the information channel), the positive effect is stronger when managers are less entrenched. Thus, the monitoring role of analysts does not appear to be a mechanism that links coverage terminations and scientific publications, at least in our settings. Despite this, the literature also presents considerable evidence that challenges the monitoring view on an a prior basis. For instance, Dyck, Morse, and Zingales (2010) suggest that analysts' monitoring incentives may be compromised due to an investment banking relationship between the firm and the analysts' investment house. Furthermore, the results of the survey by Brown, Call, Clement, and Sharp (2015) show the analysts themselves admitting to the fact that they do not focus much on detecting corporate fraud or intentional misreporting.

5.6.2. Analyst coverage and long-term investments

A drop in analyst coverage could result in an increase in scientific publications if analyst coverage has a "dark side" in that analysts ignore firms' innovation outputs when making stock recommendations, as suggested by He and Tian (2013). Under this view, short-term oriented analysts exert pressure on managers to meet short-term goals, and in response to such pressure, managers boost current earnings by sacrificing long-term investments in innovation. If this effect is common, then the increase in scientific publications could reflect the decision to increase investments in innovation following an exogenous reduction in coverage.

While the dark side of analyst coverage is intuitively appealing and has received significant attention in the academic literature, the evidence is mixed. For example, Clarke, Dass, and Patel (2015) show that the negative relationship between analyst coverage and innovation obtained by He and Tian (2013) is driven by the subsample of firms with low innovation efficiency. They conclude that analysts are discerning in their view toward long-term investments by firms and do not discourage innovation across the board. Similarly, Guo, Pérez-Castrillo, and Toldrà-Simats (2019) document that analyst coverage has no effect on overall innovation outputs when taking into account firm's innovation strategy. We also refer to evidence from analyst reports on this matter. Specifically, Bellstam, Bhagat, and Cookson (2020) propose a new measure of corporate innovation derived from textual descriptions of firm activities by analysts and find that this measure correlates strongly with patenting and R&D intensity among patenting firms. In contrast to analysts having a dark side, this suggests that analysts provide rather valuable information on firm innovation to investors.[16]

The key implication of this hypothesis is that long-term investments should increase following coverage terminations. Hence, one way to assess this assertion is to check whether we observe changes in long-term investment inputs or other innovation outcome measures. We report the results from this exercise in Table 11. Column 1 reports the regression results from estimating Eq. (2) with the dependent variable replaced by innovation inputs (R&D expenditure scaled by total assets) and shows an insignificant DiD estimator.[17] In the next two columns, the dependent variable is replaced by two established measures of innovation outputs: the number of patent applications a firm files in a given year that are eventually granted (column 2) and the number of citations these patents received in subsequent years (column 3). In both cases, the treatment coefficient is small and insignificant. Thus, it is not possible for us to assert that analyst coverage has a dark side based on standard measures of innovation investments.

A potential concern in our study is whether the effect of a decrease in analyst coverage on scientific publications changes in response to composition of patenting activities. Thus, instead we consider the use of scientific research in patents. Following Marx and Fuegi (2020), we consider patents as science-based if they contain at least one citation to science on their front page. As these files also include the DOI publication identifiers, we can match them against the focal firms' publication stocks. We calculate two variables: the fraction of science-based patents (column 4) and the fraction of science-based patents with external references (column 5). In both cases, we observe a pattern for the treatment effect that is very similar to the previous results: the estimated DiD coefficient is small and statistically insignificant. We interpret these tests as indicating that the increase in scientific publications (due to the analyst loss) does not lead to an increase in the use of scientific knowledge in firms' downstream inventions, and therefore in the composition of firm's R&D investments. Overall, this further corroborates the

[16]As an aside, why analysts would be willing to cover innovative firms in the first place if they do not have sufficient interest in innovation or lack of scientific knowledge (therefore ignoring innovation) also remains an open question. For instance, the survey by Brown, Call, Clement, and Sharp (2015) finds that analysts consider industry knowledge as the single most important input to their performance and career outcomes. Swarup (2008) describes the typical career transitions for analysts in technology-intensive industries; they tend to move into finance at a late stage in their careers and rely heavily on their scientific knowledge. Thus, the dark side hypothesis also appears to be ambiguous conceptually when taking into account analysts' incentives.

[17]If we use the natural logarithm of (one plus) R&D expenditure as a dependent variable instead of R&D intensity, the DiD coefficient remains insignificant (a coeff. of 0.044 with a standard error of 0.029).

perspective that firms do not alter their long-term investment strategies following the loss of a short-term-oriented analyst, but rather change their scientific disclosure practices to quickly counteract the consequences of an increase in information asymmetry.

[Table 11 here]

6. Conclusions

In this paper we examine how firm's financial information environment affects its scientific disclosure policies. We conduct a DiD analysis using exogenous shocks on analyst coverage caused by broker house mergers and closure events. A reduction in analyst coverage leads to a prompt and persistent increase in scientific publication outputs, suggesting that asymmetric information in capital markets impacts the incentives to disclose research outcomes. Attempts to attract investors in order to offset the consequences of an increase in information asymmetry appear to be a plausible underlying economic mechanism. Cross-sectional analysis further reveals that our causal effects are concentrated in the following subsamples: firms with greater ex ante growth opportunities, firms with greater ex ante financial constraints, firms run by managers with ex ante greater career concerns, and firms run by managers with ex ante less personal wealth tied to firm performance. This evidence is in line with the notion that innovative firms only choose to voluntary disclose scientific knowledge when the perceived benefits are large enough. We also find that treated firms do not invest more in innovation compared to their peers; nor do we see significant increases in patenting or the use of science in firms' downstream inventions.

The possibility that firm's information environment guides managers in their scientific disclosure decisions has important implications. There is widespread agreement among policymakers to create an environment with more abundant public information, allowing investors to make more informed capital allocation decisions.[18] The academic literature, however, is quite ambiguous about the effect of more public information and its overall desirability. An unintended consequence to consider is that a more transparent information environment may crowd out other market participants' information production. In particular, recent theoretical work by

[18]For instance, the Financial Accounting Standards Board (FASB) states: "The benefits of financial reporting information include better investment, credit, and similar resource allocation decisions, which in turn result in more efficient functioning of the capital markets and lower costs of capital for the economy as a whole." Examples of regulations that focus on information provision include the Sarbanes–Oxley Act, which attempts to increase the mandated reporting of firms; the Williams Act of 1968, which limits the ability of investors to trade anonymously on their private (optimistic) information; the regulation on analyst certification (Reg AC), which requires analysts to disclose possible conflicts of interest and prevent biased reports; the Dodd–Frank Act, which includes several measures aimed at improving the transparency and viability of credit ratings.

Frenkel, Guttman, and Kremer (2020) predicts that an increase in the overall amount of public information has the potential to crowd out corporate voluntary disclosure. Whether or not regulators should care about this has been far from clear. Our findings support concerns that the improvements in the information environment could be offset by lower voluntary disclosure. Given the importance of open science for cumulative knowledge production and economic growth, diminishing incentives to disclose valuable research findings are clearly not socially desirable.

In addition, we contribute to the broader discussions on the apparent decline in R&D productivity (Bloom, Jones, Van Reenen, and Webb, 2020) and the associated slowdown in economic growth. If firms can quickly increase scientific publication levels in response to changes in their information environment, it implies that many valuable ideas are hidden from the marketplace. While not unexpected in light of the theoretical discussion by Bhattacharya and Ritter (1983) and others, providing systematic evidence is nonetheless important given the wide-spread concerns over the shift from government-funded research to the private sector. However, in contrast to most current policy and academic debates regarding government-funded versus privately funded research, which focus on differences in the level and the nature of research conducted, we emphasize the point that differences in the incentives to openness also matter. As Furman and Stern (2011) put it, "the ability of a society to stand on the shoulders of giants depends not only on generating knowledge, but also on the quality of mechanisms for storing, certifying, and accessing that knowledge." The link between a potential decline in R&D productivity and the incentives to openness presents a useful line of further inquiry.

Finally, our empirical approach highlights an important but often overlooked problem in the measurement of firm-level innovation. Simply put, innovation outcomes, once observable to the manager, are not automatically disclosed to the public. While valuable insights have been developed over the past decade about the relationship between finance and innovation, these prior studies have not been able to clarify whether increases or decreases in observable innovation outputs are the results of changes in investment strategies or modifications in policies governing the disclosure of research results. However, these results have fueled extensive academic and policy debates on whether performance pressures of the capital markets cause a myopic focus on short-term profits and stifle innovation. We believe that a more careful exploitation of experiments and a deeper analysis of patent- and non-patent based measures of innovation are likely to yield novel insights into the relationship between finance and innovation. Our measure of scientific publications is a first step in that direction.

References

Abdi, F. and A. Ranaldo (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies 30*(12), 4437–4480.

Aghion, P., J. Van Reenen, and L. Zingales (2013). Innovation and institutional ownership. *American Economic Review 103*(1), 277–304.

Anton, J. J. and D. A. Yao (2004). Little patents and big secrets: Managing intellectual property. *The RAND Journal of Economics 35*(1), 1–22.

Arora, A., S. Belenzon, and L. Sheer (2020). Knowledge spillovers and corporate investment in scientific research. The American Economic Review, forthcoming.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson (Ed.), *The Rate and Direction of Inventive Activity*, pp. 609–626. Princeton University Press: Princeton, NJ.

Audretsch, D. B. and P. E. Stephan (1996). Company-scientist locational links: The case of biotechnology. *The American Economic Review 86*(3), 641–652.

Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? *Journal of Economics & Management Strategy 11*(4), 551–594.

Baker, M. and J. Wurgler (2004). A catering theory of dividends. *The Journal of Finance 59*(3), 1125–1165.

Balakrishnan, K., M. B. Billings, B. Kelly, and A. Ljungqvist (2014). Shaping liquidity: On the causal effects of voluntary disclosure. *The Journal of Finance 69*(5), 2237–2278.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies 21*(2), 785–818.

Baruffaldi, S. H. and F. Poege (2020). A firm scientific community: Industry participation and knowledge diffusion. Unpublished working paper, University of Bath, Max Planck Institute for Innovation and Competition, and IZA Institute of Labor Economics.

Bebchuk, L., A. Cohen, and A. Ferrell (2009). What matters in corporate governance? *The Review of Financial Studies 22*(2), 783–827.

Bebchuk, L. A. and L. A. Stole (1993). Do short-term objectives lead to under- or overinvestment in long-term projects? *The Journal of Finance 48*(2), 719–729.

Bellstam, G., S. Bhagat, and J. A. Cookson (2020). A text-based analysis of corporate innovation. Management Science, forthcoming.

Bernstein, S. (2015). Does going public affect innovation? *The Journal of Finance 70*(4), 1365–1403.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics 50*(2), 296–343.

Bhattacharya, S. and J. R. Ritter (1983). Innovation and communication: Signalling with partial disclosure. *The Review of Economic Studies 50*(2), 331–346.

Billett, M. T., J. A. Garfinkel, and M. Yu (2017). The effect of asymmetric information on product market outcomes. *Journal of Financial Economics 123*(2), 357–376.

Björk, B.-C. and D. Solomon (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics 7*(4), 914 – 923.

Blanco, I. and D. Wehrheim (2017). The bright side of financial derivatives: Options trading and firm innovation. *Journal of Financial Economics 125*(1), 99–119.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020). Are ideas getting harder to find? *American Economic Review 110*(4), 1104–44.

Blumenthal, D., N. Causino, E. Campbell, and K. S. Louis (1996). Relationships between academic institutions and industry in the life sciences — an industry survey. *New England Journal of Medicine 334*(6), 368–374.

Brav, A., W. Jiang, S. Ma, and X. Tian (2018). How does hedge fund activism reshape corporate innovation? *Journal of Financial Economics 130*(2), 237 – 264.

Brown, J. R. and G. Martinsson (2019). Does transparency stifle or facilitate innovation? *Management Science 65*(4), 1600–1623.

Brown, L. D., A. C. Call, M. B. Clement, and N. Y. Sharp (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research 53*(1), 1–47.

Chen, T., J. Harford, and C. Lin (2015). Do analysts matter for governance? Evidence from natural experiments. *Journal of Financial Economics 115*(2), 383–410.

Clarke, J., N. Dass, and A. Patel (2015). When do analysts impede innovation? Unpublished working paper, Georgia Institute of Technology and Wake Forest University.

Core, J. and W. Guay (2002). Estimating the value of employee stock option portfolios and their sensitivities to price and volatility. *Journal of Accounting Research 40*(3), 613–630.

Darrough, M. N. and N. M. Stoughton (1990). Financial disclosure policy in an entry game. *Journal of Accounting and Economics 12*(1), 219–243.

Dasgupta, P. and P. A. David (1994). Toward a new economics of science. *Research Policy 23*(5), 487–521.

Derrien, F. and A. Kecskés (2013). The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *The Journal of Finance 68*(4), 1407–1440.

Dyck, A., A. Morse, and L. Zingales (2010). Who blows the whistle on corporate fraud? *The Journal of Finance 65*(6), 2213–2253.

Edmans, A., X. Gabaix, and A. Landier (2009). A multiplicative model of optimal ceo incentives in market equilibrium. *The Review of Financial Studies 22*(12), 4881–4917.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov (2012). Dynamic CEO compensation. *The Journal of Finance 67*(5), 1603–1647.

Fang, V. W., X. Tian, and S. Tice (2014). Does stock liquidity enhance or impede firm innovation? *The Journal of Finance 69*(5), 2085–2125.

Farre-Mensa, J., D. Hegde, and A. Ljungqvist (2020). What is a patent worth? Evidence from the U.S. patent "lottery". *The Journal of Finance 75*(2), 639–682.

Fich, E. M., J. L. Juergens, and M. S. Officer (2018). Analyst coverage and acquisition returns: Evidence from natural experiments. Unpublished Working Paper, Drexel University, U.S. Securities and Exchange Commission and Loyola Marymount University.

Frenkel, S., I. Guttman, and I. Kremer (2020). The effect of exogenous information on voluntary disclosure and market quality. *Journal of Financial Economics 138*(1), 176–192.

Furman, J. L. and S. Stern (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review 101*(5), 1933–1963.

Gervais, S., R. Kaniel, and D. H. Mingelgrin (2001). The high-volume return premium. *The Journal of Finance 56*(3), 877–919.

Gibbons, R. and K. J. Murphy (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy 100*(3), 468–505.

Glaeser, S., J. Michels, and R. Verrecchia (2020). Discretionary disclosure and manager horizon: Evidence from patenting. *Review of Accounting Studies 25*(2), 597–635.

Gompers, P., J. Ishii, and A. Metrick (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics 118*(1), 107–156.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics 40*(1-3), 3–73.

Guo, B., D. Pérez-Castrillo, and A. Toldrà-Simats (2019). Firms' innovation strategy under the shadow of analyst coverage. *Journal of Financial Economics 131*(2), 456–483.

Guo, R.-J., B. Lev, and N. Zhou (2004). Competitive costs of disclosure by biotech ipos. *Journal of Accounting Research 42*(2), 319–355.

Hall, B. H. and D. Harhoff (2012). Recent research on the economics of patents. *Annual Review of Economics 4*(1), 541–565.

Hall, B. H. and J. Lerner (2010). The financing of R&D and innovation. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of The Economics of Innovation* (1st ed.). North-Holland.

He, J. J. and X. Tian (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial*

Economics *109*(3), 856–878.

Henderson, R. and I. Cockburn (1994). Measuring competence? exploring firm effects in pharmaceutical research. *Strategic Management Journal* *15*(S1), 63–84.

Hicks, D. (1995). Published papers, tacit competencies and corporate management of the public/private character of knowledge. *Industrial and Corporate Change* *4*(2), 401–424.

Hoberg, G. and V. Maksimovic (2015). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies* *28*(5), 1312–1352.

Hong, H. and M. Kacperczyk (2010). Competition and bias. *The Quarterly Journal of Economics* *125*(4), 1683–1725.

Iacus, S. M., G. King, and G. Porro (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis* *20*(1), 1–24.

Irani, R. M. and D. Oesch (2013). Monitoring and corporate disclosure: Evidence from a natural experiment. *Journal of Financial Economics* *109*(2), 398–418.

Kelly, B. and A. Ljungqvist (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies* *25*(5), 1366–1413.

Kerr, W. R. and R. Nanda (2015). Financing innovation. *Annual Review of Financial Economics* *7*(1), 445–462.

Lang, M. H. and R. J. Lundholm (1996). Corporate disclosure policy and analyst behavior. *The Accounting Review* *71*(4), 467–492.

Leuz, C. and P. D. Wysocki (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research* *54*(2), 525–622.

Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies* *27*(6), 1797–1829.

Lucas, R. E. and B. Moll (2014). Knowledge growth and the allocation of time. *Journal of Political Economy* *122*(1), 1–51.

Marx, M. and A. Fuegi (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal* *41*(9), 1572–1594.

Mongeon, P. and A. Paul-Hus (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* *106*(1), 213–228.

Narayanan, M. P. (1985). Managerial incentives for short-term results. *The Journal of Finance* *40*(5), 1469–1484.

Polidoro, F. and M. Theeke (2012). Getting competition down to a science: The effects of technological competition on firms' scientific publications. *Organization Science* *23*(4), 1135–1153.

Prendergast, C. and L. Stole (1996). Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning. *Journal of Political Economy* *104*(6), 1105–1134.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* *98*(5), S71–S102.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy* *19*(2), 165–174.

Saidi, F. and A. Žaldokas (2020). How does firms' innovation disclosure affect their banking relationships? Management Science, forthcoming.

Simeth, M. and M. Cincera (2016). Corporate science, innovation, and firm value. *Management Science* *62*(7), 1970–1981.

Simeth, M. and J. D. Raffo (2013). What makes companies pursue an open science strategy? *Research Policy* *42*(9), 1531–1543.

Stein, J. C. (1988). Takeover threats and managerial myopia. *Journal of Political Economy* *96*(1), 61–80.

Stein, J. C. (1996). Rational capital budgeting in an irrational world. *The Journal of Business* *69*(4), 429–455.

Stephan, P. E. (1994). Differences in the post-entry value of biotech firms: The role of human capital. Unpublished working paper, Georgia State University.

Stern, S. (2004). Do scientists pay to be scientists? *Management Science* *50*(6), 835–853.

Swarup, A. (2008). Analyzing scientific investments. *Science* *322*(5905), 1266–1267.

Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics* *5*, 179–194.

Tables

Table 1
Definitions of variables.
This table provides definitions of the main variables.

Variable	Definition	Data Source
Corporate scientific publication measures:		
PUB	Natural logarithm of (one plus) the total number of scientific publications (articles and conference papers) per firm in a given year	Elsevier's Scopus
PUB_CIT	Natural logarithm of (one plus) the total number of publications weighted by the number of citations each publication receives in a five-year window	Elsevier's Scopus
PUB_JIF	Natural logarithm of (one plus) the total number of publications weighted by the Journal Impact Factor	Elsevier's Scopus, SCImago Journal Rank
PUB_COL	Fraction of publications with academic institutions to firm's total number of publications	Elsevier's Scopus
PUB_GROWTH	Change in the number publications over the three-year period before the event year, defined as the number of publications in year −1 minus the number of publications in year −3	Elsevier's Scopus
Analyst coverage and control variables:		
COV	Analyst coverage, defined as the number of analysts covering the firm during the fiscal year. LN_COV is the natural logarithm (one plus) of COV.	I/B/E/S
LN_AT	Natural logarithm of book value of total assets (Compustat item #6) at the end of the fiscal year	Compustat
RDTA	R&D expenditure (#46) divided by book value of total assets (#6), set to zero if missing	Compustat
LN_AGE	Natural logarithm of the number of years the firm is included in Compustat	Compustat
PPETA	Net property, plant & equipment (#8) divided by book value of total assets (#6)	Compustat
CAPEXTA	Capital expenditures (#128) divided by book value of total assets (#6)	Compustat
ROA	Return on assets, defined as operating income before depreciation (#13) divided by book value of total assets (#6)	Compustat
LEV	Leverage, defined as book value of debt (#9+#34) divided by book value of total assets (#6)	Compustat
Q	Market-to-book ratio, calculated as market value of equity (#199 × #25) plus book value of total assets (#6) minus book value of equity (#60) minus balance sheet deferred taxes (#74, set to zero if missing) divided by book value of total assets (#6)	Compustat
DELAYCON	Financial constraints, calculated based on a text-based index of financial constraints from Hoberg and Maksimovic (2015), which scores firm that are at risk of delaying their investments due to liquidity issues	Gerard Hoberg's website
INSTOWN	Institutional ownership, calculated as the arithmetic mean of the four quarterly institutional holdings reported through form 13F over the fiscal year	CDA/Spectrum
ILLIQ	Natural logarithm of bid-ask spreads measured over the fiscal year, where bid-ask spreads are obtained using Abdi and Ranaldo's (2017) close-high-low measure	Farshid Abdi's website

(Continued)

(*Continued*)

Variable	Definition	Data Source
HINDEX	Herfindahl index of the four-digit SIC industry sales to which a firm belongs. HINDEX^2 is the square of *HINDEX*.	Compustat
Additional variables used in robustness tests and the cross-sectional tests:		
FDISP	Analysts' forecast dispersion, calculated as the standard deviation of analysts' EPS forecasts at the month of the forecast fiscal year-end divided by the absolute value of actual EPS. We use EPS forecasts for current year (FPI = 1) from the I/B/E/S detail file and exclude the estimates which are stopped or excluded or stale forecasts. We keep the last forecast for each analyst and then calculate the standard deviation of the forecasts for each fiscal year. We drop the standard deviation of forecasts for firms with less than three analysts following during the fiscal year	I/B/E/S
RET	Past returns, calculated as the average monthly stock return during the fiscal year	CRSP
EQUITYDELAYCON	Equity financial constraints, calculated based on a text-based index of financial constraints from Hoberg and Maksimovic (2015), which scores firm that are at risk of delaying their investments due to liquidity issues and that indicate plans to issue equity (presumably to address their liquidity challenges)	Gerard Hoberg's website
CREDIT RATING	Equals one if the firm has a Standard & Poor's (S&P) Domestic Long Term Issuer Credit Rating (splticrm) from "AAA" to "C"	S&P credit ratings
PAYOUT	Equals one if the firm has a positive dividend payout (#21)	Compustat
GINDEX	Governance index, which is based on the 24 shareholder rights introduced by Gompers, Ishii, and Metrick (2003)	IRRC/RiskMetrics
EINDEX	Entrenchment index, which is based on the six shareholder rights introduced by Bebchuk, Cohen, and Ferrell (2009)	IRRC/RiskMetrics
DUAL CEO	Equals one if the CEO is also the chairman of the firm's board	IRRC/RiskMetrics
CEO AGE	Age of the CEO	ExecuComp
PPS	Pay–performance sensitivity, calculated as the sum of the expected dollar changes in the CEO's stock and options holdings for a 1% change in the aggregate value of the firm's equity following the procedure in Core and Guay (2002)	ExecuComp
Corporate patenting measures:		
PAT	Natural logarithm of (one plus) the total number of granted patents per firm in a given year, dated by the patent application year	DISCERN, Patstat (2019 Autumn Edition)
PAT_CIT	Natural logarithm of (one plus) the total number patents weighted by the number of citations each patent receives in a five-year window	DISCERN, Patstat (2019 Autumn Edition)
PAT_SCI	Fraction of science-based patents to firm's total number of patents, where science-based patents are defined as patents with at least one reference to science on their front page	DISCERN, Patstat (2019 Autumn Edition)
PAT_SCI_EXT	Fraction of science-based patents with external references to firm's total number of patents	DISCERN, Patstat (2019 Autumn Edition)

Table 2
Summary statistics for the baseline sample.

This table presents summary statistics for the baseline sample using U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period 1997-2014. Variable definitions are provided in Table 1.

Variable	Mean	StdDev	10%	25%	Median	75%	90%	Observations
PUB	1.232	1.503	0.000	0.000	0.693	2.079	3.401	20,342
LN_COV	1.577	0.990	0.000	0.773	1.626	2.335	2.904	20,342
LN_AT	5.909	2.126	3.242	4.307	5.738	7.387	8.828	20,342
RDTA	0.106	0.137	0.005	0.019	0.060	0.132	0.256	20,342
LN_AGE	2.650	0.769	1.609	2.079	2.708	3.296	3.664	20,342
PPETA	0.179	0.143	0.033	0.069	0.141	0.252	0.387	20,342
CAPEXTA	0.040	0.036	0.008	0.016	0.029	0.051	0.085	20,342
ROA	0.027	0.243	-0.269	-0.009	0.098	0.158	0.219	20,342
LEV	0.150	0.161	0.000	0.000	0.105	0.256	0.383	20,342
Q	2.323	1.806	0.982	1.239	1.721	2.679	4.340	20,342
DELAYCON	-0.009	0.086	-0.120	-0.068	0.000	0.031	0.104	20,342
HINDEX	0.243	0.189	0.063	0.113	0.187	0.310	0.508	20,342
$HINDEX^2$	0.095	0.160	0.004	0.013	0.035	0.096	0.258	20,342
INSTOWN	0.533	0.287	0.096	0.288	0.583	0.778	0.887	20,342
ILLIQ	-4.335	0.585	-5.093	-4.768	-4.351	-3.927	-3.562	20,342

Table 3
Number and percentage of firms with and without publications by industry.

This table presents the number and percentage of firms that have published as well as the number and percentage of firms that have not published over the sample period from 1997 to 2014 in each Fama-French (FF) 12 industry. The classification is available at Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

FF	Industry name	Description	Firms with positive publications	Firms with zero publications	Total no. of firms
1	Nodur	Consumer nondurables (food, tobacco, textiles, apparel, leather, toys)	47 (61%)	30 (39%)	77
2	Durbl	Consumer durables (cars, TVs, furniture, household appliances)	58 (55%)	48 (45%)	106
3	Manuf	Manufacturing (machinery, trucks, planes, office furniture, paper, commercial printing)	275 (70%)	116 (30%)	391
4	Enrgy	Oil, gas, and coal extraction and products	23 (85%)	4 (15%)	27
5	Chems	Chemicals and allied products	80 (85%)	14 (15%)	94
6	BusEq	Business equipment (computers, software, electronic equipment)	698 (72%)	277 (28%)	975
7	Telcm	Telephone and television transmission	20 (57%)	15 (43%)	35
9	Shops	Wholesale, retail, and some services (laundries, repair shops)	12 (44%)	15 (56%)	27
10	Hlth	Healthcare, medical equipment, and drugs	554 (85%)	98 (15%)	652
12	Other	Mines, construction, building, materials, transportation, hotels, business services, entertainment	64 (60%)	43 (40%)	107

Table 4
Pre-event characteristics for treatment and control sample.
This table presents summary statistics for the treatment and control sample. *Panel A* reports summary statistics for the treatment sample. *Panel B* reports summary statistics for the control sample. The treatment sample is a combination of the broker merger and closure events from Kelly and Ljungqvist (2012) and Fich, Juergens, and Officer (2018). Our sample includes 43 broker disappearances from 2000 to 2010. We follow the procedure put forth in He and Tian (2013) to identify firms whose analyst coverage is reduced due to merger/closure events. The control sample is the remainder of firms in the CRSP/Compustat-merged universe with the required data that are not covered by either the merging or closing brokers before the event. Variable definitions are provided in Table 1.

Panel A: Treatment sample

Variable	Mean	StdDev	10%	25%	Median	75%	90%	Observations
PUB	2.537	1.905	0.000	1.099	2.303	3.970	5.545	750
COV	16.554	8.196	5.333	9.846	16.641	22.333	27.917	750
LN_AT	8.081	1.790	5.760	6.836	8.078	9.417	10.455	750
RDTA	0.074	0.074	0.003	0.018	0.056	0.107	0.150	750
LN_AGE	2.787	0.795	1.609	2.197	2.833	3.584	3.726	750
PPETA	0.205	0.156	0.049	0.086	0.165	0.274	0.445	750
CAPEXTA	0.048	0.035	0.014	0.023	0.037	0.063	0.095	750
ROA	0.127	0.129	0.010	0.087	0.135	0.196	0.256	750
LEV	0.169	0.154	0.000	0.017	0.150	0.276	0.386	750
Q	2.938	2.094	1.162	1.543	2.246	3.517	5.636	750
DELAYCON	-0.014	0.074	-0.105	-0.064	0.000	0.007	0.091	750
HINDEX	0.218	0.179	0.060	0.081	0.161	0.281	0.483	750
HINDEX^2	0.079	0.142	0.004	0.007	0.026	0.079	0.234	750
INSTOWN	0.684	0.175	0.455	0.561	0.703	0.821	0.900	750
ILLIQ	-4.560	0.527	-5.244	-4.934	-4.607	-4.160	-3.797	750

Panel B: Control sample

Variable	Mean	StdDev	10%	25%	Median	75%	90%	Observations
PUB	0.959	1.182	0.000	0.000	0.693	1.609	2.708	20,539
COV	4.429	3.592	1.000	1.833	3.500	6.000	9.083	20,539
LN_AT	5.538	1.530	3.618	4.447	5.446	6.580	7.595	20,539
RDTA	0.097	0.125	0.005	0.019	0.056	0.124	0.229	20,539
LN_AGE	2.529	0.749	1.386	1.946	2.565	3.135	3.555	20,539
PPETA	0.186	0.140	0.036	0.076	0.150	0.263	0.392	20,539
CAPEXTA	0.042	0.037	0.009	0.017	0.032	0.053	0.089	20,539
ROA	0.047	0.210	-0.213	0.001	0.103	0.162	0.220	20,539
LEV	0.141	0.155	0.000	0.000	0.090	0.243	0.363	20,539
Q	2.438	1.905	1.012	1.272	1.787	2.837	4.637	20,539
DELAYCON	-0.012	0.093	-0.129	-0.081	-0.004	0.035	0.112	20,539
HINDEX	0.246	0.185	0.063	0.117	0.193	0.317	0.496	20,539
HINDEX^2	0.095	0.154	0.004	0.014	0.037	0.100	0.246	20,539
INSTOWN	0.550	0.255	0.172	0.350	0.578	0.759	0.875	20,539
ILLIQ	-4.284	0.515	-4.963	-4.648	-4.298	-3.904	-3.618	20,539

Table 5
Baseline OLS regressions of corporate publications on analyst coverage.

This table presents the baseline ordinary least squares (OLS) regression results of corporate publications on analyst coverage over the sample period from 1997 to 2014. Robust standard errors are clustered by firm (in parentheses). Variable definitions are provided in Table 1. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)
LN_COV	0.768***	0.145***	0.126***	0.036**	0.039***	0.048***
	(0.034)	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)
LN_AT				0.246***	0.241***	0.257***
				(0.021)	(0.021)	(0.023)
RDTA				0.508***	0.534***	0.536***
				(0.114)	(0.114)	(0.115)
LN_AGE				-0.092**	-0.094**	-0.086*
				(0.046)	(0.046)	(0.046)
PPETA				0.380***	0.358***	0.339***
				(0.129)	(0.129)	(0.129)
CAPEXTA				-0.164	-0.104	-0.054
				(0.227)	(0.226)	(0.226)
ROA				-0.106*	-0.089	-0.083
				(0.056)	(0.056)	(0.056)
LEV				0.159**	0.145**	0.130*
				(0.067)	(0.067)	(0.068)
Q					-0.012**	-0.010**
					(0.005)	(0.005)
DELAYCON					0.186*	0.184*
					(0.095)	(0.095)
HINDEX					-0.723**	-0.704**
					(0.325)	(0.323)
HINDEX^2					0.712**	0.696**
					(0.286)	(0.285)
INSTOWN						-0.139*
						(0.072)
ILLIQ						0.027
						(0.024)
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes
Number of obs.	20,342	20,342	20,342	20,342	20,342	20,342
Adjusted R^2	0.256	0.896	0.898	0.903	0.903	0.903

Table 6
Baseline DiD regressions of corporate publications on the analyst coverage shock.

This table presents the baseline difference-in-difference (DiD) regression results of corporate publications on the analyst coverage shock. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. Variable definitions are provided in Table 1. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	COV	FDISP	PUB						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TREATED × POST	-1.334***	0.010**	0.092**	0.133***	0.124***	0.125***	0.114***	0.113***	0.115***
	(0.246)	(0.005)	(0.041)	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)
POST	0.482***	0.000	0.015	0.040***	-0.001	-0.003	0.003	0.002	0.002
	(0.070)	(0.002)	(0.026)	(0.012)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
TREATED	12.140***	-0.018***	1.630***	-0.060	-0.024	-0.022	-0.035	-0.035	-0.038
	(0.479)	(0.005)	(0.123)	(0.044)	(0.042)	(0.042)	(0.039)	(0.039)	(0.039)
LN_AT							0.198***	0.200***	0.217***
							(0.020)	(0.021)	(0.025)
RDTA							0.437***	0.461***	0.476***
							(0.110)	(0.108)	(0.110)
LN_AGE							-0.055	-0.048	-0.037
							(0.055)	(0.054)	(0.054)
PPETA							0.200	0.196	0.184
							(0.143)	(0.143)	(0.143)
CAPEXTA							0.044	0.047	0.088
							(0.258)	(0.259)	(0.258)
ROA							-0.064	-0.052	-0.044
							(0.061)	(0.062)	(0.062)
LEV							0.107	0.104	0.085
							(0.081)	(0.082)	(0.084)
Q								-0.006	-0.004
								(0.006)	(0.006)
DELAYCON								0.139	0.136
								(0.112)	(0.112)
HINDEX								-0.458	-0.460
								(0.335)	(0.334)
$HINDEX^2$								0.640**	0.637**
								(0.291)	(0.291)
INSTOWN									-0.062
									(0.076)
ILLIQ									0.042
									(0.028)
Firm fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Event fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes
Number of obs.	114,020	78,951	114,020	114,020	114,020	114,020	114,020	114,020	114,020
Adjusted R^2	0.207	0.001	0.063	0.873	0.876	0.876	0.880	0.880	0.880

Table 7
Baseline DiD regressions of additional corporate publication measures on the analyst coverage shock.

This table presents the baseline difference-in-difference (DiD) regression results of additional corporate publication measures on the analyst coverage shock. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	PUB CIT (1)	PUB JIF (2)	PUB JIF⩾p50 (3)	PUB JIF<p50 (4)	PUB COL (5)
TREATED × POST	0.186***	0.223***	0.173***	0.032**	0.032**
	(0.052)	(0.031)	(0.027)	(0.013)	(0.013)
POST	-0.004	-0.005	-0.003	0.001	0.007
	(0.019)	(0.009)	(0.007)	(0.005)	(0.006)
TREATED	-0.017	-0.127***	-0.098***	-0.006	-0.005
	(0.074)	(0.045)	(0.038)	(0.016)	(0.015)
Controls included	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	114,020	114,020	114,020	114,020	70,958
Adjusted R^2	0.822	0.862	0.844	0.783	0.516

Table 8
Robustness tests for baseline DiD: Coarsened Exact Matching (CEM).
This table presents the baseline difference-in-difference (DiD) regression results of corporate publications on the analyst coverage shock by balancing the sample on pre-treatment covariates using Coarsened Exact Matching (CEM). Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4 with valid matching variables in year −1. See Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Mean Treated (1)	StdDev Treated (2)	Mean Control (3)	StdDev Control (4)	Difference p-value (5)	PUB (6)
TREATED X POST						0.146***
						(0.056)
POST						-0.123**
						(0.048)
TREATED						-0.101
						(0.075)
Matching variables:						
LN_AT	6.648	1.405	6.545	1.356	0.303	
Q	2.285	1.194	2.284	1.202	0.995	
COV	11.658	5.710	11.111	5.530	0.182	
RET	0.010	0.043	0.013	0.044	0.421	
PUB_GROWTH	0.073	0.608	0.039	0.559	0.412	
Weights						CEM
Firm fixed effects						Yes
Year fixed effects						Yes
Event fixed effects						Yes
Number of obs.	225		1,044			6,789
Adjusted R^2						0.910

Table 9

Baseline DiD: Conditioning on financial constraints.

This table presents the difference-in-difference (DiD) regression results on corporate publications when conditioning on pre-event financial constraints measures (in year -1). For each of the variables used to capture financial constraints, we split the sample and run our baseline specification for the two sub-samples. Columns 1 and 2 split the sample by the median value of the delaycon measure. Columns 3 and 4 split the sample by the median value of the equitydelaycon measure. Delaycon and equitydelaycon are from Hoberg and Maksimovic (2015). Columns 5 and 6 split the sample into firms with and without S&P long-term credit rating. Columns 7 and 8 split the sample into firms with and without payout. Columns 9 and 10 split the sample by the median value of firm age. All variables are explained in detail in Table 1. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	DELAYCON		EQUITY DELAYCON		CREDIT RATING		PAYOUT		FIRM AGE	
Sample	Low	High	Low	High	Yes	No	Yes	No	Young	Old
Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(10)
TREATED × POST	0.047	0.177***	0.048	0.215***	0.051	0.187***	0.035	0.215***	0.252***	0.038
	(0.053)	(0.042)	(0.059)	(0.046)	(0.053)	(0.044)	(0.052)	(0.044)	(0.049)	(0.045)
POST	0.016	-0.011	0.008	0.006	-0.033*	0.011	-0.014	0.015	0.016	-0.015
	(0.014)	(0.015)	(0.015)	(0.015)	(0.018)	(0.010)	(0.013)	(0.012)	(0.013)	(0.012)
TREATED	-0.006	-0.088	0.014	-0.077	0.011	-0.052	0.021	-0.093*	-0.130*	0.036
	(0.058)	(0.065)	(0.061)	(0.069)	(0.057)	(0.049)	(0.058)	(0.054)	(0.068)	(0.051)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	57,202	55,927	52,454	50,519	25,005	88,124	42,282	70,847	56,936	56,193
Adjusted R^2	0.874	0.889	0.871	0.885	0.924	0.837	0.906	0.860	0.869	0.892

Table 10
Baseline DiD: Conditioning on managerial incentives.

This table presents the difference-in-difference (DiD) regression results on corporate publications when conditioning on pre-event managerial incentives measures (in year −1). For each of the variables used to capture managerial incentives, we split the sample and run our baseline specification for the two sub-samples. Columns 1 and 2 split the sample by the median value of the Gompers, Ishii, and Metrick (2003) governance index. Columns 3 and 4 split the sample by the median value of the Bebchuk, Cohen, and Ferrell (2009) entrenchment index. Columns 5 and 6 split the sample into firms depending on whether the CEO is also the chairman of the board of directors. Columns 7 and 8 split the sample by the median value of CEO age. Columns 9 and 10 split the sample the median value of the change in the sensitivity of pay to stock price measure of Core and Guay (2002). All variables are explained in detail in Table 1. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	GINDEX		EINDEX		DUAL CEO		CEO AGE		PPS	
Sample	Low	High	Low	High	No	Yes	Young	Old	Low	High
Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(10)
TREATED × POST	0.163***	0.043	0.161***	0.042	0.187***	0.039	0.163***	0.005	0.173**	0.054
	(0.049)	(0.051)	(0.054)	(0.046)	(0.051)	(0.055)	(0.046)	(0.071)	(0.072)	(0.048)
POST	-0.013	-0.000	-0.023	0.003	-0.003	-0.001	0.016	-0.003	-0.012	0.013
	(0.024)	(0.020)	(0.021)	(0.020)	(0.020)	(0.018)	(0.018)	(0.022)	(0.020)	(0.020)
TREATED	-0.058	0.035	-0.063	0.076	0.030	-0.037	-0.066	0.007	0.036	0.072
	(0.082)	(0.061)	(0.069)	(0.062)	(0.081)	(0.063)	(0.059)	(0.104)	(0.080)	(0.055)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	24,695	25,361	23,096	26,960	26,772	32,175	29,370	25,598	28,971	29,444
Adjusted R^2	0.905	0.909	0.910	0.905	0.877	0.912	0.907	0.894	0.839	0.921

Table 11
Baseline DiD regressions of R&D inputs, corporate patenting activities and the use of science on the analyst coverage shock.

This table presents the difference-in-difference (DiD) regression results of innovation inputs and patent-based measures of corporate innovation on the analyst coverage shock. Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for variable definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	RDTA (1)	PAT (2)	PAT CIT (3)	PAT SCI (4)	PAT SCI EXT (5)
TREATED × POST	0.003	0.029	-0.031	-0.004	-0.003
	(0.002)	(0.045)	(0.065)	(0.010)	(0.010)
POST	0.001	-0.067***	-0.122***	-0.006	-0.002
	(0.001)	(0.011)	(0.022)	(0.004)	(0.005)
TREATED	-0.000	-0.062	-0.059	0.001	0.007
	(0.003)	(0.055)	(0.079)	(0.012)	(0.012)
Controls included	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	114,020	114,020	114,020	77,639	77,639
Adjusted R^2	0.863	0.852	0.769	0.725	0.614

Figures

Panel A. Business R&D expenditures *Panel B. Patents and scientific publications*

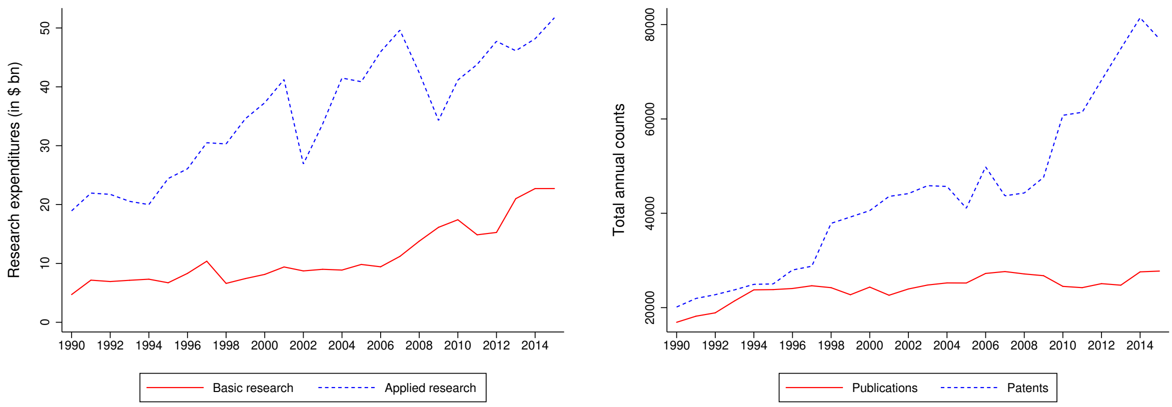

Fig. 1. Aggregate data on R&D inputs and outputs. This figure presents aggregate data on corporate R&D inputs and outputs. *Panel A* depicts the total annual expenditures by businesses in the United States by type of R&D of performed. Data for this plot come from the National Science Foundation, National Center for Science and Engineering Statistics. NSF 20-307. National Patterns of R&D Resources: 2017–18 Data Update. Available at https://ncses.nsf.gov/pubs/nsf20307. *Panel B* depicts the total annual publications and patents produced by U.S. publicly-traded firms in the Compustat universe. Patent data come from Arora, Belenzon, and Sheer (2020) and publication data from Elsevier's Scopus database.

Panel A. Distribution of broker disappearances *Panel A. Distribution of treated firms*

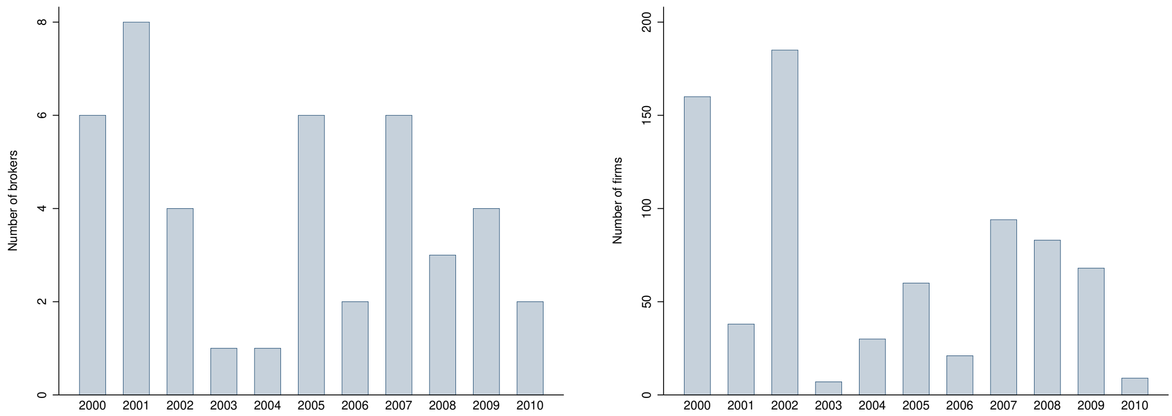

Fig. 2. Distribution of broker disappearances and treated firms by calendar year. This figure presents the distribution of broker house disappearances and treated firms by calendar year. *Panel A* depicts the distribution of brokerage house disappearances by calendar year. *Panel B* depicts the distribution of firms that lose an analyst due to brokerage house disappearances by calendar year. The sample compromises 750 treated firms that lose an analyst between 2000 and 2010. mergers. These firms are U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period 1997-2014.

Panel A. Difference (treated-control) in coverage *Panel B. Difference (treated-control) in publications*

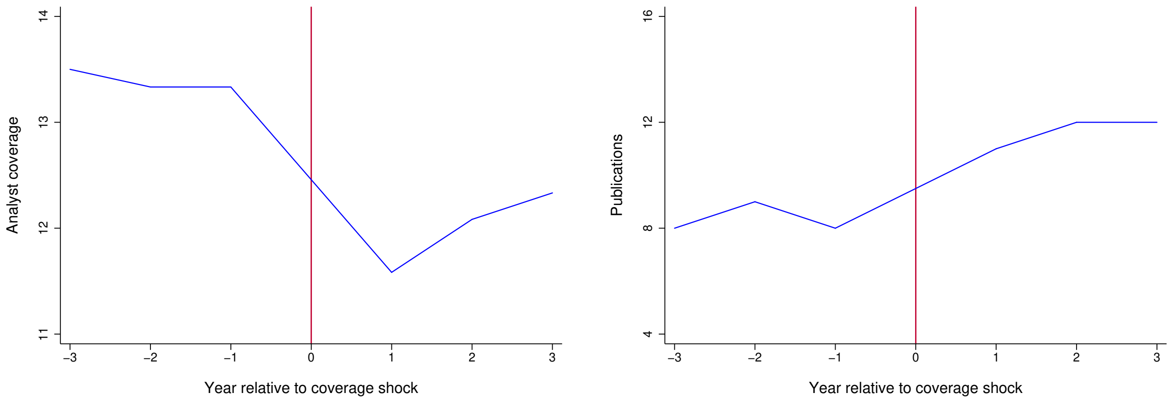

Fig. 3. Trends in analyst coverage and corporate publications in the treatment sample. This figure presents differences between treatment and control firms in analyst coverage and scientific publications, from the three years before the coverage shock to the three years after the coverage shock. The coverage shock year is denoted as year 0. *Panel A* depicts the median difference in analyst coverage between treated and control firms. *Panel B* the median difference in scientific publications between treated and control firms. The sample compromises 750 treated firms that lose an analyst between 2000 and 2010. These firms are U.S. public firms with at least one year of positive R&D expenditures and at least one patent over the period 1997-2014.

Internet Appendix for "The real effects of financial markets on scientific disclosure: Evidence from a quasi-natural experiment"

This internet appendix provides supplemental information and robustness tests that accompany the main article. In Section A.1, we provide details on the matching procedure between North American Centre for Research in Security Prices (CRSP)/Compustat Merged Database records and scientific publication data from Elsevier's Scopus database. In Section A.2, we show additional difference-in-difference (DiD) robustness results in tables that do not appear in the article.

A.1. Matching scientific publications to Compustat

In this section, we describe the methodology used to match firms in the Compustat universe to scientific publications data from Scopus for the sample period between 1997 and 2014. The process is divided into five distinctive steps.

Step 1: Sample selection and accounting for name changes

We begin with records from Compustat and select all firms with positive assets (Compustat Annual Item #6), sales (#12), and equity (#60), and require firms to have positive R&D expenses (#46) for at least one year during our sample period. We exclude financial and utility firms [with Standard Industrial Classification (SIC) codes 4900-4999 and 6000-6999] as well as firms that are not headquartered in the U.S. based on their current headquarters location (LOC). After matching the remaining firms to the Duke Innovation & Scientific Enterprises Research Network (DISCERN) patent dataset created by Arora, Belenzon, and Sheer (2020), we further restrict our sample to Compustat firms with at least one patent during our sample period. Before matching those firms to the scientific publication data in Scopus, one important issue to consider is that firms listed in Compustat appear under the their most recent name ($CONM$), while scientific publication records contain the firm name at the time of their publication. Without correcting for this, we may undercount scientific publications from earlier sample years. A prominent example is Google. Google is listed in the Compustat database under the name "Alphabet" but has published in the past (and is still publishing) under the name "Google". Following Arora, Belenzon, and Sheer (2020), we complement the most recent firm names listed in Compustat with the historical firm names listed in the CRSP Monthly Stock File. We find that our sample contains approximately 30% of Compustat records with more than one related name.[1]

Step 2: Cleaning and standardizing firm names

Firm names available in the Compustat database contain legal identifiers and indistinctive name components that have little meaning for identifying false positives while possibly reducing

[1] This is the same figure as in Arora, Belenzon, and Sheer (2020), who cover the period between 1980 and 2015.

the recall rate tremendously when querying the Scopus database. Therefore, we develop and execute an extensive cleaning script to remove all these indistinctive name components. This step is especially important as we download publications using an online source. The Scopus application programming interface (API) accesses a rather flexible search engine that allows for partial matches, word permutations, boolean search queries, and the use of wildcards. However, it does not allow for fuzzy matching of words. This implies that names have to be carefully pre-cleaned to ensure that recall is complete. Common legal extension terms include, for instance, "CO", "CORP", "CORPORATION", "INC", "HLDG", and "HOLDINGS", among many others (including various combinations thereof). For instance, the raw name "LILLY (ELI) & CO" is cleaned toward "ELI LILLY" (and the permutation thereof). We further harmonize various abbreviations of firm name extensions, as displayed below. We follow a hybrid approach for these name terms since they possibly reduce the recall but also have the potential to decrease noise when querying firm names in the Scopus API. As a general rule, these extensions have been removed, but if pretests and manual inspections indicate problematic remaining names creating huge noise, the extensions are retained and harmonized to limit false positives. Examples include "COMM", "ELECTR", "INSTR", "LAB", "PHARM", "SYS", and "TECH", among others. Importantly, beyond generic corrections, we screen exhaustively all firm names and made various manual corrections of idiosyncratic firm names. Moreover, possibly problematic names are pretested in the standard Scopus API to identify whether they create false positives or false negatives.

Step 3: Querying the Scopus API with the standardized firm names

Once the firm names are prepared for the publication download, we query the Scopus API to search for all possible publications by our sample firms and additional information to improve the matching precision. First, we download all publications that match any firm name in the affiliation name field. Each author of a scientific publication is tied to an affiliation, and this target field allows for searching publications by institutions such as firms or universities. In the Scopus query syntax, it is possible to restrict the search to the part of the affiliation field referring to the name of the affiliation (i.e., the address can be excluded) where the name of a firm would be reported (Scopus query command: "Affilorg"). This yields a list of publications where at least one author has an affiliation that matches the search criteria. After having executed a complete test download in October 2019, we downloaded all scientific publications that contain the names of our sample firms in the affiliation name field between April 23 and 27, 2020. To reduce false positives from the outset, we impose the criterion that in case of multiple name components, multiple name components must follow consecutively (e.g., "General Electric") and without permutations for the publication to be downloaded. However, in cases where we suspect that names might be permuted by the authors (e.g., "Eli Lilly" or "Lilly Eli"), we query both name orders. Second, we complete the affiliation information in the downloaded publications with an additional specific search from the Scopus Affiliation Search. Publication records, when downloaded via the Scopus API, report the Scopus affiliation identifier and only partial information regarding the affiliation. For instance, location information is reported in a separate field and cannot be unambiguously matched at the affiliation level. To circumvent this limitation, we reconstruct the complete affiliation information by searching for it, starting from the Scopus affiliation identifiers, for all affiliations present in our full set of downloaded publications. This search yields a list of affiliations with well-structured information on affiliation name, name variants, nearly complete and clean location data (city and country), and a total count of publications attributed to each affiliation in Scopus. Third, we classify all affiliations as either a firm or other types (academic, government, etc.). We do this by applying a combination of methods. To begin with, we search for a list of keywords that rather unambiguously imply that one affiliation belongs to a category (e.g., "university", "college", "school", "faculty"; against "inc.", "corp.", etc.). Then, we match at a highly precise level of string similarity all our affiliations with affiliations in the Global Research Identifier Database (GRID; https://www.grid.ac), which contains classifications for different categories. Finally, for the remaining affiliations not yet classified, we retrieve the Scopus classification, which is available via an affiliation-level API (Scopus Affiliation Retrieval API). The remaining unclear affiliation types are sorted by frequency, screened, and assigned manually up to a frequency of 15 observations. For the remainder, we prioritize recall and classify ambiguous cases as firms. Next, we reshape the publication data to identify and maintain for each publication only the affiliation that is associated with the focal firm searched (whether the match is correct or not). We do this by maintaining for each publication only affiliations classified as firms unless all affiliations are classified as non-firm. For any remaining multiple affiliations, we retain the one with the highest string similarity with the searched firm. We verify that this process identifies the affiliation responsible for the retrieved publication in virtually all cases.

Step 4: Identifying incorrect matches using a machine learning approach

The above steps yield a long list of affiliations paired to firm-name pairs, which we deduplicate at the level of the firm only. At this stage, this list is still characterized by a high recall (close to 100% in practice) but low precision. We proceed further by using a machine learning (ML)-based approach to identify incorrect or incomplete matches. In particular, we use a supervised ML algorithm based on the Python package dedupe[2] to train a model for estimating the probability of a correct match on the basis of the following variables: the similarity of the affiliation name and company names, the classification of the affiliation being a firm or otherwise, the coincidence or similarity of the affiliation and company country (U.S.) and, separately, the city. Moreover, we added a number of additional predictors to increase the accuracy of a match. First, we use the total number of publications attributed to one affiliation in Scopus. We compare this number with the number of publications associated to the same affiliation that our search has yielded for a given focal firm. We compute the ratio of these two numbers. The intuition behind this exercise is that the two numbers should be balanced if the match is correct and may be largely unbalanced if it is not correct[3]. Second, we use the ratio between the total number

[2] https://github.com/dedupeio/dedupe

[3] One illustrative example is the name "Laser Technology" (gvkey 25475), which has led to retrieving affiliations from several academic institutions with departments that contain a name component similar to "Laser Technology", such as those part of Fraunhofer Institute, Osaka University, and the Chinese Academy of Sciences. In this case, the total count of publications associated with the affiliation would be very large. On the contrary, in correspondence to the correct affiliation, "Laser Technology Inc.", the number of total publications would have to be identical, or at least more similar, to the number of publications returned by the search.

of publications retrieved for a company and the total R&D expenditure of the company. A disproportionate ratio could signal that the company has been matched to incorrect affiliations. Finally, we include interactions between all variables. We set up the algorithm to train the model on chunks of the sample of firms obtained from the deciles in total R&D expenditureof the company. This is akin to further considering that the relevance of some indicators may differ for different companies. For instance, the match in terms of location will be a more important indicator for small companies, while for large multinational companies, it may have very little predictive power. Dedupe returns examples recursively to classify them manually as correct or incorrect matches, focusing progressively on more ambiguous cases. We provide approximately 200 examples for each chunk. The algorithm returns a list of pairs of potential matches over the full sample with a probability score for each chunk. We bring back this information to our original list of companies-affiliations pairs.

Step 5: Cleaning matches: Firm-publication combinations

The probability score estimated according with the procedure above is used to trim out matches that can be considered correct or incorrect with high confidence. For the ambiguous matches we refrain from setting an arbitrary threshold of acceptance of matches and complete the matching exercise with extensive semi-manual cleaning. This additional cleaning is still supported by the information constructed, as detailed in the previous sections and existing benchmark data on companies publications. Specifically, we use as benchmarks scientific publication information provided by Arora, Belenzon, and Sheer (2020), which stems from the alternative raw data source of Thomson's Web of Science, and scientific publication numbers based on an earlier manual matching exercise of Compustat names with the Scopus database (Simeth and Cincera, 2016). We proceed as follows for defining whether the publication-firm matches are correct or not. Using information on the affiliation type, we impose the criterion that the name component causing the match actually relates to an affiliation classified as a company. This crucial automatic cleaning step ensures that false matches are excluded where an academic institution has a similar name component than our Compustat firms of interest. While this intervention has proven powerful in identifying false positive matches, the presence of publications matched due to other firms with similar name components poses a concern. Such cases are more difficult to detect. To identify false positives based on similar firm names, we first take advantage of various heuristics and plausibility checks using aggregate firm-level indicators. We create flags for differences in firm-level publication counts with the two publication benchmark indicators, unusual publication-R&D productivity ratios, and a low percentage of firms' publications that have i) a US affiliation and ii) a city overlap with the firm's headquarters location. A low geographic overlap can indicate foreign R&D locations as well as point to unrelated foreign companies with similar name components. Subsequently, based on the created flags, we screen manually potentially problematic cases and mark incorrect affiliation names to be removed when computing final publication counts. A prominent example for a firm with both many correct and incorrect matches is Johnson & Johnson. The firm name "Johnson" matches with records such as "Johnson Matthey Biomedical Research", "Johnson Controls", "Robert Wood Johnson", "Johnson Matthey", or "R.W. Johnson", among others. In such cases, we leverage on further information from corporate webpages to disambiguate correct matches from the incorrect ones.

A.2. Additional robustness tests

In this section we report the results of additional difference-in-difference regressions. The results are discussed in Sections 5.4.2 and 5.4.3 of the article.

- Table A1 shows the difference-in-difference regression results when conditioning on various broker house merger/closure characteristics.
- Table A2 shows the difference-in-difference regression results from miscellaneous robustness tests.
- Table A3 shows the difference-in-difference regression when collapsing the panel at the firm-event level.
- Table A4 shows the difference-in-difference regression results from a dynamic specification and when using alternative sample frames.

Table A1

Robustness tests for baseline DiD: Conditioning on merger/closure characteristics.

This table replicates the baseline difference-in-difference (DiD) regression results from Table 6, column 9, but splits the sample of brokerage house closures/mergers according to the following characteristics: depending on whether the broker disappeared during a crisis period (calendar years 2000, 2001, 2002, 2008 and 2009) or not (columns 1 and 2), depending on whether the broker is among the top 10 ranked by the number of firms that lose analyst or not (columns 3 and 4), depending on whether the broker disappearance is due to a merger or closure (columns 5 and 6), and depending on whether the broker is included in the Kelly and Ljungqvist (2012) list (columns 7 and 8). Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Crisis period?		Top 10 broker?		Merger?		KL?	
Sample	Yes	No	Yes	No	Yes	No	Yes	No
Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATED × POST	0.102**	0.165***	0.161***	0.097***	0.101**	0.130***	0.136***	0.077*
	(0.041)	(0.055)	(0.052)	(0.037)	(0.048)	(0.035)	(0.039)	(0.046)
POST	0.014	-0.001	-0.011	0.008	-0.009	0.003	-0.006	-0.008
	(0.020)	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.020)
TREATED	-0.001	-0.159*	-0.101*	0.015	0.073	-0.082*	-0.055	0.123*
	(0.055)	(0.086)	(0.052)	(0.049)	(0.070)	(0.042)	(0.045)	(0.073)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Number of obs.	64,484	49,536	41,815	72,205	62,341	51,679	80,809	33,211
Adjusted R^2	0.881	0.891	0.885	0.879	0.880	0.884	0.880	0.889

Table A2
Robustness tests for baseline DiD: Miscellaneous robustness tests.
This table replicates the baseline difference-in-difference (DiD) regression results from Table 6, column 9, but requires firms to be present in the sample over the entire pre- and post-treatment period (column 1), augments the specification by including industry × year fixed effects (column 2), uses scientific publication data from Web of Science instead of Scopus, and alternative clustering schemes (columns 4 though 6). Web of Science data are from Arora, Belenzon, and Sheer (2020). Industries are defined at the two-digit SIC code level. The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Test	Balanced Sample	Industry × Year FEs	WoS Data	Alternative clustering schemes		
Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)
TREATED × POST	0.116***	0.125***	0.088***	0.115***	0.115***	0.115***
	(0.042)	(0.029)	(0.032)	(0.014)	(0.022)	(0.022)
POST	-0.008	0.004	-0.001	0.002	0.002	0.002
	(0.011)	(0.009)	(0.008)	(0.008)	(0.005)	(0.008)
TREATED	-0.010	-0.034	-0.057	-0.038**	-0.038	-0.038
	(0.052)	(0.038)	(0.043)	(0.016)	(0.041)	(0.029)
Controls included	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Event fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year fixed effects	No	Yes	No	No	No	No
Clustered by firm	Yes	Yes	Yes	No	No	No
Clustered by event	No	No	No	No	Yes	No
Clustered by event × firm	No	No	No	No	No	Yes
Number of obs.	78,894	114,020	114,020	114,020	114,020	114,020
Adjusted R^2	0.883	0.886	0.879	0.880	0.880	0.880

Table A3
Robustness tests for baseline DiD: Collapsing the panel at the event-firm level.

This table replicates the baseline difference-in-difference (DiD) regression results, but collapses the panel at the event-firm level into two effective periods: before and after the coverage shock. The dependent variable, PUB^* is the natural logarithm of (one plus) the average number of scientific publications per year in the three-year window either preceding or following the coverage shock. The control variables are measured in the pre-event year (year -1) or the post-event year (year $+1$). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: PUB*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TREATED × POST	0.121***	0.134***	0.126***	0.125***	0.125***	0.125***	0.125***
	(0.038)	(0.037)	(0.037)	(0.037)	(0.037)	(0.023)	(0.024)
POST	0.049***	0.045***	0.023***	0.021	0.021	0.021	0.021
	(0.012)	(0.012)	(0.009)	(0.018)	(0.018)	(0.015)	(0.017)
TREATED	1.561***	-0.066	-0.029	-0.025	-0.025	-0.025	-0.025
	(0.120)	(0.045)	(0.042)	(0.042)	(0.042)	(0.045)	(0.031)
Controls included	No	No	No	No	Yes	Yes	Yes
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Event fixed effects	No	No	No	Yes	Yes	Yes	Yes
Clustered by firm	Yes	Yes	Yes	Yes	Yes	No	No
Clustered by event	No	No	No	No	No	Yes	No
Clustered by event × firm	No	No	No	No	No	No	Yes
Number of obs.	41,074	41,074	41,074	41,074	41,074	41,074	41,074
Adjusted R^2	0.060	0.928	0.930	0.930	0.930	0.930	0.930

Table A4
Robustness tests for baseline DiD: Dynamic effects and alternative sample frames.
This table presents the dynamic effects of the analyst coverage shock on scientific publications (column 1), and replicates the baseline difference-in-difference (DiD) regression results from Table 6 using alternative pre- and post-treatment contrasts (columns 2 through 8). Robust standard errors are clustered by firm (in parentheses). The sample includes the treated and control firms following the sample construction described in Section 4. See Table 6 for control variables and Table 1 for definitions. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Pre period	$[-3,-1]$	$[-4,-2]$	$[-5,-3]$	$[-3,-1]$	$[-3,-1]$	$[-2,-1]$	$[-4,-1]$	$[-5,-1]$
Post period	$[+1,+3]$	$[+1,+3]$	$[+1,+3]$	$[+2,+4]$	$[+3,+5]$	$[+1,+2]$	$[+1,+4]$	$[+1,+5]$
Dependent Variable: PUB	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATED × POST		0.141***	0.171***	0.138***	0.163***	0.079***	0.131***	0.156***
		(0.039)	(0.042)	(0.041)	(0.049)	(0.030)	(0.037)	(0.041)
POST		-0.002	-0.005	-0.003	-0.001	0.003	-0.001	0.001
		(0.012)	(0.015)	(0.012)	(0.014)	(0.010)	(0.008)	(0.008)
TREATED		-0.064*	-0.093**	-0.054	-0.065	-0.010	-0.051	-0.071**
		(0.039)	(0.037)	(0.039)	(0.040)	(0.041)	(0.037)	(0.035)
TREATED × SHOCK_{t-3}	-0.076							
	(0.049)							
TREATED × SHOCK_{t-2}	-0.019							
	(0.043)							
TREATED × SHOCK_{t-1}	0.023							
	(0.041)							
TREATED × SHOCK_{t+1}	0.067*							
	(0.039)							
TREATED × SHOCK_{t+2}	0.108***							
	(0.039)							
TREATED × SHOCK_{t+3}	0.137***							
	(0.042)							
Controls included	No	Yes	Yes	Yes	Yes	Yes	Yes	
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Event fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	
Number of obs.	114,020	107,736	99,900	109,792	104,678	78,963	144,582	169,521
Adjusted R^2	0.876	0.877	0.876	0.878	0.877	0.885	0.876	0.874