Informing the Market: The Effect of Modern Information Technologies on Information Production*

Meng Gao† Jiekun Huang‡

This Draft: December 2018

*We thank Heitor Almeida, Ferhat Akbas, Paul Brockman, Mara Faccio, Mariassunta Giannetti, Itay Goldstein, Kathleen Hanley, Jack He, David Hirshleifer, Jianfeng Hu, Paul Irvine, Tim Johnson, Eric Kelley, Paul Koch, Mathias Kronlund, Laura Li, Lei Li, Wei Li, Vikram Nanda, Neil Pearson, Joshua Pollet, Jeff Pontiff, Matt Spiegel, Avanidhar Subrahmanyam, Alexei Tchistyi, Paul Tetlock, Stijn Van Nieuwerburgh, Alex Wagner, Chengwei Wang, Rong Wang, Scott Weisbenner, Baozhong Yang, Liyan Yang, Hao Zhang, Haibei Zhao, Mengxin Zhao, Wei Zhu, and seminar participants at the 2017 China International Conference in Finance, 2018 SFS Cavalcade, 2018 Financial Intermediation Research Society (FIRS) Conference, the Securities and Exchange Commission, Kansas University, Lehigh University, Nanjing University, University of Illinois at Urbana-Champaign, University of St.Gallen, University of Zurich, and Xiamen University for comments and helpful discussions. Zhongnan Xiang and Jianzhang Lin provided excellent research assistance. We retain responsibility for any remaining errors.

†Department of Finance, University of Illinois at Urbana-Champaign; e-mail: menggao4@illinois.edu.

‡Department of Finance, University of Illinois at Urbana-Champaign; e-mail: huangjk@illinois.edu.

Informing the Market: The Effect of Modern Information Technologies on Information Production

Abstract

Modern information technologies have fundamentally changed how information is disseminated in financial markets. Using the staggered implementation of the EDGAR system in 1993–1996 as a shock to information dissemination technologies, we find evidence that internet dissemination of corporate disclosures increases information production by corporate outsiders. Trades by individual investors, especially those with access to the internet, become more informative about future stock returns following the EDGAR implementation. The amount and accuracy of information produced by sell-side analysts increase after the implementation. These results suggest that greater and broader information dissemination facilitated by modern information technologies improves information production.

JEL Classification: G12, G14

Keywords: Information production, internet, informational efficiency, individual investors, financial analysts

1 Introduction

A well-functioning securities market requires that a broad base of investors have access to corporate information and process such information to promote price efficiency and facilitate capital formation. The advent of modern information technologies has dramatically changed how information is disseminated in financial markets by making a large amount of information available to a broad base of financial market participants in real time at low costs. Investors nowadays can get immediate access to corporate disclosures as well as other market participants' opinions disseminated through the internet to gain insights into firms' fundamental value. In the past few decades, a series of regulatory changes have been made to make use of modern information technologies to improve the accessibility of information to the public. For example, the SEC launched the EDGAR system in 1993 to move corporate disclosure from the print era to the digital age, and in 2013 the SEC allowed public companies to use social media sites to announce key information to investors. Yet, despite the dramatic changes brought about by modern information technologies in the dissemination of information, the effects of modern information technologies on information production by market participants remain underexplored.

Modern information dissemination technologies can have two opposite effects on information production by corporate outsiders.[1] On the one hand, more timely and extensive dissemination of information facilitated by modern information technologies may crowd out information production by market participants. This may arise because of at least three reasons. First, when public information is widely disseminated (i.e., more investors become informed about the information), prices may reveal more of the information (Grossman and Stiglitz, 1980). Since information processing takes time, the advantage of becoming an information processor decreases, resulting in reduced intensity of information processing

[1] Greater dissemination facilitated by modern information technologies is expected to have a *direct* positive effect on investors' information and market efficiency. In this paper, we focus on indirect information production effects.

activities (e.g., Dugast and Foucault, 2018). Second, since public information can serve as a coordinating device for investors' beliefs, greater dissemination of public information may cause investors to overweight public information and underweight private information. This may reduce stock price efficiency when the precision of private information is high (Morris and Shin, 2002; Amador and Weill, 2010). For example, Shiller (2006) argues that mass dissemination of information by the media may negatively impact the efficiency of asset prices by creating similar thinking among large groups of people, causing "an avoidance of individual assessment of quantitative data". Third, the availability of large amounts of information may create an information overload problem (e.g., Barber and Odean, 2001; Shapiro and Varian, 1999), reducing the attention allocated to information processing. These considerations suggest that the advent of modern information technologies may dampen the incentive to produce information and therefore reduce pricing efficiency.

On the other hand, there are at least two reasons for a crowding-in effect. First, modern information technologies can reduce the cost of accessing corporate disclosures and extracting value-relevant information from the disclosures, which may induce greater intensity of information production by market participants. Other things equal, the net profit information producers derive from producing information increases as the cost of information production declines (see, e.g., Verrecchia, 1982; Kim and Verrecchia, 1994). As Verrecchia (1982) argues, "[a]s technological improvements permit more information to be obtained at the same cost, traders' increased information acquisition results in prices revealing more information." Second, greater dissemination of corporate disclosures can reduce the uncertainty traders face by allowing stock prices to reflect more of the information contained in the disclosures, which may cause traders to acquire and trade on information about other fundamentals of the firms (Goldstein and Yang, 2015). Thus, greater dissemination of information facilitated by modern information technologies may increase the incentives of market participants to produce information and, as a result, improve pricing efficiency.

Therefore, the net effect of modern information technologies on information production is ultimately an empirical question. In this paper, we investigate this question by exploiting the staggered implementation of the EDGAR system in 1993–1996 as a shock to information dissemination technologies. Before the implementation of EDGAR in 1993, publicly traded corporations had to transmit multiple paper copies of filings to the SEC, and the three public reference rooms of the SEC (in Washington DC, New York, and Chicago) were the ultimate sources of these filings. The SEC introduced the EDGAR system in February 1993 to enable companies to file electronically to facilitate the dissemination of information to the public in a timely manner. Importantly, the SEC required that all public companies began filing to EDGAR in 10 discrete groups, with companies in the first group starting to file on EDGAR in April 1993 and companies in the last group starting in May 1996. Thus, the staggered nature of the implementation of the EDGAR system provides a set of counterfactuals for how information production would have changed in the absence of a change in information dissemination technologies and so allows us to disentangle the effect of information technologies on information production from other confounding factors. For an omitted variable to explain our findings, it would have to affect different groups of companies at discrete points in time as specified in the phase-in schedule.

In this paper, we focus on information production by two groups of market participants, namely individual investors and sell-side financial analysts, for two reasons. First, both individual investors and sell-side analysts play the role of information producers in the financial markets. Specifically, there is growing evidence suggesting that individual investors produce information about stocks (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013, 2017).[2] Since EDGAR makes corporate filings, which were

[2] As Kaniel, Liu, Saar, and Titman (2012) and Kelley and Tetlock (2013) argue, individuals' trades may contain information because, while each individual investor may have only noisy information, aggregating the information through the trades of a large number of individuals may result in signals that are relatively precise. In addition, individuals might be especially well positioned to exploit private information through their trades, because they tend to trade in small quantities and are not subject to the agency problems,

particularly costly to obtain for individual investors before the implementation, readily accessible on the internet, it might have a relatively large impact on individual investors, especially those with internet access. There is also a large literature on the role of sell-side financial analysts as information intermediaries in the stock market (see, e.g., Healy and Palepu, 2001, for a comprehensive review of this literature). Second, for both groups, we can directly observe their behavior at a relatively high frequency, which enables us to construct proxies of information production around specific points in time. In particular, we use the trading data from a large discount brokerage database (the LDB dataset) used by Barber and Odean (2000) and analyst forecasts data from I/B/E/S database.[3] More important for our purposes, the LDB dataset allows us to identify investors with access to the internet who are directly affected by the EDGAR shock.

Using a comprehensive set of firms covered in the phase-in schedule of the EDGAR system, we find evidence suggesting that the crowding-in effect dominates the crowding-out effect for both individual investors and sell-side analysts. Specifically, we find that individual investors' net buying following an earnings announcement of a stock becomes more informative about future stock returns after the stock becomes subject to mandatory filing on EDGAR. The economic magnitude is nontrivial. For example, a one-standard-deviation increase in net buying by individual investors during the 20 trading days post-announcement is associated with 1.093 percentage points higher subsequent three-month cumulative abnormal returns after the stock becomes an EDGAR filer than before, which is economically nontrivial considering that the three-month CAR has a mean of 0.592% and a standard deviation of 21.668%. Importantly, we are able to identify which investors have access to the internet based on whether they placed a trade through the internet in the past. While internet users account for only 12% of the investors in our sample, the

career concerns, or liquidity constraints that institutional managers typically face.

[3] We do not examine information production by institutional investors, because the 13F institutional holdings data, commonly used in institutional investor studies, provide only quarterly snapshots of institutions' holdings and hence do not allow us to infer institutions' trades at a relatively high frequency in a specific window.

increase in stock return predictability after the EDGAR implementation is driven primarily by trades placed by these investors. We also find evidence suggesting that the increase in trade informativeness post-EDGAR is driven mainly by investors that are presumably more skilled in information production. These results suggest that the crowding-in effect dominates the crowding-out effect, thereby resulting in more information production by individual investors, especially those with ready access to the internet.

Turning to sell-side analysts, we find evidence suggesting that both the amount and accuracy of information produced by sell-side analysts increase following the EDGAR implementation. Specifically, the number of analysts covering a firm increases and the forecast accuracy of analysts improves after the firm becomes subject to mandatory filing on EDGAR. In terms of economic magnitudes, the average firm experiences an increase of 0.223 analysts post-EDGAR, which is large considering that the mean and standard deviation of the number of analysts covering a firm are 2.489 and 3.922, respectively. Similarly, the average firm experiences an increase of 0.00138 in analysts' forecast accuracy, representing 15.1% (1.7%) of the mean (standard deviation) of the variable. Perhaps more important, stock market responses to analysts' revisions become significantly stronger after the firm becomes an EDGAR filer, suggesting that the market perceives analyst research as more informative. These results are consistent with the crowding-in effect dominating the crowding-out effect for sell-side analysts.

We conduct several additional tests to assess the robustness of our results. First, to address the concern that assignment to groups is not random, we construct a control sample using a propensity-score matching approach. Specifically, for each firm that switches from being a non-filer to an EDGAR filer in a given month, we identify a non-switching firm that has statistically the same size, book-to-market, firm age, profitability, leverage, R&D, etc. We find that the above results continue to hold, suggesting that the observed effects are not driven by firm characteristics that are associated with assignment to groups. Second, we include cohort-specific time trends as additional controls in the regressions. In this case, the identification of the effects of the EDGAR implementation comes from whether the implementation leads to deviations from preexisting cohort-specific trends. We find that the observed effects continue to hold with the inclusion of these time trends. Third, our results are robust to redefining the post-EDGAR period for the first four groups of firms to start from January 1994 when the EDGAR system became publicly available to internet users without additional charges. Last, we conduct a falsification test using a period preceding the actual EDGAR implementation. We find insignificant changes in information production during this period, suggesting that the parallel trends assumption is likely to hold in our setting.

Last but not least, we examine the effect of the EDGAR implementation on trading volume and stock price efficiency. We find that the trading volume of individual investors in our sample increases post-EDGAR and internet users account for a disproportionately large fraction of the increase. Using three inverse measures of pricing efficiency, namely stock price synchronicity (Morck, Yeung, and Yu, 2000), the absolute value of stock return autocorrelation, and the standard deviation of the pricing error of Hasbrouck (1993), we find evidence that the EDGAR implementation improves stock price efficiency. These results are again consistent with the crowding-in effect dominating the crowding-out effect.

As the first paper to exploit the staggered timing of the implementation of the EDGAR system, our study highlights the impacts of advances in information dissemination technologies on information production in financial markets. Our findings have important policy implications. Government regulations that aim to promote the accessibility of corporate disclosures, such as earnings reports and other corporate releases, to a broad base of investors in real time are likely to enhance the resource allocation role of financial markets by increasing the supply of information by corporate outsiders.

The rest of the paper is organized as follows. Section 2 provides a review of related research as well as background information on the implementation of EDGAR. Section 3 describes the data and summary statistics. Section 4 presents the empirical results, and Section 5 concludes.

2 Literature Review and Institutional Background

2.1 Related literature

Our paper contributes to three strands of literature. The first is the theoretical literature on costly information acquisition in financial markets. As noted in the introduction, existing theories provide ambiguous predictions regarding whether greater information dissemination and lower information acquisition costs brought about by modern information technologies crowd in or crowd out information production by market participants (e.g., Grossman and Stiglitz, 1980; Verrecchia, 1982; Kim and Verrecchia, 1994; Morris and Shin, 2002; Amador and Weill, 2010; Goldstein and Yang, 2015, 2019; Dugast and Foucault, 2018).[4] By exploiting the staggered implementation of the EDGAR system as plausibly exogenous shocks to information dissemination technologies, our paper provides evidence suggesting that greater and broader dissemination of fundamental information facilitated by modern information technologies positively impacts information production by corporate outsiders. Our findings are consistent with the crowding-in effect dominating the crowding-out effect, highlighting technological advances in information dissemination as a contributing factor to the informational efficiency of stock prices.

The second literature our paper is related to is a growing literature examining the effects of the information dissemination process on financial market outcomes. A number of studies examine the effects of regulatory shocks to corporate disclosure, such as Regulation

[4] See Goldstein and Yang (2017) for a thorough review of the theoretical literature on the effects of information disclosure on market quality and information production in financial markets.

Fair Disclosure (Reg FD) and the adoption of the eXtensible Business Reporting Language (XBRL), both of which, like the EDGAR implementation, aim to level the playing field among investors. Reg FD has been shown to reduce the informational advantage of financial analysts (e.g., Gintschel and Markov, 2004; Agrawal, Chadha, and Chen, 2006), reduce bid-ask spreads (Eleswarapu, Thompson, and Venkataraman, 2004), increase cost of capital (Duarte, Han, Harford, and Young, 2008), and increase the volume of small trades (Bushee, Matsumoto, and Miller, 2004).[5] Studies on XBRL adoption show that it leads to higher bid-ask spreads and lower volume of small trades (Blankespoor, Miller, and White, 2014) and lower stock return synchronicity (Dong, Li, Lin, and Ni, 2016). Recent studies investigate the role of the media on information production (see, e.g., Engelberg and Parsons, 2011; Dougal, Engelberg, Garcia, and Parsons, 2012; Peress, 2014). For example, Engelberg and Parsons (2011) use extreme weather events as exogenous shocks that disrupt the delivery of daily newspapers to identify the causal impact of media coverage on investor trading. Using newspaper strikes as shocks to information dissemination by the media, Peress (2014) finds evidence that the media improve stock pricing efficiency. Our paper adds to this literature by focusing on a technological/regulatory shock, namely the implementation of the EDGAR system, that significantly increases the accessibility of corporate disclosures to a broad base of investors.

Last, our paper is related to the empirical literature on the role of corporate outsiders as information producers in financial markets. Recent studies find evidence suggesting that individual investors produce information about stocks (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013, 2017). For instance, Kaniel, Liu, Saar, and Titman (2012) show that intense buying (selling) by individual investors in the 10 days prior to an earnings announcement predicts large positive (negative) abnormal returns following the earnings announcement. Coval, Hirshleifer, and Shumway (2005), Ivković and Weisbenner (2005), and Ivković, Sialm, and Weisbenner (2008) use the LDB dataset and

[5] See Beyer, Cohen, Lys, and Walther (2010) for a thorough review of the studies.

find evidence suggesting that some individual investors possess an informational advantage about stocks. Also, it has been well established that sell-side financial analysts are among the most important information intermediaries in the stock market (see, e.g., Bhushan, 1989; O'Brien and Bhushan, 1990; Lang and Lundholm, 1996; Healy and Palepu, 2001). Our paper contributes to this literature by focusing on the effect of a plausibly exogenous shock to information dissemination technologies on information production by corporate outsiders. Our findings highlight the importance of timely and broad dissemination of information in influencing the extent of information production by individual investors and financial analysts.

The implementation of the EDGAR system is relatively underexplored. The closest paper to ours is Asthana, Balsam, and Sankaraguruswamy (2004), who use TAQ data to identify small trades (i.e., trades less than $\$5,000$) and show in a univariate setting that the correlation between net buying based on signed small trades around 10-K filings and subsequent short-term (i.e., five-day) stock returns increases when the 10-K reports are filed electronically through EDGAR for the first time.[6] Unlike our paper, Asthana et al. (2004) treat the implementation of the EDGAR system as a one-time shock and do not exploit the staggered timing of the implementation. Moreover, since trade size does not necessarily provide a good indicator for whether the trader is an individual or institution (Hirshleifer, Myers, Myers, and Teoh, 2008), a possible explanation of their results is that EDGAR filings attract the attention of institutional investors who split orders and make small trades to minimize the price impacts (Bernhardt and Hughson, 1997). Therefore, their evidence does not establish that *individuals'* trades become more informative after the implementation of EDGAR. By analyzing actual trades of individual investors using the LDB data, our paper provides direct tests of the informativeness of

[6]Asthana et al.'s (2004) inference in the multivariate setting that small trades become more informed when firms initiate filings to EDGAR is invalid, because their regressions include interaction terms combining an indicator for initial EDGAR filers and changes in market capitalization, which only allows for testing the interaction effect but not the main effect of the EDGAR implementation.

individual investors' trades around the EDGAR implementation. The LDB data also enable us to identify individual investors that have access to the internet, which allows for sharper identification.[7]

2.2 The implementation of the EDGAR system

Prior to the implementation of EDGAR in 1993, public firms had to transmit multiple paper copies of filings to the SEC by mail, by courier, or by personal delivery. These paper copies of filings would then be filed in the SEC public reference rooms for public viewing after being reviewed by the SEC examiners. Thus, the three locations of the public reference rooms (in Washington DC, New York, and Chicago) were the ultimate source of corporate disclosures for the investing public. Since the paper filings can be inspected by one reader at a time, the limited availability of paper copies for each filing (typically one or two copies at each location) makes it hard for the information to reach a large audience. Moreover, the large volume of filings being filed with the SEC makes it difficult for the investing public to find and analyze specific data. For example, a *New York Times* (1982) article quotes reference room users as saying that,"[i]t's just incredible the number of problems you can run into trying to find something you need. [...] The place can be a zoo." To make things worse, because of the difficulty of maintaining a huge volume of paper filings, "files are often misplaced or even stolen." Accessing the filings could be difficult even for financial professionals. For example, in a *New York Times* (1994a) article, a financial analyst writes that "our group experimented once with going directly to the S.E.C. offices for copies of corporate 10-K filings. We got a daylong runaround and did not manage to acquire the reports that we needed before the S.E.C. offices closed for the

[7]Further, because Asthana et al. (2004) focus on short-term stock returns, they cannot distinguish between an information effect (in which electronic dissemination of information facilitates information acquisition by market participants) and an attention effect (in which greater dissemination of information through EDGAR causes investors to respond in a naïve fashion). In contrast, our paper provides cleaner tests of the information story by examining relatively long-term stock returns. Our paper also provides corroborating evidence on the information effect from financial analysts.

day."

To meet the objective of providing information to the public in a timely and efficient manner, the SEC developed an automated system, the Electronic Data Gathering, Analysis and Retrieval (EDGAR) system, for electronic submission of company filings. The main goal of EDGAR was to enable companies to file electronically to facilitate the dissemination of information to the public in real time. By disseminating information through the internet, the EDGAR system increases the accessibility of corporate filings and thus significantly reduces corporate outsiders' information acquisition costs. Moreover, corporate outsiders can more readily process information in electronic filings than in paper filings, e.g., by using the search function to locate specific information in an electronic document.

On February 23, 1993, the SEC issued rules requiring corporate filings be transmitted electronically to EDGAR. These rules specified a phase-in schedule for all public firms to begin filing to EDGAR. Specifically, the rules categorized public firms into 10 groups and each group was phased in at different times.[8] Companies in the first group, i.e., Group CF–01, had to commence mandated electronic filing to EDGAR in April 1993, and those in the last group, i.e., Group CF–10, became EDGAR filers in May 1996. The time-lapse between the starting date of one group and that of the next group ranges from three to six months. Figure 1 plots the number of firms that are subject to mandatory filing through EDGAR at each point in time from January 1993 through December 1996. Appendix A provides a timetable for the implementation of the EDGAR system.

[Insert Figure 1 about here]

Because of the staggering of the different groups over time, firms in the sample are both treatment and control firms. For example, firms in Groups CF–02 through CF–10 serve

[8] We filed a Freedom of Information Act request to the SEC for information on how companies are assigned to different groups. The SEC responded that their staff "conducted a thorough search of the SEC's various systems of records, but did not locate or identify any information responsive to [the] request."

as the control firms when firms in Group CF–01 switch from being non-EDGAR filers to EDGAR filers in April 1993, and firms in Group CF–01 as well as those in Groups CF–03 through CF–10 serve as the control firms when firms in Group CF–02 become subject to mandatory filings to EDGAR in July 1993. Thus, the staggered implementation of the EDGAR system mitigates the concern that the phase-in schedule may coincide with other firm-level shocks that may affect information production by corporate outsiders. In other words, for an omitted variable to explain our findings, it would have to affect different groups of companies at discrete points in time as specified in the phase-in schedule. Also, it is unlikely that the phase-in schedule is designed in such a way that it anticipates changes in information production up to three years into the future, which casts doubt on reverse-causality stories.[9]

3 Data and Summary Statistics

We retrieve the list of firms on the phase-in schedule for the implementation of the EDGAR system from Appendix B of SEC Release No. 33-6977 (released on February 23, 1993) and code a firm as being subject to mandatory filing to EDGAR based on the phase-in schedule.[10] The list provides the firm name, the Central Index Key (CIK), and group number (from 1 through 10). We match the companies on the list to Compustat by CIK and company name. We are able to match $5,212$ firms with common stocks traded on

[9] We address the concern that the assignment to groups may be correlated with firm characteristics in Section 4.3.

[10] According to the Release, the SEC may, in its discretion, grant or deny a request by a firm to participate in a phase-in group other than the group assigned in the phase-in schedule. Since a firm's decision to change the timing of the implementation may be endogenous to its information environment, we use the pre-specified timing instead of the actual timing of the implementation. It is worth noting that if the actual implementation date of a firm is different from that specified in the phase-in schedule, it will result in misclassifications in our coding and bias against finding significant results. As revealed in subsequent SEC documents, about 3% of our sample firms participate in a phase-in group other than the group assigned in the phase-in schedule in the Release. For robustness, we repeat our main tests using the actual timing of the implementation. The results, reported in Table IA-1 in the Internet Appendix, are qualitatively unchanged, and if anything, the coefficient of interest is generally slightly larger than that obtained in the baseline regressions.

the NYSE, NASDAQ, and AMEX that are on the phase-in schedule and have financial information available in Compustat as of January 31, 1993, i.e., the month-end immediately before the release of the rules regarding the EDGAR implementation. We thus exclude firms that go public after the release of the rules to eliminate the possibility that the timing of IPOs could be endogenous to the implementation. For most of our analysis, we focus on quarterly earnings announcements since they are accompanied by mandatory disclosure of quarterly financial results. Our sample period starts in April 1991 (i.e., two years before the starting date of the first batch of EDGAR filers) and ends in May 1998 (i.e., two years after the starting date of the last batch).

We obtain trading data from the large discount brokerage database used by Barber and Odean (2000), which cover the trades by 77,795 households between 1991 and 1996. The dataset is particularly appropriate for assessing the impact of internet dissemination of information on individual investors' trading decisions, because about a quarter of the investors in the dataset reside in California, which was one of the states with the highest rates of internet penetration in the early years of the internet (e.g., Greenstein, 1998). Therefore, individual investors in our sample may be more tech savvy and better positioned to take advantage of the internet technology than the average individual investor in a general sample.

We use the informativeness of individual investors' trades about subsequent stock returns to capture their information production activities. If investors produce information about a stock that is not yet incorporated into stock prices and trade on such information, their trades in the stock should be positively correlated with the subsequent stock returns.[11] We focus on individuals' trades in a 20-trading-day window immediately fol-

[11] Theories on costly information acquisition posit that there should be an "equilibrium degree of disequilibrium" (Grossman and Stiglitz, 1980) in the sense that stock prices gradually reflect informed investors' information so that investors who expend costly effort to produce information receive compensation. Therefore, positive (negative) abnormal returns subsequent to investors' buy (sell) trades can be viewed as the compensation for the aggregate information production by these investors.

lowing quarterly earnings announcements.[12] Since earnings announcements are followed by the release of financial statements that are critical for assessing the fundamental value of the firms (Kim and Verrecchia, 1994), we expect that investors should be especially active in processing financial statements into actionable information when they are released. We calculate net buying by individual investors during the first 20 trading days following an earnings announcement (i.e., from day +1 to +20, with day 0 being the earnings announcement date) as the total number of shares bought by individual investors during the period minus the total number of shares sold by individual investors during the same period normalized by the total number of shares outstanding. While individual investors' trades may be driven by non-informational factors such as liquidity shocks and behavioral biases, aggregating the trades of a large number of investors can result in relatively precise signals about the information that investors possess insofar as the non-informational factors are not systematically correlated.

We compute cumulative abnormal returns (CARs) following the trading window (i.e., starting from day +21) as the sum of daily DGTW characteristics-adjusted returns. We consider four holdings horizons, i.e., three months (i.e., 63 trading days from day +21 to +83), six months (i.e., 126 trading days from day +21 to +146), 12 months (i.e., 252 trading days from day +21 to +272), and 18 months (i.e., 378 trading days from day +21 to +398). The use of relatively long holding horizons enables us to focus on the permanent change in stock prices due to information effects and minimize the noise introduced by non-informational factors such as temporary price pressure and liquidity effects. If information disseminated through EDGAR attracts investor attention and increases unin-

[12] Ideally, one would like to look at a window immediately following the release of quarterly reports (i.e., 10-Qs). However, the filing dates of these reports are not readily available before the implementation of EDGAR, which is why we focus on a window following earnings announcements. To guide our choice of the length of the window, we retrieve the filing dates of 10-Qs of our sample firms that are available on EDGAR and compute the time lag between a quarterly earnings announcement and the filing of the corresponding 10-Q report. The time lag has a median of 17 calendar days and a 95th percentile of 29 calendar days (about 20 trading days), suggesting that the release of quarterly reports is likely to occur within a 20-trading-day window immediately following earnings announcements for the vast majority of our sample firms.

formed trading by these investors (e.g., Barber and Odean, 2001, 2008), one may expect short-run, but not long-run, return predictabilities of investors' trades. Thus, focusing on relatively long holding horizons provides a cleaner test of the information story. Panel A of Table 1 shows that individual net buying has a mean of 0.034% and a standard deviation of 3.560%. The three-month (six-month) cumulative abnormal returns starting from the 21st day post-announcement have a mean of 0.592% (1.189%) and a standard deviation of 21.668% (32.062%).

Since EDGAR makes information publicly accessible through the internet, it may have a direct impact on information production by investors who have access to the internet. We use a novel approach to infer investors' access to the internet. We classify investors into internet users and non-users by making use of the information on the channel through which investors place trades (i.e., by phone or internet). Internet users are those that placed a trade through the internet in the past and non-users are otherwise. About 12.049% of the investor-month observations are classified as internet users.[13] We then calculate net buying by internet users and non-users separately. The means for post-announcement net buying by internet users and non-users are 0.008% and 0.025%, and the standard deviations are 1.388% and 3.115%, respectively.

To distinguish investors that trade primarily on raw publicly released information from those that engage in information production, we adapt the reliance on public information (RPI) measure proposed by Kacperczyk and Seru (2007) to identify individual investors that rely primarily on publicly available information. Specifically, RPI of an individual investor is the R^2 of the regression of the investor's net buying of a stock on a day on changes in financial analysts' recommendations on the stock in the last 60 trading days. We obtain stock recommendation data from First Call. We run the regression for each investor-month using the investor's trades in the past 12 months. We exclude investor-

[13] According to the Current Population Survey conducted in 1994 (the earliest year in which internet access is being surveyed), about 11.4% of the U.S. households owned a personal computer with a modem.

months that have less than five trades in the past 12 months. We classify investor-months into a high and a low group based on the median RPI in that month. The means for post-announcement net buying by high- and low-RPI investors are -0.003% and 0.004%, respectively, and the standard deviations are 1.605% and 2.032%, respectively.

Panel A of Table 1 also shows the summary statistics for the control variables, including the decile rank of earnings surprises, total assets, book-to-market ratio, firm age, prior stock return, profitability (ROA), leverage, institutional ownership, etc. Earnings surprise is measured as the difference in EPS before extraordinary items between the current quarter and the same quarter of the previous year normalized by stock price (following Jegadeesh and Livnat, 2006). Firm age is the number of years since the first trading date on CRSP (with 1925 as the earliest possible year).

We retrieve quarterly earnings forecasts made within 90 days of the quarterly earnings report date from I/B/E/S. We construct three measures to capture information production by sell-side analysts. The first is the number of analysts following a firm, calculated as the number of quarterly earnings forecasts made by distinct analysts. The second is the forecast accuracy of analysts, calculated as the negative of the absolute value of the difference between the actual earnings per share and the median analyst forecast normalized by stock price (following Lang, Lins, and Miller, 2003). The third is market responses to analyst revisions. The idea is that if analyst revisions contain information that is not yet reflected in stock prices, the market should react positively (negatively) to upward (downward) revisions. We calculate analyst revision as the difference between two consecutive quarterly earnings forecasts of an analyst for the same stock-quarter scaled by stock price (following Clement and Tse, 2003). We calculate cumulative abnormal returns during a three-day window around the revision (i.e., from -1 to $+1$, with day 0 being the earnings revision date) as the sum of daily DGTW characteristics-adjusted returns.

Panel B of Table 1 shows the summary statistics of the analyst sample. The mean and standard deviation of the number of analysts following a firm are 2.489 and 3.922, respectively. The mean and standard deviation of forecast accuracy are −0.009 and 0.079, respectively. The mean revision is −0.00184 and the mean revision CAR is −0.232%.

[Insert Table 1 about here]

4 Empirical Results

4.1 Informativeness of individual investors' trades

Our first test examines the effect of the EDGAR implementation on individual investors' trade informativeness. Anecdotal evidence suggests that individual investors make use of EDGAR to acquire information in the early years when the system was launched, as illustrated by the following two quotes:

> Investors Alliance, a personal-investment club in Fort Lauderdale, Fla., for example, downloads the 10 megabytes or so of new S.E.C. material posted daily on the Internet and makes it available to users of its electronic bulletin board system. This saves individual members of the alliance from having to seek out the data themselves. (*New York Times*, 1994b)
>
> Edgar contains enough financial documents to cause fibrillations of delight in the heart of any number cruncher. From the search screen you simply enter a company name, and Edgar returns with an interactive list of 10-Ks, 10-Qs, 8-Ks, X-17A-5s, and the whole gamut of other disclosure forms. [...] Jules Garfunkel, of Morristown, New Jersey, says he uses Edgar regularly to track down the financial fundamentals of companies whose securities he might consider buying. Recently he used the system to get information about Intel.

Edgar gave him the financial statements and balance sheets he was looking for. (*Fortune Magazine*, 1995)

Since the implementation of the EDGAR system changes how corporate disclosures are disseminated in the financial markets, we focus on the informativeness of individual investors' trades *following* the release of corporate disclosures. As mentioned above, earnings announcements are accompanied by the release of financial statements, which are of crucial importance to investors in evaluating the fundamental value of the firms (Kim and Verrecchia, 1994). Therefore, trades during the period following earnings announcements are likely to be motivated by informational reasons rather than other considerations.[14] If greater and broader information dissemination enables individual investors to produce information that is not yet incorporated into prices (i.e., when the crowding-in effect dominates), their trades in a firm's stock following earnings announcements should become more informative about future stock price movements after the firm becomes an EDGAR filer. On the other hand, if the crowding-out effect dominates the crowding-in effect, we should expect that individual investors' trades become less informative following the EDGAR implementation.

We construct a firm-quarter panel and run the following regression:

$$CAR_{i,q} = c_i + c_q + \beta_1 \times Netbuy_{i,q} \times Post\text{-}EDGAR_{i,q} + \beta_2 \times Netbuy_{i,q} + \beta_3 \times Post\text{-}EDGAR_{i,q} \left[+ \gamma \times \mathbf{X}_{i,q-1} + \vartheta \times Netbuy_{i,q} \times \mathbf{X}_{i,q-1} \right] + \varepsilon_{i,q}, \quad (1)$$

where $CAR_{i,q}$ is the cumulative DGTW-adjusted abnormal returns of stock i during a three-, six-, 12-, or 18-month window starting from the 21st trading day after quarter q's

[14] Individuals' trades may be motivated by non-informational reasons such as liquidity shocks, hedging, taxes, and behavioral biases, which may explain the observation that the overall performance of individual investors' trades is insignificant or even negative (e.g., Barber and Odean, 2000). Restricting the analysis to trades placed following earnings releases, therefore, allows us to focus on a period during which there are public information releases that may prompt individual investors to process and trade on information.

earnings announcement;[15] $Netbuy_{i,q}$ is the net buying by individual investors in stock i during the 20-trading-day period immediately following the earnings announcement, $Post\text{-}EDGAR_{i,q}$ is an indicator that equals one if the firm-quarter is subject to mandatory filing on EDGAR; c_i and c_q are firm and year-quarter fixed effects, respectively; and $\mathbf{X}_{i,q-1}$ is a vector of lagged firm characteristics that are commonly used to predict stock returns, including the decile rank of earnings surprises, firm size, book-to-market ratio, firm age, past stock return, ROA, leverage, and so on. The firm fixed effects and year-quarter fixed effects control for time-invariant differences across firms and aggregate fluctuations in stock returns over time, respectively. Since the time-varying firm characteristics are likely affected by the EDGAR implementation, controlling for the terms in brackets, i.e., firm characteristics and their interaction terms with $Netbuy_{i,q}$, might confound the estimates of the effects of the implementation on the informativeness of individuals' trades (Angrist and Pischke, 2009, pp. 64–66). We therefore run all of our regressions with and without the terms in brackets. We cluster standard errors by firm and by year-quarter to account for likely correlation in errors (Petersen, 2009). The coefficient on the interaction term combining $Netbuy_{i,q}$ and $Post\text{-}EDGAR_{i,q}$ captures the incremental effect of filings to EDGAR on the informativeness of individuals' trades. If the crowding-in effect dominates the crowding-out effect, we should expect the coefficient to be positive and significant. On the other hand, if the crowding-out effect dominates the crowding-in effect, we should expect a negative and significant coefficient on the interaction term.

Panel A of Table 2 reports the baseline results for all trades by our sample of individual investors. The coefficient on the interaction term, $Netbuy \times Post\text{-}EDGAR$, is positive

[15] Using future stock returns allows us to get closer to detecting information production effects. Specifically, if individual investors simply trade on *raw* publicly released information by corporations or other *public* information sources such as the media, one might expect a weaker predictability of their trades for subsequent long-run stock returns after the EDGAR implementation. This arises because as more investors become informed about the same public information, competition among these homogeneously informed investors leads to faster incorporation of the information into the prices, giving rise to weaker return predictability (Holden and Subrahmanyam, 1992; Back, Cao, and Willard, 2000).

and significant in all specifications.[16] Notably, the magnitude of the coefficient estimates increases only slightly when we lengthen the holding period. For example, the coefficient is 0.307 when three-month CARs are used, as compared to 0.389 when 18-month CARs are used. This pattern suggests that much of the information possessed by individual investors is impounded into stock prices during the first three months. In terms of economic magnitudes, model 1 shows that a one-standard-deviation increase in net buying by individual investors during the 20 trading days post-announcement is associated with 1.093 percentage points higher subsequent three-month cumulative abnormal returns after the stock becomes an EDGAR filer than before, which is economically nontrivial considering that the three-month CAR has a mean of 0.592% and a standard deviation of 21.668%. Furthermore, the sum of the coefficient on $Netbuy \times Post\text{-}EDGAR$ and that on $Netbuy$ is positive and significant in all specifications, indicating that individual investors' trades during the post-period are based on information not yet impounded into stock prices.[17] These results are consistent with the hypothesis that the crowding-in effect dominates the crowding-out effect, resulting in more information production by individual investors.[18] As alluded to in the introduction, the information individual investors trade on could be information extracted from firm disclosures through costly effort (e.g., Kim and Verrecchia, 1994) and/or information on the dimensions of firm fundamentals that are not covered by the disclosures (Goldstein and Yang, 2015). We note that the production of both types of information can contribute to more efficient stock prices, although it is beyond the scope

[16] The coefficient on *Post-EDGAR* itself is negative and significant, indicating that, for the subset of stocks with *zero net buying by individual investors* (about 1.5% of the sample), the subsequent returns tend to be lower after EDGAR implementation than before.

[17] The coefficient on $Netbuy$ is positive and largely insignificant, suggesting that individual investors' trades do not contain information about future stock returns before the EDGAR implementation. This result is consistent with prior findings that the abnormal *gross* returns earned by individual investors are generally insignificantly different from zero (e.g., Barber and Odean, 2000).

[18] Since the EDGAR implementation leads to more efficient stock prices (as we show in Section 4.4), in the absence of an increase in information production by individual investors, the predictive ability of individual investors' trades for subsequent stock returns should decrease post-EDGAR because more of the informed individuals' information is incorporated into stock prices during the period when the trades occur. Therefore, the observed increase in the predictive ability of individuals' trades suggests that the increase in information production by individual investors induced by the EDGAR implementation dominates the accompanying increase in pricing efficiency.

of our paper to empirically disentangle them.

If individuals trade based on firm characteristics that are correlated with future stock returns and, for some reason, this tendency becomes stronger after the EDGAR implementation, this might explain a positive and significant coefficient on $Netbuy \times Post\text{-}EDGAR$. We thus include the terms in brackets in Eq. (1), i.e., lagged firm characteristics that are likely to be correlated with future stock returns and their interaction terms with $Netbuy$, as additional controls in the regression. Panel B of Table 2 shows that the magnitude of the coefficient estimates on the interaction term is largely unchanged, suggesting that the results are not explained by observed firm characteristics driving individual investors' trades. The stability of the coefficients also suggests that the likelihood of unobservable firm characteristics driving the results is low (Oster, 2017).[19]

We exploit heterogeneity across investors in terms of internet access to shed light on the sources of the increase in the informativeness of individual investors' trades after the implementation of the EDGAR system. Panel C of Table 2 replaces net buying by all individual investors with that by internet users and that by non-users separately. The coefficient on the interaction term combining the post-EDGAR indicator and net buying by internet users is positive and significant in all four specifications, whereas that combining the post-EDGAR indicator and net buying by non-users is smaller in magnitude and generally insignificant. The difference in the two coefficients is significant at the 5% level when we look at 12- and 18-month abnormal returns. Thus, although internet users account for a relatively small fraction (i.e., about 12%) of the sample of investors, they account for the bulk of the observed increase in the informativeness of individual investors' trades post-EDGAR. This finding strengthens the interpretation that the EDGAR implementation enables individual investors, especially those with ready access to the internet,

[19] We conduct an omitted variable bias test suggested by Oster (2017). The results, reported in Table IA-2 of the Internet Appendix, show that the identified set for the true effect in all of our main tests safely excludes zero under reasonable assumptions, suggesting that the likelihood that omitted variable bias drives our conclusions is low.

to acquire and process information.

The implementation of EDGAR can have a *direct* positive effect on information dissemination by making corporate filings readily available to investors, which might enable individual investors to trade on raw (i.e., unprocessed) public information. Thus, the above positive effect might be driven by individual investors' use of public information, rather than information production. The inclusion of firm-level controls, which are public information, and their interactions with the net buying measure mitigates this concern to some extent. Nevertheless, in an attempt to more reliably disentangle the direct effect from the indirect (information production) effect, we replace net buying by all individual investors with that by high-RPI investors and that by low-RPI investors separately and reestimate the regressions. Panel D of Table 2 shows that the coefficient on the interaction term combining the post-EDGAR indicator and net buying by low-RPI investors is positive and significant in three out of four specifications, whereas that combining the post-EDGAR indicator and net buying by high-RPI investors is insignificant. The difference in the two coefficients is significant at the 5% level when we use three- and six-month abnormal returns. To the extent that low-RPI investors are more skillful in information production (Kacperczyk and Seru, 2007), these results suggest that the observed effects are likely indirect effects.

[Insert Table 2 about here]

We then examine how the effect of the EDGAR implementation on the informativeness of individuals' trades varies across stocks facing different levels of information asymmetry. If a firm faces a low level of information asymmetry (e.g., it is heavily covered by financial analysts and the news media), the implementation is likely to have a relatively modest effect on the informativeness of individual investors' trades because information about such firms is available from other sources. On the other hand, the implementation of EDGAR is likely to significantly improve the information environment of firms that face a high level of information asymmetry in the equity market, i.e., those whose information is otherwise costly to obtain, by increasing the amount of information that investors can access at low costs. We thus expect the effect of the EDGAR implementation on the informativeness of individuals' trades to be driven mainly by firms with a high level of information asymmetry. We use analyst coverage and market capitalization to proxy for the level of information asymmetry. We measure analyst coverage and market cap as of the quarter-end immediately before the earnings announcements. We classify a firm as opaque if the firm has no analyst coverage and the market capitalization of the firm is below the median. We interact the opaque indicator with the main variables, i.e., $Netbuy$, *Post-EDGAR*, and their interaction term, and repeat the regressions.

The results, reported in Panel A of Table 3, show that the triple interaction term combining $Netbuy$, *Post-EDGAR*, and the opaque indicator is positive and significant at conventional levels across all specifications. These findings provide suggestive evidence that the EDGAR implementation enables individual investors to produce novel information about stocks facing a high level of information asymmetry. Since opaque firms generally receive little coverage from financial analysts and the media, these results suggest that the observed increase in the informativeness of individual investors' trades is unlikely to be driven by individual investors trading on information obtained from these alternative sources.

It might be tempting to speculate that since markets must clear, individual investors as a whole gain an informational advantage over institutional investors post-EDGAR. This reasoning, however, is invalid, because our data only cover a *subset* of individual investors and hence do not allow us to draw conclusions regarding individual investors as a whole. Nevertheless, to explore how the effect varies with the presence of institutional investors, we construct an indicator for stocks with high institutional presence, which equals one if the firm has above-the-median level of institutional ownership and above-the-median number of institutional shareholders and zero otherwise. We interact our main variables, i.e., *Netbuy*, *Post-EDGAR*, and their interaction term, with the indicator for high institutional presence, and reestimate the regressions. The results, reported in Panel B of Table 3, show that the coefficient on the triple interaction term combining *Netbuy*, *Post-EDGAR*, and the high institutional presence indicator is negative across all specifications and significant at 1% or 5% levels when three- and six-month abnormal returns are used, indicating that the effect becomes weaker in stocks with a high presence of institutional investors. In contrast, the coefficient on the interaction term combining *Netbuy* and *Post-EDGAR* is positive and significant at conventional levels across all specifications, suggesting that the observed increase in individual investors' trade informativeness is concentrated in stocks with a low institutional presence.

These results are consistent with the view that institutions are better positioned to produce information about stocks than individual investors (see, e.g., Sias and Starks, 1997; Boehmer and Kelley, 2009; Hendershott, Livdan, and Schürhoff, 2015). Since stocks with a greater presence of institutional investors are likely associated with greater information production by these more sophisticated investors, the incremental benefit to individual investors from expending costly effort to produce information about such stocks is likely low even after the implementation of EDGAR. Therefore, the increase in the trade informativeness of individual investors in our sample post-EDGAR does not seem to occur at the expense of institutional investors. Instead, it appears that EDGAR enables individual investors in our sample to gain an informational advantage over other individual investors that presumably have less access to the internet. Since, as mentioned above, individual investors in our sample come disproportionately from areas with high internet penetration, they are likely to benefit to a greater extent from the EDGAR implementation than individual investors in general.

[Insert Table 3 about here]

The main premise of our analysis is that individual investors make use of EDGAR filings during our sample period. The two quotes at the beginning of this section indicate that they do. To provide more direct evidence on individual investors' use of EDGAR, we obtain data on the server logs of the EDGAR system hosted by New York University (`https://town.hall.org/govt/tuttle/stats_edgar_domain_073095.html`). The log provides a breakdown of WWW access to the system by domain name during the week ending July 30, 1995. The system received over $100,000$ server requests during the week.[20] Since investors can access the internet at home through internet service providers (ISPs), visits from domains associated with ISPs are likely to be made by individual investors. We thus manually identify domain names that are registered to ISPs during our sample period such as America Online (AOL), CompuServe, Prodigy, and Netcom. Because many of the domain names have become defunct or transferred to other entities, we use the Wayback Machine to access the earliest archived web pages of the domain names to ascertain whether the registrars of the domain names are ISPs. We identify 683 domain names that are registered to ISPs. Requests by users of ISPs represent over 24.45% of the total number of requests and 31.39% of the total amount of data requested. For example, users of AOL sent $5,812$ requests for over 565 megabytes of filings to EDGAR during the week, representing 5.79% of the total number of requests and 10.99% of the number of bytes requested. These numbers suggest that individual investors account for a substantial fraction of the user base of EDGAR in during our sample period when there are few alternative sources of corporate filings.[21] It should be noted that these numbers likely underestimate the actual

[20] To put this number in perspective, Yahoo!, which was "[c]onsistently rated as one of the 10 most-visited websites", had an average of $500,000$ visits each day in 1995 (*New York Times*, 1995).

[21] Using EDGAR server logs from 2003 through 2012, Loughran and McDonald (2017) show that the number of requests for 10-Ks through EDGAR is surprisingly low. Their finding, however, does not necessarily invalidate the premise that EDGAR serves as an important conduit of information for investors. As Loughran and McDonald (2017) point out, alternative distribution channels that provide access to repackaged EDGAR filings have proliferated in more recent years, e.g., FreeEDGAR, EDGAR Online, EdgarScan, and Capital IQ, which may explain the low magnitude of requests on EDGAR itself during that period. In contrast, a search of the names of these alternative sources in the *New York Times* archives before December 1996 yields no results, suggesting that these sources are largely unavailable during our sample period.

usage of EDGAR filings by individual investors, because they do not include access through other protocols such as WAIS, FTP, Gopher, and email server. These numbers also do not include requests by individual investors from work or school (educational institutions account for 10.07% of the total number of requests and 14.22% of the number of bytes requested). Further, as the above quote from *New York Times* (1994b) suggests, filings obtained via a single download can be redistributed to many individual investors.

Our results on the informativeness of individual investors' trades do not necessarily contradict previous studies on individual investors' trading behavior and performance. For example, Barber and Odean (2000) show that individual investors' stock portfolios deliver largely insignificant abnormal gross returns and that high trading levels lead to worse performance, suggesting that individual investors exhibit overconfidence in trading. It is important to note that Barber and Odean's (2000) results are based on the unconditional performance of individuals' trades, whereas our paper focuses on the performance of individuals' trades during a period that is likely associated with information releases by companies. It is possible that while individual investors on average exhibit behavioral biases that adversely affect their trading decisions, there are times when individual investors process newly released information and trade on such processed information. Using the LDB dataset, Hirshleifer et al. (2008) find that individuals' trades during a five-day window following earnings announcements do not predict subsequent abnormal stock returns. Since they focus on a relatively short window during which the corresponding financial reports are typically not made available to the public, Hirshleifer et al. (2008) likely capture naïve reactions of individual investors to earnings news. In contrast, the use of a 20-trading-day window following earnings announcements enables us to focus on trades that are likely to be motivated by the processing of information contained in newly released financial statements.

4.2 Sell-side analyst research

While the crowding-in effect dominates the crowding-out effect for individual investors, it is not immediately clear whether the same conclusion would hold for sell-side analysts. On the one hand, analysts may already have access to some corporate filings such as 10-K and 10-Q reports before the advent of EDGAR, thereby resulting in limited effects of the EDGAR implementation on analysts' information production activities. On the other hand, the implementation can positively affect analysts' information production because of at least two reasons.[22] First, corporate filings other than 10-Ks and 10-Qs, such as those on insider trades (Form 4) and material corporate events (Form 8-K), may contain important information for forecasting future performance, but are not readily accessible to analysts. The limited availability of paper copies of these filings and the difficulty in maintaining physical copies (e.g., they may easily get lost, misplaced, or even stolen) make it costly and time-consuming to gain access to these filings (see Section 2.2 for more discussions). The implementation of EDGAR thus substantially eases access to all corporate filings. Second, the EDGAR implementation can lower information processing costs by making information searchable and retrievable from anywhere connected to the internet at any time.

To examine the effect of the EDGAR implementation on information production by sell-side financial analysts, we conduct two sets of tests. The first examines analyst coverage and analyst forecast accuracy at the firm-quarter level, and the second examines market responses to analyst forecast revisions using analyst-level revision events. Specifically, for

[22] As an analogy, academic researchers had access to physical copies of journals and working papers through brick-and-mortar libraries before the advent of electronic article repositories such as JSTOR and SSRN. Most of us would agree that these online repositories greatly facilitate information acquisition and processing for researchers, so much so that they have largely replaced physical libraries as the sources of information for academic research.

the first test, we construct a firm-quarter panel and run the following regression:

$$Analyst\ research_{i,q} = c_i + c_q + \theta_1 \times Post\text{-}EDGAR_{i,q} \left[+\gamma \mathbf{X}_{i,q-1}\right] + \varepsilon_{i,q}, \tag{2}$$

where $Analyst\ research_{i,q}$ is either the number of analysts making quarterly forecasts for stock i's quarter q earnings per share or the forecast accuracy of analysts; $Post\text{-}EDGAR_{i,q}$ is an indicator that equals one if the firm-quarter is subject to mandatory filing on EDGAR; c_i and c_q are firm and year-quarter fixed effects, respectively; and $\mathbf{X}_{i,q-1}$ includes the same set of firm characteristics used in Eq. (1) except the decile rank of earnings surprises. We use the regression without the term in brackets, i.e., lagged firm characteristics, as the baseline specification, because these characteristics are likely affected by the EDGAR implementation. For example, the implementation may enable firms to grow faster and fetch higher valuation, which in turn could affect the quantity and quality of analysts' research. Therefore, including time-varying firm characteristics could confound the estimate of the total impact of the implementation. Year-quarter fixed effects absorb common variation over time in analysts' information production (Veldkamp, 2005). We again cluster standard errors by firm and by year-quarter. If the crowding-in effect dominates the crowding-out effect, the coefficient on the *Post-EDGAR* indicator should be positive and significant. On the other hand, if the crowding-out effect dominates the crowding-in effect, we should expect a negative and significant coefficient.

The results, reported in Table 4, show that both the number of analysts covering a firm and the forecast accuracy of analysts increase significantly after the firm becomes subject to mandatory filing on EDGAR. The coefficient estimates on the *Post-EDGAR* indicator remain reasonably stable when we control for firm size, market-to-book, prior stock return, ROA, and other variables that could be correlated with analysts' research, suggesting that the effect of the EDGAR implementation is largely independent of that of time-varying firm characteristics. In terms of economic magnitudes, the baseline specification in column

1 shows that the average firm experiences an increase of 0.223 analysts post-EDGAR, which is large considering that the mean and standard deviation of the number of analysts covering a firm are 2.489 and 3.922, respectively. Similarly, column 3 shows that the average firm experiences an increase of 0.00138 in analysts' forecast accuracy, representing 15.1% (1.7%) of the mean (standard deviation) of the variable.

[Insert Table 4 about here]

If financial analysts are able to produce more accurate information after a firm becomes an EDGAR filer, the market should respond more strongly to analysts' forecasts. Thus, our second set of tests investigates the impact of the EDGAR implementation on market responses to analysts' forecast revisions. We estimate the following regression using the sample of revision events:

$$CAR_{i,a,d} = c_{a,q} + c_{i,q} + \kappa_1 \times Revision_{i,a,d} \times Post\text{-}EDGAR_{i,q} + \kappa_2 \times Revision_{i,a,d} + \varepsilon_{i,a,d}, \quad (3)$$

where $CAR_{i,a,d}$ is the three-day cumulative DGTW-adjusted abnormal returns of stock i around analyst a's forecast revision on day d; $Revision_{i,a,d}$ is the price-scaled changes in analyst a's earnings forecasts for stock i on day d; $Post\text{-}EDGAR_{i,q}$ is an indicator that equals one if the firm-quarter is subject to mandatory filing on EDGAR; $c_{a,q}$ and $c_{i,q}$ are analyst $\times$ year-quarter and firm $\times$ year-quarter fixed effects, respectively. In some specifications, we include firm fixed effects and the same set of firm characteristics as used in Eq. (2) instead of firm $\times$ year-quarter fixed effects. In the most stringent specification, we include both analyst $\times$ year-quarter and firm $\times$ year-quarter fixed effects, which completely absorb time-varying analyst attributes (e.g., experience of the analyst, areas of expertise, and broker resources) and time-varying firm attributes (e.g., prior performance, information asymmetry, and ownership structure). The inclusion of analyst $\times$ year-quarter fixed effects forces identification of the coefficient on the interaction term to come from variations across firms covered by a given analyst in a given quarter, enabling us to compare market reactions to forecast revisions of the same analyst in the same quarter across stocks that are EDGAR filers and those that are not. Standard errors are three-way clustered to allow for arbitrary correlation within firm, analyst, and year-quarter.

Table 5 reports the results. In all specifications, the coefficients on the interaction terms are positive and highly significant, suggesting that the market perceives analysts' research on a firm as more informative after the firm becomes an EDGAR filer. The economic magnitudes are large: for example, model 4 shows that for a one-standard-deviation increase in the magnitude of the revisions, the three-day CAR is 0.416 percentage points ($= 0.00715 \times 0.582$) higher after the EDGAR implementation than before. This finding suggests that the implementation improves the information content of analysts' research presumably by facilitating their access to all corporate filings and lowering information processing costs. This result, combined with the above finding of increased forecast accuracy, provides suggestive evidence that the implementation enables analysts to produce new information rather than to simply disseminate public information.

Overall, the two sets of regressions in Tables 4 and 5 show consistent patterns in the effect of the EDGAR implementation on information production by sell-side analysts. These results suggest that greater dissemination of information facilitated by modern information technologies increases both the quantity and quality of sell-side analyst research. These findings are in contrast to the negative effect of Reg FD on analyst research, which is expected given that EDGAR lowers information acquisition and processing costs for all market participants, including analysts, whereas Reg FD curtails selective disclosure to analysts.

[Insert Table 5 about here]

4.3 Robustness checks

In this subsection, we perform a number of additional tests to assess the robustness of the main results.

Propensity-score matching. As suggested by the cost-benefit analysis in SEC Release No. 33-6944 (Proposed Rulemaking for EDGAR System) published in July 1992, compliance costs are a key factor that determines the assignment to groups.[23] Since these costs do not vary with firm size, the regulation would impose a relatively greater burden on smaller firms if they were required to start filing electronically at the same time as larger firms. As the Release states, "the proposed changes would affect persons that are small entities, as defined by the Commission's rules." Therefore, small companies are assigned to the last phase-in group, i.e., Group CF–10. According to the Release, "the Commission has designed the EDGAR system to accommodate small entities to the greatest degree possible while still carrying out its mandate to develop a system for the electronic dissemination of information to the public. Small companies will be the last group phased into the system, allowing them to take advantage of the substantial body of experience gained by those who precede them."[24]

To address the concern about nonrandom assignment of groups, we use a propensity-score matching approach. We first construct a sample of control firms that are statistically identical to firms that switch from being a non-filer to an EDGAR filer. Specifically, for each month in which a group of firms start to become subject to mandatory filings to EDGAR, we create a cohort consisting of treatment firms, i.e., firms that switch from

[23] As stated in the Release, "It is expected that the proposed changes to paper and electronic format related requirements for submitting documents to the Commission may result in some costs to filers and investors exceeding that which would have been incurred under a continuing paper-based system. Filers without the equipment to submit electronically would need to purchase such equipment or hire agents to submit electronically on their behalf. Those subscribing to the EDGAR electronic mail/bulletin board service also will incur the cost of subscription and other attendant expenses."

[24] For robustness, we reestimate our main regressions excluding firms in the last phase-in group. The results, reported in Table IA-3 in the Internet Appendix, are qualitatively unchanged.

being a non-filer to an EDGAR filer in that month, and control firms, i.e., those that do not switch in that month or in the 12 months before or 12 months after. Note that a control firm can be an EDGAR filer or a non-filer as long as the firm retains that status during the 25-month period around the month under consideration. We then stack the 10 cohorts into a panel and run a logistic regression to predict whether a firm becomes treated. We use a comprehensive list of firm characteristics, including the full set of control variables in Eq. (2) as well as industry fixed effects and cohort fixed effects, as the explanatory variables. Column 1 of Table 6, Panel A, reports the results. Treatment firms tend to be older, smaller, more levered, and less growth-oriented than control firms. We use the predicted probabilities, or propensity scores, from this logit estimation and perform a one-to-one nearest-neighbor matching with replacement. After excluding the few observations outside the common support, we are able to match $4,194$ treatment firms to similar control firms. To assess the quality of matching, we repeat the logit regression on the propensity-score matched sample. The results, reported in column 2 in the same panel, show that none of the coefficients on the matching variables are statistically significant at conventional levels, suggesting that the matching process is effective in removing meaningful observable differences between the two groups of firms.

We compare the change in various information production proxies between treatment firms and matched control firms. We use the four quarters immediately before the switching event (i.e., quarters -4 through -1, with quarter 0 being the switching quarter) as the pre-period and a four-quarter period after the switching event (i.e., quarters $+3$ through $+6$) as the post-period. We skip the first two quarters immediately following the event to allow time for market participants to start processing information.

To test whether individual investors' trades in treatment stocks, relative to those in matched control stocks, become more informative about subsequent stock returns after the implementation than before, we pool the treatment and matched control stocks and regress the subsequent three-month cumulative abnormal returns on net buying by individual investors, an indicator for treatment stocks, an indicator for whether the observation is from the post-event period, and interaction terms for each of these variables. The coefficient on the triple interaction term is the difference-in-differences estimator comparing the change in trade informativeness between treatment and matched control firms. Panel B of Table 6 shows that the coefficient on the triple interaction term is 0.373 and significant at the 5% level, which is comparable to the magnitude obtained in our baseline specification in Table 2.

We conduct similar tests for analyst research. Panel B of Table 6 shows that, compared to matched control stocks, treatment stocks experience an increase in the number of analysts, forecast accuracy, and market responses to forecast revisions after the firms become EDGAR filers. The difference-in-differences estimates for these outcome variables are again significant at conventional levels with magnitudes similar to those obtained in the baseline specifications in Tables 4 and 5. These results mitigate the concern that the observed effects are driven by observable characteristics such as firm size and age that determine the assignment to groups.

[Insert Table 6 about here]

Controlling for group-specific time trends. It is possible that time trends in our outcome variables may be different across groups that become subject to filings to EDGAR at discrete points in time. To account for this possibility, we include group-specific time trends as well as their interactions with $Netbuy$ or $Revision$ as additional controls in the regressions (e.g., Angrist and Pischke, 2009, pp. 238). The identification of the effects of the EDGAR implementation thus comes from whether the implementation leads to deviations from preexisting group-specific trends. We report the regression results using the baseline specification for each test, i.e., Eqs. (1) through (3), in Table 7. The results show that the effects of the EDGAR implementation on various outcomes continue to be positive and significant and the magnitude of the effects is little changed by the inclusion of these trends. These results suggest that the observed effects are not driven by differential time trends across groups.[25]

[Insert Table 7 about here]

Ease of access to EDGAR filings. When the EDGAR system first got started, corporate filings on EDGAR were available electronically through Mead Data Central, a commercial data vendor, which provided access to the information for a fee (*New York Times*, 1993). The Internet Multicasting Service, a nonprofit organization, secured a National Science Foundation grant to New York University, which made EDGAR filings publicly accessible to internet users without additional charges starting from January 17, 1994. Therefore, for the first four groups of companies, there is an interim period when the filings are electronically filed but are available at a cost, which may limit the accessibility of these filings. We thus redefine the *Post-EDGAR* indicator for the first four groups to take the value of one if the firm-quarter is after January 17, 1994 and zero otherwise, and create a new variable, *Interim*, which takes the value of one if the firm-quarter falls in the interim period for the first four groups of companies and zero otherwise. About 1% of the firm-quarters in the sample are classified as being in the interim period.

Table 8 reports the results when we replace the original *Post-EDGAR* indicator with the redefined *Post-EDGAR* indicator and the *Interim* indicator. The results show that the effects of the redefined *Post-EDGAR* indicator on various outcomes continue to be positive and significant and the magnitude of the effects is little changed from that obtained using the baseline specifications. Interestingly, we find positive, although statistically

[25] To mitigate the concern that the results may be driven by differential time trends across firms of different sizes or ages, we reestimate all the regressions allowing firms of different sizes and ages to have different time trends. Specifically, we partition firms into size (age) deciles based on total assets (firm age) as of January 31, 1993 and include time trends specific to each size (age) group as well as their interactions with *Netbuy* or *Revision* in the regressions. Table IA-4 in the Internet Appendix shows that the results are qualitatively unchanged, suggesting that the observed effects are unlikely to be driven by firms of different sizes or ages exhibiting differential time trends.

insignificant, effects of the interim period on our information production proxies. For example, model 2 shows that the number of analysts covering a firm in the first four groups increases by 0.282 when the firm moves from the pre-EDGAR period to the interim period. The insignificant results may be due to the low statistical power of the test given that only about 1% of the observations are in the interim period.

[Insert Table 8 about here]

Pre-trends. Our identification strategy assumes that absent the EDGAR implementation, trends in information production are the same between firms that become mandatory EDGAR filers and firms that remain as non-filers. To verify the parallel-trends assumption, we repeat the tests using a period preceding the actual EDGAR implementation. We define pseudo-events as occurring two years prior to the actual implementation and restrict the sample for this test to firm-quarters during a four-year window before the implementation; thus none of the firm-quarters in this test actually switches during the four-year period. The "*Post-EDGAR*" indicator takes the value of one if the firm-quarter is in the two-year period after the pseudo-event dates and zero if it is in the two-year period before. If information production at these firms exhibits parallel trends, we should expect insignificant change in information production around these pseudo-events.

Table 9 reports the results from the falsification tests. The coefficients on our variables of interest are statistically insignificant and generally close to zero. For example, the coefficient estimates on the *Post-EDGAR* indicator are 0.085 and 0.00026, respectively, in the regressions of the number of analysts and forecast accuracy, as compared to 0.223 and 0.00138 in the baseline specifications in Table 4. These results show that there is little change in information production in the absence of shocks to information dissemination, suggesting that the parallel trends assumption is likely to hold in our setting.

Similarly, we conduct falsification tests by defining pseudo-events as occurring two years *after* the actual implementation. We again find no significant changes in information production around these pseudo-events (see Table IA-5 in the Internet Appendix). These results alleviate the concern that the observed effects may be driven by firm characteristics that are generally correlated with both the relative timing of the EDGAR implementation and changes in information production. For example, one possibility is that certain types of firms may attract information production earlier than others and such firms happen to be assigned to groups that are phased in early. This story, however, would not be able to explain why the observed change occurs only around the implementation dates but not before or after the implementation.

[Insert Table 9 about here]

Transitional filers. Prior to the mandatory phase-in of the EDGAR system starting in April 1993, the SEC tested the system by allowing volunteers to file electronically. These voluntary filers are assigned to Group CF–01 in the phase-in schedule and are referred to as "transitional" filers in the SEC release adopting the rules for the EDGAR implementation. Since transitional filers elect to switch to electronic filings on a voluntary basis, they are not required to submit *all* filings electronically before the mandatory phase-in (see SEC Release No. 33-6977). Also, transitional filers can choose not to file electronically at any time and submit all filings in paper format until mandated to file electronically. Once phased in, however, firms are required to submit all documents electronically and will not be permitted to file in paper absent a hardship exemption. Since the mandated phase-in to electronic filings limits the discretion of transitional filers in their filing decisions (i.e., whether to file electronically and, if so, what documents to file electronically), it still represents a shock to the dissemination of these firms' disclosures. Therefore, we include these transitional filers in the main tests. Nevertheless, to mitigate the concern that these firms drive the observed effects, we conduct a robustness check by excluding firms assigned to Group CF–01.

Table 10 reports the regression results using the baseline specification for each test when Group CF–01 firms are excluded. The effects of the EDGAR implementation on various outcomes continue to hold and the magnitude of the effects remains qualitatively unchanged. For example, models 2 and 3 show that the coefficient estimates for the post-EDGAR indicator in the regression of the number of analysts and forecast accuracy are 0.237 and 0.00155, respectively, as compared to 0.223 and 0.00138 obtained using the full sample of firms reported in Table 4.

[Insert Table 10 about here]

4.4 Other outcomes

In this subsection, we examine the effect of the EDGAR implementation on other outcomes, i.e., the trading volume of individual investors and stock pricing efficiency.

Trading volume of individual investors. If the crowding-in effect dominates the crowding-out effect for individual investors, individual investors not only should become more informed in their trading, but they should also trade more. In other words, since the EDGAR implementation increases information production by individual investors, these investors should trade more actively to exploit their informational advantage.

To test the effect of the implementation on individual investors' trading volume, we estimate a specification similar to Eq. (2) with trading volume by individual investors following earning announcements as the dependent variable. We measure trading volume as the total number of shares traded by our sample of individual investors (purchases plus sales) during the first 20 trading days following an earnings announcement, i.e., the same window we use to measure the informativeness of individual investors' trades, scaled by the number of shares outstanding. The results, reported in Panel A of Table 11, show that the coefficient on the post-EDGAR indicator is positive and statistically significant, suggesting that individual investors trade more actively in a stock after the filings of the stock are disseminated through the internet. In terms of economic magnitudes, the baseline specification in model 1 shows that the trading volume of individual investors in a stock increases by 0.394 basis points after the stock becomes an EDGAR filer, which is nontrivial considering that the mean and standard deviation of the trading volume are 2.408 and 5.504 basis points, respectively.[26] Combined with the above findings on the informativeness of individuals' trades, these results suggest that the EDGAR implementation increases both the informativeness of individuals' trades and the trading volume of these investors.

To shed light on the sources of the increase in trading volume post-EDGAR, we partition individual investors' trading volume in two ways. First, since the EDGAR implementation directly affects internet users, we decompose the variable into the trading volume of internet users and that of non-users and run separate regressions. Panel B of Table 11 shows that the coefficient on the post-EDGAR indicator is positive and statistically significant at the 1% level when the trading volume of internet users is used as the dependent variable, but it becomes insignificant for the trading volume of non-users. Comparing the magnitude of the coefficient between these two groups of investors indicates that internet users account for a disproportionately large fraction of the increase in the trading volume even though they represent only 12% of the investors in our sample. For example, columns 1 and 3 show that internet users account for 43.6% of the increase in trading volume (i.e., 0.171 for internet users and 0.221 for non-users). These results are consistent with the idea that the implementation of EDGAR enables investors with internet access to acquire and process information more effectively.

Second, to better understand whether the increase in trading volume post-EDGAR is driven by increased trading on public information (i.e., a direct effect of greater dissemination) or increased information production (i.e., an indirect net crowding-in effect), we

[26] These results are similar to the findings in Bushee, Matsumoto, and Miller (2004), who show that Reg FD leads to increased volume of small trades to the extent that small trades are likely placed by individual investors.

decompose the total trading volume of individual investors in our sample into that by high-RPI investors and that by low-RPI investors. Panel C of Table 11 shows that the coefficient on the post-EDGAR indicator is positive and statistically significant at the 1% level when we look at the trading volume of low-RPI investors, but it becomes insignificant for the trading volume of high-RPI investors. In terms of economic magnitude, while investors are evenly split into high- and low-RPI categories, low-RPI investors account for a large majority of the increase in trading volume post-EDGAR. For example, columns 1 and 3 show that 79.6% of the increase in trading volume is driven by low-RPI investors (i.e., 0.319 for low-RPI investors and 0.082 for high-RPI investors). Since low-RPI investors are likely more skilled in producing information (Kacperczyk and Seru, 2007), these results suggest that the observed increase in trading volume is more likely driven by increased information production than by increased trading on public information.

[Insert Table 11 about here]

Stock pricing efficiency. Since greater information dissemination facilitated by modern information technologies increases information production by corporate outsiders, it may lead to more efficient stock prices. To test this, we use three inverse measures of stock price efficiency, namely stock price synchronicity (Morck, Yeung, and Yu, 2000), the absolute value of stock return autocorrelation, and Hasbrouck's (1993) pricing error. Price synchronicity is the R-squared from the regression of a stock's daily return on the contemporaneous market return and industry return (following Chen, Goldstein, and Jiang, 2007). Durnev, Morck, Yeung, and Zarowin (2003) show that firms with low stock price synchronicity are associated with a stronger predictive ability of current stock returns for future earnings, suggesting that the current stock price reflects more information about future earnings. We compute stock return autocorrelation for a stock-month as the first-order autocorrelation coefficient for the daily stock return series. A lower absolute value of return autocorrelation implies more efficient stock pricing (e.g., Lo and MacKinlay, 1988).

To construct the pricing error measure, we first decompose the log transaction price p_t as $p_t = m_t + s_t$, where m_t is a random walk process representing the market efficient price conditional on all public information available at t; s_t is a zero-mean covariance-stationary process capturing the transient deviation of the transaction price from the efficient price due to factors such as inventory control by market makers, price discreteness, and temporary liquidity effects. The standard deviation of the pricing error, denoted as $\sigma(s_t)$, captures the extent to which the transaction price deviates from the efficient price and thus can be interpreted as an inverse measure of market efficiency. We follow Boehmer and Kelley (2009) to use a vector autoregressive (VAR) system to obtain estimates for s_t. Specifically, we use intraday transaction data from NYSE Trade and Quote (TAQ) data from 1993–1998 and Institute for the Study of Security Markets (ISSM) data from 1991–1992. We exclude stock-months with less than 200 transactions. We use trades and quotes during regular hours and discard overnight price changes. For all transaction, we only include transactions with positive prices, positive sizes, and positive bid and ask prices with bid minus ask being positive and less than 25% of the mid quote. To make the measure comparable across stocks and over time, we normalize the standard deviation of the pricing error by the standard deviation of the log transaction price and use this ratio as an inverse measure of pricing efficiency, i.e., $\mathit{PricingError} = \sigma(s_t)/\sigma(p_t)$. We construct the pricing error measure at a monthly frequency.

To test the effect of the implementation of EDGAR on stock pricing efficiency, we run the following regression:

$$\mathit{InverseEfficiency}_{i,m} = c_i + c_m + \theta_1 \times \mathit{Post\text{-}EDGAR}_{i,m} \left[+\gamma \mathbf{X}_{i,m-1}\right] + \varepsilon_{i,q}, \tag{4}$$

where $\mathit{InverseEfficiency}_{i,m}$ is one of the three inverse measures of information efficiency for stock i in month m; $\mathit{Post\text{-}EDGAR}_{i,m}$ is an indicator set to zero before the stock becomes subject to mandatory EDGAR filing and one afterward; c_i and c_m are firm and year-month fixed effects, respectively; and $\mathbf{X}_{i,q-1}$ is the same set of firm characteristics used in Eq. (2). We again run all of our regressions with and without the term in brackets, i.e., lagged firm characteristics. We cluster standard errors by firm and by year-month. If the EDGAR implementation increases pricing efficiency, we expect a significant and negative coefficient on the post-EDGAR indicator.

The results, reported in Table 12, show that the coefficient on the post-EDGAR indicator is negative and statistically significant across all six specifications, suggesting that the EDGAR implementation leads to more efficient stock pricing.[27] These results are similar to the findings in Dong, Li, Lin, and Ni (2016), who show that XBRL adoption reduces stock price synchronicity. The economic magnitude is nontrivial: for example, since the mean (standard deviation) of stock price synchronicity is 0.155 (0.158), model 1 shows that stock price synchronicity decreases by 5.2% (5.1%) relative to its mean (standard deviation) after the implementation of EDGAR. These results are consistent with our above findings of increased information production by corporate outsiders post-EDGAR, although they are also consistent a direct effect of greater dissemination.

[Insert Table 12 about here]

5 Conclusion

Modern information technologies have greatly facilitated the dissemination of information in financial markets. In this paper, we investigate the impact of internet dissemination of corporate disclosures on information production by corporate outsiders, namely individual investors and financial analysts. Using the staggered implementation of the EDGAR system in 1993–1996 as a shock to information dissemination technologies, we find evidence

[27] We run similar regressions to examine the effect of the EDGAR implementation on stock liquidity. Consistent with EDGAR improving pricing efficiency and reducing information asymmetry, we find evidence that the implementation of EDGAR leads to improved stock liquidity. We report the results in Table IA-6 in the Internet Appendix.

that greater information dissemination facilitated by modern information technologies increases information production by these two sets of market participants. Specifically, trades by individual investors in a stock become more informative about future stock returns after the stock becomes subject to mandatory filing on EDGAR. This effect is driven primarily by investors who have access to the internet and those who are more skilled in information production. We also find evidence that the trading volume of individual investors increases after a stock becomes an EDGAR filer. As for financial analysts, we find that both the amount and accuracy of information produced by sell-side analysts increase following the EDGAR implementation. Market responses to analyst revisions become stronger after firms start to file electronically on EDGAR. Furthermore, stock pricing efficiency improves after a firm becomes an EDGAR filer. Overall, these results suggest that advances in information technologies that facilitate greater and broader information dissemination improve information production and stock pricing efficiency.

This paper contributes to our understanding of the effects of modern information technologies on financial markets. Our findings suggest that regulations that aim at promoting the accessibility of corporate disclosures to a broad base of investors in real time at low costs are likely to enhance the resource allocation role of financial markets by increasing the supply of information by corporate outsiders. Given the profound effects of modern information technologies on stock pricing efficiencies, future research should investigate whether and, if so, how information technologies influence the real decisions of firms.

Appendix A: Timetable for Implementation of EDGAR Division of Corporation Finance Filings

April 26, 1993: Phase-in of Group CF–01.

July 19, 1993: Phase-in of Group CF–02.

October 4, 1993: Phase-in of Group CF–03.

December 6, 1993: Phase-in of Group CF–04.

August 1994: Phase-in of Group CF–05.

November 1994: Phase-in of Group CF–06.

May 1995: Phase-in of Group CF–07.

August 1995: Phase-in of Group CF–08.

November 1995: Phase-in of Group CF–09.

May 1996: Phase-in of Group CF–10.

References

[1] Agrawal, Anup, Saliba Chadha, and Mark A. Chen, 2006, Who is afraid of Reg FD? The behavior and performance of sell-side analysts following the SEC's fair disclosure rules, *Journal of Business* 79, 2811–2834.

[2] Amador, Manuel, and Pierre-Olivier Weill, 2010, Learning from prices: Public communication and welfare, *Journal of Political Economy* 118, 866–907.

[3] Angrist, Joshua D., and Jörn-Steffen Pischke, 2009, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, NJ.

[4] Asthana, Sharad, Steven Balsam, and Srinivasan Sankaraguruswamy, 2004, Differential response of small versus large investors to 10-K filings on EDGAR, *The Accounting Review* 79, 571–589.

[5] Back, Kerry, H. Henry Cao, and Gregory A. Willard, 2000, Imperfect competition among informed traders, *Journal of Finance* 55, 2117–2155.

[6] Barber, Brad, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

[7] Barber, Brad, and Terrance Odean, 2001, The Internet and the investor, *Journal of Economic Perspectives* 15, 41–54.

[8] Barber, Brad, and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

[9] Bernhardt, Dan, and Eric Hughson, 1997, Splitting orders, *Review of Financial Studies* 10, 69–101.

[10] Beyer, Anne, Daniel A. Cohen, Thomas Z. Lys, and Beverly R.Walther, 2010, The financial reporting environment: Review of the recent literature, *Journal of Accounting and Economics* 50, 296–343.

[11] Bhushan, Ravi, 1989, Firm characteristics and analyst following, *Journal of Accounting and Economics* 11, 255–274.

[12] Blankespoor, Elizabeth Gregory S. Miller, and Hal D. White, 2014, Initial evidence on the market impact of the XBRL mandate, *Review of Accounting Studies* 19, 1468–1503.

[13] Boehmer, Ekkehart, and Eric K. Kelley, 2009, Institutional investors and the informational efficiency of prices, *Review of Financial Studies* 22, 3563–3594.

[14] Bushee, Brian, Dawn A. Matsumoto, and Gregory S. Miller, 2004, Managerial and investor responses to disclosure regulation: the case of Reg FD and conference calls, *The Accounting Review* 79, 617–643.

[15] Chen, Qi, Itay Goldstein, and Wei Jiang, 2007, Price informativeness and investment sensitivity to stock price, *Review of Financial Studies* 20, 619–650.

[16] Clement, Michael B., and Senyo Y. Tse, 2003 Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *Accounting Review* 78, 227–249.

[17] Coval, Joshua D., David A. Hirshleifer, and Tyler Shumway, 2005, Can individual investors beat the market? Working paper.

[18] Dong, Yi, Oliver Z. Li, Yupeng Lin, and Chenkai Ni, 2016, Does information-processing cost affect firm-specific information acquisition? Evidence from XBRL adoption, *Journal of Financial and Quantitative Analysis* 51, 435–462.

[19] Dougal, Casey, Joseph Engelberg, Diego García, and Christopher A. Parsons, 2012, Journalists and the stock market, *Review of Financial Studies* 25, 639–679.

[20] Duarte, Jefferson, Han, Xi, Harford, Jarrad, and Young, Lance, 2008, Information asymmetry, information dissemination, and the effect of Regulation FD, *Journal of Financial Economics* 87, 24–44.

[21] Dugast, Jérôme, and Thierry Foucault, 2018, Data abundance and asset price informativeness, *Journal of Financial Economics* 130, 367–391.

[22] Durnev, Artyom, Randall Morck, Bernard Yeung, and Paul Zarowin, 2003, Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41, 797–836.

[23] Eleswarapu, Venkat, Rex Thompson, and Kumar Venkataraman, 2004, The impact of Regulation Fair Disclosure: trading costs and information asymmetry, *Journal of Financial and Quantitative Analysis* 39, 209–225.

[24] Engelberg, Joseph, and Christopher A. Parsons, 2011, The causal impact of media in financial markets, *Journal of Finance* 66, 67–97.

[25] Gintschel, Andreas, and Stanimir Markov, 2004, The effectiveness of Regulation FD, *Journal of Accounting and Economics* 37, 293–314.

[26] Goldstein, Itay, and Liyan Yang, 2015, Information diversity and complementarities in trading and information acquisition, *Journal of Finance* 70, 1723–1765.

[27] Goldstein, Itay, and Liyan Yang, 2017, Information disclosure in financial markets, *Annual Reviews of Financial Economics* 9, 101–125.

[28] Goldstein, Itay, and Liyan Yang, 2019, Good disclosure, bad disclosure, *Journal of Financial Economics* 131, 118–138.

[29] Greenstein, Shane, 1998, Universal service in the digital age: The commercialization and geography of US internet access, NBER Working Paper 6453.

[30] Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

[31] Hasbrouck, Joel, 1993, Assessing the quality of a security market: A new approach to transaction-cost measurement, *Review of Financial Studies* 6, 191–212.

[32] Healy, Paul M., and Krishna G. Palepu, 2001, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* 31, 405–440.

[33] Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2015, Are institutions informed about news? *Journal of Financial Economics* 117, 249–287.

[34] Hirshleifer, David, James N. Myers, Linda A. Myers, and Siew Hong Teoh, 2008, Do individual investors drive post-earnings announcement drift? Direct evidence from personal trades, *The Accounting Review* 83, 1521–1550.

[35] Holden, Craig W., and Avanidhar Subrahmanyam, 1992, Long-lived private information and imperfect competition, *Journal of Finance* 47, 247–270.

[36] Ivković, Zoran, and Scott Weisbenner, 2005, Local does as local is: Information content of the geography of individual investors' common stock investments, *Journal of Finance* 60, 267–306.

[37] Ivković, Zoran, Clemens Sialm, and Scott Weisbenner, 2008, Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis* 43, 613–656.

[38] Jegadeesh, Narasimhan, and Joshua Livnat, 2006, Revenue surprises and stock returns, *Journal of Accounting and Economics* 41, 147–171.

[39] Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, 2012, Individual investor trading and return patterns around earnings announcements, *Journal of Finance* 67, 639–680.

[40] Kelley, Eric K., and Paul C. Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229–1265.

[41] Kelley, Eric K., and Paul C. Tetlock, 2017, Retail short selling and stock prices, *Review of Financial Studies* 30, 801–834.

[42] Kim, Oliver, and Robert E. Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41–67.

[43] Lang, Mark. H., Karl V. Lins, and Darius P. Miller, 2003, ADRs, analysts, and accuracy: Does cross listing in the United States improve a firm's information environment and increase market value?, *Journal of Accounting Research* 41, 317–345.

[44] Lang, Mark H., and Russell J. Lundholm, 1996, Corporate disclosure policy and analyst behavior, *Accounting Review* 71, 467–492.

[45] Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41–66.

[46] Loughran, Tim, and Bill McDonald, 2017, The use of EDGAR filings by investors, *Journal of Behavioral Finance* 2, 231–248.

[47] Morck, Randall, Bernard Yeung, and Wayne W. Yu, 2000, The information content of stock markets: Why do emerging markets have synchronous stock price movements?, *Journal of Financial Economics* 58, 215–260.

[48] Morris, Stephen, and Hyun Song Shin, 2002, Social value of public information, *American Economic Review* 92, 1521–1534.

[49] *New York Times*, 1982, S.E.C. Data: Difficult hunt, May 19.

[50] *New York Times*, 1993, Plan opens more data to public, October 22.

[51] *New York Times,* 1994a, The S.E.C. and public filings, July 17.

[52] *New York Times*, 1994b, Need timely S.E.C. corporate filings? Look on Internet, June 19.

[53] *New York Times*, 1995, Making searches easier in the Web's sea of data, October 2.

[54] O'Brien, Patricia C., and Ravi Bhushan, 1990, Analyst following and institutional ownership, *Journal of Accounting Research* 28, 55–76.

[55] Oster, Emily, 2017, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business and Economic Statistics*, forthcoming.

[56] Peress, Joel, 2014, The media and the diffusion of information in financial markets: Evidence from newspaper strikes, *Journal of Finance* 69, 2007–2043.

[57] Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

[58] Shapiro, Carl, and Hal R. Varian, 1999, *Information rules: A strategic guide to the network economy*, Harvard Business Press, Boston, MA.

[59] Shiller, Robert, 2006, *Irrational Exuberance*, Crown Business Publishing, Danvers, MA.

[60] Sias, Richard W., and Laura T. Starks, 1997, Return autocorrelation and institutional investors, *Journal of Financial Economics* 46, 103–131.

[61] Veldkamp, Laura L., 2005, Slow boom, sudden crash, *Journal of Economic Theory* 124, 230–257.

[62] Verrecchia, Robert E., 1982, Information acquisition in a noisy rational expectations economy, *Econometrica* 50, 1415–1430.

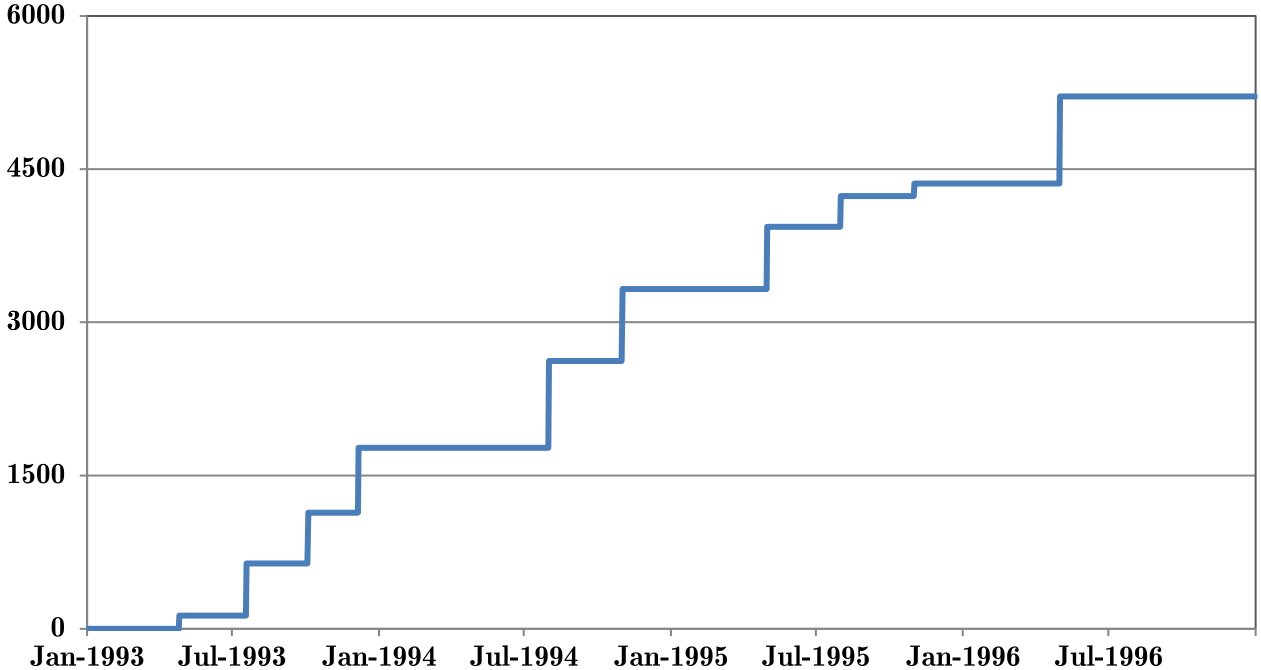

Figure 1. Staggered implementation of mandatory filing through EDGAR

This figure plots the number of firms that are subject to mandatory filing through EDGAR during the period from January 1993 through December 1996.

Table 1: Summary statistics

This table reports the summary statistics for the individual trading sample (Panel A) and analyst sample (Panel B). *Post-EDGAR* is an indicator that equals one after a firm-quarter becomes a mandatory EDGAR filer and zero otherwise. $Netbuy_{[+1,\ +20]}$ is the number of shares bought minus the number of shares sold by individual investors as a fraction of the number of shares outstanding during a 20-day window after an earnings announcement. *Netbuy by internet users*$_{[+1,\ +20]}$ and *Netbuy by non-users*$_{[+1,\ +20]}$ are similarly defined for individual investors with access to the internet and those without, respectively. We identify an investor as an internet user if she placed a trade through the internet in the past. $CAR_{[+21,\ +83]}$, $CAR_{[+21,\ +146]}$, $CAR_{[+21,\ +272]}$, and $CAR_{[+21,\ +398]}$ are cumulative DGTW-characteristics adjusted returns during a three-, six-, 12-, and 18-month window starting from the 21st day after an earnings announcement, respectively. *# of analysts* is the number of analysts making quarterly earnings forecasts in the I/B/E/S database for a stock in a given quarter. *Forecast accuracy* is negative of the absolute value of the difference between the actual earnings per share and the median analyst forecast normalized by stock price (following Lang, Lins, and Miller, 2003). *Revision* is the difference between two consecutive quarterly earnings forecasts of an analyst for the same stock-quarter scaled by stock price (following Clement and Tse, 2003). $CAR_{[-1,\ +1]}$ is the cumulative DGTW characteristics-adjusted returns during a three-day window around an earnings forecast revision by an analyst. *SUE decile rank* is the decile rank of earnings surprises, defined as the difference in EPS before extraordinary items between the current quarter and the same quarter of the previous year normalized by stock price (following Jegadeesh and Livnat, 2006). The rankings are from zero to nine, with zero (nine) representing firms in the bottom (top) decile of earnings surprises in a quarter. *Total assets* is the book value of assets of the firm. *Book-to-market* is the book value of common equity divided by the market value of common equity. *Prior stock return* is the buy-and-hold stock return during the past 12 months skipping the most recent month. *ROA* is the ratio of income before extraordinary items to book value of assets. *Book leverage* is the ratio of the book value of total debt to the book value of total assets. *Asset tangibility* is the ratio of net property, plant, and equipment to total assets. *Sales growth* is the percentage change in quarterly sales from four quarters earlier to the current quarter. *CapEx* is the ratio of capital expenditure to total assets. *R&D* is the ratio of R&D expenses to total assets. *Institutional ownership* is the number of shares held by institutional investors as a fraction of the number of shares outstanding. All variables are winsorized at the 0.1% and 99.9% levels to minimize the effect of outliers.

Panel A: Summary statistics for the individual trading sample

	# of obs	Mean	Standard deviation	P10	Median	P90
Main variables						
Post-EDGAR	29,364	0.431	0.495	0.000	0.000	1.000
$\text{Netbuy}_{[+1,\ +20]}$ (%)	29,364	0.034	3.560	-1.702	-0.019	1.771
Netbuy by internet $\text{users}_{[+1,\ +20]}$ (%)	29,364	0.008	1.388	-0.202	0.000	0.200
Netbuy by non-users $_{[+1,\ +20]}$ (%)	29,364	0.025	3.115	-1.427	0.000	1.459
Netbuy by high-RPI $\text{investors}_{[+1,\ +20]}$ (%)	29,364	-0.003	1.605	-0.452	0.000	0.457
Netbuy by low-RPI investors $_{[+1,\ +20]}$ (%)	29,364	0.004	2.032	-0.587	0.000	0.570
$\text{CAR}_{[+21,\ +83]}$ (%)	29,364	0.592	21.668	-21.529	-0.071	22.778
$\text{CAR}_{[+21,\ +146]}$ (%)	29,364	1.189	32.062	-29.907	-0.054	32.337
$\text{CAR}_{[+21,\ +272]}$ (%)	29,364	2.881	49.607	-41.635	0.652	48.226
$\text{CAR}_{[+21,\ +388]}$ (%)	29,364	4.319	62.061	-50.425	1.559	60.346
Control variables						
SUE decile rank	29,364	4.556	2.874	1.000	5.000	8.000
Total assets ($ mil)	29,364	2532.060	7039.280	24.647	258.121	5585.520
Book-to-Market	29,364	0.652	0.483	0.186	0.548	1.232
Firm age (years)	29,364	17.932	15.675	4.000	13.000	35.000
Prior stock return	29,364	0.284	0.609	-0.301	0.158	1.000
ROA	29,364	0.027	0.127	-0.062	0.043	0.130
Book leverage	29,364	0.495	0.231	0.186	0.490	0.833
Asset tangibility	29,364	0.295	0.230	0.034	0.239	0.667
Sales growth	29,364	0.176	0.376	-0.093	0.100	0.487
CapEx	29,364	0.077	0.084	0.005	0.054	0.162
R&D	29,364	0.046	0.090	0.000	0.000	0.154
Institutional ownership	29,364	0.395	0.221	0.090	0.392	0.694

Panel B: Summary statistics for the analyst sample

	# of obs	Mean	Standard deviation	P10	Median	P90
Main variables						
Post-EDGAR	103,929	0.512	0.500	0.000	1.000	1.000
# of analysts	103,929	2.489	3.922	0.000	1.000	8.000
Forecast accuracy (x100)	56,447	-0.915	7.943	-1.688	-0.199	0.000
Revision (%)	358,443	-0.184	0.715	-0.642	-0.042	0.193
$\text{CAR}_{[-1,\ +1]}$ (%)	358,443	-0.232	4.999	-4.790	-0.146	4.549
Control variables						
Total assets ($ mil)	103,929	2368.320	12143.340	12.676	152.707	3630.230
Book-to-Market	103,929	0.723	23.129	0.160	0.576	1.368
Firm age (years)	103,929	16.725	14.805	3.000	12.000	33.000
Prior stock return	103,929	0.240	0.777	-0.364	0.125	0.857
ROA	103,929	-0.001	0.191	-0.123	0.032	0.115
Book leverage	103,929	0.539	0.308	0.193	0.517	0.909
Asset tangibility	103,929	0.292	0.239	0.024	0.231	0.678
Sales growth	103,929	0.356	7.398	-0.124	0.087	0.472
CapEx	103,929	0.083	0.943	0.002	0.048	0.158
R&D	103,929	0.045	0.273	0.000	0.000	0.133
Institutional ownership	103,929	0.319	0.236	0.030	0.285	0.660

Table 2: Staggered implementation of EDGAR and the informativeness of trades by individual investors

This table reports regression analysis of the impact of EDGAR on the informativeness of individual investors' trades about subsequent stock returns. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-, six-, 12-, or 18-month window starting from the 21st day after an earnings announcement. Panel A estimates baseline specifications using net buying by all individual investors in our sample. Panel B includes firm characteristics, i.e., *SUE decile rank*, *Log of total assets*, *Book-to-Market*, *Log of firm age*, *Prior stock return*, *ROA*, *Book leverage*, *Asset tangibility*, *Sales growth*, *CapEx*, *R&D*, and *Institutional ownership*, and the interaction terms combining net buying and each of these firm characteristics as additional controls. The coefficients of the control variables are omitted for brevity. Panel C decomposes the net buying measure into net buying by internet users and that by non-users. We classify an investor as an internet user if she placed a trade through the internet in the past. Panel D decomposes the net buying measure into net buying by high-RPI investors and that by low-RPI investors. We adapt the reliance on public information (RPI) measure proposed by Kacperczyk and Seru (2007) to classify individual investors into two groups based on the median RPI. All other variables are defined in Table 1. Numbers in parentheses are t-statistics based on standard errors two-way clustered by firm and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated. Numbers in square brackets are p-values for the null that the coefficients on the two interaction terms are equal.

Panel A: Baseline specifications

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
$Netbuy_{[+1, +20]}$ × Post-EDGAR	0.307	0.454	0.407	0.389
	(3.30)***	(2.53)**	(2.11)**	(1.90)*
$Netbuy_{[+1, +20]}$	0.017	0.062	0.098	0.223
	(0.41)	(1.13)	(1.13)	(2.28)**
Post-EDGAR	-0.019	-0.036	-0.086	-0.096
	(2.06)*	(3.37)***	(5.53)***	(5.55)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.05	0.13	0.27	0.38

Panel B: Controlling for firm characteristics and their interaction terms with *Netbuy*

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
$Netbuy_{[+1, +20]}$ × Post-EDGAR	0.318	0.474	0.457	0.434
	(3.42)***	(2.68)**	(2.82)***	(2.19)**
$Netbuy_{[+1, +20]}$	0.309	0.634	0.366	0.777
	(1.57)	(1.90)*	(0.67)	(1.41)
Post-EDGAR	-0.016	-0.028	-0.074	-0.080
	(1.91)*	(2.69)**	(5.52)***	(5.33)***
Firm controls	Yes	Yes	Yes	Yes
$Netbuy_{[+1, +20]}$ × Firm controls	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.10	0.19	0.36	0.47

Panel C: Internet users vs. non-users

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
Netbuy by internet users × Post-EDGAR (b_1)	0.493	0.874	1.128	1.306
	(1.87)*	(2.31)**	(2.98)***	(2.79)***
Netbuy by non-users × Post-EDGAR (b_2)	0.250	0.276	0.146	0.071
	(2.43)**	(1.30)	(0.55)	(0.23)
Netbuy by internet users	0.016	0.139	0.175	0.182
	(0.14)	(0.83)	(0.86)	(0.78)
Netbuy by non-users	0.022	0.058	0.091	0.244
	(0.52)	(0.87)	(0.77)	(1.89)*
Post-EDGAR	-0.019	-0.036	-0.086	-0.097
	(2.39)**	(3.58)***	(5.92)***	(5.79)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
p-value for $b_1 = b_2$	[0.407]	[0.213]	[0.049]**	[0.044]**
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.05	0.13	0.27	0.38

Panel D: High- vs. low-RPI investors

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
Netbuy by high-RPI investors × Post-EDGAR (b_1)	-0.030	-0.074	0.044	-0.080
	(0.16)	(0.24)	(0.14)	(0.19)
Netbuy by low-RPI investors × Post-EDGAR (b_2)	0.581	0.974	0.862	0.796
	(3.09)***	(3.23)***	(2.07)**	(1.60)
Netbuy by high-RPI investors	0.100	0.155	0.183	0.344
	(0.94)	(1.03)	(1.04)	(1.71)*
Netbuy by low-RPI investors	-0.034	-0.006	-0.074	0.154
	(0.40)	(0.04)	(0.33)	(0.58)
Post-EDGAR	-0.019	-0.036	-0.086	-0.096
	(2.38)**	(3.55)***	(5.87)***	(5.75)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
p-value for $b_1 = b_2$	[0.024]**	[0.012]**	[0.123]	[0.239]
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.05	0.13	0.27	0.38

Table 3: Staggered implementation of EDGAR and the informativeness of trades by individual investors: Cross-sectional tests

This table reports regression analysis of how the impact of EDGAR on the informativeness of individual investors' trades varies with the level of information asymmetry and the presence of institutional investors. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-, six-, 12-, or 18-month window starting from the 21st day after an earnings announcement. Panel A interacts the main variables with *Opaque*, which is an indicator that equals one if the firm has no analyst coverage and below-the-median market capitalization and zero otherwise. Panel B interacts the main variables with an indicator for high institutional presence (*High IP*), which equals one if the firm has above-the-median level of institutional ownership and above-the-median number of institutional shareholders and zero otherwise. Analyst coverage, market cap, and institutional presence are measured as of the quarter-end immediately before the earnings announcements. All other variables are defined in Table 1. Numbers in parentheses are t-statistics based on standard errors two-way clustered by firm and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated. Numbers in square brackets are p-values for the null that the coefficients on the two interaction terms are equal.

Panel A: High vs. low information asymmetry stocks

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
$Netbuy_{[+1, +20]}$ × Post-EDGAR × Opaque	0.468	1.119	0.817	0.869
	(2.29)**	(3.67)***	(2.14)**	(2.22)**
$Netbuy_{[+1, +20]}$ × Post-EDGAR	0.027	-0.224	-0.113	-0.176
	(0.22)	(1.13)	(0.39)	(0.52)
$Netbuy_{[+1, +20]}$ × Opaque	-0.092	-0.480	-0.485	-0.580
	(0.83)	(3.63)***	(2.64)**	(2.58)**
Post-EDGAR × Opaque	-0.005	0.002	-0.001	0.012
	(0.47)	(0.15)	(0.02)	(0.36)
$Netbuy_{[+1, +20]}$	0.077	0.363	0.406	0.594
	(1.06)	(4.57)***	(3.71)***	(4.04)***
Post-EDGAR	-0.017	-0.035	-0.084	-0.095
	(1.96)*	(3.20)***	(5.37)***	(5.33)***
Opaque	0.079	0.132	0.202	0.246
	(8.89)***	(11.36)***	(12.88)***	(10.31)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.06	0.14	0.29	0.39

Panel B: High vs. low institutional presence

Dependent =	$CAR_{[+21, +83]}$ (1)	$CAR_{[+21, +146]}$ (2)	$CAR_{[+21, +272]}$ (3)	$CAR_{[+21, +398]}$ (4)
$Netbuy_{[+1, +20]}$ × Post-EDGAR × High IP	-0.568	-0.868	-0.571	-0.397
	(2.90)***	(3.20)***	(1.54)	(0.94)
$Netbuy_{[+1, +20]}$ × Post-EDGAR	0.431	0.640	0.515	0.448
	(4.62)***	(2.85)***	(2.21)**	(2.08)**
$Netbuy_{[+1, +20]}$ × High IP	0.167	0.413	0.488	0.467
	(1.45)	(4.07)***	(3.25)***	(2.08)**
Post-EDGAR × High IP	0.002	-0.002	0.005	0.007
	(0.25)	(0.18)	(0.28)	(0.30)
$Netbuy_{[+1, +20]}$	-0.015	-0.021	0.004	0.141
	(0.33)	(0.33)	(0.04)	(1.17)
Post-EDGAR	-0.020	-0.034	-0.088	-0.098
	(1.95)*	(3.00)***	(4.54)***	(4.47)***
High IP	-0.068	-0.115	-0.194	-0.262
	(12.28)***	(12.08)***	(13.15)***	(12.15)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,058	29,058	29,058	29,058
Adj. R-squared	0.06	0.14	0.29	0.40

Table 4: Staggered implementation of EDGAR and sell-side analyst research

This table reports regression analysis of the impact of EDGAR on analyst coverage and analysts' forecast accuracy. The dependent variable in the first (last) two columns is the number of analysts (forecast accuracy). All variables are defined in Table 1. Numbers in parentheses are *t*-statistics based on standard errors two-way clustered by firm and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	# of analysts		Forecast accuracy	
	(1)	(2)	(3)	(4)
Post-EDGAR	0.223	0.230	0.138	0.128
	(3.07)***	(3.27)***	(2.81)***	(3.35)***
Log(Total assets)		0.853		0.227
		(13.64)***		(3.55)***
Book-to-Market		-0.000		-0.004
		(2.65)**		(1.37)
Log(Firm age)		-0.032		-0.365
		(0.21)		(3.45)***
Prior stock return		-0.078		0.686
		(2.49)**		(13.87)***
ROA		0.131		1.265
		(1.76)*		(5.55)***
Book leverage		-0.340		-0.364
		(4.35)***		(2.60)**
Asset tangibility		0.732		-0.171
		(4.01)***		(0.67)
Sales growth		0.000		0.000
		(1.24)		(0.67)
CapEx		0.005		0.078
		(1.37)		(2.28)**
R&D		0.001		0.464
		(0.06)		(2.33)**
Institutional ownership		1.812		1.072
		(10.44)***		(6.70)***
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	103,866	103,866	56,281	56,281
Adj. R-squared	0.81	0.81	0.30	0.32

Table 5: Staggered implementation of EDGAR and market responses to analysts' forecast revisions

This table reports regression analysis of the impact of EDGAR on the informativeness of analysts' forecast revisions. The dependent variable is the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions. The unit of observation is a revision event. All variables are defined in Table 1. Numbers in parentheses are t-statistics based on standard errors three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	Revision $CAR_{[-1, +1]}$			
	(1)	(2)	(3)	(4)
Revision × Post-EDGAR	0.641	0.629	0.655	0.582
	(4.51)***	(4.57)***	(4.65)***	(4.23)***
Revision	0.368	0.221	0.202	0.255
	(4.80)***	(3.05)***	(2.63)**	(3.21)***
Post-EDGAR	0.001	0.002	0.002	
	(1.13)	(2.41)**	(2.62)**	
Log(Total assets)		-0.003	-0.002	
		(3.19)***	(1.93)*	
Book-to-Market		0.004	0.004	
		(3.87)***	(3.65)***	
Log(Firm age)		-0.004	-0.004	
		(2.27)**	(2.28)**	
Prior stock return		0.010	0.011	
		(11.85)***	(11.32)***	
ROA		-0.009	-0.009	
		(2.03)*	(1.81)*	
Book leverage		-0.001	0.000	
		(0.31)	(0.01)	
Asset tangibility		-0.007	-0.008	
		(1.60)	(2.00)*	
Sales growth		-0.000	-0.000	
		(0.33)	(0.67)	
CapEx		-0.001	-0.002	
		(0.23)	(0.82)	
R&D		0.003	0.002	
		(0.63)	(0.38)	
Institutional ownership		0.005	0.004	
		(1.53)	(1.42)	
Firm FEs	Yes	Yes	Yes	No
Analyst FEs	Yes	Yes	No	No
Year-quarter FEs	Yes	Yes	No	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	No	No	Yes	Yes
# of observations	358,080	354,853	349,114	342,824
Adj. R-squared	0.06	0.07	0.07	0.10

Table 6: Propensity-score matching

This table reports propensity-score matching diagnostics and the diff-in-diff tests using the propensity-score matched sample. Panel A compares firm characteristics across treatment firms (i.e., firms that switch from a non-filer to an EDGAR filer in a month) and control firms (i.e., firms remain as a filer or non-filer in the 12 months before the month under consideration and 12 months after). *Treated* is an indicator that takes the value of one for treatment firms and zero for control firms. Column 1 of Panel A uses the pre-matching sample to estimate a logit regression of *Treated* on firm characteristics. We use the predicted probabilities, or propensity scores, from the estimation and perform one-to-one nearest-neighbor matching with replacement. Column 2 of Panel A repeats the regression on the propensity-score matched sample. Panel B reports the diff-in-diff tests of the impact of EDGAR on information production using the propensity-score matched sample. All other variables are defined in Table 1. Numbers in parentheses are *t*-statistics. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Panel A: Pre- and post-matching regressions

Dependent =	Treated	
	Pre-matching (1)	Post-matching (2)
Log(Total assets)	-0.035	-0.021
	(4.08)***	(1.31)
Book-to-Market	0.149	-0.043
	(5.72)***	(1.36)
Log(Firm age)	0.348	0.042
	(24.76)***	(1.34)
Prior stock return	0.133	0.060
	(1.27)	(0.47)
ROA	0.053	-0.036
	(1.65)*	(1.11)
Book leverage	0.315	-0.009
	(5.57)***	(0.11)
Asset tangibility	0.002	0.154
	(0.03)	(1.26)
Sales growth	-0.520	-0.129
	(2.46)**	(0.52)
CapEx	-0.050	-0.004
	(1.83)*	(0.39)
R&D	0.008	-0.167
	(0.04)	(0.71)
Institutional ownership	0.047	0.181
	(0.67)	(1.48)
Cohort FEs	Yes	Yes
Industry FEs	Yes	Yes
# of observations	26,097	8,388
Pseudo R-squared	0.09	0.00

Panel B: Diff-in-diff tests using the propensity-score matched sample

	Informativeness of individual trades (1)	# of analysts (2)	Forecast accuracy (3)	Market responses to revisions (4)
DiD estimate	0.373	0.243	0.097	0.380
	(2.22)**	(1.92)*	(2.04)**	(2.45)**

Table 7: Controlling for group-specific time trends

This table reports regression analysis of the impact of EDGAR on various information production measures after adding controls for group-specific time trends. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn $_{[+21,\ +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $CAR_{[-1,\ +1]}$ (4)
$Netbuy_{[+1,\ +20]}$ × Post-EDGAR	0.360 (3.69)***			
$Netbuy_{[+1,\ +20]}$	0.138 (1.26)			
Post-EDGAR	-0.019 (2.00)**	0.216 (2.94)***	0.143 (3.20)***	
Revision × Post-EDGAR				0.613 (4.66)***
Revision				0.291 (1.73)*
Group × time trends	Yes	Yes	Yes	No
Netbuy × group × time trends	Yes	N/A	N/A	N/A
Revision × group × time trends	N/A	N/A	N/A	Yes
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	29,058	103,866	56,281	342,824
Adj. R-squared	0.05	0.81	0.30	0.10

Table 8: Ease of access to EDGAR filings

This table reports regression analysis of the impact of EDGAR on various information production measures when we partition the post-EDGAR period for the first four groups into two periods based on the ease of access to EDGAR filings. Specifically, we redefine the *Post-EDGAR* indicator for the first four groups to take the value of one if the firm-quarter is after January 17, 1994 (when the filings became available to internet users without additional charges) and zero otherwise. *Interim* is an indicator variable that takes the value of one if the firm-quarter falls in the interim period, i.e., the time from the starting date of mandated electronic filing to EDGAR to January 16, 1994, for the first four groups of companies and zero otherwise. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All other variables are defined in Table 1. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn $_{[+21, +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision CAR $_{[-1, +1]}$ (4)
Netbuy $_{[+1, +20]}$ × Post-EDGAR	0.308 (3.28)***			
Netbuy $_{[+1, +20]}$ × Interim	0.269 (0.83)			
Netbuy $_{[+1, +20]}$	0.017 (0.42)			
Post-EDGAR	-0.031 (2.73)***	0.208 (2.89)***	0.178 (3.09)***	
Interim	0.010 (1.11)	0.282 (1.43)	0.020 (0.42)	
Revision × Post-EDGAR				0.619 (4.37)***
Revision × Interim				0.088 (1.29)
Revision				0.255 (3.21)***
Firm Fes	Yes	Yes	Yes	No
Quarter Fes	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	29,058	103,866	56,281	342,824
Adj. R-squared	0.05	0.81	0.30	0.10

Table 9: Falsification tests

This table reports regression analysis of information production activities using a four-year period preceding the actual EDGAR implementation. We define pseudo-events as occurring 24 months prior to the actual implementation. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. "*Post-EDGAR*" is an indicator that equals one for firm-quarters that are in the two-year window after the pseudo-event date and zero for firm-quarters that are in the two-year window immediately before the pseudo-event date. All other variables are defined in Table 1. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn $_{[+21, +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $CAR_{[-1, +1]}$ (4)
$Netbuy_{[+1, +20]}$ × "Post-EDGAR"	-0.050			
	(0.56)			
$Netbuy_{[+1, +20]}$	0.103			
	(1.61)			
"Post-EDGAR"	0.007	0.085	0.026	
	(0.67)	(0.62)	(0.47)	
Revision × "Post-EDGAR"				0.146
				(1.47)
Revision				0.059
				(1.11)
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	18,489	67,828	30,983	122,755
Adj. R-squared	0.06	0.83	0.39	0.07

Table 10: Excluding Group CF–01 firms

This table reports regression analysis of the impact of EDGAR on various information production measures after excluding firms assigned to Group CF–01 (which consists mostly of transitional filers) on the phase-in schedule. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn$_{[+21,\ +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision CAR$_{[-1,\ +1]}$ (4)
Netbuy$_{[+1,\ +20]}$ × Post-EDGAR	0.296			
	(3.10)***			
Netbuy$_{[+1,\ +20]}$	0.018			
	(0.46)			
Post-EDGAR	-0.021	0.237	0.155	
	(2.12)**	(2.85)***	(3.38)***	
Revision × Post-EDGAR				0.567
				(4.03)***
Revision				0.261
				(3.18)***
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	27,695	100,674	53,514	314,634
Adj. R-squared	0.05	0.80	0.30	0.10

Table 11: Staggered implementation of EDGAR and the trading volume of individual investors

This table reports regression analysis of the impact of EDGAR on the trading volume of individual investors following earnings announcements. The dependent variable in Panel A is individual investors' trading volume measured as the total number of shares traded by our sample of individual investors (purchases plus sales) during the first 20 trading days following an earnings announcement scaled by the number of shares outstanding. Panel B partitions the trading volume variable into trading volume of internet users and that of non-users. All other variables are defined in Table 1. Numbers in parentheses are *t*-statistics based on standard errors two-way clustered by firm and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Panel A: All individual investors

Dependent =	Trading volume$_{[+1,\ +20]}$	
	(1)	(2)
Post-EDGAR	0.394 (2.28)**	0.340 (2.17)**
SUE decile rank		0.004 (0.35)
Log(Total assets)		-0.522 (2.54)**
Book-to-Market		0.573 (2.61)**
Log(Firm age)		-0.050 (0.13)
Prior stock return		0.797 (7.09)***
ROA		1.682 (2.04)**
Book leverage		1.599 (1.91)*
Asset tangibility		-1.239 (1.02)
Sales growth		-0.009 (0.06)
CapEx		1.843 (2.53)**
R&D		0.148 (0.11)
Institutional ownership		-2.596 (4.30)***
Firm FEs	Yes	Yes
Year-quarter FEs	Yes	Yes
# of observations	29,418	29,418
Adj. R-squared	0.24	0.25

Panel B: Internet users vs. non-users

Dependent =	Trading volume of internet users $_{[+1, +20]}$		Trading volume of non-users$_{[+1, +20]}$	
	(1)	(2)	(3)	(4)
Post-EDGAR	0.171	0.161	0.221	0.177
	(2.78)***	(2.71)***	(1.51)	(1.34)
Firm controls	No	Yes	No	Yes
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,418	29,418	29,418	29,418
Adj. R-squared	0.13	0.14	0.22	0.23

Panel C: High- vs. low-RPI investors

Dependent =	Trading volume of high-RPI investors$_{[+1, +20]}$		Trading volume of low-RPI investors$_{[+1, +20]}$	
	(1)	(2)	(3)	(4)
Post-EDGAR	0.082	0.068	0.319	0.270
	(1.25)	(1.05)	(3.19)***	(3.05)***
Firm controls	No	Yes	No	Yes
Firm FEs	Yes	Yes	Yes	Yes
Year-quarter FEs	Yes	Yes	Yes	Yes
# of observations	29,418	29,418	29,418	29,418
Adj. R-squared	0.13	0.13	0.16	0.17

Table 12: Staggered implementation of EDGAR and stock pricing efficiency

This table reports regression analysis of the impact of EDGAR on stock pricing efficiency. The dependent variable is one of the inverse measures of stock pricing efficiency, namely stock price synchronicity (i.e., R-squared), the absolute value of stock return autocorrelation, and the standard deviation of the pricing error divided by the standard deviation of the log transaction price. The unit of observation is a firm-month. All other variables are defined in Table 1. Numbers in parentheses are t-statistics based on standard errors two-way clustered by firm and by year-month. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	Price synchronicity		Abs(Stock return autocorrelation)		Pricing error	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-EDGAR	-0.008	-0.009	-0.011	-0.011	-0.007	-0.005
	(2.31)**	(2.61)**	(5.76)***	(6.08)***	(2.36)**	(2.20)**
Log(Total assets)		0.012		-0.014		-0.022
		(5.34)***		(9.64)***		(10.39)***
Book-to-Market		-0.013		0.013		0.023
		(9.33)***		(8.89)***		(8.83)***
Log(Firm age)		-0.014		-0.009		-0.006
		(3.29)***		(2.56)**		(1.36)
Prior stock return		0.014		-0.026		-0.036
		(10.86)***		(21.45)***		(26.47)***
ROA		-0.001		-0.006		-0.015
		(0.49)		(1.52)		(2.41)**
Book leverage		-0.017		0.015		0.033
		(5.75)***		(3.45)***		(5.76)***
Asset tangibility		-0.012		0.007		0.001
		(2.18)**		(1.07)		(0.06)
Sales growth		0.002		-0.001		-0.002
		(3.03)***		(1.56)		(1.77)*
CapEx		0.022		-0.031		-0.038
		(3.68)***		(4.73)***		(4.04)***
R&D		0.006		-0.012		-0.022
		(0.88)		(1.21)		(1.63)
Institutional ownership		0.054		-0.035		-0.096
		(7.98)***		(6.80)***		(13.05)***
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	300,395	300,395	300,370	300,370	158,319	158,319
Adj. R-squared	0.43	0.43	0.17	0.18	0.39	0.43

Internet Appendix for Informing the Market: The Effect of Modern Information Technologies on Information Production

(Not Intended for Publication)

Table IA-1: Robustness check using the actual timing of the implementation

This table reports regression analysis of the impact of EDGAR on various information production measures using actual implementation dates to define the Post-EDGAR indicator. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1 in the paper. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn $_{[+21,\ +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision CAR $_{[-1,\ +1]}$ (4)
Netbuy $_{[+1,\ +20]}$ × Post-EDGAR	0.309 (3.24)***			
Netbuy $_{[+1,\ +20]}$	0.017 (0.43)			
Post-EDGAR	-0.019 (2.04)**	0.233 (3.28)***	0.136 (2.79)***	
Revision × Post-EDGAR				0.598 (4.31)***
Revision				0.249 (3.10)***
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	29,057	103,822	56,255	341,921
Adj. R-squared	0.05	0.80	0.30	0.10

Table IA-2: Omitted variable bias tests

This table presents the results from an omitted variable bias test suggested by Oster (2017). The idea behind the test is that if selection on the observables is proportional to the selection on the unobservables, an identified set, which includes the true effect, can be constructed. We perform this test by calculating the identified set as

$$[\tilde{\beta}, \tilde{\beta} - \tilde{\delta}\,\frac{(\beta_0 - \tilde{\beta})(R_{max} - \tilde{R})}{\tilde{R} - R_0}],$$

where β_0 and R_0 denote the estimate and R-squared for the baseline model without controls for firm characteristics, respectively; $\tilde{\beta}$ and $\tilde{R}$ denote the estimate and R-squared for the model with the full set of controls, respectively; and $\tilde{\delta}$ denotes the relative importance of observed versus unobserved variables in generating selection bias. Following Knupfer, Rantapuska, and Sarvimaki (2017), we set $\tilde{\delta} = 1$ and $R_{max} = min(1, 1.3\tilde{R})$. The last row provides the identified set for the true effect in each of our main tests.

Dependent =	AbnReturn$_{[+21,\ +83]}$	# of analysts	Forecast accuracy	Revision CAR$_{[-1,\ +1]}$
	(1)	(2)	(3)	(4)
Baseline model	Column 1 of Table 2, Panel A	Column 1 of Table 4	Column 3 of Table 4	Column 1 of Table 5
β_0	0.307	0.223	0.138	0.641
R_0	0.142	0.814	0.340	0.075
Full model	Column 1 of Table 2, Panel B	Column 2 of Table 4	Column 4 of Table 4	Column 4 of Table 5
$\tilde{\beta}$	0.318	0.240	0.128	0.582
$\tilde{R}$	0.209	0.823	0.363	0.283
R_{max}	0.272	1.000	0.472	0.367
Identified set	[0.318, 0.328]	[0.240, 0.598]	[0.128, 0.082]	[0.582, 0.558]

Table IA-3: Robustness check excluding Group CF–10

This table reports regression analysis of the impact of EDGAR on various information production measures after excluding firms assigned to Group CF–10 (which consists mostly of small firms) on the phase-in schedule. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1 in the paper. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	$\text{AbnReturn}_{[+21,\ +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $\text{CAR}_{[-1,\ +1]}$ (4)
$\text{Netbuy}_{[+1,\ +20]}$ × Post-EDGAR	0.345 (3.40)***			
$\text{Netbuy}_{[+1,\ +20]}$	-0.008 (0.19)			
Post-EDGAR	-0.010 (1.11)	0.295 (2.78)***	0.119 (1.98)*	
Revision × Post-EDGAR				0.667 (4.88)***
Revision				0.206 (2.58)**
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	26,367	90,503	48,731	226,966
Adj. R-squared	0.05	0.82	0.29	0.10

Table IA-4: Controlling for time trends that are common across firm size groups and age groups

This table reports regression analysis of the impact of EDGAR on various information production measures after adding controls for group-specific time trends. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1 in the paper. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	$AbnReturn_{[+21, +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $CAR_{[-1, +1]}$ (4)
$Netbuy_{[+1, +20]}$ × Post-EDGAR	0.301 (2.77)***			
$Netbuy_{[+1, +20]}$	-0.316 (0.69)			
Post-EDGAR	-0.020 (2.09)**	0.219 (2.89)***	0.160 (3.61)***	
Revision × Post-EDGAR				0.533 (4.62)***
Revision				-0.150 (0.59)
Size group × time trends	Yes	Yes	Yes	No
Age group × time trends	Yes	Yes	Yes	No
Netbuy × size group × time trends	Yes	N/A	N/A	N/A
Netbuy × age group × time trends	Yes	N/A	N/A	N/A
Revision × size group × time trends	N/A	N/A	N/A	Yes
Revision × age group × time trends	N/A	N/A	N/A	Yes
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	28,895	101,282	55,307	344,751
Adj. R-squared	0.06	0.81	0.30	0.10

Table IA-5: Additional falsification tests

This table reports regression analysis of information production activities using a four-year period following the actual EDGAR implementation. We define pseudo-events as occurring 24 months after the actual implementation. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. "*Post-EDGAR*" is an indicator that equals one for firm-quarters that are in the two-year window after the pseudo-event date and zero for firm-quarters that are in the two-year window immediately before the pseudo-event date. All other variables are defined in Table 1 in the paper. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	$\text{AbnReturn}_{[+21, +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $\text{CAR}_{[-1, +1]}$ (4)
$\text{Netbuy}_{[+1, +20]}$ × "Post-EDGAR"	0.072 (0.38)			
$\text{Netbuy}_{[+1, +20]}$	0.287 (2.11)**			
"Post-EDGAR"	0.026 (2.87)**	-0.116 (1.54)	-0.017 (0.51)	
Revision × "Post-EDGAR"				0.063 (0.36)
Revision				0.750 (5.88)
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	12,305	55,404	31,479	222,285
Adj. R-squared	0.06	0.85	0.34	0.13

Table IA-6: Staggered implementation of EDGAR and stock market liquidity

This table reports regression analysis of the impact of EDGAR on stock pricing efficiency. The dependent variable is relative effective spread, Amihud illiquidity ratio, and share turnover, respectively. The relative effective spread is twice the signed difference between the transaction price and the midpoint of the bid and ask quotes at the time of the transaction, normalized by the midpoint of the bid and ask quotes. Amihud illiquidity ratio is the average ratio of the absolute daily stock return to dollar trading volume (in $ millions). Turnover is the ratio of the total number of shares traded to the number of shares outstanding. The unit of observation is a firm-month. All other variables are defined in Table 1 in the paper. Numbers in parentheses are *t*-statistics based on standard errors two-way clustered by firm and by year-month. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	Relative effective spread		Amihud illiquidity ratio		Share turnover	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-EDGAR	-0.002	-0.002	-1.263	-1.088	0.007	0.006
	(4.17)***	(4.84)***	(4.24)***	(3.78)***	(3.20)***	(3.05)***
Log(Total assets)		-0.005		-2.152		0.014
		(7.79)***		(6.70)***		(5.63)***
Book-to-Market		0.007		3.368		-0.014
		(9.58)***		(8.95)***		(9.50)***
Log(Firm age)		-0.001		0.526		0.006
		(0.98)		(1.28)		(1.56)
Prior stock return		-0.010		-3.958		0.042
		(26.27)***		(18.18)***		(16.85)***
ROA		-0.012		-5.221		0.014
		(6.90)***		(5.77)***		(2.38)**
Book leverage		0.018		8.036		0.003
		(10.49)***		(8.28)***		(0.62)
Asset tangibility		0.004		2.167		-0.031
		(1.39)		(1.51)		(3.30)***
Sales growth		-0.001		-0.398		0.003
		(1.54)		(2.20)**		(2.45)**
CapEx		-0.011		-7.737		0.064
		(5.19)***		(5.43)***		(6.25)***
R&D		-0.012		-2.462		0.025
		(3.20)***		(1.77)*		(1.41)
Institutional ownership		-0.008		2.469		0.098
		(5.08)***		(3.47)***		(11.95)***
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	203,350	203,350	321,613	321,613	321,613	321,613
Adj. R-squared	0.74	0.77	0.45	0.47	0.41	0.45

Table IA-7: Robustness check using firms that have above-the-median age as of January 1993

This table reports regression analysis of the impact of EDGAR on various information production measures using the sample of firms that have above-the-median age as of January 1993. The dependent variables are cumulative DGTW-characteristics adjusted returns during a three-month window starting from the 21st day after an earnings announcement, the number of analysts, forecast accuracy, and the cumulative DGTW characteristics-adjusted returns during a three-day window around analyst revisions, respectively. All variables are defined in Table 1 in the paper. Numbers in parentheses in the first three columns are *t*-statistics based on standard errors two-way clustered by firm and year-quarter, and those in the last column are three-way clustered by firm, analyst, and year-quarter. Significance at the 10% (*), 5% (**), or 1% level (***) is indicated.

Dependent =	AbnReturn $_{[+21,\ +83]}$ (1)	# of analysts (2)	Forecast accuracy (3)	Revision $CAR_{[-1,\ +1]}$ (4)
$Netbuy_{[+1,\ +20]}$ × Post-EDGAR	0.563 (3.44)***			
$Netbuy_{[+1,\ +20]}$	-0.087 (1.29)			
Post-EDGAR	-0.004 (0.58)	0.312 (3.39)***	0.092 (2.21)**	
Revision × Post-EDGAR				0.667 (4.88)***
Revision				0.206 (2.58)**
Firm FEs	Yes	Yes	Yes	No
Year-quarter FEs	Yes	Yes	Yes	No
Firm × year-quarter FEs	No	No	No	Yes
Analyst × year-quarter FEs	N/A	N/A	N/A	Yes
Clustering SEs	Firm, year-quarter	Firm, year-quarter	Firm, year-quarter	Firm, analyst, year-quarter
# of observations	17,828	58,311	33,390	226,966
Adj. R-squared	0.05	0.83	0.30	0.10