Review of Accounting Studies (2024) 29:2350–2390
https://doi.org/10.1007/s11142-022-09737-6

The productivity effect of digital financial reporting

Zheng Liu[1] · Ning Zhang[2]

Accepted: 1 November 2022 / Published online: 27 May 2023

Abstract
We examine the effect of digital financial reporting on firm productivity. Information frictions represent a constraint that impedes efficient resource allocation and a major source of such frictions stems from the fact that firms' production functions (the conversion from inputs to outputs) are not observable to corporate outsiders. Digital communication of corporate financial data fundamentally changes how firm-specific information is disclosed, released, and disseminated by mitigating information asymmetry between corporate insiders and outsiders and facilitates the processing of such information. We use the staggered implementation of the SEC's Electronic Data Gathering and Analysis Retrieval (EDGAR) system to investigate the impact of digital financial reporting on firms' productivity. We show that the implementation of EDGAR results in an economically meaningful and statistically significant increase on firms' productivity, measured by total factor productivity (TFP). By focusing on the role of information dissemination in coordinating investments and production, our findings provide evidence on the real effects of "going digital" in corporate reporting.

Keywords Digital technology · Financial reporting · EDGAR · TFP · Information asymmetry

JEL Classification G14 · G30

✉ Ning Zhang
nz8@queensu.ca

1 School of Business, Hong Kong Baptist University, Hong Kong, China

2 Smith School of Business, Queen's University, Kingston, ON, Canada

1 Introduction

The past two decades have witnessed the application of digital technologies in almost every corner of the financial world. One of the main purposes of this digital shift in corporate financial reporting is to provide easily accessible information to investors at lower cost. In this study, we examine the impact of a specific information technology—the implementation, by the U.S. Securities and Exchange Commission (SEC), of the Electronic Data Gathering Analysis and Retrieval (EDGAR) system—on firms' productivity. To the extent that technological productivity is a vital source of long-term economic growth (Solow 1957), information technology can add great value to the economy if it enhances firms' productivity.

We motivate our research question by observing the emerging use of digital technology in the capital market. Digital reporting, particularly the transition from a paper-based to a digitalized information dissemination system, fundamentally changes the ways of information production, dissemination, acquisition, and utilization. Users of digital financial information access most public information at their own computer terminals without having to physically travel. Corporate financial reporting is arguably one of the most important types of public information released periodically to investors. The introduction of the EDGAR system—a major milestone in the history of the SEC that took place between 1993 and 1996—substantially improved the accessibility of such information to corporate outsiders. The EDGAR system performs automated collection, validation, indexing, acceptance, and forwarding of submissions by companies and others who are required by law to file forms with the SEC. After almost 30 years of operations, the EDGAR platform now maintains a complete history of corporate filings that enables investors to conduct fundamental analysis, valuation, and investment analysis by employing both times-series information within the firm and cross-sectional information within and across industries. Recent studies have examined the impact of adopting such digital information technology on retail investors. For example, Gao and Huang (2020) show that individual investors become more informed about future stock returns in the post-EDGAR stage. Their study focuses on an economic consequence accruing to the *users* of corporate information. In this study, we investigate the impact of digital information technology on the *producers* of corporate information—that is, the adopting firms themselves. We illuminate the economic implications of this digital technology in financial reporting by providing evidence as to whether and how the way corporate financial information is presented and disseminated improves firms' productivity— arguably the driver of economic growth more broadly.

Our investigation is also motivated by a fundamental question in economics: What determines a firm's productivity? In economics, productivity captures the efficiency of converting production inputs into final outputs. Researchers typically employ total factor productivity (TFP) to empirically measure productivity. Also known as the "Solow's residual," Solow (1956, 1957) shows that TFP is the key source of long-term economic growth. Hall and Jones (1999) show that TFP explains a considerable portion of the cross-country variation in income levels. While studies by economists demonstrate that firm characteristics and operating environments (such as management practice, institutions, and political economy) collectively shape TFP, less is known about how the financial reporting environment affects firms' productivity. Our study responds to the call for more research on the effect of the big data technology on investment choices (Roychowdhury et al. 2019). Specifically, in their review, Roychowdhury et al. argue that "recent technological advances and the availability of big data and sophisticated data analytic tools can influence internal and external reporting decisions which can ultimately influence investment via agency costs, managerial learning or behavioral biases, and thus offer promising opportunities for both empirical and analytical research." As such, our investigation is pertinent not only to researchers, but also to regulators and practitioners.

We develop our hypothesis by focusing on the role of accounting information in shaping firms' operational outcomes. Productivity is achieved when a firm uses economic inputs to generate outputs. Information frictions represent a constraint that impedes efficient resource allocation, and a major source of such frictions stems from the fact that firms' production functions (the conversion from inputs to outputs) are not observable to corporate outsiders. When a firm's financial information is unavailable or costly to acquire and analyze, moral hazard and adverse selection problems arise. Moral hazard occurs because managers have a tendency to overinvest when they have positive free cash flows (Jensen 1986). Adverse selection occurs because, when exact corporate information is hard to acquire, capital providers may price the firm into the "bad" type, given information asymmetry (Myers and Majluf 1984). As such, when a firm needs additional capital, potential investors may be reluctant to supply capital, or they may charge a higher required rate of return upon supplying it. The implementation of EDGAR reduces the cost of financial information acquisition and improves corporate transparency. With digital corporate financial information provided in the EDGAR system, capital providers have more readily accessible information and are able to monitor the manager more frequently and more proactively. Higher monitoring efficiency leads to higher investment efficiency, mitigating the problem of overinvestment. Timely information through EDGAR also mitigates the information asymmetry between the manager and capital providers, relieving the problem of underinvestment and capital rationing. As such, to the extent that funds (i.e., inputs) are better employed in investment projects and operations, we expect that the adoption of the EDGAR technology increases a firm's productivity.

To test the prediction, we employ the universe of firms listed in the United States and exploit the staggered implementation of the EDGAR system between 1993 and 1996. Specifically, the three-year phase-in program requires companies to electronically file their corporate reports to the SEC on the EDGAR system beginning on April 26, 1993, and ending on May 6, 1996. Prior to the phase-in schedule, firms had to submit hardcopies of their fillings to the SEC. Upon receiving these hardcopies, the staff at the SEC reviewed and then transferred them to reference rooms located in Washington, DC, New York, and Chicago. Since the implementation of the EDGAR system, the online system has stored all corporate fillings in digital format and has made them freely available online to allow immediate access to arguably everyone.

The implementation of EDGAR serves as a powerful setting with clear identification for our research question for two reasons. First, this phase-in project represents a natural exogenous shock to the way in which a firm discloses and disseminates financial information by holding the firm's fundamentals constant. In other words, the innate factors that shape a firm's financial reporting quality, such as its business model and operating environment, remain unchanged. Second, the adoption is implemented over a three-year period, allowing us to infer causality. Specifically, the SEC divides all filers into ten groups, with the filings by the first group going online, at the earliest, in April 1993 and the filings by the last group going online in May 1996. This quasi-random, staggered feature of the adoption helps mitigate the confounding effect of concurrently occurring events related to the years in the phase-in period.

We use total factor productivity (TFP) as our measure of productivity. Conceptually, TFP measures the overall efficiency with which capital and labor are employed in the production process (Bennett et al. 2020). We follow Ackerberg et al. (2015) and estimate a log-linearized production function to calculate firm-level TFP. We find that the implementation of EDGAR has an economically meaningful and statistically significant impact on firms' productivity. On average, a firm's move from paper-based filings to electronic filings at EDGAR results in an increase of 3.5% in productivity after controlling for a handful of firm characteristics, firm fixed effects, and year fixed effects. This finding provides support to our main hypothesis that digital information technology enhances firms' productivity.

We perform a battery of robustness checks. First, we employ a propensity score matching algorithm to mitigate the concern that the SEC's grouping may not be completely random. Second, we employ an alternative firm-level TFP measure, calculated following İmrohoroğlu and Tüzel (2014) and based on the specification in Olley and Pakes (1996). The impact of the EDGAR implementation remains statistically significant, and the economic magnitude is similar to the result in the baseline test. Third, we use a dynamic specification that includes seven year indicators around the year of EDGAR implementation. We find that the impact of EDGAR implementation is statistically significant in all years after the implementation and that the effect persists in post-EDGAR years. Fourth, we remove firms that voluntarily provide electronic filings in the "transition period." According to the official SEC schedule, these voluntary EDGAR filers are all placed in the first of the ten groups. Results are robust to removing these voluntary adopters. Finally, we perform a test to capture the costs of accessing corporate information in EDGAR with a fee when the system was initially introduced. The effect of implementing EDGAR on TFP remains economically large and statistically significant after controlling for the initial fee-based EDGAR access.

To assess the confounding effect of time trends on firms' productivity, we conduct a falsification analysis in which we randomly generate pseudo-adoption dates and define a pseudo *Post-EDGAR* indicator using these randomly produced dates. We repeat the baseline regression 100, 500, and 1,000 times and find that the mean coefficient on the pseudo *Post-EDGAR* indicator is statistically insignificant at conventional levels in each of these falsification specifications.

The implication underlying our prior analysis is that the EDGAR implementation enhances a firm's TFP by reducing information asymmetry from the input side. We next show that the implementation of EDGAR reduces both overinvestment and underinvestment, consistent with the idea that more readily available corporate information enables investors to better monitor the firm and provide capital when needed in the post-EDGAR period. Finally, we show that the implementation of the EDGAR system plays a more important role when alternative information from financial analysts is scarce.

Our study contributes to the finance and accounting literature in several ways. First, our study joins the recent literature on the impact of capital markets on firm productivity. For example, David et al. (2016) consider a theoretical framework where firms choose inputs (to make operational decisions) under limited information about their idiosyncratic fundamentals. They show analytically that informational friction leads to a *misallocation* of factors across firms in an *ex post* sense, reducing productivity and output. Bennett et al. (2020) document that stock price informativeness has a positive effect on a firm's productivity. Our study shows that a firm's productivity is also a function of the acquisition cost of corporate financial information.

Second, by focusing on the role of information dissemination in coordinating investments and production, our findings collectively provide evidence on the real effect of going digital in corporate reporting. A common thread of concurrent studies using the EDGAR adoption setting is that researchers typically evaluate the economic benefits accruing to *investors* due to convenient information access through EDGAR. In this study, we focus on the economic consequences accruing to the *EDGAR-adopting firms* themselves by examining how this change to digital reporting affects firms' investment and operation outcomes. Our study also responds to the call for research in Bushman and Smith (2001) by showing that it is not only financial information in and of itself, but also the way such information is presented to external users, that plays a role in shaping economic outcomes.

Third, our study also complements recent findings in financial technology. Using a historical context in China between 1881 and 1936, Lin et al. (2021) show that the introduction of the telegraph facilitates information exchange and business dealings between bank headquarters and their branches, expanding both the number and geographic scope of banks' branch networks. While the telegraph they study primarily reduces the time and cost for the banks to obtain information about each transaction, the adoption of EDGAR presents a fuller picture of corporate information in the digital format. We use a more recent technological innovation that takes place in the United States—the digital presentation and dissemination of financial information—and show that it improves productivity.

The remainder of the paper proceeds as follows. Section 2 discusses related literature on the EDGAR system and develops the hypothesis. Section 3 discusses the econometric estimation of TFP and lays out the research design. Section 4 describes the sample and discusses the main empirical results. Section 5 performs robustness tests, and Section 6 conducts additional analyses. Section 7 concludes.

2 Background information and literature review

2.1 The EDGAR implementation

Since the 1930s, the primary mission of the U.S. SEC has been to protect investors; maintain fair, orderly, and efficient markets; and facilitate capital formation. As part of this effort, public firms are mandated to disclose accurate and complete information about their operations on a regular basis, as well as any event that may materially impact them, in public filings. In the pre-EDGAR era, the enormous volume of information received by the SEC created a very heavy workload for SEC staff members, and also an onerous process for public users who wanted to access such information. Until the invention of the EDGAR system, hardcopies of corporate filings were displayed only in three reference rooms, located in Washington, D.C., Chicago, and New York, for public access. Public access was available via request, making obtaining such information inconvenient.[1] Given the large volume and limited copies of each document as well as the inconvenient search cost of physical visits, this traditional way of displaying and disseminating corporate-specific information was subject to criticism.

The EDGAR system is the first of its kind to perform automated collection, validation, indexing, acceptance, and forwarding of submissions by corporations to security regulators around the world. The key feature of the EDGAR system is to enable immediate and wide access to company fillings through the internet to corporate outsiders. The number of internet visits to EDGAR has been steadily increasing, averaging 20 million per month between 2005 and 2012, with a peak of almost 50 million in November 2012 (Loughran and McDonald 2017). Furthermore, as a result of the electronic platform, the information processing costs of corporate filings are much lower than they were for paper copies. For instance, EDGAR users can simply enter keyword strings into a computer software (e.g., a web browser) to search for and locate specific information, rather than having to manually read through the entire text. In addition, the application of artificial intelligence, such as machine learning technologies, also facilitates simultaneously accessing and processing a large number of corporate filings within EDGAR using computer-automated algorithms.

The adoption of the EDGAR system took a phase-in approach. As discussed in Gao and Huang (2020), on February 23, 1993, the SEC issued the final rules requiring all registered firms to submit their required fillings electronically through EDGAR. Firms were categorized into ten groups, from Group CF-01 to Group CF-10. Companies in the first group began to file electronically in April 1993, and the companies in the last group were required to file in May 1996 at the latest. The

[1] A 1982 *New York Times* article quoted Maryann Wismer—then a researcher at Disclosure Incorporated, a company specializing in retrieving financial information for private customers—as exclaiming "it's just incredible the number of problems you can run into trying to find something you need." See the full text at: https://www.nytimes.com/1982/05/19/business/sec-data-difficult-hunt.html.

ten groups were generally evenly spaced over the three-year period with the gap between each neighboring group being approximately three to six months.[2]

To the extent that the implementation of EDGAR has changed the way in which market participants access firm-specific information, researchers have evaluated the capital market consequence of this transition. For example, using TAQ data, Asthana et al. (2004) show that the switch to EDGAR filings results in significant increases in the volume of small, but not large, trades during the short window around a filing. Gao and Huang (2020) exploit the staggered implementation of EDGAR and show that the implementation results in an increase in the trading volume and profits of retail investors, as well as an increase in the frequency and accuracy of sell-side analyst reports. Another strand of literature exploits the EDGAR search traffic log data and investigates the impact of the usage of EDGAR fillings on capital market outcomes. For example, Lee et al. (2015) develop a "co-search" based peer firm membership using EDGAR search traffic data. Their approach is to capture the idea that investors perform searches for firms with perceived similarities (along multiple dimensions) in chronologically adjacent points in time.

Recently, there has been an emerging literature that uses the EDGAR implementation setting to help infer causality of digital financial reporting on capital market outcomes. For example, Goldstein et al. (2021) explore the setting of EDGAR adoption and show that EDGAR implementation leads to a decrease in investment-to-price sensitivity. Lai et al. (2020) show that EDGAR implementation leads to a reduction in the cost of equity capital. Gomez (2020) documents that the implementation of EDGAR reduces information asymmetry between managers and investors but increases information asymmetry between more- and less-sophisticated investors. Guo et al. (2019) show that firms experience a decrease in stock price crash risk after they file electronically in EDGAR. Our study joins this literature by placing an emphasis on examining the benefits accruing to EDGAR-adopting firms' operating and investment behaviors.[3]

2.2 Hypothesis development

2.2.1 Productivity

The production process converts resources (i.e., inputs) into products (i.e., outputs), and the TFP measures the efficiency of this conversion process. In an editorial review, Restuccia and Rogerson (2017) summarize that this efficiency is primarily determined by two factors: technology and resource misallocation. While

[2] The SEC's effort to promote capital market information efficiency is ongoing. Since the adoption of the EDGAR system, the SEC has been actively planning and developing additional technological innovations to improve reporting transparency. For example, to provide a level playing field to different types of investors, in April 2009, the SEC mandated that firms use XBRL when preparing their financial statements over three phase-in periods.

[3] We carefully review variables of interest in these studies and draw their connections to ours in Sect. 4.3.

technology is arguably more homogeneous across firms within the same industry in a given country (or countries with similar levels of economic development), individual firms exhibit variations in resource misallocation. In an ideal, frictionless state of the world, production factors are optimally allocated to arrive at the maximum productivity. Deviations from such allocation (e.g., some degree of capital or labor immobility) result in a misallocation of resources and generate a depressed level of productivity (Hsieh and Klenow 2009). For example, Bloom and Van Reenen (2007) show that management practice affects TFP using 732 medium-sized manufacturing firms in the United States, France, Germany, and the United Kingdom. Bennett et al. (2020) document a positive relationship between a firm's productivity and the price informativeness of the underlying stock.

While these studies demonstrate that firm-level characteristics collectively shape TFP, perhaps less is known about how a firm's financial reporting affects productivity. Two important recent exceptions are Banker et al. (2021) and Hann et al. (2020). Banker et al. (2021) examine the impact of the adoption of International Financial Reporting Standards (IFRS) on TFP. They show that mandatory IFRS-adopting countries experience significant increases in TFP and labor productivity. Hann et al. (2020) argue that high-quality financial reporting facilitates efficient resource allocation and show that industries with higher reporting quality exhibit lower within-industry productivity dispersion.

2.2.2 The role of EDGAR implementation on productivity

We develop our hypothesis by analyzing the role of information in the production process. In a neoclassical model, a firm keeps investing until the marginal cost exceeds the marginal benefit of a potential project. However, due to moral hazard and adverse selection, firms are likely to either overinvest or underinvest (Biddle et al. 2009). Overinvestment is primarily due to the incentive misalignment between the principal (i.e., shareholders) and the agent (the manager). The manager has an incentive to maximize personal utility and is likely to overinvest in projects that are not in the best interest of shareholders (e.g., Harford and Li 2007), especially in the case of positive free cash flow (Jensen 1986). Meanwhile, underinvestment occurs due to information asymmetry between shareholders and the manager. The manager has an information advantage regarding the firm's true state of operations over capital providers. This information advantage provides an opportunity for the manager to strategically time the event of raising capital, for example, when the firm's security is overpriced. Anticipating this strategic timing, investors may be reluctant to supply capital or charge a high hurdle rate (Myers and Majluf 1984).

The implementation of EDGAR mitigates both overinvestment and underinvestment problems. First, there is substantial reduction in the cost of information acquisition. Firm-specific information contained in periodic reports (e.g., 10-K and 10-Q reports) and event-based reports (e.g., 8-K reports) is intensively used by stakeholders in a variety of monitoring mechanisms, including proxy fights (DeAngelo 1988), debt covenants (Smith and Warner 1979), and takeovers (Palepu 1986). To the extent that the implementation of the EDGAR system allows investors to access timely firm-specific information (e.g., economic inputs and outputs) with relatively little cost, monitoring costs for the corporate outsiders are substantially reduced. As such, investors are likely to monitor more *frequently* and more *proactively* after the adoption of the EDGAR system.[4] In addition, users of the EDGAR system can employ artificial intelligence technologies (e.g., machine-based computing approaches) to download, read, analyze, and compare the corporate information of interest to them over time, and can perform comparisons between firms within and across industries (Bertomeu et al. 2020; Loughran and McDonald 2017). As such, to the extent that monitoring by stakeholders is more likely to take place with better efficacy after the introduction of EDGAR, firms are less likely to overinvest.

Second, investors rely primarily on publicly disclosed financial reports to understand operations and to coordinate investments (Leuz and Verrecchia 2004). The EDGAR system allows managers to communicate firm-specific information to capital providers in a timely manner and thus mitigates information asymmetry between managers and capital providers. Without timely corporate-specific information, investors are not able to differentiate a good firm from a bad firm and thus price the security into the "bad" type in the pooling equilibrium due to uncertainty. In contrast, with better access to firm-specific information, investors are able to identify firms with good investment opportunities and therefore provide funds accordingly. As such, firms with more readily available financial information are capable of conducting flexible financial policies by attracting financing (Chang et al. 2006) and are therefore less likely to bypass good investment opportunities due to capital insufficiency. In addition, a firm can use the new capital raised to upgrade its production technology, acquire additional state-of-the-art production lines, and improve its management practices (Bloom and Van Reenen 2007). Taken together, the implementation of the EDGAR system alleviates the problems of moral hazard and adverse selection by reducing both overinvestment and underinvestment. This improved investment efficiency in turn leads to higher productivity.

Finally, digital financial reporting makes a firm's operating results available to almost everyone in the capital market. As such, managerial performance (of the agent) is more visible to shareholders (the principal), and a poorly performing manager is more likely to be replaced (Lehn and Zhao 2006). Meanwhile, digital reporting allows peer firms (in addition to investors) to see firm-specific information more easily, at a lower cost, across a wider range of competitors. This means that any technological or other productivity advantage a firm currently has is likely to diminish with reporting in EDGAR (Glaeser 2018). Anticipating these considerations, a manager who cares about the existing contract or implicit contracts (in the context of labor market reputation) should be motivated to work harder (to search for new ways of keeping and regaining comparative advantages) and make the best use of corporate resources for value creation.[5]

[4] Shleifer and Vishny (1986) show that, monitoring will only occur when the expected benefit of monitoring outweighs the related cost of doing so.

[5] An alternative explanation is that investors, upon observing firms' operating outcomes, provide capital to the most efficient ones. We overcome this concern by employing a firm fixed effects regression strategy so that the coefficient on the EDGAR implementation indicator captures the within-firm trends in TFP.

Notwithstanding of our main prediction, it is also likely that we may not observe a positive impact of digital reporting in EDGAR on TFP.[6] Morris and Shin (2002) study the welfare effect of disclosure in the context of "a beauty test," where agents not only take actions to the underlying fundamentals but also have a coordination motive (i.e., they second-guess other agents' actions). In this setting, public information serves as a coordinating device for investors' beliefs, and greater dissemination of public information may cause investors to *over-weigh* public information and *under-weigh* private information.[7] As such, more precise public information does not necessarily lead to an increase in social welfare, especially when private information is of very high quality. In addition, the digital financial reporting platform, which makes public financial reports publicly available at no or little cost, may actually discourage investors from discovering firm-specific private information (Dugast and Foucault 2018). Taken together, it is likely that the provision of high-quality public information in EDGAR may crowd out private information discovery, production, and dissemination, making the net effect of digital financial information unclear. As such, it is an open empirical question as to whether and how digital financial reporting in EDGAR affects firms' productivity. Taken together, we state our hypothesis in the alternative form as follows:

H1: A firm's productivity increases after it first has its financial reports filed electronically in the EDGAR system.

3 Research design

3.1 Econometric issues

Our main dependent variable is total factor productivity (TFP). Conceptually, TFP measures a firm's overall effectiveness in the production process. In other words, TFP captures the efficiency of how inputs are converted into final outputs in a production function. To calculate TFP, we first consider a Cobb–Douglas production function with two inputs, as follows:

$$Y_{i,t} = A_{i,t} * {K_{i,t}}^{\beta_K} * {L_{i,t}}^{\beta_L} \tag{1}$$

To facilitate comparison, we use mathematical notation similar to Bennett et al. (2020). In Eq. (1), Y represents the value of outputs produced. K and L are inputs (i.e., capital and labor) deployed in the production function. β_K and β_L represent the output elasticities of capital and labor, respectively. A is total factor productivity. Subscripts i and t indicate firm i in year t, respectively.

[6] We thank an anonymous reviewer for this suggestion.

[7] Morris and Shin (2002) show that in the absence of private information, greater provision of public information always increases social welfare (i.e., the sum of all traders' utilities).

Taking the natural logarithm on both sides and suppressing the subscripts for simplicity, Eq. (1) is equivalent to a log-linearized variant:

$$y = \ln(A) + \beta_K * k + \beta_L * l$$

The lower-case k and l represent capital and labor in natural logarithm, respectively. To estimate this log-linearized equation, we have:

$$y = \beta_0 + \beta_K * k + \beta_L * l + u \tag{2}$$

It is straightforward that TFP (in natural logarithm) can be expressed as $\beta_0 + u$. As discussed in Ackerberg et al. (2015, henceforth "ACF"), there are two econometric unobservables in u— ω and ε. The ε term represents shocks to production or productivity that are *not* observable (or predictable) by firms *before* the firms make their input decisions at time t. Such shocks include exogenous shocks due to any natural disaster. In contrast, ω represents shocks to production that are potentially or partially observable by firms (but not observable by us as researchers/econometricians) *before* the firms make their input decisions at t. Such shocks, for example, could be the overall ability of the management team in a firm, expected defection rate of machine breakdown, expected labor disruptions due to strikes, or returned raw materials due to quality issues. As such, Eq. (2) is decomposed into:

$$y = \beta_0 + \beta_K * k + \beta_L * l + \overbrace{\underbrace{\omega}_{\substack{\textit{observable to}\\ \textit{the manager}}} + \underbrace{\varepsilon}_{\substack{\textit{unobservable to}\\ \textit{the manager}}}}^{u} \tag{3}$$

When making investment decisions, upon observing ω or a noisy measure of ω, the manager can choose k and l. In other words, under this situation, the independent variables (k and l) are correlated with the error term (u), making the coefficient estimates from an OLS regression biased. To solve endogeneity, we follow the procedures outlined in ACF to tackle the econometric issues. We explain the econometrics in detail in the appendix.

We use Compustat data to estimate the production function parameters. As explained earlier, the key research question that we ask centers on the change in TFP around the phase-in years (1993 to 1996) when EDGAR is implemented for U.S.-listed firms. As such, we use Compustat data between 1980 and 2010 to estimate TFP, with approximately 15 years before the first group's adoption and another 15 years after the last group's adoption.[8] Note that the sample that we use to estimate

[8] Results are similar if we use alternative time windows (from 1990 to 2000 and from 1960 to 2019) to estimate TFP.

TFP is "wider"—containing more years—than the test sample that we later use to test our prediction, to allow reliable estimation of the production function.

We apply the procedures described in the appendix to perform the estimation. Our estimated β_K and β_L are approximately 0.14 and 0.82, respectively, and these estimates are very close to the results reported in recent studies.[9] After the estimation, we calculate firm-level (in natural logarithm) TFP as $y - \widehat{\beta_K} * k - \widehat{\beta_L} * l$ (subscripts suppressed). We use this TFP as the main dependent variable in our empirical tests.

3.2 The baseline empirical test

Our primary prediction is that a firm's TFP increases after it has its financial reports filed in digital format in the EDGAR system. Because the implementation of EDGAR is staggered into ten groups between 1993 and 1996, we use a generalized difference-in-differences (DiD) approach (deHaan 2020). Specifically, we estimate:

$$\begin{aligned} TFP_{i,t} =& \alpha + \beta_1 Post-EDGAR_{i,t} + Firm\ characteristics \\ & + Firm\ Fixed\ effects + Year\ Fixed\ effects \end{aligned} \tag{4}$$

We designate the first partial year when a firm electronically files to the SEC as the EDGAR-adoption year and designate the subsequent year as the first year in the post-EDGAR period. In other words, the key variable of interest, *Post-EDGAR*, takes a value of one if a full firm-year is subject to mandatory filing in the EDGAR platform, and zero otherwise.[10] If the implementation of EDGAR improves firm productivity, we expect a positive coefficient on *Post-EDGAR*. Following prior studies (e.g., Bennett et al. 2020), we include a set of firm characteristic variables that may affect total factor productivity in the baseline equation. Specifically, we include *Size* (defined as the natural logarithm of total assets), *Tobin's Q* (defined as the sum of total assets plus market value of equity minus book value of equity, divided by total assets), *Leverage* (book leverage, defined as the sum of long-term debt plus the current portion of long-term debt, scaled by total assets), and *Cash* (defined as cash and cash equivalents scaled by total assets). Firm size tends to be positively associated with a firm's productivity due to economies of scale, "learning-by-doing" effects, and better human resources. We include Tobin's Q to capture a firm's growth opportunities. In a neoclassical model, Tobin's Q is the key determinant of a firm's investment policy (e.g., Yoshikawa 1980), and a higher Q is associated with a higher level of productivity.[11]

[9] İmrohoroğlu and Tüzel (2014) report their estimated β_K and β_L as being 0.23 and 0.75, respectively, using the Olley and Pakes method (discussed in Sect. 5.1).

[10] For all of the empirical tests, this definition of *Post-EDGAR* allows the financial information users to observe the publicly disseminated financial information through the EDGAR system.

[11] Ideally, we would have used the marginal Tobin's Q. However, due to the empirical challenge of estimating the marginal Tobin's Q documented in prior studies, we instead follow Biddle et al. (2009) and use the average Tobin's Q in the regression.

The impact of leverage on TFP is *ex ante* unclear. On the one hand, leverage helps discourage overinvestment by self-serving managers because of the pre-committed interest payment (Hart and Moore 1995), thus increasing efficiency. On the other hand, there are costs associated with higher leverage, such as the debt overhang problem and bankruptcy costs (Myers 1977). Such costs will have a negative impact on a firm's productivity, as the firm has to forgo positive NPV projects when additional debt financing becomes infeasible. We include cash to capture financial capacity (Chaney et al. 2012). We include capital expenditures (*Investment*, defined as capital expenditures scaled by lagged net PP&E) to capture investments in hard capital. We also include soft spending—R&D expenditure (*R&D*, defined as R&D expense scaled by sales) and advertising expenditure (*Advertising*, defined as advertising expense scaled by sales)—to control for a firm's efforts in innovation and marketing activities, respectively.[12] To capture a firm's business model and operating environment, we include *Loss incidence* (defined as a percentage of loss-making years over a five-year rolling window), *Std*(*sales*) (sales volatility, defined as the standard deviation of sales scaled by total assets over a five-year rolling window), and *Std*(*CFO*) (cash flow volatility, defined as the standard deviation of operating cash flow scaled by total assets over a five-year rolling window). Because To et al. (2018) show that analyst coverage increases TFP, we also capture the information environment using *Analysts* (defined as the natural logarithm of one plus the number of analysts provided in I/B/E/S).

Bennett et al. (2020) show that stock price informativeness positively affects TFP. Following their design, we estimate daily stock returns (e.g., $r_{j,\,i,\,t}$ for firm j in industry i in day t) on the market portfolio's returns ($r_{M,\,t}$) and industry returns ($r_{i,\,t}$).

$$r_{j,i,t} = \beta_{j,0} + \beta_{j,M} r_{M,t} + \beta_{j,i} r_{i,t} + \varepsilon$$

We collect the explanatory power R^2 from the above regression and perform a logarithm transformation on R^2 (i.e., map R^2 into $\ln(\frac{1-\boldsymbol{R}^2}{\boldsymbol{R}^2})$). Following the approach in Bennett et al. (2020) and Durnev et al. (2004), we define the average of the previous three years' logarithm-transformed R^2 as our measure of price informativeness (*PSI*). We also control for financial reporting quality following Biddle et al. (2009) and Francis et al. (2004). Specifically, we estimate a modified Dechow and Dichev (2002) regression (McNichols 2002), collect the residual term from the regression, and define accruals quality (*AQ*) as the standard deviation of the residuals over a five-year rolling window. A higher *AQ* indicates poorer mapping between accruals and cash flows, thus representing lower financial reporting quality. Our measure of financial reporting quality (*FRQ*) takes the negative of *AQ* so that a higher value of *FRQ* indicates better quality financial reporting. We also include ΔFRQ, defined as the first-order difference of financial reporting quality (*FRQ*).[13] Lastly, we include the level of institutional ownership. We obtain institutional ownership data from

[12] When firms report missing R&D or advertising expenses, we fill a value of zero for these observations. To the extent that Koh and Reeb (2015) show that firms that indeed engage in innovation may report missing R&D expenses, we assess the sensitivity of this design choice by excluding firms that report missing R&D. Results (untabulated) and inferences are qualitatively similar.

[13] We thank an anonymous reviewer for this suggestion.

Thomson financial and define *%Institutional* as the percentage of shares owned by institutional investors scaled by total shares outstanding at the end the fiscal year. For all regressions, we include firm fixed effects to make sure we are comparing within-firm trends in TFP caused by the scheduled EDGAR adoption (Bertrand and Mullainathan 2003).[14] We also include year fixed effects to help control for any sample-wide systematic differences across years. To mitigate the concern about outliers, we winsorize all continuous variables at the 1st and 99th percentiles.

4 Sample selection and baseline results

4.1 Sample selection

Following Gao and Huang (2020), we first retrieve the full list of firms in the SEC's official phase-in schedule for the implementation of the EDGAR system from Appendix B of the SEC Release No. 33–6977 (released on February 23, 1993). Because this list is only available in a scanned PDF file with a total of 159 pages, we use the text recognition function in Adobe's Acrobat software to restore firm names and CIK (Central Index Key) numbers into a digital format. To ensure text recognition quality, we manually inspect the places where the original characters in the scanned PDF file are unclear or create ambiguous text recognition outcomes. To the extent that the SEC recycles and reassigns CIK numbers, to match firms in the schedule, we employ the official dictionary of firm identifiers from WRDS. This dictionary provides detailed matching information between each firm's historical name and historical CIK number, with validated start and end dates. After this step, each firm in the EDGAR implementation schedule is matched with a unique GVKEY identifier. We remove firms without a matched GVKEY because they cannot be matched with fundamental data. We then merge this matched sample with Compustat annual data.

We next remove firms in financial and utility industries, as these firms may have a production function distinct from that of industrial firms. Finally, we keep observations between 1990 and 2000, leaving, in our final sample, four years of data prior to the start of the implementation of EDGAR and four years of data after the completion of the implementation. After requiring that the necessary data be available for all test variables, the final sample comprises 21,342 firm-year observations from 1990 through 2000. We present the sample selection procedures in Panel A of Table 1. We also present the detailed phase-in schedule of the EDGAR implementation and the number of unique firms in each group in the schedule in Panel B of Table 1. Overall, the distribution of firms across groups is very similar to the one reported in Fig. 1 in Gao and Huang (2020).

Table 2 presents the summary statistics for the main variables. The average *TFP* (in natural logarithm) is −0.22, and the magnitude is consistent with those reported by recent studies using similar approaches, such as To et al. (2018). The indicator

[14] In alternative specifications, we use industry and group fixed effects. Results and inferences are qualitatively similar. Because firm fixed effects are finer than and subsume both the group fixed effects and industry fixed effects, we use firm fixed effects in all tests.

Table 1 Sample selection and the distribution of observations

Panel A: Sample selection procedures

Procedures	# of unique firms	# of firm-years
Obtain firms covered in the SEC's phase-in schedule	11,354	–
Less sample firms without valid GVKEY	(4,165)	–
Less sample firms not matched with Compustat	(2,641)	41,596
Less firms in financial and utility industries	(1,059)	(9,487)
Less firms without necessary financial characteristics or TFP measures	(608)	(10,767)
The final sample	2,881	21,342

Panel B: The phase-in schedule of the EDGAR implementation

Group	Mandatory Date	# of identities in the SEC schedule	# of unique firms in the final sample	# of firm-years in the final sample
CF-01	04/26/93	232	88	829
CF-02	07/19/93	727	308	2,831
CF-03	10/04/93	682	335	2,840
CF-04	12/06/93	892	437	3,765
CF-05	08/01/94	1,383	578	4,841
CF-06	11/01/94	1,375	459	3,341
CF-07	05/01/95	1,367	216	1,178
CF-08	08/01/95	1,348	84	342
CF-09	11/01/95	1,253	24	100
CF-10	05/01/96	2,095	352	1,275
Total		11,354	2,881	21,342

Panel A describes the sample selection process

Panel B describes the ten groups specified in the SEC official release. We list the number of unique firms and unique firm-year observations for each group, respectively

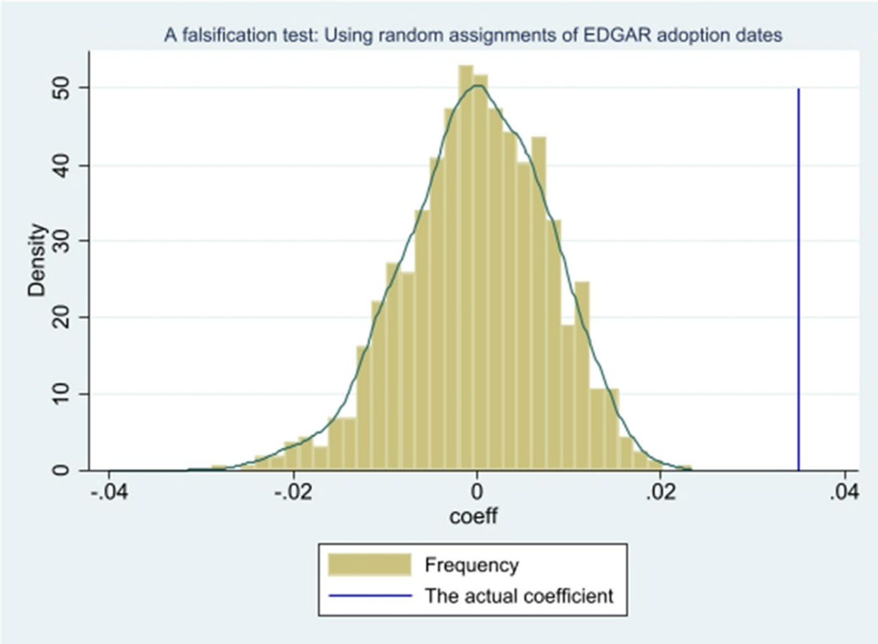

Fig. 1 The falsification analysis using random assignments of EDGAR adoption dates. This figure presents the results from the falsification analysis where we randomly generate pseudo-EDGAR adoption dates. We apply the pseudo-EDGAR adoption dates and re-estimate the baseline equation (Eq. (4)). We then repeat the exercise 1,000 times. We obtain the coefficient estimate on *Pseudo Post-EDGAR* and plot the distribution using a histogram with 30 bins. For comparison, we also use the vertical line (in blue) to plot the regression coefficient on *Post-EDGAR* obtained using the actual EDGAR adoption dates (from column (2) in Panel A of Table 3)

variable *Post-EDGAR* has a mean value of 0.48, suggesting that, as desired, approximately half of the firm-year observations are associated with electronic filings in the EDGAR platform. Overall, the distribution of key firm characteristic variables is comparable to that of other studies (e.g., Bennett et al. 2020; To et al. 2018).

4.2 Baseline results

We present the results of estimating Eq. (4) in Panel A of Table 3. The dependent variable is *TFP*. The key variable of interest is *Post-EDGAR*. We estimate *TFP* on *Post-EDGAR* with a subset of firm characteristics plus firm fixed effects and year fixed effects in column (1). In column (2), we add stock price informativeness (*PSI*), financial reporting quality (*FRQ*), the change in financial reporting quality (Δ*FRQ*), and institutional ownership (*%Institutional*). The coefficients on *Post-EDGAR* are positive and statistically significant at the 1% level or better in both columns. This is consistent with H1, insofar as firm productivity improves when firms electronically file in the EDGAR system. In particular, in column (2), the coefficient on *Post-EDGAR* is 0.035. Because Eq. (4) is in semi-elasticity form, this coefficient suggests

Table 2 Summary statistics

Variables	N	Mean	P25	Median	P75	Std Dev
TFP	21,342	−0.22	−0.46	−0.20	0.04	0.50
Post-EDGAR	21,342	0.48	0.00	0.00	1.00	0.50
Size	21,342	5.54	4.11	5.36	6.79	1.88
Tobin's Q	21,342	1.69	1.05	1.36	1.92	1.07
Cash	21,342	0.11	0.02	0.05	0.15	0.13
Leverage	21,342	0.24	0.08	0.23	0.36	0.19
R&D	21,342	0.03	0.00	0.00	0.03	0.05
Advertising	21,342	0.01	0.00	0.00	0.01	0.02
Investment	21,342	0.28	0.13	0.21	0.35	0.25
Analysts	21,342	1.39	0.00	1.39	2.30	1.12
Loss incidence	21,342	0.20	0.00	0.00	0.40	0.27
Std(Cash flow)	21,342	0.06	0.03	0.05	0.08	0.05
Std(Sales)	21,342	0.28	0.11	0.19	0.35	0.29
PSI	21,342	3.19	2.08	3.27	4.40	1.60
FRQ	21,342	−0.05	−0.07	−0.04	−0.02	0.04
ΔFRQ	21,342	0.00	−0.01	0.00	0.01	0.02
%Institutional	21,342	0.38	0.16	0.37	0.57	0.24

This table presents descriptive statistics for the baseline sample. *TFP* is the total factor productivity in natural logarithm calculated following Ackerberg et al. (2015). *Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing in the EDGAR platform and zero otherwise. *Size* is natural logarithm of total asset. *Tobin's Q* is the sum of total assets plus market value of equity minus book value of equity divided by total assets. *Cash* is defined as cash equivalent scaled by total assets. *Leverage* is defined as the sum of long-term debt plus the current portion of long-term debt, scaled by total assets. *Investment* is defined as capital expenditure scaled by lagged net PP&E. *R&D* and *Advertising* are the firm's R&D expense and advertising expense scaled by sales, respectively. *Loss incidence* is defined as the percentage of loss-making years over a five-year rolling window. *Analysts* is defined as the natural logarithm of one plus the number of analysts provided in I/B/E/S. *Std*(*Cash flow*) is cash flow volatility, defined as the standard deviation of operating cash flow scaled by total assets over a five-year rolling window. *Std*(*Sale*) is sales volatility, defined as the standard deviation of sales scaled by total assets over a five-year rolling window. *PSI* is stock price informativeness, defined following Bennett et al. (2020). *FRQ* captures financial reporting quality and is defined as the negative of accruals quality (*AQ*). We calculate accruals quality following Francis et al. (2004). Δ*FRQ* is defined as the first-order difference of financial reporting quality (*FRQ*). *%Institutional* is the percentage of shares owned by institutional investors scaled by total shares outstanding at the end the fiscal year

that the implementation of the EDGAR system increases TFP by 3.5%. This magnitude is economically significant, comparable to the effect of a one standard deviation increase in the management practice score (Bloom and Van Reenen 2007).[15] Results in Panel A of Table 3 also reveal that the control variables are associated

[15] We also benchmark our coefficient on *Post-EDGAR* against To et al. (2018). In their baseline regression, the coefficient on their main independent variable, *LnCoverage*, is 0.037 (column (3) in Table 3, with firm-level characteristics, year, and industry fixed effects included). In their Table 2, they report that the standard deviation of *LnCoverage* is 0.933. That is, a one standard deviation change in *LnCoverage* results in an increase in TFP of 0.034. In other words, the introduction of digital reporting in EDGAR results in an increase in TFP that is very comparable to, and slightly higher than, a one standard deviation change in analyst coverage.

Table 3 The baseline test

	(1) TFP	(2) TFP
Panel A: Using an indicator variable Post-EDGAR		
EDGAR indicator		
Post-EDGAR	0.037***	0.035***
	(3.60)	(3.40)
Firm characteristics		
Size	0.109***	0.110***
	(11.59)	(11.30)
Tobin's Q	0.079***	0.079***
	(19.21)	(19.25)
Cash	0.116***	0.116***
	(2.73)	(2.72)
Leverage	–0.163***	–0.159***
	(–4.95)	(–4.83)
R&D	–2.629***	–2.540***
	(–13.19)	(–12.77)
Advertising	–0.588*	–0.603*
	(–1.72)	(–1.76)
Investment	0.090***	0.087***
	(6.41)	(6.25)
Analysts	0.017***	0.018***
	(3.14)	(3.12)
Loss incidence	–0.247***	–0.240***
	(–11.60)	(–10.90)
Std(Cash flow)	–0.101	–0.107
	(–0.98)	(–0.98)
Std(Sales)	0.062***	0.062***
	(3.79)	(3.70)
PSI		0.011***
		(2.88)
FRQ		0.105
		(0.75)
ΔFRQ		0.721***
		(4.53)
%Institutional		0.067**
		(2.31)
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
N	21342	21342
adj. R-sq	0.766	0.767

Table 3 (continued)

	(1) TFP	(2) TFP
Panel B: Using a dynamic window		
EDGAR indicators		
$EDGAR^{-3}$	–0.012	
	(–0.97)	
$EDGAR^{-2}$	−0.019	
	(–0.90)	
$EDGAR^{-1}$	0.013	
	(0.45)	
$EDGAR^{0}$	0.037	
	(1.02)	
$EDGAR^{+1}$	0.076*	
	(1.78)	
$EDGAR^{+2}$	0.105**	
	(2.17)	
$EDGAR^{3+}$	0.113**	
	(2.09)	
Firm characteristics	Yes	
Firm fixed effects	Yes	
Year fixed effects	Yes	
N	21342	
adj. R-sq	0.767	

Panel A presents the baseline result of estimating productivity on the EDGAR-implementation indicator. The dependent variable is *TFP*, calculated following Ackerberg et al. (2015). *Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing in the EDGAR platform and zero otherwise. *Size* is natural logarithm of total asset. *Tobin's Q* is the sum of total assets plus market value of equity minus book value of equity divided by total assets. *Cash* is defined as cash equivalent scaled by total assets. *Leverage* is defined as the sum of long-term debt plus the current portion of long-term debt, scaled by total assets. *Investment* is defined as capital expenditure scaled by lagged net PP&E. *R&D* and *Advertising* are the firm's R&D expense and advertising expense scaled by sales, respectively. *Loss incidence* is defined as the percentage of loss-making years over a five-year rolling window. *Analysts* is defined as the natural logarithm of one plus the number of analysts provided in I/B/E/S. Std(*Cash flow*) is cash flow volatility, defined as the standard deviation of operating cash flow scaled by total assets over a five-year rolling window. Std(*Sale*) is sales volatility, defined as the standard deviation of sales scaled by total assets over a five-year rolling window. *PSI* is stock price informativeness, defined following Bennett et al. (2020). *FRQ* captures financial reporting quality and is defined as the negative of accruals quality (*AQ*). We calculate accruals quality following Francis et al. (2004). Δ*FRQ* is defined as the first-order difference of financial reporting quality (*FRQ*). *%Institutional* is the percentage of shares owned by institutional investors scaled by total shares outstanding at the end the fiscal year. We include firm fixed effects and year fixed effects. *T*-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Panel B presents the effect of EDGAR-implementation on firm productivity using a dynamic approach. Specifically, we define a set of seven year indicators. $EDGAR^{-3}$, $EDGAR^{-2}$, $EDGAR^{-1}$, $EDGAR^{0}$, $EDGAR^{+1}$, $EDGAR^{+2}$, and $EDGAR^{3+}$ take a value of one if the firm will adopt EDGAR in three years, will adopt EDGAR in two years, will adopt EDGAR in one year, adopted EDGAR in the current year, adopted EDGAR in the previous year, adopted EDGAR two years ago, and adopted EDGAR three years or more than three years ago, respectively, and zero otherwise. We include all firm characteristic control variables with firm fixed effects and year fixed effects. Variable definitions are presented in Table 2. *T*-statistics are in parentheses and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively with productivity in manners that are consistent with the findings in prior studies. The coefficient on *Size* is positive and statistically significant at the 1% level, consistent with findings, in prior studies, that larger firms tend to have higher productivity. The coefficient on *Tobin's Q* is positive and statistically significant at the 1% level, consistent with the idea that future growth opportunities enhance productivity. The coefficient on *Cash* is positive and statistically significant at the 1% level. The coefficient on *R&D* is negative and statistically significant at the 1% level. This may be consistent with the research design, in which both *TFP* and *R&D* are measured in the concurrent period. Conceptually, because the innovation cycle of an R&D activity usually extends beyond one fiscal year, innovations induced by *successful* R&D activities enhance the *long-term* TFP but not necessarily the contemporaneous year's TFP.[16] The coefficient on *Advertising* is also negative, and is statistically significant at the 10% level. The coefficient on *Analysts* is positive, consistent with findings in To et al. (2018). The coefficient on *Loss incidence* is negative, consistent with the idea that firms that frequently incur losses have lower TFP. The coefficient on *Std*(*Sales*) is positive and statistically significant at the 1% level, suggesting that firms with higher sales volatility have higher TFP. The coefficient on *FRQ* is positive and statistically insignificant at the conventional levels. The coefficient on ΔFRQ is positive and statistically significant at the 1% level, consistent with the findings in Biddle et al. (2009) that firms with deteriorating financial reporting quality suffer the agency problem and thus experience lower TFP.[17] Taken together, the results in Panel A of Table 3 provide support to H1, suggesting that the implementation of EDGAR increases firms' TFP when they file their financials electronically in the post-EDGAR period.

To show the dynamics of the effect of EDGAR implementation on TFP, we introduce several year indicator variables around the specified dates in the implementation schedule. Specifically, $EDGAR^{-3}$, $EDGAR^{-2}$, $EDGAR^{-1}$, $EDGAR^{0}$, $EDGAR^{+1}$, $EDGAR^{+2}$, and $EDGAR^{3+}$ take a value of one if the firm will adopt EDGAR in three years, will adopt EDGAR in two years, will adopt EDGAR in one year, adopted EDGAR in the current year, adopted EDGAR in the previous year, adopted EDGAR two years ago, and adopted EDGAR three or more years ago, respectively, and zero otherwise. If EDGAR implementation enhances TFP, we should then observe positive coefficients for $EDGAR^{+i}$ but insignificant coefficients for $EDGAR^{-i}$ and $EDGAR^{0}$. We replace *Post-EDGAR* in Eq. (4) with seven indicator variables

[16] Lev and Sougiannis (1996) show that imputed R&D capitalization contains value-relevant information for investors, but contemporaneous stock prices do not *fully* reflect R&D capital. Stock returns in subsequent years are also positively associated with their imputed R&D capital.

[17] In untabulated tests, we capture financial reporting quality using an alternative, disclosure-based index motivated by Chen et al. (2015). We define *Disclosure* as the simple average of balance sheet disclosure quality (DQ^{BS}) and income statement disclosure quality (DQ^{IS}). We follow the procedures outlined in Chen et al. (2015) to calculate DQ^{BS} and DQ^{IS}. A higher value of *Disclosure* signifies better disaggregation of accounting data and thus higher disclosure quality. For completeness, we also include the first-order difference of this disclosure score (Δ*Disclosure*). The coefficient on *Disclosure* is positive and statistically significant at the 1% level, consistent with firms with better reporting quality experiencing higher productivity (Biddle et al. 2009). Most importantly, the coefficient on *Post-EDGAR* remains statistically significant at the 1% level.

to capture the years around the adoption year and present the results in Panel B of Table 3. Whereas the pre-EDGAR years are employed as the base years in Table 3 in Panel A, the benchmark period in Panel B includes all years prior to year –3. The coefficients on $EDGAR^{+1}$, $EDGAR^{+2}$, and $EDGAR^{3+}$ are positive and statistically significant at the 10% level or better. When a firm electronically files its financials in EDGAR in year 0, the impact of EDGAR on TFP occurs in year + 1, and the effect persists in the post-EDGAR years. Overall, the results support the interpretation that the implementation of EDGAR results in an increase in TFP, and the improvement is not transient.

4.3 Including additional variables examined in concurrent studies

As discussed earlier, a number of concurrent studies also use the EDGAR implementation setting to help infer causality of digital financial reporting on capital market outcomes. *Ex ante*, prior to any empirical testing, we acknowledge that it is possible that the implementation of EDGAR improves firms' TFP through economic constructs such as cost of equity capital, stock price informativeness, and financial reporting quality. To further guard against the possibility that the effect of EDGAR implementation on TFP *completely* goes through variables that have been shown to be affected by EDGAR, we add these variables gradually as additional controls in the baseline regression.

We control for changes in equity financing (*Equity Proceeds*) using the proceeds from the sale of common shares scaled by total assets. We control for liquidity using the bid-ask spread (*Spread*). Following Butler et al. (2005) and Welker (1995), we define the bid-ask spread (*Spread*) using the mean daily difference between ask and bid prices scaled by the midpoint of ask and bid prices. We control for the incidence of stock price crashes following Kim et al. (2011) and define *Crash* as an indicator variable that takes a value of one if a firm-year experiences at least one crash week.[18] We control for informed trading using the probability of informed trading (PIN). The PIN score is the probability of an information-based trade derived from a structural market microstructure model (Easley and O'Hara 1992; Easley et al. 2010).[19] Following Bennett et al. (2020), we define *PIN* as the average of the previous three years' PIN scores as our measure of informed trading. We control for analyst forecast quality using analyst forecast accuracy (*Accuracy*), defined as the

[18] We first compound daily returns into weekly returns for each firm and year. We then estimate an expanded market model regression (with weekly returns) with a lead and a lag term for the market index return. The firm-specific weekly return is defined as the natural log of one plus the residual return. We define a week as a crash week when a firm experiences a firm-specific weekly return that is 3.2 standard deviations below the mean firm-specific weekly return over the entire fiscal year, with 3.2 chosen to generate a frequency of 0.1% in the normal distribution.

[19] We estimate the PIN score using a combined approach. For observations after the availability of NYSE Trade and Quote (TAQ) data (since 1993), we use the intraday transaction-level data provided by TAQ to estimate the PIN score. For observations prior to 1993 (when the transaction-level data from the Institute for the Study of Security Markets (ISSM) is not available), we directly use the PIN data (Easley, Hvidkjaer, and O'Hara 2010) provided directly from Professor Soeren Hvidkjaer's website.

negative of the absolute value of the mean analyst forecast error scaled by the last fiscal year-end's stock price. Adding this variable requires that a given firm-year has at least one analyst forecast. Finally, we control for cost of equity capital (*CoE*) using the implied cost of equity estimate based on the Claus and Thomas (2001) approach.[20]

We present the results in Table 4 in the text. Because including these additional variables results in additional sample attrition, we add these variables in an order that preserves the largest sample size first, then include additional variables one at a time. Column (1) reproduces the results shown in column (2) in Panel A of Table 3 and serves as the benchmark. We add additional variables one at a time in subsequent columns. The coefficients on these additional variables are, in general, consistent with theory and intuition. For example, consistent with Bennett et al. (2020), the coefficient on *PIN* in column (5) is positive and statistically significant at the 5% level. The coefficient on *CoE* is negative (albeit less significant at the conventional levels), consistent with the idea that firms with cheaper financing experience higher TFP. Most importantly, the coefficient on *Post-EDGAR* remains economically large and statistically significant at the 5% level or better. As such, we conclude cautiously that the effect of EDGAR implementation on TFP does not completely go through these variables.

4.4 A design to tackle potential non-randomness in the SEC's grouping

Another possible concern with the baseline design is that the SEC's grouping may not be completely random.[21] Bertrand and Mullainathan (2003) and Atanasov and Black (2016) discuss the potential non-random assignment to treatment groups, which is common in much of the literature that exploits regulatory shocks, and establish that the non-randomness per se does not necessarily invalidate the DiD design. In particular, Chang et al. (2021, p. 2) provide that "Scott Bauguess, then Acting Chief Economist of the SEC, informed us that the wave assignments were *randomized conditional on firm size* [emphasis added]." In other words, when implementing the adoption, the SEC grouping assignment is *conditionally* random. To further mitigate the potential effect of firm size in this quasi-randomized grouping, we employ a size-based matching approach, i.e., matching a treatment firm (i.e., an EDGAR switcher) with a similar-sized control firm (a non-switcher). To do so, for each wave (i.e., the treatment group with the scheduled switch date) out of the ten wave groups, we construct a control group that consists of firms that are similar in size but do not experience any switch within 12 months after the implementation date of that group. We repeat this exercise

[20] Claus and Thomas (2001) employ a residual-income valuation model that adds current book value per share with discounted expected residual earnings per share up to five years. They assume residual earnings grow at an expected inflation rate minus 3% after the five-year forecast horizon and that 50% of earnings are paid out as dividends each period. We obtain the inflation data from the Federal Reserve Bank of St. Louis and then solve the valuation equation to obtain the discount factor as an estimate of the implied cost of equity.

[21] Both Gao and Huang (2020) and Chang et al. (2021) describe their (private) correspondences with the SEC staff.

for Group 1 to Group 7, with each group of treatment firms matched with the one most similar in size (measured using the market value of equity). Because all firms switch to digital reporting sooner or later over the four-year window between 1993 and 1996, the available universe of control firms becomes smaller as we move from earlier waves (e.g., Groups 1 and 2) to later waves (e.g., Groups 6 and 7). As such, we perform the matching algorithm with the first seven groups. We leave the treatment firms in Groups 8 to 10 unmatched because firms in the last three groups cannot be matched with similar-sized peer firms (since the time between the implementation date for Group 8 and the implementation date for Group 10 is less than one year).

We report the summary statistics for firm size (measured using the market value of equity in natural logarithm) in Panel A of Table 5. Before the matching, it is clear that firm size differs between the treatment firms (the EDGAR adopters) and the control firms (firms that have not yet adopted reporting in EDGAR). This substantial difference is consistent with the statement made in Chang et al. (2022). The difference in firm size becomes insignificant after the matching procedure, suggesting an effective control for firm size using the procedures described in the last paragraph.

We employ the propensity-matched sample to re-estimate the baseline Eq. (4) with *Post-EDGAR* replaced by *EDGAR Adoption*. The treatment firms are EDGAR adopters, and the control firms are firms that have not yet adopted digital reporting. *EDGAR Adoption* takes a value of one for treatment firms and zero otherwise. Because we perform the matching with replacement, it is likely that our final matched sample contains control observations that appear more than once.[22] To assess the potential effect of this approach, we further create a more restrictive sample by removing repeating control observations. In other words, although, in the matching algorithm, we allow for the possibility that one treatment firm can be matched with more than one control observation, each control observation receives a weight of $1/N$ when we run the regression on this more restrictive sample, where N is the number of *unique* firm-year observations. We alternately use two calipers (0.05 and 0.1) and present the results in Panel B of Table 5. The coefficients on *EDGAR Adoption* remain positive and statistically significant at the 1% level across all columns, mitigating the concern that the results are driven by the impact of firm size on the SEC's group assignment.

5 Robustness analyses

5.1 Alternative specifications using the approach in Olley and Pakes (1996)

In the baseline test, we have followed Ackerberg et al. (2015) to estimate TFP. We next assess the robustness of our results using alternative econometric specifications. First, we employ the approach proposed by Olley and Pakes (1996, henceforth

[22] Shipman et al. (2017) provide excellent discussions on matching with and without replacement. Matching without replacement may result in lower-quality matches.

Table 4 The connection to concurrent studies

	(1) TFP	(2) TFP	(3) TFP	(4) TFP	(5) TFP	(6) TFP	(7) TFP
EDGAR indicator							
Post-EDGAR	0.035***	0.034***	0.034***	0.035***	0.027***	0.028**	0.032**
	(3.40)	(3.33)	(3.30)	(3.39)	(2.62)	(2.36)	(2.50)
Additional controls							
Equity financing		−0.343***	−0.350***	−0.342***	−0.776***	−1.227***	−1.683***
		(−3.01)	(−3.08)	(−2.95)	(−4.76)	(−4.36)	(−5.97)
Spread			−0.156***	−0.153***	−0.133**	−0.080	−0.001
			(−2.66)	(−2.61)	(−2.04)	(−1.09)	(−0.01)
Crash				−0.019***	−0.014**	−0.009	−0.008
				(−3.50)	(−2.49)	(−1.52)	(−1.31)
PIN					0.133**	0.029	−0.056
					(2.05)	(0.30)	(−0.52)
Accuracy						0.153***	0.113
						(3.00)	(1.60)
CoE							−0.059
							(−1.06)
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	21,342	21,342	21,342	21,317	18,715	13,718	11,444
adj. R-sq	0.767	0.767	0.768	0.768	0.771	0.789	0.803

This table presents the result of adding additional variables examined in concurrent studies using the EDGAR implementation setting. *Equity Proceeds* is proceeds from the sale of common shares scaled by total assets. *Spread* is the bid-ask spread (*Spread*), defined as the mean daily difference between ask and bid prices scaled by the midpoint of ask and bid prices. *Crash* is an indicator variable that takes a value of one if a firm-year experiences at least one crash week, following Kim et al. (2011). *PIN* is the probability of informed trading derived from a structural market microstructure model (Easley and O'Hara 1992; Easley et al. 2010). *Accuracy* is defined as the negative of the absolute value of the median analyst forecast error scaled by the last fiscal year-end's stock price. *CoE* is cost of equity capital using the implied cost of equity estimate based on the Claus and Thomas (2001) approach. The dependent variable is *TFP*, calculated following Ackerberg et al. (2015). *Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing in the EDGAR platform and zero otherwise. We include all firm characteristics, firm fixed effects, and year fixed effects. T-statistics are in parentheses and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

"OP"). Ackerberg et al. (2015) explain that their own estimation is free of the functional dependence problem and provides an improved estimation of the production function over the specification in Olley and Pakes (1996). Nevertheless, to assess the sensitivity of our results, we repeat the exercise in the baseline test using the

Table 5 Propensity score matching

Panel A: The mean values of firm size before and after matching

	(1)	(2)	(3)	(4)	(5)	(6)
			Caliper = 0.05		Caliper = 0.1	
	Pre-Match		Post-Match		Post-Match	
ln(*MVE*)	Treatment	Control	Treatment	Control	Treatment	Control
	5.25	3.77	4.89	4.89	4.88	4.88
	(30.62)		(0.00)		(−0.00)	

Panel B: The baseline test using the matched sample

	(1)	(2)	(3)	(4)
	TFP			
	Caliper = 0.05		Caliper = 0.1	
Allow repeating control observations	Yes	No	Yes	No
EDGAR indicator				
EDGAR Adoption	0.039***	0.037***	0.038***	0.036***
	(2.87)	(2.84)	(3.24)	(3.18)
Firm characteristics	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	12,374	11,429	15,755	14,615
adj. R-sq	0.732	0.730	0.730	0.727

Panel A presents the mean firm size (measured using the market value of equity, ln(*MVE*)) before and after the size-based matching algorithm. The treatment firms are EDGAR adopters, and the control firms have not yet adopted reporting in EDGAR. We alternately use two calipers, 0.05 and 0.1. T-statistics presented in the parentheses, and ***, **, and * indicate that the difference is significantly different from zero at the 1%, 5%, and 10% levels, respectively

Panel B presents the results of estimating baseline Eq. (4) using the matched sample, with *Post-EDGAR* replaced by *EDGAR Adoption*. The treatment firms are EDGAR adopters, and the control firms are firms that have not yet adopted reporting in EDGAR. *EDGAR Adoption* takes a value of one for treatment firms and zero otherwise. We alternately use two calipers, 0.05 and 0.1. For each caliper, we perform the matching with replacement. We also create a more restrictive sample by removing repeating control observations. In other words, although in the matching algorithm we allow for the possibility that one treatment firm can be matched with more than one control observation, each control observation receives a weight of 1/*N* when we run the regression on this restrictive sample, where *N* is the number of *unique* firm-year observations. Variable definitions are presented in Table 2. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

OP approach to estimate the production function and designate the TFP estimate as TFP_{OP}. We present the results in Panel A of Table 6. Because we require capital investment in estimating TFP_{OP}, the sample in this test is slightly smaller than our baseline sample. The coefficient of *Post-EDGAR* is positive, of a similar magnitude, and statistically significant at the 1% level.

Table 6 Robustness tests

Panel A: Using an alternative measurement of TFP

	(1)
	TFP_{OP}
EDGAR indicator	
Post-EDGAR	0.031***
	(3.07)
Firm characteristics	Yes
Firm fixed effects	Yes
Year fixed effects	Yes
N	20,858
adj. R-sq	0.889

Panel B: Removing transitional filers

	(1)
	TFP
	Remove firms in the CF-01 group
EDGAR indicator	
Post-EDGAR	0.034***
	(3.22)
Firm characteristics	Yes
Firm fixed effects	Yes
Year fixed effects	Yes
N	20,513
adj. R-sq	0.762

Panel C: Correcting for initial availability of EDGAR filings

	(1)	(2)
	TFP	TFP
EDGAR indicator		
Post-EDGAR	0.036***	0.036***
	(3.53)	(3.47)
Interim		−0.061
		(−1.08)
Firm characteristics	Yes	Yes
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
N	21,342	21,342
adj. R-sq	0.799	0.799

Panel D: Adding additional variables to control for information environment

	(1)	(2)	(3)
	TFP		
EDGAR indicator			
Post-EDGAR	0.035***	0.038***	0.038***
	(3.40)	(3.20)	(3.21)
Information environment controls			
ΔAnalysts	0.022***		0.021***
	(3.51)		(2.92)
Accuracy		0.156**	0.156**
		(2.21)	(2.19)

Table 6 (continued)

ΔAccuracy		0.000	−0.001
		(0.01)	(−0.03)
Firm characteristics	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
N	21,342	15,130	15,130
adj. R-sq	0.799	0.825	0.825
Panel E: Falsification analysis			
	(1)	(2)	(3)
	TFP		
	100 rounds	500 rounds	1000 rounds
EDGAR indicator			
Pseudo Post-EDGAR	−0.001	−0.000	−0.000
	(−1.24)	(−1.10)	(−0.21)
Significant at the 5% level	0	1	7
Significant at the 1% level	0	0	1
Firm characteristics	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes

Panel A presents the result of estimating productivity on the EDGAR-implementation indicator. We use an alternative method of estimating productivity following Olley and Pakes (1996) and indicate the dependent variable as TFP_{OP}. *Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing in the EDGAR platform and zero otherwise. We include firm fixed effects and year fixed effects. Variable definitions are described in Table 3. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Panel B presents the results after removing transitional filers. We remove firms assigned to Group CF-01 and re-estimate Eq. (4). We include firm characteristic control variables with firm fixed effects and year fixed effects. Variable definitions are presented in Table 2. T-statistics are in parentheses and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Panel C presents the results of accounting for the initial cost of accessing EDGAR online. In column (1), we employ an alternative date of January 17, 1994, as the adoption date for the first four groups of firms to define *Post-EDGAR*. In column (2), we keep the original coding of *Post-EDGAR* (using the SEC scheduled dates described in the baseline test) and employ a separate indicator variable, *Interim*, which takes a value of one if the firm-year falls into the interim periods between the dates in the SEC scheduled dates and January 17, 1994, for the first four groups of companies, respectively, and zero otherwise. We include firm characteristic control variables with firm fixed effects and year fixed effects. Variable definitions are presented in Table 2. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Panel D presents the result of adding additional variables to control for information environment. We include Δ*Analysts* (defined as the first-order difference of the number of analysts following), *Accuracy* (defined as the negative of the absolute value of the median analyst forecast error, scaled by the stock price at the last fiscal year-end), and Δ*Accuracy* (defined as the first-order difference of *Accuracy*). We include firm characteristic control variables with firm fixed effects and year fixed effects. Variable definitions are presented in Table 2. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Panel E presents the result of a falsification analysis using randomly generated pseudo adoption dates. Specifically, we randomly select a pseudo adoption date between April 26, 1993 (the scheduled adoption date of the first group), and May 1, 1996 (the scheduled adoption date of the last group), and use the same procedure outlined in Sect. 4.1 to define the indicator variable *Pseudo Post-EDGAR*. *Pseudo Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing on EDGAR using the randomly generated pseudo date and zero otherwise. We replace *Post-EDGAR* with *Pseudo Post-EDGAR* and perform the estimation 100, 500, and 1,000 times, respectively. We report the test of the significance of coefficients using Zellner's seemingly unrelated regressions. Variable definitions are presented in Table 2. T-statistics are in parentheses, and superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

5.2 Remove transitional filers

While testing the EDGAR system before its scheduled official start in 1993, the SEC encouraged a small set of firms to voluntarily submit fillings through EDGAR. These voluntary filers were assigned to Group CF-01 exclusively in the SEC's EDGAR adoption schedule. During this "test drive" period, these "transitional filers" were given the opportunity to decide which forms to file electronically to the SEC and then file the remainder in paper-based forms. In April 1993, when the "test drive" period ended, all voluntary filers were mandated to deliver the required fillings to the SEC electronically through the EDGAR system. To mitigate the concern that some firms in Group CF-01 are transitional filers and that their discretion to file electronically or on paper introduces self-selection bias, we remove firms assigned to Group CF-01 and re-estimate Eq. (4).[23] We present the result in Panel B of Table 6. The coefficient on *Post-EDGAR* remains positive and statistically significant at the 1% level, alleviating the concern about self-selection by transitional filers.

5.3 Correct for initial availability of EDGAR filings

EDGAR is now accessible for free to all users. However, this was not the case when the EDGAR system was first introduced to the public. As discussed in Kambil and Ginsburg (1998), when the SEC started EDGAR, corporate filings on EDGAR were electronically accessible, for a usage fee, through the commercial data vendor Mead Data Central (MDC) (renamed Lexis-Nexis in 1994).[24] Free EDGAR access started later, when New York University's Stern School of Business obtained a grant from the National Science Foundation in late 1993 in collaboration with Internet Multicasting Services (IMS), a non-profit organization. Beginning on January 17, 1994, IMS provided servers for hosting the data through a high-speed internet connection to EDGAR at the NYU-hosted URL http://edgar.stern.nyu.edu (no longer hosted).

To account for the possibility that the subscription fee charged by MDC may in fact have prevented some users from accessing EDGAR, we perform two sensitivity tests using alternative designs. First, we employ January 17, 1994, as the adoption date for the first four groups of firms. In other words, we use the later date (January 17, 1994) when EDGAR became accessible free of charge to define *Post-EDGAR* for the first four groups. We present the results in column (1) of Panel C of Table 6. Second, we employ a separate indicator variable, *Interim*, which takes a value of one if the firm-year falls into the interim periods between the dates specified in the schedule and January 17, 1994, for the first four groups of companies, and zero otherwise. This variable captures the effect of EDGAR becoming available to *some* users (for a usage fee) but not accessible for *all* users. We present the results in column (2) of Panel C of Table 6. In both columns, the coefficients on

23 This design choice is conservative. While *not all* firms in Group CF-01 are transitional filers, *all* transitional filers are assigned to Group CF-01.

24 The annual subscription fee for real-time data through Mead Data Central was \$150,000. The non-real-time subscription fee was \$75,000. The pay-as-you-go per filing rate ranged between \$20 and \$30.

Post-EDGAR remain positive and statistically significant at the 1% level. The coefficient on *Interim* is negative and statistically insignificant at conventional levels in column (2), again confirming the baseline findings.

5.4 Additional controls for information environment

Our baseline test controls for analyst coverage. It is likely that not only the level of analyst coverage but also the change in analyst coverage plays a role in monitoring. In addition, we capture analyst quality using forecast accuracy (*Accuracy*), defined as the negative of the absolute value of the median analyst forecast error scaled by the stock price at the last fiscal year-end. Finally, we include the change of analyst forecast accuracy (Δ*Accuracy*) to capture the possible improvement in the information environment induced by EDGAR implementation. We present the results in Panel D of Table 6. Including Δ*Accuracy* results in additional sample attrition (about 30% of the sample size in the baseline sample), as it requires that an observation has analyst coverage in both periods (the contemporary and the lagged year). In column (1), the coefficient on Δ*Analysts* is positive and statistically significant at the 1% level, consistent with the idea that firms with improved coverage experience higher TFP. In columns (2) and (3), the coefficients on *Accuracy* are positive and statistically significant at the 5% level, consistent with the idea that firms with a better information environment, provided by analysts, experience higher TFP. Most importantly, the coefficient on *Post-EDGAR* remains economically large and statistically significant at the 1% level.

5.5 Falsification analysis

Our baseline test shows that the implementation of EDGAR increases firms' productivity when the firms file their financials electronically in the post-EDGAR period. When we code the indicator variable *Post-EDGAR*, within a given firm, later years (i.e., years after 1996) are always associated with *Post-EDGAR* = 1 and earlier years (i.e., years before 1993) are always associated with *Post-EDGAR* = 0. As such, a potential concern is that the observed relation between *Post-EDGAR* and *TFP* may simply be driven by technological progress occurring between 1990 and 2000, such as the technologies related to the internet boom.[25]

To rule out this alternative explanation, we perform a falsification analysis using randomly generated pseudo adoption dates. To ensure that the pseudo-date falls in the actual EDGAR phase-in period, we select the two boundary dates from the SEC's actual policy. Specifically, we randomly generate a pseudo adoption date between April 26, 1993 (the scheduled adoption date for the first group), and May 1, 1996 (the scheduled adoption date for the last group). We then use the same procedures outlined in Section 4.1 to define the indicator variable, *Pseudo Post-EDGAR*.

[25] This mechanical relation is unlikely to be the explanation for regressions with year fixed effects, while *Post-EDGAR* is defined using staggered adoption dates.

Pseudo Post-EDGAR takes a value of one if a full firm-year is subject to mandatory filing on EDGAR using this pseudo date, and zero otherwise. We then re-estimate Eq. (4) by replacing *Post-EDGAR* with *Pseudo Post-EDGAR* and perform the estimation 100, 500, and 1,000 times. We tabulate the results in Panel E of Table 6. Coefficient estimates are the mean of estimated coefficients out of 100, 500, and 1,000 times. In all columns, the coefficients on *Pseudo Post-EDGAR* are statistically indistinguishable from zero at conventional levels. We also plot the distribution of the coefficient estimates in Fig. 1. For comparison, we also include a line for the actual coefficient on *Post-EDGAR* based on the actual scheduled EDGAR adoption events. It is evident that the pseudo EDGAR adoption, based on randomly assigned dates, does not trigger any systematic increase in TFP, whereas the actual EDGAR adoption indeed has an economically meaningful and statistically significant impact on TFP. As such, we conclude that the mechanical explanation due to technological progress is unlikely to be the source of improvement in TFP.

6 Additional analyses

6.1 Investment efficiency

One corollary of our prior analysis is that the implementation of EDGAR enhances investment efficiency.[26] We next directly test this idea. Following Choi et al. (2020), we first estimate an expected level of investment using the following specification:

$$\begin{aligned} Investments_{i,t} =& \alpha + \beta_1 Investments_{i,t-1} + \beta_2 CFO_{i,t} \\ &+ \beta_3 Tobin' sQ_{i,t-1} + \beta_4 Asset\ Growth_{i,t-1} + \varepsilon \end{aligned} \tag{5}$$

where $Investment_{i,\,t}$ is (firm *i*'s) capital expenditure scaled by lagged net PP&E; $Tobin's\ Q_{\ i,\,t-1}$ is the sum of total assets plus market value of equity minus book value of equity divided by total assets measured in year $t-1$; $CFO_{i,\,t}$ is cash flow from operations scaled by lagged total assets; and $Asset\ Growth_{i,\,t-1}$ is the growth rate of assets from year $t-2$ to $t-1$. Lagged investments capture the patterns in which large projects often require capital investments across consecutive years in the time series. *Tobin's Q* captures the feedback effect in which managers also learn from stock prices (Chen et al. 2007). Finally, *CFO* at time *t* captures the investment–cash flow sensitivity (one empirical regularity of financial constraints) documented in Fazzari et al. (1988).[27]

[26] Shroff et al. (2014) argue that TFP captures an economic construct similar to investment efficiency.

[27] In concurrent work, Goldstein et al. (2021) also explore the setting of EDGAR adoption and investigate whether managerial learning is affected by its implementation. Using investment-to-price sensitivity to proxy for "revelatory price efficiency" (i.e., the extent to which prices reveal new information to managers), they show that the EDGAR implementation leads to a decrease in investment-to-price sensitivity. Their work is incrementally different from ours in that we focus on *investment efficiency* (i.e., how corporate investment may deviate from its optimal level) rather than on the level of investment per se or how sensitive the level of investment is, related to firm valuation.

We estimate Eq. (5) by industry-year (defined as the two-digit SIC code). Following Choi et al. (2020), we employ the absolute value of the residuals from this regression as a measure of investment inefficiency, with higher absolute values indicating larger deviations from the model-implied optimal level of investments. The signed residuals represent (directional) inefficiency. Negative residuals represent insufficient investments, whereas positive residuals represent overinvestments. We designate firm-years with positive (negative) residuals into the overinvestment (underinvestment) subsample.[28]

To investigate the effect of the implementation of EDGAR on investment efficiency, we estimate:

$$\begin{aligned} Inefficiency_{i,t} =& \alpha + \beta_1 Post-EDGAR_{i,t} + Firm\ characteristics \\ &+ Firm\ Fixed\ effects + Year\ Fixed\ effects \end{aligned} \tag{6}$$

Inefficiency $_{i,\,t}$ is the absolute value of the residuals from Eq. (5). As before, *Post-EDGAR* takes a value of one if a full firm-year is subject to mandatory filing on EDGAR and zero otherwise. As in Choi et al. (2020), we include *MTB* (defined as the market-to-book ratio), *Age* (defined as the number of years that a firm has appeared in the Compustat database, in natural logarithm), *OpCycle* (defined as the sum of days in accounts receivable and days in inventory), *Tangibility* (defined as net PP&E scaled by total assets), *Bankruptcy* (defined as the Altman Z-score), *CFO* (defined as cash flow from operations scaled by sales), *Loss* (the loss indicator to denote whether the current year has negative earnings), *Leverage* (book leverage, defined as the sum of long-term debt plus the current portion of long-term debt, scaled by total assets), *Slack* (financial slack, defined as cash and cash equivalent, scaled by net PP&E), *Dividend* (an indicator variable that captures whether a firm pays dividend or not), Std(*Sales*) (sales volatility, defined as the standard deviation of sales scaled by total assets over a five-year rolling window), Std(*CFO*) (cash flow volatility, defined as the standard deviation of operating cash flow scaled by total assets over a five-year rolling window), and Std(*Investment*) (investment volatility, defined as the standard deviation of capital expenditures scaled by total assets over a five-year rolling window). We also add stock price informativeness (*PSI*), financial reporting quality (*FRQ*), the change in financial reporting quality (Δ*FRQ*), and institutional ownership (*%Institutional*). We estimate Eq. (6) for the full sample, the overinvestment subsample, and the underinvestment subsample. As before, we continue to include firm fixed effects so that the coefficient on *Post-EDGAR* captures within-firm trends in investment (in)efficiency that are caused by the scheduled EDGAR adoption dates.

We present the results in Table 7. In column (1), the dependent variable is *Inefficiency*. The coefficient on *Post-EDGAR* is −0.022 and statistically significant at the 1% level, consistent with the idea that the implementation of

[28] The mean investment inefficiency in the full sample is 0.13, whereas the inefficiency measures are 0.15 and 0.11 for the over-(under-) investment subsamples, respectively.

EDGAR reduces investment inefficiency. In columns (2) and (3), we continue to use *Inefficiency* as the dependent variable but perform the tests on the overinvestment and underinvestment subsamples, respectively. Because *Inefficiency* takes the absolute value of the residual term in Eq. (6), we expect the coefficient on *Post-EDGAR* to be negative in both columns (2) and (3). In column (2), the coefficient on *Post-EDGAR* is −0.027 and statistically significantly at the 5% level, consistent with the idea that the implementation of EDGAR constrains overinvestment. In column (3), the coefficient on *Post-EDGAR* is −0.019 and statistically significantly at the 1% level, consistent with the idea that the implementation of EDGAR mitigates underinvestment.

6.2 A cross-sectional test

We perform our cross-sectional tests by revisiting the tension in our hypothesis development. Morris and Shin (2002) show that, in the absence of private information, greater provision of public information always increases social welfare (i.e., the sum of all traders' utilities). They also show that this is not the case when investors also receive private information. Specifically, when public information serves as a coordinating device for investors' beliefs, more precise public information does not necessarily lead to an increase in social welfare, especially when private information is of very high quality. As such, it is likely that the provision of high-quality public information in EDGAR may crowd out private information discovery, production, and dissemination.

Sell-side stock analysts perform at least two roles for capital market participants. First, they discover, aggregate, interpret, and compile publicly disclosed financial information into an easily understandable format for investors. Second, they produce additional information through their own channels, for example, via private communication with managers or site visits (Huang et al. 2018; Cheng et al. 2016). The EDGAR platform, which makes financial reports publicly available at no or little cost, may actually discourage investors from discovering private information. As such, we expect that the effect of EDGAR implementation on productivity is more pronounced for firms with less analyst coverage, where alternative information is scarce.

One key research design issue is that analyst coverage is endogenously determined by many factors (e.g., Bhushan 1989). To alleviate the concern of endogeneity, we follow Das et al. (2006) and Yu (2008) and estimate the following regression first:

$$\begin{aligned} Coverage = &\beta_0 + \beta_1 \ln MVE + \beta_2 NYSE + \beta_3 NASDAQ \\ &+ \beta_4 ROA + \beta_5 ExFin + \beta_6 Asset\ Growth \\ &+ \beta_7 Cash\ flow\ volatility + Fixed\ effects + \varepsilon \end{aligned}$$

We obtain analyst data from I/B/E/S. To avoid stale forecasts, we keep the forecasts issued within one year prior to the fiscal year-end. *Coverage* is the

Table 7 Investment efficiency

	(1)	(2)	(3)
	Investment inefficiency		
	Full	Overinvestment	Underinvestment
EDGAR indicators			
Post-EDGAR	−0.022***	−0.027**	−0.019***
	(−3.32)	(−2.14)	(−2.58)
Firm characteristics			
Size	0.017***	0.037***	−0.000
	(3.89)	(4.16)	(−0.06)
MTB	0.004***	0.007***	0.001
	(5.24)	(4.36)	(1.44)
Firm age	−0.020***	−0.026**	−0.010*
	(−3.27)	(−2.20)	(−1.77)
Tangibility	0.113***	0.277***	−0.007
	(5.00)	(5.81)	(−0.33)
Leverage	0.004	−0.020	0.023*
	(0.29)	(−0.59)	(1.76)
Slack	−0.003	−0.003	−0.002
	(−1.42)	(−0.76)	(−0.83)
Bankruptcy	0.006***	0.010***	0.003*
	(3.82)	(3.19)	(1.90)
CFO	−0.008	−0.072*	0.031
	(−0.36)	(−1.81)	(1.43)
Loss	−0.009***	−0.020**	0.003
	(−2.75)	(−2.47)	(0.91)
OpCycle	0.007	0.024	−0.011*
	(1.05)	(1.63)	(−1.75)
Dividend	0.007	−0.004	0.008
	(1.32)	(−0.36)	(1.50)
Std(Cash flow)	0.126**	0.192*	−0.021
	(2.43)	(1.87)	(−0.49)
Std(Sale)	0.025***	0.032*	0.014**
	(2.90)	(1.90)	(1.97)
Std(Investment)	−0.088***	−0.188***	0.029***
	(−8.14)	(−8.53)	(2.87)
PSI	0.003	0.012***	−0.004**
	(1.44)	(2.83)	(−2.04)
FRQ	−0.146**	−0.377***	−0.092
	(−2.26)	(−2.94)	(−1.52)
ΔFRQ	0.030**	0.031	−0.001
	(2.46)	(1.20)	(−0.07)
%Institutional	−0.074	−0.132	−0.047
	(−0.87)	(−0.73)	(−0.58)

Table 7 (continued)

	(1)	(2)	(3)
	Investment inefficiency		
	Full	Overinvestment	Underinvestment
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
N	20,563	8,136	12,427
adj. R-sq	0.352	0.516	0.426

This table presents the result of estimating investment efficiency on the EDGAR-implementation indicator. The dependent variable is *Inefficiency*, defined as the absolute value of the residual from Eq. (5). A higher value of this measure indicates less efficient investment. The key variable of interest is *Post-EDGAR*. *Size* is natural logarithm of total asset. *MTB* is defined as the market-to-book ratio, calculated as market value of equity scaled by book value of equity. *Firm age* is the number of years that a firm has appeared in the Compustat database. *Tangibility* is defined as net PP&E scaled by total assets. *Leverage* is defined as the sum of long-term debt plus the current portion of long-term debt, scaled by total assets. *Slack* is financial slack, defined as cash and cash equivalent scaled by net PP&E. *Bankruptcy* is the Altman Z-score. For manufacturing firms, Altman's Z-score = [4.34 + 0.08 × working capital/total assets-0.04 × retained earnings/total assets + 0.1 × EBIT/total assets + 0.22 × market value of equity/book value of total liabilities-0.06 × sales/total assets] (Hillegeist et al. 2004); for non-manufacturing firms, Altman's Z-score = [6.56 × working capital/total assets + 3.26 × retained earnings/total assets + 6.72 × EBIT/total assets + 1.05 × book value of equity/book value of total liabilities] (Altman 2013). CFO_{sale} is defined as cash flow from operations scaled by sales. *Loss* is an indicator variable to capture whether the current year has negative earnings. *OpCycle* is operation cycle, defined as the sum of days in accounts receivable and days in inventory. *Dividend* is an indicator variable that captures whether a firm pays dividend. *Std*(*Cash flow*) is cash flow volatility, defined as the standard deviation of operating cash flow scaled by total assets over a five-year rolling window. *Std*(*Sale*) is sales volatility, defined as the standard deviation of sales scaled by total assets over a five-year rolling window. *Std*(*Investment*) is investment volatility, defined as the standard deviation of capital expenditure scaled by total assets over a five-year rolling window. In column (1), we present the result of estimating Eq. (6) in the full sample. In columns (2) and (3), we present the result of estimating Eq. (6) using the overinvestment sample subsample and underinvestment subsample, respectively. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

number of analysts who made EPS forecasts in any given firm-year. ln*MVE* is the natural logarithm of market value of equity at the end of the fiscal year. *NYSE* (*NASDAQ*) is an indicator variable that takes a value of one if the stock is listed on NYSE (NASDAQ). Following Yu (2008), *ROA* and *ExFin* capture past performance and the amount of external financing, respectively. *Asset growth* is measured using the growth rate of total assets from year $t-1$ to year t. *Cash flow volatility* is the volatility of operating cash flow. We include industry fixed effects and year fixed effects. We extract the residual term from the above regression as *residual coverage* and use it as the partitioning variable.

We designate firms into the low (high) coverage subsample if their residual coverage is lower (higher) than the sample median. We present the results in Table 8. In column (1), the coefficient on *Post-EDGAR* is 0.013 and statistically insignificant at the conventional levels for the high coverage subsample. In contrast, for the low coverage subsample in column (2), the coefficient on *Post-EDGAR* (0.041) is statistically significant at the 1% level. The difference is statistically significant at the 10%

Table 8 A cross-sectional test

	(1)	(2)
	TFP	
	Analyst coverage	
	High	Low
EDGAR indicator		
Post-EDGAR	0.013	0.041***
	(0.94)	(2.84)
Firm characteristics	Yes	Yes
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
p-value $[\beta_{\text{Post-EDGAR}}{}^{(2)} > \beta_{\text{Post-EDGAR}}{}^{(1)}]$	0.08	
N	10,671	10,671
adj. R-sq	0.838	0.807

This table presents the result of a cross-sectional test partitioned by analyst coverage. We first estimate the first-stage analyst coverage deterministic regression described in Sect. 6.2. We then extract the residual term from the regression as *residual coverage* and use this residual coverage as the partitioning variable. We designate firms into the low (high) coverage subsample if their residual coverage is lower (higher) than the sample median. We include firm characteristic control variables with firm fixed effects and year fixed effects. Variable definitions are presented in Table 2. T-statistics are in parentheses, and standard errors are clustered by firm. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

level (based on a one-tail test), consistent with the idea that the implementation of the EDGAR system plays a more important role when alternative information from analysts is scarce.[29]

[29] Another possible cross-sectional test is to investigate the potential moderating role of institutional ownership. This idea is motivated by the possibility that some institutional investors may have access to subscribed third-party services to observe firms' fundamental information prior to the adoption of the EDGAR system. However, we may not observe a *dampened* effect of digital reporting in EDGAR on firms' productivity for firms with high institutional ownership for at least two reasons. First, prior to the adoption of the EDGAR system, machine-readable format data provided by commercial vendors may not deliver exactly the same quality of data as the full-scale digital financial reports do. Even if institutional investors have access to such third-party data, there is still a lot of information in the 10-K filings that would have not been captured in the third-party data. For example, the most popular corporate financial database is Standard and Poor's Compustat, and most of the data fields in Compustat are tabulated numbers that are *recognized* in financial statements (Schipper 2007). Instead, numbers (together with their detailed contexts) that are disclosed only in the accompanying footnotes as well as other places in the 10-K (e.g., management discussion and analysis – MD&A) are often uncaptured by such a product. As such, institutional investors still may not be able to uncover a firm's full picture of corporate operations and financial positions based on their access to a third-party database. Second, the intensity of monitoring provided by institutional investors versus retail investors is different. There is a vast literature documenting that the institutional investors perform a more active role in monitoring a firm's management. For example, Hartzell and Starks (2003) show that institutional ownership concentration is positively related to the pay-for-performance sensitivity of executive compensation and negatively related to the level of compensation. As such, compared with firms with concentrated institutional ownership, firms with lower institutional ownership may not experience a higher increase in TFP as they receive less intense monitoring from their investor base (primarily made up of retail investors).

7 Conclusion

Motivated by the emerging application of digital technologies in financial reporting, we examine the effect of digital corporate financial reporting on firms' productivity. We argue that information frictions represent a constraint that impedes efficient resource allocation and that a major source of such frictions stems from the fact that firms' production functions (the conversion from inputs to outputs) are not observable to corporate outsiders prior to digital financial reporting. The implementation of EDGAR mitigates this resource allocation inefficiency by providing timely firm-specific information and leads to an improvement in productivity. Employing the staggered implementation of the SEC's EDGAR that took place between 1993 and 1996, we document a statistically significant and economically meaningful increase in total factor productivity caused by this shift to digital financial reporting. By focusing on the role of information dissemination in coordinating investments and production, our findings provide evidence on the economic consequences accruing to EDGAR-adopting firms.

Appendix: Econometrics on the estimation of TFP

This appendix provides the technical details for the estimation of TFP.

- Econometrics on TFP estimation

To calculate TFP, we start with the log-linearized Cobb–Douglas production function:

$$y = \beta_0 + \beta_K * k + \beta_L * l + u \tag{A1}$$

As discussed in the main text, u can be decomposed into two terms ω and ε:

$$y = \beta_0 + \beta_K * k + \beta_L * l + \overbrace{\underbrace{\omega}_{\substack{\textit{observable to} \\ \textit{the manager}}} + \underbrace{\varepsilon}_{\substack{\textit{unobservable to} \\ \textit{the manager}}}}^{u} \tag{A2}$$

When making investment decisions, the manager can choose k and l upon observing ω or a noise measure of ω. In other words, under this situation, the independent variables (k and l) are correlated with the error term (u), making the estimation of Eq. (3) from an ordinary least squares (OLS) regression biased.

To alleviate the concern, we follow Ackerberg et al. (2015) to estimate the production function. There are several assumptions in this model:

– Assumption 1: (Information set) The firm's information set at time t is I_t. It includes both current and past production shocks but does not include any future production shock.

- Assumption 2: (First-order Markov) Productivity shocks evolve according to the distribution $p(\omega_{t+1}|I_t) = p(\omega_{t+1}|\omega_t)$.
- Assumption 3: Intermediate input (m) is chosen at the same time or after l_t is chosen.
- Assumption 4: m_t is strictly increasing in ω_t.

Based on Assumption 3, m_t is written as:

$$m_t = f(k_t, l_t, \omega_t) \tag{A3}$$

To the extent that m_t is strictly increasing in ω_t, we invert Equation (A3) and represent ω_t as a function of k_t, l_t and m_t:

$$\omega_t = f^{-1}(k_t, l_t, m_t) \tag{A4}$$

Thus, we rewrite equation (A2) (with time subscripts) by substituting ω_t using Equation (A4):

$$y_t = \beta_0 + \beta_K * k_t + \beta_L * l_t + f^{-1}(k_t, l_t, m_t) + \varepsilon_t \tag{A5}$$

Now, the output y_t is now a non-parametric function of k_t, l_t, and m_t.

$$y_t = \varphi(k_t, l_t, m_t) + \varepsilon_t \tag{A6}$$

where $\varphi(k_t, l_t, m_t) = \beta_0 + \beta_K * k_t + \beta_L * l_t + f^{-1}(k_t, l_t, m_t)$.

Apply the first moment condition here:

$$E[\varepsilon_t|I_t] = E\left[y_t - \varphi(k_t, l_t, m_t)\middle|I_t\right] = 0 \tag{A7}$$

We use a second-order polynomial function to re-express $\varphi(k_t, l_t, m_t)$:

$$\begin{aligned}\varphi(k_t, l_t, m_t) = \theta_0 + \theta_K{*}k_t \\ &+ \theta_L{*}l_t + \theta_M{*}m_t \\ &+ \theta_{K^2}{*}k_t^2 + \theta_{L^2}{*}l_t^2 \\ &+ \theta_{M^2}{*}m_t^2 + \theta_{KL}{*}k_t{*}l_t + \theta_{KM}{*}k_t{*}m_t \\ &+ \theta_{LM}{*}l_t{*}m_t\end{aligned}$$

In the first stage regression, we estimate the above equation and obtain the fitted value $\widehat{\varphi}(k_t, l_t, m_t)$.[30] Assumptions 1 and 2 imply that ω_t can be decomposed into the expected value at time $t-1$ and a residual term:

$$\omega_t = E[\omega_t|I_{t-1}] + \mu_t = E[\omega_t|\omega_{t-1}] + \mu_t = g(\omega_{t-1}) + \mu_t \tag{A8}$$

[30] We also use a third-order polynomial function to represent $\varphi(k_t, l_t, m_t)$. Results and inferences are qualitatively similar.

Therefore, Equation (A2) can be re-written (with time subscripts) into:

$$y_t = \beta_0 + \beta_K * k_t + \beta_L * l_t + g(\omega_{t-1}) + \mu_t + \varepsilon_t \quad \text{(A9)}$$

Given that $E[\mu_t | I_{t-1}] = 0$ and $E[\omega_t | I_t] = 0$, the second moment condition is:

$$E[\mu_t + \varepsilon_t | I_{t-1}] = E[y_t - \beta_0 - \beta_K * k_t - \beta_L * l_t - g(\omega_{t-1}) | I_{t-1}] = 0 \quad \text{(A10)}$$

Rewrite ω_{t-1} into:

$$\omega_{t-1} = \varphi(k_{t-1}, l_{t-1}, m_{t-1}) - \beta_0 - \beta_K * k_{t-1} - \beta_L * l_{t-1} \quad \text{(A11)}$$

By inserting Equation (A11) into Equation (A10), Equation (A10) becomes:

$$E\left[y_t - \beta_0 - \beta_K * k_t - \beta_L * l_t - g\left(\varphi(k_{t-1}, l_{t-1}, m_{t-1}) - \beta_0 - \beta_K * k_{t-1} - \beta_L * l_{t-1}\right) \middle| I_{t-1}\right] = 0 \quad \text{(A12)}$$

Now, replace $\varphi(k_{t-1}, l_{t-1}, m_{t-1})$ with the fitted value $\widehat{\varphi_{t-1}}$. Following İmrohoroğlu and Tüzel (2014), we set $\omega_t = g(\omega_{t-1}) = \rho\omega_{t-1} + \mu_t$, and Equation (A12) can be re-written as:

$$E\left[y_t - \beta_0 - \beta_K * k_t - \beta_L * l_t - \rho\left(\widehat{\varphi_{t-1}} - \beta_0 - \beta_K * k_{t-1} - \beta_L * l_{t-1}\right) \middle| I_{t-1}\right] = 0 \quad \text{(A13)}$$

Following İmrohoroğlu and Tüzel (2014), we use the non-linear least squares model to estimate β_0, β_K, β_L, and ρ.

- Data and variable definitions in TFP estimation

To calculate *TFP* for the firms in our sample, we follow Bennett et al. (2020) and obtain accounting data from Compustat. We obtain the price index for gross domestic product (*GDP*) and the price index for private fixed investment from the Bureau of Economic Analysis. Value added (*Y*) is defined as sales (*SALE*) minus materials scaled by GDP deflator. Material (*M*) is defined as the difference between total expense and labor expense deflated by the GDP price index. Total expense is revenue (*REVT*) minus operation income before depreciation and amortization (*OIBDP*). We use staff expense (*XLR*) in Compustat to calculate total labor expense. When this value is missing, we first calculate the average wage per employee within a Fama–French-12 industry using all non-missing wages in that specific industry, then impute a firm's labor cost using the number of employees in the firm times the industry-average wage per employee. Capital stock (*K*) is defined as gross property, plant, and equipment (*PPEGT*) scaled by the price deflator and adjusted for the age of the capital stock, following Hall (1990). Labor (*L*) is the number of employees.

Acknowledgements We thank Stephen Penman (Editor) and two anonymous reviewers for very helpful comments. We are indebted to Michael Welker, who provided constant encouragement, supervision, and support while we worked on this project. We also thank Paul Calluzzo, Yolande Chan, Linda Myers, Nancy L. Su, Kai Sun, Changqiu Yu, Ying Zhang, Steven Zheng, and seminar participants at Queen's University and the University of Manitoba. Zhang and Liu acknowledge supports from the Commerce '83 Fellowship at Queen's University and the Chartered Professional Accountants Research Fellowship at University of Manitoba, respectively. Liu also thanks the financial support provided by Queen's University, where the initial draft was completed during his doctoral study at the Smith School of Business.

References

Ackerberg, D., K. Caves, and G. Frazer. 2015. Identification properties of recent production function estimators. *Econometrica* 83 (6): 2411–2451.

Altman, E. 2013. Predicting financial distress of companies: revisiting the Z-score and ZETA® models. In *Handbook of research methods and applications in empirical finance*. Edward Elgar Publishing.

Asthana, S., S. Balsam, and S. Sankaraguruswamy. 2004. Differential response of small versus large investors to 10-K filings on EDGAR. *The Accounting Review* 79 (3): 571–589.

Atanasov, V., and B. Black. 2016. Shock-based causal inference in corporate finance and accounting research. *Critical Finance Review* 5: 207–304.

Banker, R., R. Huang, Y. Li, and S. Zhao. 2021. Do Accounting Standards Matter for Productivity? *Production and Operations Management* 30 (1): 68–84.

Bennett, B., R. Stulz, and Z. Wang. 2020. Does the stock market make firms more productive? *Journal of Financial Economics* 136 (2): 281–306.

Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. 2020. Using machine learning to detect misstatements. *Review of Accounting Studies* 26 (2): 468–519.

Bertrand, M., and S. Mullainathan. 2003. Enjoying the quiet life? Corporate governance and managerial preferences. *Journal of Political Economy* 111 (5): 1043–1075.

Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11 (2–3): 255–274.

Biddle, G., G. Hilary, and R. Verdi. 2009. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics* 48 (2–3): 112–131.

Bloom, N., and J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics* 122 (4): 1351–1408.

Bushman, R., and A. Smith. 2001. Financial accounting information and corporate governance. *Journal of Accounting and Economics* 32 (1–3): 237–333.

Butler, A., G. Grullon, and J. Weston. 2005. Stock market liquidity and the cost of issuing equity. *Journal of Financial and Quantitative Analysis* 40 (2): 331–348.

Chaney, T., D. Sraer, and D. Thesmar. 2012. The collateral channel: How real estate shocks affect corporate investment. *American Economic Review* 102 (6): 2381–2409.

Chang, X., S. Dasgupta, and G. Hilary. 2006. Analyst coverage and financing decisions. *Journal of Finance* 61 (6): 3009–3048.

Chang, Y., P. Hsiao, A. Ljungqvist, and K. Tseng. 2022. Testing disagreement models. *Journal of Finance* 77 (4): 2239–2285.

Chang, Y., A. Ljungqvist, and K. Tseng. 2021. Do corporate disclosures constrain strategic analyst behavior? *Working Paper*.

Chen, Q., I. Goldstein, and W. Jiang. 2007. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies* 20 (3): 619–650.

Chen, S., B. Miao, and T. Shevlin. 2015. A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research* 53 (5): 1017–1054.

Cheng, Q., F. Du, X. Wang, and Y. Wang. 2016. Seeing is believing: Analysts' corporate site visits. *Review of Accounting Studies* 21 (4): 1245–1286.

Choi, J., R. Hann, M. Subasi, and Y. Zheng. 2020. An empirical analysis of analysts' capital expenditure forecasts: Evidence from corporate investment efficiency. *Contemporary Accounting Research* 37 (4): 2615–2648.

Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *Journal of Finance* 56 (5): 1629–1666.

Das, S., R. Guo, and H. Zhang. 2006. Analysts' selective coverage and subsequent performance of newly public firms. *Journal of Finance* 61 (3): 1159–1185.

David, J., H. Hopenhayn, and V. Venkateswaran. 2016. Information, misallocation, and aggregate productivity. *Quarterly Journal of Economics* 131 (2): 943–1005.

DeAngelo, L. 1988. Managerial competition, information costs, and corporate governance: The use of accounting performance measures in proxy contests. *Journal of Accounting and Economics* 10 (1): 3–36.

Dechow, P., and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77 (s-1): 35–59.

deHaan, E. 2020. Using and interpreting fixed effects models. *Working paper*. University of Washington.

Dugast, J., and T. Foucault. 2018. Data abundance and asset price informativeness. *Journal of Financial Economics* 130 (2): 367–391.

Durnev, A., R. Morck, and B. Yeung. 2004. Value-enhancing capital budgeting and firm-specific stock return variation. *Journal of Finance* 59 (1): 65–105.

Easley, D., and M. O'hara. 1992. Time and the process of security price adjustment. *Journal of Finance* 47 (2): 577–605.

Easley, D., S. Hvidkjaer, and M. O'Hara. 2010. Factoring information into returns. *Journal of Financial and Quantitative Analysis* 45 (2): 293–309.

Fazzari, S., R. Hubbard, and B. Petersen. 1988. Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity* 1: 141–195.

Francis, J., R. LaFond, P. Olsson, and K. Schipper. 2004. Costs of equity and earnings attributes. *The Accounting Review* 79 (4): 967–1010.

Gao, M., and J. Huang. 2020. Informing the market: The effect of modern information technologies on information production. *Review of Financial Studies* 33 (4): 1367–1411.

Glaeser, S. 2018. The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics* 66 (1): 163–193.

Goldstein, I., S. Yang, and L. Zuo. 2021. The real effects of modern information technologies. Cornell University Working Paper.

Gomez, E. 2020, The Effect of mandatory disclosure dissemination on information asymmetry: evidence from the implementation of the EDGAR System. *Working Paper*.

Guo, F., L. Lisic, M. Stuart, and C. Wang. 2019. The impact of information technology on stock price crash risk: evidence from the EDGAR implementation. *Working Paper*.

Hall, R., and C. Jones. 1999. Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* 114 (1): 83–116.

Hall, B. 1990. The Manufacturing Sector Master File: 1959–1987. National Bureau of Economic Research. *Working paper*.

Hann, R., H. Kim, W. Wang, and Y. Zheng. 2020. Information frictions and productivity dispersion: The role of accounting information. *The Accounting Review* 95 (3): 223–250.

Harford, J., and K. Li. 2007. Decoupling CEO wealth and firm performance: The case of acquiring CEOs. *Journal of Finance* 62 (2): 917–949.

Hart, O., and J. Moore. 1995. Debt and seniority: An analysis of the role of hard claims in constraining management. *American Economic Review* 85 (3): 567.

Hartzell, J., and L. Starks. 2003. Institutional investors and executive compensation. *Journal of Finance* 58 (6): 2351–2374.

Hillegeist, S., E. Keating, D. Cram, and K. Lundstedt. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9 (1): 5–34.

Hsieh, C., and P. Klenow. 2009. Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics* 124 (4): 1403–1448.

Huang, A., R. Lehavy, A. Zang, and R. Zheng. 2018. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science* 64 (6): 2833–2855.

İmrohoroğlu, A., and Ş Tüzel. 2014. Firm-level productivity, risk, and return. *Management Science* 60 (8): 2073–2090.

Jensen, M. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review* 76 (2): 323–329.

Kambil, A., and M. Ginsburg. 1998. Public access Web information systems: Lessons from the Internet EDGAR project. *Communications of the ACM* 41 (7): 91–97.

Kim, J., Y. Li, and L. Zhang. 2011. Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics* 100 (3): 639–662.

Koh, P., and D. Reeb. 2015. Missing R&D. *Journal of Accounting and Economics* 60 (1): 73–94.

Lai, S., C. Lin, and X. Ma. 2020. Regtech Adoption and the Cost of Capital, *Working Paper*.

Lee, C., P. Ma, and C. Wang. 2015. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics* 116 (2): 410–431.

Lehn, K., and M. Zhao. 2006. CEO turnover after acquisitions: Are bad bidders fired? *Journal of Finance* 61 (4): 1759–1811.
Leuz, C., and R. Verrecchia. 2004. Firms' capital allocation choices, information quality, and the cost of capital. *Working paper*. University of Pennsylvania.
Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21 (1): 107–138.
Lin, C., C. Ma, Y. Sun, and Y. Xu. 2021. The telegraph and modern banking development, 1881–1936. *Journal of Financial Economics* 141 (2): 730–749.
Loughran, T., and B. McDonald. 2017. The use of EDGAR filings by investors. *Journal of Behavioral Finance* 18 (2): 231–248.
McNichols, M. 2002. Discussion of the quality of accruals and earnings: the role of accrual estimation errors. *The Accounting Review* 77 (s-1): 61–69.
Morris, S., and H. Shin. 2002. Social value of public information. *American Economic Review* 92 (5): 1521–1534.
Myers, S. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5 (2): 147–175.
Myers, S., and N. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13 (2): 187–221.
Olley, G., and A. Pakes. 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64 (6): 1263–1297.
Palepu, K. 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics* 8 (1): 3–35.
Restuccia, D., and R. Rogerson. 2017. The causes and costs of misallocation. *Journal of Economic Perspectives* 31 (3): 151–174.
Roychowdhury, S., N. Shroff, and R. Verdi. 2019. The effects of financial reporting and disclosure on corporate investment: A review. *Journal of Accounting and Economics* 68 (2–3): 101246.
Schipper, K. 2007. Required disclosures in financial reports. *The Accounting Review* 82 (2): 301–326.
Shipman, J., Q. Swanquist, and R. Whited. 2017. Propensity score matching in accounting research. *The Accounting Review* 92 (1): 213–244.
Shleifer, A., and R. Vishny. 1986. Large shareholders and corporate control. *Journal of Political Economy* 94 (3–1): 461–488.
Shroff, N., R. Verdi, and G. Yu. 2014. Information environment and the investment decisions of multinational corporations. *The Accounting Review* 89 (2): 759–790.
Smith, C., and J. Warner. 1979. On financial contracting: An analysis of bond covenants. *Journal of Financial Economics* 7 (2): 117–161.
Solow, R. 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70 (1): 65–94.
Solow, R. 1957. Technical change and the aggregate production function. *Review of Economics and Statistics,* 312–320.
To, T., M. Navone, and E. Wu. 2018. Analyst coverage and the quality of corporate investment decisions. *Journal of Corporate Finance* 51: 164–181.
Welker, M. 1995. Disclosure policy, information asymmetry, and liquidity in equity markets. *Contemporary Accounting Research* 11 (2): 801–827.
Yoshikawa, H. 1980. On the "q" theory of investment. *American Economic Review* 70 (4): 739–743.
Yu, F. 2008. Analyst coverage and earnings management. *Journal of Financial Economics* 88 (2): 245–271.