Causal Effect of Information Costs on Asset Pricing Anomalies☆

Yong Hyuck Kim
California State University, Long Beach

Zoran Ivković
Eli Broad College of Business
Michigan State University

Dmitriy Muravyev
Eli Broad College of Business
Michigan State University and
CDI Research Fellow

September 6, 2022

Abstract

Active investors strive to beat the market by obtaining an information edge, a costly enterprise that reduces their net profits (Grossman and Stiglitz (1980)). We are the first to both causally identify and assess how these information costs affect stock anomalies. The SEC's EDGAR slashed the costs of acquiring and trading on accounting information. Using the staggered EDGAR introduction, we show that average alphas for 125 accounting anomalies decline substantially; and the decline entirely explains pre-EDGAR alphas. By contrast, alphas for 80 non-accounting anomalies do not change significantly. Information costs are as substantial and important as other limits to arbitrage.

Keywords: Information costs, stock anomalies, EDGAR, limits to arbitrage

☆We thank Charles Hadlock, Joshua Pollet, Andrei Simonov, Scott Weisbenner, as well as the participants at the 2022 China International Conference in Finance (especially the discussant Si Cheng) for comments and suggestions.

1. Introduction

Traditional asset pricing theories such as the CAPM (Sharpe (1964), Lintner (1965), Mossin (1966)) and the APT (Ross (1976)) assume frictionless markets, including costless trading and information gathering. Theoretical contributions by Grossman and Stiglitz (1980) and Verrecchia (1982) are quick to point out that costly information acquisition, an inevitable reality of financial markets, affects investor decisions and market outcomes. Investors purchase data, hire analysts to clean and process it, and provide them with the necessary hardware and software tools to do so; each of these steps is costly.[1] Nonetheless, most prior asset pricing studies ignore information costs and instead focus on trading costs (e.g., Novy-Marx and Velikov (2016)) or short sale costs (e.g., Chu, Hirshleifer, and Ma (2020)).[2] Consequently, there continues to be a lack of empirical studies that undertake the challenging task of quantifying the costs of acquiring information.

This paper fills that gap in the literature by estimating information costs in the U.S. equity markets. Information costs are understood to be all the costs of gathering the information (be it through subscriptions or through diligent searches of electronic or other information sources), hiring and compensating the team of analysts processing and interpreting the information, acquiring the computing equipment for them to do so, arranging for physical space in which to perform such activities, and communicating their findings to fund management. More specifically, we use a quasi-natural experiment—the SEC's staggered implementation of the Electronic Data Gathering, Analysis and Retrieval (EDGAR) system from February 1993 to May 1996—and study

[1] Indeed, although recent technological advances lowered information gathering and disseminating costs, a typical hedge fund spent over $1 million on data subscriptions alone in 2019 (Whyte (2020)).
[2] Other notable studies of the effect of regular costs on stock anomalies include Keim and Madhavan (1997), Korajczyk and Sadka (2004), Lesmond, Schill, and Zhou (2004), Frazzini, Israel, and Moskowitz (2018), Patton and Weller (2020) for trading costs, and Geczy, Musto, and Reed (2002) and Drechsler and Drechsler (2014) for short sale costs.

how stock anomalies responded to the EDGAR shock that drastically lowered the cost of gathering the accounting information.

There are several reasons why the introduction of EDGAR is well-suited for our inquiry. First, EDGAR is an online system that enables companies to report their corporate filings electronically and investors to download them freely from anywhere in the world. Its adoption drastically lowered investors' information acquisition costs. Before EDGAR, a comprehensive analysis of a broad cross-section of stocks was cost-prohibitive or impractical for most investors (Chang, Ljungqvist, and Tseng (2021)). An investor could physically visit one of the SEC's reference rooms in Washington DC, New York, or Chicago and go through paper financial statements; pursue costly subscriptions to commercial data vendors such as Compustat, Value Line, or Dialog, which were often delayed and contradicted one another (Kern and Morris (1994)), or request companies to mail the filing documents. Second, most known stock anomalies rely on accounting information. EDGAR made this information convenient and inexpensive to access. Third, the SEC's adoption design allows us to harness a staggered difference-in-difference framework. Finally, Grossman and Stiglitz (1980) establish that, in a competitive equilibrium, the drop in trading profitability matches the decrease in information costs. Accordingly, studying anomaly profitability enables us to estimate information costs for the marginal investor.

We estimate the effect of EDGAR introduction on anomaly alphas in a staggered difference-in-difference framework. The SEC adopted EDGAR following a phase-in schedule over three years, randomly assigning each public firm to one of ten implementation phases. This adoption design helps us identify a causal effect of the decline in information costs on anomaly profitability, with identification coming from firms entering EDGAR at different times and from anomalies requiring—or not requiring—accounting information. We analyze a comprehensive set of anomalies documented in Chen and Zimmermann (2020). Our baseline results are grounded in a panel of long-short portfolio monthly returns for ten implementation phases and 205 asset-pricing anomalies, 125 of which require accounting information and 80 do not.

EDGAR enables convenient access to firms' accounting information; accordingly, we find that the Fama-French six-factor (the Fama and French (2015) five factors and the Carhart (1997) momentum factor) alphas for the accounting-based anomaly portfolios decline on average by 47 to 62 basis points per month (or 5.7% to 7.4% per year) in response to the EDGAR introduction. The average Fama-French alpha of the accounting anomaly portfolios in the treatment group before the EDGAR implementation is 44.9 basis points per month. Therefore, the 47 to 62 basis points drop accounts for virtually all of the pre-EDGAR alphas. By contrast, EDGAR has not lowered the costs of gathering other, non-accounting information. Indeed, we find that the alphas for non-accounting anomalies have not been significantly affected by the EDGAR introduction. We estimate information acquisition costs as the 41.9 to 49.4 basis points per month difference between the effect of EDGAR implementation on accounting anomalies and non-accounting anomalies. The results are robust to using alternative specifications and factor models.

The EDGAR-prompted decline in profitability of accounting anomalies should be more pronounced among the stocks for which the information was more difficult to gather in the pre-EDGAR period. We use two empirical proxies for information availability—analyst coverage and market capitalization (e.g., Kelly and Ljungqvist (2012))—to show that, indeed, the accounting anomalies' profitability decline is primarily driven by stocks with low information availability (stocks followed by fewer analysts or stocks with below-median market capitalization)—approximately 70 to 71 basis points per month, or about 8.5% per year. By contrast, the EDGAR-prompted profitability decline for accounting anomalies for high information availability stocks is statistically indistinguishable from zero. We also consider the complexity of trading strategies required to exploit the accounting anomalies. We differentiate among low-complexity, medium-complexity, and high-complexity accounting anomalies (requiring increasingly larger numbers of contemporaneous accounting variable to execute the strategy). The magnitude of the decline of the post-EDGAR return gap grows monotonically with the increase in anomaly complexity; the post-EDGAR return gaps for high- and low-complexity anomalies are statistically significant and economically large, in the range between 31 and 51 basis points per month.

Splitting each anomaly portfolio into short and long legs reveals that profit attenuation effects are concentrated among the short legs of the accounting anomaly portfolios. This finding is consistent with the fact that many anomalies are concentrated in the short legs (Stambaugh, Yu, and Yuan (2012)). Thus, activity by short-sellers, proxied by short interest, should be more informative than activity by long-only investors, proxied by institutional ownership. Indeed, we find that, as stocks in the short leg of accounting anomalies became more efficiently priced, their short interest declined as these stocks became less attractive to short sellers, while institutional ownership did not change significantly. We also find that trading activity increased less for stocks in top/bottom portfolios of accounting anomalies than for stocks in top/bottom portfolios of non-accounting anomalies, as the former became less attractive to arbitrageurs.

If the EDGAR introduction made stocks in the top and bottom portfolios of accounting anomalies less mispriced, these stocks should feature less information asymmetry and should, therefore, become less volatile and more liquid. We confirm these hypotheses in the data. The EDGAR introduction is associated with a decline in stock returns' idiosyncratic volatility and, therefore, an increase in price efficiency. This decline in volatility is present for both accounting anomalies and non-accounting anomalies, but it is stronger for accounting anomalies. Similarly, following a general trend in increased liquidity, the stocks' Amihud illiquidity measures declined for both types of anomalies, but liquidity increased more for non-accounting anomalies.

We conclude the list of our empirical tests with a range of robustness checks. We confirm that our results are robust to controlling for differences in stocks characteristics, most importantly firm size, across the implementation phases. We check that the results are not driven by one or few phases. Moreover, pre-trends and falsification tests validate our difference-and-difference analysis. We also confirm that the post-publication effect (McLean and Pontiff (2016)) is not driving our results.

Prior literature documents that limits to arbitrage such as noise trader risk, trading costs, and short sale costs partially explain anomaly returns, but no study of which we are aware examines the effect of information acquisition costs *per se* on anomaly returns.[3] However, investors incur the costs of acquiring information even before they pay trading or short sale costs—they need first to identify which stocks to buy or sell before they trade. To our knowledge, we are the first to quantify information acquisition costs and show that the costs can be as high as 42 to 49 basis points per month, explaining virtually all of the accounting anomalies' pre-EDGAR alphas.[4] Overall, our results suggest that information costs are large.

Our results are highly relevant for today's markets. Indeed, while utilizing accounting information in active portfolio management was innovative in the mid-1990s, it became quickly

[3] For instance, Geczy, Musto, and Reed (2002) argue that stock borrowing costs explain little of anomaly alphas. By contrast, Chu, Hirshleifer, and Ma (2020) rely on the Reg-SHO pilot program to document that relaxed short sale constraints reduce abnormal returns on 11 anomaly portfolios by 72 basis points per month. Novy-Marx and Velikov (2016) use bid-ask spreads from TAQ to show that the average trading costs range from 20 to 57 basis points per month for the mid-turnover anomalies. By contrast, Frazzini, Israel, and Moskowitz (2018) argue that institutional trading costs are much smaller than those implied by the TAQ data.

[4] Lower costs attracted new arbitrageurs and thus increased the capital pool involved in correcting mispricing caused by accounting anomalies. As a result, the annual alphas associated with trading on such anomalies that require EDGAR information decreased by 6-7%. In the Grossman and Stiglitz (1980) model, this decline in profitability would be attributed solely to the lower information costs.

commoditized post-EDGAR, reducing its alpha-generating ability. In response, hedge funds and other active managers expanded into new types of data. Such data are presently expensive and hard to process, a circumstance similar to the status of accounting information pre-EDGAR. Thus, the same principles we uncovered for circumstances surrounding the EDGAR introduction likely apply to these data.

This paper contributes to four strands of the literature. First, if contributes to the literature concerning information costs and market outcomes. Merton (1987) and Shapiro (2002) point out that costly information constraints compel investors to trade only the securities regarding which they possess adequate information and show how these constraints affect the general equilibrium process and outcomes. Grossman and Stiglitz (1980) show that perfect market efficiency is elusive because information is costly to collect. Easley and O'Hara (2004) argue that private information presents a systematic risk, for which uninformed investors are compensated. Our paper contributes to this literature by estimating information costs in the context of the U.S. equity markets.

Second, we also contribute to the stock anomaly literature by identifying the causal effect of information constraints on anomaly returns. Only a few papers study how limits to arbitrage affect anomalies using exogenous shocks to address the endogeneity (e.g., Chu, Hirshleifer, and Ma (2020)). However, unlike our study, they do not explore costly information constraints as limits to arbitrage. McLean and Pontiff (2016) document that the portfolio alphas decline by 58% on average after publication. Whereas they argue that investors learn about anomalies from academic research, we point out that even investors who discover anomalies before academics do incur substantial information costs of computing the anomaly signals.

Third, we contribute to the literature that considers the effects of a change in the information environment on a range of outcomes. Dong et al. (2016) show that stock return synchronicity decreases for firms that started to file with the SEC using a machine-friendly format XBRL. Chen, Kelly, and Wu (2020) show that reduction in analyst coverage prompts hedge funds to acquire more information through other means (participate in earnings conference calls and search EDGAR). The hedge fund participation mitigates the impairment of market efficiency caused by coverage reductions. Whereas both studies focus on the change in the information environment and provide a range of results, unlike our study neither seeks to quantify the change in information costs in the aftermath of the information environment change in terms of changes in stock returns.

Finally, a strand of the literature focuses on the effect of EDGAR on financial markets. Gao and Huang (2020) were the first to apply the staggered EDGAR implementation and show that internet dissemination of information prompts corporate outsiders to produce more information. The EDGAR adoption also improves equity financing (Goldstein, Yang, and Zuo (2020)), reduces the information asymmetry between managers and investors (Gomez (2020)), reduces investor disagreement, and mitigates crash risk, especially among the stocks with binding short sale constraints and high investor optimism (Chang *et al*. (2022)). Our paper is the first to study how the EDGAR implementation affects the anomaly profitability and to estimate the costs of acquiring information investors bear in the absence of readily accessible accounting information.

2. Implementation of the EDGAR system

A. Costs of Information Acquisition before EDGAR

Prior to the EDGAR adoption in the mid-1990s, the costs of acquiring the information contained in corporate filings were prohibitively large. Investors were mostly limited to three options. The first option was to visit one of the reference rooms in Washington DC, New York, or Chicago where the SEC kept the paper financial statements. The second option was to subscribe to the commercial data vendors' services such as Compustat, Disclosure, Value Line, or Dialog. Lastly, current shareholders could request that the companies mail their filing documents to them.

Anecdotal evidence confirms that the first option was costly and unreliable. Investors had to be physically present in one of the SEC's reference rooms and make a painstaking effort to acquire information on the corporate filings. Occasionally, investors could not even access the information they needed because some of the paper files in the SEC's reference rooms were lost.[5]

The second option was also costly because the pre-EDGAR data aggregators charged high fees. A petition filed to the SEC and the U.S. House of Representatives in 1992 documents the related complaints. The petition demands free public access to corporate filings, pointing out that the Compustat CD-ROM database with historical filings for just 7,200 companies cost $18,000 (Love (1992)).[6] Depending on the coverage, annual subscription fees ranged between $5,000 and $50,000.[7] Value Line Database cost $1,700 per quarter and covered only 1,650 companies. Mead Data Central was only available for a considerable fee that consisted of a $125 monthly fixed fee, a $39 hourly connection fee, and a search fee ranging from $6 to $51 per search.[8]

[5] A *Wall Street Journal* article reports in 1991 that "…nowadays the SEC is being hit by a tidal wave of paper, receiving some 700,000 paper filings every year, amounting to about five million pieces of paper. Those documents are warehoused in the SEC's crowded public reference room, where investors, journalists and financial research organizations routinely comb through stacks of file folders in search of hot documents – and don't always find them."
[6] According to Love (1992), the CD-ROM was called "COMPUSTAT PC Plus." A less expensive product, "COMPUSTAT Corporate Text," was available for $9,000, but was limited in its coverage to only 3,200 firms.
[7] SEC: Oversight of the Edgar System (March 14, 1985), pp. 51.
[8] The petition also reveals that Dialog charged $84 per hour on top of a $1 per page search fee. Compact Disclosure was another popular commercial database at the time. Richards (1988) documents that Compact Disclosure had quarterly updated financial and management information on 10,150 public companies, and cost around $4,500 per year for commercial institutions. However, Richards (1988) notes that Compact Disclosure's access software had technical issues retrieving time-series data, and was missing information on brokerage houses, foreign companies, and microcap stocks with less than $5 million in assets.

Aside from high fees, Compustat suffered from production lag and inaccuracy, which also pushed up the costs of acquiring accurate financial information. D'Souza, Ramesh, and Shen (2010) find that Compustat had an average dissemination lag of 24.7 weekdays prior to EDGAR (that lag dropped by almost 50% once EDGAR was adopted). Moreover, even if investors had subscribed to commercial data vendor services, there existed a significant mismatch between their databases. Kern and Morris (1994) compare two popular commercial databases at the time, Value Line and Compustat, and find material disagreements between the two datasets from 1985 to 1990. More importantly, they replicate Porcano (1986) using each database to show that empirical research could have different outcomes depending on the database used. Kothari, Shanken, and Sloan (1995) explore the implications of a selection bias in the Compustat data for return predictability. Therefore, the costs of obtaining *accurate* financial information were still very high, even after paying the stiff fees that the commercial data vendors had charged.

Lastly, in principle, investors could have received the financial documents directly from the companies by mail. Besides the costs of a long wait, this was not a viable option for an investor who intended to perform cross-sectional firm characteristics analysis because such analyses require simultaneous availability of financial information concerning many public companies.

B. Introduction of EDGAR

Responding to the call for more transparency and easier accessibility of corporate filings by publicly traded companies, the SEC harnessed the advances in information technology by developing and introducing the EDGAR system. The SEC began developing the system in 1983. Eventually, after extensive testing, on February 23, 1993, the Commission issued four releases adopting the rules that required filers to file electronically. The process began on April 26, 1993, gradually bringing all filers onto the EDGAR system. EDGAR allows the public firms to disclose their financial information electronically, and investors or any other information consumers to access the filed corporate information instantaneously via the internet without charge.

The introduction of EDGAR significantly lowered the costs of information acquisition by expediting electronic filing and information dissemination via the internet. The SEC website points out that EDGAR "… benefits investors, corporations, and the U.S. economy overall by increasing the efficiency, transparency, and fairness of the securities markets... Access to EDGAR's public database is *free*—allowing you to research, for example, a public company's financial information and operations by reviewing the filings the company makes with the SEC." Furthermore, EDGAR's search function and other interface features allowed the users to retrieve specific information in electronic documents that may not be available in commercial databases.

A feature of EDGAR implementation, of paramount importance to our empirical design, is that the SEC adopted EDGAR following a phase-in schedule. The schedule assigned each public firm that required filing to one of ten phases (from Group CF-01 to Group CF-10). Each phase had a designated date as of which electronic filling was mandated (SEC Release No. 33-6977). The firms in the first group were mandated to start uploading filings through EDGAR on April 26, 1993, and those in the last group on May 1, 1996. Table I shows the implementation schedule.

We estimate the extent to which the investor information costs decreased. The staggered nature of EDGAR implementation helps us better identify the effect of information costs, alleviate alternative explanations, and control for other confounding factors. For example, one alternative explanation could be that the equity market is becoming increasingly efficient and non-information costs decrease over time. However, to explain our results, these trends would have to discontinuously change for each firm at exactly the time at which it starts filling with EDGAR, a highly implausible set of circumstances. We also check that the ten implementation phases were similar pre-EDGAR in terms of anomaly alphas.

3. Data and Methodology

A. The SEC EDGAR Implementation Data

To construct anomaly portfolios for firms in each implementation phase, we first identify the date each firm becomes an EDGAR filer by examining the SEC Release No. 33-6977. We also incorporate all the subsequent changes and corrections to the initial phase-in list.[9] The SEC Release documents provide the list of company names and their Central Index Key (CIK). We manually match each firm to their record in Compustat using the company name and the CIK. We then use the linking file provided by the WRDS to link Compustat with CRSP. The last column of Table 1 reports the number of firms in each phase that we were able to match to the two databases.

B. The Anomalies

We start by examining a total of 320 anomalies replicated and shared by Chen and Zimmermann (2020), covering almost all the return signals that researchers have discovered to date.[10] By analyzing a comprehensive set of anomalies, we capture the full ramification of the information cost-saving effect of the EDGAR introduction on the anomalies' profitability. We follow Chen and Zimmermann (2020), who in turn follow the original academic papers that introduced each anomaly, their filters and datasets including CRSP, Compustat, IBES, the SEC's Form 13Fs, and

[9] The subsequent changes and corrections to the initial EDGAR phase-in list reported in SEC Release No. 33-6977 can be found in the SEC Release documents No. 33-7063, No. 34-34097, No. 33-7156, No. 34-35572, No. 33-7258, No. 34-36737, No. 33-7215, and No. 34-36220.

[10] Specifically, Chen and Zimmermann (2020) documents all the anomalies in Hou, Xue, and Zhang (2020), 98% of the anomalies in McLean and Pontiff (2016), 90% of anomalies in Green, Hand, and Zhang (2017), and 90% of the anomalies in Harvey, Liu, and Zhu (2016). We thank Andrew Chen and Tom Zimmerman for sharing the anomaly signal generating codes.

the Federal Reserve Economic Data (FRED). Chen and Zimmermann (2020) provide the quarterly versions of the anomalies by modifying the original characteristics to incorporate quarterly instead of annual information (assuming the standard one-quarter lag for quarterly data availability). Following this approach, we convert nine additional anomalies from annual to quarterly versions.[11]

We exclude penny stocks, that is, firms with a market capitalization below $50 million or a stock price lower than $5, because these stocks are not sufficiently liquid to be traded by institutional investors. Applying these two stock-level filters also mitigates the concern that the microcap returns shape our results (Hou, Xue, and Zhang (2020)). Also, our results primarily rely on value-weighted anomaly portfolio alphas, thus further mitigating the microcap concern. We adjust stock returns for delisting bias following the approach of Shumway (1997).

We eliminate anomalies that rely on binary signals, are unprofitable pre-EDGAR, or are too correlated with each other (we keep one of the pair). We first compute the Fama and French three-factor alphas (Fama and French (1992, 1993)) and the pairwise return correlation of the decile equal-weighted anomaly portfolio returns over a ten-year period pre-EDGAR, from October 1983 to September 1993.[12] Then, in the spirit of Green, Hand, and Zhang (2017), we exclude the 58 anomalies that have negative pre-EDGAR alphas. Arbitrageurs would have been unlikely to trade anomalies without positive alphas, and it is not clear how to capture alpha attenuation caused by EDGAR for such anomalies. Next, to ensure that we focus on relatively independent anomalies, we identify "twin" anomalies that have a pairwise return correlation above 0.9 and eliminate 28 anomalies by dropping one of the twins. Finally, we drop 29 anomalies with binary signals, such

[11] The nine anomalies are: accruals, sales growth over inventory growth, sales growth over overhead growth, change in sales vs change in receivables, revenue growth rank, change in depreciation to gross PPE, change in gross margin versus sales, change in sales to inventory, net income/book equity.

[12] We use the Fama and French *three*-factor alphas to filter out unprofitable anomalies because the Fama and French *three*-factor model was known to the investors at the time of EDGAR implementation whereas the Fama and French *five*-factor model was still not known to the public. Therefore, using the Fama and French *three*-factor alphas allows us to better capture the actual investors' trading activity prior to the introduction of EDGAR.

as whether a firm had paid dividends last month, because we cannot form decile portfolios for such anomalies. Our final sample includes 205 anomalies.

Next, we compute the benchmark-adjusted anomaly monthly returns over January 1992 to December 1997 sample period for the final sample of 205 anomalies, after controlling for the Fama-French five-factor alphas adjusted for momentum (henceforth the Fama-French six-factor alpha) for the equal-weighted and value-weighted decile and quintile portfolio returns. Jensen, Kelly, and Pedersen (2021) emphasize the importance of focusing on anomaly alphas instead of anomaly returns. We focus on alphas, but our results hold if we do not risk-adjust anomaly returns.

4. Results

A. Baseline Difference-in-Difference Results

EDGAR provides free and instant online access to SEC filings and thus lowers information costs, making it easier for arbitrageurs to identify mispriced stocks. Accordingly, the profitability of anomaly portfolios constructed from stocks that started to file with EDGAR should weaken. However, only the anomalies that rely on accounting information from EDGAR should attenuate.

We first compute the alpha for top-minus-bottom portfolio for a given anomaly, implementation phase, and month in two steps. First, we compute the difference between the top and bottom decile (quintile) portfolio returns, aggregated in the equal-weighted (value-weighted) manner, for each anomaly, phase, and month. Second, the alpha is calculated in a standard way as the sum of the residuals and the average alpha (intercept) from a regression of top-minus-bottom portfolio return on Fama-French factors, estimated over the sample period.

Our baseline specification estimates the effect of EDGAR implementation on the anomaly portfolio profitability using a standard difference-in-difference framework:

$$\widehat{\alpha_{a,p,t}} = \gamma_t + \gamma_a + \beta_1 * Post_{p,t} + \beta_2 * Post_{p,t} * ACC_a + \epsilon_{a,p,t}. \quad (1)$$

The dependent variable, $\widehat{\alpha_{a,p,t}}$, is the Fama-French six-factor alpha of the anomaly a top-minus-bottom portfolio for phase p in month t; γ_t are monthly time fixed effects; γ_a are anomaly fixed effects; $Post_{p,t}$ is an indicator variable equal to one if month t is on or after the effective date for phase p, and equal to zero before that date; and ACC_a is an indicator variable equal to one if a is an accounting anomaly, and equal to zero if a is a non-accounting anomaly. The standard errors are clustered by anomaly and month to address the potential correlation in errors (Petersen (2009)).

Presented in Panel A of Table II, the results are remarkably consistent across alternative specifications for the dependent variable—the Fama-French six-factor alphas for decile or quintile, equal-weighted or value-weighted portfolios. The portfolio alphas and the coefficients are expressed in percentages. As shown at the bottom of Table II, Panel A, accounting-based anomaly alphas decline by 47 to 62 basis points per month (or 5.7% to 7.4% per year) because of the EDGAR introduction. This decline completely offsets the average accounting anomaly alphas of 44.9 basis points per month from the pre-EDGAR period. By contrast, as shown in the top row of Table II, Panel A, non-accounting alphas do not decline post-EDGAR. The difference between the two, captured by the difference-in-difference coefficient, β_2, is between 42 and 49 basis points per month across different specifications, statistically significant at the one-percent level. It measures the amount of information costs investors faced in the absence of EDGAR.

These findings show that the information costs can be as important as other limits to arbitrage. There is a debate about the extent to which short sale costs affect anomaly profitability. Geczy, Musto, and Reed (2002) show that stock borrow fees explain a small portion of anomaly returns. By contrast, Chu, Hirshleifer, and Ma (2020) show that relaxed short sale constraints by Regulation SHO reduce abnormal returns of 11 anomalies by 72 basis points per month. A similar debate is ongoing about the effect of trading costs on anomaly profitability. Using TAQ data, Novy-Marx and Velikov (2016) show that the average trading costs range from 20 to 57 basis points for the mid-turnover anomalies. By contrast, Frazzini, Israel, and Moskowitz (2018) argue that institutional trading costs are much smaller than the effective bid-ask spreads in TAQ. Although our results do not speak to the two debates, the 42 to 49 basis point per month information costs that we estimate are comparable to the upper bounds for the trading and short sale costs. Also, investors need to acquire information to identify which stocks to buy or sell before they start trading. Therefore, investors incur information costs even before they pay trading or short sale costs.

One potential concern is that stock characteristics differ across EDGAR implementation stages, which could lead to changes in return predictability. To address this concern, Panel B of Table II estimates Equation (1) with a range of additional controls, especially logarithm of firm size.[13] The results in Panel B of Table II closely mirror those from Panel A, suggesting that additional controls do not affect our results.

Figure 1 illustrates how the average anomaly alphas of the treatment and control groups responded to the EDGAR implementation. It reiterates the salient features of our regression results from Table II. The treatment group—accounting anomalies—experiences a sharp decline in average alphas, from 0.73, 0.78, two and one months before EDGAR implementation, to 0.58 percent at the EDGAR implementation date, to the substantially lower values of EDGAR implementation for the affected stocks to -0.13, 0.23, and 0.22 percent per month during the first three months following EDGAR implementation. At the same time, the average alphas of non-

[13] The control variables include: the monthly mean of log market capitalization for all the stocks assigned to a given phase, Amihud illiquidity, earning surprise, book-to-market ratio, firm age, book leverage, return on assets, asset tangibility, sales growth, R&D relative to sales, total assets, institutional ownership, and capital expenditures.

accounting anomalies—the control group—did not experience such a large decline. If anything, they increased by about 15 basis points per month, as documented by the full blue lines covering pre- and post-EDGAR implementation periods in Figure 1.

B. EDGAR Effects and Information Availability

EDGAR prompted a decline in profitability of accounting anomalies because accounting information became more easily and readily available. The profitability decline should be more pronounced among the stocks for which the information was more difficult to gather in the pre-EDGAR period. To test this hypothesis, we use two empirical proxies for information availability—analyst coverage and market capitalization (e.g., Kelly and Ljungqvist (2012))—to classify stocks into high or low information-availability groups. For example, full-service broker-dealers provided their clients with analysts' research and opinions in addition to executing trades as part of an overall package of services (the so-called "soft" dollar arrangements). Thus, information for stocks with high analyst coverage is much easier to acquire.

We also confirm the main results from the previous section (based on decile/quintile portfolio sorts) using a two-stage approach inspired by Fama-MacBeth regression methodology. In the first stage, we estimate a cross-sectional regression of monthly returns on an anomaly signal for each phase and month. In the second stage, we estimate the standard difference-in-difference regression in Equation (1), except the dependent variable is the linear slope from the first stage instead of the top-minus-bottom portfolio alpha. We conduct this analysis separately for stocks with high and low information availability.

We first outline the methodology for this test. To gauge information availability, we first compute the average analyst coverage and market capitalization pre-EDGAR, from January 1990 to December 1992, and then classify each stock i as high-information, h (above-median analyst coverage; above-median market capitalization of equity), or low-information, l (below-median analyst coverage; below-median market capitalization of equity). For each of the 125 accounting anomalies, for each implementation phase, and for every month from 1992 to 1997, we estimate two first-pass regressions of the form

$$R_{i,a,p,t+1} = \alpha + \beta_{a,p,t} * SignalPercentile_{i,a,p,t} + \epsilon_{i,a,p,t}, \tag{2}$$

separately for high and low information availability stock groups. For each group, firm i is assigned to phase p for accounting anomaly a in month t; $R_{i,a,p,t+1}$ is the next-month return for stock i; $SignalPercentile_{i,a,p,t}$ is anomaly signal's percentile within stocks in phase p for anomaly a in month t. $\beta_{a,p,t}$ is the coefficient of interest. This first-pass regression step creates a panel of $\widehat{\beta_{h/l,a,p,t}}$, monthly beta estimates for information availability groups h/l (high or low). Next, we estimate the second-pass panel regression, similar to Equation (1):

$$\widehat{\beta_{h/l,a,p,t}} = \gamma_t + \gamma_a + \delta_1 * Post_{p,t} + \delta_2 * LoInfo_{h/l,a,p,t} + \delta_3 * Post_{p,t} * LoInfo_{h/l,a,p,t} + \epsilon_{h/l,a,p,t}, \tag{3}$$

where $\widehat{\beta_{h/l,a,p,t}}$ are the monthly beta estimates from the first-pass regressions; γ_t and γ_a are monthly and anomaly fixed effects; $Post_{p,t}$ is an indicator variable equal to one if month t is after the effective date for phase p, and equal to zero otherwise; and $LoInfo_{a,p,t}$ is an indicator variable equal to one for all $\widehat{\beta_{h/l,a,p,t}}$ associated with low information availability groups, and equal to zero for all $\widehat{\beta_{h/l,a,p,t}}$ associated with high information availability groups.

Presented in Table III, the results confirm our hypothesis for both analyst coverage and market capitalization. The EDGAR-prompted profitability decline for accounting anomalies is solely concentrated in low information availability stocks. For these stocks, monthly alphas decline by approximately 70 to 71 basis points per month, or about 8.5% per year, with the corresponding *t*-statistic of -3.2. By contrast, the EDGAR-prompted profitability decline for accounting anomalies among high information availability stocks is statistically indistinguishable from zero. These findings confirm the intuition that the effects of EDGAR introduction are particularly pronounced in the domain of stocks for which information was particularly costly to acquire pre-EDGAR.

C. EDGAR Effects and Implementation Complexity

When accounting information is scarce and, thus, information costs are high, investors may find it difficult to compute anomaly signals especially if those signals require multiple accounting inputs (e.g., if the inputs require purchasing multiple datasets). We test the hypothesis that the performance for anomalies that require many accounting inputs attenuate more dramatically post-EDGAR introduction. Intuitively, EDGAR expanded the set of accounting variables that investors can use and made them freely available in one place.

To test this hypothesis, we utilize a simple measure of implementation complexity: the number of distinct accounting variables required to compute a given anomaly signal. For each accounting anomaly, we count the number of contemporaneous accounting variables required and rank-order them accordingly. We then divide these accounting anomalies into low-complexity, medium-complexity, and high-complexity terciles. Similar to the main test in Section 4.A, we benchmark accounting anomalies against non-accounting ones. Four accounting anomalies require no contemporaneous accounting information. Therefore, we add them to the benchmark non-accounting anomalies.

We estimate the difference-in-difference regression similar to Equation (1), except for splitting the accounting anomaly indicator into three indicator variables that correspond to low, medium, and high complexity for accounting anomalies. These indicators are always zero for non-accounting anomalies. The following equation summarizes the setup:

$$\widehat{\alpha_{a,p,t}} = \gamma_t + \gamma_a + \beta_1 * Post_{p,t} + \sum_{Comp\in\{Low,Med,High\}} \beta_{Comp} * Post_{p,t} * Comp_a + \epsilon_{a,p,t} \quad (4)$$

The results in Table IV confirm the hypothesis that harder-to-implement anomalies attenuate more as EDGAR makes them easier to implement. First, consistent with the EDGAR introduction making a real difference across all accounting anomalies, abnormal returns for all three implementation complexity categories decrease significantly post-EDGAR. Specifically, the low-complexity anomalies feature post-EDGAR returns lower by 13 to 26 basis points per month than the returns to the reference, non-accounting anomalies (the regression estimates are economically large but not statistically significant).

Second, as the following rows of Table IV show, the magnitude of the decline of the post-EDGAR return gap grows monotonically with the increase in anomaly complexity; high-complexity anomalies experience large, highly statistically significant post-EDGAR gaps in the range between 55 and 66 basis points per month. The subsequent post-estimation test directly compares the post-EDGAR return gaps for high- and low-complexity anomalies; it reports statistically significant gaps in the range between 31 and 51 basis points per month.

Overall, the performance of harder-to-implement accounting anomalies attenuates more post-EDGAR. Our results are consistent with the notion that the EDGAR introduction decreased information costs for all accounting anomalies but disproportionately so for the anomalies that require many contemporaneous accounting inputs and thus are particularly challenging to implement.

D. Baseline Analyses Revisited: Long and Short Anomaly Portfolio Legs

In this section, we study how the attenuation of accounting anomaly portfolio profitability propagates through the equity market and affects different outcomes. We first focus on a well-known asymmetry between the short and long legs of equity anomalies. For example, Stambaugh, Yu, and Yuan (2012) find that 11 anomalies they study are concentrated in the short legs.

The results, presented in Table IV, confirm that profit attenuation effects are concentrated among the short legs of the accounting anomaly portfolios. Across all four columns of Table IV, the difference-in-difference coefficient estimates for the long leg accounting anomaly portfolios are small and statistically indistinguishable from zero. By contrast, the difference-in-difference coefficient estimates, statistically significant at the 1-percent level across all columns of the table, are 32 to 50 basis points per month. These estimates are comparable to the difference-in-difference coefficient estimates of 42 to 49 basis points from the baseline specification from Table II. These results suggest that short sale constraints can interact and elevate information costs.

E. Understanding the Mechanism behind the Attenuation of Anomaly Profitability

Lower costs of acquiring accounting information should make stock prices more efficient, reduce information asymmetry, and improve liquidity. We confirm this hypothesis using idiosyncratic volatility as a proxy for price (in)efficiency and Amihud (2002) measure for illiquidity.[14] We also examine trading turnover, a broad measure of the effect of EDGAR on investors' trading activity.

[14] Idiosyncratic volatility is the standard deviation of the regression residual from the Fama-French six-factor model, estimated from daily return data from the past month. Amihud illiquidity measure is defined as the past twelve-month average of daily return divided by dollar volume.

We first estimate how stocks in the top and bottom anomaly portfolios differ from stocks in other portfolios for the three measures of interest. We estimate a cross-sectional regression for every anomaly a, phase p, and month t:

$$Measure_{i,a,p,t} = \beta_0 + \beta_{a,p,t} * Treated_{i,a,p} + \epsilon_{i,a,p,t}\,, \tag{5}$$

where $Measure_{i,a,p,t}$ is idiosyncratic volatility, Amihud illiquidity, or stock turnover, and $Treated_{i,a,p}$ is equal to one if stock i is in a top or bottom decile portfolio and is equal to zero otherwise.

Next, we estimate for each of the three measures the second-pass cross-sectional regression similar to that from Equation (1):

$$\widehat{\beta_{a,p,t}} = \gamma_t + \gamma_a + \delta_1 * Post_{p,t} + \delta_2 * Post_{p,t} * ACC_a + \epsilon_{a,p,t}. \tag{6}$$

The dependent variable, $\widehat{\beta_{a,p,t}}$, is the sensitivity estimate from the first-stage cross-sectional regression in Equation (5); γ_t and γ_a are monthly and anomaly fixed effects; $Post_{p,t}$ is an indicator variable equal to one if month t is on or after the effective date for phase p, and equal to zero otherwise; and ACC_a is an indicator variable equal to one if a is an accounting anomaly, and equal to zero if a is a non-accounting anomaly. The standard errors are clustered by anomaly and month.

Table VI reports the results. The results for *non*-accounting anomalies reflect general trends in market variables—idiosyncratic volatility and Amihud illiquidity decrease and trading volume increases as the markets become more efficient and liquid over time.[15] However, we seek to evaluate the effects on accounting anomalies relative to non-accounting anomalies. Among the

[15] Whereas we do not explore liquidity measures based on intraday data such as the bid-ask spread and price impact, Goyenko, Holden, and Trzcinka (2009) show that Amihud illiquidity is strongly correlated with these measures.

EDGAR filer stocks in the decile portfolios for accounting anomalies (treatment group), idiosyncratic volatility decreased even more than it did for non-accounting anomalies. This result is consistent with the EDGAR introduction making stocks in extreme portfolios of accounting anomalies more efficiently priced. Similarly, these stocks became more liquid—Amihud illiquidity decreased much more for accounting anomalies than it did for non-accounting anomalies. More efficiently priced stocks experience less information asymmetry, a major contributor to illiquidity (e.g., adverse-selection component of the bid-ask spread). Finally, trading volume increases less for accounting anomalies, consistent with less trading by active institutional investors. The institutional investors appear to have recognized that these stocks once in EDGAR, provided fewer future opportunities for alpha generation.

F. Short Interest and Institutional Ownership

In this section, we study how investors' trading activity measures respond to the EDGAR implementation. Table IV shows that accounting anomaly profit attenuation is driven by portfolio short legs. An immediate implication of this finding is that the accompanying activity by short-sellers, reflected in stock short interest, ought to become more informative than the accompanying activity by long-only investors, reflected in stock institutional ownership. To test this hypothesis, we analyze long and short sides of the anomaly portfolios separately. For the long leg, we first estimate the following cross-sectional regression:

$$Measure_{i,a,p,t} = \beta_0 + \beta_{a,p,t} * TreatedLong_{i,a,p} + \epsilon_{i,a,p,t} \quad . \tag{7}$$

Analogously, for the short leg, we first estimate:

$$Measure_{i,a,p,t} = \beta_0 + \beta_{a,p,t} * TreatedShort_{i,a,p} + \epsilon_{i,a,p,t} \quad . \tag{8}$$

In these regressions, $Measure_{i,a,p,t}$ is short interest or institutional ownership, and $TreatedLong_{i,a,p,t}$ ($TreatedShort_{i,a,p,t}$) is an indicator variable equal to one if stock i is an EDGAR filer in the long (short) leg of the decile portfolio for anomaly a, in phase p, in month t, and equal to zero otherwise. Next, we estimate for each of the two measures the second-pass cross-sectional regression, separately for the long and short legs, as in Equation (5).

The results are reported in Table VII. Panel A presents the results for both long and short legs of anomaly portfolios, revealing no changes in institutional ownership for the long leg, as well as a sizeable, eight-percent increase in institutional holdings for the short leg for both accounting and non-accounting anomalies, with their difference near zero and statistically insignificant. As for short interest, its level has dropped for the short leg considerably more for accounting anomalies than for non-accounting anomalies: -12.59 percent versus -8.93 percent, a -3.65 percent difference, statistically significant at the one-percent level. The short interest increased for the long leg for both anomaly types, but the increase is considerably larger for accounting anomalies: 5.77 percent versus 3.23 percent.

Overall, that short interest increased for the long anomaly leg and decreased for the short leg, while institutional ownership changed similarly for the treatment and control groups is quite intuitive. First, because anomaly attenuation is concentrated in the short leg, we expect that short sellers respond more than long only investors do, suggesting that short interest should respond more than institutional ownership does. Second, pre-EDGAR, many stocks in the short leg of anomalies had had negative expected alphas, which short sellers had exploited. The EDGAR introduction made prices more efficient and fewer stocks in the short leg have had negative expected alphas in its aftermath, which, in turn, prompted short sellers to reduce their short positions. Put differently, short interest declined in response to the EDGAR introduction because of fewer opportunities on the short side.

Short interest increased for the long leg of anomalies for the same reason. Pre-EDGAR, most stocks in the long leg had had positive expected alphas and only a few of them had had negative expected alpha, resulting in few opportunities for short sellers and, accordingly, low short interest. The EDGAR introduction made the average expected alpha for the long leg of accounting anomalies close to zero, which can be viewed as about an even split between positive and negative alphas for individual stocks in the long leg. Therefore, the EDGAR introduction provided more opportunities for short sellers of long-leg stocks and short interest has increased relative to pre-EDGAR.

5. Robustness Tests

A. Concerns Regarding the First Implementation Phase

Randomized assignment of firms into implementation phases is crucial for the difference-in-difference methodology. If phase assignment is not fully random, a cross-phase comparison could be affected. The assignment was not perfectly random for the first phase. Before the EDGAR rollout in April 1993, the SEC called for volunteers to file electronically. This trial confirmed the integrity of the EDGAR system before engaging in a full-fledged implementation. The volunteer firms were subsequently assigned primarily to the first phase.

Also, accounting information for the first phase was delayed. The public could freely access EDGAR only after January 17, 1994 (Goldstein, Yang, and Zuo (2020), Chang, Ljungqvist, and Tseng (2021)); before that date, investors had access to EDGAR through Mead Data Central. Given the standard three-month information lag assumption we introduce, if EDGAR were not easily available prior to January 1994, the first phase would have little cost-saving effect because its effective date (October 1, 1993) falls before January 1994. The remaining implementation phases are unaffected by these issues because their effective dates are after January 1994.

We address these concerns by repeating the baseline difference-in-difference analysis after dropping the first phase. The results, presented in Table VIII, are very similar to the main results reported in Table II, thus alleviating concerns about the first phase. Indeed, as shown toward the bottom of Table VIII, accounting anomaly alphas decline by 46 to 62 basis points per month (or 5.5% to 7.4% per year) because of the EDGAR introduction. By contrast, as shown in the top row of Table VIII, non-accounting alphas do not decline post-EDGAR. The difference between the two, captured by the difference-in-difference coefficient, β_2, is between 41 and 53 basis points per month across different specifications, statistically significant at the one-percent level.

A related concern is that a small subset of implementation phases could be driving the results. Table A.I in the Appendix documents the contribution of each phase. Specifically, we interact the difference-in-difference indicator ($Post_{p,t} * ACC_a$) with indicators for each implementation phase except the first (the reference phase). The first column shows the results for the equally weighted quantile anomaly portfolios. While the effect of EDGAR introduction on anomaly returns varies across phases, the difference-in-difference coefficient is negative for every phase. The 8th phase features the largest alpha attenuation of 112 basis points, while the alphas declined the least (by 2 basis points) for the last phase. Thus, excluding any one phase has negligible effect on the overall results.

B. Pre-trends and Falsification Tests

The difference-in-difference analysis assumes parallel trends before the treatment. We formally test this assumption following the methodology from Gao and Huang (2020). Specifically, we estimate the baseline difference-in-difference regression over a four-year period *prior to* the actual EDGAR implementation, using pseudo-event dates. The pseudo-events of each EDGAR implementation phase are assumed to take place two years *before* the actual phase dates. Accordingly, the indicator variable $Post_{p,t}$ is redefined to equal one if month t is after the first *pseudo*-event date on which investors presumably trade on the latest EDGAR information related to the new EDGAR filers and is equal to zero if month t is before that pseudo-event date.

Panel A of Table IX presents the result for the pre-trend test. The results show that the parallel trend assumption is likely to hold in our difference-in-difference setting. The difference-in-difference coefficient switches sign and becomes positive but insignificant (depending on the specification, *t*-statistics range from 0.66 to 1.65). That is, accounting anomalies become slightly more profitable relative to non-accounting anomalies on the pseudo implementation dates, the opposite of the effect we find for the actual EDGAR implementation.

We also run a falsification test in a similar fashion as the pre-trends test. Again, following Gao and Huang (2020), we estimate the baseline difference-in-difference regression over a four-year period *following* the actual EDGAR implementation, using the pseudo-event dates from two years *after* the actual phase dates. The indicator variable $Post_{p,t}$ is redefined accordingly. Panel B of Table IX reports the results for the falsification test. Similar to the pre-trends test, the difference-in-difference coefficient switches sign and becomes positive but insignificant in most specifications. These tests are known to have low statistical power, thus the swings in coefficient estimates across specifications are not surprising.

These results indicate that accounting anomaly alphas decline shortly around the EDGAR implementation rather than long time before or after EDGAR. While neither test may have sufficient statistical power to estimate the effects precisely, the fact that the difference-in-difference coefficient have the "wrong" sign in both tests is encouraging. Accounting anomalies become slightly more profitable around pseudo EADGAR dates. Figure 1 further shows that accounting alphas decline discontinuously around the actual phase implementation dates, thus further supporting the difference-and-difference identification.

C. Dropping annual siblings, thin portfolios

In this section, we address two potential issues: annual and quarterly versions of the same anomaly conceivably could be highly correlated, and some anomaly portfolios could contain only a few stocks. First, as discussed in Section III.B, we constructed the sample of 205 anomalies by following the process of Chen and Zimmermann (2020). That process resulted in 23 anomalies based on both annual and quarterly portfolio formation, introducing the issue of potential double-counting. At the outset, returns for these "sibling" anomalies pass the correlation filter described in Section III.B and, thus, contain independent information. Nonetheless, we further address the issue of potential double-counting by estimating our baseline results from Table II on the sample of 182 anomalies, obtained from the full sample of 205 anomalies by dropping the annual "sibling" anomalies. The results, presented in Panel A of Table X, are virtually identical to those from Table II, indicating that the presence of annual sibling anomalies does not drive our results.

Second, the portfolio construction of long and short legs of an anomaly in each implementation phase could result in "thin" portfolios, consisting of relatively few stocks. These thin portfolios could make our estimates more variable and thus imprecise. However, this issue affects only a few observations because the median number of stocks in top/bottom portfolio is 25. To alleviate this concern, we replicate our baseline results from Table II with the added step of dropping all the observations based on "thin" portfolios consisting of fewer than five stocks in either long or short portfolio leg. This step creates a gently unbalanced panel by marking some anomaly-phase-months with few stocks as missing (13.4% of observations are affected). Once again, the results, presented in Panel B of Table X, are virtually identical to those from Table II. Therefore, the issue of thin portfolios does not affect our results.

6. Conclusion

In this paper, we investigate the causal effect of the information acquisition costs on the anomaly portfolio returns. We use the SEC's EDGAR implementation as a quasi-exogenous shock that lowers the costs of acquiring accounting information. Using the difference-in-difference framework, we find that alphas of accounting anomalies attenuate on average by 47 to 62 basis points per month (or 5.7% to 7.4% per year) in response to the EDGAR introduction. This decline explains away all of the pre-EDGAR accounting anomaly alphas. By contrast, the profitability of the non-accounting anomalies remains largely unaffected by the EDGAR introduction.

The profitability attenuation translates to the costs of acquiring accounting information (Grossman and Stiglitz (1980)) that investors had to bear in the absence of the EDGAR system. As Grossman and Stiglitz (1980) state in general and prove for CARA utility functions and normal return distributions, a decline in the cost of information prompts an increase in the fraction of informed traders participating in the markets (relative to uninformed traders), whose trading activities in turn make the price system more informative. Therefore, by lowering the accounting information costs, the EDGAR introduction increased price informativeness, which, in turn, eroded the profitability of trading strategies based on accounting anomalies. To the best of our knowledge, this paper is the first to estimate information costs in terms of returns (Dong et al. (2016) focuses on stock return synchronicity, and Chen, Kelly, and Wu (2020) study the hedge funds' changes in information-gathering efforts and trading activities) and show that they are comparable to trading or short sale costs. While non-information costs are extensively studied (e.g., Keim and Madhavan (1997), Korajczyk and Sadka (2004), Lesmond, Schill, and Zhou (2004), Novy-Marx and Velikov (2016), Frazzini, Israel, and Moskowitz (2018), and Patton and Weller (2020) study trading costs; Geczy, Musto, and Reed (2002), Drechsler and Drechsler (2014), and Chu, Hirshleifer, and Ma (2020) study short sale costs), the study of information costs is in its nascent stage.

We further explore how the effect of EDGAR on anomalies propagates to other market variables. First, we find that this effect is concentrated in stocks for which accounting information was particularly hard to acquire pre-EDGAR and is more pronounced among the more complex accounting anomalies. Second, the alpha attenuation is concentrated in the short leg of anomaly portfolios. Next, we confirm that stocks in the top/bottom portfolios became more efficiently priced and more liquid because of EDGAR. Finally, as these stocks are less mispriced, they attract fewer short sellers and other arbitrageurs.

Our results remain highly relevant for today's markets. To generate alpha, investors strive to establish information advantage by acquiring and analyzing novel data. Pre-EDGAR, accounting data was at the cutting edge of investors' data exploration efforts. The EDGAR implementation made accounting data widely available and thus less useful for generating alpha. Arbitrageurs move on to other, more costly, and thus less explored data. Our conclusions likely extrapolate to the alternative data.

Data, as a source of information, are central to arbitrageurs' success. For example, Citadel CEO Ken Griffin notes that, "our ability to leverage big data effectively in our investment process is critical to our success as a firm" (Randle (2018)). Many hedge funds have introduced Chief Data Officer positions to highlight the importance of data. Our paper offers an initial systematic attempt to understand the role that information costs play in the investment process.

References

Amihud, Yakov, (2002), "Illiquidity and stock returns: cross-section and time series effects", *Journal of Financial Markets,* 5, 31-56.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57-82.

Chang, Y. C., Ljungqvist, A., & Tseng, K. (2021). Do corporate disclosures constrain strategic analyst behavior? *Available at SSRN*.

Chang, Y. C., Hsiao, P.J., Ljungqvist, A., & Tseng, K. (2022). Testing disagreement models. *The Journal of Finance*, forthcoming.

Chen, Y., Kelly, B., and Wu, W., 2020, Sophisticated investors and market efficiency: Evidence from a natural experiment. *Journal of Financial Economics* 138, 316–341.

Chen, A. Y., & Zimmermann, T. (2020). Open source cross-sectional asset pricing. *Available at SSRN*.

Chu, Y., Hirshleifer, D., & Ma, L. (2020). The causal effect of limits to arbitrage on asset pricing anomalies. *The Journal of Finance*, *75*(5), 2631-2672.

Dong, Y., Li, O. Z., Lin, Y., and Ni, C. (2016). Does Information-Processing Cost Affect Firm-Specific Information Acquisition? Evidence from XBRL Adoption. *Journal of Financial and Quantitative Analysis* 51(2), 435-462.

Drechsler, I., & Drechsler, Q. F. (2014). *The shorting premium and asset pricing anomalies* (No. w20282). National Bureau of Economic Research.

D'Souza, J. M., Ramesh, K., & Shen, M. (2010). The interdependence between institutional ownership and information dissemination by data aggregators. *The Accounting Review*, *85*(1), 159-193.

Easley, D., & O'hara, M. (2004). Information and the cost of capital. *The journal of finance*, *59*(4), 1553-1583.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607-636.

Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). Trading costs. *Available at SSRN 3229719*.

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

Geczy, C. C., Musto, D. K., & Reed, A. V. (2002). Stocks are special too: An analysis of the equity lending market. *Journal of Financial Economics*, *66*(2-3), 241-269.

Goldstein, I., Yang, S., & Zuo, L. (2020). *The real effects of modern information technologies* (No. w27529). National Bureau of Economic Research.

Gomez, E. (2020). The Effect of Mandatory Disclosure Dissemination on Information Asymmetry: Evidence from the Implementation of the EDGAR System. *Available at SSRN 3577162*.

Goyenko, R.Y., Holden, C.W. and Trzcinka, C.A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2), pp.153-181.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies*, *30*(12), 4389-4436.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393-408.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5-68.

Hirshleifer, D., Hou, K., & Teoh, S. H. (2012). The accrual anomaly: risk or mispricing?. *Management Science*, *58*(2), 320-335.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, *33*(5), 2019-2133.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2021). *Is there a replication crisis in finance?* (No. w28432). National Bureau of Economic Research.

Kelly, B. and Ljungqvist, A. (2012). Testing Asymmetric-Information Asset Pricing Models. *The Review of Financial Studies, 25(5)*, 1366–1413.

Keim, D. B., & Madhavan, A. (1997). Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. *Journal of Financial Economics*, *46*(3), 265-292.

Kern, B. B., & Morris, M. H. (1994). Differences in the Compustat and expanded Value Line databases and the potential impact on empirical research. *Accounting Review*, 274-284.

Korajczyk, R. A., & Sadka, R. (2004). Are momentum profits robust to trading costs?. *The Journal of Finance*, *59*(3), 1039-1082.

Kothari, S. P., Shanken, J., & Sloan, R. G. (1995). Another look at the cross-section of expected stock returns. *The journal of finance*, *50*(1), 185-224.

Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of financial economics*, *71*(2), 349-380.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47*(1), 13-37. doi:10.2307/1924119

Love, J. P. (1992, January 14). Public Access to SEC information. *IU Bio-Archive. http://www.bio.net/bionet/mm/ag-forst/1992-January/000187.html*

Merton, R. C. (1987) A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3): 483–510

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability?. *The Journal of Finance*, *71*(1), 5-32.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica, 34*(4), 768-783. doi:10.2307/1910098

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, *29*(1), 104-147.

Patton, A. J., & Weller, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics*, *137*(2), 515-549.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, *22*(1), 435-480.

Porcano, T. (1986). Corporate tax rates: Progressive, proportional, or regressive. *Journal of the American Taxation Association*, *7*(2), 17-31.

Randle, C. (2018). Alternative Data: A New Source for Alpha. Available from www.futuresmag.com/2018/02/15/alternative-data-new-source-alpha .

Richards, D. (1988). Compact Disclosure. *RQ*, *27*(3), 413-415.

Stephen A. Ross. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, **13** (December 1976), 341–360.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425-442.

Shapiro, A. (2002). The investor recognition hypothesis in a dynamic general equilibrium: Theory and evidence. *The Review of Financial Studies*, *15*(1), 97-141.

Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance*, *52*(1), 327-340.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288-302.

Verrecchia, R. E. (1982). Information Acquisition in a Noisy Rational Expectations Economy, *Econometrica* 50, 1415−1430.

Whyte, A. (2020, July 9). How Much Are Managers Paying for Data? *Institutional Investors.* *https://www.institutionalinvestor.com/article/b1mf7j9918j9lg/How-Much-Are-Managers-Paying-for-Data* (accessed 25 July 2021).

Table I
EDGAR Implementation Schedule

This table shows the EDAGR implementation timeline. EDGAR was implemented in ten phases over three years. The SEC mandated the assigned firms to start filing electronically via EDGAR on the "implementation date." The anomaly literature assumes a one-quarter lag before accounting information is available to investors. "Effective date" accounts for this lag and is the first date when investors start trading the EDGAR filers stocks using the latest information retrieved from EDGAR. The last column reports the number of stocks in our sample for each phase (that we match successfully with Compustat and CRSP).

Implementation Phase	Implementation Date	Effective Date	Number of Stocks
1	4/26/1993	10/1/1993	149
2	7/19/1993	1/1/1994	541
3	10/4/1993	4/1/1994	564
4	12/6/1993	4/1/1994	737
5	8/1/1994	1/1/1995	1,033
6	11/1/1994	4/1/1995	866
7	5/1/1995	10/1/1995	858
8	8/1/1995	1/1/1996	756
9	11/1/1995	4/1/1996	386
10	5/1/1996	10/1/1996	2,723

Table II

Difference-in-Difference Estimates of EDGAR Effect on Anomalies

This table presents the coefficients from the main difference-in-difference regression from Equation (1). The coefficient associated with $Post_{p,t} * ACC_a$ reflects the decline in information costs post-EDGAR. It captures the gap in the extent to which anomaly portfolio alphas change in response to EDGAR for accounting anomalies relative to non-accounting anomalies. To facilitate interpretation, we also report the coefficient sum ($Post_{p,t} + Post_{p,t} * ACC_a$) and the mean of the dependent variable (portfolio alpha). The main regression is estimated on an anomaly-phase-month panel, where $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase p, and equal to zero if month t is before that date; ACC_a is an indicator variable that equals one if a is an accounting anomaly, and equals zero if a is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. All specifications contain anomaly and month fixed effects. The sample extends over 205 anomalies from January 1992 to December 1997. Panel A provides estimates without additional control variables. Panel B provides estimates with the full set of control variables including firm size, book-to-market, and illiquidity. The full list is described in the main text (Section 4.A). The portfolio alphas and the coefficients are expressed in percentages. The standard errors are clustered by anomaly and month. Robust *t*-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
Panel A: EDGAR Effect on Anomalies, no additional controls				
$Post_{p,t}$	-0.066	-0.042	-0.120	-0.123
(non-accounting anomalies)	(-0.63)	(-0.34)	(-0.95)	(-0.88)
$Post_{p,t} * ACC_a$	-0.419***	-0.429***	-0.421***	-0.494***
(difference-in-difference)	(-3.55)	(-3.69)	(-3.11)	(-3.47)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.485***	-0.471***	-0.541***	-0.617***
(accounting anomalies)	(-4.85)	(-4.03)	(-3.78)	(-4.09)
Mean of Dependent Variable	0.301	0.231	0.332	0.281
Panel B: EDGAR Effect on Anomalies, full set of controls				
$Post_{p,t}$	0.108	0.051	0.027	-0.026
(non-accounting anomalies)	(0.87)	(0.36)	(0.18)	(-0.16)
$Post_{p,t} * ACC_a$	-0.419***	-0.425***	-0.431***	-0.500***
(difference-in-difference)	(-3.57)	(-3.64)	(-3.24)	(-3.54)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.312***	-0.374***	-0.404***	-0.526***
(accounting anomalies)	(-2.79)	(-2.71)	(-2.71)	(-3.17)
Mean of Dependent Variable	0.301	0.231	0.332	0.281

Table III
Information Availability and EDGAR Effect on Accounting Anomalies

This table reports the coefficients from the regression in Equation (3), estimated separately for stocks with high and low information availability. Analyst coverage (Panel A) and market capitalization (Panel B) proxy for information availability. A stock is classified as low information if its analyst coverage (or market capitalization) is below the cross-sectional median pre-EDGAR (January 1990 to December 1992). For each of the 125 accounting anomalies, phase, and month, we estimate the first-pass regression from Equation (2) separately for high- and low-information availability stocks. This step creates an anomaly-phase-month panel of beta estimates for how well an anomaly signal predicts future stock returns. Next, we estimate the second-pass regression from Equation (3) that estimates how the EDGAR introduction affects anomaly predictability. The table reports the coefficients of interest for each regression and the difference between the two groups (last column). All specifications contain anomaly and month fixed effects. The standard errors are clustered by anomaly and by month. Robust *t*-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

	Panel A: Analyst coverage		
	(1) Low analyst coverage $\delta_1 + \delta_2 + \delta_3$	(2) High analyst coverage δ_1	(1) – (2) Difference $\delta_2 + \delta_3$
$Post_{p,t}$	-0.706***	0.032	-0.739***
	(-3.27)	(0.17)	(2.74)
	Panel B: Market capitalization		
	(1) Small stocks $\delta_1 + \delta_2 + \delta_3$	(2) Large stocks δ_1	(1) – (2) Difference $\delta_2 + \delta_3$
$Post_{p,t}$	-0.701***	-0.100	-0.608*
	(-3.23)	(-0.46)	(1.91)

Table IV

Baseline Difference-in-Difference by Implementation Complexity

This table documents the way that attenuation of accounting anomalies affected by EDGAR varies with the complexity of the anomalies' implementation. Specifically, the table presents the coefficients from the difference-in-difference anomaly portfolios regression in Equation (4) except the accounting anomaly indicator is split into three indicators for low, medium, and high complexity accounting anomalies. The table also presents the estimates of the statistical differences of the DiD coefficients between the high- and the low- complexity groups. We measure the implementation complexity as the number of distinct contemporaneous accounting variables required to compute the anomaly signal. Based on this measure, we split accounting anomalies into low, medium, and high complexity terciles. Non-accounting anomalies and four accounting anomalies that do not require contemporaneous accounting information serve as the control group (benchmark). For example, the $LowComp_a$ indicator is set to one if a is a low-complexity accounting anomaly, and is set to zero otherwise. The standard errors are clustered by anomaly and by month. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The last row reports the average monthly alphas for all anomalies pre-EDGAR.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
$Post_{p,t} \times LowComp_a$	-0.255	-0.239	-0.130	-0.195
	(-1.66)	(-1.42)	(-0.69)	(-0.95)
$Post_{p,t} \times MedComp_a$	-0.353***	-0.417***	-0.349***	-0.451***
	(-3.33)	(-3.32)	(-2.99)	(-2.79)
$Post_{p,t} \times HighComp_a$	-0.585***	-0.550***	-0.641***	-0.661***
	(-3.74)	(-3.72)	(-3.31)	(-3.40)
Post-estimation test:				
$Post_{p,t} \times HighComp_a - Post_{p,t} \times LowComp_a$	-0.330**	-0.311*	-0.510**	-0.466*
	(-2.51)	(-1.78)	(-2.60)	(-1.93)
Mean of Dependent Variable	0.301	0.231	0.332	0.281

Table V
EDGAR Effect on Long versus Short Anomaly Portfolio Legs

This table presents the coefficients from the baseline difference-in-difference regression from Equation (1), estimated separately for the long (Panel A) and short (Panel B) legs of the 205 anomaly portfolios. $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase p, and equal to zero if month t is before that date; ACC_a is an indicator variable that equals one if a is an accounting anomaly and equals zero if a is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor long (Panel A) or short (Panel B) alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. All specifications contain anomaly and month fixed effects. The sample extends over 205 anomalies in the period from January 1992 to December 1997. The portfolio alphas and the coefficients are expressed in percentages.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
	Panel A: Long Leg Anomaly Portfolios			
$Post_{p,t}$	-0.221	0.031	-0.238	-0.032
(non-accounting anomalies)	(-1.29)	(0.25)	(-1.28)	(-0.21)
$Post_{p,t} * ACC_a$	-0.0004	-0.126	0.059	-0.0445
(difference-in-difference)	(-0.00)	(-1.26)	(0.51)	(-0.35)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.221	-0.095	-0.179	-0.076
(accounting anomalies)	(-1.18)	(-0.71)	(-0.85)	(-0.49)
Mean of Dependent Variable	0.674	0.566	0.668	0.569
	Panel B: Short Leg Anomaly Portfolios			
$Post_{p,t}$	0.128	-0.088	0.099	-0.100
(non-accounting anomalies)	(0.70)	(-0.56)	(0.48)	(-0.55)
$Post_{p,t} * ACC_a$	-0.431***	-0.324***	-0.499***	-0.482***
(difference-in-difference)	(-4.07)	(-3.16)	(-3.77)	(-3.44)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.303	-0.412**	-0.400	-0.582***
(accounting anomalies)	(-1.45)	(-2.30)	(-1.64)	(-2.68)
Mean of Dependent Variable	-0.291	-0.288	-0.250	-0.230

Table VI
EDGAR Effect on Market Quality

This table shows the response of stock return idiosyncratic volatility, liquidity, and trading volume to EDGAR implementation. We first estimate a cross-sectional regression from Equation (5) for each of the three market outcomes, for each anomaly, phase, and month. This step estimates how the outcome (idiosyncratic volatility, liquidity, or volume) differs for the top/bottom portfolios versus the other portfolios. We then estimate the second-pass panel regression from Equation (6), a standard difference-in-difference regression. Idiosyncratic volatility is the standard deviation of the regression residual from the Fama-French six-factor model, estimated from daily return data from the past month. Amihud illiquidity is defined as the past twelve-month average of daily return divided by dollar volume. Log trading volume is the logarithm of average monthly trading volume. The sample period is from January 1992 to December 1997. The second-pass regressions contain anomaly and month fixed effects. The standard errors are clustered by anomaly and month. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

	(1) Accounting Anomalies $\delta_1+\delta_2$	(2) Non-Accounting Anomalies δ_1	(1) – (2) Difference δ_2
	Panel A: Idiosyncratic Volatility		
$Post_{p,t}$	-0.0075***	-0.0070***	-0.0005***
	(-20.36)	(-20.24)	(-2.97)
Number of Anomalies	125	80	205
	Panel B: Amihud Illiquidity		
$Post_{p,t}$	-0.2171***	-0.2754***	0.0583***
	(-10.05)	(-11.38)	(4.23)
Number of Anomalies	125	80	205
	Panel C: Log of Trading Volume		
$Post_{p,t}$	0.5982***	0.6931***	-0.0949***
	(8.89)	(10.21)	(-5.02)
Number of Anomalies	125	80	205

Table VII
Investor Response to EDGAR Implementation

This table shows how institutional ownership (Panel A) and short interest (Panel B) in the long and the short leg of the anomaly portfolios respond to the EDGAR implementation. For the long (short) leg, we first estimate the respective cross-sectional regressions from Equations (7) and (8). We then estimate long (short) side panel regressions from Equation (5). The sample period is from January 1992 to December 1997. The second-pass regressions contain anomaly and month fixed effects. The standard errors are clustered by anomaly and month. Robust *t*-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

Treatment portfolio (1-10)	(1) Accounting Anomalies $\delta_1 + \delta_2$	(2) Non-Accounting Anomalies δ_1	(1) – (2) Difference δ_2
	Panel A: Institutional Ownership		
Long leg			
$Post_{p,t}$	0.0035	0.0025	0.0010
	(0.82)	(0.59)	(0.19)
Short leg			
$Post_{p,t}$	0.0797***	0.0794***	0.0003
	(8.54)	(9.24)	(0.09)
Number of Anomalies	125	80	205
	Panel B: Short Interest		
Long leg			
$Post_{p,t}$	0.0577***	0.0323***	0.0254***
	(5.06)	(2.77)	(2.91)
Short leg			
$Post_{p,t}$	-0.1259***	-0.0893***	-0.0365***
	(-6.02)	(-4.92)	(-4.56)
Number of Anomalies	125	80	205

Table VIII
Robustness – Excluding the First Implementation Phase

This table presents the coefficients from the baseline difference-in-difference anomaly portfolios regression from Equation (1). $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase p, and equal to zero if month t is before that date; ACC_a is an indicator variable that equals one if a is an accounting anomaly and equals zero if a is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. All specifications contain anomaly and monthly fixed effects. The sample extends over 205 anomalies in the period from January 1992 to December 1997, with the observations associated with the first implementation phase excluded from the sample. The portfolio alphas and the coefficients are expressed in percentages. The standard errors are clustered by anomaly and month. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
$Post_{p,t}$	-0.035	-0.0265	-0.095	-0.095
(non-accounting anomalies)	(-0.33)	(-0.21)	(-0.74)	(-0.67)
$Post_{p,t} * ACC_a$	-0.426***	-0.462***	-0.413***	-0.534***
(difference-in-difference)	(-3.45)	(-3.78)	(-2.86)	(-3.57)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.461***	-0.489***	-0.508***	-0.625***
(accounting anomalies)	(-4.74)	(-4.01)	(-3.62)	(-4.19)
Mean of Dependent Variable	0.327	0.253	0.358	0.304

Table IX
Robustness – Pre-trends test, falsification test

This table presents the coefficients from the pre-trends and falsification tests of the baseline difference-in-difference regression results reported in Table II. Following Gao and Huang (2020), the regression reported in Panel A (Panel B) is estimated over a four-year period *prior to* (*following*) the actual EDGAR implementation, and the pseudo-event dates are assumed to take place 2 years *before* (*after*) the actual dates. $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase p, and equal to zero if month t is before that date; ACC_a is an indicator variable that equals one if a is an accounting anomaly and equals zero if a is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. The sample extends over 205 anomalies in the period from January 1992 to December 1997. The portfolio alphas and the coefficients are expressed in percentages. All specifications contain anomaly and month fixed effects. The standard errors are clustered by anomaly and month. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
	Panel A: Pre-Trends Test			
$Post_{p,t}$	-0.436***	-0.345**	-0.485**	-0.412*
(non-accounting anomalies)	(-2.73)	(-2.04)	(-2.50)	(-1.80)
$Post_{p,t} * ACC_a$	0.250	0.112	0.326	0.197
(difference in difference)	(1.52)	(0.66)	(1.65)	(0.89)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.186	-0.233*	-0.159	-0.215
(accounting anomalies)	(-1.59)	(-1.69)	(-1.03)	(-1.12)
Mean of Dependent Variable	0.256	0.182	0.300	0.225
	Panel B: Falsification Test			
$Post_{p,t}$	-0.176	-0.829*	-0.272	-0.422
(non-accounting anomalies)	(-0.50)	(-1.92)	(-0.59)	(-0.75)
$Post_{p,t} * ACC_a$	0.659	1.056**	0.684	0.494
(difference in difference)	(1.63)	(2.21)	(1.23)	(0.77)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	0.483**	0.227	0.412	0.072
(accounting anomalies)	(2.08)	(0.63)	(1.22)	(0.17)
Mean of Dependent Variable	0.464	0.172	0.556	0.235

Table X

Robustness – Drop Sibling Anomalies, Thin Portfolios

This table presents the coefficients from the two robustness tests of the baseline difference-in-difference regression results reported in Table II. Panel A features a modified sample of anomalies. We first identify the 23 sibling anomalies in our sample of 205 anomalies, that is, pairs of anomalies that exploit the same investment idea, but have portfolios formed based on annual signals and quarterly signals, respectively. We then drop the 23 annual siblings and estimate Equation (1) on the sample of 183 anomalies. Panel B features a full sample of 205 anomalies, but thin portfolio observations are dropped from the anomaly-phase-month panel. That is, observations are removed from the sample if the portfolio construction of either the long leg or the short leg for a given anomaly and implementation phase was based on fewer than five stocks. $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase *p*, and equal to zero if month *t* is before that date; ACC_a is an indicator variable that equals one if *a* is an accounting anomaly, and equals zero if *a* is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. The portfolio alphas and the coefficients are expressed in percentages. All specifications contain anomaly and month fixed effects. The standard errors are clustered by anomaly and month. Robust *t*-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
	Panel A: Excluding Annual Siblings			
$Post_{p,t}$	-0.070	-0.049	-0.136	-0.140
(non-accounting anomalies)	(-0.68)	(-0.40)	(-1.09)	(-1.02)
$Post_{p,t} * ACC_a$	-0.424***	-0.419***	-0.439***	-0.508***
(difference in difference)	(-3.77)	(-3.52)	(-3.43)	(-3.57)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.494***	-0.468***	-0.575***	-0.648***
(accounting anomalies)	(-4.69)	(-3.92)	(-3.81)	(-4.23)
Mean of Dependent Variable	0.256	0.182	0.300	0.225
	Panel B: Excluding Thin Portfolios (with < 5 stocks)			
$Post_{p,t}$	-0.214**	-0.181	-0.169	-0.141
(non-accounting anomalies)	(-2.38)	(-1.65)	(-1.56)	(-1.07)
$Post_{p,t} * ACC_a$	-0.223**	-0.240**	-0.269**	-0.312**
(difference in difference)	(-2.13)	(-2.19)	(-2.20)	(-2.26)
Post-estimation test:				
$Post_{p,t} + Post_{p,t} * ACC_a$	-0.437***	-0.421***	-0.438***	-0.453***
(accounting anomalies)	(-4.61)	(-3.93)	(-4.06)	-3.41
Mean of Dependent Variable	0.464	0.172	0.556	0.235

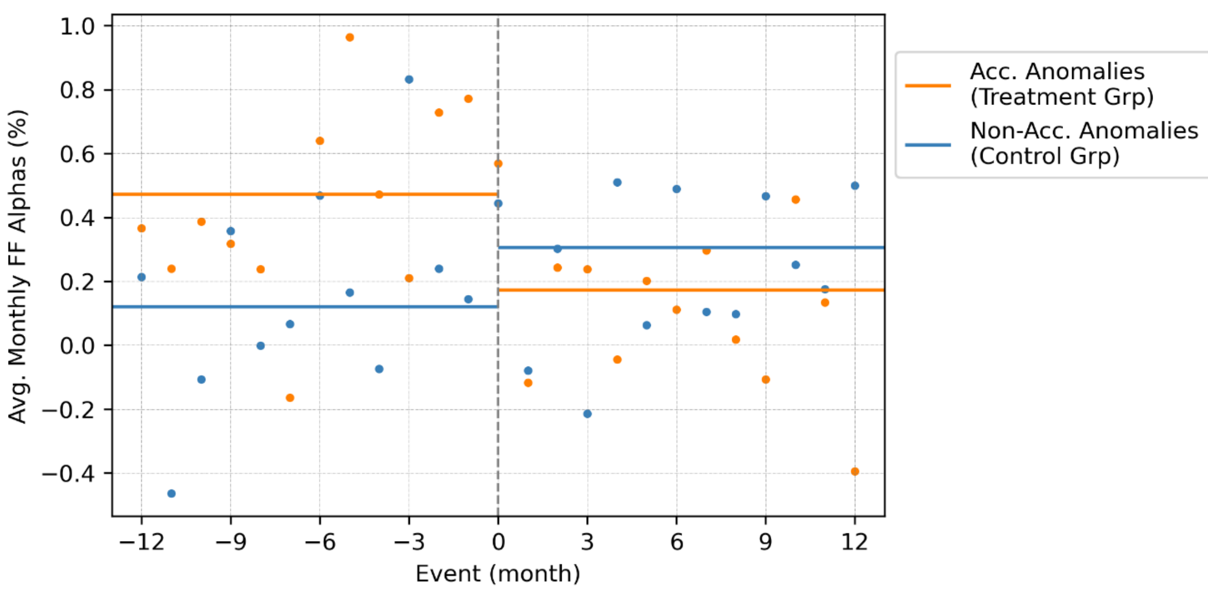

Figure 1: Effect of EDGAR on Anomaly Profitability

This figure shows the average alphas for anomaly portfolios around the staggered EDGAR implementation. The orange and blue lines show the average alphas for accounting and non-accounting anomalies, respectively, pre- and post-EDGAR implementation dates. Because implementation dates differ for each of the ten phases, we center each phase's implementation date at zero and average over implementation phases. Dots show alphas for individual months relative to implementation dates. Y axis reports the average monthly alphas estimated using six-factor Fama-French model across all four portfolio specifications we investigate in the main analysis.

Appendix

Table A.I

Baseline Difference-in-Difference by Implementation Phase

This table presents the coefficients from the difference-in-difference anomaly portfolios regression similar to that from Equation (1). The estimated specification features additional interaction terms of the form $Post_{p,t} * ACC_a$* Phase i, where i = 1, 2, 3/4, 5, 6, 7, 8, 9, 10, thus enabling the estimation of the difference-in-difference coefficient separately for each phase (phase 3 has the same effective date as phase 4). $Post_{p,t}$ is an indicator equal to one if month t is on or after the implementation date for phase p, and equal to zero if month t is before that date; ACC_a is an indicator variable that equals one if a is an accounting anomaly and equals zero if a is a non-accounting anomaly. The dependent variables $\widehat{\alpha_{a,p,t}}$ in the four columns of the table are the Fama-French six-factor alphas for quintile (1-5) or decile (1-10), equal-weighted (EW) or value-weighted (VW) portfolios. The sample extends over 205 anomalies in the period from January 1992 to December 1997. The portfolio alphas and the coefficients are expressed in percentages. All specifications contain anomaly and month fixed effects. The standard errors are clustered by anomaly and month. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	FF6 Alpha 1-5 EW	FF6 Alpha 1-5 VW	FF6 Alpha 1-10 EW	FF6 Alpha 1-10 VW
$Post_{p,t}$	0.0136	-0.0682	0.0466	-0.150
	(0.07)	(-0.29)	(0.20)	(-0.49)
$Post_{p,t} * ACC_a$* Phase 1	-0.286	0.0492	-0.371	0.154
	(-1.54)	(0.21)	(-1.61)	(0.52)
$Post_{p,t} * ACC_a$* Phase 2	-0.347	-0.234	-0.252	-0.209
	(-1.58)	(-1.22)	(-1.09)	(-1.05)
$Post_{p,t} * ACC_a$* Phase 3/4	-0.0585	-0.122	-0.0308	-0.0492
	(-0.41)	(-0.77)	(-0.19)	(-0.26)
$Post_{p,t} * ACC_a$* Phase 5	-0.171	-0.0310	-0.160	-0.0869
	(-0.91)	(-0.13)	(-0.61)	(-0.26)
$Post_{p,t} * ACC_a$* Phase 6	-0.428*	-0.438	-0.478	-0.785**
	(-1.94)	(-1.61)	(-1.65)	(-2.19)
$Post_{p,t} * ACC_a$* Phase 7	-0.748**	-1.026***	-0.940**	-1.434***
	(-2.19)	(-2.65)	(-2.23)	(-3.16)
$Post_{p,t} * ACC_a$* Phase 8	-1.122**	-0.615	-1.160*	-0.701
	(-2.33)	(-1.10)	(-1.85)	(-1.12)
$Post_{p,t} * ACC_a$* Phase 9	-0.775*	-1.230***	-1.008*	-1.302**
	(-1.84)	(-3.19)	(-1.93)	(-2.45)
$Post_{p,t} * ACC_a$* Phase 10	-0.0233	-0.562**	0.0742	-0.352
	(-0.10)	(-2.24)	(0.24)	(-1.13)
Mean of Dependent Variable	0.301	0.231	0.332	0.281