THE UNIVERSITY OF CHICAGO

# PROCESSING INDUSTRY CLASSIFICATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
KALASH JAIN

CHICAGO, ILLINOIS
JUNE 2023

## TABLE OF CONTENTS

# List of Figures

# List of Tables

## ACKNOWLEDGMENTS

I am grateful for the insights and support of my dissertation committee: Philip Berger (chair), Valeri Nikolaev, Haresh Sapra, and Chris Stewart. I thank Ray Ball, Jonas Dalmazzo, June Huang, Sinja Leonelli, Christian Leuz, Mike Minnis, Maxi Muhn, Fabian Nagel, Doug Skinner, Gurpal Sran, and workshop participants at Columbia, Northwestern, Purdue, Southern Methodist, University of Chicago, UVA Darden, and Washington University in St Louis for their comments and suggestions. I also thank Gerard Hoberg and Gordon Phillips for publicly sharing their Text-based Network Industry Classification data. All errors are my own.

# ABSTRACT

External users of financial reporting often rely on industry classification providers to reduce information processing costs. I study the economic consequences of capital market participants relying on the Global Industrial Classification Standard (GICS). I find that firms with higher quality classifications exhibit stronger liquidity, a lower cost of capital, and higher investment. When exploiting staggered GICS reclassifications as a shock to classification quality, I find relative improvements for firms with lower pre-treatment classification quality. The results are stronger for conglomerate firms that are more difficult to process and for firms that lack sell-side analyst coverage. Finally, I find increased investment for spillover firms after reclassified firms join their GICS industry group. Taken together, I provide evidence that industry classification providers generate capital market, real, and spillover effects through their information processing.

# CHAPTER 1

# INTRODUCTION

The objective of financial reporting is to facilitate decision making. External users of financial statements commonly compare a focal firm with a relevant peer group, often based on the focal firm's industry. However, industry peer groups are costly to construct and are generally outsourced to external classification providers (Fama and French, 1997; Bhojraj et al., 2003; Hovenkamp, 2021). In this paper, I study the economic consequences of agents relying on industry classification providers as information intermediaries.

Although many settings rely on external industry classification providers, I focus on the information processing role the Global Industrial Classification Standard (GICS) plays in the US capital markets, where GICS is used as the default provider (O'Shaughnessy, 2021).[1] Four of the five largest US asset managers, collectively managing over $24 trillion, rely on GICS classifications for benchmarking their passive sector funds.[2] In active funds, analysts often cover a single industry group - as defined by GICS - and their compensation is dependent on how they perform relative to the GICS benchmark. However, classification is difficult, and functional constraints prevent GICS from classifying firms as accurately as possible. Given GICS' role as a benchmarking and information intermediary, constraints on classification can limit the information processing GICS provides to capital market agents, which can affect firm outcomes.

I construct a measure of industry classification quality by comparing GICS to the Hoberg and Phillips (2010, 2016) (HP) text-based industry classification system. HP show that their measure captures more variation than other classifications in four characteristics that are key

1. Dennis Lynch, a portfolio manager managing approximately $100B at Counterpoint Global, discusses the critical role GICS plays as the default classification service in this interview.

2. The five largest asset managers are Blackrock, Charles Schwab, Vanguard, Fidelity, and State Street. Charles Schwab, Morgan Stanley, Fidelity, Vanguard, State Street, and BNY Mellon all use GICS sector benchmarks. Blackrock uses a mix of Thomson Reuters, GICS, and proprietary benchmarks, depending on the fund.

to valuation modeling: competition, profitability, market beta, and growth. If HP captures more explanatory power than GICS in these characteristics, agents using GICS may miss relevant peer characteristics that indicate changes in the focal firm's economics, creating information processing frictions.

HP is able to improve upon GICS because it relaxes two constraints GICS classifications face: transitivity and non-overlapping sets. In a transitive classification system, if firm A and firm B both have firm C as a peer, all three firms are grouped together. A non-overlapping system forces each firm to appear in a group only once. Although these constraints simplify the structure of the classification, it can lead to imprecise groupings. Consider, for example, Amazon as a focal firm. While Amazon is competing with Barnes & Noble in books and with Google in cloud computing, transitivity would force all three companies to be grouped into a single industry. Comparing Google with Barnes & Noble likely does not provide a relevant peer group to the agent studying these firms, rendering this classification methodology less useful for processing firm financials. In an intransitive overlapping system, each firm can have its own peer group, and each firm's peers need not be peers with each other. Intransitive systems have been heavily used in the recent literature that develops improved classification methodologies (Hoberg and Phillips, 2010, 2016; Lee et al., 2015; Kaustia and Rantala, 2021). These methodologies have been shown to explain more variation in key firm characteristics than the traditional non-overlapping transitive systems (Hoberg and Phillips, 2016; Li et al., 2013).

GICS is unable to adopt an intransitive methodology, due to the benchmarking role that GICS classifications play in the capital market. Asset managers use GICS to split the universe of firms across a team of analysts, which requires a transitive non-overlapping system. Such a system forces each stock's performance to be attributable to a *single analyst*, which creates accountability to prevent top performing stocks from being underweighted in or missing from the fund's portfolio (Bessembinder, 2018).[3] The benchmarking constraints prevent GICS from adopting intransitive systems and thus lead to information processing costs for agents relying on GICS.

I measure classification quality by identifying the overlap between the GICS and HP classifications. I validate the assumption that HP captures incrementally relevant information to GICS by rerunning tests as in Hoberg and Phillips (2016) in which they compare SIC, NAICS, and their own measure's explanatory power on a variety of firm characteristics. Similar to their SIC and NAICS findings, I show that the HP measure has significantly more explanatory power for competition, profitability, market beta, and growth than GICS. Given these characteristics are the first order inputs in firm valuation, these findings suggest that HP classifications provide incrementally useful information to market participants. Thus, I proceed with the assumption that GICS provides higher classification quality and more information processing to the capital market for firms with a greater overlap between the HP and GICS classifications.

I establish associative relations between classification quality and capital market outcomes. Because classification helps agents process firm information, it can reduce the information asymmetry between constrained investors who are reliant on GICS and unconstrained investors who are not. Thus, I hypothesize that firms with higher classification quality have lower information asymmetry, which leads to improved liquidity and a lower cost of capital (Diamond and Verrecchia, 1991). In addition to the indirect effect via liquidity, increasing the relevance of a peer group directly impacts the cost of capital due to the reduction in assessed covariances between a firm's expected cash flows and the rest of the market (Lambert et al., 2007). I find that a one standard deviation (SD) increase in classification quality is associated with liquidity improvements of 1.4% and a cost of capital reduction of 11 basis points (bps).

3. Similar issues exist for SIC and NAICS, which also use transitive non-overlapping systems.

If the cost of capital is affected, the hurdle rate for firms to take on new investment projects is also affected, generating real effects from classification quality. I find that a one SD increase in classification quality is associated with a Capital Expenditures/Assets (Capex) increase of 0.2 pp, and an R&D/Assets (R&D) increase of 0.9 pp. The increase is significantly larger in R&D, in line with information processing improving investment efficiency for more uncertain investments (Arrow and Fisher, 1974; Dixit and Pindyck, 1994). Furthermore, given the evidence from Martens and Sextroh (2021) that firms use analysts as an information intermediary to learn about peer activity, the real effects suggest GICS is used in a similar manner.

To improve the identification of the capital market and real effect findings, I exploit the staggered timing of large-scale reorganizations to the GICS classification system. I validate the shock by showing that after reclassification, firms on average have large increases in classification quality, and firms with low pre-period classification quality have larger increases. I use a triple-differences design that exploits variation across time, treatment, and pre-period classification. I show that relative to reclassified firms with high pre-period classification quality, reclassified firms with low pre-period classification quality exhibit relative increases in liquidity, decreases in the cost of capital, and increases in Capex and R&D, in line with the benefits of classification quality accruing to firms with worse classifications.

To further pin down the mechanism, I perform cross-sectional splits on the sample based on the availability of analyst coverage and the degree of conglomeration of the firm. Firms that have analyst coverage have other information processing services available to them, which can decrease the reliance on industry classification providers for information processing. I find that firms with no analyst coverage respond more strongly to classification improvements for liquidity and R&D. Highly conglomerated firms are more difficult to understand and have been shown to create processing difficulty (Cohen and Lou, 2012), leading to a larger potential impact for classification quality. I find that conglomerates are more sensitive to classification quality for the cost of capital, liquidity, Capex, and analyst forecast errors (AFEs).

Given that industry classification is inherently providing information about peers, I also study the spillover effects of reclassification. To satisfy the stable unit treatment value assumption (SUTVA), I construct three mutually exclusive groups: firms are either reclassified by GICS (treatment); are in a GICS group that receives reclassified firms, but the firm itself does not move (spillover); or no movement by the firm or peers in its group occurs (control). I estimate the magnitude of a spillover effect from an exogenous change in classification quality by estimating a difference-in-differences (DD) between the spillover and control groups. I find spillover firms see improvements in the cost of capital and liquidity, along with increases in Capex and R&D, suggesting that peers receive information spillover effects from focal firm reclassification.

Finally, as a placebo test, I reconstruct the classification quality measure, but replace GICS classifications with SIC classifications. Because capital market participants do not rely on SIC codes, SIC classification quality is not relevant to the participants' decision making. However, if the results are not due to information processing costs but are instead based on confounding industry-level characteristics, low overlap between SIC and HP constructed groups should load similarly to the GICS results.[4] Instead, I find insignificant or inconsistent relations for all specifications, providing suggestive evidence that I am not identifying an industry-level or firm complexity effect.

I contribute to the burgeoning literature on information processing costs, specifically on the effects of information intermediaries. Industry classification providers are ubiquitous, with academics, regulators, and capital market participants all relying on them. Despite their prominent role as an intermediary in many settings, the effects of their information

4. In particular, an alternate hypothesis could be that firms with more complexity are harder to classify, and thus have lower classification quality. That is, the measure developed in this paper proxies for firm complexity. If complexity is the driving factor behind the results, SIC classification quality should also proxy for firm complexity and load similarly to the GICS results, which I do not find.

processing have not been studied in the literature. Industry classification providers are most similar to data providers (e.g., I/B/E/S), which have been heavily studied. However, data providers are defined by Blankespoor et al. (2020) as intermediaries that offer information "without significant curation or interpretation." The key motivation behind this paper is that industry classification *requires* curation and interpretation. These intermediaries do not cleanly fit the existing taxonomy and have been understudied. I contribute by performing some of the first work on these intermediaries and demonstrate the capital market and real effects of their information processing.

I also contribute to the literature on the spillover effects of peer information. Bustamante and Frésard (2021) shows firms learn from peer investment. However, they do not study the effects from the quality of industry classification. I show improved classification quality can increase the learning from peers, which increases investment. Furthermore, Martens and Sextroh (2021) shows real effects arise from common analyst coverage. Analysts act as intermediaries and transport information about innovative peer activities to focal firms. I show industry classification providers are another intermediary through which such information spillovers can occur.

Finally, I contribute to the literature on industry classification. The prior literature has primarily focused on horse-races between existing classification systems (Bhojraj et al., 2003), or on developing new systems (Hoberg and Phillips, 2010, 2016; Kaustia and Rantala, 2021; Lee et al., 2015). I instead focus on the institutional constraints that GICS faces, providing evidence on why their classifications underperform recently developed methods. Motivated by the use of GICS as a default system, I construct a novel measure of classification quality that compares the classifications used by market participants with a benchmark developed recently in the literature. To my knowledge, I am also the first to exploit staggered GICS reclassifications as shocks to information.[5] The measurement and reclassification methods

5. Some papers have exploited *geographic* relocation as a shock to market efficiency, but not industry classification (Engelberg et al., 2018).

developed in this paper can be leveraged in many other settings - such as antitrust regulators relying on NAICS for market definitions - to further our understanding of industry classification providers.

# CHAPTER 2

# HYPOTHESIS DEVELOPMENT

## 2.1 Capital Market Effects

In a Kyle (1989)-style model, information asymmetry between investors reduces liquidity, as adverse selection concerns reduce trading between parties and increase bid-ask spreads. Diamond and Verrecchia (1991) show reducing information asymmetry can improve liquidity and reduce the cost of capital, and Amihud (2002) provides empirical evidence of the linkage between illiquidity and higher expected returns. Lambert et al. (2007) show that information asymmetry can directly affect the cost of capital by affecting the assessed covariance between a firm's and the market's expected cash flows. Using this literature, I generate testable hypotheses for the relation between classification quality and both liquidity and the cost of capital.

If industry classification is used as an information processing service, the quality of the classification can help determine information asymmetry. If the classification quality for a particular firm is high, this can reduce the information asymmetry between unconstrained investors, who can validate industry classification themselves and are thus not as affected by the quality of classification, and constrained investors, who rely on the classification service to sort firms. The opposite will be true for poorly classified firms, leading to cross-sectional variation in the amount of information processing GICS provides. Thus, I hypothesize that firms with stronger classification quality will have lower information asymmetry, which will cause an improvement in liquidity. Through liquidity, the cost of capital will reduce.[1]

Separate from the liquidity-driven cost of capital effect, improved classification quality can directly improve the cost of capital by providing investors more relevant peer groups.

1. In theory, any change in the covariances of a firm and the market should flow through a forward-looking measure of beta. However, if there exists measurement error, or beta does not fully capture forward-looking information, it is possible for there to be a separable effect of classification quality on cost of capital (Lambert et al., 2007).

Peer groups are used to benchmark a firm's performance to determine what portion of a firm's financial performance is driven by industry-level versus firm-specific factors. Higher quality classification can improve the industry-adjustment and thus offer more information on the *relative* performance of a firm compared with its industry. Better industry adjustment allows investors to extract firm specific information, which lowers the assessed covariances of the focal firm's expected cash flows with those of other firms, reducing the systematic risk and cost of capital of the focal firm (Lambert et al., 2007). Thus, I hypothesize that firms with stronger classification quality will have a lower cost of capital.

## 2.2 Real Effects

A lower cost of capital decreases a firm's hurdle rate for taking on new investment projects. Evidence also suggests illiquidity can reduce investment (Amihud and Levi, 2022). Thus, I expect firms with higher classification quality to be able to take on more investment projects, for both Capex and R&D.[2]

Beyond the real effects through the cost of capital and liquidity, industry classification can directly affect investment through *managerial learning*, which can heterogeneously affect different types of firm investment. When a manager is evaluating an investment, she is weighing the present value of the expected cash flows against both the explicit cost and opportunity cost of not delaying the investment (Arrow and Fisher, 1974). If I assume both Capex and R&D are partially irreversible, a real option value framework suggests there is value in delaying investment to a future period, as the delay allows the manager

2. Note that information asymmetry between the *firm manager* and the investor base can also generate real effects, as in Stein (1989). That is, if I assume that the firm manager is partially myopically focused on near term stock prices and has information that the market does not have, improved classification quality can reduce the information asymmetry between the manager and the market. The resulting improved price informativeness can improve the manager's incentives to invest efficiently. If the manager is underinvesting to inflate near-term earnings and consequently, near-term stock prices, the improved incentives will lead to increased investment. This agency story is difficult to separate from a cost of capital story, as both rely on capital market pricing informativeness to increase.

to learn more about the distribution of potential outcomes before investing (Badertscher et al., 2013; Dixit and Pindyck, 1994). This value is increasing in the uncertainty of the investment, because the benefits from delaying the investment will increase as marginal learning increases. Evidence shows that when facing uncertainty, managers underinvest relative to the neoclassical optimal level (Bloom et al., 2007; Arif et al., 2016). The degree of underinvestment is larger for R&D than Capex, given R&D's greater uncertainty. Thus, if classification quality decreases uncertainty by providing the manager with more information, I hypothesize that higher classification quality will increase investment. Due to the greater uncertainty in R&D, the increase in investment will be larger for R&D than Capex.

## 2.3 Analyst Forecast Error Effects

There is a significant literature on the ability of sell-side analysts to play an information intermediary role in the capital market as it relates to industry information. Boni and Womack (2006) show analysts largely provide information on variation within-industry. That is, their expertise is stronger to explain firm level within-industry information, as opposed to across-industry information. Furthermore, Boni and Womack (2006) and Merkley et al. (2017) suggest analysts rely on GICS classifications to organize firms into industries, implying that GICS plays a *complementary* role to analysts in information production. Complementarity suggests that if the classification quality of a firm is higher, analysts will have a better information environment to conduct within-industry analysis, thus lowering AFEs.

However, evidence also shows analysts offer their own industry-level expertise, which may act as a *substitute* and reduce their reliance on GICS. Ali et al. (2020) show analysts have an information advantage over managers for firms that are more sensitive to industry-level forces. Bradley et al. (2017) show analysts that have prior work experience in an industry have better forecast accuracy. Kadan et al. (2012) show analysts display both within-industry and across-industry expertise. If analysts have industry-level expertise, when GICS provides less information about industry classification, analysts can perform their own analysis to supplement the weak information environment. This substitution of analysis may attenuate the relation between classification quality and AFEs.

Finally, sell-side analysts may face differential benchmarking pressures than buy-side analysts, which may affect the degree of sensitivity to GICS information processing frictions. Buy-side analysts rely on GICS to define the benchmark upon which they are compensated. Sell-side analysts do not actively invest capital and thus have lower benchmarking pressures. If sell-side analysts do not have the incentives to meet or beat a GICS-defined benchmark, they may not face GICS-driven information processing costs, and analyst forecasts may not respond to classification quality. Overall, the past literature provides evidence for a negative or null relation between classification quality and AFEs.

## 2.4 Information Spillover Effects

In addition to improving agents' information processing for a focal firm, higher quality industry classification can also provide similar benefits through *information spillovers* to the focal firm's *peer group*. That is, if reclassification improves the focal firm's classification quality, this occurs because the focal firm is more similar to its new peer group than its old peer group. Consequently, the new peer group has received a relevant new peer and should also experience improved classification quality. Thus, any capital market or real effects from improved information processing for the focal firm should also accrue to its peers, as a result of the focal firm's reclassification. I hypothesize that through the capital market channel, firms that receive information spillovers will demonstrate lower cost of capital, improved liquidity, and greater investment.

Given existing evidence of *peer learning*, I hypothesize that separate from the capital market channel, improved classification quality will affect investment. Bustamante and Frésard (2021) show that firms learn from peer investment. Firm managers do not have perfect information on the fundamentals of investment projects they can take on, leading them to use peer investment decisions as information for their own investment decisions. This logic can be extended to within-industry variation in information. If classification quality increases a manager's ability to identify the most representative peers, this effect may improve their information set and increase investment more than a manager with low quality industry classification. Indeed, Bustamante and Frésard (2021) show that the information spillovers occur through product-market information, which is exactly the information that HP conveys. Thus, the measure of classification quality used in this paper will precisely measure variation in this kind of information spillover. Evidence also shows peer information can affect investment through an information intermediary. Martens and Sextroh (2021) find firms rely on the coverage of peers by analysts, and the likelihood of peer patent citation is higher if the focal firm and peer are covered by the same analyst. A similar effect could exist for GICS. If a GICS grouping contains relevant peer firms that are in the same product market, the focal manager can learn more about the investment behavior of these peers.

Based on the peer learning literature, I hypothesize that increased information from improved classification quality generates increases in both Capex and R&D. Additionally, given the reduction in investment uncertainty due to managerial learning (Bloom et al., 2007; Arif et al., 2016), I hypothesize that the increase in R&D will be larger than the increase in Capex. Finally, as is the case for directly affected firms, I hypothesize a negative or null effect of classification quality on spillover firms' AFEs.

# CHAPTER 3

# BACKGROUND

GICS is the dominant industry classification provider used in the investment management space. GICS was jointly created by Standard & Poor's (S&P) and Morgan Stanley Capital International (MSCI) in 1999 to assign firms to sectors, industry groups, industries, and sub-industries.[1] They created the service "due to the needs for a standardized, global, accurate classification offering" (Chan, 2021). GICS is "widely accepted as the industry analysis framework for investment research, portfolio management and asset allocation. Its universal approach to industries worldwide has contributed to increased transparency and efficiency in the investment process" (MSCI, 2005), suggesting that GICS is as a critical information intermediary in the capital market.

## 3.1 GICS Methodology and Reclassifications

GICS assigns firms to the narrowest designation in their structure, 8-digit sub-industries, based on their fundamentals. They classify firms into the sub-industry where most of their revenue is generated, and if revenue does not generate a clean classification, they rely on earnings. If neither revenue nor earnings fit into a single sub-industry, they rely on the "market perception" of the firm's sub-industry (Chan, 2021; Bhojraj et al., 2003). GICS aggregates the 8-digit sub-industries into a nested structure of 6-digit industries, 4-digit industry groups, and 2-digit sectors.

GICS performs annual reviews of their classification systems to see if any large-scale changes that need to be made. GICS has undergone 10 major reclassifications since beginning in 1999, as can be seen in Figure A1 of the Appendix. Aside from these reclassifications,

1. Prior to the creation of GICS, capital market participants largely used SIC codes. NAICS was introduced in 1997, and its adoption in the capital market has been minimal, partially due to the contemporaneous creation of GICS.

firms can idiosyncratically be reclassified if they undergo a business transformation such as a merger or spin-off. However, GICS tries to minimize changes in the industry classification and disregards fluctuations in business activities when possible (Chan, 2021).

When exploiting GICS reclassifications as plausibly exogenous shocks to industry classification quality, I use only the major reclassifications. I do not exploit idiosyncratic reclassifications, due to the likelihood that they are occurring concurrently with changing economics. For example, if a reclassification occurs due to a spin-off or merger, the underlying economics of a firm will likely change with the GICS classification. Thus, any change in the outcome variable will reflect the bundle of changing economics and classification. For large-scale changes, many firms are reclassified at the same time, reducing the concerns about idiosyncratic economic changes.

A potential concern on using large-scale changes is that they may reflect changing economics at the *industry* level. For example, the Communications sector was created in October 2018 due to the changing landscape in telecommunications, media, and advertising as a result of the internet. These firms were spread across three separate sectors but were increasingly becoming intertwined, suggesting that the reclassification is a function of the changing economics of this space. However, the specific timing of the change is likely to be uncorrelated with the underlying industry changes. The decision to create a Communications sector was first proposed in June 2014, as can be seen in Figure A2 of the Appendix. This reclassification took four years to take effect, due to repeated votes for approving the sector's creation. Thus, I only need to assume the *specific* timing of the reclassification is plausibly exogenous to the underlying economics of the industry. Furthermore, in my DD and DDD tests, I construct short-windows around the shock to ensure that the economics are relatively stable within the DD time period.

### 3.2 GICS' Role as an Information Intermediary

GICS plays a dominant role as an information intermediary in the capital market for many reasons. First, because GICS is created by index providers, it is embedded into index funds. For example, the S&P 500 has sector indices that are based on GICS methodology, due to S&P's involvement in GICS. Many sector-specific funds follow the GICS S&P indices as a benchmark. As stated in the introduction, four of the five largest investment advisors use GICS classifications for benchmarking their sector funds. In 2005, MSCI and S&P estimate that $3 trillion in funds are benchmarked to GICS (MSCI, 2005); the amount is likely significantly higher today. In addition to directly benchmarked assets, individual buy-side analysts who cover specific sectors are also internally benchmarked to GICS sectors. This benchmarking is internal and thus does not show up in fund prospectuses. Thus, the total assets tied to GICS are at least $3 trillion, but the total assets influenced by GICS are many multiples of that amount.

Second, internal benchmarking can exacerbate information processing frictions that vary with GICS classification quality. Benchmarking shapes active investor decision making, based on evidence of active managers closet indexing due to career concerns (Petajisto, 2013; Cremers and Petajisto, 2009). If an analyst is incentivized to closet index, the focus on firms that sit outside the industry group but have similar product-markets may decrease, which can create information awareness costs (Blankespoor et al., 2020; Merton, 1987). For firms that sit within the industry group but have distinct product-markets, the analyst covering the industry group will face greater fixed costs to understand the firms' distinct operations, which can create information integration costs (Blankespoor et al., 2020). Thus, due to the benchmarking pressures created by GICS, poor industry classification quality can exacerbate information processing costs.[2]

2. Separate from information processing costs, reclassifications can also create changes in incentives that can have capital market and real effects. Kashyap et al. (2021) identify a "benchmark inclusion subsidy", where firms inside a benchmark have inelastic, forced buying that increases the market price of the firm,

Third, although analysts can construct industry groupings themselves, outsourcing this process to an external information intermediary provides two key benefits. First, *within* an asset management firm, if each analyst constructs their own classification methodology, overlap in where a firm gets classified may occur, which hampers accountability, as attribution of the stock's performance will lie with multiple analysts. Outsourcing the classification to a transitive non-overlapping system can eliminate such incentive issues. Second, *across* asset managers, comparability benefits arise from all managers using the same classification system. For example, if all technology analysts are compared to a standard benchmark, determining the quality of the analyst will be easier for limited partners who have investments across multiple managers. Given these reasons, GICS plays a first-order role in supplying information about industry classification to the capital market, and variation in the quality of this information could lead to variation in processing costs.

## 3.3 Advantages of GICS and the Capital Market Setting

Relative to SIC and NAICS, studying the information processing role GICS plays has a few advantages. First, as shown by Bhojraj et al. (2003), GICS classifications explain significantly more variation in stock returns, valuation multiples, growth rates, and R&D than SIC or NAICS. Because this paper is about the information processing frictions that result from suboptimal classifications, I am interested in studying the costs that arise while using the best transitive, non-overlapping classification system.

---

lowering the cost of capital and thereby increasing the optimal level of investment for the firm. A similar dynamic can take place with GICS classifications. Buy-side analysts often cover a single industry for information processing efficiencies and their performance is benchmarked to that industry. Given the dominance of GICS in the asset management space, the definition of the industry benchmark is often determined by GICS. Applying the Kashyap et al. (2021) mechanism to an industry-specific benchmark would suggest that firms sitting within the GICS benchmark should benefit from a subsidy and thus have a lower cost of capital. This mechanism is distinct from the information processing mechanism, which lowers the cost of capital through lower information asymmetry. In this paper, I focus on the effects driven by information processing, not the benchmark inclusion subsidy. In the research design section, I provide details on how I disentangle these two effects.

Second, whereas NAICS and SIC are industry codes created by the US Government Census Bureau, GICS is a private sector classification system run as a joint-venture between Standard & Poor's (S&P) and MSCI, two large index providers. As a consequence, whereas the Census can only update NAICS every five years, S&P and MSCI are able to update the classifications at an annual or bi-annual basis. As a result, many firms get reclassified by GICS in a staggered fashion throughout my sample, which can be exploited as plausibly exogenous shocks to the quality of industry classification.[3]

Finally, due to the dominance of S&P and MSCI in the index construction market, most capital market participants use GICS as a default classification, which allows me to run placebo tests using SIC codes in the place of GICS codes. Because these agents do not rely on SIC, if the results are driven by information processing frictions, any variation in SIC classification quality should not vary with capital market or firm investment outcomes.

3. SIC codes are no longer being maintained by the government, but were historically updated at a maximum frequency of five years.

# CHAPTER 4

# DATA

## 4.1 Classification Quality

To construct a measure of GICS classification quality, I require GICS classifications and the benchmark HP classifications. The HP data come from the Hoberg Phillips data library. The HP data cover all firms with 10-K's in the EDGAR database for fiscal years from 1989 to 2019. The dataset captures product market similarity at the firm-year pair level by calculating the cosine similarity of the "Item 1. Business Description" text between two firms.

To briefly describe the HP process, they collect all 10-K's in a given year and extract the nouns and proper nouns from the "Item 1. Business Description" section. The union of every unique word in this set can be represented as a vector, and each firm-year can be represented by a vector of 1's and 0's, where the element of the vector is filled as 1 if the noun appears in the firm-year's Item 1. A vector is constructed for each firm-year, and the cosine similarity can be calculated as the dot product of the normalized vectors for any two pairs of firms $i, j$. The higher the cosine similarity, the greater the commonality is in the pair's Item 1, which is used as a proxy for product market similarity. The firms are considered peers if the cosine similarity is higher than a threshold determined by the unconditional likelihood of two firms being SIC 3-digit peers.

$$\text{Product Cosine Similarity }_{i,j} = \left(V_i \cdot V_j\right), \quad \text{where } V_i = \frac{P_i}{\sqrt{P_i \cdot P_i}} \quad \forall i, j \tag{1}$$

I observe GICS classifications from the historical GICS file on WRDS. The file is at the event-level, capturing the day a reclassification occurs for every firm covered by GICS. To merge this information with the CRSP monthly file, I convert the file into a firm-month level dataset.

Hoberg and Phillips (2010, 2016) show their measure captures variation in characteristics that are relevant to firm valuation. Thus, a measure of GICS classification quality should be higher when GICS classifications are more similar to HP classifications. I use Jaccard similarity as a proxy for classification quality. Jaccard similarity is defined as the intersection of two sets scaled by the union of those sets. That is, the greater the number of common peers the two classifications have, the higher the similarity score. This measure can be calculated at the firm-year level, because the HP measure updates for every new 10-K that is released.

For example, to calculate the Jaccard similarity score for salesforce.com in FY2006 (CRM,2006), I use the number of HP peers for salesforce.com (248), the number of GICS peers in the Software & Services industry group (466), and the overlap between these two sets (175). I can then calculate:

$$Similarity = \frac{HP_{CRM,2006} \cap GICS_{CRM,2006}}{HP_{CRM,2006} \cup GICS_{CRM,2006}} = \frac{175}{248 + 466 - 175} = 0.325 \qquad (2)$$

Summary statistics and histograms of the measure can be found in Table 1 and Figure 1. Figure 1A shows the distribution of group sizes for HP, GICS 4-digit industry groups, and GICS 6-digit industry classifications. HP follows a Pareto distribution, with a large mass of firms having fewer than 50 peers and a long tail of firms having upwards of 800 peers. The information processing costs generated by GICS are driven by benchmarking pressures that buy-side analysts face. Thus, the impact of classification quality is highest when calculated at the levels analysts are benchmarked to. Although there is variation in the benchmarking level, there is evidence that analysts work at the GICS industry group level (Merkley et al., 2017). Furthermore, from discussions with analysts in investment management, the benchmark generally seemed to exist anywhere from the industry to sector level. As a middle-ground, and to keep in line with the prior literature, I use GICS industry groups in my tests. [1] [2]

### 4.2 Remaining Data

The remaining data come from the CRSP Daily, CRSP Monthly, Compustat Annual, and IBES Quarterly files. I merge all data to the CRSP monthly file with varying lags. Compustat is merged on a six-month lag, as in Ball et al. (2015, 2016). Because the *Similarity* scores are based on 10-K data, they are also merged on a six-month lag. As in Frankel and Lee (1998), IBES files are merged on a one-month lag from the "time available" date of the estimates, to ensure the estimates are publicly available when calculating implied cost of capital. Since GICS begins in 1999, and I require the full cross-section of 10-K's to calculate HP scores across all pairs of firms, Jaccard Similarity is not calculated until the end of 1999. After lagging this measure by six months, the sample begins in July 2000. HP data are available until 2019, so the sample ends in June 2021.[3]

I study five primary outcome variables in this paper: cost of capital, liquidity, R&D, Capex, and analyst forecast errors. Cost of capital is calculated based on the methods in Easton (2004).[4] Liquidity is calculated as in Christensen et al. (2013, 2016) from the

1. The variation in the level of benchmarking depends on the size of the fund and the seniority of the analyst. As the fund grows larger or the analyst becomes more junior, the benchmark becomes more narrow. To exploit this variation for identification purposes, I would require an inside view of the benchmarking procedure for each individual fund. Without this information, assuming the industry group 4-digit level is the best alternative.

2. In addition to the conceptual argument, GICS 4-digit industry groups are better because they provide greater variation in *Similarity*, given their greater mass in the right tail of the distribution. Figure 1B shows direct evidence of this. While the *Similarity* distribution for GICS industry ends by around 0.6, GICS Group has significant variation past 0.8. This increased cross-sectional variation in *Similarity* improves the sensitivity of the tests by separating firms more effectively.

3. *Similarity* is calculated annually, so the 2019 scores run from July 2020 - June 2021.

4. Cost of capital is often calculated based on the average of four implied cost of capital models: Gebhardt et al. (2001); Claus and Thomas (2001); Easton (2004); Ohlson and Juettner-Nauroth (2005). However, the cost of capital that I am trying to capture is the firm manager's hurdle rate that is used when evaluating new investments. Prior literature has shown that there is a large buffer between the market implied cost of capital and the internal hurdle rate used by managers (Jagannathan et al., 2016; Graham, 2022). Therefore, I require an implied cost of capital that is closer to the manager hurdle rate. Larocque et al. (2018) show

first principal component of four measures of liquidity: bid-ask spread, percentage of zero trading days, Amihud (2002) illiquidity, and transaction costs estimated by the methodology in Corwin and Schultz (2012).[5] The measures are constructed using the CRSP Daily file and then aggregated to the monthly level. R&D and Capex are scaled by lagged total assets, all coming from the Compustat Annual file.[6] Finally, AFEs are calculated from the IBES Quarterly file as the absolute deviation between the mean estimate and the actual EPS reported.[7]

I perform sample-splitting mechanism tests based on analyst coverage and the conglomeration of firms. Analyst coverage is an indicator variable equal to one if a firm-month appears within the IBES Quarterly file. GICS defines conglomerates as firms with multiple business-lines spread across three or more GICS sectors in which no single sector contributes the majority of revenue or profits. They are classified in the sub-industries "Industrial Conglomerates" or "Multi-Sector Holdings." Given that *Similarity* is calculated at the 4-digit group level, I define conglomerate as an indicator variable equal to one for any firm in the same 4-digit industry group as the two conglomerate sub-industries. Finally, I perform a placebo test by measuring classification quality for SIC, instead of GICS. I use the same methodology to calculate *Similarity*, but replace 4-digit GICS industry group codes with 2-digit SIC codes.

---

that the Easton (2004) measure is most correlated with manager hurdle rates. I thus use this estimation procedure in my tests.

5. Christensen et al. (2013, 2016) use the regression based method from Lesmond et al. (1999) (LOT) to estimate transaction costs. However, Fong et al. (2017) perform horse-races across many measures and find that the closed-form Corwin and Schultz (2012) method outperforms the LOT method.

6. Koh and Reeb (2015) find that firms with missing R&D have different patent properties than firms with reported zero R&D. When using R&D as a covariate, they recommend adding an indicator variable for "Missing R&D" and replacing missing R&D with the industry average value or 0. Given that R&D is an outcome variable in this analysis, an indicator variable will not add to the sample. Furthermore, given most analyses include industry-time fixed effects, industry averages will be absorbed by the fixed effects. The R&D results do not change significantly whether replacing R&D with zero or dropping the missing observations. Given the different properties that Koh and Reeb (2015) find between 0 and missing R&D firms, I do not replace R&D values and leave them as missing.

7. To study the change in forecast accuracy, the natural log of AFEs is used in the empirical tests.

# CHAPTER 5

# RESEARCH DESIGN

The research design consists of three model specifications: associative tests, a triple-differences test around GICS reclassifications for directly treated firms, and a spillover DD test around reclassifications for indirectly affected firms.

## 5.1 Associative Tests

I first estimate associative tests between capital market or firm investment outcomes and classification quality:

$$Y_{ijt} = \beta Similarity + \gamma X_{it} + \nu_i + \eta_{jt} + \epsilon_{ijt} \quad (3)$$

In this specification, $i, j, t$ correspond to firm, industry, and month, respectively, $X_{it}$ are firm-month controls, and *Similarity* is the Jaccard similarity measure that is used as a proxy for GICS classification quality. The outcome variables are measures of liquidity, cost of capital, R&D, Capex, and AFEs. I argue that due to the reduction in processing costs from improved classification quality, higher *Similarity* will result in stronger liquidity, a lower cost of capital, and greater firm investment. Thus, I hypothesize that $\beta < 0$ for illiquidity and cost of capital, and $\beta > 0$ for R&D and Capex. Given evidence that GICS may play a substitute or complementary role to analysts when conducting industry analysis, I hypothesize either a negative or null relation between *Similarity* and AFEs.

To focus on the effect of *Similarity* on the outcome variables, I use both firm- and industry-year fixed effects. Firms with high or low classification quality are likely inherently different from each other on multiple dimensions; I use firm fixed effects to capture these time-invariant differences. I use industry-year fixed effects to control for time-varying changes in underlying industry economics. Moreover, because *Similarity* is measured as the overlap between HP and GICS classifications, these fixed effects will compare firms with similar industry classification measures. For example, the "Banks" industry group has the highest average *Similarity* score in the sample, due to the group's homogeneous, stable structure. On the other hand, the "Capital Goods" industry group has the lowest score, due to the heterogeneous and dynamic industry structure. Thus, industry fixed effects not only control for time-varying economics, but also prevent comparisons across firms with different latent *Similarity* score distributions.[1]

### 5.2 Triple-Differences Tests

To establish a causal effect of classification quality, I exploit staggered GICS reclassifications as plausibly exogenous shocks to classification quality. I depart from the traditional difference-in-differences (DD) design in two ways. First, I use a difference-in-difference-in-differences (DDD) design (Olden and Møen, 2022; Granja, 2018). A simple DD design would compare firms that have been reclassified by GICS to matched firms that were not reclassified. I hypothesize that the causal effect stems from an improvement in classification quality. Thus, I further difference between treated firms with low and high pre-treatment *Similarity* scores. This identifies the differential treatment effect on the outcome variables for treated firms with more to gain from reclassification.

Second, to address negative-weighting and dynamic treatment effect concerns brought up in the recent staggered DD literature, I construct a "stacked regression" dataset (Barrios, 2021; Baker et al., 2022; Cengiz et al., 2019). Stacked designs avoid using already-treated (to-be-treated) observations as control observations later (earlier) in the time-period of the sample. The literature provides no consensus on the best method to address these concerns,

1. As a robustness check, to better control for differences in economics across industries, Table A1 provides similar analyses with a variety of overlapping industry fixed effects. All permutations of GICS, SIC, and the Hoberg and Phillips (2010, 2016) Fixed Industrial Classification (FIC) are used. The results are robust to the permutations.

but stacked designs address the key issues regarding treatment timing while being simple to implement, transparent, and efficient (Baker et al., 2022).

To implement this design, for each of the five major reclassifications, I construct an independent dataset including both treated firms and never-treated, coarsened exact matched firms. I coarsen five variables - profitability, book to market, investment, size, and *Similarity* - into quintiles and match on the values in the year before treatment for each cohort dataset.[2] I keep all observations within the [-3,+5] year window of the treatment year for both groups.[3] I then append these five datasets together to construct a single "stacked" dataset. The DDD coefficient is a weighted average of each group's DDD coefficients.

The specification is as follows:

$$\begin{aligned} Y_{ijkt} =& \beta_1 \text{TreatPost}_{it} + \beta_2 \text{Low Pre-Period Similarity}_i \\ & + \beta_3 \text{Low Pre-Period Similarity}_i \times \text{TreatPost}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{jkt} + \epsilon_{ijkt} \end{aligned} \tag{4}$$

As in the prior specification, $i, j, t$ correspond to firm, industry, and month, respectively, whereas $k$ corresponds to the cohort dataset for a given reclassification. Pre-Period Similarity is an indicator variable that is fixed across-time for every firm and equals 1 if the firm has a below industry median *Similarity* score in the five years leading up to a reclassification event, and 0 otherwise.[4] By keeping the variable fixed, I ensure I am comparing low *Similarity* treated firms with low *Similarity* control firms. Treat Post is equal to one if the firm has

2. Given the focus of this paper on how capital market participants use industry information to value firms, I match on the five Fama and French (2015) factors to best control for characteristics known to vary with expected returns. I do not match on market beta, due to the measurement error and noise inherent in the measure. Bloomfield (2021) matches on the pre-treatment value of ln(Market Share), which measures heterogeneous exposure to treatment in their setting. I analogously match on the pre-treatment value of *Similarity*.

3. Firms are only considered treated during their first reclassification in order to avoid overlap situations in the [-3,+5] window. That is, with this design, I avoid classifying a firm as untreated in the pre-period when it was previously treated in a prior reclassification.

4. If a firm has fewer than five filled observations prior to treatment, I average *Similarity* over the available data.

been reclassified, and 0 otherwise. $X_{it}$ are firm-month controls as before.

$\nu_{ik}, \eta_{jkt}$ correspond to cohort-firm and cohort-FIC-month fixed effects, respectively. Unlike the firm and industry-month structure used previously, due to the stacked regression design, the fixed effects are interacted with a cohort fixed effect, corresponding to the independent dataset that the observation comes from. Furthermore, given that reclassifications occur *across* industry groups, I cannot use time-varying GICS industry group classifications for my fixed effects, because they will absorb some of the treatment variation. Instead, I use the Hoberg and Phillips (2010) FIC classification system. By maintaining the transitivity and non-overlapping constraints, FIC can be used in a fixed effect structure. This allows me to flexibly control for time-varying industry economics without absorbing variation related to GICS reclassifications.[5] I also show DDD results with no industry fixed effects, instead using only cohort-firm and cohort-month fixed effects. To satisfy SUTVA restrictions, I drop all spillover observations, such that I am comparing treated observations with never-treated observations within the cohort-specific window.[6][7]

The coefficient of interest is the DDD coefficient, $\beta_3$, which measures the incremental treatment effect for the low pre-period classification quality group, relative to the DD treatment effect for the high treatment quality group. $\beta_1$ measures the main treatment effect for reclassified firms with high pre-period classification quality, while $\beta_2$ is subsumed by the cohort-firm fixed effects.

5. Although FIC does not provide as much explanatory power as the unconstrained HP method, the latter cannot be used in a fixed effect structure, due to its intransitivity.

6. Spillover firms are firms that receive new peers in their GICS industry group as a result of reclassification; I do not classify the "left behind" firms that lose peers as spillover. Although left behind firms do experience changes in their GICS peer groups as a result of departing treated firms, untabulated analyses show *Similarity* is largely unchanged for these firms, suggesting the classification quality effect on these firms is minimal.

7. Berg et al. (2021) shows that many designs in corporate finance face spillover concerns, and the SUTVA restriction is often violated. By dropping the peers of reclassified firms that face these spillover concerns, I maintain the SUTVA assumption when estimating the direct effect.

### 5.3 Spillover Tests

Finally, to study information spillover effects, I identify the treatment effect for firms that receive information spillovers from reclassifications compared with matched control firms that are not treated and do not receive any new peers in their industry groups. A firm is classified as "Spillover" if it does not reclassify during one of the major reclassifications, but a new firm enters the spillover firm's industry group during that time.

Similar to the DDD design, I construct a stacked dataset by appending independent datasets for each of the five major reclassifications, but this dataset includes only spillover firms and never-treated, coarsened exact matched firms. That is, I drop all treated firms to ensure I am comparing only spillover firms with never-treated firms. This structure satisfies SUTVA restrictions, because the spillover firms, treated firms, and never-treated firms all sit in mutually exclusive groups. The specification makes two changes to the stacked DDD specification above. First, I replace Treat Post with Spill Treat Post. Second, due to the tighter identification the spillover design creates, I now use a simple DD design, i.e., I don't interact pre-period *Similarity* with Spill Treat Post:

$$Y_{ijkt} = \beta_1 \text{SpillTreatPost}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{jkt} + \epsilon_{ijkt} \quad (5)$$

I primarily focus on the effects driven by information processing, not the effects driven by benchmark-based analyst incentives. Thus, I require a setting that can disentangle the two mechanisms. Shocks to classification quality on their own are not sufficient. When a firm is reclassified, it experiences both a change in its classification quality and its benchmark, which can affect buyside analyst incentives to closet index towards their GICS benchmark. Spillover tests can disentangle these two mechanisms. Firms that receive newly reclassified peers experience changes in classification quality, due to the changing peer environment. However, because these firms remain in the same GICS group, their benchmarking does not change significantly.[8] [9]

8. The benchmarking effect will not be exactly zero, as some reweighting may occur as a result of peers moving groups. However, this effect is likely much smaller than a DD on directly treated firms.

9. I can also use the triple-differences design to partially address the benchmark inclusion subsidy concern. In the DDD design, I difference between two groups of treated firms. Thus, under the assumption that both the high and low pre-treatment classification quality firms are similarly exposed to changes in the benchmark inclusion subsidy, the effects will be differenced out. To the extent these firms have systematically different exposures, DDD will not address this concern.

# CHAPTER 6

# RESULTS

Before estimating models of the relation between capital market or investment outcomes and *Similarity*, I validate that *Similarity* is a reasonable measure of classification quality. Implicit in using *Similarity* is the assumption that the HP classification system captures incremental information relative to GICS classifications. Hoberg and Phillips (2016) show that HP captures incremental information relative to SIC and NAICS codes on the key components of firm valuation: competition, profitability, systematic risk, and growth. I replicate these results and show similar findings for the incremental information of HP relative to GICS.

The results can be found in Table 2. HP's intransitive Text-based Network Industry Classification (TNIC) measure captures more variation, as measured by R-squared, than GICS, SIC, or NAICS codes. This is true for both fixed effects regressions and univariate regressions using the average characteristic value within the industry grouping. Since TNIC is intransitive and overlapping, it cannot be used in fixed effects regressions. Instead, I use the HP fixed industry classification (FIC) grouping, which is meant to combine the HP methodology with the transitive, non-overlapping restrictions placed upon SIC, NAICS, and GICS. I show that TNIC demonstrates greater explanatory power than FIC as well, in line with the benefits of TNIC being driven by the relaxation of these constraints. Thus, I show that the HP classification contains incremental information to GICS and continue with the assumption that *Similarity* is a reasonable proxy for classification quality.

## 6.1 Associative Results

I next run associative tests between the outcome variables of interest and *Similarity*. I hypothesize that higher classification quality will reduce information asymmetry and improve liquidity. In column (2) of Table 3, I show a one SD increase in classification quality is associated with a 3.5% improvement in liquidity.[1] After including firm fixed effects to address time-invariant latent differences in firm characteristics, the one SD increase is associated with a still-significant 1.4% improvement in column (7).[2]

For the cost of capital, I anticipate higher classification quality will reduce the cost of capital in two ways. First, an improvement in liquidity should reduce the illiquidity premium documented in the literature and reduce expected returns (Amihud, 2002). Second, based on the logic in Lambert et al. (2007), the increased firm-related information from improved classification quality will lower the assessed covariances between a focal firm and other firms' expected cash flows, lowering assessed systematic risk and the cost of capital. Column (1) shows that a one SD increase in classification quality is associated with a 96 bps decrease in the cost of capital, or a 6.5% decrease relative to the mean. After controlling for firm fixed effects, this reduces to a marginally significant 11 bps decrease in column (6), or a 0.7% decrease relative to the mean. Given that the cost of capital effect is motivated by cross-sectional variation in the covariances of expected firm cash flows, firm fixed effects may be absorbing critical variation in the cost of capital. However, significant latent differences between firms with high and low *Similarity* are likely, beyond the information effects. Thus, these estimates provide a reasonable upper and lower bound for the effect.[3]

1. Note the liquidity factor is composed of four measures of illiquidity, so a negative coefficient corresponds to an improvement in liquidity.

2. To more directly test price efficiency, I also conduct short-window capital market response tests for focal firms around GICS peer and non-GICS peer earnings announcements (EAs). The results can be seen in Table A5 in the appendix. I find that when a focal firm shares a GICS classification with a peer, the focal firm has stronger price efficiency around the peer EA relative to non-GICS peers. Furthermore, the sensitivity of price efficiency to the closeness of the firms, as measured by the HP Score, only exists for non-GICS peers. That is, when a focal firm and a peer are not GICS peers, price efficiency increases in the economic similarity of the two firms. However, when they are GICS peers, this economic similarity does not matter. Instead, there is a fixed increase in price efficiency for all GICS peers, regardless of the degree of similarity. This suggests that investors rely heavily on GICS peers classifications when processing relevant information to the focal firm, instead of determining closeness in a more detailed manner.

3. As an alternate measure of how investors value firms, I use valuation multiples as the dependent variable in Table A2 of the Appendix. Across many valuation multiples, I show that firms with higher *Similarity* have higher multiples, in line with investors bidding up the prices of firms with better information environments.

After establishing clear capital market effects, I turn to real effects on firm investment. Given the reduction in the cost of capital, the threshold for which a manager can green-light new NPV¿0 projects has decreased, implying the manager can now take on more projects. I therefore hypothesize that investment will be associated positively with classification quality. Beyond this first moment effect, improved information environments will further benefit investments with greater uncertainty, because the real option value of delaying the investment is higher for more uncertain investments (Dixit and Pindyck, 1994; Arrow and Fisher, 1974). This observation implies an increase in information will reduce the incentive to delay investment by more for more uncertain investments. Thus, I hypothesize the effect is greater for R&D than Capex. The results for Capex and R&D can be seen in columns (8) and (9), respectively. After including firm fixed effects, a one SD increase in classification quality is associated with a 0.2 pp (3.4% relative to mean) increase in Capex and a significantly larger 0.9 pp (8.0% relative to mean) increase in R&D. This finding offers suggestive evidence of real effects emanating from improved classification quality, especially for highly uncertain investments.[4]

Finally, given the prior evidence that sell-side analysts also rely to some extent on external GICS classifications, I turn to the effect on analyst forecast errors (Merkley et al., 2017). If analysts rely on GICS classifications and offer expertise on only within-industry variation (Boni and Womack, 2006), higher quality classifications should improve their forecasts.[5] However, this effect may be attenuated if analysts have industry-level expertise and convey

4. To further demonstrate the information processing that *intransitivity* specifically provides, I replicate the main cross-sectional results from Table 3 by including *Similarity* measures based on both the intransitive HP and transitive FIC measures. Since both measures are based on the text analysis methods in Hoberg and Phillips (2016), they similarly capture classification quality, aside from variation due to transitivity and non-overlapping sets. In Table A8, the intransitive *Similarity* measure remains significant across all capital market and real investment outcome variables, demonstrating the incremental information processing that the unconstrained, intransitive classifications provide to the capital market.

5. Huang et al. (2022) show that information sharing across analysts that cover economically linked firms improves analyst performance. If GICS can capture these economic linkages and better organize analysts to promote information sharing, analyst performance may improve as a result of higher classification quality.

separate industry-level information to the capital market (Ali et al., 2020; Kadan et al., 2012). Additionally, benchmarking pressures may not be as relevant for sell-side analysts as buy-side analysts, reducing the relative information processing fictions they face. In columns (5) and (10), I find classification quality does not covary meaningfully with AFEs, suggesting that due to industry-level expertise or reduced benchmarking frictions, classification quality is not a first-order determinant of analyst forecasts.[6]

## 6.2 Triple-Differences Results

To provide causal evidence on the relation between the outcome variables discussed above and classification quality, I exploit staggered industry reclassifications by GICS, which are used as plausibly exogenous shocks to classification quality. The GICS methodology documentation highlights 10 major reclassifications that have occurred since the service launched in 1999, usually occurring within two years of the previous reclassification (see Figure A1). However, I find five of these reclassifications largely affect firms at the industry or sub-industry level, yielding few reclassifications at the industry group level. To avoid having very few treated firms in some of the cohorts in the stacked regression design, I drop these reclassifications and proceed with the other five reclassifications.

I first demonstrate the impact that reclassifications have on the *Similarity* measure in Figure 2. Across all firms, I note an increase of 10% in classification quality, slowly rising to a 30% increase by time t+4 relative to the reclassification. For firms that begin with low classification quality, I note an incremental effect of about 10%, slowly increasing to 25% by

6. To further test the benchmarking mechanism, in Table A3 of the appendix, I split the sample on the percentage ownership from the Big 3 index fund companies (Blackrock, State Street, and Vanguard). Given the passive investment approach of index funds, I would not expect these funds to be responsive to classification quality. In line with this hypothesis, in Panel A, I find that for firms with higher index fund ownership, cost of capital and liquidity are less sensitive to classification quality. In Panel B, I identify the segment of firms that should be most sensitive to classification quality: firms with low index fund ownership and high active mutual fund ownership. I find that these firms are incrementally more sensitive to classification quality for all capital market and real outcomes.

time t+4. This is in line with reclassifications helping relatively poorly classified firms more.

I next demonstrate the reclassification results on the five key outcome variables using the two sets of coefficients in Table 4. Treat Post measures the DD coefficient for firms with above average pre-period classification quality that are reclassified, relative to matched control firms that are not reclassified with above average classification quality. The interacted DDD coefficient measures the incremental treatment effect for below average pre-period classification quality firms that are reclassified, relative to below-average matched control firms. To control for time-varying differences in underlying industry economics, I focus on the specifications with cohort-time-FIC and cohort-firm fixed effects in columns (6)-(10).

Focusing on the DD coefficients first, I show that, on average, the main reclassification effect goes in the opposite direction of the associative results. The cost of capital increases by 2.6 pp, illiquidity increases by 8%, Capex does not significantly change, and R&D decreases by 1.4 pp. The opposing findings are likely driven by a change in benchmarking that occurs during reclassification. When a firm is reclassified, it is no longer in the initial buy-side analyst's benchmark, so they may be incentivized to sell their position. There may be a delay in the new analyst picking up this position, due to a lack of familiarity with the firm. This can lead to a short-term window with higher bid-ask spreads and lower prices, creating an environment with low liquidity and high cost of capital, which can cause lower investment. I find a reduction in analyst forecast errors, which is inconsistent with the benchmarking effect causing a negative impact on the information environment. However, given sell-side analysts are less prone to benchmarking incentives, this may still be consistent with an incentive-based mechanism. Overall, the results suggest negative main effects from reclassification due to benchmarking frictions.

The incremental effect estimated in the DDD coefficients suggests that previously poorly classified firms are more positively affected than previously well classified firms. For the low classification group, liquidity improves by 6% relative to the high classification group treatment effect. For cost of capital, I find an insignificant but economically meaningful 50 bps reduction relative to the high classification group treatment effect. These results translate to a 1.5 pp increase in R&D and a 1.9 pp increase in Capex.[7] Given the economically meaningful effects on cost of capital and liquidity, the findings suggest that the effects are at least partially driven by the capital market channel. Moreover, given the insignificantly different response between R&D and Capex, there is less clear evidence of the uncertainty channel driving these results.[8]

Figure 3 shows the generalized DDD coefficients for each point in event time for R&D and Capex. For both outcome variables, I find a sharp increase in investment around the time of reclassification, although the effect seems to decay in future years. This is consistent with the uncertainty channel, as the uncertainty reduction may have a short-term "pull forward" effect on investment, as the option value of delaying investment decreases. [9]

7. In Table A4 of the appendix, I test for heterogeneous investment response by interacting *Similarity* with a proxy for growth opportunities, as in Badertscher et al. (2013). In Panel A's associative results, I find that as growth opportunities increase (decrease), firm R&D investment becomes more (less) sensitive to classification quality, in line with firms differentially responding to classification quality depending on their investment opportunities. I further find that in equilibrium, the efficiency of R&D investment, as measured by Cooper et al. (2022)'s Research Quotient (RQ) metric, is higher for firms with better classification quality, in line with the peer learning channel. In Panel B's DD tests, I find that after reclassification, firms with more growth opportunities increase R&D by more. I also find decreases in the RQ metric after reclassification, in line with an increase in R&D investment resulting in diminishing marginal R&D efficiency.

8. This may also partially be driven by peer learning effects, that is, that managers learn from new peers about optimal investment and increase investment after being reclassified (Martens and Sextroh, 2021; Bustamante and Frésard, 2021). However, this possibility cannot be explicitly tested in this specification, given the bundle of direct treatment and information spillovers. I formally test the peer learning mechanism in the "Spillover Effects" section.

9. In Table A6 in the Appendix, I show broadly consistent results using a continuous DD design as in Bloomfield (2021). In this design, I interact Treat Post with a continuous pre-treatment measure of *Similarity*. The continuous DD design allows for the identification of heterogeneous treatment effects by exploiting all variation in *Similarity*, as opposed to only exploiting the median cutoff of pre-treatment *Similarity* in the DDD design. However, because I am interacting Treat Post with a pre-treatment measure of *Similarity*, the signs should be interpreted in an inverted fashion to the associative and DDD results. For example, the negative coefficients on the interacted variable for Capex and R&D suggest that the treatment effect is *smaller* as the pre-treatment classification quality increases. This observation is consistent with the hypotheses above.
I use a DDD design in the main analysis instead of a continuous DD design for two reasons. First, as discussed above, the interpretation is more consistent with the association tests when doing a median split. Second, as mentioned in Bloomfield (2021), DDD benefit from more lenient identifying assumptions. Because a third dimension exists upon which the specification is differencing, the only assumption required is that

Overall, these results provide suggestive evidence that for well classified firms, the benchmarking incentive effect causes negative shocks to liquidity, cost of capital, and investment. The DDD differences out the benchmarking effect and shows that firms with low pre-treatment classification have improvements in cost of capital and liquidity, along with increases in investment. This suggests that for poorly classified firms, reclassifications improve information processing and provide capital market and real effects.

### 6.3 Mechanism Tests

I conduct cross-sectional sample splitting tests to test the mechanism of classification quality improving capital market and investment outcomes. I first split the sample based on whether or not a firm has analyst coverage. Analysts are information intermediaries that help reduce information asymmetry by providing information to a large segment of the market (Bradshaw et al., 2017). The presence of additional information intermediaries would likely lower the sensitivity to GICS classification, which would lower the sensitivity of the analyst-covered firms to classification quality relative to the non-analyst-covered firms. I thus hypothesize that the sensitivity to *Similarity* will be lower in the analyst group.

Table 5 Panel A provides evidence of this hypothesis. For liquidity, the sensitivity to *Similarity* is greater for firms without analyst coverage. A one SD increase in similarity is associated with an insignificant 0.4% change in liquidity for the analyst group, and a significant 1.5% change for the non-analyst group. Similarly, R&D exhibits a 1.4 pp increase for firms without analyst coverage, compared with a 0.6 pp increase for firms with analyst coverage. However, for Capex, the result goes in the opposite direction. The non-analyst group has a *negative* association with *Similarity*, whereas the analyst group has the positive

absent a GICS reclassification, the sensitivity of the outcome variable to *Similarity* is parallel for low and high pre-treatment firms. Importantly, this assumption is a weaker one than a DD design, which requires that the sensitivity of the outcome variable to *Similarity* is parallel for treated and untreated firms. Given the parallel trends must now only hold across two groups of treated firms, the identifying assumptions are more likely to hold.

association seen in the previous results.[10] Overall, the effect of classification quality is stronger for firms lacking analyst coverage. The effects attenuate for the analyst group but do not disappear. This finding suggests that although analysts may offer independent information on industry classification, they also partially rely on GICS classifications and pass through this information to their readers.

In addition to a cross-sectional split on the information environment of the capital market, I also split the sample on a firm-level characteristic of conglomeration. As evidenced by the anecdote on Amazon, Barnes & Noble, and Google in the introduction, firms that have a high degree of conglomeration can be difficult to categorize, which can lower classification quality given the transitive and non-overlapping restrictions of GICS classifications. Thus, conglomerates likely have lower classification quality on average and higher sensitivity to improvements in classification quality. Moreover, Cohen and Lou (2012) show conglomerates are complicated firms with information processing frictions that lead to mispricing, which further suggests that these firms may have greater sensitivity to classification quality.

GICS defines a conglomerate as any firm that is diversified across three or more sectors such that none of the sectors contribute the majority of revenue or earnings (Bhojraj et al., 2003). These firms are placed in the "Industrial Conglomerates" or "Multi-Sector Holdings" sub-industry groups. I construct an indicator variable equal to one if the firm is in the same industry group as either of the two conglomerate sub-industries, and 0 otherwise. I use industry groups as opposed to sub-industries, given that industry group is the level at which I measure *Similarity*. I hypothesize that the sensitivity to *Similarity* is higher for GICS-defined conglomerates.

Table 5 Panel B provides evidence consistent with this hypothesis. For all variables except R&D, the sensitivity to *Similarity* is larger for the conglomerate group. The sensitivity differences are about two times larger for Capex, four times larger for liquidity, and an order

10. ICC and AFE require analyst coverage to be calculated, so they could not be tested in this cross-sectional split.

of magnitude larger for cost of capital and AFEs. For AFEs, the sign flips from a significant *negative* relation with *Similarity* for conglomerates to a small *positive* relation for standalone firms. The large cost of capital difference relative to liquidity suggests the reduction in the assessed expected cash flow covariance between the focal firm and the rest of the market is particularly acute for conglomerates. This observation aligns with the interpretation that conglomerates are usually seen to be market-proxies and thus have high systematic risk, unless *Similarity* can help identify firm-specific information about expected cash flows to reduce this assessment.[11] R&D may go in the opposite direction, due to the greater variance in R&D expense across stand-alone firms relative to conglomerates. Overall, there exists strong evidence of conglomeration increasing information processing costs and the reliance on classification quality.

## 6.4 Information Spillover Results

Given industry classification is inherently providing peer information, information spillover effects may arise from improving classification for a focal firm. That is, a focal firm moving industries will not only improve its own classification, but may also improve the classification of the new peers it joins. As in the directly treated mechanism, if firms receive improved classifications, they should display improved liquidity and cost of capital. For real effects, given prior evidence on managerial learning through peer investment and that this learning flows through information intermediaries, firms should also display increases in investment (Bustamante and Frésard, 2021; Martens and Sextroh, 2021). Given that the reduction in uncertainty should be especially large for R&D, the increase should be larger for R&D than Capex (Bloom et al., 2007; Arif et al., 2016).

I use GICS reclassifications to test for such spillover effects. I define spillover firms as any

11. Given the vast majority of firms are stand-alone, the precision of the non-conglomerate group is higher, which causes that group to have significant coefficients while the conglomerate group has some insignificant coefficients.

firm that is not reclassified and is in an industry that receives new firms during one of the GICS reclassifications. In Figure 4, I show focal firm reclassifications increase classification quality for newly joined peers. The entire spillover sample shows a 5% improvement in classification quality relative to never-treated firms. As expected, this improvement is smaller than that of the directly treated firms, but is still significant.

Table 6 shows the DD effects for the Spillover firms. Unlike the DD main effects in Table 4, there are minimal benchmarking frictions present here, as spillover firms are not moving benchmarks themselves. Consequently, the effect is in the predicted direction of the hypotheses. Looking at columns (6)-(10), spillover firms have reductions in cost of capital of 78 bps, improvements in liquidity of 7%, and increases in Capex of 0.7 pp and R&D of 2.4 pp. As with the associative results, there is an insignificant change in AFE. Figure 5 shows the generalized DD coefficients for Capex and R&D. As in the DDD results in Figure 3, I observe increases in both Capex and R&D around the time of reclassification. The increase in R&D is significantly larger than Capex but decays through the window of time. These results are in line with reclassifications generating information spillovers for peer firms, reducing their investment uncertainty and increasing their investment, especially for R&D.

## 6.5 Placebo Test

I have shown *Similarity* leads to improved capital market outcomes, which generates real effects by increasing firm investment, due to the reduced processing costs that a higher quality classification generates. Specifically, this improvement is due to the reliance on GICS as the default classification provider in the capital market. Because agents rely on GICS, any misclassification will generate processing costs and reduce capital market efficiency.

However, the *Similarity* measure captures not only reductions in processing costs, but also other characteristics that vary with the overlap between GICS and HP classifications. For example, firms likely have latent differences in their complexity, and firms that are more complex are more difficult to classify, lowering their *Similarity* score. More complex firms being harder to understand could generate similar results to what I have shown. Although the DDD results suggest these effects are not purely driven by latent differences in firm complexity, the possibility that they are a driving factor of my results could still be a concern.[12] More generally, many industry-level characteristics could vary with *Similarity*, which brings up concerns of alternate hypotheses driving the results.

To address these concerns, I run a placebo test using SIC classifications in place of GICS classifications. That is, I reconstruct the *Similarity* measure, but instead of identifying the overlap between GICS and HP industry classifications, I calculate the overlap between SIC and HP.[13] If the results are driven by firm complexity, complex firms will similarly have low *Similarity* scores when constructed using the overlap between SIC and HP classifications.[14] If the results are driven by an industry-level characteristic, the industry classifications of SIC should identify these characteristics in the same manner as GICS. The results can be seen in Table 7, where I repeat the cross-sectional analysis of Table 3. For SIC codes, all coefficients are insignificant except AFE, which shows an economically small change of 1.7% in AFEs. Overall, the placebo test provides evidence that the previous results are not driven by confounding industry-level or firm complexity characteristics. Instead, they are likely to be driven by the information processing effects of relying on GICS as the default provider for the capital market.

12. To specifically address the complexity story, in Table A7 of the Appendix, I repeat the association tests from Table 3 but add controls for two measures of complexity from the finance and accounting literatures (Hoitash and Hoitash, 2018; Loughran and McDonald, 2020). I show that the relation between all outcome variables and *Similarity* remains significant.

13. Chen et al. (2016) find evidence that conglomerate firms strategically manipulate SIC codes in order to extract capital market benefits, by reclassifying themselves into a more favorable industry classification. To remove this endogenous mechanism from the placebo test, I drop all conglomerate firms that are most susceptible to this behavior, based on the definitions in Chen et al. (2016). Specifically, I drop all firms whose top industry segment contributes between 50-60% of the combined sales of the top two segments. The placebo results are robust to this restriction.

14. To account for possible correlations between GICS and SIC *Similarity*, in untabulated tests, I repeat the tests after residualizing GICS *Similarity* from SIC *Similarity* and vice-versa. GICS *Similarity* maintains significant results after residualization, while SIC *Similarity* maintains insignificant results.

# CHAPTER 7

# CONCLUSION

GICS is an information intermediary that the capital market relies on as its default industry classification provider. When GICS provides better industry classifications to the capital market, the information available to the market increases, reducing information asymmetry and improving liquidity and cost of capital. These capital market effects lead to real effects from a change in the cost of capital. The lower cost of capital results in increases in both Capex and R&D. R&D experiences an incremental increase, due to the greater reduction in opacity and uncertainty, which decreases firm incentives to delay investment.

These results are shown in associative tests and are further strengthened by mechanism tests that show stronger results for firms with weaker information environments and more complex structures. Causal and spillover tests that exploit staggered GICS reclassifications suggest that managers use classification to identify appropriate peers, which can reduce uncertainty and increase R&D investment. Placebo tests replacing GICS classifications with SIC - which the capital markets do not rely on - show insignificant or inconsistent results. Overall, I demonstrate the capital market and real effects emanating from the quality of GICS industry classifications.

This paper is the first to study industry classification providers as information intermediaries. I focus on GICS, the industry classification provider for the US capital markets, but future work can study other intermediaries, both within and outside of the capital market. Within the US capital market, heterogeneity in the reliance on GICS across hedge funds, mutual funds, passive funds, and retail investors could create heterogeneity in firm outcomes, depending on their shareholder composition. Outside of the US, given the heterogeneity in the global capital market, a dominant industry classification provider may not exist, which may create different frictions and effects.

Beyond the capital market, other settings have default classification systems, and the reliance on these systems can lead to unique frictions. For example, one could study the real effects from government regulators using NAICS when determining antitrust outcomes. Antitrust regulators require market definitions when determining if a merger will significantly increase concentration and market power. Imperfections in the quality of NAICS classifications can affect which mergers are allowed to pass through and which are blocked by antitrust rules. Similar frictions may exist for other government agencies, researchers, or agents that rely on outsourced industry classification. I leave the investigation of these topics to future research.

## REFERENCES


Ashiq Ali, Dan Amiram, Alon Kalay, and Gil Sadka. Industry sensitivity to external forces, and the industry-level information advantage of analysts over managers when forecasting earnings, November 2020. URL `https://papers.ssrn.com/abstract=2663611`.

Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, January 2002. ISSN 1386-4181. doi: 10.1016/S1386-4181(01)00024-6. URL `https://www.sciencedirect.com/science/article/pii/S1386418101000246`.

Yakov Amihud and Shai Levi. The Effect of Stock Liquidity on the Firm's Investment and Production. *The Review of Financial Studies*, page hhac036, June 2022. ISSN 0893-9454. doi: 10.1093/rfs/hhac036. URL `https://doi.org/10.1093/rfs/hhac036`.

Salman Arif, Nathan Marshall, and Teri Lombardi Yohn. Understanding the relation between accruals and volatility: A real options-based investment approach. *Journal of Accounting and Economics*, 62(1):65–86, August 2016. ISSN 0165-4101. doi: 10.1016/j.jacceco.2016.04.005. URL `https://www.sciencedirect.com/science/article/pii/S0165410116300337`.

Kenneth J. Arrow and Anthony C. Fisher. Environmental Preservation, Uncertainty, and Irreversibility*. *The Quarterly Journal of Economics*, 88(2):312–319, May 1974. ISSN 0033-5533. doi: 10.2307/1883074. URL `https://doi.org/10.2307/1883074`.

Brad Badertscher, Nemit Shroff, and Hal D. White. Externalities of public firm presence: Evidence from private firms' investment decisions. *Journal of Financial Economics*, 109 (3):682–706, September 2013. ISSN 0304-405X. doi: 10.1016/j.jfineco.2013.03.012. URL `https://www.sciencedirect.com/science/article/pii/S0304405X13000925`.

Andrew C. Baker, David F. Larcker, and Charles C. Y. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2): 370–395, May 2022. ISSN 0304-405X. doi: 10.1016/j.jfineco.2022.01.004. URL `https://www.sciencedirect.com/science/article/pii/S0304405X22000204`.

Ray Ball, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri Nikolaev. Deflating profitability. *Journal of Financial Economics*, 117(2):225–248, August 2015. ISSN 0304405X. doi: 10.1016/j.jfineco.2015.02.004. URL `https://linkinghub.elsevier.com/retrieve/pii/S0304405X15000203`.

Ray Ball, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri Nikolaev. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, 121(1):28–45, July 2016. ISSN 0304405X. doi: 10.1016/j.jfineco.2016.03.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0304405X16300307`.

John Manuel Barrios. Staggeringly Problematic: A Primer on Staggered DiD for Accounting Researchers, February 2021. URL `https://papers.ssrn.com/abstract=3794859`.

Tobias Berg, Markus Reisinger, and Daniel Streitz. Spillover effects in empirical corporate finance. *Journal of Financial Economics*, 142(3):1109–1127, December 2021. ISSN 0304405X. doi: 10.1016/j.jfineco.2021.04.039. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X21001690.

Hendrik Bessembinder. Do stocks outperform Treasury bills? *Journal of Financial Economics*, 129(3):440–457, September 2018. ISSN 0304-405X. doi: 10.1016/j.jfineco.2018.06.004. URL https://www.sciencedirect.com/science/article/pii/S0304405X18301521.

Sanjeev Bhojraj, Charles M. C. Lee, and Derek K. Oler. What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research*, 41(5):745–774, 2003. ISSN 1475-679X. doi: 10.1046/j.1475-679X.2003.00122.x. URL http://onlinelibrary.wiley.com/doi/abs/10.1046/j.1475-679X.2003.00122.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1475-679X.2003.00122.x.

Elizabeth Blankespoor, Ed deHaan, and Iván Marinovic. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, 70(2):101344, November 2020. ISSN 0165-4101. doi: 10.1016/j.jacceco.2020.101344. URL https://www.sciencedirect.com/science/article/pii/S016541012030046X.

Nick Bloom, Stephen Bond, and John Van Reenen. Uncertainty and Investment Dynamics. *The Review of Economic Studies*, 74(2):391–415, April 2007. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2007.00426.x. URL https://doi.org/10.1111/j.1467-937X.2007.00426.x.

Matthew J. Bloomfield. Compensation disclosures and strategic commitment: Evidence from revenue-based pay. *Journal of Financial Economics*, 141(2):620–643, August 2021. ISSN 0304405X. doi: 10.1016/j.jfineco.2021.04.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X21001331.

Leslie Boni and Kent L. Womack. Analysts, Industries, and Price Momentum. *Journal of Financial and Quantitative Analysis*, 41(1):85–109, March 2006. ISSN 1756-6916, 0022-1090. doi: 10.1017/S002210900000243X. URL http://www.cambridge.org/core/journals/journal-of-financial-and-quantitative-analysis/article/analysts-industries-and-price-momentum/9DB760B3F0392A1943C1A673A173AFE7. Publisher: Cambridge University Press.

Daniel Bradley, Sinan Gokkaya, and Xi Liu. Before an Analyst Becomes an Analyst: Does Industry Experience Matter? *The Journal of Finance*, 72(2):751–792, 2017. ISSN 1540-6261. doi: 10.1111/jofi.12466. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12466. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12466.

Mark Bradshaw, Yonca Ertimur, and Patricia O'Brien. Financial Analysts and Their Contribution to Well-Functioning Capital Markets. *Foundations and Trends® in Accounting*, 11 (3):119–191, December 2017. ISSN 1554-0642, 1554-0650. doi: 10.1561/1400000042. URL http://www.nowpublishers.com/article/Details/ACC-042. Publisher: Now Publishers, Inc.

M. Cecilia Bustamante and Laurent Frésard. Does Firm Investment Respond to Peers' Investment? *Management Science*, 67(8):4703–4724, August 2021. ISSN 0025-1909. doi: 10.1287/mnsc.2020.3695. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2020.3695. Publisher: INFORMS.

Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics*, 134(3):1405–1454, August 2019. ISSN 0033-5533. doi: 10.1093/qje/qjz014. URL https://doi.org/10.1093/qje/qjz014.

Wendy Chan. Methodology of GICS. page 10, October 2021.

Huaizhi Chen, Lauren Cohen, and Dong Lou. Industry Window Dressing. *Review of Financial Studies*, 29(12):3354–3393, December 2016. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhw020. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhw020.

Hans B. Christensen, Luzi Hail, and Christian Leuz. Mandatory IFRS reporting and changes in enforcement. *Journal of Accounting and Economics*, 56(2, Supplement 1):147–177, December 2013. ISSN 0165-4101. doi: 10.1016/j.jacceco.2013.10.007. URL https://www.sciencedirect.com/science/article/pii/S0165410113000645.

Hans B. Christensen, Luzi Hail, and Christian Leuz. Capital-Market Effects of Securities Regulation: Prior Conditions, Implementation, and Enforcement. *The Review of Financial Studies*, 29(11):2885–2924, November 2016. ISSN 0893-9454. doi: 10.1093/rfs/hhw055. URL https://doi.org/10.1093/rfs/hhw055.

James Claus and Jacob Thomas. Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets. *The Journal of Finance*, 56(5):1629–1666, 2001. ISSN 1540-6261. doi: 10.1111/0022-1082.00384. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.00384. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00384.

Lauren Cohen and Dong Lou. Complicated firms. *Journal of Financial Economics*, 104(2): 383–400, May 2012. ISSN 0304405X. doi: 10.1016/j.jfineco.2011.08.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X11001899.

Michael Cooper, Anne Marie Knott, and Wenhao Yang. RQ Innovative Efficiency and Firm Value. *Journal of Financial and Quantitative Analysis*, 57 (5):1649–1694, August 2022. ISSN 0022-1090, 1756-6916. doi: 10.1017/

S0022109021000417. URL https://www.cambridge.org/core/product/identifier/S0022109021000417/type/journal_article.

Shane A. Corwin and Paul Schultz. A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, 67(2):719–760, 2012. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2012.01729.x. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2012.01729.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2012.01729.x.

K. J. Martijn Cremers and Antti Petajisto. How Active Is Your Fund Manager? A New Measure That Predicts Performance. *The Review of Financial Studies*, 22(9):3329–3365, September 2009. ISSN 0893-9454. doi: 10.1093/rfs/hhp057. URL https://doi.org/10.1093/rfs/hhp057.

Douglas W. Diamond and Robert E. Verrecchia. Disclosure, Liquidity, and the Cost of Capital. *The Journal of Finance*, 46(4):1325–1359, 1991. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1991.tb04620.x. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1991.tb04620.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1991.tb04620.x.

Avinash K. Dixit and Robert S. Pindyck. *Investment under Uncertainty*. Princeton University Press, Princeton, N.J, 1st edition edition, January 1994. ISBN 978-0-691-03410-2.

Peter D. Easton. PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review*, 79(1):73–95, January 2004. ISSN 0001-4826. doi: 10.2308/accr.2004.79.1.73. URL https://doi.org/10.2308/accr.2004.79.1.73.

Joseph Engelberg, Arzu Ozoguz, and Sean Wang. Know Thy Neighbor: Industry Clusters, Information Spillovers, and Market Efficiency. *Journal of Financial and Quantitative Analysis*, 53(5):1937–1961, October 2018. ISSN 0022-1090, 1756-6916. doi: 10.1017/S0022109018000261. URL https://www.cambridge.org/core/product/identifier/S0022109018000261/type/journal_article.

Eugene F. Fama and Kenneth R. French. Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193, February 1997. ISSN 0304-405X. doi: 10.1016/S0304-405X(96)00896-3. URL https://www.sciencedirect.com/science/article/pii/S0304405X96008963.

Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, April 2015. ISSN 0304405X. doi: 10.1016/j.jfineco.2014.10.010. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X14002323.

Kingsley Y L Fong, Craig W Holden, and Charles A Trzcinka. What Are the Best Liquidity Proxies for Global Research?*. *Review of Finance*, 21(4):1355–1401, July 2017. ISSN 1572-3097. doi: 10.1093/rof/rfx003. URL https://doi.org/10.1093/rof/rfx003.

Richard Frankel and Charles M. C. Lee. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, 25(3):283–319, June 1998. ISSN 0165-4101. doi: 10.1016/S0165-4101(98)00026-3. URL https://www.sciencedirect.com/science/article/pii/S0165410198000263.

William R. Gebhardt, Charles M. C. Lee, and Bhaskaran Swaminathan. Toward an Implied Cost of Capital. *Journal of Accounting Research*, 39(1):135–176, 2001. ISSN 1475-679X. doi: 10.1111/1475-679X.00007. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/1475-679X.00007. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-679X.00007.

John R. Graham. Presidential Address: Corporate Finance and Reality. *The Journal of Finance*, 77(4):1975–2049, 2022. ISSN 1540-6261. doi: 10.1111/jofi.13161. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13161. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13161.

João Granja. Disclosure Regulation in the Commercial Banking Industry: Lessons from the National Banking Era. *Journal of Accounting Research*, 56(1):173–216, March 2018. ISSN 00218456. doi: 10.1111/1475-679X.12193. URL https://onlinelibrary.wiley.com/doi/10.1111/1475-679X.12193.

Gerard Hoberg and Gordon Phillips. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies*, 23(10):3773–3811, October 2010. ISSN 0893-9454. doi: 10.1093/rfs/hhq053. URL https://doi.org/10.1093/rfs/hhq053.

Gerard Hoberg and Gordon Phillips. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465, October 2016. ISSN 0022-3808. doi: 10.1086/688176. URL https://www.journals.uchicago.edu/doi/abs/10.1086/688176. Publisher: The University of Chicago Press.

Rani Hoitash and Udi Hoitash. Measuring Accounting Reporting Complexity with XBRL. *Accounting Review*, 93(1):259–287, January 2018. ISSN 00014826. doi: 10.2308/accr-51762. URL http://proxy.uchicago.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=bah&AN=127415668&site=ehost-live&scope=site. Publisher: American Accounting Association.

Herbert Hovenkamp. Antitrust Error Costs. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3853282. URL https://www.ssrn.com/abstract=3853282.

Allen H. Huang, An-Ping Lin, and Amy Y. Zang. Cross-industry information sharing among colleagues and analyst research. *Journal of Accounting and Economics*, 74(1):101496, August 2022. ISSN 0165-4101. doi: 10.1016/j.jacceco.2022.101496. URL https://www.sciencedirect.com/science/article/pii/S0165410122000192.

Ravi Jagannathan, David A. Matsa, Iwan Meier, and Vefa Tarhan. Why do firms use high discount rates? *Journal of Financial Economics*, 120(3):445–463, June 2016. ISSN

0304-405X. doi: 10.1016/j.jfineco.2016.01.012. URL https://www.sciencedirect.com/science/article/pii/S0304405X16000179.

Ohad Kadan, Leonardo Madureira, Rong Wang, and Tzachi Zach. Analysts' industry expertise. *Journal of Accounting and Economics*, 54(2):95–120, October 2012. ISSN 0165-4101. doi: 10.1016/j.jacceco.2012.05.002. URL https://www.sciencedirect.com/science/article/pii/S0165410112000419.

Anil K Kashyap, Natalia Kovrijnykh, Jian Li, and Anna Pavlova. The benchmark inclusion subsidy. *Journal of Financial Economics*, 142(2):756–774, November 2021. ISSN 0304405X. doi: 10.1016/j.jfineco.2021.04.021. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X21001513.

Markku Kaustia and Ville Rantala. Common Analysts: Method for Defining Peer Firms. *Journal of Financial and Quantitative Analysis*, 56(5):1505–1536, August 2021. ISSN 0022-1090, 1756-6916. doi: 10.1017/S0022109020000514. URL http://www.cambridge.org/core/journals/journal-of-financial-and-quantitative-analysis/article/common-analysts-method-for-defining-peer-firms/C104EC5F6422620532C9753FA21C2E5A. Publisher: Cambridge University Press.

Ping-Sheng Koh and David M. Reeb. Missing R&D. *Journal of Accounting and Economics*, 60(1):73–94, August 2015. ISSN 0165-4101. doi: 10.1016/j.jacceco.2015.03.004. URL https://www.sciencedirect.com/science/article/pii/S0165410115000324.

Albert S. Kyle. Informed Speculation with Imperfect Competition. *The Review of Economic Studies*, 56(3):317–355, 1989. ISSN 0034-6527. doi: 10.2307/2297551. URL https://www.jstor.org/stable/2297551. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Richard Lambert, Christian Leuz, and Robert E. Verrecchia. Accounting Information, Disclosure, and the Cost of Capital. *Journal of Accounting Research*, 45(2):385–420, 2007. ISSN 1475-679X. doi: 10.1111/j.1475-679X.2007.00238.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-679X.2007.00238.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-679X.2007.00238.x.

Stephannie Larocque, Alastair Lawrence, and Kevin Veenstra. Managers' Cost of Equity Capital Estimates: Empirical Evidence. *Journal of Accounting, Auditing & Finance*, 33(3):382–401, July 2018. ISSN 0148-558X. doi: 10.1177/0148558X16654034. URL https://doi.org/10.1177/0148558X16654034. Publisher: SAGE Publications Inc.

Charles M. C. Lee, Paul Ma, and Charles C. Y. Wang. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, 116(2):410–431, May 2015. ISSN 0304-405X. doi: 10.1016/j.jfineco.2015.02.003. URL https://www.sciencedirect.com/science/article/pii/S0304405X15000197.

David A. Lesmond, Joseph P. Ogden, and Charles A. Trzcinka. A New Estimate of Transaction Costs. *The Review of Financial Studies*, 12(5):1113–1141, October 1999. ISSN 0893-9454. doi: 10.1093/rfs/12.5.1113. URL https://doi.org/10.1093/rfs/12.5.1113.

Feng Li, Russell Lundholm, and Michael Minnis. A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research*, 51(2):399–436, 2013. ISSN 1475-679X. doi: 10.1111/j.1475-679X.2012.00472.x. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-679X.2012.00472.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-679X.2012.00472.x.

Tim Loughran and Bill McDonald. Measuring Firm Complexity. *SSRN Electronic Journal*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3645372. URL https://www.ssrn.com/abstract=3645372.

Tim Martens and Christoph J. Sextroh. Analyst Coverage Overlaps and Interfirm Information Spillovers. *Journal of Accounting Research*, 59(4):1425–1480, 2021. ISSN 1475-679X. doi: 10.1111/1475-679X.12391. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/1475-679X.12391. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-679X.12391.

Kenneth Merkley, Roni Michaely, and Joseph Pacelli. Does the Scope of the Sell-Side Analyst Industry Matter? An Examination of Bias, Accuracy, and Information Content of Analyst Reports. *The Journal of Finance*, 72(3):1285–1334, 2017. ISSN 1540-6261. doi: 10.1111/jofi.12485. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12485. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12485.

Robert C. Merton. A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, 42(3):483–510, 1987. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1987.tb04565.x. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1987.tb04565.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1987.tb04565.x.

MSCI. Frequently Asked Questions about GICS, 2005. URL https://www.msci.com/documents/10199/5973a128-47f0-4317-b083-716a10207b50.

James A. Ohlson and Beate E. Juettner-Nauroth. Expected EPS and EPS Growth as Determinantsof Value. *Review of Accounting Studies*, 10(2):349–365, September 2005. ISSN 1573-7136. doi: 10.1007/s11142-005-1535-3. URL https://doi.org/10.1007/s11142-005-1535-3.

Andreas Olden and Jarle Møen. The triple difference estimator. *The Econometrics Journal*, 25(3):531–553, September 2022. ISSN 1368-4221. doi: 10.1093/ectj/utac010. URL https://doi.org/10.1093/ectj/utac010.

Patrick O'Shaughnessy. Invest Like the Best with Patrick O'Shaughnessy: Dennis Lynch - Delivering Alpha in Adapting Markets on Apple Podcasts, May 2021. URL https://podcasts.apple.com/us/podcast/

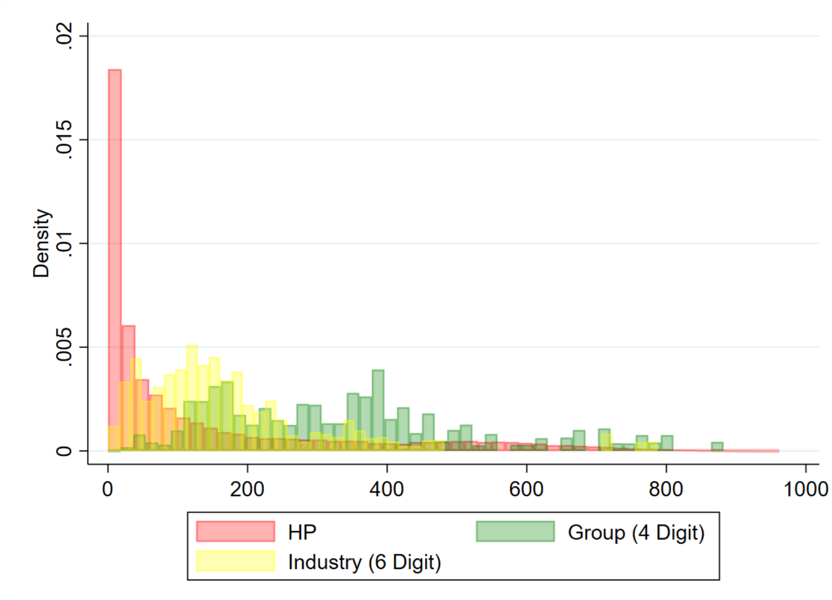
dennis-lynch-delivering-alpha-in-adapting-markets/id1154105909?i=1000523756181.

Antti Petajisto. Active Share and Mutual Fund Performance. *Financial Analysts Journal*, 69(4):73–93, July 2013. ISSN 0015-198X. doi: 10.2469/faj.v69.n4.7. URL https://doi.org/10.2469/faj.v69.n4.7. Publisher: Routledge _eprint: https://doi.org/10.2469/faj.v69.n4.7.

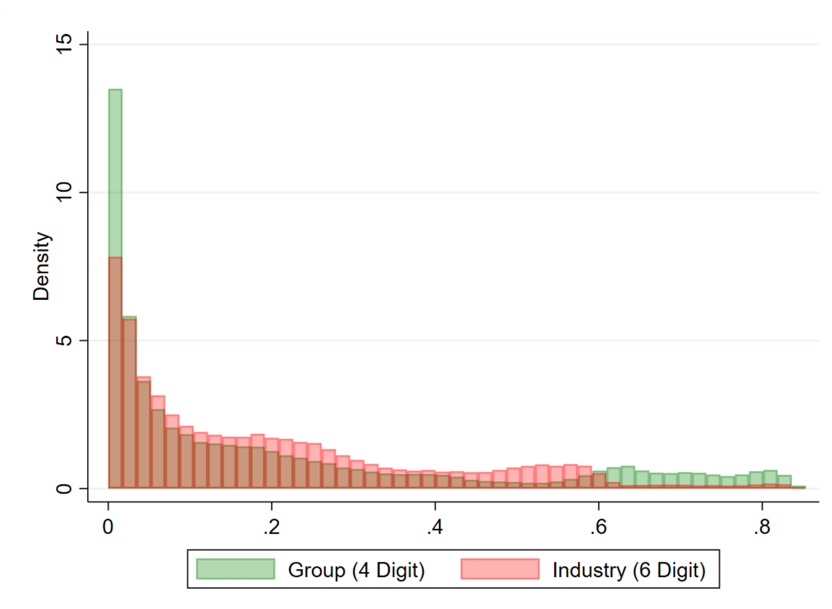
Jeremy C. Stein. Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior*. *The Quarterly Journal of Economics*, 104(4):655–669, November 1989. ISSN 0033-5533. doi: 10.2307/2937861. URL https://doi.org/10.2307/2937861.

Figure 1: Classification Quality Summary Statistics

A: Distribution of Classification Group Size



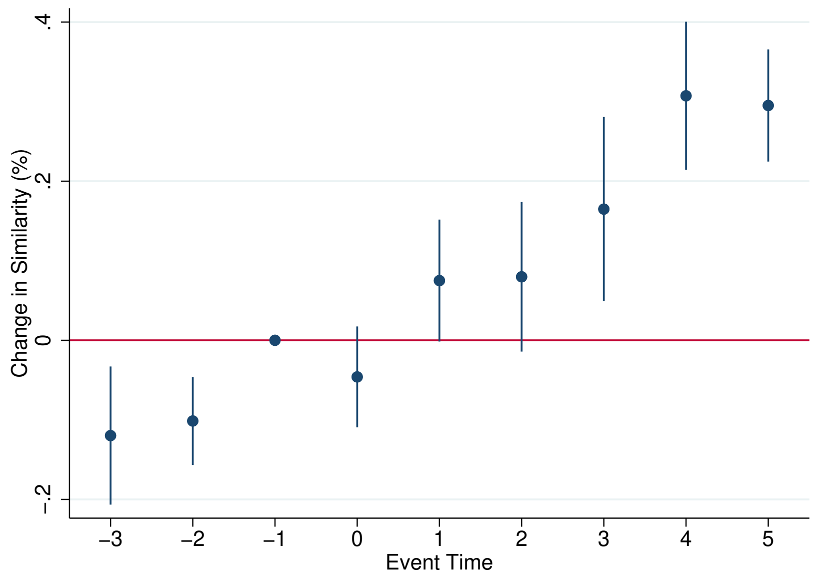

B: Distribution of Classification Quality



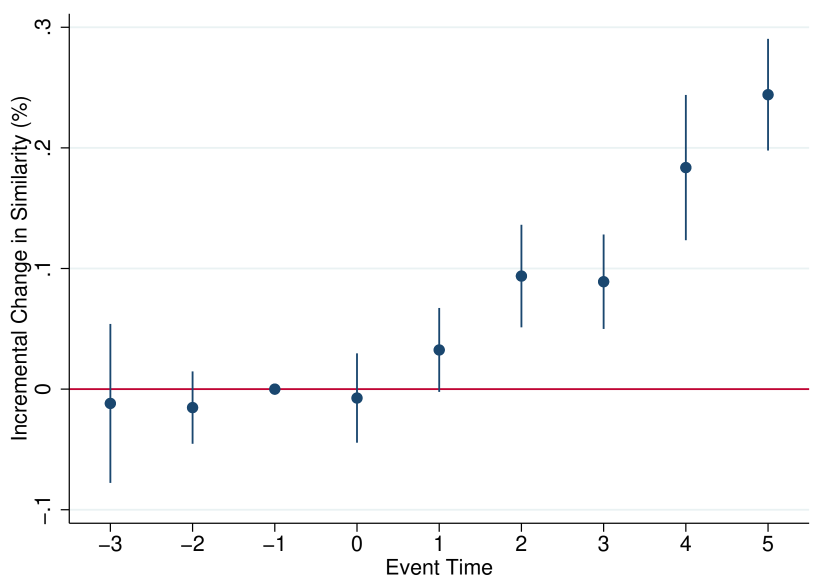

*Notes:* Figure 1A shows histograms of the size of the classification groups for Hoberg Phillips, GICS 4-digit Group Level, and GICS 6-digit Industry Level. Figure 1B shows histograms of the Classification Quality measure for the GICS 4-digit Group Level and GICS 6-digit Industry Level. Classification Quality measurement explained in Section 3.1.

Figure 2: The Effect of Reclassifications on Classification Quality

A: Main Effect




B: Incremental Effect for Low Classification Quality Firms

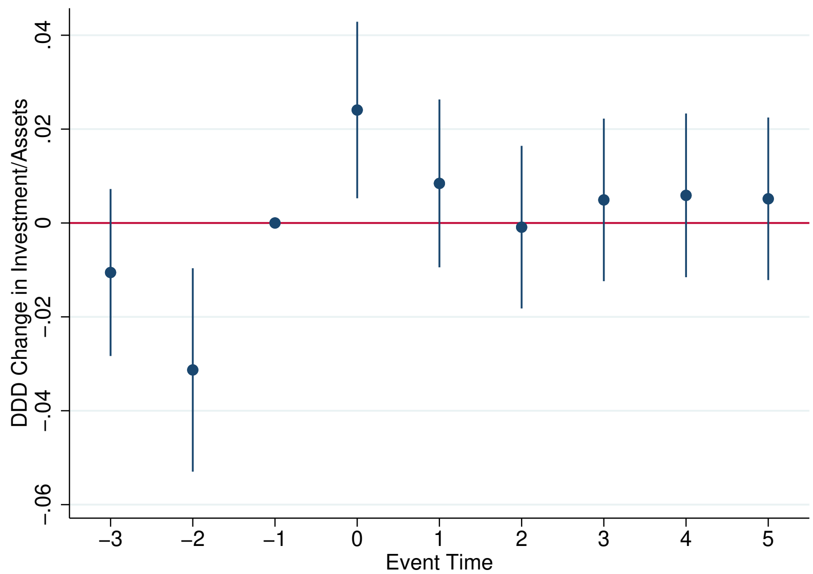



*Notes:* Figure 2A shows the $\beta_1$ coefficients, representing the main effect of GICS reclassifications on classification quality (*Similarity*). Figure 2B shows the $\beta_3$ coefficients, representing the incremental effect for the low pre-period *Simliarity* group of GICS reclassifications on classification quality (*Similarity*). The specification is as follows:

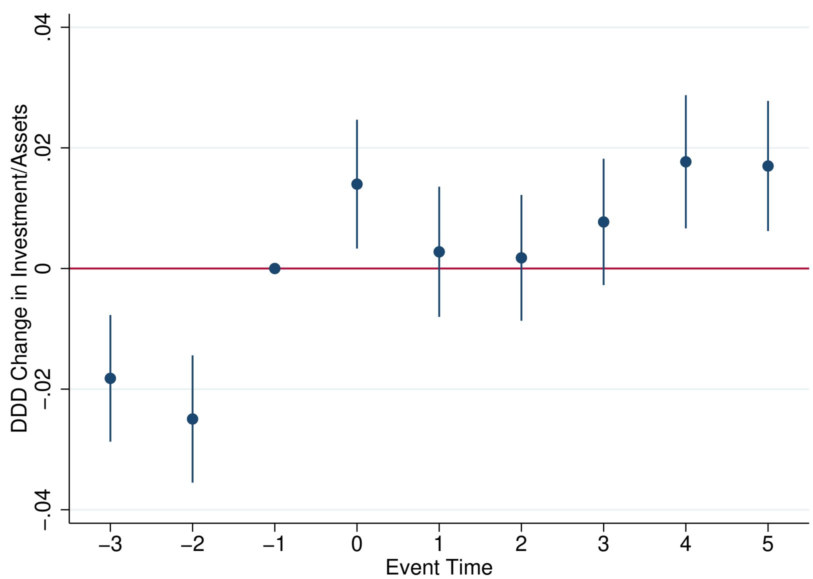
$$\begin{aligned} Y_{ijkt} =& \beta_1 \text{Treat} \times \text{Time}_{it} + \beta_2 \text{Low Pre-Period Similarity}_i \\ &+ \beta_3 \text{Low Pre-Period Similarity}_i \times \text{Treat} \times \text{Time}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{kt} + \epsilon_{ijkt} \end{aligned}$$

Figure 3: Real Effects of Reclassifications for Low Classification Quality Firms

A: Effect on R&D



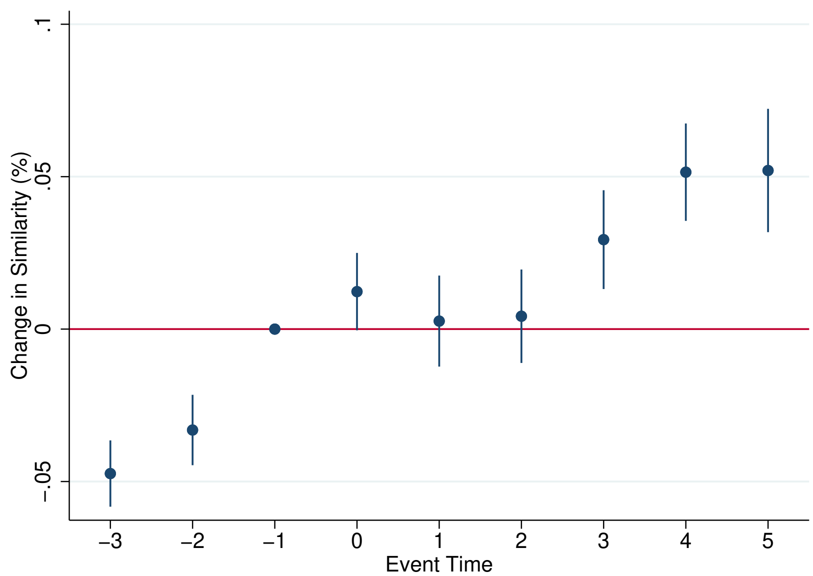

B: Effect on Capex




*Notes:* Figure 3 shows the DDD $\beta_3$ coefficients around reclassifications for treated firms with below median pre-period classification quality, relative to treated firms with above median pre-period classification quality. Figure 3A shows the change in R&D / Assets, while Figure 3B shows the change in Capex / Assets. The specification is as follows:

$$\begin{aligned} Y_{ijkt} =& \beta_1 \text{Treat} \times \text{Time}_{it} + \beta_2 \text{Low Pre-Period Similarity}_i \\ &+ \beta_3 \text{Low Pre-Period Similarity}_i \times \text{Treat} \times \text{Time}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{jkt} + \epsilon_{ijkt} \end{aligned}$$

Figure 4: Spillover Effect of Reclassifications on Classification Quality

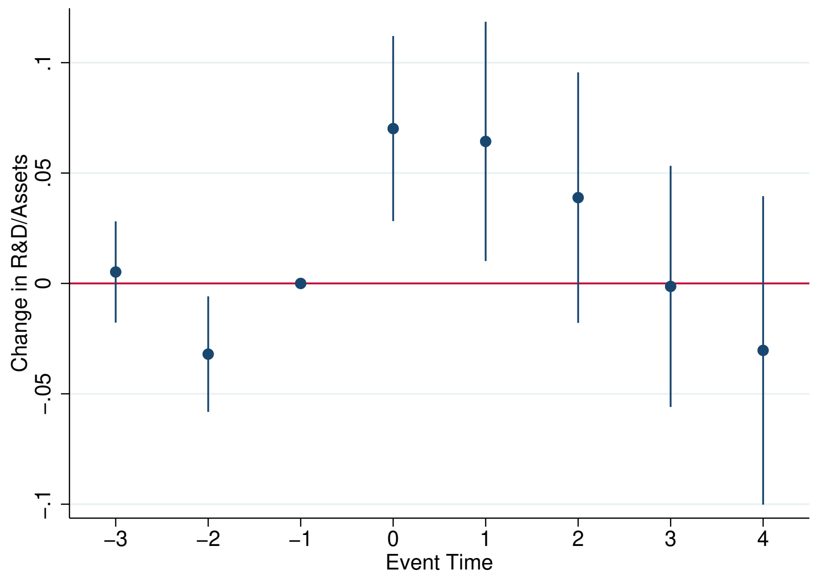



*Notes:* Figure 4 shows the difference-in-differences $\beta_1$ coefficients for the full sample of spillover firms (firms that were in GICS groups that received new peers) around staggered GICS reclassifications, relative to matched never-treated firms. The specification is as follows:

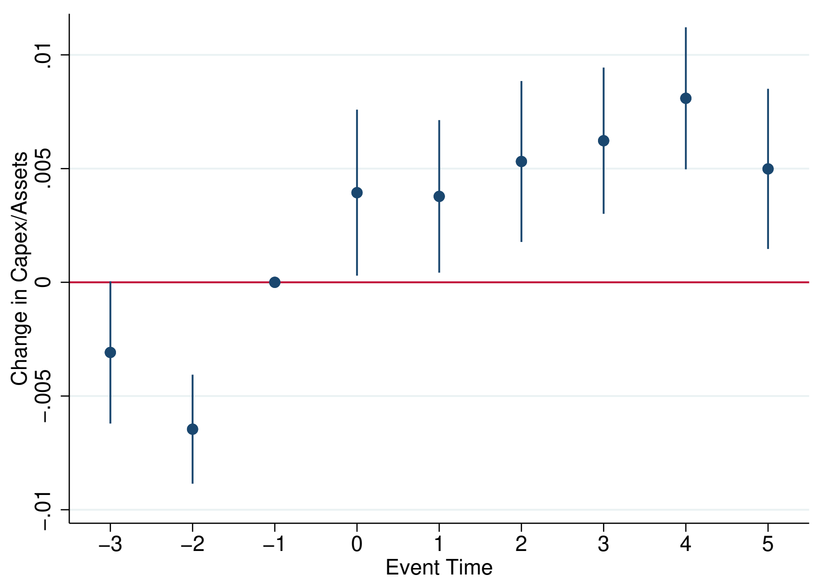
$$Y_{ijkt} = \beta_1 \text{Treat} \times \text{Time}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{kt} + \epsilon_{ijkt}$$

Figure 5: Spillover Real Effects of Reclassification

A: Effect on R&D




B: Effect on Capex




*Notes:* This shows the difference-in-differences coefficients around reclassifications for spillover firms, relative to matched never-treated firms. Figure 5A shows the change in R&D / Assets, while Figure 5B shows the change in Capex / Assets. The specification is as follows:

$$Y_{ijkt} = \beta_1 \text{Treat} \times \text{Time}_{it} + \gamma X_{it} + \nu_{ik} + \eta_{jkt} + \epsilon_{ijkt}$$

Table 1: Sample Statistics

Panel A: Summary Statistics

| | Mean | SD | P10 | Med | P90 | N |
|---|---|---|---|---|---|---|
| **Control** | | | | | | |
| Similarity | 0.133 | 0.175 | 0.005 | 0.050 | 0.406 | 673081 |
| $\beta$ | 0.013 | 0.009 | 0.003 | 0.012 | 0.024 | 673,081 |
| LN(BTM) | -7.715 | 1.097 | -9.072 | -7.655 | -6.434 | 673,081 |
| OP/AT | 0.029 | 0.274 | -0.262 | 0.089 | 0.244 | 673,081 |
| INV | 1.132 | 0.459 | 0.808 | 1.046 | 1.464 | 673,081 |
| LN(MKTEQ) | 12.902 | 2.190 | 10.009 | 12.897 | 15.801 | 673,081 |
| ICC | 14.832 | 12.749 | 5.196 | 11.851 | 26.896 | 402,156 |
| ln(Liq) | -0.922 | 1.009 | -2.067 | -0.928 | 0.318 | 642,325 |
| Capex | 0.059 | 0.101 | 0.006 | 0.032 | 0.128 | 673,081 |
| R&D | 0.114 | 0.308 | 0.000 | 0.038 | 0.294 | 451,522 |
| AFE | 0.666 | 14.510 | 0.030 | 0.250 | 1.210 | 509,285 |
| **Treatment** | | | | | | |
| Similarity | 0.132 | 0.128 | 0.007 | 0.085 | 0.341 | 66,371 |
| $\beta$ | 0.015 | 0.009 | 0.005 | 0.014 | 0.027 | 66,371 |
| LN(BTM) | -7.784 | 1.008 | -9.031 | -7.765 | -6.585 | 66,371 |
| OP/AT | 0.086 | 0.167 | -0.063 | 0.091 | 0.256 | 66,371 |
| INV | 1.129 | 0.436 | 0.832 | 1.045 | 1.435 | 66,371 |
| LN(MKTEQ) | 13.531 | 2.091 | 10.749 | 13.552 | 16.292 | 66,371 |
| ICC | 14.185 | 11.249 | 5.717 | 11.487 | 25.746 | 44,727 |
| ln(Liq) | -1.102 | 0.936 | -2.136 | -1.093 | 0.028 | 63,161 |
| Capex | 0.047 | 0.076 | 0.006 | 0.028 | 0.105 | 66,371 |
| R&D | 0.107 | 0.106 | 0.001 | 0.092 | 0.213 | 42,611 |
| AFE | 0.552 | 1.417 | 0.030 | 0.250 | 1.200 | 54,770 |

Panel B: Treatment Selection

| | Treated | | Spillover | |
|---|---|---|---|---|
| | Obs | Firms | Obs | Firms |
| 2003 | 2794 | 206 | 29519 | 2374 |
| 2005 | 703 | 53 | 5861 | 465 |
| 2006 | 442 | 34 | 5237 | 443 |
| 2016 | 372 | 36 | 3171 | 398 |
| 2018 | 1604 | 132 | 958 | 141 |
| Control | 64367 | 8324 | 25536 | 4961 |

*Notes:* Panel A. reports sample summary statistics at the firm-month level from July 2000 to June 2021. All variables are defined in the Appendix. Panel B reports observations at the firm-year level and unique firms that have been reclassified or indirectly affected by a reclassification. Treated represents firms that have been reclassified. Left (Join) Spillover represents all firms who did not get reclassified, but are in the industry group that the treated firms are departing (joining).

Table 2: Validating HP TNIC Benchmark Classification

| | OP/Sales | OP/Assets | Sales Growth | Market $\beta$ | Asset $\beta$ |
|---|---|---|---|---|---|
| GICS Industry Avg | 0.233 | 0.157 | 0.0463 | 0.232 | 0.263 |
| GICS FE | 0.216 | 0.143 | 0.0216 | 0.113 | 0.16 |
| SIC Industry Avg | 0.267 | 0.0968 | 0.0505 | 0.169 | 0.187 |
| SIC FE | 0.246 | 0.0798 | 0.0218 | 0.0794 | 0.113 |
| NAICS Industry Avg | 0.289 | 0.0866 | 0.0399 | 0.125 | 0.134 |
| NAICS FE | 0.269 | 0.0716 | 0.0182 | 0.0603 | 0.0784 |
| **TNIC Industry Avg** | **0.444** | **0.307** | **0.184** | **0.391** | **0.42** |
| FIC FE | 0.191 | 0.138 | 0.021 | 0.101 | 0.149 |

*Notes:* This table reports adjusted R-squared values from regressions of firm-level characteristics on industry fixed effects or industry-year averages, as in Table 3 of Hoberg and Phillips (2016). GICS is classified at the group level. SIC and NAICS are classified at the 2 digit level. FIC is the fixed industrial classification system from Hoberg and Phillips (2016); unlike TNIC, it is intransitive and non-overlapping, allowing for a fixed effects regression structure. The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021.

Table 3: Cross-Sectional Results

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Similarity | -0.958*** | -0.035*** | 0.014*** | 0.056*** | -0.001 | -0.108* | -0.014*** | 0.002*** | 0.009*** | 0.011** |
| | (-26.98) | (-18.41) | (32.28) | (50.34) | (-0.17) | (-1.89) | (-3.83) | (3.72) | (5.55) | (2.21) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 446699 | 705330 | 739296 | 493820 | 554229 | 446586 | 705245 | 739211 | 493768 | 554128 |
| $R^2$ | 0.303 | 0.426 | 0.209 | 0.167 | 0.111 | 0.533 | 0.528 | 0.531 | 0.437 | 0.356 |

*Notes:* This table reports coefficient estimates from association tests of classification quality on a variety of outcome variables. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table 4: Triple-Differences Results

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Treat Post | 0.530 | 0.111*** | 0.008** | -0.038*** | -0.290*** | 2.623*** | 0.083* | 0.003 | -0.014*** | -0.444*** |
| | (0.67) | (2.88) | (2.23) | (-4.94) | (-5.15) | (2.62) | (1.73) | (0.62) | (-3.26) | (-6.05) |
| Treat Post × Low Pre-Treat Similarity | 0.083 | -0.125*** | 0.016*** | 0.010* | 0.173*** | -0.498 | -0.063** | 0.019*** | 0.015*** | 0.130*** |
| | (0.19) | (-3.81) | (4.83) | (1.89) | (3.39) | (-1.16) | (-2.03) | (5.43) | (3.07) | (2.86) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Cohort-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes | Yes | No | No | No | No | No |
| Cohort-FIC-Time FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Cohort-Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 44532 | 72081 | 76016 | 43101 | 55699 | 42789 | 70448 | 74437 | 41360 | 54190 |
| $R^2$ | 0.616 | 0.527 | 0.582 | 0.385 | 0.447 | 0.681 | 0.588 | 0.684 | 0.673 | 0.536 |

*Notes:* This table reports coefficient estimates from DDD estimates from staggered GICS reclassifications. The coefficient of interest is the interacted variable of Pre-Period Similarity × Treat Post. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the cohort-month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table 5: Splitting on Information Processing Variables

Panel A: Analyst Coverage

| | No Analyst Coverage | | | Analyst Coverage | | |
|---|---|---|---|---|---|---|
| | (1) ln(Liq) $\beta$ / t | (2) Capex $\beta$ / t | (3) R&D $\beta$ / t | (4) ln(Liq) $\beta$ / t | (5) Capex $\beta$ / t | (6) R&D $\beta$ / t |
| Similarity | -0.015* | -0.004** | 0.014** | -0.004 | 0.002*** | 0.006*** |
| | (-1.96) | (-2.54) | (2.10) | (-1.05) | (3.52) | (3.49) |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 159979 | 163321 | 104714 | 544015 | 574598 | 387654 |
| $R^2$ | 0.582 | 0.462 | 0.400 | 0.405 | 0.591 | 0.539 |

Panel B: Conglomerate

| | Conglomerate | | | | | Standalone | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
| Similarity | -2.017*** | -0.048 | 0.004** | 0.005 | -0.369*** | -0.118** | -0.012*** | 0.002*** | 0.009*** | 0.019*** |
| | (-3.90) | (-1.10) | (2.52) | (1.51) | (-3.58) | (-2.04) | (-3.40) | (3.62) | (5.48) | (3.60) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 50006 | 79738 | 83798 | 59999 | 62450 | 393589 | 620277 | 650041 | 433152 | 488278 |
| $R^2$ | 0.553 | 0.545 | 0.593 | 0.734 | 0.291 | 0.533 | 0.527 | 0.532 | 0.429 | 0.363 |

*Notes:* In Panel A, columns 1-3 (4-6) contain firms with no analyst coverage (analyst coverage). ICC and AFE are not in Panel A as they require analyst coverage to compute. Analyst coverage defined as appearance in IBES Quarterly file. In Panel B, columns 1-5 (6-10) columns contain conglomerate (stand-alone) firms as defined by GICS. Conglomerate firms are defined as firms with business spread across three or more sectors in which no single sector contributes the majority of revenue or profits. They are classified in the sub-industries "Industrial Conglomerates" or "Multi-Sector Holdings". Given *Similarity* is calculated at the 4-digit group level, I define conglomerate as any firm in the same 4-digit industry group as the two conglomerate sub-industries. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table 6: Spillover Difference-in-Differences Results

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Treat Post | -0.391* | -0.040*** | 0.008*** | -0.042* | 0.043** | -0.782*** | -0.072*** | 0.007*** | 0.027 | -0.036 |
| | (-1.84) | (-3.10) | (8.52) | (-1.85) | (2.02) | (-3.04) | (-4.06) | (6.16) | (1.02) | (-1.21) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Cohort-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes | Yes | No | No | No | No | No |
| Cohort-FIC-Time FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Cohort-Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 194097 | 323754 | 338944 | 228809 | 244731 | 192455 | 322388 | 337591 | 227034 | 243232 |
| $R^2$ | 0.572 | 0.535 | 0.557 | 0.339 | 0.369 | 0.604 | 0.558 | 0.598 | 0.480 | 0.413 |

*Notes:* This table reports coefficient estimates from spillover DD estimates from staggered GICS reclassifications. The coefficient of interest is Treat Post. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the cohort-month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table 7: Placebo Test with SIC Industry Classification System

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|
| Similarity | -0.074 | -0.000 | 0.000 | 0.003 | -0.022*** |
| | (-1.19) | (-0.02) | (0.65) | (0.41) | (-3.63) |
| FF Ctrl | Yes | Yes | No | No | No |
| Firm FE | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No |
| Observations | 413187 | 647877 | 678866 | 455797 | 510328 |
| $R^2$ | 0.548 | 0.559 | 0.557 | 0.450 | 0.377 |

*Notes:* This table repeats the cross-sectional analysis from Table 3 with measures of classification quality for SIC. The methodology is identical to that of GICS classification quality, replacing 4-digit GICS classifications with 2-digit SIC codes. To account for Chen et al. (2016)'s findings of strategic manipulation by conglomerates with similar sized segments, I drop all firms whose top industry segment contributes between 50-60% of the combined sales of the top two segments. Industry is defined by 2-digit SIC codes, as in Chen et al. (2016). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

# Appendices

# APPENDIX A - VARIABLE DEFINITIONS

| Variable | Definition |
|---|---|
| Similarity | Defined as the Jaccard Similarity between the GICS Industry Group (4 digit) and the firm-specific HP TNIC peer group. Jaccard Similarity calculated as the intersection of two sets scaled by the union of two sets. |
| Beta | Defined as in Fama and French (1992). Beta calculated on monthly data over the trailing 60 months, with a minimum of 24 months required for calculation. |
| LN(BTM) | Defined as in Fama and French (1992). BE is the book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Market value comes from CRSP and is price times shares outstanding. |
| OP/Assets | Defined as in Fama and French (2015). Operating Profit (OP) equals revenue less the sum of COGS, SG&A, and Interest Expense, scaled by lagged total assets. |
| INV | Defined as in Fama and French (2015). Growth in total assets from year t-1 to t. |
| LN(MKTEQ) | Defined as in Fama and French (1992). Market value comes from CRSP and is price times shares outstanding. |
| ICC | Defined as in Easton (2004). |
| ln(Liq) | Defined as in Christensen et al. (2013, 2016) and Daske et al. (2008). The natural log of one plus the first principal components of bid-ask spreads, Amihud (2002) liquidity, percent of zero trading days, and a measure of transaction costs from Corwin and Schultz (2012). |
| Capex | Capex scaled by lagged total assets. |
| R&D | R&D scaled by lagged total assets. |
| ln(AFE) | The natural log of the absolute value of the difference between the mean EPS estimate and the actual EPS. |
| OP/Sales | OP defined as in Fama and French (2015) scaled by contemporaneous revenue. |
| Sales Growth | Change in revenue from year t-1 to t. |
| Asset Beta | Unlevered Beta based on 35% tax rate and total debt to book equity ratio. Book equity calculated as in Fama and French (1992). |
| ARC | Accounting reporting complexity from Hoitash and Hoitash (2018). |
| Complexity | Business complexity measure from Loughran and McDonald (2020). |

# APPENDIX B - ADDITIONAL FIGURES AND TABLES

Figure A1: Major GICS Reclassifications

## Changes to the GICS Structure

The four-tier GICS structure accurately reflects equities in today's global investment environment yet is flexible enough to capture tomorrow's developments. The eight-digit GICS coding system is designed to adapt easily to the changing investment world. As the global economy changes, sectors, industry group, industries, and sub-industries can be added or divided.

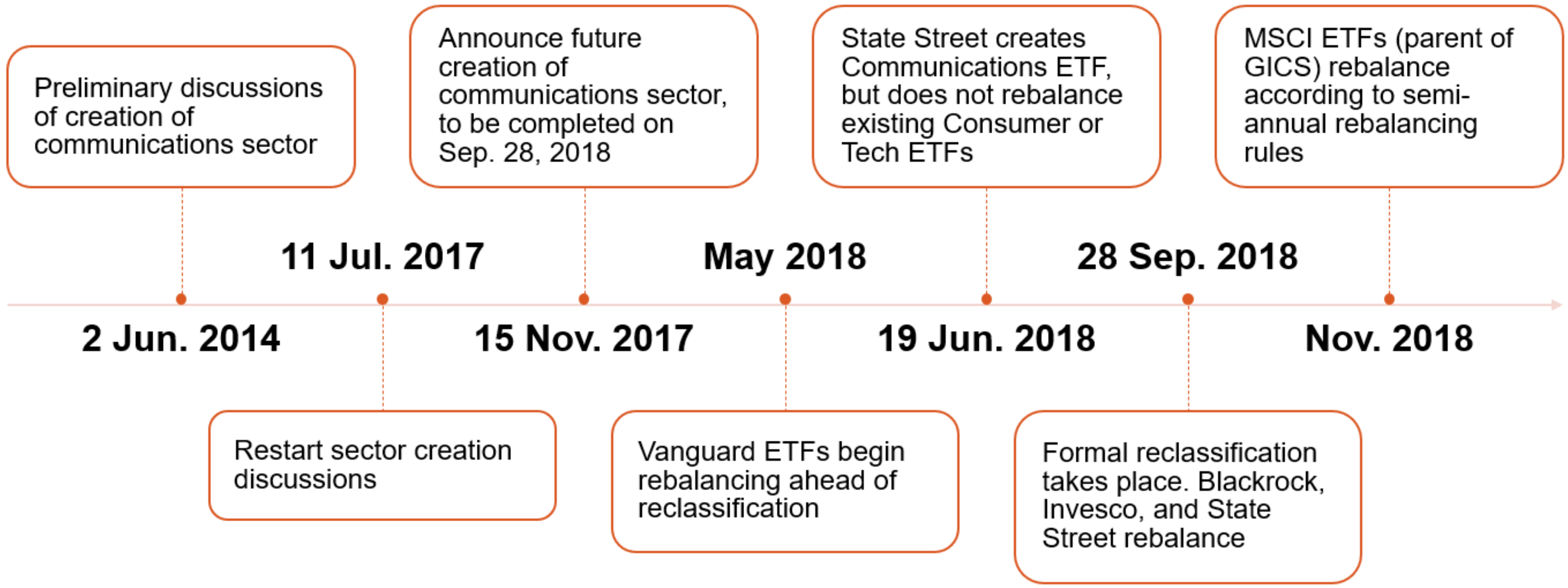
**Chronology of Events**

| Date | GICS Event |
| --- | --- |
| August 1999 | GICS introduced. GICS comprised of 10 sectors, 23 industry groups, 59 industries and 123 sub-industries. |
| March 2002 | The total number of sub-industries in the GICS structure reduced to 122 from 123. Name and description changes to other GICS categories. GICS comprised of 10 sectors, 23 industry groups, 59 industries and 122 sub-industries. |
| April 2003 | Changes included the creation of new industry groups, industries and sub-industries, the merger of some existing sub-industries and significant name and definition changes to others. GICS comprised of 10 sectors, 24 industry groups, 62 industries and 132 sub-industries. |
| April 2004 | Changes limited to definition changes for some of the GICS sub-industries. GICS comprised of 10 sectors, 24 industry groups, 62 industries and 132 sub-industries. |
| April 2005 | Changes included renaming and redefining one industry group, creating two industries and seven sub-industries, and making significant name and definition changes to others. GICS comprised of 10 sectors, 24 industry groups, 64 industries and 139 sub-industries. |
| April 2006 | Changes included renaming one industry group, creating four new industries and ten new sub-industries. Changes also included discontinuing one industry and two sub-industries, and making name and definition changes to others. GICS comprised of 10 sectors, 24 industry groups, 67 industries and 147 sub-industries. |
| August 2008 | Changes included renaming one industry group, creating one new industry and eight new sub-industries. Changes also included discontinuing one sub-industry, and making name and definition changes to others. GICS comprised of 10 sectors, 24 industry groups, 68 industries and 154 sub-industries. |
| June 2010 | Changes were made to definitions only. GICS comprised of 10 sectors, 24 industry groups, 68 industries and 154 sub-industries. |
| February 2014 | Changes included renaming four industries and discontinuing one, adding six new sub-industries and discontinuing 4, and making name and definition changes to others. GICS comprised of 10 sectors, 24 industry groups, 67 industries, and 156 sub-industries. |
| September 2016 | Changes included the creation of a Real Estate sector, as well as adding one industry group, three industries, and 15 sub-industries. One industry group was discontinued, along with two industries and 14 sub-industries. One industry and one sub-industry were renamed, and there were two definition changes. GICS comprised of 11 sectors, 24 industry groups, 68 industries, and 157 sub-industries. |
| September 2018 | Changes included the renaming of the Telecommunication Services Sector to Communication Services, removing the Media Industry Group from Consumer Discretionary and adding the Media Industry Group to the Communication Services Sector under the name Media & Entertainment. Under the Media & Entertainment Industry Group, three Industries were created, with seven Sub-Industries. Two Sub-Industries from Information Technology were discontinued and one was created. Several definitions were updated. GICS comprised of 11 sectors, 24 industry groups, 69 industries, and 158 sub-industries. |

*Detailed information on changes to the GICS structure can be found on S&P Dow Jones Indices' Web site at www.spdji.com.*

*Notes:* This figure shows the 10 major reclassifications GICS has instituted since its creation in 1999. Given that some reclassifications occur at the sub-industry level, I only exploit 5 of the 10: 2003, 2005, 2006, 2016, and 2018. Details can be found here.

Figure A2: Timeline of the Creation of the GICS Communications Sector




*Notes:* This figure shows the timeline from the initial proposal to create a new Communications sector, through the creation and index-rebalancing at the end of 2018.

Table A1: Fixed Effects Robustness on Cross-Sectional Results

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| FIC FE | -1.010*** | -0.014*** | 0.010*** | 0.071*** | 0.061*** | -0.216*** | -0.001 | 0.001** | 0.012*** | 0.003 |
| | (-24.00) | (-8.11) | (32.70) | (65.65) | (18.87) | (-3.97) | (-0.28) | (1.99) | (8.65) | (0.49) |
| SIC FE | -1.117*** | -0.007*** | 0.002*** | 0.088*** | 0.014*** | -0.267*** | -0.007* | 0.002*** | 0.012*** | -0.007 |
| | (-26.74) | (-3.10) | (5.64) | (70.50) | (3.31) | (-4.28) | (-1.95) | (3.41) | (9.22) | (-1.18) |
| GICS+FIC FE | -1.063*** | -0.036*** | 0.012*** | 0.055*** | 0.004 | -0.127** | -0.012*** | 0.001 | 0.008*** | 0.003 |
| | (-25.95) | (-19.20) | (28.26) | (46.71) | (1.12) | (-2.21) | (-3.34) | (1.43) | (4.63) | (0.57) |
| GICS+SIC FE | -1.056*** | -0.036*** | 0.008*** | 0.058*** | 0.004 | -0.074 | -0.015*** | 0.002*** | 0.007*** | 0.004 |
| | (-30.33) | (-18.36) | (21.05) | (53.11) | (0.90) | (-1.27) | (-4.39) | (3.97) | (4.47) | (0.67) |
| GICS+FIC+SIC FE | -1.121*** | -0.038*** | 0.007*** | 0.054*** | 0.014*** | -0.108* | -0.015*** | 0.001** | 0.006*** | -0.002 |
| | (-28.41) | (-19.71) | (19.37) | (45.60) | (3.69) | (-1.84) | (-4.12) | (2.02) | (3.42) | (-0.32) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |

*Notes:* This table reports coefficient estimates from association tests of classification quality (*Similarity*) on a variety of outcome variables. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

## Table A2: Cross-Sectional Results on Valuation Multiples

| | (1) EV/EBIT | (2) EV/EBITDA | (3) Price/Earnings | (4) Price/Sales | (5) Price/Book | (6) EV/EBIT | (7) EV/EBITDA | (8) Price/Earnings | (9) Price/Sales | (10) Price/Book |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t |
| Similarity | 1.235*** | 0.243*** | 2.446*** | 2.752*** | 0.295*** | 1.353*** | 0.215*** | 2.707*** | -0.384* | -0.161*** |
| | (9.43) | (4.14) | (8.74) | (29.19) | (11.71) | (7.37) | (3.54) | (6.79) | (-1.72) | (-5.02) |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 531961 | 531343 | 462946 | 718963 | 732026 | 531861 | 531244 | 462885 | 718872 | 731933 |
| $R^2$ | 0.081 | 0.153 | 0.069 | 0.135 | 0.097 | 0.416 | 0.550 | 0.321 | 0.539 | 0.457 |

*Notes:* This table reports coefficient estimates from association tests of classification quality on valuation multiples. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table A3: Benchmarking Mechanism - Splitting on Fund Ownership

Panel A: Big 3 Index Fund Ownership

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|
| Similarity | -0.481*** | -0.017*** | 0.003*** | 0.003 | 0.002 |
| | (-6.91) | (-3.84) | (5.39) | (1.10) | (0.28) |
| High Index × Similarity | 0.390*** | 0.006* | -0.000 | 0.000 | 0.015*** |
| | (6.75) | (1.89) | (-1.19) | (0.27) | (2.78) |
| Inst Own (%) | -2.856*** | -0.455*** | 0.015*** | 0.012*** | 0.022 |
| | (-13.57) | (-30.89) | (20.49) | (3.27) | (1.14) |
| FF Ctrl | Yes | No | No | No | No |
| Firm FE | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No |
| Observations | 361197 | 485334 | 511882 | 350091 | 438550 |
| $R^2$ | 0.521 | 0.404 | 0.612 | 0.404 | 0.367 |

Panel B: Active Mutual Fund and Big 3 Index Fund Ownership

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|
| Similarity | -0.106* | -0.010*** | 0.002*** | 0.007*** | 0.015*** |
| | (-1.84) | (-2.78) | (3.19) | (4.28) | (2.93) |
| Low Index High Active MF | -0.052 | -0.009*** | 0.001*** | 0.011*** | -0.017*** |
| × Similarity | (-1.05) | (-3.36) | (3.60) | (9.56) | (-3.51) |
| Inst Own (%) | -2.656*** | -0.471*** | 0.016*** | -0.002 | 0.042** |
| | (-15.00) | (-35.13) | (26.57) | (-0.96) | (2.50) |
| FF Ctrl | Yes | No | No | No | No |
| Firm FE | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No |
| Observations | 446586 | 705245 | 739211 | 493768 | 554128 |
| $R^2$ | 0.533 | 0.532 | 0.531 | 0.437 | 0.356 |

*Notes:* In Panel A, "High Index" refers to firms who have above median ownership from the Big 3 index fund companies (Vanguard, Blackrock, State Street), measured as a percentage of total shares outstanding. In Panel B, "Low Index High Active MF" refers to firms who have below median ownership from the Big 3 index fund companies and above median active mutual fund ownership. Active and mutual fund definitions come from Bushee (2001) classifications. The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table A4: Heterogeneous Investment Response

Panel A: Associative Results

| | (1) RQ β / t | (2) R&D β / t | (3) R&D β / t | (4) RQ β / t | (5) R&D β / t | (6) R&D β / t |
|---|---|---|---|---|---|---|
| Similarity | 0.007*** | 0.041*** | 0.000 | 0.001*** | 0.004** | 0.003 |
| | (29.31) | (35.08) | (0.10) | (5.46) | (2.40) | (1.04) |
| Q | | 0.014*** | | | 0.007*** | |
| | | (25.34) | | | (12.23) | |
| Similarity × Q | | 0.004*** | | | 0.002*** | |
| | | (10.28) | | | (5.30) | |
| Similarity × Ind Q | | | 0.017*** | | | 0.002** |
| | | | (17.29) | | | (2.30) |
| Firm FE | No | No | No | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 238342 | 493820 | 493820 | 238319 | 493768 | 493768 |
| $R^2$ | 0.181 | 0.189 | 0.171 | 0.727 | 0.440 | 0.437 |

Panel B: DD Results

| | (1) RQ β / t | (2) R&D β / t | (3) R&D β / t | (4) RQ β / t | (5) R&D β / t | (6) R&D β / t |
|---|---|---|---|---|---|---|
| Treat Post | -0.013*** | -0.045*** | -0.043*** | -0.007 | -0.022*** | -0.017*** |
| | (-7.13) | (-6.00) | (-5.65) | (-1.58) | (-6.88) | (-4.27) |
| Q | | 0.003*** | | | 0.002*** | |
| | | (4.62) | | | (4.52) | |
| Treat Post × Q | | 0.006*** | | | 0.006*** | |
| | | (7.38) | | | (6.93) | |
| Ind Q | | | 0.000** | | | 0.000*** |
| | | | (2.55) | | | (4.80) |
| Treat Post × Ind Q | | | 0.007*** | | | 0.006*** |
| | | | (4.10) | | | (4.46) |
| Cohort-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-FIC-Month | No | No | No | Yes | Yes | Yes |
| Cohort-Month | Yes | Yes | Yes | No | No | No |
| Observations | 21060 | 44960 | 44960 | 19483 | 43141 | 43141 |
| $R^2$ | 0.812 | 0.388 | 0.386 | 0.846 | 0.684 | 0.681 |

*Notes:* RQ is a measure of R&D efficiency from Cooper et al. (2022). Columns (2), (3), (5), and (6) compare the sensitivity of R&D investment to classification quality and Q ratios. The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table A5: Capital Market Responses Around Peer Earnings Announcements

| | (1) IPT $\beta$ / t | (2) IPT $\beta$ / t | (3) IPT $\beta$ / t | (4) IPT $\beta$ / t |
|---|---|---|---|---|
| GICS Peer | 0.042*** | 0.030*** | 0.024*** | 0.028*** |
| | (5.51) | (5.12) | (3.70) | (4.70) |
| HP Score | | | 0.258** | 0.177 |
| | | | (2.58) | (1.62) |
| GICS Peer $\times$ HP Score | | | -0.122 | -0.261* |
| | | | (-1.16) | (-1.90) |
| Firm-Year FE | Yes | No | Yes | No |
| Peer Firm-Date FE | No | Yes | No | Yes |
| Observations | 7423329 | 7424768 | 7361070 | 7362483 |
| $R^2$ | 0.046 | 0.027 | 0.045 | 0.026 |

*Notes:* The analyses are conducted at the pair-earnings announcement level. That is, each observation contains the capital market response for a focal firm around an HP peer earnings announcement. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table A6: Continuous Difference-in-Differences Results

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Pre-Period Similarity | -0.905** | -0.020 | -0.003 | -0.064*** | -0.010 | -1.664*** | -0.017 | -0.013*** | -0.015*** | -0.006 |
| | (-1.98) | (-0.77) | (-0.80) | (-3.61) | (-0.26) | (-3.50) | (-0.65) | (-4.48) | (-3.34) | (-0.14) |
| Treat Post | 0.689 | 0.065** | 0.019*** | -0.025*** | -0.153*** | 2.345*** | 0.077* | 0.020*** | -0.002 | -0.305*** |
| | (1.17) | (2.49) | (9.48) | (-3.96) | (-3.68) | (2.72) | (1.78) | (3.94) | (-0.95) | (-4.08) |
| Treat Post × Pre-Period Similarity | -0.545*** | 0.046*** | -0.008*** | 0.001 | -0.039** | -0.493** | 0.036*** | -0.006*** | -0.011*** | -0.076*** |
| | (-2.91) | (3.23) | (-6.75) | (0.19) | (-2.11) | (-2.31) | (2.61) | (-3.26) | (-2.63) | (-3.61) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Cohort-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-FIC-Time FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes | Yes | No | No | No | No | No |
| Observations | 44208 | 71511 | 75403 | 43003 | 55331 | 42573 | 70052 | 73993 | 41277 | 53899 |
| $R^2$ | 0.617 | 0.476 | 0.582 | 0.387 | 0.447 | 0.681 | 0.547 | 0.684 | 0.674 | 0.536 |

*Notes:* This table reports coefficient estimates from Continuous DD estimates from staggered GICS reclassifications. The coefficient of interest is the interacted variable of Pre-Period Similarity × Treat Post. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the cohort-month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

## Table A7: Controlling for Complexity

Panel A: Business Complexity

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Similarity | -0.925*** | -0.027*** | 0.014*** | 0.054*** | -0.014*** | -0.140** | -0.006* | 0.002*** | 0.005*** | 0.014** |
| | (-24.17) | (-13.37) | (33.24) | (46.90) | (-3.00) | (-2.22) | (-1.66) | (2.79) | (2.92) | (2.35) |
| Complexity | 21.222*** | -0.603*** | -0.033*** | -0.105*** | 2.191*** | 7.465*** | 0.053 | 0.005 | -0.059** | 1.182*** |
| | (40.64) | (-16.38) | (-18.50) | (-17.20) | (45.29) | (15.18) | (1.62) | (1.20) | (-2.48) | (19.89) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 383643 | 588347 | 614330 | 415484 | 469888 | 383591 | 588305 | 614288 | 415463 | 469845 |
| $R^2$ | 0.297 | 0.376 | 0.224 | 0.177 | 0.123 | 0.516 | 0.486 | 0.552 | 0.413 | 0.363 |

Panel B: Accounting Complexity

| | (1) ICC $\beta$ / t | (2) ln(Liq) $\beta$ / t | (3) Capex $\beta$ / t | (4) R&D $\beta$ / t | (5) ln(AFE) $\beta$ / t | (6) ICC $\beta$ / t | (7) ln(Liq) $\beta$ / t | (8) Capex $\beta$ / t | (9) R&D $\beta$ / t | (10) ln(AFE) $\beta$ / t |
|---|---|---|---|---|---|---|---|---|---|---|
| Similarity | -0.950*** | -0.028*** | 0.015*** | 0.056*** | 0.008 | -0.691*** | -0.017*** | 0.006*** | -0.009 | -0.018* |
| | (-14.30) | (-8.79) | (21.67) | (20.16) | (0.93) | (-5.88) | (-2.68) | (6.73) | (-1.62) | (-1.74) |
| ARC | 0.010*** | -0.000*** | -0.000*** | -0.000*** | 0.001*** | 0.004*** | 0.000** | -0.000* | -0.000*** | 0.000*** |
| | (17.97) | (-16.19) | (-29.41) | (-31.25) | (26.52) | (7.52) | (2.41) | (-1.84) | (-7.74) | (4.29) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 179889 | 255017 | 267008 | 182250 | 214706 | 179774 | 254903 | 266896 | 182167 | 214666 |
| $R^2$ | 0.303 | 0.360 | 0.212 | 0.168 | 0.126 | 0.564 | 0.463 | 0.633 | 0.546 | 0.428 |

*Notes:* I repeat the cross-sectional analysis from Table 3 with controls for firm complexity. In Panel A, I control for the business complexity measure from Loughran and McDonald (2020), which is a scaled count variable of the number of unique complexity-related words that appear in a firm's 10-K. The measure is available from 2001-2018. In Panel B, I control for the accounting reporting complexity measure from Hoitash and Hoitash (2018), which measures the number of XBRL tags in a firm's 10-K. The data are available from 2011 through June 2020. Standard errors are clustered at the year level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.

Table A8: Effect of Transitivity on Information Processing

| | (1) ICC | (2) ln(Liq) | (3) Capex | (4) R&D | (5) ln(AFE) | (6) ICC | (7) ln(Liq) | (8) Capex | (9) R&D | (10) ln(AFE) |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t | $\beta$ / t |
| Similarity | -0.929*** | -0.032*** | 0.013*** | 0.057*** | 0.007* | -0.107* | -0.014*** | 0.001** | 0.008*** | 0.017*** |
| | (-23.79) | (-16.57) | (30.66) | (48.62) | (1.74) | (-1.81) | (-3.92) | (2.52) | (4.21) | (3.19) |
| Transitive FIC Similarity | -0.081*** | -0.008*** | 0.003*** | -0.004*** | -0.021*** | -0.005 | 0.002 | 0.003*** | 0.006*** | -0.024*** |
| | (-3.37) | (-4.39) | (20.44) | (-10.20) | (-7.50) | (-0.13) | (0.98) | (8.03) | (3.65) | (-5.52) |
| FF Ctrl | Yes | No | No | No | No | Yes | No | No | No | No |
| Firm FE | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Group-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Size-Time FE | No | Yes | No | No | No | No | Yes | No | No | No |
| Observations | 446699 | 705330 | 739296 | 493820 | 554229 | 446586 | 705245 | 739211 | 493768 | 554128 |
| $R^2$ | 0.303 | 0.426 | 0.209 | 0.167 | 0.111 | 0.533 | 0.528 | 0.531 | 0.437 | 0.356 |

*Notes:* This table reports coefficient estimates from association tests of classification quality on a variety of outcome variables. Transitive FIC Similarity is calculated as the Jaccard Similarity between GICS Industry Groups and the Hoberg and Phillips (2016) FIC codes. ICC is the implied cost of capital, ln(Liq) is the natural log of the liquidity factor, Capex is Capex/Assets, R&D is R&D Expense/Assets, AFE is analyst forecast error. FF Ctrl are controls for characteristics known to explain expected returns: the 5 Fama and French (2015) factors, short-term reversal (returns for the most recent month), and momentum (last 12 month returns, excluding the most recent month). The analyses are conducted at the firm-month level. The sample spans from July 2000 to June 2021. Standard errors are clustered at the month level. All variables are defined in the Appendix. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels (two-tailed), respectively.