STOCKHOLM SCHOOL OF ECONOMICS
Department of Finance
Master Thesis in Finance
Spring 2021

Peering into the future

A study of the relationship between listed industry peer populations and stock price informativeness

C.A.L.M. Planting-Gyllenbåga *

May 31, 2021

Abstract

This thesis explores the relationship between US firms' stock price informativeness and the number of US public companies belonging to their same primary industry. The study is conducted on an unbalanced panel of 94,714 firm-year observations over the period from 1976 to 2016. Using an extended FERC methodology with fixed effects, I find that the association between price informativeness and industry listing counts is insignificantly negative in my main specification. Furthermore, the inverse relationship becomes statistically significant in recent years and across a couple of robustness checks. Taken at face value, the non-positive results suggest that the informativeness of most US stock prices may not have suffered from their having fewer listed industry peers in the last two-and-a-half decades. Additionally, the persistently negative coefficients hint that investors may be employing intra-industry information improperly, especially in more recent times for firms with high market-to-book ratios.

Supervisor: Dong Yan

Keywords: Price informativeness, Industry population, Intra-industry information, Declining listings, Market efficiency

*E-mail: 41669@student.hhs.se

Acknowledgements

This thesis would not have taken its current form without the support of many people. I must first thank Dong Yan, my supervisor, for nudging me out the door and guiding me since. For their patient advice on both qualitative and quantitative matters, I am also hugely indebted to Daniel Evans and Adam Gill. Your comments were invaluable. Additionally, I must express my gratitude to the many friends, and Simran Pachnanda in particular, who lightened these heavy months. Finally, I have only managed to get this far because of my family. Thank you for the lifelong support and for putting up with very long phone calls recently. Not to mention that your visit was especially timely Peach, thank you.

Contents

List of Figures

List of Tables

1 Introduction

This thesis explores the relationship between the information content (informativeness) of the stock prices of firms based in the United States (US) and the number of US public companies belonging to their same primary industry [1]. My motivation for this study initially derived from the dramatic decline in the number of domestically-listed US public firms since 1997 (Figure A.1 in the Appendix), a fall that is disproportionate even relative to other countries (Doidge et al. 2017). The phenomenon is also present across the vast majority of US industries, classified by 3-digit Standard Industrial Classification (SIC) codes (Figure A.2 in the Appendix). This raises concerns that the informativeness of many US stock prices may have suffered from there being less public information generated by listed peers, with possible knock-on effects on the investment efficiency of both public and private US firms.

However, existing research suggests no straightforward relationship between industry populations and informativeness. On the one hand, some literature suggests that an additional public firm could generate positive informational externalities for its peers. It may attract more analysts and liquidity to the industry, as well as deepen the pool of information useful for evaluating other firms in the industry (intra-industry information) through its disclosures and pricing signals. The plunge in listings may have reduced these informational benefits amongst peers. On the other hand, market inefficiencies may be impeding the constructive use of additional intra-industry signals, in which case a declining number of peers may not necessarily have adversely affected informativeness. These different possibilities create cause for empirical investigation. Thus, my research question can be stated as follows: *Ceteris paribus, do US firms in industries with fewer listed domestic companies have stock prices that are less informative?*

Various aspects of this study may appeal to different stakeholders. As implied earlier, the consequences of the decline in US listings have not been exhaustively studied. My thesis helps fill some of this research gap, which may be of interest to academics in the field and to regulators worried about the shrinking US market. I also contribute to our understanding of the determinants of stock price informativeness. A large body of research has identified factors like corporate disclosures, business uncertainty, analyst coverage, and arbitrage costs, but the role of industry population has not yet been examined, to my knowledge. Despite this, some literature has already assumed that informativeness is increasing with the number of industry peers, further motivating this study.

For many readers, the concept of price informativeness will need more explanation. Since

[1] When I refer to "industry listing counts", "industry lists", "industry population", the "number of listed peers", or similar expressions in this thesis, I mean the number of US firms, listed on a US exchange, which are classified within the same primary industry.

Hayek (1945), literature has mostly argued that the information investors trade on is, in-aggregate, reflected in market prices. Informativeness reflects the amount of information about a firm's prospects that has been aggregated into its stock price by market activity. As it is inherently unobservable, measuring informativeness is an inaccurate science. In line with many other papers in the field, I use the future earnings response coefficient (FERC) for this task. The philosophy of FERC is based upon the classic valuation framework that explains current excess stock returns as a function of unexpected current dividends and changing expectations about future dividends. Price efficiency implies that the discovery of new information should alter investors' expectations, causing the market price to change in response. Thus, the stronger the link between the stock price and changing expectations, the more information is presumed to have been aggregated and the more informative the stock price. As changing expectations are themselves not measurable, the calculation of FERC entails a variation of this relationship which estimates informativeness by regressing future earnings and future stock returns on current stock returns. The derivation of FERC is discussed in more detail in section 4.

I formulate two hypotheses that allow me to test and draw inferences about the avenues through which industry populations may affect informativeness – analyst coverage, liquidity, and intra-industry information. To begin with, I address the measurable factors – analyst coverage and liquidity. Based on Subrahmanyam & Titman (1999) and Veldkamp (2006), my first hypothesis expects these mediator variables to increase with industry listing counts. Since research has shown that analyst coverage and liquidity improve informativeness (Kerr et al. 2020, Chordia et al. 2008, Piotroski & Roulstone 2004), evidence in favour of my first hypothesis would also suggest that informativeness may be increasing with industry population. However, there is also an unobservable factor at play – intra-industry information – which may have a positive or negative influence on informativeness, depending on the market's efficiency. Thus, my second hypothesis expects a positive association between industry lists and informativeness if the evidence favours the first hypothesis and if the market behaves efficiently with respect to intra-industry information. As I cannot observe the latter condition, this hypothesis is not fully dismissible. Instead, I use it to propose inferences about the market's treatment of intra-industry information, under the assumption that all other avenues through which listing counts may affect informativeness have been already addressed.

To investigate my hypotheses, I use data from the Center for Research in Security Prices (CRSP), Compustat, and the Institutional Brokers Estimate System (I/B/E/S) to form an unbalanced panel of firm and industry characteristics across 9,690 unique US firms with a domestically listed security during the period from 1976 to 2016. In aggregate, the dataset consists of 94,714 firm-year observations across 378 unique primary 3-digit SIC [2] industries. Using this sample, I first run univariate and multivariate ordinary least squares (OLS) regressions with time fixed effects of analyst coverage and liquidity on industry listing counts. As expected by my first hypothesis, the multivariate models yield significantly positive coefficients on industry listing counts. However, the explanatory power of industry populations is very small for both mediator variables. Next, I examine the relationship between industry listing counts and informativeness by modifying and extending Collins et al. (1994)'s FERC model, as done by Lundholm & Myers (2002) and others. The method requires each variable of interest and control to be interacted with each independent variable in the basic FERC model. Thus, one helps ensure an unbiased estimation of the coefficient on the interaction term between future earnings and the variable of interest. For my study, the relevant interaction is between future earnings and industry listing counts.

The results of the extended FERC model show a statistically insignificant, negative coefficient on the interaction of interest, implying that informativeness is not meaningfully related to industry population. The inclusion of analyst coverage and liquidity mediator variables makes next to no difference to my results. I subsequently perform various cross-sectional tests on subsamples of the main dataset to investigate this finding. First, I check whether an additional listing is beneficial in smaller industries but irrelevant in larger industries. There may be a threshold beyond which the incremental information generated becomes negligible, or even damaging, for FERC. Second, I test whether the informativeness of different types of firms respond differently to industry listings. Literature suggests that intra-industry signals may be more used in younger firms, larger firms, those with a higher market-to-book (M/B) ratio, or those with a greater proportion of intangible assets. All these cross-sectional tests yield insignificant differences, implying that the insignificant relationship between industry listing counts and informativeness is not due to opposing behaviours between cross-sections.

However, an analysis of the coefficient of interest across time reveals some interesting results as I find a structural break in 2011. The relation between industry populations and informativeness remains insignificantly negative in the prior period from 1976 to 2010 but becomes significantly negative from 2011 to 2016, particularly for firms with high M/B ratios [3]. Finally, I perform multiple robustness checks to see how sensitive my results are to my modelling choices. None of these reveals a positive relation between industry lists and informativeness. Instead, some show statistically significant negative coefficients even across the full dataset.

[2] See section 6.2 for a an explanation of SIC classifications.

[3] In the 2011 to 2016 period, the coefficient of interest is also significantly negative for a cross-section based on a high proportion of intangible assets. However, the evidence for this is mixed across robustness tests.

I propose several inferences from these findings. First, the very limited effect of the mediators imply that additional industry lists appear to have essentially no influence on informativeness through their attraction of more analysts and trading activity. Second, the statistical insignificance of the relation between industry listing counts and informativeness in models that account for the mediators suggest that intra-industry signals may be relatively unimportant for the evaluation of a stock. However, the occasional significance of the persistently negative coefficients on industry listing counts may also be due to the market reacting inappropriately to the greater quantities of intra-industry information available in larger industries, especially in more recent years. Thus, I speculate that investors may often be susceptible to the noise in intra-industry signals or trade on a smaller subset of information when evaluating firms in larger industries. This thesis may therefore also contribute to literature studying intra-industry information and the market's reaction to it, which may interest investors who are evaluating their use of industry signals or looking to better understand recent market developments.

Regardless of its causes, the non-positive association between industry listings and informativeness has a couple of implications. First, it suggests that the informativeness of most US stock prices may not have suffered from their having fewer industry peers. Regulators may have one less reason to worry about the declining number of listed in the US. Also, my non-positive results question the assumption made by papers, such as Chemmanur et al. (2010), that investors can more easily evaluate firms with a greater number of public industry peers. Consequently, some literature may need to reconsider using industry listing counts as a proxy for informativeness. All this said, readers should keep in mind that imperfections and limitations in my specification, such as difficulties in accurately encompassing industries, could be obscuring the true relationship between informativeness and industry listing counts.

The remainder of the thesis is structured as follows. In Section 2, I summarise academic literature that is directly and adjacently relevant to this thesis, as well as describe the research gap I am filling. Partially based on this review, I explain my hypotheses in Section 3. Next, in Section 4 I go into more detail about the philosophy and estimation of FERC, my measure of price informativeness. This gives the reader the background needed to understand how I construct my main model's specification in Section 5. The data I employ in my chosen model is described in depth in Section 6. My main results are described in Section 7, along with the cross-sectional investigations I perform. Subsequently, Section 8 presents the motivations and results of multiple robustness checks used to assess the validity of my findings. I discuss the possible reasons for my results, including limitations, and offer suggestions for future research in Section 9. Finally, in Section 10 I summarise the conclusions of this paper.

2 Literature Review

This thesis examines the relationship between US firms' stock price informativeness and the number of public firms within their industries. The literature review sets the stage for this analysis by first summarising the body of previous work investigating why the number of public firms has declined in the US. Next, it turns to academic studies regarding the effects that falling listings may have had on firms and the economy. Surveying these consequences shows that little research has been conducted into the effects on market efficiency and price informativeness, the focus of this thesis. This leads to an examination of the literature that studies the determinants of price informativeness, which has a similar gap. To complete the picture and provide further significance to the topic, the review also looks at how informativeness could, in turn, influence the economy. Finally, I summarise how I contribute to existing knowledge within the academic landscape sketched by this review.

2.1 Backdrop: The declining number of listed US firms

Doidge et al. (2017) show that the US has experienced a declining number of public firms, both in absolute terms and relative to other countries. They, and other academics, study this phenomenon through a cost-benefit trade-off framework of a firm's decision to go public. The less attractive this trade-off is, the lower a firm's listing propensity - the likelihood that it will opt to go or remain public. Expressing listing counts as a function of listing propensity and the total number of firms able to list, Doidge et al. (2017) show that falling US listings have primarily resulted from a lower listing propensity.

A large body of literature (e.g. Subrahmanyam & Titman (1999), Pagano et al. (1998), Bushee & Miller (2012), Chemmanur & Fulghieri (1999)) explores the costs and benefits that define the going-public decision and listing propensities. Some more well-defined costs include the fees paid to the exchange and consultants, the efforts and risks associated with mandated disclosures, the separation of ownership and control, and the political pressures that accompany greater visibility. On the other hand, key benefits to pursuing an IPO usually include access to funding from public markets, a share-based currency for acquisitions and employee compensation, price discovery, and liquidity for pre-IPO shareholders. Ritter & Welch (2002), Röell (1995) and Djama et al. (2012) provide a deeper review of the advantages and disadvantages of going public.

Literature falls into two main camps when explaining why the cost-benefit trade-off and listing propensities have shifted in the US. One side argues that regulatory changes to public markets have caused them to fall into disfunction by making public ownership more costly. The other instead suggests that firms and their competitive environment have evolved in such a way that public capital confers fewer benefits and is less attractive than before. As a result, firms have chosen to merge or raise private capital instead of listing.

In the first camp, Weild & Kim (2009*a*) have called the US markets 'broken'. They primarily blame changes to trading regulation that eroded broker-dealer margins, such as the Manning Rule and Order Handling Rules. The introduction of such laws may have pushed banks to prioritise high-frequency trading customers over long-term investors, adversely affecting the quality of research and support for smaller firms looking to list because they would generate less order flow. Furthermore, Weild & Kim (2009*a*) point to tightening regulatory requirements for listed firms, such as the Sarbanes–Oxley Act. IPOs may have thus become more cumbersome, especially for smaller companies less able to absorb these higher fixed costs. In essence, new regulation may have increased the cost of listing, especially for smaller companies.

However, Gao et al. (2013) and Doidge et al. (2017) do not find evidence of regulatory overreach and challenge these claims. Doidge et al. (2017) instead base their argumentation on the existence of an industry-specific firm size threshold. Firms whose size exceeds the threshold prefer public ownership because its benefits start to outweigh the costs. If this threshold were to increase, some firms that were once large enough to prefer listing may no longer be. Symptoms of this would include fewer listed firms and an increase in average size of the remaining listed firms. In line with this theory, Doidge et al. (2017) find that the US's listing gap cannot be explained by industry factors, changes in listing requirements, public-to-private transactions, regulatory reforms in the early 2000s, or a deterioration in the quality of newly listed firms. Instead, they show evidence that firms have been merging more frequently and that the average size of listed firms has increased in tandem, suggesting that the size threshold in the US has been raised.

Gao et al. (2013) suggest a fundamental reason for this - the profit-maximising company size may have changed. They argue that the importance of economies of scope and product-to-market times has grown over the years, pushing firms to seek a competitive advantage by merging into larger organisations that are better placed to compete in these areas. Going public and remaining independent would not allow firms to grow as quickly, causing them to choose a merger instead. Adjacent to this theory, Grullon et al. (2019) provide evidence of increasing industry concentration levels in the US, which they suppose may be linked to the increasing importance of economies of scale following technological change. Lattanzio et al. (2020)'s cross-country analysis also provides empirical support for Gao et al. (2013) and Doidge et al. (2017), showing that M&A activity is a main driver behind the US' listing gap.

Doidge et al. (2018) present an alternative explanation for the increasing size of public firms. They argue that entrepreneurs need less capital than before, making public fundraising necessary only for progressively larger firms. They hypothesise that this shift is the result of technological development that has enabled the outsourcing of many, otherwise fixed, costs and facilitated business scalability. Funds raised through IPOs may have become less beneficial to entrepreneurs. Furthermore, an expanding pool of private capital may have decreased the costs of remaining private. Contrary to Gao et al. (2013) and Doidge et al. (2017), Ewens & Farre-Mensa (2020) argue that "IPOs have not been replaced by an increase in the number of firms that rely on capital provided by a publicly listed acquirer to fund their growth". Instead, they show that show that start-ups are staying private longer while raising large sums of private capital. Ewens & Farre-Mensa (2020) provide further evidence that some regulatory changes, such as the National Securities Markets Improvement Act, made it easier for private firms to raise so much capital and stay private. Given an increasing supply of funds, founders may have greater bargaining power versus private equity sponsors. This makes the remaining private decision less costly to founders, who may also prefer it to an IPO as it can preserve their control and ownership over the firm (Brau & Fawcett 2006, Helwege & Packer 2009, Boot et al. 2006).

Private ownership may also have become a more efficient organisational form. Jensen (1997) argued that public ownership creates conflicts between owners and managers, which lead to economic inefficiencies. Instead, he theorised that a new organisational form embodied by private equity ownership would be more effective at solving this agency problem, making it a more efficient governance structure for some firms. Doidge et al. (2018) challenge this view though, arguing that the performance of large US public corporations shows they still represent a relevant form of governance. Instead, they propose that public markets have become less efficient because they are unable to properly evaluate younger firms with increasing amounts of intangible assets. Standard accounting principles used in public disclosures are not as informative about firms with a large proportion of intangible assets. Furthermore, companies may fear the expropriation of their intangible assets following the disclosures they would be mandated to make upon listing. By seeking private capital instead, companies with important intangible assets benefit from having to convince and interact with a much more concentrated and specialised investor base that may be more likely to properly evaluate their opportunities without divulging sensitive information to the public. In contrast to this speculation about private capital though, Lattanzio et al. (2020) find that private equity activity actually decreases the US listing gap. They suggest this may reflect how private equity's substitution of public equity is outweighed by the support private equity sponsors provide to firms, which helps them grow and eventually IPO.

2.2 The consequences of falling US listings

Ruminations about the consequences of falling US listings are inherently linked to beliefs about the phenomenon's causes. If we assume the IPO market is 'broken', the implication is that private firms would like to list but are facing more obstacles to doing so. This raises concerns about whether these companies are suffering from less efficient access to capital. Weild & Kim (2009*b*) worry that a less active US IPO market could adversely impact GDP and employment growth by restricting small firms' access to equity, and consequently debt - going public may allow companies to borrow more cheaply (Pagano et al. 1998). That said, Pagano et al. (1998)'s work on the Italian market also suggests that companies choose to go public to exploit periods of overvaluation, rather than to fund future growth. If this is the case in the US as well, fewer listings may not have as much of an impact on corporate funding opportunities as is supposed by Weild & Kim (2009*b*).

An ineffective IPO market may also adversely affect innovation. Black & Gilson (1998) suggest that a vibrant public market is key to promoting entrepreneurship and venture capital in the US. They argue that IPOs are an important exit and liquidity option for venture capitalists and entrepreneurs since they allow entrepreneurs to potentially remain in control of their company – a possibility that other exit options offer with more difficulty. Without the chance of entrepreneurs retaining control after exit, more frictions may arise in negotiations between venture capitalist and entrepreneur, with a negative impact on venture financing and innovation. However, Ewens & Farre-Mensa (2017) cast doubt on this theory by showing that venture-capital-backed start-ups in the US have continued grow and raise capital to an extent usually reserved for public companies, despite remaining private. One reason for this may that private funding has become more flexible towards the entrepreneur, as argued by Ewens & Farre-Mensa (2020). This view is more consistent with studies that believe declining listings may not reflect a broken market, but rather a movement towards more efficient private funding solutions. Doidge et al. (2017) and Ewens & Farre-Mensa (2017) espouse this possibility, suggesting that developments in financial markets, particularly in the private equity market, may have made it easier for firms to succeed without being listed.

However, declining listings may have adverse consequences even if the phenomenon reflects a shift towards a new financing equilibrium. The presence of fewer public firms may have reduced both the amount of publicly available information and market participation, with effects on firm and market efficiency. Although there is no research specifically exploring declining US listings in this way, adjacent research suggests such consequences may exist. Public firms are mandated to make disclosures, such as 10-K and 10-Q filings, that contain a large amount of information about their businesses and their competitive environment, which provides external stakeholders with insights into firm's future performance and the outlook of their industry. Furthermore, the behaviour of peers' stock prices facing a common shock may provide information to managers, investors, and other stakeholders (Foucault & Fresard 2014). As the number of public firms declines, so does the amount of publicly available information they produce, with potentially negative repercussions. Badertscher et al. (2013), for example, show that private US firms face greater business uncertainty and worse investment efficiency if there is less public firm presence in their industry.

The amount of information that is accurately reflected by market prices may also be affected. Foster (1981) and Han et al. (1989), amongst others, show that stock prices of US firms react to the earnings releases and management forecasts of US firms within the same industry, implying that intra-industry information is used in the pricing of firms. Ramnath (2002) and Thomas & Zhang (2008) suggest the immediate reaction to new information may cause mispricing through. They find that investors seem to assume that good (bad) news from firms making the first announcements imply good (bad) news for subsequent announcers. However, this mispricing is corrected quickly once later announcers release their earnings (Thomas & Zhang 2008). Thus, as the amount of disclosures falls with the number of listed firms, the information content of stocks may have decreased in the US.

Furthermore, theoretical work by Subrahmanyam & Titman (1999) suggests that IPOs may have positive effects on the price efficiency of other public firms. They propose a model whereby analysts examining a public company may discover "serendipitous" information that is useful for the evaluation of another company as well, usually a peer. Individual snippets of serendipitous information may be diverse and noisy but could provide useful signals to managers and investors once aggregated into market prices (see section 2.3). Subrahmanyam & Titman (1999) thus argue that an additional public firm creates more opportunities for analysts to uncover serendipitous information, potentially allowing them to better evaluate similar companies. Also, they propose that an additional IPO increases the size and liquidity of the market by attracting more investors. Assuming the market's ability to aggregate information increases with its liquidity, as later shown by Chordia et al. (2008), this would make the market more efficient.

Hence, theory suggests that a market with more firms is likely to contain more information and function more efficiently. The declining number of public firms in the US may thus see stocks losing some informativeness. Lower price informativeness may consequently have negative knock-on effects on capital allocation across the economy, as discussed in section 2.4.

2.3 Stock price informativeness and its determinants

An extensive body of literature, such as Hayek (1945), Grossman & Stiglitz (1980), and Glosten & Milgrom (1985), argues that a key purpose of financial markets is to allow the aggregate activity of traders to summarise dispersed information about economic fundamentals into stock prices. Individually, investors trade on specific signals they have gathered. These naturally differ but as a multitude of trades occur based on investors' different beliefs, market prices end up reflecting the collective information used by investors. Stock price informativeness derives from this concept. At its core, informativeness represents the amount of information about a firm's prospects that has been accumulated into its stock price. Since prices' information content is not observable though, various methods have been used to indirectly measure informativeness. These mainly revolve around stock price non-synchronicity (Roll 1988) or the relation between current returns (or prices) and future cash flows, usually represented by FERC (Collins et al. 1994). The latter approach is the one I will use for this paper because it seems better than the former at differentiating real information from noise. FERC is discussed in detail in section 4 while stock-price non-synchronicity will be briefly covered in this section and section 8 as a robustness check.

Current literature supposes that the information about fundamentals derives from two main sources: the disclosure of information to the public, such as earnings reports or economic statistics, and the activities of risk arbitrageurs who gather and trade on private information. Studying the former avenue, Gelb & Zarowin (2002) and Lundholm & Myers (2002) find evidence that US firms with better disclosures, represented by AIMR-FAF scores [4], tend to have more informative stock prices, as measured by FERC. The ability to compare disclosures easily is also useful as it allows investors to process information at a lower cost. Choi et al. (2019) find that US firms whose financial statements are more comparable to those of peers tend to have prices that better reflect future earnings. Other practices that appear to make disclosures more informative include income smoothing Tucker & Zarowin (2006), capitalisation of R&D Oswald & Zarowin (2007), and direct method cash flow statements Orpurt & Zang (2009).

The main purveyors of disclosed information are financial analysts and institutional investors. Ayers & Freeman (2003) suggest that these agents accelerate the pricing in of future earnings, which Piotroski & Roulstone (2004) theorise is mainly the result of their analysis of public information rather than their generation of private information. However, other literature has found evidence in favour of the relevance of private information. Roll (1988) conjectured about its importance when observing that firm-specific stock price movements

[4] AIMR-FAF scores are an annual ranking of firms based the informativeness of their disclosures, as judged by analysts belonging to what is now the CFA institute.

were often not related to the introduction of new public information through news releases. He suggested this may be because private information was particularly important in the capitalisation of firm-specific information. Assuming that variations in stock returns can be broken down into market-related variations, industry-related variations, and firm-specific variations, he further proposed that stocks with higher non-synchronous price movements – return variations unexplained by market or industry variations – contained more private information. As a result, non-synchronicity has been used by many studies as a proxy for stock price informativeness. However, Roll also acknowledged that greater non-synchronicity could be the result of greater noise, irrational stock variations based on sentiment, or incomplete or inaccurate data. Although the differentiation between the noise and information components of non-synchronous prices is unclear, Durnev et al. (2003) show that US stocks with higher price synchronicity do indeed tend to have a higher FERC, supporting the use of non-synchronicity as a proxy for informativeness.

Noise trading may also directly impact informativeness by affecting the ease with which private information is priced into stocks. Classical theories of market efficiency posit that the activity of arbitrageurs, sophisticated investors who know the fundamental value of securities and trade to profit on the difference between actual and fundamental prices, ensures that prices reflect all known information. However, De Long et al. (1990) suggest that such arbitrage trades are not risk-free – the arbitrageur is exposed to the risk that irrational investors trade the price away from fundamentals. This is termed 'noise trader risk'. Wurgler & Zhuravskaya (2002) suggest that arbitrageurs facing greater noise trader risk require greater returns to justify their arbitrage trades, meaning that they only engage in arbitrage if a stock's price is further from its fundamentals. Greater arbitrage risk is thus thought to result in stock prices that are mispriced for longer periods of time. Shleifer (2000) show that arbitrage risk is more significant for smaller and less liquid firms, suggesting that these stocks' prices may be less informative. In line with this theory, Dávila & Parlatore (2018) find that US stocks traded on the NYSE, with higher market capitalisation and greater share turnover, tend to have more informative prices. Also studying the US, Kerr et al. (2020) find further support that prices of more liquid stocks have greater FERCs. Additionally, Jiambalvo et al. (2002) and Fan et al. (2019) show that firms with a greater proportion of institutional ownership, also related to less noise trading, tend to have more informative stock prices.

Arbitrage risk may also affect whole markets. In their investigation of how macro factors are related markets' aggregation of private information, Morck et al. (2000) find more synchronous stock returns in emerging economies than in developed economies. They show this phenomenon is linked more to the protection of property rights rather than to the structural characteristics of economies, such as market size, volatility of fundamentals, country size or economic diversification. Drawing from De Long et al. (1990), Morck et al. (2000) conjecture that stronger property rights may facilitate arbitrage, which in turn allows more information to be priced into stocks. On the other hand, weaker property rights may make arbitrage riskier and leave more room for noise trading.

Developments in technology and finance may be other examples of macro factors relevant to stocks' informativeness. Declining trading costs, the proliferation of information, cheaper and faster data processing, as well as increasing market liquidity, institutional ownership, and greater spending on price discovery, may all have facilitated the discovery, analysis, and aggregation of public and private information. Bai et al. (2016) investigate this by studying the development of informativeness in the US from 1960 to 2012. They provide evidence that the relation between current prices and future earnings has increased in the S&P500 over their study period. Additionally, they argue that this has partially been due to investors producing more private information independently of management disclosures. However, they show that informativeness has not improved for firms outside the S&P500 – an observation they associate with the influx of younger and smaller firms listed on the Nasdaq (Fama & French 2004). Farboodi et al. (2018) investigate this finding further and suggest that it may be due S&P500 firms growing relatively faster than other public companies. Their model implies that investors allocate more data processing shifts to larger firms' data because it is more valuable to process. Thus, rapidly increasing amounts of data could be reinforcing the informativeness of larger firms to the expense of smaller ones.

Finally, it is worth pointing out that inherently uncertain performance may cause a firm's stock price to have poor informativeness regardless of superior disclosure, low arbitrage risk, and favourable macro factors. Fan et al. (2019) show that US firms with more volatile earnings have a weaker relation between current prices and future returns. In a similar vein, Lee (2018) argues that firms with stronger market power and those that invest less in long-term assets or R&D are likely to face less business uncertainty and thus to have more predictable future cash flows. In line with this theory, he shows that US firms with these characteristics tend to have higher FERCs. Following the same logic, he also finds evidence that industries facing greater uncertainty following deregulation tend to suffer a decline in informativeness.

2.4 The economic relevance of price informativeness

By this point, a natural question to ask is: why do stocks' price informativeness matter? In short, the answer lies in research which suggests that information production by public markets can facilitate efficient investment and capital allocation by firms. Dow & Gorton (1997), for example, propose that managers look to stock prices for information regarding their decision-making and adjust their decisions to maximise future stock prices. Thus, a circular feedback mechanism is formed by which corporate decisions cause stock pricing to adjust, which in turn causes corporate decisions to adjust.

In line with this theory, Chen et al. (2007) find that more non-synchronous stock prices have a greater influence on corporate investment, presumably because they contain more private information unknown to managers. However, more non-synchronous stocks are not necessarily priced closer to fundamental value as their fluctuations may not only reflect private information but also noise. Therefore, less synchronicity may also lead to worse investment decisions (Morck et al. 1990). Economic efficiency is only benefitted if stock prices are giving managers new, true information. This is what the FERC measure of informativeness aims to reflect.

As discussed previously, work by Badertscher et al. (2013) implies that private firms' investments are also affected because their managers utilise information generated by public peers. This possibility is also supported by a study of UK private markets by Yan (2020), which shows that investment by private firms increases with market valuations of peer firms. Therefore, structural changes in the level of price informativeness across a market or industry are likely to affect the allocation of capital by both public and private firms.

2.5 Research gaps

This thesis empirically examines the relationship between the price informativeness of US firms and the number of other public firms in their industry. My research adds to discussions about the consequences of the declining number of listings in the US. If there is a link between the number of public peer firms and price informativeness, the US may be experiencing deteriorating market informativeness since most industries have lost public firms. Lower informativeness may, in turn, have real effects on capital allocation and investment efficiency within both the public and private spheres of its economy. In such a case, there would be additional cause to worry about the decline in US listings, regardless of whether it has been caused by new regulation 'breaking' markets or by developments in alternative sources of funding.

Furthermore, a study of how industry listing counts are related to price informativeness has not been performed before, to my knowledge, and fills a gap amongst studies of the determinants of price informativeness. An answer to this question is especially warranted because the relationship has been assumed to be positive by some papers. Chemmanur et al.

(2010), for example, argue that *"the more firms already listed in an industry, the easier it is for outside investors to evaluate a firm in that industry"* when justifying one of their controls. Some subsequent papers have used this to motivate the employment of listing counts as a proxy for informativeness. However, related literature provides hints that informativeness may also face negative pressures from greater industry populations. While examining which forces dominate, my analysis acts as a test of some papers' assumptions too.

Finally, the relation between industry listing counts and informativeness may hint at how information generated by peers is priced into stocks. Thus, my thesis may contribute to literature examining intra-industry information and the way in which investors employ it.

3 Hypotheses

The indications provided by relevant literature are not unanimous about the direction of the relationship between industry listing counts and informativeness. Some suggest that stock price informativeness could be higher in more populated industries because an additional listed firm may attract more analyst coverage and liquidity, as well as increase the quantity of public information relevant to peers. In turn, each of these factors may be positively related to informativeness. However, there are also indications that potential market inefficiencies may mean that additional intra-industry information is not necessarily beneficial to informativeness. Since analyst coverage and liquidity can be estimated but intra-industry information cannot be observed, I develop a two-step hypothesis that allows me to examine the relation between industry listing counts and informativeness while enabling to me to draw inferences about intra-industry information.

The concept that analysts and investors are more likely to research and trade firms in larger industries is mainly based on theoretical models by Subrahmanyam & Titman (1999) and Veldkamp (2006). Subrahmanyam & Titman (1999)'s model implies that analysts should find more serendipitous information when researching firms in a larger industry. Since the serendipitous information helps analysts in their work of evaluating similar firms, they have an incentive to research firms with more peers. This proposition is similar to that of a model developed by Veldkamp (2006). She argues that producing information entails a high fixed cost and thus increasing returns to scale, meaning that, as producers of information, analysts are incentivised to generate information that is more in demand. This is often data that is applicable to more than one firm. Investors like this kind of information because they pay for signals that can be used for multiple trades, reducing their costs. Additionally, Veldkamp (2006) suggest that the more demand there is for a piece of data, the less information producers need to charge for it to cover their fixed costs, implying that investors can buy popular data more cheaply, further boosting demand for it. In essence, the suggestion is that information about firms whose fundamentals better predict the fundamentals of a greater number of other firms is likely to be more used by investors to trade and more covered by analysts. Hameed et al. (2015) show this has been the case in the US.

More populated industries are prime ground for such mechanics. Firms belonging to more populated industries are more likely to have a larger number of similar companies, offering greater opportunities for serendipitous information discovery and data cost amortisation. Thus, they may attract more analysts and more trades from investors. Since these variables are all measurable, I can phrase my first hypothesis as:

H1: *I expect the liquidity and analyst coverage of US firms to be positively related to their industry listing count after the inclusion of relevant control variables.*

If there is evidence in favour of this first hypothesis, I would also expect industry populations to have some positive effect on informativeness. Both of analyst coverage and liquidity have been shown to boost informativeness by Kerr et al. (2020), Chordia et al. (2008), Piotroski & Roulstone (2004), amongst others.

However, there are more mechanisms at play that may disrupt the overall association. An additional listed firm likely increases the pool of intra-industry information. Industries with more public firms may provide analysts and investors with a greater quantity of information useful for evaluating other companies within the industry. Foster (1981) and Han et al. (1989) evidence an instance of investors using such intra-industry information by showing that stock prices respond to the earnings disclosures and management forecasts of industry peers. Assuming the market efficiently reflects new public information (the semi-strong form of market efficiency theorised by Fama (1970)), market participants' aggregate reactions to additional intra-industry signals would cause peer stock prices to contain more information. Thus, stocks in more populated industries would be more informative given their larger information pool. If there is evidence in favour of H1 too, each avenue through which listing counts may affect informativeness would probably have a beneficial effect, and I would expect the overall association to be positive.

Unfortunately, market efficiency has often been challenged (e.g. Watts (1978), Beaver (1968)). If one relaxes the semi-strong version, it becomes possible for the market to incorrectly aggregate intra-industry signals and for the market to only reflect subsets of the available information. Therefore, more information does not necessarily entail more informativeness.

The first scenario implies that market participants, in aggregate, are unable to properly utilise data generated by peers, which is likely to be noisy. Ramnath (2002) and Thomas

& Zhang (2008) identify a temporary instance of this by showing that firms suffer from mispricing in the interval between the first and last firms in their industry disclose quarterly earnings. My thinking goes as follows. While some investors correctly analyse peer data, others may be unable to differentiate between true information and mere noise. If investors are disproportionately influenced by the noise, they will, in aggregate, price in new information incorrectly. Their active and incorrect use of what they believe are useful signals causes their valuations to diverge from fundamentals. Consequently, they may be more likely to misprice the stocks for which they discover more "information" from industry peers, such as those present in more populated industries. In this case, the relation between price informativeness and industry listing counts would have a negative component.

Investors may also just be using a smaller subset of the public information available for firms in more populated industries. Veldkamp (2006) and Hameed et al. (2015) suggest a mechanism through which this may occur. As discussed earlier, Veldkamp (2006)'s model implies that analysts produce more research about firms with more peers, and that this information is cheaper and more demanded by investors because they can use it to trade on a greater number of similar companies. However, she also suggests that, as most investors purchase the low-cost, high-demand data, they perpetuate a scenario in which traders are using the same subset of information. As more investors use the same information to evaluate a given asset type, news about one asset affects other assets' prices, causing asset prices to comove. In line with this thought, Hameed et al. (2015) show that the returns of the more neglected firms in an industry tend to comove most with the returns of peers that have the most analyst coverage and trading activity. They call these peers the "bellweather" firms of the industry. This finding leads me to speculate that more populated industries may have particularly prominent "bellweather" firms because their information is considered so widely applicable. Instead of using the additional signals from peers to complement other information they have about a firm, investors may be more likely to place excessive weight on the signals of certain peers while neglecting other firm-specific information. In this way, the information content of stocks may again decrease as industry population increases.

Given these considerations, my second hypothesis can be expressed as:

H2: *If there is evidence in favour of H1 and the US market is semi-strong efficient in its aggregation of additional public intra-industry information, I expect the price informativeness of US stocks to be positively related to their industry listing counts after the inclusion of relevant control variables.*

As I cannot observe whether the market efficiently aggregates intra-industry information, this hypothesis is not empirically dismissible. Instead, I will use it to propose some inferences about the market's treatment of intra-industry information, under the assumption that I have addressed all the avenues through which listing counts may otherwise affect informativeness.

4 Measuring stock price informativeness with FERC

Since informativeness cannot be directly observed, past literature has developed several methods for its measurement. I follow the FERC approach developed by Collins et al. (1994), Kothari & Sloan (1992), Lundholm & Myers (2002), and Durnev et al. (2003), amongst others.

FERC is based on the classic valuation framework, which explains current excess stock returns as the result of unexpected current dividends and the discovery of new information that adjusts the market's expectations of future dividends. With this framework, the stock prices with more information content are those which better reflect the adjustment of current returns to changes in expected future dividends. By assuming that revisions in expected dividends are correlated with revisions in expected earnings, Collins et al. (1994) express current stock returns as a function of current unexpected earnings and changes in expected future earnings:

$$r_t = b_0 + b_1 UX_t + \sum_{\tau=1}^{\infty} b_{1+\tau} \Delta\xi_t \left(X_{t+\tau}\right) + \varepsilon_t \tag{1}$$

where r_t is the annual buy-and-hold stock return, UX_t is current unexpected earnings growth, and $\Delta\xi_t \left(X_{t+\tau}\right)$ is the change in expectations between time $t-1$ and time t regarding future earnings growth τ years after time t. ε_t are the residual error terms.

Since these independent variables are not observable, Collins et al. (1994) use the current change in earnings (X_t) and the changes in reported future earnings ($X_{t+\tau}$) to proxy for current unexpected earnings and changes in expected future earnings, respectively. However, this approximation introduces an error-in-variables problem. Actual changes in current and future earnings are not equal to changes in their unexpected and expected value, respectively. A portion of actual current earnings is expected, and a portion of future earnings is unexpected. Collins et al. (1994) and Kothari & Sloan (1992) account for this by using future stock returns ($r_{t+\tau}$) as control variables. Since an unexpected shock to future earnings should induce future stock returns, future returns are correlated with the error in measuring expected future earnings. Furthermore, future returns are largely uncorrelated with current returns, the dependent variable, making them a good control for the model's measurement error.

Finally, Collins et al. (1994) show that future earnings beyond year 3 add little explanatory power to the model. They thus cap τ to 3 years. Their model for the relationship between current returns, unexpected current earnings, and changes in expected future earnings therefore takes the form:

$$r_t = b_0 + b_1 X_t + \sum_{\tau=1}^{3} b_{\tau+1} X_{t+\tau} + \sum_{\tau=1}^{3} b_{\tau+4} r_{t+\tau} + \varepsilon_t \tag{2}$$

Drawing from this same underlying literature, Lundholm & Myers (2002) and Durnev et al. (2003)'s approaches to analysing stock price informativeness refer to more informative stocks as those with higher coefficients on future earnings ($b_{\tau+1}$ to $b_{\tau+3}$). These coefficients are named the "future earnings response coefficients", or FERCs for short. A single value for FERC is then calculated by summing each future earnings coefficient.

Admittedly, the FERC method has been criticised for trying to measure changes in expectations in a way that remains exposed to bias by factors such as trading constraints and management withholding information. Alternative measures based on the relation between price and future earnings have occasionally been used, such as by Bai et al. (2016), but are less researched and are also not without flaws. As it remains the most studied and commonly used measure in price informativeness literature, I choose to employ FERC for this paper.

Lundholm & Myers (2002)'s method for analysing informativeness:

Specifically, I will use Lundholm & Myers (2002)'s variation of Collins et al. (1994)'s model in this thesis. To study the relation between corporate disclosures, their variable of interest, and informativeness, Lundholm & Myers (2002) extend model 2 by interacting their disclosure variable and their controls with firm-year earnings and returns. As the number of parameters quickly gets very large with this approach, they condense the three years of future returns and future earnings into two variables: E3 and R3. Also, under the assumption that earnings follow an autoregressive process, they use past earnings (E_{t-1}) to control for the measurement error in unexpected current earnings. The regressions they run therefore take the form:

$$r_t = b_0 + b_1 E_{t-1} + b_2 E_t + b_3 E3_t + b_4 R3_t + \varepsilon_t \tag{3}$$

for the basic relationship between current returns and future earnings, and

$$\begin{aligned} r_t = b_0 &+ b_1 E_{t-1} + b_2 E_t + b_3 E3_t + b_4 R3_t \\ &+ b_5 \Phi_{t,i_3} + b_6 \Phi_{t,i_3} * E_{t-1} + b_7 \Phi t, i_3 * E_t + b_8 \Phi t, i_3 * E3_t + b_9 \Phi t, i_3 * R3_t \\ &+ \sum b_n \Psi_t * (1 + E_{t-1} + E_t + E3_t + R3_t) + \varepsilon_t \end{aligned} \tag{4}$$

for the models used to examine their variable of interest.

E_t are current earnings scaled by the market value of equity in $t-1$, and $E3_t$ is the sum of earnings for the three years following year t, scaled by the market value of equity in $t-1$. FERC has therefore been condensed into one coefficient: b_3. $R3_t$ is the buy-and-hold return from t to $t+3$. Φ and Ψ are the variable of interest and vector of controls respectively. If the coefficient on the interaction term between E3 and Φ is positive, the variable of interest has a positive effect on FERC and thus on informativeness.

5 Methodology

To test my hypotheses, I run several ordinary least squares (OLS) regressions. In this section, I will detail the specifics of these models. My first regressions examine how industry listings are related to liquidity and analyst coverage. Then, I focus on the main purpose of this thesis by testing the association between industry listings and informativeness.

5.1 Specifications for the analyst coverage and liquidity models

To test whether analyst coverage and stock liquidity increase with industry population, I run two OLS regressions with the forms:

$$NEST_t = b_0 + b_1 LC_t + b_2 S_t + b_3 TO_t + b_4 Age_t + YEAR + \varepsilon_t \tag{5}$$

and

$$TO_t = b_0 + b_1 LC_t + b_2 S_t + b_3 NEST_t + b_4 M/B_t + b_5 IndM/B_t + YEAR + \varepsilon_t \tag{6}$$

where $NEST_t$ is the natural logarithm of the average number of analyst estimates made for a firm at various points in year t, representing analyst coverage. TO_t is the natural logarithm of a firm's share turnover in year t, representing liquidity. LC_t is the natural logarithm of the number of public companies present within a firm's primary industry at the end of year t. S_t is a firm's market capitalisation at the end of year t. M/B_t is a firm's market-to-book (M/B) ratio at the end of year t, and $IndM/B_t$ is the average M/B ratio of the firm's industry the same year. $YEAR$ is a vector of year dummy variables and ε_t are the residual error terms. The specific descriptions and calculations of each of these features are detailed in section 6.3. If listings counts are positively related to analyst coverage and liquidity, as I expect, the b_2 coefficients in both models should be positive.

Choice of control variables

To obtain unbiased estimates of the coefficients I am studying, I control for variables that may cause changes in analyst coverage and liquidity while also being associated with industry listings. I include size as a control in both regressions because literature has shown that larger firms tend to attract more analysts and be more liquid (Bhushan 1989, Lipson & Mortal 2007). Changes in size have also been shown to have an association with declining listing counts in the US (Doidge et al. 2017, Gao et al. 2013). Liquidity and analyst coverage are added to each other's regressions because Alford & Berger (1999) suggest that more analysts follow more traded firms, possibly due to greater commissions. Age is controlled for in the analyst coverage regression as the number of analysts following a stock may accumulate over time, like industry listings. I also adjust for M/B ratios in the liquidity regressions because higher industry valuations may attract more industry IPOs (Pagano et al. 1998, Ritter 1984, Loughran et al. 1994) while being linked to greater investor sentiment and activity (Lowry 2003).

Year fixed effects

In addition, I include time fixed effects in the form of year dummies to both models to account for any time trends in industry listings, analyst coverage, and liquidity. Breusch & Pagan (1979)'s Lagrange-Multiplier test reveals the presence of significant time fixed effects in both regressions, indicating that a fixed effect model is better than pooled OLS here.

Standard errors

Finally, I adjust my models' standard errors. Inadequate standard errors may influence the statistical significance associated with the independent variables, potentially leading to spurious conclusions being drawn from the results. First, I test for heteroskedasticity, which arises when the variance of the error term is not constant across an independent variable's values, using Breusch & Pagan (1979)'s heteroskedasticity test. The null hypothesis of homoscedasticity is rejected, so I choose to employ heteroskedasticity-robust covariance estimates and standard errors, following White (1980).

However, heteroskedasticity-robust standard errors may still fall short when clustered observations have correlated error terms. I follow Abadie et al. (2017), who suggest that standard errors should be clustered when sampling is performed at the clustered level or when groups of units are assigned to a treatment. This study does not use cluster sampling, so the first motivation for clustered standard errors is not applicable. However, groups of firm-year observations are assigned to industries in the formation of LC_t, my variable of interest. Therefore, I choose to cluster my model's standard errors by the industry classification used for calculating industry listing counts.

5.2 Specification for the informativeness models

As discussed in section 4, I choose to use Lundholm & Myers (2002)'s extension of the FERC regression developed by Collins et al. (1994) as a base for my investigation of the relation between industry listing counts and firm informativeness. Therefore, my OLS models can generally be described by:

$$\begin{aligned} r_t = {} & b_0 + b_1 E_{t-1} + b_2 E_t + b_3 E3_t + b_4 R3_t \\ & + b_5 LC_t + b_6 LC_t * E_{t-1} + b_7 LC_t * E_t \\ & + b_8 LC_t * E3_t + b_9 LC_t * R3_t \\ & + \sum b_n X_t * (1 + E_{t-1} + E_t + E3_t + R3_t) + \varepsilon_t \end{aligned} \tag{7}$$

where r_t, E_t, $E3_t$, and $R3_t$ are current returns, current earnings, 3 years of future earnings, and future returns over 3 years, respectively. LC_t is the variable I am studying - the natural logarithm of the number of public companies present within a firm's primary industry at the end of year t. X represents a vector of control variables, including firm size, industry concentration, market-to-book ratio, industry market-to-book-ratio, proportion of intangible assets, firm age, average industry age, a loss indicator, 1-digit SIC industry dummies, and year dummies. The specific descriptions and calculations of each of these features are detailed in section 6.3. ε_t are the residual error terms. As each control and variable of interest must be interacted multiple times, my specification quickly gets very lengthy. However, more compact regressions featuring cross-sectional industry estimations of FERC as a dependent variable, as employed by Durnev et al. (2003), are unfortunately not meaningful for this analysis. [5]

The main interpretation of model 7 relevant to my analysis relies on coefficients b_3 (FERC) and b_8. As mentioned earlier, the listing count variable, LC_t is a natural logarithm. Therefore, its minimum value will be 0 when a firm's industry's listing count is 1. In a model without controls, b_3 is thus the FERC of firms that are alone in their industry. b_8 instead represents the relation between a 1% increase in industry listing count and a hundredth change in FERC. Therefore, a coefficient of 10, for example, would signify that a 1% increase in industry listing count is related to a 0.1 ($\frac{10}{100}$) increase in FERC. If the informativeness of a firm's stock improves with the number of public companies in its industry, b_8 should be positive. In combination, b_3 and b_8 allow us to determine the expected magnitude of FERC given different levels of industry listing counts.

[5] The independent variable of interest, industry listing counts, would be extremely correlated with the number of observations used to form the industry observations of FERC, and thus with the bias in industry informativeness estimates. The coefficients on listing counts would thus not be reliable. Lundholm & Myers (2002)'s more bulky specification side-steps this issue by not using FERC estimates as dependent variables.

Choice of control variables

To isolate the effect of industry listing counts on informativeness, variables that cause informativeness to change while also being correlated with industry listing counts should be controlled for. Since the coefficient of interest aims to reflect the interaction term between *expected* future earnings and listing counts, an unbiased specification requires that controls are also applied to the variables in the basic FERC model that are used to adjust for various measurement errors. Therefore, each control variable is interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. While obtaining a perfect set of controls is not possible due to unavailable data and the risk of overspecification, the following ones are included in my models.

My first set of controls relate to firm size and market power. Doidge et al. (2017) and Gao et al. (2013), amongst others, show that listings have declined in the US due to public companies performing more acquisitions, causing them to become larger and their industries to become more concentrated (Grullon et al. 2019). Thus, these variables likely exhibit a negative relationship with listing counts. Larger firms are also likely to have a more developed information environment (Farboodi et al. 2018), gain attention from more institutional investors and research analysts (Bhushan 1989, Lipson & Mortal 2007), as well as become less exposed to noise trader risk (Shleifer 2000). On the other hand, firms in more concentrated industries may have less uncertain cash flows (Lee 2018). Each of these relationships suggests that size and concentration affect FERC. In fact, Dávila & Parlatore (2018) and Lee (2018) find that larger firms and those in more concentrated industries tend to have more informed stock prices. Therefore, lower industry listings may coincide with, but not cause, greater firm size and industry concentration, which may lead to greater informativeness.

Second, I control for factors related to earnings timeliness, predictability, and growth prospects. This includes M/B ratios and the proportion of intangible assets. The M/B ratio is often used as a proxy for the growth opportunities of a firm as the market is offering a strong valuation to the firm's future earnings relative to its current assets. Durnev et al. (2003) argue that higher growth prospects may be linked to higher FERC measures as the market gives more weight to future earnings. The stocks of firms with a higher M/B ratio may thus have a weaker relation with current earnings, a stronger relation with future earnings, and hence a greater FERC. However, firms with greater growth prospects and intangible assets also tend to have riskier future cash flows, possibly making their businesses harder to value and their stock prices less informative (Lee 2018). Furthermore, standard accounting tends to be less informative about intangible assets, meaning investors face more difficulty evaluating such companies (Doidge et al. 2018), hurting their FERC. Intangible assets and M/B ratios may also be related to industry listing counts. Doidge et al. (2018) suggest that firms facing more uncertainty, particularly those with more intangible assets, draw more benefit from private rather than public markets, potentially making their industries less populated. On the other hand, the serendipitous information generated by public markets may also be more useful to firms facing greater uncertainty, increasing their industries' size (Subrahmanyam & Titman 1999). Given the seemingly ambiguous and intertwined relationship between M/B ratios, intangible assets, informativeness, and industry listings, I control for these factors.

Furthermore, the M/B ratio also controls for company valuations, which may affect listing counts and informativeness at both the firm and industry-level. Lowry (2003) shows that greater IPO volume in the US is related to periods of overvaluation and stronger investor sentiment. This seems particularly true within industries. Studying markets across the world, Pagano et al. (1998), Ritter (1984) and Loughran et al. (1994) amongst others, document periods of IPO clustering amongst firms in sectors for which valuations were high. Additionally, declining industries with poor prospects may see both lower valuations and listing counts. Valuations may also be linked to informativeness. High prices seem to reflect stronger investor sentiment and a weaker adherence to company fundamentals (Baker & Wurgler 2007, Lowry 2003), possibly causing mispricing. Pagano et al. (1998) also provide evidence of this within clusters of Italian IPOs. Given these effects, a measure of industry M/B ratios is added to the model too.

Firm age is included as a control at both the firm and industry level. Older firms are more likely to have an established information environment, which may make it easier for investors to evaluate new information and accurately price a stock (Farboodi et al. 2018). Older industries may also have accumulated a deeper pool of information. Furthermore, they may contain more listed firms, as there has been more time for firms to grow and go public, or fewer listed firms, if the sector is in decline.

Following Lundholm & Myers (2002), I also include a loss indicator, which identifies whether earnings are negative or not, as it has been shown to affect the returns-earnings relation. It appears that it is harder for investors to anticipate future earnings in loss-making companies.

Finally, to account for time-invariant industry differences that may affect FERC, such as industry-specific regulation or systematic uncertainty, 1-digit SIC industry dummies are included in the regression. This captures broad industry differences while retaining an important number of degrees of freedom.

Year fixed effects

Despite the inclusion of control variables, unobserved year-specific effects or time trends in FERC and informativeness may still introduce bias to my models' coefficients of interest. These could include phenomena such as changes in macroeconomic conditions (Lowry 2003, La Porta et al. 1997) or technological developments which facilitate the distribution and analysis of information (Bai et al. 2016). To deal with this, I choose to employ year fixed effects, which demean observations for each variable by subtracting their average value across all observations that year. Breusch & Pagan (1979)'s Lagrange-Multiplier test reveals the presence of significant time fixed effects, indicating that a time fixed effect model is better than a pooled OLS model for this analysis.

Standard errors

As previously, I test for heteroskedasticity using Breusch & Pagan (1979)'s heteroskedasticity test. The null hypothesis of homoscedasticity is rejected, so I use White (1980)'s standard errors. Since an interaction with LC_t is my variable of interest, I also cluster at the industry level used to calculate listing counts again.

Mediator variables

A final point to address is the inclusion of mediator variables - factors through which industry listing counts may affect price informativeness. As discussed in the formation of my hypotheses in section 3, analyst coverage and liquidity are likely mediator variables. Whether to include these factors in my main model is a dilemma with no correct answer. On the one hand, adding them may prevent the model from revealing the full effect of industry populations. Part of the relationship between industry listings and informativeness would be absorbed by the mediators. On the other hand, excluding them risks introducing omitted variable bias because variations in the mediators may be caused by factors other than industry listings. Given these issues, I will run models both with and without mediators. Examining the models that include mediators will be particularly useful when drawing possible inferences about intra-industry information, since these regressions will have accounted for the other avenues through which industry populations may affect informativeness.

6 Dataset and Variables

In this section, I will describe the variables used for my chosen specifications and the sources from which I have retrieved the necessary data.

6.1 General sample construction and main exclusions

The underlying dataset used for this thesis is comprised of the characteristics and financials of US firms that have had common stock listed on the NYSE, Nasdaq or Amex at any time during the period from 1976-2016. This data is then transformed into the various features used in my models.

I begin by pulling monthly observations for all listed securities in the Center for Research in Security Prices (CRSP) database from 1975-2020. From these, I select only year-end observations, effectively removing any securities that were listed and delisted in the same year. Since informativeness measures are based on annual observations, the exclusion of these securities would have occurred anyway. Next, I filter based on CRSP's exchange codes and share codes. Stocks listed on the NYSE, Nasdaq or Amex have exchange codes 1, 2 or 3. Common stocks of firms incorporated in the US have share codes 10 or 11. At this point, many firms (identified by CRSP's "PERMCO" variable) in the dataset still have multiple stocks per time period. I drop duplicate firm-year observations to ensure companies are only included once per year.

The resulting annual, firm-level dataset from CRSP is then merged with annual observations of firms' financials (and other characteristics) from Compustat using the CRSP/ Compustat merged database. Unfortunately, CRSP and Compustat do not use a shared identification method for firms. In cases where the merged database suggests multiple Compustat identifiers for a single CRSP company, the primary links are chosen based on the merged database's LINKPRIM variable (equal to P or C) the link's validity for a given year (LINKDT and LINKENDDT). The "Link research complete" (LC) LINKTYPE is then preferred for any remaining duplicates. This approach helps ensure that only the most reliable connections are made between the two databases. However, out of the 22,380 unique firms in the final CRSP dataset, 1,918 were not reliably matched with a Compustat identifier. To ensure that they are not understated, industry listing counts are calculated on the full CRSP dataset. The calculation and provenance of this variable, and my other ones, are described in more detail later in this section. For my measure of analyst coverage, I use a linking table provided by Wharton Data Research Services (WRDS) for matching CRSP firms to observations from I/B/E/S, Refinitiv's Institutional Brokers Estimate System.

After the creation of the required features, I exclude several industries, based on Standard Industry Classification (SIC) codes. For a more detailed explanation of SIC codes, please see section 6.2. As FERC is estimated with company earnings, the accounting measures used for the companies in my sample must be relatively similar. As is common in the past literature (Durnev et al. (2003), Gelb & Zarowin (2002) etc.), I exclude financial companies (SIC codes 6000-6799) because their accounting standards and return measures differ significantly from those of other firms. I drop non-classifiable industries (SIC codes 9900-9999) as well, given the extremely eclectic nature of the firms included therein.

Finally, firm-year observations with missing values for the variables needed in my model are removed. The most relevant of these are firm-year observations without the five years of consecutive earnings or four year of consecutive returns required to estimate FERC. The resulting dataset is an unbalanced panel of 94,714 firm-year observations, consisting of years ranging from 1976 to 2016 and 9,690 unique firms present within 378 unique primary 3-digit SIC industries. Table 1 presents the number of observations available for each year.

Table 1: Number of industry and firm-year observations by year

Year	Unique 3-digit industries	Firm-year observations
1976	230	1617
1977	230	1563
1978	227	1493
1979	225	1422
1980	220	1359
1981	216	1287
1982	235	2029
1983	239	2188
1984	240	2367
1985	257	2427
1986	267	2437
1987	270	2552
1988	270	2618
1989	269	2534
1990	268	2525
1991	267	2544
1992	273	2609
1993	279	2733
1994	284	2864
1995	287	2821
1996	289	2789
1997	290	2800
1998	290	2693
1999	289	2645
2000	286	2681
2001	275	2883
2002	269	2872
2003	273	2753
2004	269	2596
2005	275	2524
2006	272	2482
2007	271	2463
2008	270	2374
2009	268	2280
2010	269	2225
2011	261	2164
2012	251	2050
2013	249	1952
2014	252	1902
2015	250	1842
2016	245	1755
Total	**378**	**94,714**

6.2 Industry classifications

When assigning a firm to an industry, I use each firm's primary Standard Industry Classification (SIC) code, imported from CRSP, for classification. Since I must use CRSP data to get an accurate estimate of listing counts, I am also bound to the two industry classification systems used in the database – SIC and NAICS (North American Industry Classification System). While it has been argued that NAICS codes are better at dealing with the emergence of new industries, CRSP only offers them from 2004. Therefore, I choose to employ SIC codes as they greatly expand my sample. However, I perform a robustness test that instead uses NAICS. It does not meaningfully alter my results.

SIC codes classify industries using 4 digits. The first two digits provide a very broad industry allocation. Each progressive digit represents a more specific industry subclassification. For example, SIC code 48 refers to a business in the "Communications" industry. A code of 481 adds a level of detail, identifying "Telephone Communications" firms, and a code of 4812 goes even deeper, representing "Radiotelephone Communications".

To give an idea of the industries in this study's dataset, Table 2 presents the numbers of observations by their 2-digit SIC code title. Unsurprisingly, the largest portion of the sample belongs to the Manufacturing and Services industries. There are no observations for Agriculture, Forestry and Fishing and Public Administration because the firm-year observations within these industries that were public were in the unclassifiable group, which was dropped, or had missing values for necessary variables. As explained previously, firms within the Finance, Real Estate and Insurance industry are excluded, and thus have no observations.

Table 2: Number of observations by broad industry grouping

SIC range	Industry group	Unique 3-digit industries	Unique firms	Firm-year obs
01-09	Agriculture, Forestry, Fishing	0	0	0
10-14	Mining	23	674	5,579
15-17	Construction	14	184	1,551
20-39	Manufacturing	156	5,023	50,658
40-49	Transportation and Public Utilities	44	900	9,195
50-51	Wholesale Trade	19	607	4,627
52-59	Retail Trade	47	873	7,521
60-67	Finance, Insurance, Real Estate	0	0	0
70-89	Services	77	2,380	15,919
91-99	Public Administration	0	0	0

6.3 Variable description and measurement

The calculation and provenance of my dataset's features are described below. However, one preliminary note to make is that when I refer to time t, I mostly mean the fiscal year-end of year t. This is necessary as some Compustat datapoints are only reported at fiscal year-end. I will specify any instances where I instead refer to the calendar year-end. The mismatch in some year-ends is later addressed by a robustness check.

Variables of interest:

Stock returns (r_t and $R3_t$): r_t is the buy-and-hold return for the stock for year t, adjusted for stock dividends and stock splits. It is calculated by dividing the sum of stock-split-adjusted share price at time t and the stock-split-adjusted dividends per share received from $t-1$ to t, by the stock-split-adjusted share price at $t-1$. $R3_t$ is the buy-and-hold return from t to $t+3$. It is calculated by dividing the sum of the stock-split-adjusted share price at time $t+3$ and sum of the stock-split-adjusted dividends per share made in the period from t to $t+3$, by the stock-split-adjusted share price at time t. To avoid extreme outliers affecting my results, I follow Lee (2018) and winsorise stock return variables at the 1st and

99^{th} percentiles. Stock prices, dividends per share, and stock split adjustment factors are all from Compustat.

Earnings (E_t) and ($E3_t$): E_t are earnings generated in the period from $t-1$ to t, scaled by the market value of equity at $t-1$. $E3_t$ is the sum of earnings generated from t to $t+3$, scaled by the market value of equity at $t-1$. To minimise discrepancies caused by accounting choices and capital structure I follow Durnev et al. (2003) and use earnings-before-interest-tax-depreciation-and-amortisation (EBITDA) as my measure of earnings. A robustness check with earnings-before-interest-and-tax (EBIT) does not meaningfully change my findings though. The market values of equity are calculated by multiplying the time $t-1$ share prices by the number of common shares outstanding at time $t-1$. Collins et al. (1994) suggest scaling earnings by either the beginning-of-period earnings or market value. The latter may further help mitigate FERC's error-in-variables problem as prices reflect market expectations (Kothari & Sloan 1992). Like Lundholm & Myers (2002) and others, I choose to use the beginning-of-period market capitalisation rather than beginning-of-period earnings, as it also avoids issues with scaling the earnings of firms that have non-positive earnings. To avoid extreme outliers affecting my results, I follow Lee (2018) and winsorise stock return variables at the 1^{st} and 99^{th} percentiles. Company earnings and shares outstanding are drawn from Compustat.

Industry listing counts (LC_t): I calculate industry listings by counting the number of unique firms within the same primary 3-digit SIC industry that have a listed security covered by CRSP at the calendar end of year t. I choose the 3-digit level as the 4-digit level may be too granular and may not capture information effects common to slightly more loosely related firms. Using 4-digit SIC industry listing counts does not alter my results though. CRSP identifies unique firms with its "PERMCO" identifier, which remains constant for firms over time even if they delist and relist. Using CRSP is likely the best way to track industry-level listings in the NYSE, Nasdaq and Amex as it maintains one of the largest and most reliable databases for these exchanges. It is very commonly used in academia. Since listing counts are very positively skewed, I adjust the distribution of this variable by taking its natural logarithm.

Mediator variables:

Analyst coverage ($NEST_t$): I estimate a company's analyst coverage with the natural logarithm of the average number of earnings estimates that analysts have produced about it during calendar year t. This data is obtained from the summary detail database of I/B/E/S. Following Piotroski & Roulstone (2004), I assume the number of forecasts is zero if I/B/E/S has not recorded any analyst estimates for a company in year t.

Liquidity (TO_t): I estimate a firm's liquidity via its share turnover. This is calculated by dividing the total number of a firm's common shares that were traded during calendar year t by its average number of outstanding common shares during the same year. As this yields a strongly positively skewed distribution, I then scale via natural logarithm. Share volumes and shares outstanding are drawn from CRSP. Firm-year observations where no shares were traded are dropped and extremely positive outliers are winsorised at the 1^{st} and 99^{th} percentiles.

Control variables:

Firm size (S_t): To measure firm size, I use inflation-adjusted market capitalisation at time t. This is calculated by multiplying time t share prices by the number of common shares outstanding at time t (from Compustat) and then correcting for US Producer Price Index (PPI) inflation, provided by the US Bureau of Labour Statistics. To adjust for the positive skew of firm size, I take its natural logarithm.

Industry concentration (HHI_t): As is common, I measure industry concentration using the Herfindahl-Hirschman Index (HHI) - the sum of the squared market share of each firm within its industry. The higher the HHI, the more concentrated the industry. Following Grullon et al. (2019), I calculate the HHI for each 2-digit SIC industry using yearly revenue data for all firms in the Compustat database over the study period. A firm's market share at time t is calculated as its sales at time t divided by its primary industry's total annual revenue in year t. To adjust for HHI's positive skew, I take its natural logarithm.

Unfortunately, this measure of concentration is not perfect. Compustat does not have extensive coverage of private firms, for example, which may have significant market shares in certain industries. The simplicity of the measure also fails to account for geographic considerations and instances where firms have an important presence in specific industry segments. However, better measures of concentration are not available for most US industries. As it is merely a control in this study, a Compustat-based estimation of HHI should be sufficient. I use a 2-digit industry classification for my HHI calculation as the HHII's limitations become more significant at more granular SIC levels. A 2-digit SIC serves its purpose as an indication of the fragmentation in a firm's broader competitive environment while not exacerbating the issues with the HHI computation.

Market-to-Book ratio by firm (M/B_t) and industry ($IndM/B_t$): M/B_t is calculated by dividing market capitalisation at time t by the book value of equity at time t. As before, market capitalisation is the product of stock price and number of outstanding shares at time t. Equity book values are retrieved from Compustat. Observations with negative book values are meaningless and are excluded. Industry M/B ratio is then calculated by taking the mean ratio within 3-digit SIC industries.

Intangible assets (Int_t): Int_t is calculated by pulling intangible balance sheet assets at time t from Compustat and scaling them by total firm assets at time t, also retrieved from Compustat, to account for differences in firm size.

Firm age (Age_t) and mean industry age ($IndTO_t$): I estimate the age of a firm by the number of years since the stock was first included in the CRSP database at the calendar end of year t, as suggested by Fama & French (2001). $IndTO_t$ is the mean age of firms within 3-digit industries.

Loss indicator (L_t): L_t is a dummy that indicates whether a firm's EBITDA was negative in year t or not. Negative EBITDAs are given a loss indicator value of 1.

6.4 Summary statistics and correlations

Table 3 displays the summary statistics for my sample. The median of the log of 3-digit listing counts is 3.43, equivalent to an actual listing count of 31. Since the minimum logged population is 0, some 3-digit industries only contain a single firm. A few industries are very populated, containing up to a maximum 721 listed firms (the log of which is 6.58). This is SIC code 737 - Computer Programming, Data Processing, and other Computer Related Services in the year 2000. The presence of such mega-industries questions whether a full-sample regression can capture a non-linear relationship characterised by declining benefits to new lists. I investigate this with a cross-sectional test. The medians of E_{t-1} and E_t are similar, and roughly a third of the size of the median of the next 3 years' earnings. Similarly, the median current returns (8%) are roughly a third of the median returns over the next 3 years. This suggests there are no structural changes in these variables over the sample.

Table 4 presents the Pearson correlations between the same variables. Some interesting relations stand out. As theorised by the classic valuation framework, earnings and returns are positively correlated. Consistent with Collins et al. (1994), future returns are also quite related to future earnings, but unrelated to current returns, confirming their role as a control in the FERC regression. As expected, listing counts are inversely related to industry concentration and positively related to industry valuations. Furthermore, they seem associated with lower and even negative earnings, suggesting that the quality of firms may decrease with industry population. It is also interesting to note that more populated industries seem to contain more young firms, possibly due to the emergence of new industries. Finally, listing counts are positively correlated with liquidity but not with analyst coverage. However, these relationships may be confounded by certain variables. In particular, liquidity and analyst coverage appear related to each other and firm size.

Table 3: Summary statistics of 94,714 firm-year observations

	Min	25^{th}	Median	75^{th}	Max	Mean	SD
r_t	-0.860	-0.190	0.080	0.380	3.230	0.180	0.640
$R3_t$	-0.950	-0.290	0.200	0.780	7.400	0.470	1.270
E_{t-1}	-0.920	0.050	0.130	0.240	1.300	0.150	0.240
E_t	-0.740	0.060	0.140	0.270	1.550	0.180	0.260
$E3_t$	-1.490	0.180	0.500	0.970	6.840	0.710	1.030
LC_t	0	2.640	3.470	4.620	6.580	3.530	1.370
$NEST_t$	0	0	1.340	2.200	4.020	1.350	1.050
TO_t	-7.970	-1.220	-0.420	0.320	2.040	-0.480	1.140
S_t	-2.220	3.280	4.780	6.360	12.630	4.860	2.150
HHI_t	4.190	5.880	6.290	6.740	9.210	6.330	0.790
MB_t	0.040	1.510	2.870	5.420	49.780	4.900	6.810
$IndMB_t$	0.140	2.570	4.260	6.490	49.780	4.890	3.310
Int_t	0	0	0.020	0.140	1	0.100	0.160
Age_t	0	6	13	24	91	17.710	16.530
$IndAge_t$	0	10.400	14.750	21.740	91	17.080	9.330
L_t	0	0	0	0	1	0.160	0.360

Table 4: Pearson correlations

	r_t	$R3_t$	E_{t-1}	E_t	$E3_t$	LC_t	$NEST_t$	TO_t	S_t	HHI_t	M/B_t	$IndM/B_t$	Int_t	Age_t	$IndAge_t$	L_t
r_t	1															
$R3_t$	-0.01	1														
E_{t-1}	0.06	0.1	1													
E_t	0.26	0.11	0.77	1												
$E3_t$	0.27	0.33	0.6	0.76	1											
LC_t	0	-0.02	-0.22	-0.21	-0.17	1										
$NEST_t$	-0.05	-0.05	0.02	-0.02	-0.06	-0.01	1									
TO_t	0.07	-0.09	-0.21	-0.2	-0.18	0.13	0.41	1								
S_t	0.11	-0.11	0.08	0.06	0	-0.08	0.83	0.34	1							
HHI_t	0.01	0.02	0.03	0.04	0.05	-0.33	-0.12	-0.02	-0.15	1						
M/B_t	0.21	-0.1	-0.26	-0.24	-0.2	0.12	0.11	0.25	0.2	-0.05	1					
$IndM/B_t$	0.07	-0.1	-0.34	-0.34	-0.31	0.24	0.13	0.36	0.16	-0.11	0.49	1				
Int_t	-0.03	-0.03	-0.06	-0.06	-0.06	-0.03	0.18	0.21	0.22	-0.05	0.11	0.28	1			
Age_t	-0.02	-0.02	0.16	0.13	0.07	-0.2	0.3	-0.02	0.44	-0.11	-0.05	-0.05	0.02	1		
$IndAge_t$	-0.01	-0.01	0.17	0.14	0.09	-0.35	0.21	-0.02	0.32	-0.2	-0.05	-0.1	-0.01	0.57	1	
L_t	-0.11	-0.04	-0.44	-0.53	-0.35	0.22	-0.23	0.07	-0.3	-0.05	0.21	0.21	-0.07	-0.21	-0.18	1

7 Results

In this section, I present and describe the results of the models specified in section 5. I also perform some cross-sectional tests of the extended FERC model to examine the relationship between industry listings and FERC over different types of firms and time periods.

7.1 Results of the analyst coverage and liquidity models

Before getting to the main results of this thesis – those reflecting the relationship between industry listing counts and informativeness – I check that the mediator variables behave in the way I expect. Namely, my first hypothesis anticipated that analyst coverage and liquidity are increasing with industry population. Table 5 presents the results of the relevant models.

I start with the regressions that have analyst coverage, represented by $NEST_t$ as their dependent variable. In the univariate regression, we see that industry listing counts, represented by LC_t, have a statistically significant inverse relationship with analyst coverage. However, after the inclusion of control variables, the coefficient on LC_t becomes positive with statistical significance beyond the 1% level. Since both $NEST_t$ and LC_t are natural logarithms, the magnitude of the coefficient takes meaning in percentage terms. For a 1% increase in industry listing counts, the number of analyst estimates made for a firm increases by 0.017%. Turning to the models for liquidity, represented by TO_t, we see a positive relationship, significant beyond the 1% level, for both the univariate regression and the one with controls. Again, both TO_t and LC_t are natural logarithms. Therefore, a 1% increase in industry listings counts appears associated with a 0.082% increase in share turnover. All-in-all, the relationships between industry population and the mediator variables appear to be consistent with my first hypothesis. These results are robust across multiple versions of the industry listings variable (Tables A.1 and A.2 in the Appendix). However, it is worth noting that the adjusted R^2 statistics on the univariate regressions for both TO_t and $NEST_t$ are extremely small – 0.00 and 0.02 respectively. Industry listing counts appear to explain a tiny portion of the variation of the mediator variables. Overall, the takeaway seems to be that, although analyst coverage and liquidity are significantly positively related to industry population, this association is unlikely to be economically relevant.

7.2 Results of the extended FERC models over the whole sample

Having found evidence in favour of the first hypothesis, I now move onto the main purpose of this thesis – examining the relationship between informativeness and industry listing counts. Table 6 displays the incrementally-constructed extended FERC model used

Table 5: Regressions of analyst coverage and liquidity on industry listing counts

$NEST_t$ represents the analyst coverage and TO_t represents liquidity. Both are natural logarithms. LC_t are logged industry listing counts. S_t is market capitalisation and Age_t is the number of years since a firm was first listed. M/B and $IndM/B$ are firm and industry M/B ratios, respectively. All variables are as described in section 6.3. The coefficient of interest is on LC_t.

	Dependent variable:			
	$NEST_t$		TO_t	
LC_t	-0.005^{**}	0.017^{***}	0.108^{***}	0.082^{***}
	(0.002)	(0.001)	(0.003)	(0.002)
S_t		0.401^{***}		-0.054^{***}
		(0.001)		(0.003)
TO_t		0.143^{***}		
		(0.002)		
Age_t		-0.002^{***}		
		(0.0001)		
$NEST_t$				0.437^{***}
				(0.006)
M/B_t				0.014^{***}
				(0.001)
$IndM/B_t$				0.017^{***}
				(0.001)
Constant	1.372^{***}	-1.018^{***}	-0.863^{***}	-1.793^{***}
	(0.009)	(0.017)	(0.010)	(0.023)
Year fixed effects	No	Yes	No	Yes
Adjusted SE	Yes	Yes	Yes	Yes
Observations	94714	94714	94714	94714
Adjusted R^2	0.00	0.73	0.02	0.37
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6: Full sample: Extended FERC model with incremental controls

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Firm size controls are based on market capitalisation. HHI controls reflect industry concentration. M/B ratio controls include firm and industry M/B ratios. Other controls include firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. The coefficient of interest is on $LC_t{*}E3_t$.

	FERC model	Incremental Models					Full model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
E_{t-1}	−0.963***	−0.532***	0.138*	−0.047	0.313	1.114***	0.434
	(0.023)	(0.061)	(0.075)	(0.252)	(0.247)	(0.299)	(0.303)
E_t	0.805***	0.602***	0.573***	0.937***	0.841***	−0.036	0.976***
	(0.026)	(0.072)	(0.085)	(0.274)	(0.267)	(0.318)	(0.329)
$E3_t$	0.188***	0.224***	0.091***	0.004	−0.032	−0.047	−0.055
	(0.005)	(0.015)	(0.018)	(0.053)	(0.050)	(0.061)	(0.062)
$R3_t$	−0.091***	−0.101***	−0.073***	−0.075***	−0.086***	−0.040	−0.036
	(0.002)	(0.005)	(0.006)	(0.020)	(0.020)	(0.027)	(0.028)
LC_t		0.029***	0.033***	0.037***	0.021***	0.010***	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
$LC_t * E_{t-1}$		−0.115***	−0.142***	−0.137***	−0.090***	−0.077***	−0.037*
		(0.017)	(0.017)	(0.018)	(0.019)	(0.019)	(0.019)
$LC_t * E_t$		0.058***	0.054***	0.046**	0.053**	0.048**	0.068***
		(0.020)	(0.020)	(0.021)	(0.021)	(0.022)	(0.023)
$LC_t * E3_t$		−0.010**	−0.007*	−0.006	−0.008*	−0.004	−0.005
		(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
$LC_t * R3_t$		0.002	0.002	0.002	0.003*	0.003*	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Size controls	No	No	Yes	Yes	Yes	Yes	Yes
HHI controls	No	No	No	Yes	Yes	Yes	Yes
M/B ratio controls	No	No	No	No	Yes	Yes	Yes
Other controls	No	No	No	No	No	No	Yes
Mediators	No	No	No	No	No	No	No
Industry dummies	No	No	No	No	No	Yes	Yes
Year fixed effects	No	No	No	No	No	Yes	Yes
Adjusted SE	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	94714	94714	94714	94714	94714	94714	94714
Adjusted R^2	0.16	0.16	0.19	0.19	0.24	0.34	0.38

Notes: *p<0.1; **p<0.05; ***p<0.01

to analyse this association.

Looking first at the standard FERC regression (regression 1 in the table), one sees that trailing earnings (E_{t-1}) and future returns ($R3_t$) are significantly decreasing with current returns, while current and future earnings (E_t and $E3_t$) have a significantly positive relation to current returns. This is consistent with Collins et al. (1994) and Lundholm & Myers (2002)'s findings. The positive coefficients demonstrate that news regarding current and future earnings is priced into stocks' current prices. The negative coefficients show that past earnings and future returns remove some of the measurement error when using current and future earnings to proxy for unexpected current earnings and changes in future earnings expectations, respectively.

The rest of 6 builds up to my full specification step by step. Regression 2 extends the FERC model by adding industry listing counts (LC_t) without any controls. The coefficient on the variable of interest $LC_t * E3_t$ is negative at a 5% significance level, showing that a firm's FERC generally decreases as industry listing counts increase. As previously, LC_t is a natural logarithm. However, $E3_t$ is not. Therefore, the magnitude of -0.010 on $LC_t * E3_t$ means that a 1% increase in the number of public firms in the same industry is associated with a -0.0001 decrease in FERC.

To make this impact more quantifiable, it may be easier to see how FERC changes when an industry moves from 1 listed firm to 10 listed firms. The coefficient on $E3_t$, 0.224, is the baseline value of FERC when LC_t is 0, which is equivalent to a 3-digit SIC industry listing count of 1. A firm in an industry with 10 listed firms would instead be expected to have a FERC of around 0.134 ($0.224 + (-0.01) * (\frac{10}{1} - 1)$). This is equivalent to a c.40% decline in FERC.

So far, this relationship ignores the presence of confounding factors though. Regressions 3 and 4 add controls for firm size (by market capitalisation) and industry concentration (HHI). Their inclusion causes the coefficient of interest to become less negative and lose its significance. The residual FERC also decreases from a statistically significant 0.224 to an insignificant 0.004, indicating that FERC is increasing in firm size and HHI. As anticipated in my argumentation for these controls, it appears that an important portion of the negative relation between industry listing counts and FERC is due to smaller industries being more concentrated and having larger firms.

Regression 5 adds firm and industry M/B ratio controls, while regression 6 adds industry dummies and year fixed effects. As discussed in section 5.2, the likely effect of M/B ratio controls was uncertain. Since the baseline FERC has decreased again in regression 5, it seems that FERC may be increasing with M/B ratios, as argued by Durnev et al. (2003), when already accounting for firm size and HHI. The coefficient on $LC_t * E3_t$ also becomes more negative and significant after the inclusion of M/B ratio controls, suggesting they correlate positively with industry listings as well as FERC. $LC_t * E3_t$ loses both magnitude and significance again in regression 6 though, revealing the importance of fixed effects.

My full specification is represented by regression 7. Compared to regression 6, regression 7 includes "Other controls", including firms' proportion of intangible assets, firm and industry age, and the loss indicator. The coefficient on the interaction between industry listing counts and future earnings now becomes -0.005, meaning that for a 1% increase in listing count, a firm's FERC decreases by 0.00005. Firms in industries that go from having 10 listed firms to 20 listed firms see their FERC falling from 0.179 $(0.224 + (-0.005) * (\frac{10}{1} - 1))$ to 0.129 $(0.224+(-0.005)*(\frac{20}{1}-1))$, a roughly 28% decrease [6]. However, the size of the standard error (also -0.005) means that the null hypothesis of the coefficient on $LC_t * E3_t$ being different from 0 cannot be rejected even at a 10% significance level [7]. Thus, there is insufficient evidence that industry listing counts are related to FERC.

As mentioned at the end of the methodology (section 5), I also add analyst coverage and liquidity mediator variables to check the robustness of this finding. Table A.3 in the Appendix shows the result of this change, including the coefficients on each mediator interaction. In essence, adding analyst coverage and liquidity makes no difference. They have a negligible impact on $LC_t * E3_t$, which merely oscillates between having a coefficient of -0.005, as in the model without mediators, and -0.006. The standard errors are also pretty much unaffected, meaning that the relation between FERC and industry population remains as statistically insignificant as before. It seems that any effect the mediators may have on the industry population-FERC relation has been mostly absorbed by the factors already included in the model. Industry listing counts do not seem to influence informativeness much through liquidity or analyst coverage. This is not surprising given the extremely low adjusted R2 statistics previously observed in Table 5.

Table A.3 also shows that $NEST_t * E3_t$ and $TO_t * E3_t$ are both positive, implying that FERC increases with firms' analyst coverage and liquidity. This is consistent with research by Kerr et al. (2020), Chordia et al. (2008), Piotroski & Roulstone (2004) used to argue my second hypothesis. The coefficient on the interaction with analyst coverage is not statistically significant though, suggesting a more tenuous link there.

Another interesting observation, consistent across every incremental model in both Table

[6] Note that the baseline value for FERC for firms in industries with only one listing is still the coefficient on $E3_t$ in regression 2. The same coefficient in regression 7 represents the baseline FERC of a firm with a 0 value not only for LC_t but also for all the controls included in the model (i.e. a firm with a \$1m market capitalisation, in an industry with an HHI of 1, with no intangible assets, etc.). This is naturally not meaningful for gauging the magnitude impact of listing counts on FERC.

[7] The similar magnitudes of the standard errors across each incremental model and the full model shows that the insignificant results are not due to a lack of degrees of freedom.

6 and Table A.3, is that there appears to be a positive relationship between industry listing counts and the response coefficient on current earnings (ERC), represented by the coefficient on E_t. This association is always significant at the 5% level, at least. Similarly to FERC, the ERC measures how important unexpected current earnings are in determining current returns. [8] While I will discuss the possible reasons for this finding in section 9, no conclusions can be drawn from it with certainty because this thesis has not focussed on controlling for omitted variables in the relation between ERC and industry population.

All-in-all, my extended FERC models suggest that the market seems to be no better, and potentially worse, at judging the future earnings of stocks with more listed peers. Instead, investors seem to increasingly pay attention to current earnings. Before drawing any real conclusions though, I check whether cross-sectional differences or issues with my specification may be causing my results.

7.3 Cross-sectional tests for the extended FERC model

One reason why the full model does not show any significance for $LC_t * E3_t$ may be that the different types of industries and firms have opposing relationships between listing counts and informativeness. To investigate this possibility, I split the main sample into top and bottom quartiles based on industry listing counts, firm market capitalisation, firm market-to-book ratio, firm proportion of intangible assets, and firm age. I then test the difference in the coefficients of interest between each pair of quartiles. These regressions are run without mediators. Therefore, if the coefficients remain insignificant, it is not because some of the effect of industry population is being absorbed by analyst coverage or liquidity variables.

First, the importance of an additional public company may hinge on the number of existing public companies in its industry. The additional intra-industry information generated by a newly listed firm may add more value when there is less existing industry information. Consequently, analysts and investors may be less attracted by an additional list in a larger industry, and the marginal impact of an additional listed firm on FERC could vary based on industry population.

Companies of different sizes may also respond differently to externalities created by listed peers because size tends to be inversely correlated with analyst and sophisticated investor following. In turn, this may affect how the market prices in intra-industry information. Ayers & Freeman (2003) and Piotroski & Roulstone (2004) suggest that greater institutional investor and financial analyst activity accelerates the pricing-in of public information. Furthermore, Farboodi et al. (2018) suggest that large firms benefit to the expense of smaller

[8] Under the assumption that earnings follow an autoregressive process, past earnings (E_{t-1}) are used to control for the measurement error in unexpected current earnings.

ones as information quantities increase because they attract more data processing shifts from investors. Therefore, information produced by newly public peer firms may be less used in the pricing of smaller firms, causing them to have a less significant relationship with industry listing counts. In my cross-section, I measure size with market capitalisation.

Another set of interesting cross-sections may be based on the amount of uncertainty firms face. If a firm's cash flows are more uncertain, signals from both outside and inside a firm become increasingly used for forecasting them (Bai et al. 2016, Mathers et al. 2017). Bai et al. (2016) propose this as an explanation for firms with higher M/B ratios having larger improvements in informativeness over their period of study – they benefitted more during a structural shift in informativeness caused by technological and financial advances that facilitated the discovery, analysis, and market price aggregation of information. In a similar manner, more uncertain businesses may respond more to the presence of additional listed peers. I proxy for uncertainty using a firm's M/B ratio and percentage of intangible assets as both tend to be related to riskier future cash flows (Lee 2018, Doidge et al. 2018).

Finally, I examine cross-sections based on firms' ages in each time period – measured by the number of years since they first listed. Older firms are likely to have more public information about themselves as they have probably made more disclosures than younger firms and been researched more thoroughly. Investors in younger firms are likely to have to rely more on outside information like intra-industry signals. Thus, younger firms' prices may be more responsive to industry populations.

The results of the regressions run on each quartile pair are presented in Table 7. It shows that the interaction between LC_t and $E3_t$ remains insignificant across all cross-sections, suggesting that cross-sectional differences are not behind the main model's null result. To confirm the differences in coefficients are not statistically different between the quartile pairs, I perform a Z-test to compare the difference in their means. The Z-test calculation for differences between regressions is based on Clogg et al. (1995) and is expressed as:

$$Z = \frac{b_{8n} - b_{8m}}{\sqrt{(SE_{b_{8n}})^2 + (SE_{b_{8m}})^2}} \tag{8}$$

where b_{8m} and b_{8n} are the $LC_t * E3_t$ coefficients for regressions run on two different quartiles, and the SEs are their standard errors.

The bottom of Table 7 shows the Z-scores for the differences in $LC_t * E3_t$ coefficients between each quartile pair. None of the Z-scores breach a significance level of 10%. Therefore, there is not enough evidence at the 10% level to reject the null hypothesis that the $LC_t * E3_t$ coefficients are the same regardless of industry listing count, firm size, M/B ratio, proportion of intangible assets, or age. These results suggest the insignificant findings of the extended

Table 7: Full sample: Cross-sectional tests

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. The sample is split into top and bottom quartiles based on industry listing counts, size, market-to-book ratio, intangible assets, and age. The coefficient of interest is on $LC_t * E3_t$.

	Listing counts		Size		M/B ratio		Intangibles		Age	
	Low	High	Small	Big	Low	High	Low	High	Young	Old
E_{t-1}	0.256	−0.148	0.116	−0.618	0.305	2.749**	0.961**	−0.430	0.334	0.368
	(0.478)	(2.424)	(0.345)	(0.975)	(0.264)	(1.200)	(0.395)	(0.979)	(0.513)	(0.633)
E_t	1.219**	4.153	1.196***	1.447	0.565**	2.587*	1.087**	1.881*	2.116***	0.273
	(0.533)	(3.079)	(0.392)	(1.021)	(0.276)	(1.497)	(0.431)	(1.106)	(0.611)	(0.707)
$E3_t$	−0.100	−0.260	−0.125	0.367*	0.026	−0.158	−0.138	−0.222	−0.097	0.027
	(0.103)	(0.639)	(0.087)	(0.222)	(0.054)	(0.269)	(0.093)	(0.174)	(0.113)	(0.149)
$R3_t$	−0.050	−0.157	−0.012	−0.055	−0.107***	0.089	−0.008	0.060	0.016	−0.005
	(0.046)	(0.153)	(0.037)	(0.091)	(0.026)	(0.083)	(0.047)	(0.066)	(0.052)	(0.068)
LC_t	−0.006	−0.079***	−0.008*	0.019***	−0.027***	0.012**	−0.013***	0.009*	−0.002	0.006
	(0.007)	(0.026)	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
$LC_t * E_{t-1}$	0.008	−0.067	−0.016	−0.142**	−0.011	−0.068	−0.006	−0.120**	−0.067**	−0.022
	(0.046)	(0.210)	(0.022)	(0.063)	(0.017)	(0.067)	(0.025)	(0.059)	(0.030)	(0.052)
$LC_t * E_t$	−0.041	0.014	0.038	0.102	0.026	0.188**	0.049*	0.166***	0.101***	0.0001
	(0.055)	(0.282)	(0.026)	(0.069)	(0.019)	(0.081)	(0.029)	(0.063)	(0.036)	(0.059)
$LC_t * E3_t$	0.001	0.073	0.003	−0.009	0.003	−0.015	−0.004	−0.016	0.001	−0.008
	(0.012)	(0.066)	(0.006)	(0.013)	(0.004)	(0.014)	(0.006)	(0.010)	(0.007)	(0.013)
$LC_t * R3_t$	−0.006	0.007	0.003	−0.010**	0.005***	0.002	0.004	−0.00000	0.002	−0.0001
	(0.005)	(0.015)	(0.002)	(0.004)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	No	No	No	No	No	No	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	25260	23677	23679	23679	23679	23679	33340	23679	26513	22516
Adjusted R^2	0.43	0.42	0.38	0.45	0.55	0.38	0.39	0.40	0.42	0.40

Note: *p<0.1; **p<0.05; ***p<0.01

Z-scores for differences in $LC_t * E3_t$ coefficients between quartile pairs

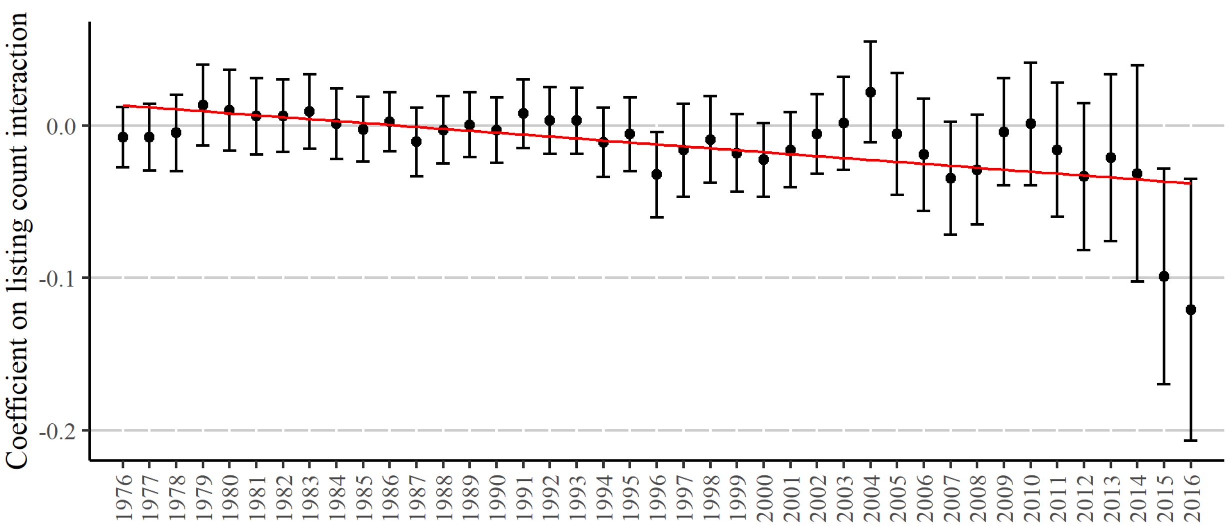
	Comparison across				
	Listing count	Size	M/B ratio	Intangibles	Age
Difference	-0.072	0.012	0.018	0.012	0.009
Z-score	-1.070	0.780	1.270	0.940	0.650
Significance	$p > 10\%$	$p > 10\%$	$p > 10\%$	$p > 10\%$	$p > 10\%$

FERC model which was run across the full sample are not due to cross-sectional differences.

7.4 Exploring the relevance of listing counts across time

Next, I look at whether industry listing counts have a stronger relationship with informativeness in certain periods of time. This analysis is primarily exploratory but also serves as a robustness check on my use of year fixed effects. For this cross-sectional test, I perform the full regression separately for each year (without year dummies) and plot the coefficient on $LC_t * E3_t$ over time. Figure 1 presents the results of this analysis, along with 95% confidence intervals for each coefficient.

Figure 1: The association between industry listings and price informativeness over time
The y axis shows the magnitude of the coefficient on $LC_t * E3_t$ in regressions (including all controls, fixed effects, and standard error adjustments) run separately for each year in the sample, with 95% confidence intervals.



Consistent with the full model run across the whole sample, the progression of the coefficient of interest over time is not significantly different from 0 for most years. However, it is interesting to note that there appears to be a negative slope since 2011, and that the regressions for 2015 and 2016 display much more negative coefficients. While the coefficients for other years range between -0.05 and +0.02, 2015's is c. -0.10 and 2016's is c. -0.12. Both years also exhibit much greater standard errors. A Chow test with a break point in 2011 is significant beyond the 1% level, further indicating the presence of a structural change.

To study this phenomenon further, I re-run size, M/B ratio, intangible asset, and age cross-sections on two subsamples of my dataset. The first subsample contains firm-year observations from 1976 to 2010 and the second has observations from 2011 to 2016. Tables A.4 (in the Appendix) and 8 present the result across the first and second subsample, respectively.

The regressions across the full subsamples show that $LC_t * E3_t$ has a coefficient of -0.003, with no statistical significance at the 10% level, in the 1976 to 2010 subsample, and a coefficient of -0.036, with significance at the 5% level, in the 2011 to 2016 sample. This supports Figure 1 in implying that an inverse relationship between industry listing counts and informativeness has mostly developed in recent years.

The cross-sectional tests across firm characteristics help locate where a change may have occurred. Table A.4 (in the Appendix) shows that, like the regressions performed on the complete dataset, the interaction between industry listings and future earnings remains negative and statistically no different from 0 across all cross-sections of the 1976-2010 subsample. However, Table 8 reveals that firms in the top quartiles for M/B ratio, proportion of intangible assets, and age display significantly negative coefficients on $LC_t * E3_t$ over the period from 2011 to 2016. The firms with the highest M/B ratio have a coefficient of -0.173, significant to the 1% level. The firms with the greatest percentage of intangible assets have a coefficient of -0.145, significant to the 1% level. And the oldest firms have a coefficient of -0.131, significant to the 1% level. The Z-tests show the cross-sectional differences are significant for each of these quartile pairs, especially between firms with high and low M/B ratios and intangible assets (significant beyond the 1% level). Furthermore, the Z-test for size also reveals that the largest firms in the 2011-2016 subsample are significantly more positively affected by industry listings than the smallest firms, at the 10% level.

Thus, the cross-sectional tests suggest that the negative relation between informativeness and listing counts from 2011 to 2016 has its root primarily amongst firms with high growth opportunities or a high percentage of intangible assets. As older firms are slightly more affected than younger ones, it seems that the change has not occurred due to new types of firms listing but rather due to changes in existing public firms. That said, the robustness tests detailed in the next section provide mixed evidence regarding the significance of the coefficient of interest across the high intangibles cross-section. The results of the high M/B ratio cross-section remain consistent though.

8 Robustness checks

Before drawing conclusions, I check whether my results are consistent across multiple different versions of my extended FERC regression. These include approaches that address limitations to my primary specification or other subjective tweaks. I run my robustness tests across the full dataset and the 2011-2016 subsample, where I discovered significantly negative relationships between FERC and industry populations particularly amongst high

Table 8: 2011-2016 subsample: Cross-sectional tests by firm characteristics

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Regressions are run on the full 2011-2016 subsample as well as the subsample split into top and bottom quartiles based on size, market-to-book ratio, intangible assets, and age. The coefficient of interest is on $LC_t * E3_t$.

	Full Sample	Size		M/B ratio		Intangibles		Age	
		Small	Big	Low	High	Low	High	Young	Old
E_{t-1}	0.349	−0.359	2.449	0.443	−3.723	−0.052	5.378*	−1.148	0.723
	(1.071)	(1.377)	(3.172)	(0.925)	(4.008)	(1.467)	(3.105)	(1.514)	(2.122)
E_t	−0.864	−2.046	2.752	0.032	8.051	0.772	−6.363	−0.713	−2.794
	(1.259)	(1.593)	(4.579)	(1.019)	(4.922)	(1.606)	(3.957)	(1.981)	(2.561)
$E3_t$	0.103	0.522	−1.247	0.068	0.880	−0.406	1.311	0.897*	1.104**
	(0.313)	(0.407)	(1.034)	(0.230)	(0.966)	(0.427)	(0.817)	(0.493)	(0.544)
$R3_t$	0.107	0.097	0.036	−0.019	0.497**	0.279*	−0.042	0.175	−0.859***
	(0.084)	(0.143)	(0.175)	(0.088)	(0.236)	(0.157)	(0.210)	(0.122)	(0.172)
LC_t	−0.003	−0.012	0.020	0.008	0.033**	−0.023*	0.009	−0.010	0.011
	(0.007)	(0.010)	(0.014)	(0.008)	(0.014)	(0.012)	(0.014)	(0.013)	(0.015)
$LC_t * E_{t-1}$	0.016	−0.017	−0.214	−0.043	0.262	0.075	−0.239	0.271**	0.101
	(0.080)	(0.110)	(0.130)	(0.058)	(0.249)	(0.116)	(0.182)	(0.107)	(0.137)
$LC_t * E_t$	0.131	0.234**	−0.100	−0.016	0.167	−0.016	0.559**	−0.072	0.089
	(0.081)	(0.100)	(0.201)	(0.056)	(0.282)	(0.112)	(0.260)	(0.124)	(0.162)
$LC_t * E3_t$	−0.036*	−0.046	0.042	−0.016	−0.173***	0.013	−0.145***	−0.055	−0.131***
	(0.021)	(0.031)	(0.047)	(0.016)	(0.056)	(0.029)	(0.054)	(0.034)	(0.044)
$LC_t * R3_t$	−0.002	0.002	0.0002	−0.001	−0.010	−0.014	0.007	−0.008	0.056***
	(0.005)	(0.009)	(0.009)	(0.007)	(0.012)	(0.009)	(0.009)	(0.008)	(0.015)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	No	No	No	No	No	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11665	2917	2916	2917	2916	2917	2916	3022	2879
Adjusted R^2	0.33	0.33	0.43	0.47	0.33	0.39	0.41	0.40	0.47

Note: *p<0.1; **p<0.05; ***p<0.01

Z-scores for differences in $LC_t * E3_t$ coefficients between quartile pairs

	Comparison across			
	Size	M/B ratio	Intangibles	Age
Difference	-0.088	0.157	0.158	0.076
Z-score	-1.550	2.690	2.600	1.370
Significance	$p < 10\%$ (*)	$p < 1\%$ (***)	$p < 1\%$ (***)	$p < 10\%$ (*)

M/B ratio and high intangible cross-sections.

Alternative industry listing count variables:

This set of checks entails changes to the industry listing count variable. First, I see whether my main results hold up when lagging LC_t by one year as the impact of a change in listing counts may not be immediate. Additionally, I employ different industry classifications when calculating industry listing counts. Rather than 3-digit SIC classifications, I instead try using 4-digit SIC classifications and 4-digit North American Industry Classification System (NAICS) codes. The 4-digit SIC listing counts check whether I have been using too broad a peer group. The NAICS classification checks whether the limitations to SIC codes, such as its arguably slow adaptation to new industries, may have influenced my results. However, NAICS data was only available on CRSP from 2004, so the sample used for this robustness check is also more limited. Tables A.5, A.6, and A.7 in the Appendix present the results of these checks. The coefficient of interest in the high intangibles cross-section in the 2011-2016 subsample loses its significance in the 4-digit SIC and NAICS variations. My other results remain unchanged regardless of which industry listing count variation is used.

Additional exclusions:

This set of checks includes additional exclusions based on firm IPO date, utilities industry affiliation, and fiscal year end. First, I try dropping firm-year observations of firms that listed within the last two years. This gives their returns time to normalise, as investors are dealing with a lot of new information after an IPO. As shown in Table A.8 in the Appendix, performing this exclusion does not significantly alter my findings.

Second, I drop firms whose primary industry belongs to SIC codes between 4000 and 4999 (the Utilities industry). Many papers studying informativeness, such as Gelb & Zarowin (2002) and Tucker & Zarowin (2006), make this exclusion, arguing that their regulated businesses make them generally incomparable to other firms. Table A.9 in the Appendix shows that dropping Utilities does not affect my findings across the full dataset. However, negative relation between industry listing counts and FERC in the 2011 to 2016 subsample gets slightly smaller and loses its significance. The high M/B ratio and high intangibles cross-sections in the 2011 to 2016 subsample maintain their significance nonetheless though.

Third, I exclude firms with fiscal years ending on dates other than the 31st of December. As mentioned in the Data section, some of my variables are necessarily calculated at fiscal year-ends, while others are necessarily calculated at calendar year-ends. By dropping all firm-years where the fiscal year-end is not also the calendar year end, I check whether this mismatch affects my results. Table A.10 in the Appendix shows that this exclusion makes the negative relation between industry listings and FERC significant at the 10% level across the whole dataset. Furthermore, the same coefficient becomes less pronounced in the 2011 to 2016 subsample and its high M/B ratio cross-section.

Changing FERC parameters:

Next, my results may be sensitive to the subjective choices I have made when selecting the earnings measure and time horizon for the underlying FERC model (Model 3). Table A.11 in the Appendix shows the results of a specification using EBIT rather than EBITDA as the measure of earnings. Table A.12 in the Appendix estimates FERC with 4 years of future earnings, rather than 3. Both robustness checks do not significantly influence my results.

Using non-synchronicity as my measure of informativeness:

For my final set of robustness checks, I use an alternative measure of price informativeness – price non-synchronicity. Non-synchronicity is the extent to which stock returns are unexplained by market or industry returns, namely firm-specific returns. As discussed in the literature review, Roll (1988) first suggested that more non-synchronous stock prices were likely to reflect the presence of more private information or noise. This led to many studies using non-synchronicity as an estimate of the amount of private information baked into a stock price. Subsequent research by Durnev et al. (2003) provided evidence in favour of this usage by showing a strong link between non-synchronicity and FERC, supporting studies using non-synchronicity as a measure of informativeness.

Non-synchronicity is calculated by taking 1-R^2, where R^2 is the R-squared of the following regression:

$$r_d = \beta_0 + \beta_m r_{m,d} + \beta_j r_{j,d} + \varepsilon_d \tag{9}$$

where r_d is the stock return at time d for a firm in industry j, $r_{m,d}$ is the market return at time d, and $r_{j,d}$ is the return of industry j at time d.

For my robustness check, I calculate non-synchronicity for all the firm-years in my full dataset. To do this, I firstly pull the daily market returns and daily stock returns from CRSP for the relevant firms from 1976 to 2016. For the market returns, I use CRSP's value-weighted index. Both market and stock returns exclude distributions. Next, I calculate daily industry returns by taking the average daily returns, weighted by market cap, of firms within the same 2-digit SIC industry. Finally, I run model 9 over the daily returns within each firm-year and extract the 1-R^2's of each. This measure of non-synchronicity is represented by the variable $NSYNCH_t$.

Thus, the specification used for this robustness check is expressed as:

$$NSYNCH_t = b_0 + b_1 LC_t + b_2 Z_{,t} + b_3 IND_t + b_4 YEAR_t + \varepsilon_t \tag{10}$$

where LC_t is the natural logarithm of the number of publicly listed firms within a firm's primary 3-digit SIC industry at the end of year t, IND_t is a vector of dummy variables for the firm's primary 1-digit SIC industry that year, $YEAR$ are year dummy fixed effects, and Z_t is a vector of control variables. Z_t contains the same controls as were argued for the extended FERC model. However, as my variable of interest, LC_t, is no longer part of an interaction (informativeness is now the dependent variable) the controls are not interacted either. As with the main models, heteroscedasticity-robust White standards errors and industry clustering are used.

Table A.13 in the Appendix displays the results of this robustness check. I give an example of their interpretation. For the full sample regression with mediators, a coefficient of -0.001 is reported for LC_t, the industry listing count. As LC_t is a natural logarithm but $NSYNCH_t$ is not, this means that for a 1% increase in industry listing count, a firm's non-synchronicity generally decreases by 0.00001. Thus, the direction of the relationship between industry listing counts and the informativeness measure is the same as for my extended FERC models. However, the negative relation with non-synchronicity across the full sample is statistically significant to the 1% level, both with and without mediators. This is unlike the FERC models, whose negative $LC_t * E3_t$ coefficients lacked significance. The negative relationship in the 2011 to 2016 subsample is also significant at the 5% level rather than the 10% level, although the intangibles cross-section again loses statistical support.

9 Discussion

Most of my results and robustness checks across the full dataset reveal no statistically meaningful relationship between informativeness and industry population. Cross-sectional tests based on industry listing count, firm size, M/B ratio, proportion of intangible assets, and age, suggest this finding is likely not explained by non-linearity or by divergent reactions across different types of firms. However, I may have missed an important cross-section. It is also interesting to note that the direction of the association between informativeness and industry listings is consistently negative across all the variations of my regression, including robustness checks. In the ones where fiscal year ends other than the 31st of December are dropped or non-synchronicity is used to represent informativeness, the negative relationship is also statistically significant. Furthermore, the inverse association is statistically meaningful in the 2011-2016 period, particularly amongst firms with high M/B ratios, but not in the 1976-2010 period. Having briefly summarised my findings, the rest of this section will discuss what meaning can be drawn from them, propose ideas for future research, and address this thesis' limitations.

Before discussing possible inferences from the results, the reader may benefit from a recap of my hypotheses' formulation (Section 3). In essence, I argued that the impact industry listing counts may have on informativeness is driven by their influence on analyst coverage, liquidity, and the quantity of public intra-industry information. Literature has shown that analyst coverage and liquidity positively affect informativeness, and that investors use some forms of public intra-industry information (Foster 1981, Han et al. 1989, Kerr et al. 2020, Chordia et al. 2008, Piotroski & Roulstone 2004). Therefore, in H2, I hypothesised that informativeness would be increasing with industry populations under two conditions. The first is that analyst coverage and liquidity (the mediator variables) are greater in more populated industries, as inferred from work by Subrahmanyam & Titman (1999) and Veldkamp (2006). Since this condition is testable, it is expressed in H1. The second condition is that the US market prices properly react to additional public intra-industry signals.

I explored H1 through the models that used analyst coverage and liquidity as dependent variables. Since all versions of these regressions reveal a strongly significant positive association between industry listing counts and the mediators, the evidence appears to be distinctly in favour of H1. However, the same regressions show that industry population explains an almost negligible percentage of the variation in analyst coverage and liquidity. Unsurprisingly then, the results of all my extended FERC models and robustness checks are fundamentally unchanged regardless of whether I include the mediator variables or not. The impact that industry population has on informativeness through its positive relation with analyst coverage and liquidity appears very limited.

Assuming there are no further mediators, the extended FERC regressions which account for analyst coverage and liquidity variables provide insights about the third avenue through which I propose industry listing counts may affect informativeness – intra-industry information. The continued insignificance of the relation between industry population and FERC once mediators are included therefore questions the utility of intra-industry signals. One reason for this may be that, although literature has shown that investors use intra-industry news, other types of information, such as firms' own disclosures, may be much more useful for pricing stocks. Alternatively, information generated by companies in the same industry may be valuable, but internationally as well as domestically. This thesis has measured industry listing counts based on the number of US firms listed on the NYSE, Nasdaq, or Amex. Since certain shocks, such as political and environmental events, are shared locally, information generated by peers is likely most relevant for evaluating firms located close by. However, economic and financial globalisation, coupled with rapid information transfer, has probably given greater importance to data from domestic and foreign peers listed abroad. Therefore, the association between domestic industry listing counts and informativeness may be dampened by the presence of foreign peers that also generate useful information. The real informational effect of an additional listing may be better studied with a global perspective.

However, if intra-industry signals were merely relatively uninformative, they would still be somewhat useful for the evaluation of stocks. I would still expect my results to suggest, without significance, that the relation between industry listing counts and informativeness is positive. Instead, the direction is negative across all informativeness models run on the full dataset, and meaningfully so in the fiscal year-end end and non-synchronicity variations. Furthermore, the inverse relationship has been gradually becoming more negative over time and is particularly prominent in 2015 and 2016. It seems that informativeness is suffering as listing counts increase. Therefore, I suspect that the second condition in H2 does not hold in many cases. The market may be improperly aggregating additional public intra-industry signals for many firms, especially in recent years.

Harkening to my hypotheses' argumentation again, I proposed two ways through which a market with semi-strong inefficiencies may cause greater industry listing counts to adversely impact peers' stock price informativeness. First, most investors may be susceptible to the noise in intra-industry signals. The more they use such signals, as is likely the case in more populated industries, the more they would be promulgating mispricing. Second, the existence of a greater number of peers may incentivise investors to increasingly rely on the information generated by certain "bellweather" firms, at the expense of their employment of firm-specific information. This could reduce the information content of prices. Both possibilities could help explain another interesting result of my extended FERC regressions – the significantly positive relation between current earnings and industry listings. If investors' expectations about future earnings are less correct or less informed, they may end up more surprised about current performance.

Of course, both mechanisms could occur simultaneously. However, the significantly negative coefficient on industry listing counts in the non-synchronicity robustness check favours the "bellweather" explanation. FERC and non-synchronicity are both used as proxies for informativeness but they fundamentally measure different things. While FERC aims to represent the extent to which prices properly react to new information, non-synchronicity is more a measure of stocks' firm-specific information content. Greater investor sensibilities to noise and "bellweather" signals should both decrease FERC, but they may have different effects on return non-synchronicity. More mispricing due to the incorporation of noise has indeterminate consequences on non-synchronicity, depending on whether the noisy signals amongst investors are correlated (Subrahmanyam & Titman 1999). On the other hand, greater adherence to "bellweather" firm signals would increase comovement and decrease non-synchronicity, as evident in my non-synchronicity robustness check.

All-in-all, several inferences and implications can be drawn from my results. First, larger industry populations do not seem to meaningfully affect informativeness by attracting more analyst coverage or liquidity. Second, intra-industry signals may be relatively irrelevant for the evaluation of a stock and may even prompt an inappropriate reaction by market prices. Further exploration of this latter possibility may be of interest to investors looking to improve their use of information or their understanding of the market. Many may be applying intra-industry information more indiscriminately in larger industries and with decreasing regard for stocks' firm-specific information, especially in recent times. Regardless of its causes though, the fact that the association between industry listings and informativeness is not positive suggests that the informativeness of most US stock prices may not have suffered from their having fewer industry peers since 1997. Regulators may have one less reason to worry about the declining number of listed in the US. Finally, my non-positive results question the assumption made by some papers, such as Chemmanur et al. (2010), that investors can more easily evaluate firms with more public industry peers. The data provides insufficient support for this - at worst the true relationship may even be the reverse. Some literature may therefore need to review its usage of industry listing counts as a proxy for informativeness.

Suggestions for future research:

My inferences regarding the proper usage of intra-industry information are very tentative given the insufficient statistical significance of most of my coefficients of interest. That said, the meaningfully negative relationship between industry listing counts and non-synchronicity deserves a more targeted study. I have merely run it as a robustness check and may have missed elements of the specification that require adjustment when altering the measure of informativeness. A more refined examination could provide concrete insights about investors' incorporation of firm-specific information given the presence of more listed peers, which may be of interest to both investors and regulators.

In a similar vein, the relation between ERC (the current, not future, earnings response coefficient) and industry listing counts may be worth a closer look as I find that it is significantly positive across all my models. Why the market seems to pay more attention to current earnings in larger industries is curious. Although I have offered one rough explanation – that investors are generally more surprised by current earnings because additional intra-industry information worsens their forecasts of future earnings – a targeted study may reveal more robust insights.

Future research could also perform a closer investigation of the seemingly changing relationship between industry populations and price informativeness in more recent years. My results suggest that the relation is significantly negative for firms with high M/B ratios in the period from 2011 to 2016, but not 1976 to 2010. Additionally, the changes do not seem to have been driven by new types of firms listing. Although one of my robustness checks assigns less significance to these findings, a study is warranted anyway because of the possible implications. An increasingly negative relation between industry lists and informativeness suggests there has recently been a structural change in existing public businesses, the way in which investors evaluate them, or the aptness of FERC as a proxy for their stocks' informativeness. For example, if these firms are investing much more in long-term growth and taking medium-term losses, investors may be pricing in expected earnings for increasingly distant horizons. If this is the case, FERC's usual 3 leads of future earnings, argued for by Collins et al. (1994), may be insufficient in the more modern environment, with repercussions on much future research involving FERC.

Furthermore, a study of the impact of industry listing counts on public firms' investment efficiency could be interesting. Badertscher et al. (2013)'s finding that the investment efficiency of private firms increases with public firm presence in their industry provides an interesting contrast to my results. Together, they suggest that managers of private firms draw useful information from peer firms, but investors do not. This may be because managers are better informed about their own firms than investors and are better able to evaluate information generated by public peers. If this is the case, the investment efficiency of public firms may still improve with industry population, contemporaneously with a decrease in their stocks' informativeness. The declining number of public firms in the US may have had adverse effects through this avenue.

Limitations:

However, before taking this thesis' results at face value, the reader should keep in mind its various limitations. A first instance of this is my use of primary industry classifications for the calculation of industry populations. My listing count measures thus unfortunately overlook the more nuanced set of subindustries that a firm may fit into. Since industry listing counts are necessarily based on CRSP data, which does not provide segment breakdowns, I am currently unaware of a way around this issue. Furthermore, industry classification is not a straightforward science. SIC codes may not accurately reflect some firms' businesses and have been criticised for not properly adapting to the emergence of new industries. The fact that my findings do not change when switching from an SIC to NAICS classification partially eases concerns about this limitation though.

Next, it is not certain that FERC is a good measure of price informativeness. As mentioned in section 4, the FERC model may imperfectly deal with issues related to the unobservable nature of expectations. Additionally, FERC assumes that new developments relevant to a firm affect its earnings within a medium-term horizon of around 3 years. However, this is not necessarily the case – effects on earnings may take longer to appear. Furthermore, annual estimations of FERC may be noisy because they are susceptible to strange movements in year-start and year-end stock prices that are unrelated to news. Relying on it as an estimate of informativeness requires one to assume that such errors are not consistent in their direction. Given the randomness of the market price movements in the absence of new information, this seems reasonable. The bottom line is that I have used FERC as it is the most researched variation of the more sophisticated informativeness measures, but better proxies may emerge.

Furthermore, my models may not be perfect. They may suffer from some remaining omitted variable bias deriving from the absence of important controls I have not identified or not managed to find data for [9]. There may also be some leftover bias due to imperfections in the controls I have included, such as HHI's approximation of industry concentration, or because of some mismatched year-ends. Additionally, my coefficients of interest in the informativeness models may contain traces of reverse causality, although this should not affect the implications of my findings. Research such as Subrahmanyam & Titman (1999) suggests that an improvement in the informativeness of a market's prices may attract more IPOs from businesses whose managers draw value from market signals. However, whether this is also true for greater industry-level informativeness is unclear. Even if it were, the relation between industry listings and informativeness would end up being more positive than it should be, if anything. Since my results show negative coefficients instead, the direction of the coefficients, and consequently my interpretations, should not change. My inferences rely on my having identified all mediators though. If new avenues are discovered through which industry listings may affect informativeness, my deductions about intra-industry information may be subject to re-evaluation.

As I was not able to perfectly resolve them at this point in time, each of the limitations discussed in this subsection may also present an opportunity for future research to improve on my study.

10 Conclusion

This thesis studies the relation between US firms' stock price informativeness and the number of domestically listed US firms in their primary industry. Indications from existing literature are inconclusive about this association. Some research suggests additional listed firms create positive informational externalities by attracting more analysts and trading activity to their industries, as well as by deepening the pool of intra-industry information.

[9]Institutional ownership, a variable I was not able to get data for, may be an example of this.

However, possible market inefficiencies make the direction of the overall relationship uncertain.

My investigation begins with regressions of industry population on analyst coverage and liquidity, possible mediator variables through which listing counts may affect informativeness. In line with my first hypothesis, these show a positive association. Despite strong statistical significance, only a tiny proportion of the mediators' variation is explained though, questioning their actual relevance as mediators. Next, I form an extended FERC model to test the association between industry listing counts and informativeness. Without the inclusion of control variables, informativeness is significantly decreasing in industry population. The association remains negative but loses statistical meaning once important confounding factors are accounted for. Further including analyst coverage and liquidity in the regression as mediating variables has a negligible effect on this result. Robustness tests also consistently reveal negative coefficients, with occasional statistical significance. I find no evidence that these findings on the full sample are due to starkly different behaviours between cross-sections based on industry population, firm size, firm M/B ratios, firm intangible assets, or age. In the period from 2011 to 2016, however, the negative relation is significant and particularly strong for firms with high M/B ratios.

Since analyst coverage and liquidity behave as expected but seem unimportant as mediators, I suggest several reasons for these findings based on the mechanics of intra-industry information, the remaining avenue through which I propose industry populations may affect informativeness. First, the general lack of statistical significance on my coefficients of interest implies that information produced by public companies may not be that useful for evaluating peers compared to other signals, such as firms' own disclosures. However, the persistently negative coefficients, with statistical significance in some cross-sections, suggest market inefficiencies may also be at play. I propose two ways in which this may occur. First, investors may often be more susceptible to the noise generated by peers than to the true information they offer, causing greater use of intra-industry signals in larger industries to perpetuate mispricing. Second, greater industry populations may incentivise investors use more information about prominent industry peers, at the expense of their employment of firm-specific data. This may reduce stocks' information content. All this said, readers should also keep in mind that certain limitations to this study could be obscuring the true dynamics of the relationship between informativeness and industry listing counts.

My results could be useful in several dimensions. Regardless of its cause, the non-positive association between industry listings and informativeness suggests that regulators should not be worried about the informativeness of most US stock prices having suffered from the presence of fewer industry peers since 1997. Furthermore, it questions the use of industry listing counts as a proxy for informativeness by some previous literature. As my findings may be due to the market's misaggregation of intra-industry signals, this thesis could also alert investors to circumstances where their collective employment of such information may be more likely to have flaws. Finally, I propose various ideas for future research.

11 References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. (2017), When should you adjust standard errors for clustering?, Technical report, National Bureau of Economic Research.

Alford, A. W. & Berger, P. G. (1999), 'A simultaneous equations analysis of forecast accuracy, analyst following, and trading volume', *Journal of Accounting, Auditing & Finance* **14**(3), 219–240.

Ayers, B. C. & Freeman, R. N. (2003), 'Evidence that analyst following and institutional ownership accelerate the pricing of future earnings', *Review of Accounting Studies* **8**(1), 47–67.

Badertscher, B., Shroff, N. & White, H. D. (2013), 'Externalities of public firm presence: Evidence from private firms' investment decisions', *Journal of Financial Economics* **109**(3), 682–706.

Bai, J., Philippon, T. & Savov, A. (2016), 'Have financial markets become more informative?', *Journal of Financial Economics* **122**(3), 625–654.

Baker, M. & Wurgler, J. (2007), 'Investor sentiment in the stock market', *Journal of economic perspectives* **21**(2), 129–152.

Beaver, W. H. (1968), 'The information content of annual earnings announcements', *Journal of accounting research* pp. 67–92.

Bhushan, R. (1989), 'Firm characteristics and analyst following', *Journal of accounting and economics* **11**(2-3), 255–274.

Black, B. S. & Gilson, R. J. (1998), 'Venture capital and the structure of capital markets: banks versus stock markets', *Journal of financial economics* **47**(3), 243–277.

Boot, A. W., Gopalan, R. & Thakor, A. V. (2006), 'The entrepreneur's choice between private and public ownership', *The Journal of Finance* **61**(2), 803–836.

Brau, J. C. & Fawcett, S. E. (2006), 'Initial public offerings: An analysis of theory and practice', *The Journal of Finance* **61**(1), 399–436.

Breusch, T. S. & Pagan, A. R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica: Journal of the econometric society* pp. 1287–1294.

Bushee, B. J. & Miller, G. S. (2012), 'Investor relations, firm visibility, and investor following', *The Accounting Review* **87**(3), 867–897.

Chemmanur, T. J. & Fulghieri, P. (1999), 'A theory of the going-public decision', *The review of financial studies* **12**(2), 249–279.

Chemmanur, T. J., He, S. & Nandy, D. K. (2010), 'The going-public decision and the product market', *The Review of Financial Studies* **23**(5), 1855–1908.

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

Choi, J.-H., Choi, S., Myers, L. A. & Ziebart, D. (2019), 'Financial statement comparability and the informativeness of stock prices about future earnings', *Contemporary Accounting Research* **36**(1), 389–417.

Chordia, T., Roll, R. & Subrahmanyam, A. (2008), 'Liquidity and market efficiency', *Journal of Financial Economics* **87**(2), 249–268.

Clogg, C. C., Petkova, E. & Haritou, A. (1995), 'Statistical methods for comparing regression coefficients between models', *American journal of sociology* **100**(5), 1261–1293.

Collins, D. W., Kothari, S. P., Shanken, J. & Sloan, R. G. (1994), 'Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association', *Journal of Accounting and Economics* **18**(3), 289–324.

Dávila, E. & Parlatore, C. (2018), Identifying price informativeness, Technical report, National Bureau of Economic Research.

De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J. (1990), 'Noise trader risk in financial markets', *Journal of political Economy* **98**(4), 703–738.

Djama, C., Martinez, I. & Serve, S. (2012), What do we know about delistings? a survey of the literature, *in* 'Comptabilités et innovation', pp. cd–rom.

Doidge, C., Kahle, K. M., Karolyi, G. A. & Stulz, R. M. (2018), 'Eclipse of the public corporation or eclipse of the public markets?', *Journal of Applied Corporate Finance* **30**(1), 8–16.

Doidge, C., Karolyi, G. A. & Stulz, R. M. (2017), 'The us listing gap', *Journal of Financial Economics* **123**(3), 464–487.

Dow, J. & Gorton, G. (1997), 'Stock market efficiency and economic efficiency: Is there a connection?', *The Journal of Finance* **52**(3), 1087–1129.

Durnev, A., Morck, R., Yeung, B. & Zarowin, P. (2003), 'Does greater firm-specific return variation mean more or less informed stock pricing?', *Journal of Accounting Research* **41**(5), 797–836.

Ewens, M. & Farre-Mensa, J. (2017), 'The evolution of the private equity market and the decline in ipos', *Available at SSRN* **3017610**.

Ewens, M. & Farre-Mensa, J. (2020), 'The deregulation of the private equity markets and the decline in ipos', *The Review of Financial Studies* **33**(12), 5463–5509.

Fama, E. F. (1970), 'Efficient capital markets a review of theory and empirical work', *The Fama Portfolio* pp. 76–121.

Fama, E. F. & French, K. R. (2001), 'Disappearing dividends: changing firm characteristics or lower propensity to pay?', *Journal of Financial economics* **60**(1), 3–43.

Fama, E. F. & French, K. R. (2004), 'New lists: Fundamentals and survival rates', *Journal of financial Economics* **73**(2), 229–269.

Fan, G. H., Joos, P. R., Monahan, S. J. & Nambiar, S. P. (2019), 'Evaluating price informativeness and its determinants'.

Farboodi, M., Matray, A. & Veldkamp, L. (2018), 'Where has all the big data gone?', *Available at SSRN 3164360* .

Foster, G. (1981), 'Intra-industry information transfers associated with earnings releases', *Journal of accounting and economics* **3**(3), 201–232.

Foucault, T. & Fresard, L. (2014), 'Learning from peers' stock prices and corporate investment', *Journal of Financial Economics* **111**(3), 554–577.

Gao, X., Ritter, J. R. & Zhu, Z. (2013), 'Where have all the ipos gone?', *Journal of Financial and Quantitative Analysis* **48**(6), 1663–1692.

Gelb, D. S. & Zarowin, P. (2002), 'Corporate disclosure policy and the informativeness of stock prices', *Review of accounting studies* **7**(1), 33–52.

Glosten, L. R. & Milgrom, P. R. (1985), 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders', *Journal of financial economics* **14**(1), 71–100.

Grossman, S. J. & Stiglitz, J. E. (1980), 'On the impossibility of informationally efficient markets', *The American economic review* **70**(3), 393–408.

Grullon, G., Larkin, Y. & Michaely, R. (2019), 'Are us industries becoming more concentrated?', *Review of Finance* **23**(4), 697–743.

Hameed, A., Morck, R., Shen, J. & Yeung, B. (2015), 'Information, analysts, and stock return comovement', *The Review of Financial Studies* **28**(11), 3153–3187.

Han, J. C., Wild, J. J. & Ramesh, K. (1989), 'Managers' earnings forecasts and intra-industry information transfers', *Journal of Accounting and Economics* **11**(1), 3–33.

Hayek, F. A. (1945), 'The use of knowledge in society', *The American economic review* **35**(4), 519–530.

Helwege, J. & Packer, F. (2009), 'Private matters', *Journal of Financial Intermediation* **18**(3), 362–383.

Jensen, M. C. (1997), 'Eclipse of the public corporation', *Harvard Business Review (Sept.-Oct. 1989), revised* .

Jiambalvo, J., Rajgopal, S. & Venkatachalam, M. (2002), 'Institutional ownership and the extent to which stock prices reflect future earnings', *Contemporary accounting research* **19**(1), 117–145.

Kerr, J., Sadka, G. & Sadka, R. (2020), 'Illiquidity and price informativeness', *Management Science* **66**(1), 334–351.

Kothari, S. P. & Sloan, R. G. (1992), 'Information in prices about future earnings: Implications for earnings response coefficients', *Journal of Accounting and Economics* **15**(2-3), 143–171.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. & Vishny, R. W. (1997), 'Legal determinants of external finance', *The journal of finance* **52**(3), 1131–1150.

Lattanzio, G., Megginson, W. L. & Sanati, A. (2020), 'Dissecting the listing gap: Mergers, private equity, or regulation?', *Private Equity, or Regulation* .

Lee, J. J. (2018), 'Economic determinants of price informativeness about future earnings', *Journal of Contemporary Accounting & Economics* **14**(1), 83–102.

Lipson, M. L. & Mortal, S. (2007), 'Liquidity and firm characteristics: evidence from mergers and acquisitions', *Journal of Financial Markets* **10**(4), 342–361.

Loughran, T., Ritter, J. R. & Rydqvist, K. (1994), 'Initial public offerings: International insights', *Pacific-Basin Finance Journal* **2**(2-3), 165–199.

Lowry, M. (2003), 'Why does ipo volume fluctuate so much?', *Journal of Financial economics* **67**(1), 3–40.

Lundholm, R. & Myers, L. A. (2002), 'Bringing the future forward: the effect of disclosure on the returns-earnings relation', *Journal of accounting research* **40**(3), 809–839.

Mathers, A. M., Wang, B. & Wang, X. (2017), 'Innovation and price informativeness', *Financial Management* **46**(2), 523–546.

Morck, R., Shleifer, A., Vishny, R. W., Shapiro, M. & Poterba, J. M. (1990), 'The stock market and investment: is the market a sideshow?', *Brookings papers on economic Activity* **1990**(2), 157–215.

Morck, R., Yeung, B. & Yu, W. (2000), 'The information content of stock markets: why do emerging markets have synchronous stock price movements?', *Journal of financial economics* **58**(1-2), 215–260.

Orpurt, S. F. & Zang, Y. (2009), 'Do direct cash flow disclosures help predict future operating cash flows and earnings?', *The Accounting Review* **84**(3), 893–935.

Oswald, D. R. & Zarowin, P. (2007), 'Capitalization of r&d and the informativeness of stock prices', *European Accounting Review* **16**(4), 703–726.

Pagano, M., Panetta, F. & Zingales, L. (1998), 'Why do companies go public? an empirical analysis', *The journal of finance* **53**(1), 27–64.

Piotroski, J. D. & Roulstone, D. T. (2004), 'The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices', *The accounting review* **79**(4), 1119–1151.

Ramnath, S. (2002), 'Investor and analyst reactions to earnings announcements of related firms: An empirical analysis', *Journal of Accounting Research* **40**(5), 1351–1376.

Ritter, J. R. (1984), 'The" hot issue" market of 1980', *Journal of Business* pp. 215–240.

Ritter, J. R. & Welch, I. (2002), 'A review of ipo activity, pricing, and allocations', *The journal of Finance* **57**(4), 1795–1828.

Röell, A. (1995), 'The decision to go public: An overview'.

Roll, R. (1988), 'R2', *The Journal of finance (New York)* **43**(3), 541–566.

Shleifer, A. (2000), *Inefficient markets: An introduction to behavioural finance*, OUP Oxford.

Subrahmanyam, A. & Titman, S. (1999), 'The going-public decision and the development of financial markets', *The Journal of Finance* **54**(3), 1045–1082.

Thomas, J. & Zhang, F. (2008), 'Overreaction to intra-industry information transfers?', *Journal of Accounting Research* **46**(4), 909–940.

Tucker, J. W. & Zarowin, P. A. (2006), 'Does income smoothing improve earnings informativeness?', *The accounting review* **81**(1), 251–270.

Veldkamp, L. L. (2006), 'Information markets and the comovement of asset prices', *The Review of Economic Studies* **73**(3), 823–845.

Watts, R. L. (1978), 'Systematic 'abnormal'returns after quarterly earnings announcements', *Journal of financial Economics* **6**(2-3), 127–150.

Weild, D. & Kim, E. (2009*a*), 'Market structure is causing the ipo crisis', *Grant Thornton, Capital Market Series* .

Weild, D. & Kim, E. (2009*b*), 'A wake-up call for america', *Grant Thornton: Capital* .

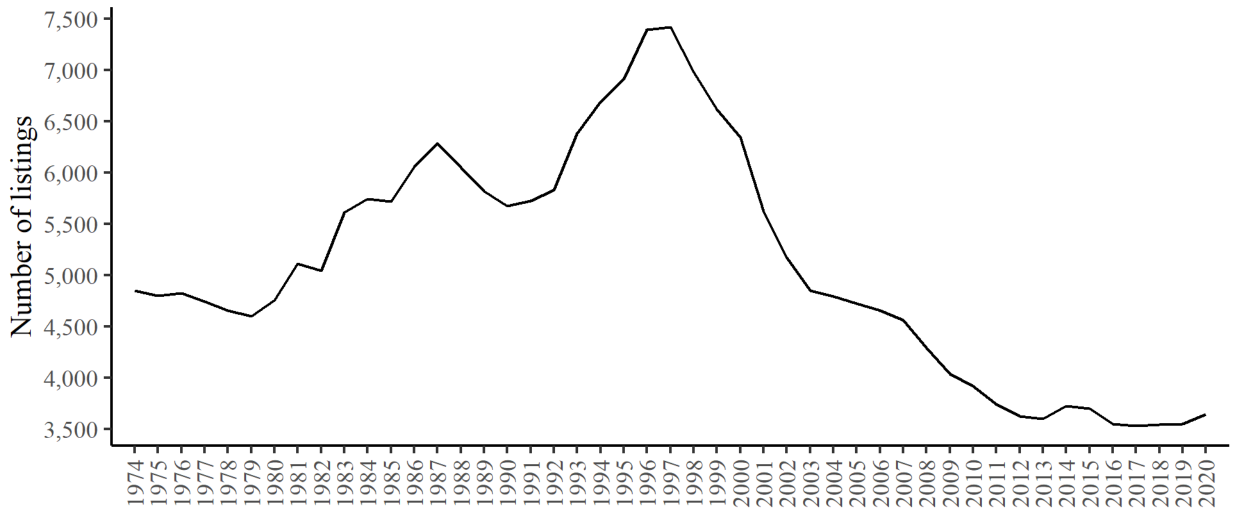
White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica: journal of the Econometric Society* pp. 817–838.

Wurgler, J. & Zhuravskaya, E. (2002), 'Does arbitrage flatten demand curves for stocks?', *The Journal of Business* **75**(4), 583–608.

Yan, D. (2020), 'Do private firms (mis) learn from the stock market?', *Swedish House of Finance Research Paper* .

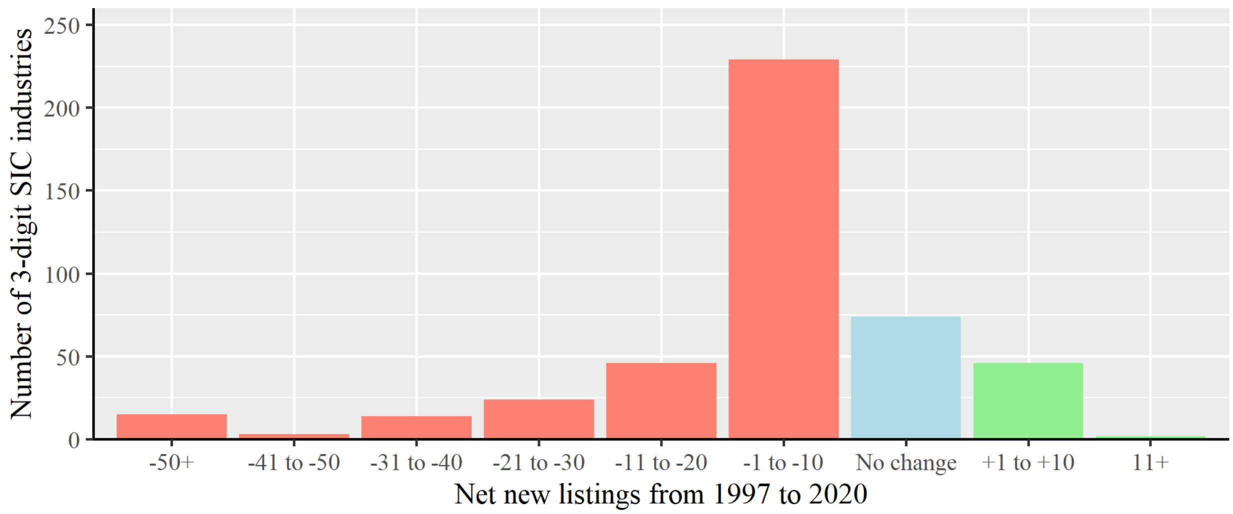
A Appendix

Figure A.1: Number of US firms listed on the NYSE, Nasdaq or Amex (1974-2020)



Source: The Center for Research in Security Prices (CRSP)
Note: Investment funds, trusts and other companies whose only business goal is to hold shares of other listed companies are excluded.

Figure A.2: Distribution of net new listings of domestically listed US firms across 3-digit SIC industries (1997-2020)



Source: The Center for Research in Security Prices (CRSP)
Note: Net new listings are calculated by subtracting the number of firms that have delisted from the number of IPOs within each industry. Firms that change industry classification are ***not*** *counted as having delisted from their industry. Investment funds, trusts and other companies whose only business goal is to hold shares of other listed companies are excluded. Firms in SIC code 9999 (unclassified) are excluded. Domestically listed firms refer to those with a common stock trading on the NYSE, Nasdaq, or Amex.*

Table A.1: Robustness checks: Analyst coverage model

$NEST_t$ represents analyst coverage. It is a natural logarithm. LC_t are logged 3-digit industry listing counts. 4-digit SIC LC_t are logged industry listing counts identified at the 4-digit SIC level. NAICS LC_t are logged industry listing counts identified at the 4-digit NAICS level. S_t is market capitalisation, TO_t is firm liquidity, and Age_t is the number of years since a firm was first listed. M/B and $IndM/B$ are firm and industry M/B ratios, respectively. All variables are as described in section 6.3. The coefficients of interest are on LC_t, LC_{t-1}, 4-digit SIC LC_t, and NAICS LC_t.

	Dependent variable: $NEST_t$			
	(1)	(2)	(3)	(4)
LC_t	0.017^{***}			
	(0.001)			
LC_{t-1}		0.019^{***}		
		(0.001)		
4-digit SIC LC_t			0.017^{***}	
			(0.001)	
NAICS LC_t				0.027^{***}
				(0.002)
S_t	0.401^{***}	0.401^{***}	0.401^{***}	0.360^{***}
	(0.001)	(0.001)	(0.001)	(0.002)
TO_t	0.143^{***}	0.144^{***}	0.142^{***}	0.222^{***}
	(0.002)	(0.002)	(0.002)	(0.004)
Age_t	-0.002^{***}	-0.003^{***}	-0.002^{***}	-0.006^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Constant	-1.018^{***}	-1.021^{***}	-1.007^{***}	-0.532^{***}
	(0.017)	(0.018)	(0.017)	(0.016)
Year fixed effects	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes
Observations	94714	94034	94714	28473
Adjusted R^2	0.73	0.73	0.73	0.74
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A.2: Robustness checks: Liquidity model

TO_t represents firm liquidity. It is a natural logarithm. LC_t are logged 3-digit industry listing counts. 4-digit SIC LC_t are logged industry listing counts identified at the 4-digit SIC level. NAICS LC_t are logged industry listing counts identified at the 4-digit NAICS level. S_t is market capitalisation, $NEST_t$ is analyst coverage, and Age_t is the number of years since a firm was first listed. M/B and $IndM/B$ are firm and industry M/B ratios, respectively. All variables are as described in section 6.3. The coefficients of interest are on LC_t, LC_{t-1}, 4-digit SIC LC_t, and NAICS LC_t.

	Dependent variable: TO_t			
	(1)	(2)	(3)	(4)
LC_t	0.082***			
	(0.002)			
LC_{t-1}		0.076***		
		(0.002)		
4-digit SIC LC_t			0.098***	
			(0.002)	
NAICS LC_t				0.023***
				(0.004)
S_t	−0.054***	−0.056***	−0.050***	−0.042***
	(0.003)	(0.003)	(0.003)	(0.004)
$NEST_t$	0.437***	0.441***	0.429***	0.603***
	(0.006)	(0.006)	(0.006)	(0.010)
M/B_t	0.014***	0.014***	0.013***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
$IndM/B_t$	0.017***	0.019***	0.016***	−0.007***
	(0.001)	(0.001)	(0.001)	(0.002)
Constant	−1.793***	−1.768***	−1.788***	−0.645***
	(0.023)	(0.023)	(0.023)	(0.029)
Year fixed effects	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes
Observations	94714	94034	94714	28473
Adjusted R^2	0.37	0.37	0.38	0.31
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A.3: Full sample: Extended FERC model including mediator variables

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). $NEST_t$ is analyst coverage and TO_t is firm liquidity. Both are logged variables. Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. The coefficient of interest is on $LC_t * E3_t$.

	Original model	With analyst coverage	With liquidity	With both
E_{t-1}	0.434	0.440	0.102	0.152
	(0.303)	(0.297)	(0.306)	(0.298)
E_t	0.976***	1.115***	1.395***	1.447***
	(0.329)	(0.320)	(0.325)	(0.316)
$E3_t$	−0.055	−0.055	−0.009	−0.019
	(0.062)	(0.061)	(0.062)	(0.061)
$R3_t$	−0.036	−0.005	−0.069**	−0.040
	(0.028)	(0.028)	(0.029)	(0.028)
LC_t	0.001	0.011***	0.001	0.008***
	(0.003)	(0.002)	(0.003)	(0.002)
$LC_t * E_{t-1}$	−0.037*	−0.042**	−0.028	−0.035*
	(0.019)	(0.019)	(0.020)	(0.019)
$LC_t * E_t$	0.068***	0.062***	0.056**	0.052**
	(0.023)	(0.022)	(0.022)	(0.022)
$LC_t * E3_t$	−0.005	−0.006	−0.005	−0.006
	(0.005)	(0.004)	(0.005)	(0.005)
$LC_t * R3_t$	0.001	0.001	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
$NEST_t$		−0.240***		−0.256***
		(0.005)		(0.005)
$NEST_t * E_{t-1}$		0.123***		0.160***
		(0.040)		(0.040)
$NEST_t * E_t$		0.093**		0.043
		(0.043)		(0.043)
$NEST_t * E3_t$		0.008		0.007
		(0.009)		(0.009)
$NEST_t * R3_t$		0.019***		0.025***

Table A.3 – continued from previous page

		(0.003)		(0.003)
TO_t			-0.002	0.034^{***}
			(0.003)	(0.003)
$TO_t * E_{t-1}$			-0.112^{***}	-0.109^{***}
			(0.021)	(0.021)
$TO_t * E_t$			0.174^{***}	0.165^{***}
			(0.023)	(0.023)
$TO_t * E3_t$			0.014^{***}	0.011^{**}
			(0.005)	(0.005)
$TO_t * R3_t$			-0.014^{***}	-0.017^{***}
			(0.002)	(0.002)
Controls	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes
Observations	94714	94714	94714	94714
Adjusted R^2	0.38	0.41	0.38	0.42

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.4: 1976-2010 subsample: Cross-sectional tests by firm characteristics

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Regressions are run on the full 1976-2010 subsample as well as the subsample split into top and bottom quartiles based on size, market-to-book ratio, intangible assets, and age. The coefficient of interest is on $LC_t * E3_t$.

	Full Sample	Size		M/B ratio		Intangibles		Age	
		Small	Big	Low	High	Low	High	Young	Old
E_{t-1}	0.419	0.197	−0.987	0.251	1.325	1.004**	−0.161	0.327	−0.254
	(0.321)	(0.364)	(1.011)	(0.270)	(1.212)	(0.408)	(0.908)	(0.580)	(0.615)
E_t	1.041***	1.320***	1.109	0.644**	3.378**	1.115**	0.872	2.386***	0.822
	(0.347)	(0.411)	(1.108)	(0.276)	(1.435)	(0.442)	(1.015)	(0.700)	(0.651)
$E3_t$	−0.066	−0.167*	0.530**	0.028	−0.170	−0.139	−0.113	−0.069	−0.022
	(0.064)	(0.091)	(0.219)	(0.054)	(0.243)	(0.094)	(0.169)	(0.126)	(0.134)
$R3_t$	−0.046	−0.003	−0.092	−0.112***	0.074	−0.012	0.053	0.004	−0.013
	(0.029)	(0.039)	(0.095)	(0.027)	(0.086)	(0.048)	(0.064)	(0.059)	(0.066)
LC_t	−0.0003	−0.007	0.016**	−0.029***	0.013**	−0.012**	0.008	−0.001	0.0001
	(0.003)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
$LC_t * E_{t-1}$	−0.044**	−0.018	−0.204***	0.001	−0.076	−0.009	−0.144***	−0.073**	0.010
	(0.021)	(0.023)	(0.073)	(0.017)	(0.069)	(0.026)	(0.053)	(0.033)	(0.053)
$LC_t * E_t$	0.070***	0.027	0.212***	0.013	0.198**	0.052*	0.178***	0.095**	−0.049
	(0.024)	(0.027)	(0.079)	(0.020)	(0.080)	(0.030)	(0.058)	(0.043)	(0.057)
$LC_t * E3_t$	−0.003	0.006	−0.013	0.004	−0.011	−0.004	−0.012	0.003	0.006
	(0.005)	(0.007)	(0.014)	(0.005)	(0.013)	(0.007)	(0.010)	(0.008)	(0.012)
$LC_t * R3_t$	0.001	0.002	−0.009**	0.005***	0.002	0.003	0.0001	0.001	−0.002
	(0.002)	(0.003)	(0.005)	(0.002)	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	No	No	No	No	No	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	83049	20763	20762	20763	20762	31499	20762	21473	20569
Adjusted R^2	0.39	0.38	0.46	0.55	0.39	0.39	0.42	0.43	0.41

Note: *p<0.1; **p<0.05; ***p<0.01

Z-scores for differences in $LC_t * E3_t$ coefficients between quartile pairs

	Comparison across			
	Size	M/B ratio	Intangibles	Age
Difference	0.019	0.015	0.012	0.003
Z-score	1.220	1.010	0.630	-0.210
Significance	$p > 10\%$	$p > 10\%$	$p > 10\%$	$p > 10\%$

Table A.5: Robustness checks: Changing to start-of-year industry listing counts

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged, year-start 3-digit SIC industry listing counts (LC_{t-1}). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. The coefficient of interest is on $LC_{t-1} * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.549^{***}	0.320	0.041	0.689	-4.965	5.825^{**}
	(0.061)	(0.297)	(0.298)	(1.073)	(4.287)	(2.955)
E_t	0.631^{***}	1.109^{***}	1.586^{***}	-1.589	6.286	-9.002^{**}
	(0.073)	(0.320)	(0.317)	(1.251)	(5.215)	(3.729)
$E3_t$	0.215^{***}	-0.083	-0.048	0.228	1.667^{*}	1.929^{**}
	(0.015)	(0.061)	(0.061)	(0.303)	(0.992)	(0.779)
$R3_t$	-0.100^{***}	-0.043	-0.048^{*}	0.060	0.240	-0.104
	(0.005)	(0.028)	(0.028)	(0.080)	(0.227)	(0.190)
LC_{t-1}	0.026^{***}	-0.001	0.007^{***}	-0.005	0.028^{**}	0.007
	(0.002)	(0.003)	(0.002)	(0.006)	(0.013)	(0.015)
$LC_{t-1} * E_{t-1}$	-0.111^{***}	-0.025	-0.024	-0.005	0.411	-0.330^{**}
	(0.017)	(0.018)	(0.018)	(0.081)	(0.267)	(0.163)
$LC_{t-1} * E_t$	0.047^{**}	0.047^{**}	0.031	0.188^{**}	0.270	0.913^{***}
	(0.020)	(0.021)	(0.021)	(0.084)	(0.302)	(0.237)
$LC_{t-1} * E3_t$	-0.007^{*}	-0.001	-0.002	-0.049^{**}	-0.236^{***}	-0.225^{***}
	(0.004)	(0.004)	(0.004)	(0.022)	(0.060)	(0.059)
$LC_{t-1} * R3_t$	0.002	0.002	0.002	0.002	0.008	0.015
	(0.001)	(0.002)	(0.002)	(0.005)	(0.011)	(0.011)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	94034	94034	94034	11655	2911	2911
Adjusted R^2	0.16	0.38	0.41	0.33	0.34	0.41

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.6: Robustness checks: Changing to 4-digit SIC industry listing counts

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 4-digit SIC industry listing counts (4-digit SIC LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. The coefficient of interest is on 4-digit SIC $LC_t * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.715^{***}	0.409	0.108	0.229	-0.626	5.137^{*}
	(0.045)	(0.279)	(0.277)	(0.937)	(3.662)	(3.039)
E_t	0.716^{***}	1.074^{***}	1.536^{***}	-0.217	6.159	-4.695
	(0.053)	(0.306)	(0.296)	(1.174)	(4.460)	(3.706)
$E3_t$	0.229^{***}	-0.057	-0.024	0.074	0.151	0.575
	(0.011)	(0.058)	(0.057)	(0.294)	(0.883)	(0.754)
$R3_t$	-0.102^{***}	-0.039	-0.046^{*}	0.056	0.345	0.003
	(0.004)	(0.026)	(0.026)	(0.077)	(0.218)	(0.197)
4-digit SIC LC_t	0.028^{***}	0.003	0.008^{***}	-0.004	0.011	0.007
	(0.002)	(0.002)	(0.002)	(0.006)	(0.013)	(0.013)
4-digit SIC $LC_t * E_{t-1}$	-0.086^{***}	-0.040^{**}	-0.035^{**}	0.027	-0.037	-0.293
	(0.016)	(0.017)	(0.017)	(0.064)	(0.198)	(0.189)
4-digit SIC $LC_t * E_t$	0.032^{*}	0.066^{***}	0.048^{**}	0.071	0.401^{*}	0.441^{*}
	(0.018)	(0.020)	(0.019)	(0.068)	(0.227)	(0.243)
4-digit SIC $LC_t * E3_t$	-0.015^{***}	-0.006	-0.006	-0.039^{**}	-0.129^{**}	-0.076
	(0.004)	(0.004)	(0.004)	(0.018)	(0.051)	(0.047)
4-digit SIC $LC_t * R3_t$	0.004^{***}	0.002	0.003^{*}	0.003	0.001	0.004
	(0.001)	(0.001)	(0.001)	(0.005)	(0.013)	(0.009)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	94714	94714	94714	11665	2916	2916
Adjusted R^2	0.16	0.38	0.42	0.33	0.33	0.40

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.7: Robustness checks: Changing to 4-digit NAICS industry listing counts

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 4-digit NAICS industry listing counts (NAICS LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. The coefficient of interest is on NAICS $LC_t * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.988^{***}	-0.816	-0.799	-0.456	-2.886	3.494
	(0.164)	(0.790)	(0.751)	(1.131)	(3.299)	(3.103)
E_t	1.143^{***}	2.140^{**}	2.316^{***}	0.038	8.593^{**}	-1.947
	(0.193)	(0.863)	(0.820)	(1.399)	(4.346)	(4.106)
$E3_t$	0.349^{***}	-0.041	-0.032	0.160	0.436	0.231
	(0.042)	(0.174)	(0.166)	(0.317)	(0.839)	(0.854)
$R3_t$	-0.153^{***}	-0.069	-0.071	0.004	0.223	-0.001
	(0.011)	(0.049)	(0.049)	(0.076)	(0.216)	(0.193)
NAICS LC_t	0.035^{***}	0.003	0.012^{***}	0.008	0.031^{***}	0.027^{**}
	(0.004)	(0.004)	(0.004)	(0.006)	(0.011)	(0.014)
NAICS $LC_t * E_{t-1}$	0.011	0.018	0.023	0.091	0.297	-0.068
	(0.046)	(0.052)	(0.049)	(0.073)	(0.186)	(0.183)
NAICS $LC_t * E_t$	-0.130^{**}	0.048	0.027	0.061	0.133	0.060
	(0.052)	(0.057)	(0.054)	(0.086)	(0.237)	(0.265)
NAICS $LC_t * E3_t$	-0.015	-0.010	-0.011	-0.047^{**}	-0.172^{***}	-0.030
	(0.012)	(0.012)	(0.012)	(0.021)	(0.047)	(0.058)
NAICS $LC_t * R3_t$	0.016^{***}	0.009^{***}	0.008^{**}	0.010^{**}	0.012	-0.002
	(0.003)	(0.003)	(0.003)	(0.005)	(0.011)	(0.010)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28473	28473	28473	11652	2911	2914
Adjusted R^2	0.19	0.41	0.44	0.33	0.33	0.40

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.8: Robustness checks: Excluding new IPOs

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. Firm-year observations of firms that listed within the last two years are excluded. The coefficient of interest is on $LC_t * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.550^{***}	0.656^{**}	0.376	0.433	-4.487	5.057
	(0.062)	(0.296)	(0.292)	(1.090)	(3.821)	(3.114)
E_t	0.609^{***}	0.698^{**}	1.153^{***}	-1.210	9.442^{*}	-4.822
	(0.074)	(0.326)	(0.312)	(1.253)	(4.848)	(3.936)
$E3_t$	0.221^{***}	-0.049	-0.002	0.162	0.646	0.888
	(0.016)	(0.063)	(0.062)	(0.316)	(0.955)	(0.800)
$R3_t$	-0.100^{***}	-0.040	-0.049^{*}	0.112	0.550^{**}	-0.050
	(0.005)	(0.029)	(0.029)	(0.085)	(0.247)	(0.219)
LC_t	0.029^{***}	0.00001	0.007^{***}	-0.003	0.034^{**}	0.008
	(0.002)	(0.003)	(0.003)	(0.007)	(0.014)	(0.015)
$LC_t * E_{t-1}$	-0.111^{***}	-0.032	-0.030	0.008	0.347	-0.245
	(0.017)	(0.020)	(0.019)	(0.080)	(0.240)	(0.181)
$LC_t * E_t$	0.055^{***}	0.063^{***}	0.048^{**}	0.151^{*}	0.067	0.625^{**}
	(0.021)	(0.023)	(0.022)	(0.081)	(0.276)	(0.261)
$LC_t * E3_t$	-0.010^{**}	-0.006	-0.007	-0.039^{*}	-0.162^{***}	-0.160^{***}
	(0.004)	(0.005)	(0.005)	(0.021)	(0.054)	(0.055)
$LC_t * R3_t$	0.002	0.002	0.002	-0.003	-0.014	0.008
	(0.001)	(0.002)	(0.002)	(0.005)	(0.013)	(0.010)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	89692	89692	89692	11380	2845	2845
Adjusted R^2	0.17	0.38	0.41	0.32	0.32	0.40

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.9: Robustness checks: Excluding utilities

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. Firms in the Utilities industry (SIC codes 4000-4999) are excluded. The coefficient of interest is on $LC_t * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.521^{***}	0.012	-0.228	0.080	0.869	6.052^{*}
	(0.063)	(0.324)	(0.319)	(1.173)	(5.184)	(3.390)
E_t	0.599^{***}	1.324^{***}	1.693^{***}	-0.871	7.869	-8.541^{*}
	(0.075)	(0.353)	(0.344)	(1.351)	(6.319)	(4.462)
$E3_t$	0.230^{***}	-0.072	-0.020	0.302	0.234	2.128^{**}
	(0.016)	(0.068)	(0.067)	(0.353)	(1.077)	(0.971)
$R3_t$	-0.106^{***}	-0.025	-0.034	0.073	0.529^{**}	-0.190
	(0.005)	(0.031)	(0.031)	(0.100)	(0.263)	(0.215)
LC_t	0.031^{***}	0.001	0.008^{***}	-0.004	0.022	0.008
	(0.002)	(0.003)	(0.003)	(0.007)	(0.015)	(0.015)
$LC_t * E_{t-1}$	-0.114^{***}	-0.033	-0.033	0.042	-0.072	-0.082
	(0.018)	(0.020)	(0.020)	(0.082)	(0.301)	(0.179)
$LC_t * E_t$	0.065^{***}	0.067^{***}	0.052^{**}	0.110	0.401	0.346
	(0.021)	(0.024)	(0.023)	(0.084)	(0.334)	(0.245)
$LC_t * E3_t$	-0.010^{**}	-0.003	-0.004	-0.034	-0.120^{**}	-0.114^{*}
	(0.004)	(0.005)	(0.005)	(0.022)	(0.060)	(0.059)
$LC_t * R3_t$	0.003^{**}	0.002	0.002	-0.001	-0.016	0.007
	(0.001)	(0.002)	(0.002)	(0.006)	(0.013)	(0.010)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84438	84438	84438	10501	2625	2625
Adjusted R^2	0.17	0.38	0.42	0.33	0.35	0.40

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.10: Robustness checks: Only 31^{st} December fiscal year-ends

Regressions of current returns (r_t) on past and current earnings (E_t), 3 years of future earnings ($E3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. Firms which have their fiscal year-end on a date other than the 31^{st} of December are excluded. The coefficient of interest is on $LC_t * E3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.504^{***}	0.441	0.193	0.907	3.553	6.056^{*}
	(0.083)	(0.401)	(0.390)	(1.293)	(5.126)	(3.449)
E_t	0.554^{***}	0.797^{*}	1.256^{***}	-0.517	2.711	-4.662
	(0.099)	(0.434)	(0.411)	(1.701)	(6.093)	(4.146)
$E3_t$	0.258^{***}	0.028	0.037	-0.044	1.272	1.199
	(0.020)	(0.081)	(0.079)	(0.364)	(1.050)	(0.778)
$R3_t$	-0.108^{***}	-0.009	-0.016	0.134	0.476^{*}	0.073
	(0.007)	(0.039)	(0.039)	(0.098)	(0.262)	(0.220)
LC_t	0.034^{***}	0.004	0.012^{***}	0.004	0.034^{**}	0.013
	(0.003)	(0.003)	(0.003)	(0.008)	(0.016)	(0.018)
$LC_t * E_{t-1}$	-0.116^{***}	-0.045^{*}	-0.041	0.035	-0.193	-0.325
	(0.023)	(0.027)	(0.026)	(0.081)	(0.304)	(0.209)
$LC_t * E_t$	0.035	0.075^{**}	0.060^{**}	-0.004	0.227	0.466^{*}
	(0.028)	(0.031)	(0.029)	(0.088)	(0.352)	(0.273)
$LC_t * E3_t$	-0.015^{***}	-0.010^{*}	-0.011^{*}	-0.016	-0.079	-0.120^{**}
	(0.006)	(0.006)	(0.006)	(0.022)	(0.066)	(0.054)
$LC_t * R3_t$	0.004^{**}	0.005^{**}	0.005^{**}	-0.004	-0.008	0.005
	(0.002)	(0.002)	(0.002)	(0.006)	(0.013)	(0.011)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56288	56288	56288	8096	2024	2024
Adjusted R^2	0.16	0.40	0.44	0.37	0.40	0.45

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.11: Robustness checks: Using EBIT rather than EBITDA for earnings

Regressions of current returns (r_t) on past and current earnings ($EBIT_t$), 3 years of future earnings ($EBIT3_t$), future returns over three years ($R3_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with $EBIT_{t-1}$, $EBIT_t$, $EBIT3_t$, and $R3_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. The coefficient of interest is on $LC_t * EBIT3_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
$EBIT_{t-1}$	-0.642^{***}	-0.077	-0.265	0.163	-2.123	1.911
	(0.064)	(0.303)	(0.296)	(1.268)	(4.154)	(3.594)
$EBIT_t$	0.710^{***}	1.547^{***}	2.127^{***}	-0.596	7.910	-3.535
	(0.077)	(0.339)	(0.330)	(1.346)	(5.192)	(4.678)
$EBIT3_t$	0.307^{***}	-0.060	-0.003	-0.028	0.577	1.582
	(0.017)	(0.073)	(0.072)	(0.363)	(1.021)	(1.071)
$R3_t$	-0.108^{***}	-0.009	-0.016	0.098	0.404^{*}	-0.066
	(0.005)	(0.030)	(0.030)	(0.092)	(0.240)	(0.216)
LC_t	0.029^{***}	0.005^{**}	0.012^{***}	-0.001	0.025^{**}	0.012
	(0.002)	(0.002)	(0.002)	(0.006)	(0.012)	(0.013)
$LC_t * EBIT_{t-1}$	-0.090^{***}	-0.007	-0.005	0.080	0.315	-0.064
	(0.017)	(0.021)	(0.020)	(0.087)	(0.262)	(0.194)
$LC_t * EBIT_t$	0.056^{***}	0.067^{***}	0.040^{*}	0.113	0.352	0.617^{**}
	(0.022)	(0.025)	(0.024)	(0.088)	(0.308)	(0.297)
$LC_t * EBIT3_t$	-0.030^{***}	-0.006	-0.008	-0.041^{*}	-0.210^{***}	-0.251^{***}
	(0.005)	(0.005)	(0.005)	(0.023)	(0.067)	(0.073)
$LC_t * R3_t$	0.005^{***}	0.002	0.002	0.001	-0.015	0.012
	(0.001)	(0.002)	(0.002)	(0.006)	(0.012)	(0.010)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93061	93061	93061	11665	2916	2916
Adjusted R^2	0.15	0.38	0.42	0.34	0.35	0.43

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.12: Robustness checks: Using four years of future earnings and returns

Regressions of current returns (r_t) on past and current earnings (E_t), four years of future earnings ($E4_t$), future returns over four years ($R4_t$) and interactions with logged 3-digit SIC industry listing counts (LC_t). Controls include firm size, industry concentration (HHI), firm and industry M/B ratios, firm proportion of intangible assets, firm and industry age, and a loss indicator. Mediators include firm analyst coverage and firm liquidity. All controls and mediators are interacted with E_{t-1}, E_t, $E4_t$, and $R4_t$. Each variable is as described in section 6.3. Models are run on the full dataset, the 2011-2016 subsample, and the top quartile of observations in the 2011-2016 subsample based on M/B ratios and proportion of intangible assets. The coefficient of interest is on $LC_t * E4_t$.

	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
E_{t-1}	-0.496^{***}	0.149	-0.067	1.182	-3.325	8.984^{**}
	(0.064)	(0.326)	(0.324)	(1.130)	(4.548)	(3.590)
E_t	0.658^{***}	1.100^{***}	1.531^{***}	-1.976	10.261^{**}	-10.865^{***}
	(0.072)	(0.338)	(0.328)	(1.292)	(5.032)	(4.153)
$E4_t$	0.148^{***}	-0.027	0.007	0.158	-0.297	1.347^{**}
	(0.010)	(0.042)	(0.042)	(0.220)	(0.714)	(0.639)
$R4_t$	-0.086^{***}	-0.023	-0.036	0.163^{**}	0.488^{**}	-0.024
	(0.004)	(0.022)	(0.022)	(0.077)	(0.218)	(0.167)
LC_t	0.028^{***}	0.002	0.009^{***}	-0.001	0.024	0.004
	(0.002)	(0.003)	(0.003)	(0.007)	(0.015)	(0.015)
$LC_t * E_{t-1}$	-0.127^{***}	-0.035^{*}	-0.037^{*}	-0.037	0.010	-0.420^{**}
	(0.018)	(0.021)	(0.021)	(0.086)	(0.294)	(0.205)
$LC_t * E_t$	0.056^{***}	0.061^{***}	0.046^{**}	0.182^{**}	0.286	0.716^{**}
	(0.021)	(0.023)	(0.022)	(0.087)	(0.341)	(0.284)
$LC_t * E4_t$	-0.005^{*}	-0.002	-0.003	-0.028^{*}	-0.077^{*}	-0.096^{**}
	(0.003)	(0.003)	(0.003)	(0.015)	(0.044)	(0.041)
$LC_t * R4_t$	0.004^{***}	0.001	0.001	-0.006	-0.014^{*}	0.003
	(0.001)	(0.001)	(0.001)	(0.004)	(0.008)	(0.007)
Controls	No	Yes	Yes	Yes	Yes	Yes
Mediators	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	87289	87289	87289	10452	2620	2630
Adjusted R^2	0.16	0.38	0.42	0.34	0.37	0.41

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.13: Robustness checks: Using price non-synchronicity rather than FERC

$NSYNCH_t$ is stock-price non-synchronicity measured annually on daily returns. LC_t are logged 3-digit SIC industry listing counts. $NEST_t$ is analyst coverage and TO_t is firm liquidity. S_t is market capitalisation, HHI_t is industry concentration, M/B_t and $IndM/B_t$ are firm and industry M/B ratios, respectively. Int_t is a firm's proportion of intangible assets. Age_t and $IndAge_t$ are firm and industry ages, respectively. L_t is a loss indicator. All variables are as described in section 6.3. The coefficient of interest is on LC_t.

	Dependent variable: $NSYNCH_t$					
	Full Sample			2011-2016 Subsample		
	No controls	With controls	With mediators	Full subsample	High M/B	High intangibles
LC_t	0.017***	−0.003***	−0.001***	−0.003**	−0.005**	−0.002
	(0.0005)	(0.0004)	(0.0004)	(0.001)	(0.002)	(0.002)
$NEST_t$			−0.017***	−0.029***	−0.033***	−0.011**
			(0.001)	(0.002)	(0.005)	(0.005)
TO_t			−0.018***	−0.002	0.023***	0.017***
			(0.0004)	(0.002)	(0.003)	(0.004)
S_t		−0.059***	−0.048***	−0.062***	−0.058***	−0.065***
		(0.0003)	(0.0004)	(0.001)	(0.002)	(0.002)
HHI_t		−0.031***	−0.029***	−0.020***	−0.060***	−0.042***
		(0.001)	(0.001)	(0.003)	(0.005)	(0.006)
M/B_t		0.001***	0.001***	0.001***	0.0003	0.0002
		(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)
$IndMB_t$		0.003***	0.002***	0.004***	0.002***	0.004***
		(0.0002)	(0.0002)	(0.0004)	(0.001)	(0.001)
Int_t		0.038***	0.034***	0.032***	0.014	0.028*
		(0.003)	(0.003)	(0.006)	(0.011)	(0.017)
Age_t		−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
		(0.00003)	(0.00003)	(0.0001)	(0.0002)	(0.0002)
$IndAge_t$		−0.0003***	−0.0004***	−0.002***	−0.001***	−0.002***
		(0.0001)	(0.0001)	(0.0002)	(0.0004)	(0.0004)
L_t		−0.012***	−0.006***	0.008**	0.007	0.001
		(0.001)	(0.001)	(0.003)	(0.006)	(0.008)
Constant	0.772***	1.323***	1.246***	0.999***	1.318***	1.085***
	(0.002)	(0.007)	(0.007)	(0.021)	(0.045)	(0.059)
Industry dummies	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Adjusted SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	91005	91005	91005	11502	2861	2867
Adjusted R^2	0.01	0.66	0.68	0.70	0.68	0.67
Note:						*p<0.1; **p<0.05; ***p<0.01