Essays on Product Market Competition

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in Finance

at Massey University

Palmerston North, New Zealand

Suxiang Zang

2023

Abstract

Product market competition is a fundamental economic mechanism and a key topic in recent decades. In this thesis, we analyse competition measurement to improve the credibility of relevant analyses, and we study relation of firm market power with investor sentiment to fill the gap in current literature.

Essay One and Essay Two investigate a typical measure of industry concentration, the Herfindahl-Hirschman index (HHI), which is widely used to gauge competition based on industrial organisation theories. To facilitate HHI application and improve its measuring accuracy, we review the existing HHI proxies and recommend two simple HHI measures. Our survey shows that the convenient but misleading Compustat HHI is most frequently employed by researchers, while Census HHI that contains the most complete market share information of US firms is less preferred mainly due to low publication frequency and narrow industry coverage. Other HHI proxies developed recently often require extra data with complicated computation and are only occasionally employed. Comparatively, the simple HHI measures we propose are strongly correlated with the comprehensive Census HHI and are available at high frequencies for wide industries. Further, compared with Compustat HHI, the simple HHI measures better approximate Census HHI in association with important firm characteristics, and lead to more similar results as Census HHI in empirical examinations.

Essay Three explores the relation between market power and stock sensitivity to investor sentiment, on which previous studies basically keep silent. We show that firms with the weakest market power have the most susceptible returns to investor sentiment, and that return spreads between firms with high and low market power are significantly higher after optimistic sentiment than pessimistic sentiment. The return patterns across market power portfolios are more evident when sentiment is more extreme, and when sentiment later weakens than strengthens. Our baseline regressions usually show significantly positive relation between the high-minus-low market power portfolio returns and the preceding sentiment levels, which pulls through a set of robustness tests. Conclusively, our finding reveals a negative relation between market power and sentiment-driven misvaluation, consistent with the argument that market power insulates profits and reduces performance uncertainties.

Acknowledgement

First of all, I would like to express my deep and heartfelt gratitude to my supervisors: Professor Hamish Anderson and Associate Professor Yuk Ying (Candie) Chang. They lead me into this brand-new PhD journey and spend great effort with infinite patience on guiding me in my research. They enlighten me with ideas, check the methods, view my writings, give me feedbacks, correct my errors, support me for presentations and examinations, and so on and so on. They also care about my life during my study and give me warm helps whenever needed. All of these make me feel so indebted. I am lucky to have them as my supervisors, and I might never complete this PhD program without their inestimable supports.

I am grateful to Professor Ben Marshall for teaching me financial knowledge and giving me valuable suggestions for my research. I thank Professor Martin Berka, Professor David Tripe, Professor Jeff Wongchoti, and the others who gave beneficial feedbacks to my presentations in the seminars. I thank Mark woods, who has been very helpful and patient in answering my questions and providing technical supports. I thank Executive assistant Maryke Bublitz, who has always been kind and cheerful providing necessities for my study and solving problems. I thank Associate Professor Jing Chi, Associate Professor Jing Liao, Dr. Mui Kuen Yuen, Fong Mee Chin, who warmly encouraged me to pursue my goal.

I am thankful to the PhD fellow team for their warm friendship, useful suggestions to my study, beneficial discussions on the research, and pleasant conversations relieving the stress. My special thanks are to Dr. Sulu Manu O'Uiha for reading through my first essay and feedbacking with detailed suggestions for the writings. All my PhD friends are smart and kind person. It is a pleasure to work with them and to spend time with them.

Finally, I want to take this chance to thank my parents, who gave me my life, raised me up without asking for anything back. They always support me, believe in me, and are proud of me. I can never return enough to them. I thank my husband who keeps encouraging me and supporting me unreservedly. He undertakes almost all the work at home which leaves me more time to do the research. My sweet daughter trusts me and admires me wholeheartedly. She is one of the reasons that I can pull through the hard times in frustration.

PhD is a special and valuable experience to me. Here I acknowledge all the people who care for me during this journey. Your warm hearts are the most precious for me during my PhD journey and during my life.

Table of Contents

Abstract ii

Acknowledgement iii

List of Figures ix

List of Tables x

1 Chapter One: Introduction 1

1.1 Backgrounds and motivations 1

1.2 Essay One: Industry concentration measures: From an overview to simple gap-filling HHI proxies 3

1.3 Essay Two: Do relations with simple HHI measures approximate relations with Census HHI empirically? 5

1.4 Essay Three: Investor sentiment, market power, and stock returns 6

1.5 Structure of the thesis 7

2 Chapter Two (Essay One): Industry concentration measures: From an overview to simple gap-filling HHI proxies 9

2.1 Introduction 10

2.2 Applications of HHI proxies 13

2.3 Pros and Cons of existing HHI proxies 17

2.3.1 Common HHI measures 20

2.3.2 Recently introduced HHI measures 25

2.4 Simple HHI measures 32

2.4.1 Introduction 32

2.4.2 The Latest Census HHI estimation 33

2.4.3 Compustat-growth-adjusted Census HHI (CGA-HHI) estimation 33

2.4.4 Correlations of Census HHI measures with the simple HHI measures 34
2.4.5 Other straightforward Census HHI proxies 38
2.5 Conclusion 39
3 Chapter Three (Essay Two): Do relations with simple HHI measures approximate relations with Census HHI empirically? 40
3.1 Introduction 41
3.2 Industry concentration and important firm characteristics. 44
3.2.1 Variables and data 44
3.2.2 Sorts 46
3.2.3 Correlation 48
3.2.4 Bootstrapping analysis 50
3.2.5 Firm-level analysis 55
3.3 Application of simple HHI measures in empirical tests 56
3.3.1 Introduction 56
3.3.2 Methods and results 57
3.4 Conclusion 59
4 Chapter Four (Essay Three): Investor sentiment, market power, and stock returns 61
4.1 Introduction 62
4.2 Data and measures 65
4.2.1 Sentiment measurement 66
4.2.2 Market power measurement 67
4.3 Main tests on the relation between market power and stock sensitivity to sentiment 69
4.3.1 Sorts 69
4.3.2 Long-short portfolio regression 73

4.3.3 Concerns of persistent predictors ... 77

4.4 Robustness tests ... 81

4.4.1 Alternative sentiment measures ... 81

4.4.2 Alternative market power measures ... 83

4.4.3 Other robustness tests ... 85

4.5 Conclusion ... 91

5 Chapter Five: Thesis conclusion ... 92

5.1 Findings and implications ... 92

5.2 Concerns and future research ... 94

References ... 96

Appendix A for Chapter Two (Essay One) ... 110

Appendix A-1: List of articles from Journal of Finance, Review of Financial Studies, and Journal of Financial Economics applying HHI. ... 110

Appendix A-2: Proportion of Compustat to Census Data (firm number, sales value) ... 120

Appendix A-3：Descriptive statistics and correlation coefficients for Census HHI and import-adjusted HHI ... 122

Appendix A-4: CGA-HHI construction reasoning ... 124

Appendix B for Chapter Three (Essay Two) ... 126

Appendix B-1: Variable definitions ... 126

Appendix B-2: Correlations of firm variables with alternative HHI proxies for non-manufacturing industries ... 129

Appendix B-3: Replication of Mitton’s (2022) summary statistics ... 131

Appendix B-4: Average values of firm variables across portfolios sorted by HHI proxies ... 132

Appendix B-5: Correlations of firm characteristics with HHI measures (firm level) ... 133

Appendix B-6: Correlations of firm characteristics with HHI measures (alternative methods for calculating the firm characteristics) 134

Appendix B-7: Replication of empirical results reported by Ali et al. (2009) and Hou and Robinson (2006) 136

Appendix C for Chapter Four (Essay Three) 139

Appendix C-1: Variable definitions 139

Appendix C-2: Returns of market power deciles in different sentiment regimes 141

Appendix C-3: Relation between sentiment level and the following return, conditional on the following change of sentiment. 143

Appendix D: Statement of contribution 145

Appendix D-1: Statement of contribution (Chapter Two) 145

Appendix D-2: Statement of contribution (Chapter Three) 146

Appendix D-3: Statement of contribution (Chapter Four) 147

List of Figures

Figure 2-1: Composition of Census and Compustat sales 21

Figure 3-1: Sorts 47

Figure 3-2: Correlation of HHI measures with firm characteristics (industry level) 49

Figure 3-3: Bootstrapped correlation coefficients of industry-level firm characteristics with HHI measures 51

Figure 3-4: Regression of Census HHI's correlation coefficients on the other HHI measures' correlation coefficients 55

Figure 4-1: Returns of market power deciles following different investor sentiment 71

List of Tables

Table 2-1: Number and citation of articles with HHI application....................15

Table 2-2: JEL codes for the articles with HHI application....................16

Table 2-3: Comparison of US HHI measures....................18

Table 2-4: Descriptive statistics and Correlation of Census HHI, Compustat HHI, and Fitted HHIs 23

Table 2-5: Industry comparisons of Census HHI and import-adjusted HHI....................28

Table 2-6: Ranges of Census indices' correlation with the latest Census proxies and with Compustat-growth-adjusted proxies.....................36

Table 3-1: Summary statistics....................46

Table 3-2: Fama–MacBeth cross-sectional regressions of industry-level stock returns....................58

Table 4-1: Summary statistics....................68

Table 4-2: Long-short return regression on sentiment....................76

Table 4-3: Regressions in twelve subsamples, each containing a specific month from every year (Fama-Macbeth regression)....................78

Table 4-4: Simulation of the long-short return regression on artificial sentiment index....................80

Table 4-5: Long-short return regression on sentiment: alternative sentiment measures....................82

Table 4-6: Long-short return regression on sentiment: alternative market power proxies....................84

Table 4-7: Long-short return regression on sentiment: other empirical choices....................87

1 Chapter One: Introduction

1.1 Backgrounds and motivations

Product market competition is a key factor in the economic system. Theoretically, firms facing inelastic residual demand curves enjoy market power (Syverson, 2019), which facilitates them to price above marginal costs and gain excess profits. Unrestrained market power depresses a firm's motivation to compete with the rivals, and insufficient competition among firms could stifle innovation, hinder economy development, and reduce the welfare of the society (e.g., Aghion et al., 2005; De Loecker & Eeckhout, 2018; De Loecker et al., 2020). In recent decades, the secular trends of rising market concentration and increasing average firm markup raise concerns for weakened competition and strengthened market power (De Loecker et al., 2018, 2020; Grullon et al., 2019; Gutiérrez & Philippon, 2017). In the meantime, a long list of studies investigate the drivers and influences of competition and market power, which could be also implicative for enhancing effectiveness of antitrust policies for the smooth functioning of the economy. (e.g., Liu et al., 2022; Khan, 2017)

However, competition is an elusive concept and is difficult to gauge. It is of vital importance to measure competition correctly to allow for solid conclusions from empirical analyses. In the existing studies, researchers often use industry concentration as an inverse indicator of market competition based on industrial organisation theories. The basic idea is, with more market share occupied by fewer firms, the industry is more concentrated and is dominated by firms with stronger market power indicating weaker competition, and vice versa.[1] Among all the indices used to measure industry concentration, Herfindahl-Hirschman Index (HHI) has theoretical appeal and is widely employed in various areas.[2] However, a "true" HHI is difficult to obtain as it requires information of all firms in

[1] As pointed out by many researchers, there are also limitations of concentration indices and other indices could be more suitable estimates of market competition under certain circumstances, with more details specified in Chapter Five, Section 5-2. However, despite all the critics, it is still a widely accepted conjecture that concentration of market shares substantially influences the strategic activities and competition pressure in an industry, and concentration indices are extensively used in approximation of market competition in existing literature. As such, improving concentration measurement is highly meaningful.

[2] Many theoretical studies state that HHI meets most of the qualifications as a concentration measure (Adelman, 1969;

the industry. In practice, the most frequently used HHI proxy is based on Compustat data (Compustat HHI) and is argued to be misleading due to incomplete and irrelevant market share information (Ali et al., 2009; Keil, 2017). On the other hand, the US Census Bureau HHI covers the most complete market share information of US firms, but has low publication frequency and narrow industry coverage, which limit its application in empirical analyses. Other HHI proxies developed in recent studies may not solve all the major problems for Census HHI and Compustat HHI, and they mostly have complex computation process. These proxies are only occasionally adopted by researchers. As such, development of easily applicable proxies that can accurately gauge concentration is particularly necessary for improving credibility of the relevant empirical conclusions, which motivates our first two essays.

Further, market power can be heterogeneous across firms even within the same industries (De Loecker et al., 2020; Syverson, 2019), and dissimilar market power could be implicative for the cross-sectional differences in asset valuation which is a central topic in financial economics. Existing studies have shed light on the relation between firm market power and stock valuation from different dimensions. For example, researchers find that firms with stronger market power require lower capital costs (Sullivan, 1978), have more susceptible firm values to investment shocks (Garlappi & Song, 2017), present faster growth in market capitalisation since the 1980s (De Loecker et al., 2020), and have greater incentive to invest in pursuit of infinite value of being permanent leaders when interest rate approaches zero (Liu et al., 2022), etc. However, to the best of our knowledge, little attention is paid to the potential association between firm market power and the susceptibility of stock valuation to investor sentiment.

There is a growing literature that casts doubt on the traditional efficient markets hypothesis and proposes that asset valuation could be notably affected by investor sentiment, which can explain the

Hay & Morris, 1991; Marfels, 1971; Schmalensee, 1977; Tirole, 1988). Empirically, there are numerous articles using HHI as competition measure in economics, finance, accounting, etc. (Aggarwal & Samwick, 1999; Aghion & Howitt, 1992; Aghion et al, 2001, 2005; Arya & Mittendorf, 2007; Baines & Langfield-Smith, 2003; Beyer et al., 2010; Birt et al., 2006; Boone et al., 2007; Bushman & Smith, 2003; Chevalier, 1995; Dixit & Stiglitz, 1977; Eaton & Grossman, 1986; Fee & Thomas, 2004; Fresard, 2010; Giroud & Mueller, 2010, 2011; Hannah & Kay, 1977; Hay & Morris, 1991; Hou, 2007; Hou & Robinson, 2006; Ittner & Larcker, 2001; Kim et al., 2011; Klemperer, 1995; Krugman, 1979; Lang & Stulz, 1992; MacKay & Phillips, 2005; Masulis et al., 2007; Romer, 1990; Salop, 1979; Schmidt, 1997; Scott, 1999; Tirole, 1988) enduring deviations of asset prices from fundamental values (e.g., Baker & Wurgler, 2006, 2007; De Long et al., 1990; Keynes, 1936; Kumar & Lee, 2006; Shiller, 2000, 2002, 2003). In a pioneering work, Baker and Wurgler (2006) argue that investor sentiment exerts cross-sectionally different effects on different firms, and that valuation difficulties matter for the misvaluation driven by investor sentiment. Baker and Wurgler (2006) suggest that firms that are small, young, volatile, unprofitable, and non-dividend paying etc., are considered by investors to be harder to value, and therefore are more sensitive to sentiment. Other firm characteristics proposed in the literature to be related to sentiment sensitivity are also mostly consistent with the hypothesis that harder-to-value firms are more vulnerable to investor sentiment. For instance, Antoniou et al. (2016) find that firms with higher market beta gain higher returns in pessimistic sentiment, but not in optimistic sentiment. Shen et al. (2017) suggest that those firms more susceptible to macroeconomic indices have higher returns in pessimistic sentiment but not in optimistic sentiment. Under this logic, as market power exerts crucial influences on performance uncertainties or risks (Abdoh & Varela, 2017; Gaspar & Massa, 2006; Hoberg et al., 2014; Irvine & Pontiff, 2009; Kubick et al. 2015; Peress, 2010), it is possible that valuation for firms with different market power may have varied sensitivity to investor sentiment. As the existing studies basically keep silent on this potential relation between firm market power and investor sentiment, Essay Three is motivated to fill this gap.

1.2 Essay One: Industry concentration measures: From an overview to simple gap-filling HHI proxies

Essay One (Chapter Two) reviews the commonly used and recently developed HHI proxies, analysing their applications in empirical studies and their benefits and drawbacks. Furthermore, we recommend two simple HHI measures that can accurately measure concentration and are convenient to apply.

We find that the HHI proxy employed in most papers (82% in our survey) is Compustat HHI which is convenient but problematic. It is easy and natural for researchers to use Compustat statistics for HHI computation when they analyse firms with financial data available in Compustat dataset. However, Compustat HHI can hardly gauge the actual concentration level in the US market, primarily because it accounts for only a small proportion of US firms and includes irrelevant components (Ali et al., 2009; Keil, 2017). Ali et al. (2009) recommend Census HHI that captures the most complete market share information of all US firms. Nevertheless, Census HHI has low frequency, only covers manufacturing industries before 2017, and is used in a minority of empirical studies (16% in our survey). Recent studies also introduce proxies that could mitigate the problems for Census HHI and Compustat HHI, such as Census HHI floor, Fitted HHIs, the characteristics-based concentration index (Bustamante & Donangelo, 2017; Hoberg & Phillips, 2010a; Keil, 2017), etc. Nevertheless, these proxies still have associated problems and some of them are difficult to compute. Besides, researchers also consider issues such as imports, common ownerships, and industry classifications in HHI computation (Hoberg & Phillip, 2016; Gutiérrez & Philippon, 2017; Feenstra & Weinstein, 2017; Salop & O'Brien, 2000; etc.), but these considerations do not solve the basic problems related to HHI construction as mentioned above. After all, our survey shows that all these newly developed HHI proxies are only occasionally employed.

In view of the problems in the commonly used HHI proxies and the indicative preference for simple and convenient concentration measures in empirical studies, we suggest two simple HHI measures: the Latest Census HHI and the Compustat-growth-adjusted Census HHI (CGA-HHI). The Latest Census HHI is the concurrent Census HHI in Census years and is the most recent Census HHI in non-Census years. The CGA-HHI also equals Census HHI in Census years and equals the product of the Latest Census HHI and the growth of Compustat HHI in non-Census years. The simple HHI measures can be applied at high frequencies and can be extended to non-manufacturing industries based on close approximates of Census HHI, such as Census HHI floor. Our empirical examinations show that the simple HHI measures are highly correlated with Census HHI (Census HHI floor) in manufacturing (non-manufacturing) industries, with correlation coefficients ranging between 0.70-0.89 (0.92-0.95) based on alternative industry classifications. We also examine whether stability of industry concentration matters to the measuring accuracy of our suggested simple HHI measures since these proxies contain historical Census HHI information. We find that when Census HHI experiences smaller changes from the previous Census year, the Latest Census HHI is more highly correlated with Census HHI than CGA-HHI. On the contrary, when Census HH is less stable, CGA-HHI tends to outperform the Latest Census HHI with more robust correlation with Census HHI, especially based on broader industry classifications.

1.3 Essay Two: Do relations with simple HHI measures approximate relations with Census HHI empirically?

Essay Two (Chapter Three) verifies whether the simple HHI measures suggested in Essay One better estimate industry concentration than the most frequently used Compustat HHI in association with firm variables and in empirical tests. For all the examinations, we set Census HHI as the benchmark of concentration measure as it covers the most complete population of US firms and is normally considered the closest to the "true" HHI (Ali et al., 2009; Hoberg & Phillips, 2010a; Keil, 2017).

First, we analyse whether the simple HHI measures better estimate Census HHI than Compustat HHI in association with commonly considered firm characteristics. We find that the simple HHI measures are often similar as Census HHI, while Compustat HHI often deviates from Census HHI in association with the firm characteristics. The conclusions hold for both industry analysis and firm analysis. Specifically, the Latest Census HHI tends to more closely approximate Census HHI in industry analysis than in firm analysis, while CGA-HHI shows high consistency with Census HHI in both industry and firm analyses.

Second, we examine whether and how the previous conclusions based on Compustat HHI change when the simple HHI measures and Census HHI are employed respectively. We reproduce the empirical tests that are originally conducted by Hou and Robinson (2006) based on Compustat HHI and later examined by Ali et al. (2009) based on Census HHI in Census years and estimates of Census HHI in non-Census years. In our tests, we employ Census HHI, Compustat HHI, and the simple HHI measures separately and we show that using the simple HHI measures leads to similar conclusions as using Census HHI, which are both different from the conclusions based on Compustat HHI.

Essay One and Essay Two contribute to the set of studies that analyse and develop HHI proxies for better measurement of industry concentration or competition (Ali et al., 2009; Bustamante & Donangelo, 2017; Feenstra & Weinstein, 2017; Gutiérrez & Philippon, 2017; Hoberg & Phillips, 2010a, 2016; Keil, 2017; Salop & O'Brien, 2000; etc.). We propose two simple HHI measures that are available at high frequencies, extendable to a wide range of industries, easy to construct and convenient to apply. The simple HHI measures we suggest are strongly correlated with Census HHI, closely approximate Census HHI in association with key corporate characteristics, and lead to similar conclusions as Census HHI does in empirical explorations. In addition, previous conclusions based on the popular Compustat HHI may be rejected using the simple HHI measures similarly as using Census HHI.

1.4 Essay Three: Investor sentiment, market power, and stock returns

Essay Three (Chapter Four) investigates relation between firm market power and stock misvaluation driven by investor sentiment. As the efficient markets hypothesis maintained by traditional capital pricing theories cannot explain notable deviations of stock prices from fundamental values that repeatedly take place in capital market, some academics turn to behavioural finance to seek for the answer. A series of studies claim that the pervasive optimistic or pessimistic emotion in capital market could affect the way investors evaluate stock values and drive subsequent reversals in asset prices. (Baker & Wurgler, 2007; De Long et al., 1990; Keynes, 1936; Kumar & Lee, 2006; Shiller, 2000, 2002, 2003). Researchers also find that different firms may have different sensitivity to market sentiment, and that many important firm characteristics are associated with sentiment sensitivities (Antoniou et al., 2016; Baker & Wurgler, 2006, 2007; Brown and Cliff, 2004; Glushkov, 2006; Shen et al., 2017; etc.). However, to our best knowledge, firm market power is paid little attention in sentiment investigations.

Based on Baker and Wurgler's (2006, 2007) argument that valuation difficulties determine the vulnerability of stock valuation to investor sentiment, we suppose market power may influence stock sensitivity to investor sentiment based on alternative conjectures. On the one hand, firms with market power could be less sensitive to investor sentiment because market power facilitates firms to transfer idiosyncratic negative shocks to customers to insulate profits (Abdoh & Varela, 2017; Gaspar & Massa, 2006; Irvine & Pontiff, 2009; Peress, 2010), and are thus easier to value. On the other hand, firms with market power could also be more vulnerable to sentiment because they allow for more risk-taking activities (e.g., Hoberg et al., 2014; Kubick et al., 2015), which increase valuation difficulties. As the "profit-insulating" and the "risk-taking" effects influence valuation difficulties in contradictory ways, we investigate the relation between market power and sentiment-driven misvaluation through empirical examinations.

Our empirical results reveal that market power is negatively related with sentiment sensitivity for individual firms. Firms with weak market power are more likely to be overestimated (underestimated) in high (low) sentiment than firms enjoying strong market power, followed by lower (higher) stock returns. With more extreme sentiment at the beginning of the period, the conditional returns differ more between firms with high and low market power. In addition, return reversals are more evident during the periods when sentiment diminishes or transits than in the periods when sentiment aggravates. Our baseline regressions show that the return spreads between firms with high and low market power often significantly increase with the preceding sentiment, with or without controlling for the common asset pricing factors. One standard deviation increase of sentiment at the beginning of the year is averagely followed by annual raw returns (abnormal returns) 2.16%-5.28% (2.28%-3.12%) higher than that in the previous year. The significant regression results basically hold when different measures of sentiment or market power are employed, and when other empirical choices are substituted. Our examinations also suggest that the regression results are not driven by the persistency in the sentiment series as the predictor.

Essay Three contributes to both sentiment and market power literature. First, we propose that market power could be significantly related to firms' sensitivity to investor sentiment, which is absent in previous studies analysing cross-sectional differences in sentiment sensitivity (Antoniou et al., 2016; Baker & Wurgler, 2006, 2007; Brown and Cliff, 2004; Glushkov, 2006; Mian & Sankaraguruswamy, 2012; Shen et al., 2017; etc.). Second, we reveal the relation of market power with valuation efficiencies and with stock returns conditional on sentiment stages. The pervasive view in recent studies focuses on negative effects of rising market power on customer welfare, economic output, and product market efficiencies (such as Aghion et al., 2005; De Loecker & Eeckhout, 2018; De Loecker et al., 2020). From a different angle, our study implies higher valuation efficiency for firms enjoying market power in capital market as they are less likely to be misvalued when the market is dominated by emotional investors.

1.5 Structure of the thesis

The rest of the thesis is structured as follows. Chapter Two summarises the applications of alternative HHI proxies in literature, analyses their pros and cons, and suggests two simple HHI measures: the Latest Census HHI and a new CGA-HHI. Chapter Three examines how Compustat HHI and the simple HHI measures suggested in Chapter Two estimate the comprehensive Census HHI in relation with key firm characteristics and in empirical studies. Chapter Four explores relation between market power and stock sensitivity to investor sentiment. Chapter Five summarises the main findings, the implications, the latent problems, and the potential future research. Chapter Two, Three, and Four are structured as articles for publication and are intended to be submitted to journals.

2 Chapter Two (Essay One): Industry concentration measures: From an overview to simple gap-filling HHI proxies

Abstract:

We first review the commonly used, as well as recently introduced Herfindahl-Hirschman index (HHI) measures, highlighting their key features, strengths and weaknesses as concentration measures. Generally, researchers use HHI published by Census Bureau (Census HHI) or HHI based on Compustat data (Compustat HHI) to measure US industry concentration. However, both proxies have significant deficiencies. Census HHI, although normally regarded as the closest to the true HHI, has a low frequency and covers a limited range of industries. Compustat dataset does not cover the population of the firms in US industries, and thus Compustat HHI does not capture actual industry concentration and is poorly correlated with Census HHI. As such, we recommend two HHI measures that are simple to compute and apply: the Latest Census HHI and a new Compustat-growth-adjusted Census HHI estimation (CGA-HHI). The simple HHI measures have higher frequency and wider industry coverage than Census HHI, and they are more closely correlated with Census HHI than Compustat HHI. When changes of Census HHI between Census years are smaller, the Latest Census HHI has a higher correlation with Census HHI than CGA-HHI. However, for larger changes of Census HHI between the Census years, CGA-HHI can (although not always) have a higher correlation with Census HHI than the Latest Census HHI.

Keywords:

Concentration; measurement; Herfindahl-Hirschman index (HHI); concentration ratio

JEL classification:

G00; L11; L16

2.1 Introduction

Industry concentration is a key concept in a multitude of topics. It attracts considerable attention in recent decades when US industries generally become more concentrated than before (e.g., Grullon et al., 2019). Researchers commonly consider concentration an inverse measure of industrial competition in various areas of financial economics. For example, studies suggest concentration is a key factor in mergers and acquisitions (e.g., Fee & Thomas, 2004; Masulis et al., 2007), corporate governance (e.g., Aggarwal & Samwick, 1999; Boone et al., 2007; Giroud & Mueller, 2010, 2011), firm financing policy (e.g., Chevalier, 1995; Fresard, 2010; MacKay & Phillips, 2005), and asset pricing analyses (e.g., Lang & Stulz, 1992; Hou, 2007; Hou & Robinson, 2006; Kim et al., 2011).[3]

Concentration occurs when a firm has significant market share. Therefore, concentration measures are generally based on firms' market share distribution. It is crucial to measure concentration appropriately, as defective concentration estimation in data analyses could lead to incorrect conclusions and decisions. Among all concentration measures, Herfindahl-Hirschman Index (HHI) introduced by Herfindahl (1950) and Hirschman (1945) is both theoretically appealing (Adelman, 1969; Hay & Morris, 1991; Marfels, 1971; Schmalensee, 1977; Tirole, 1988), and frequently used by researchers and policymakers (Bikker & Haaf, 2002; Keil, 2017).[4] HHI is calculated as the sum of all firms' squared market shares[5], with the value ranging between 0 and 1.[6] Higher HHI value normally denotes fewer firms and less evenly distributed market shares, which implies greater industry concentration.

---

[3] Industry concentration also has a long history in other economic areas, particularly in industrial organization, innovation and invention, productivity, and international economics (Aghion & Howitt, 1992; Aghion et al, 2001, 2005; Dixit & Stiglitz, 1977; Eaton & Grossman, 1986; Hannah & Kay, 1977; Hay & Morris, 1991; Klemperer, 1995; Krugman, 1979; Scott, 1999; Salop, 1979; Romer, 1990; Schmidt, 1997; Tirole, 1988). Besides, industry concentration also attracts scholars' attention in accounting areas, such as financial reporting and managerial accounting (Arya & Mittendorf, 2007; Baines & Langfield-Smith, 2003; Beyer et al., 2010; Birt et al., 2006; Bushman & Smith, 2003; Ittner & Larcker, 2001).

[4] Typically, the US Department of Justice uses HHI as competition measure in their 1982 merger guidelines and the subsequent versions. Other examples of government agencies adopting HHI include the Council of European Union publishing guidelines on the assessment of horizontal mergers related to the Council Regulation (EC) No 139/2004 Act, and the Australian Competition & Consumer Commission carrying out the 2008 Merger Guidelines.

[5] Researchers could use different variables to capture firm's share in a market, including sales, assets, employment, stockholder equity, and market value. These proxies could be highly correlated with each other (Hannah & Kay, 1977).

[6] HHI is also computed in percentages (rather than decimals), with the possible range 0 and 10000 (i.e. 100^2).

However, HHI requires information of all market competitors, which is often difficult to obtain. It is particularly problematic when competitors with significant market shares are missing. In the US, the industrial HHI index published by the US Census Bureau (denoted as Census HHI) is generally considered the closest to the true HHI (e.g., Hoberg & Phillips, 2010a), as its underlying data is the most representative of all US firms' population. However, Census HHI is only available every 5 years and previously only for manufacturing industries. It is not until 2017 when Census HHI is extended to non-manufacturing industries. As such, researchers often use Compustat dataset to calculate HHI (denoted as Compustat HHI) which is available at higher frequencies and for more industries. However, Compustat HHI is poorly correlated with Census HHI, primarily due to the omission of most US firms and inclusion of irrelevant components. It could lead to different or even opposite conclusions compared with Census HHI (Ali et al., 2009; Keil, 2017).

Consequently, academics propose alternative HHI proxies that improve on Census HHI's data limitations, and approximate Census HHI more closely than Compustat HHI. Nevertheless, some of these proxies still have low frequency, such as Keil's (2017) Fitted HHI, while others often require extra data and/or computational effort, such as Hoberg and Phillips' (2010a) Fitted HHI. Moreover, in recent studies researchers also consider additional dimensions to better capture underlying concentration or competition. For instance, Gutiérrez and Philippon (2017) and Feenstra and Weinstein (2017) suggest import-adjusted HHI to reflect influence of imports on US industry concentration. Salop and O'Brien (2000) develop an HHI measure involving common owners' control and interest that may potentially influence industry competitiveness. Hoberg and Phillips (2016) build HHI based on a text-based industry classification (TNIC HHI) that identifies competitors by product similarity rather than by production process typically used for industry classifications.

Even though Compustat HHI is poorly correlated with Census HHI, we show that 82% of the studies in our literature survey use Compustat HHI to measure concentration, while 16% use Census HHI and even fewer studies adopt the recently developed HHI proxies. Our survey reveals a pervasive usage of the poorly performing Compustat HHI that could weaken research credibility, and implies a preference for simple HHI proxies. As such, we consider two HHI proxies that are simple to compute with commonly accessible datasets: the Latest Census HHI and a new Compustat-growth-adjusted Census HHI proxy (CGA-HHI). These simple HHI measures are available at frequencies as high as quarterly and can be extended to non-manufacturing industries using other Census Bureau concentration measures. Furthermore, they are significantly more strongly correlated with Census HHI (0.70-0.89) than Compustat HHI (0.04-0.43), and more robust than Hoberg and Phillips' Fitted HHI (0.43-0.78). Our results also suggest that CGA-HHI (the Latest Census HHI) tends to correlate with Census HHI more closely than the Latest Census HHI (CGA-HHI) when industry concentration is less (more) stable.

Another popular concentration measure is concentration ratio (CRn), usually defined as the sum of market shares for the largest "n" firms in an industry. Although CRn is criticised as being less informative (e.g., Hannah & Kay, 1977), it can be more broadly available than HHI. Researchers can also combine the information of multiple CRn to estimate HHI values. For example, Keil (2017) constructs Census HHI floor using a set of Census CRn and shows high correlations between Census HHI floor and Census HHI (0.91-0.96) across different Census years.[7] We use Census HHI floor to construct the simple HHI measures for non-manufacturing industries where Census HHI is not available, and also find strong correlations between these simple measures and Census HHI floor.

This paper contributes to the set of studies that develop HHI construction methods to better measure industry concentration or competition intensity (Bustamante & Donangelo, 2017; Feenstra & Weinstein, 2017; Gutiérrez & Philippon, 2017; Hoberg & Phillips, 2010a, 2016; Keil, 2017; Salop & O'Brien, 2000; etc.). Compared with Census HHI, the simple HHI measures we suggest are available at higher frequencies and can be extended for wider industries. Compared with Compustat HHI, the simple HHI measures have much higher correlations with Census HHI, robust for different industry classifications. Compared with some newly developed proxies (Bustamante & Donangelo, 2017; Hoberg & Phillips, 2010a), the simple HHI measures are easier to construct and apply. Besides, the simple HHI measures can also be applied in construction of import-adjusted HHI to account for the influence of imports on domestic market (Feenstra & Weinstein, 2017; Gutiérrez & Philippon, 2017).[8]

[7] Census CRn is available for the largest four (CR4), eight (CR8), twenty (CR20), and fifty (CR50) firms' market shares for manufacturing industries since 1947 and for many non-manufacturing industries since 1987. The industrial coverage of Census CRn changes over time but is always wider than Census HHI. The equation for HHI floor is: HHI floor $=4*[CR4/4]^2+(8-4)*[(CR8-CR4)/(8-4)]^2+(20-8)*[(CR20-CR8)/(20-8)]^2+(50-20)*[(CR50-CR20)/(50-20)]^2$.

[8] Feenstra and Weinstein (2017) and Gutiérrez and Philippon (2017) construct different import-adjusted HHI proxies by

Note that the Latest Census HHI has been occasionally used in previous studies but its performance has not been formally tested, to the best of our knowledge. Our research thus also lends empirical support to those studies that already use the Latest Census HHI as the concentration measure (e.g., Bustamante, 2014; Giroud & Mueller, 2011; Hartman-Glaser et al., 2019).

This study also reviews major and recently proposed US HHI estimations, in particular with respect to advantages, disadvantages, characteristics and usage counts in finance research.[9] We present the frequency of HHI application in our surveyed studies classified by JEL code. Also, we show the prevalence of different existing HHI proxies, and highlight the necessity of developing simple and robust HHI estimation in academic research.

Theoretically, there are many alternative concentration measures (e.g., the indices introduced by Hall and Tideman, 1967, Hannah and Kay, 1977, and others). While these measures are generally analytically and statistically associated with HHI (see Bikker & Haaf, 2002 and Curry & George, 1983, for surveys of concentration measures), they are less widely adopted. As we focus on empirical HHI construction, rather than comparing different theoretical concentration measures, these additional measures are out of the scope of this study.

The rest of the paper is structured as follows. Section 2.2 summarises applications of the commonly used Census HHI, Compustat HHI, and some other HHI proxies introduced recently. Section 2.3 discusses characteristics, advantages and disadvantages of the alternative HHIs. Section 2.4 introduces the simple HHI measures and assesses their robustness as Census HHI approximation. Section 2.5 concludes.

2.2 Applications of HHI proxies

In this section, we provide an overview of how prevalent different HHI proxies are in finance and in different areas of finance. For this purpose, we search all articles published in Journal of Finance (JF),

---

combining domestic HHI with different import adjustment factors. In both cases, the simple HHI measures can work as suitable domestic HHI estimates in computation of the import-adjusted HHI proxies.

[9] Industry concentration has long been a subject of inquiry. We emphasize the papers that we are more familiar with and that are, as a rule, more finance oriented. We have surely omitted reference to a number of noteworthy papers – to those authors, we apologize.

Review of Financial Studies (RFS), and Journal of Financial Economics (JFE) by the end of 2019. We identify 164 papers, published from 1992 to 2019, that measure industrial concentration or competition with at least one of the discussed HHI proxies.[10] We list these papers in Appendix A-1.

As shown in Table 2-1, our survey indicates that the most frequently used concentration measure is Compustat HHI, which appears in 82% of the papers. Compustat HHI is employed in more papers in more recent decades. This is somewhat surprising given that Ali et al. (2009) have highlighted serious flaws of Compustat HHI as a measure of concentration. The next most frequently used measure is Census HHI, employed in 16% of the surveyed papers. We also notice that when researchers use Census HHI, a few of them restrict the analysis to Census years, while others fill in non-Census years with Census HHI for the latest or nearest Census year, but do not formally test whether such adaptions are appropriate. Concerning the other HHI proxies, only a few researchers use Hoberg and Phillips' (2010a) Fitted HHI and TNIC HHI. Even fewer studies modify HHI proxies to account for special factors like the anticompetitive influence of common ownership. In short, the simple and convenient Compustat HHI is more prevalent than the more complex or inconvenient measures which may gauge concentration better. This highlights the needs of robust and convenient proxies for HHI application.

Following the 1-digit JEL classification system, Table 2-2 reveals that in the field of finance, HHI measures are most frequently employed in Corporate Finance and Governance (G3), and then in General Financial Markets (G1) according to our survey. At the level of 2-digit JEL code, HHI is the most frequently used in the categories of G32 (financing policy, financial risk, firm value, etc.) and G34 (mergers, acquisitions, restructuring and corporate governance). This is perhaps expected given the close association of competition or market power with merger and acquisition, with firm strategies and performances, and with the corporate governance functions for manager discipline. Other areas with relatively high HHI usage include asset pricing, information, and market efficiency (covering the JEL codes of G12, G14), etc.

[10] Those studies focusing on a particular industry are excluded from our searched results because specific industries may have specific datasets and/or specific variables that are more appropriate for constructing their concentration/competition measure. Examples include the market only for banks (Berger & Bouwman, 2009, 2013; Berger et al., 2005; Houston et al., 2010; Sapienza, 2004), discount firms (Khanna & Tice, 2001, 2005), carriers (Slovin et al., 1991), funds (Cremers et al., 2016; Feldman et al., 2020; Moreno et al., 2018), etc.

Table 2-1: Number and citation of articles with HHI application

The table shows the number of articles that apply at least one of our discussed HHI proxies as concentration measure, and are published in Journal of Finance, Journal of Financial Economics and Review of Financial Studies by the end of 2019. The table also reports the corresponding total number of Google Scholar citations as of 28/08/2020. A blank cell indicates a value of 0. As one paper may contain more than one HHI proxy, the paper for each HHI alternative adds up to more than the total 164 searched papers in number.

HHI Alternatives	**Census HHI**		**Compustat HHI**		**Fitted HHI (HP)**		**Common ownership adjusted HHI**		**TNIC HHI**		**HHI**	
Year of first application	**1992**		**1992**		**2010**		**2018**		**2012**		**1992**	
Journal & Period	**Number**	**Citation**	**Number**	**Citation**	**Number**	**Citation**	**Number**	**Citation**	**Number**	**Citation**	**Number**	**Citation**
Journal of Finance	**8**	**3,186**	**28**	**14,040**	**1**	**369**	**1**	**352**	**1**	**545**	**37**	**17,278**
1990-1999	2	1,173									2	1,173
2000-2009	1	125	13	9,787							14	9,912
2010-2019	5	1,888	15	4,253	1	369	1	352	1	545	21	6,193
Review of Financial Studies	**7**	**2,208**	**27**	**4,503**	**2**	**984**	**0**	**0**	**3**	**101**	**38**	**7,367**
2000-2009	2	1,339	5	1,504							6	2,422
2010-2019	5	869	22	2,999	2	984			3	101	32	4,945
Journal of Financial Economics	**12**	**4,151**	**79**	**21,688**	**5**	**1,379**	**0**	**0**	**5**	**863**	**89**	**24,360**
1990-1999			1	1,098							1	1,098
2000-2009	4	1,774	16	6,212							20	7,986
2010-2019	8	2,377	62	14,378	5	1,379			5	863	68	15,276
Sum of 3 Journals	**27**	**9,545**	**134**	**40,231**	**8**	**2,732**	**1**	**352**	**9**	**1,509**	**164**	**49,005**
1990-1999	2	1,173	1	1,098	0	0	0	0	0	0	3	2,271
2000-2009	7	3,238	34	17,503	0	0	0	0	0	0	40	20,320
2010-2019	18	5,134	99	21,630	8	2,732	1	352	9	1,509	121	26,414

Table 2-2: JEL codes for the articles with HHI application

The following tables show categorisation of papers using HHI as the concentration measure published by Journal of Finance, Journal of Financial Economics and Review of Financial Studies. Panel A classifies the papers into broadest JEL categories, and Panel B groups the papers into one-digit and two-digit JEL codes under the category of Financial Economics (G).

Panel A

JEL		Paper Number			
Code	**Definition**	**JF**	**RFS**	**JFE**	**Sum**
C	Mathematical and Quantitative Methods			2	2
D	Microeconomics	6	5	17	28
E	Macroeconomics and monetary Economics	3	1	3	7
F	Internation Economics	1		3	4
G	Financial Economics	34	38	86	158
H	Public Economics		1	2	3
J	Labor and Demographic Economics	3	3	10	16
K	Law and Economics	1	1	9	11
L	Industrial Organization	17	10	20	47
M	Business Administration and Business Economics; Marketing; Accounting; Personnel Economics	3	5	12	20
N	Economic History			1	1
O	Economic Development, Innovation, Technological Change, and Growth	2	2	12	16
P	Economic Systems			1	1
Z	Other Special Topics			1	1
	Total	37	38	89	**164**

Panel B

JEL		Paper Number			
Code	**Definition**	**JF**	**RFS**	**JFE**	**Sum**
G ▾	**Financial Economics**	**34**	**38**	**86**	**158**
G1 ▾	**General Financial Markets**	**16**	**12**	**21**	**49**
G10	General	0	1	2	3
G11	Portfolio Choice; Investment Decisions	0	1	4	5
G12	Asset Pricing; Trading Volume; Bond Interest Rates	11	6	7	24
G13	Contingent Pricing; Futures Pricing	1	1	0	2
G14	Information and Market Efficiency; Event Studies; Insider Trading	6	6	11	23
G15	International Financial Markets	0	1	0	1
G17	Financial Forecasting and Simulation	0	0	1	1
G18	Government Policy and Regulation	1	1	5	7
G2 ▾	**Financial Institutions and Services**	**4**	**12**	**15**	**31**
G20	General	0	0	1	1
G21	Banks; Depository Institutions; Micro Finance Institutions; Mortgages	1	2	5	8
G22	Insurance; Insurance Companies; Actuarial Studies	0	0	2	2
G23	Non-bank Financial Institutions; Financial Instruments; Institutional Investors	2	7	4	13
G24	Investment Banking; Venture Capital; Brokerage; Ratings and Ratings Agencies	1	4	4	9
G28	Government Policy and Regulation	1	0	2	3
G3 ▾	**Corporate Finance and Governance**	**29**	**34**	**78**	**141**
G30	General	0	4	11	15
G31	Capital Budgeting; Fixed Investment and Inventory Studies; Capacity	5	9	9	23
G32	Financing Policy; Financial Risk and Risk Management; Capital and Ownership Structure; Value of Firms; Goodwill	19	15	42	76
G33	Bankruptcy; Liquidation	0	2	8	10
G34	Mergers; Acquisitions; Restructuring; Corporate Governance	16	15	44	75
G35	Payout Policy	1	0	4	5
G38	Government Policy and Regulation	1	3	9	13
G39	Other	0	1	0	1

2.3 Pros and Cons of existing HHI proxies

In this section, we discuss the strengths and weaknesses of the commonly used and recently introduced HHI measures. Table 2-3 provides an overview.

Table 2-3: Comparison of US HHI measures

This table shows the qualities of existing HHI proxies. Census HHI is published by the US Census Bureau, summing the squared market shares of the 50 largest companies (or all companies when there are fewer than 50 firms) of each industry. Census HHI floor is computed with various Census concentration ratios (CRn), with the following equation: HHI floor = $4*[CR4/4]^2+(8-4)*[(CR8-CR4)/(8-4)]^2+(20-8)*[(CR20-CR8)/(20-8)]^2+(50-20)*[(CR50-CR20)/(50-20)]^2$. Compustat HHI is constructed on data from Compustat Fundamental files. Keil's (2017) Fitted HHI is constructed based on the regression of Census HHI on Census HHI floor and its square. Hoberg and Phillips' (2010a) Fitted HHI is calculated as the following equation: Fitted HHI = a + b*Compustat HHI + c*BLS employee number + d*Compustat employee number + e*Compustat HHI*BLS employee number + f*Compustat HHI*Compustat employee number, where a, b, c, d, e, and f are the estimated coefficients of the regression of Census HHI on Compustat HHI and employee data from both Bureau of Labor Statistics (BLS) and Compustat for manufacturing industries. Import-adjusted HHI considers import data in HHI construction, with import data from Schott's and Feenstra's websites for example. Common ownership adjusted HHI considers anti-competition effects of the common ownership, with the ownership data originated, for example, from firms' 13-F dataset. TNIC HHI is constructed on a Text-based Network Industry Classification, developed by Hoberg and Phillips (2010b, 2016), and is available up to 2017. The correlation coefficients in the last row are extracted from the studies of Keil (2017), Ali et al. (2009), and Hoberg and Phillips (2010a). The qualities include the coverage of firm information, the availability, and some specific features for each HHI proxy. "√" stands for affirmative, "×" for negative, and "-" indicates that information is unavailable or indefinite.

HHI		**Census HHI**	**Census HHI Floor**	**Compustat HHI**	**Keil's Fitted HHI**	**Hoberg and Phillips' Fitted HHI**	**US Import-adjusted HHI**	**Common Ownership Adjusted HHI**	**TNIC HHI**
A. Information Coverage	A1. US establishments of US public firms	√	√	√	√	√	√	√	√
	A2. US establishments of US non-public firms	√	√	×	√	√	-	×	×
	A3. US establishments of foreign firms	√	√	×	√	√	-	×	×
	A4 Non-US sales excluded	√	√	×	√	-	-	×	×
B. Availability	B1. Manufacturing	√	√	√	√	√	√	√	√
	B2. Non-manufacturing	×	√	√	√	√	-	√	√
	B3. Annual frequency	×	×	√	×	√	-	√	√
	B4. Starting year	1982	1947	1950	1947	-	-	-	1997
C. Specific factors	C1. Impact of imports on US local market	×	×	×	×	×	√	×	×
	C2. Impact of Co-ownership	×	×	×	×	×	×	√	×
	C3. Identify competitors by product similarity	×	×	×	×	×	×	×	√
	C4. Annual classification of competitors for each firm	×	×	×	×	×	×	×	√
D. Correlation coefficient with Census HHI		1	(0.940, 0.952)	(0.110, 0.129)	0.953	0.542	-	-	-

2.3.1 Common HHI measures

2.3.1.1 Census HHI

Census HHI is calculated by summing the squared market shares of the 50 largest companies (or all companies when there are fewer than 50 firms) of each industry, where a company is defined as having one or more establishments under common ownership or control. As discussed, Census HHI is generally recognized as the closest to the true HHI, primarily due to being based on compulsory nationwide surveys of all US firms (Ali et al., 2009). However, Census HHI is not perfect. It is only available for Census years at intervals of five years since 1982 (i.e., 1982, 1987, 1992, etc.), and the publication year is typically three years after the Census year. Also, Census HHI is restricted to manufacturing industries before 2017. As such, Census HHI is difficult to use in normal investigations that generally require information at least at a yearly basis and beyond manufacturing industries. To solve this problem, previous studies suggest proxies of Census HHI, but some of these proxies are still restricted to Census publication years (Keil, 2017) while others usually require extra information and/or computational effort (Hoberg & Phillips, 2010a; Bustamante & Donangelo, 2017).[11] Our literature survey of HHI usage tells that these HHI proxies are only occasionally adopted in the literature.

There are other limitations of Census HHI. For example, Census HHI ignores the competition from foreign firms that sell to the US market. It also overlooks the anticompetitive effect of common owners. Besides, Census HHI is based on traditional industry classifications which may not correctly classify firms from the demand side. To mitigate these problems, researchers have proposed specific methods in recent studies, which we will discuss in Section 2.3.2.[12]

Despite the imperfections, Census HHI is still commonly regarded as the benchmark for the US

---

[11] More details are discussed in Section 2.3.2.1.

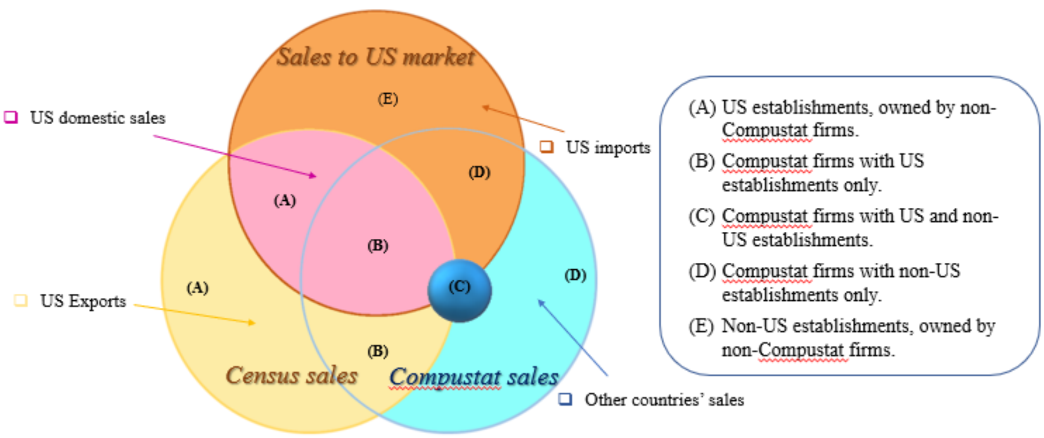
[12] Additional problems related to Census data exist with uncertain influence on HHI construction. For instance, White et al. (2012) question the accuracy of the Census data. They report that on average 27% of Census sales value for each industry are imputed, and Census imputation method underestimates the missing data's variability. However, while the quantity of missing data is large, they mention this primarily occurs in smaller establishments, which are less influential on HHI values.

industry concentration due to its wide coverage of operating establishments in the US. Therefore, we use Census HHI as our comparison basis in the following discussion of alternative measures.

2.3.1.2 Compustat HHI

The most commonly used HHI proxy in the literature is Compustat HHI, which is calculated with accounting data from Compustat Fundamental files covering all US listed companies.[13] There are several advantages of Compustat HHI compared to Census HHI. Most importantly, Compustat HHI is available at a higher frequency (annually or quarterly), across a broader range of industries, and from an earlier starting year (since 1950 for the annual file, and 1961 for the quarterly file). However, Ali et al. (2009) suggest that Compustat HHI is unlikely to capture actual industry concentration or competition intensity because it represents an incomplete list of market participants. Keil (2017) further addresses other defects of Compustat HHI and agrees with Ali et al. (2009) that empirical analysis using Compustat HHI may result in misleading conclusions. As highlighted by Ali et al. (2009) and Keil (2017), the main reasons Compustat HHI deviates from the Census HHI can be depicted in the following Venn diagram as Figure 2-1 (Sales volume is not proportional to the area shown in the diagram):

Figure 2-1: Composition of Census and Compustat sales



First, Census and Compustat datasets cover different firms. Census surveys involve all firms with operating establishments in the US while Compustat datasets include mainly public companies. As

[13] Analysis of US public companies normally relies on Compustat North American files, which records both the US and Canada public companies accounting data.

most of the firms in the US are privately owned, the number of Compustat firms is much smaller than that of Census firms. Second, Census records only include sales from the US operating establishments, while Compustat sales could be from the US and/or other countries.[14] As many public companies have businesses around the world, a large proportion of their revenues included in Compustat data are made outside of the US and not included in the US Census data. According to our calculation in Appendix A-2, Compustat firms account for around 1% of the US firms in number (for manufacturing industries), but generate 152%-215% of the Census counterparts' total revenues.[15] This mismatch between firm number and sales volume plausibly explains the main deviations between Compustat HHI and Census HHI.[16] Next, we analyse value difference and correlation between these two HHI proxies.[17,18]

As shown in Table 2-4, Compustat HHI is always much higher than Census HHI, with the mean of the former 3.9-7.8 times that of the latter. Compustat HHI is also poorly correlated with Census HHI, with correlation coefficients normally below 0.10 for finer industry classifications. However, their

---

[14] One example is Nike Inc, with the revenue around two-thirds of the total Compustat sales of industry NAICS 316 (the leather and allied product manufacturing) for the year 2012. However, Nike's major product footwear is all produced outside the US and should be excluded from Census sales.

[15] The number of firms from each industry in Census publication may add up to more than the actual total firm number because firms with establishments in multiple industries are repeatedly counted for each involved industry. However, this problem is unlikely to affect our analysis materially, since the sum of Census firm number by our calculation is at the same magnitude order as the total firm number from manufacturing Institute website (http://www.themanufacturinginstitute.org/Research/Facts-About-Manufacturing/Facts.aspx).

[16] There are other Compustat data defects such as misclassification of firms' minor operating segments, questionable industry classification, and double counting of listed subsidiaries and holding companies. More details are discussed in Keil (2017).

[17] We calculate Compustat HHI with all positive sales data obtained from Compustat Universe. As robustness check, we also calculate Compustat HHI using merged Compustat and CRSP dataset since researchers often rely on this merged data (e.g., Abdoh & Varela, 2017; Ali et al., 2009; Baker & Wurgler, 2006; Fama & French, 1992; Gaspar & Massa, 2006; Hoberg & Phillips, 2010a, 2010b; Hou & Robinson, 2006; Irvine & Pontiff, 2009; Peress, 2010). The result is similar and not reported for brevity.

[18] Previous studies show that industry classifications matter. Influentially, Bhojraj et al. (2003) mention that governmental institutions, marketers and academics widely use Standard Industry Classification (SIC) and North American Industry Classification System (NAICS), whereas economists and financial communities also commonly use the Fama-French industry classifications and Global Industry Classification Standard (GICS) for research. They suggest 2-digit SIC, 3-digit NAICS, FF48, and 6-digit GICS as commonly employed. Hence, we use these relatively broader classifications in Table 2-4, Panels A-D. We also report finer ones from Census publications in Panels E-G as supplements.

Table 2-4: Descriptive statistics and Correlation of Census HHI, Compustat HHI, and Fitted HHIs

This table shows descriptive statistics and correlation coefficients of Census HHI, Compustat HHI, and Fitted HHI. The sample consists of manufacturing industries in Census Publication years (1982, 1987, 1992, 1997, 2002, 2007, 2012). Within each panel, we keep the industry years with non-missing data for all HHI measures. The table reports Pearson correlation coefficient in the bottom left matrix and Spearman in the top right matrix. *, ** and *** indicate the significance levels at 10%, 5% and 1%, respectively. Census HHI is provided on basis of 4-digit SIC before 1997 and of 3-, 4-, 5-, 6-digit NAICS in/after 1997. Fitted HHI (Keil) sources from Keil's (2017) online appendix, Fitted HHI (HP) is computed with Hoberg and Phillips' (2010a) method as mentioned in Section 2.3.2.1. We follow Ali et al (2009) to construct HHIs based on boarder classifications by adding up the component finer industries' HHIs weighted by their squared sales share in the corresponding broader industry. The conversion tables between SIC and NAICS are obtained from the US Census website, the allocation of SIC industries to FF48 sectors follows the methods from French's website, and the transition rules from NAICS to GICS is available on Alison Weingarden's website.

Panels on industry classification	Descriptive statistics								Correlation			
	Obs	Mean	Median	SD	10%	25%	75%	90%	Census HHI	Compustat HHI	Fitted HHI(Keil)	Fitted HHI(HP)
Panel A: By 2-Digit SIC (1982-2012)												
Census HHI[iii]	126	0.017	0.010	0.022	0.004	0.006	0.018	0.037	1	0.172*	0.692***	0.657***
Compustat HHI	126	0.121	0.077	0.127	0.038	0.050	0.133	0.240	0.123***	1	0.131	0.444***
Fitted HHI(Keil)[iv]	126	0.013	0.008	0.019	0.001	0.004	0.015	0.033	0.837***	0.138	1	0.448***
Fitted HHI(HP)	126	0.019	0.013	0.018	0.005	0.009	0.022	0.040	0.673***	0.250***	0.522***	1
Panel B: By 3-Digit NAICS (1982-2012)												
Census HHI	142	0.022	0.013	0.025	0.004	0.006	0.026	0.057	1	0.308***	0.956***	0.558***
Compustat HHI	142	0.135	0.089	0.145	0.033	0.054	0.159	0.256	0.196**	1	0.245***	0.086
Fitted HHI(Keil)	142	0.022	0.013	0.024	0.004	0.006	0.025	0.061	0.960***	0.201**	1	0.554***

Fitted HHI(HP)	142	0.040	0.040	0.020	0.014	0.031	0.048	0.063	0.600***	0.376***	0.648***	1
Panel C: By FF48 (1982-2012)												
Census HHI	140	0.031	0.017	0.040	0.005	0.010	0.031	0.073	1	0.287***	0.818***	0.542***
Compustat HHI	140	0.240	0.150	0.245	0.045	0.077	0.291	0.635	0.430***	1	0.237***	0.644***
Fitted HHI(Keil)	140	0.027	0.013	0.039	0.002	0.006	0.029	0.074	0.953***	0.472***	1	0.537***
Fitted HHI(HP)	140	0.030	0.023	0.024	0.009	0.014	0.036	0.063	0.735***	0.707***	0.732***	1
Panel D: By 6-Digit GICS (1982-2012)												
Census HHI	168	0.031	0.017	0.038	0.004	0.009	0.038	0.075	1	0.371***	0.792***	0.804***
Compustat HHI	168	0.183	0.125	0.187	0.052	0.064	0.221	0.371	0.139*	1	0.365***	0.424***
Fitted HHI(Keil)	168	0.028	0.013	0.037	0.001	0.005	0.033	0.071	0.936***	0.180**	1	0.677***
Fitted HHI(HP)	168	0.034	0.025	0.035	0.005	0.011	0.041	0.070	0.779***	0.283***	0.804***	1
Panel E: By 4-Digit NAICS (1997-2012)												
Census HHI	338	0.039	0.029	0.039	0.006	0.013	0.051	0.082	1	0.128**	0.989***	0.620***
Compustat HHI	338	0.300	0.237	0.236	0.072	0.122	0.385	0.686	0.038	1	0.132**	0.296***
Fitted HHI(Keil)	338	0.040	0.029	0.038	0.006	0.013	0.053	0.087	0.981***	0.05	1	0.636***
Fitted HHI(HP)	338	0.045	0.039	0.024	0.030	0.034	0.048	0.067	0.676***	0.100*	0.701***	1
Panel F: By 5-Digit NAICS (1997-2012)												
Census HHI	665	0.058	0.039	0.057	0.009	0.018	0.078	0.137	1	0.121***	0.990***	0.602***
Compustat HHI	665	0.448	0.372	0.281	0.139	0.229	0.624	0.956	0.097**	1	0.117***	0.304***
Fitted HHI(Keil)	665	0.058	0.039	0.053	0.009	0.020	0.084	0.133	0.954***	0.075*	1	0.627***
Fitted HHI(HP)	665	0.058	0.049	0.028	0.038	0.042	0.060	0.084	0.535***	0.181***	0.595***	1
Panel G: By 6-Digit NAICS (1997-2012)												
Census HHI	1047	0.076	0.056	0.067	0.013	0.027	0.107	0.171	1	0.075**	0.987***	0.500***
Compustat HHI	1047	0.616	0.571	0.305	0.213	0.355	0.991	1.000	0.084***	1	0.078**	0.556***
Fitted HHI(Keil)	1047	0.076	0.059	0.060	0.013	0.029	0.111	0.165	0.938***	0.091***	1	0.523***
Fitted HHI(H&P)	1047	0.073	0.068	0.023	0.055	0.061	0.078	0.094	0.432***	0.302***	0.503***	1

correlations are generally higher for broad-level industry classifications. One possible reason is that more large firms are included in Compustat files in a broader industry than in a finer industry, and that share distribution should be closer to reality when information of more large firms is available.

Despite the drawbacks of Compustat HHI, in Section 2.4 we show that incorporating Compustat growth information into Census data provides a good approximation of Census HHI at high frequencies. One reason is that the degree of deviation between the Census and Compustat datasets tends to keep steady within relatively short horizons. We provide more details in Sections 2.4.3-2.4.4.

2.3.2 Recently introduced HHI measures

2.3.2.1 Fitted HHI

In view of the limited industry coverage and low frequency of Census HHI, as well as the flaws of Compustat HHI, Fitted HHIs are proposed. For instance, Keil (2017) develops a Fitted HHI based on the regression of Census HHI on Census HHI floor and its square. This Fitted HHI covers a broad range of industries and is unsurprisingly highly correlated (0.95) with Census HHI since all the data come from Census datasets. Alternatively, Hoberg and Phillips (2010a) compute another Fitted HHI with data from US Census Bureau, Bureau of Labour Statistics (BLS) and Compustat files,[19] capturing both public and private firms' information. Their Fitted HHI is available annually for manufacturing and non-manufacturing industries.

To compare the two Fitted HHIs with Compustat HHI in approximating Census HHI, we present their summary statistics and correlation coefficients with Census HHI in Table 2-4. As expected, Keil's Fitted HHI is the closest approximation of Census HHI among the three proxies, in terms of both correlation (typical above 0.90) and value difference (with mean differences between 0.000-0.004). Correlation of Hoberg and Phillips' Fitted HHI with Census HHI is moderate to strong. The mean value differences are slightly larger than that between Keil's Fitted HHI and Census HHI. Finally, Compustat HHI is the least correlated with Census HHI, with the greatest mean value differences.

[19] To construct their Fitted HHI, Hoberg and Phillips (2010a) estimate the coefficients by regressing Census HHI on Compustat HHI, Compustat employee number, BLS employee number (covering public and private firms), and interaction terms between Compustat HHI and each of the two employee numbers using manufacturing industries.

However, for Keil's Fitted HHI, a major drawback is that it is only available for the Census publication years, typically once every five years. For Hoberg and Phillips' Fitted HHI, it requires extra data and involves a relatively more complex process. Further, it assumes that the relation between concentration and employment is the same for all industries, regardless of whether the industries are labour- or capital-intensive. This assumption is yet to be formally tested.

There are other proxies introduced by researchers for similar purposes and with similar flaws as the Fitted HHIs discussed above, including the Census HHI floor introduced by Keil (2017), the characteristics-based concentration index invented by Bustamante and Donangelo (2017), etc. Here we save similar discussions of such proxies for conciseness. In addition to the Fitted HHIs and the proxies alike, recent studies also consider other factors relevant to competition in HHI estimation, including imports, common ownership, and industry classification. We briefly discuss some of these innovative proxies in the following Sections 2.3.2.2 to 2.3.3.4.

2.3.2.2 Import-adjusted HHI

International trade may intensify domestic market competition by bringing foreign rivals to the domestic stage. Some earlier studies already consider imports as necessary sales components in local markets (Curry & George, 1983; Lyons, 1981). These studies argue that omission of import information from HHI proxies, such as Census HHI, is likely to distort the true degree of competition in the local markets. Consistently, some recent studies add import information in HHI construction to measure competition in the US market. For example, Gutiérrez and Philippon (2017) construct an import-adjusted HHI using domestic HHI times the proportion of US sales in global sales to the US market.[20] Feenstra and Weinstein (2017) further consider industry concentration in the countries of the exporters that sell to US market, but this information is typically difficult and costly to obtain. In unreported tests, we correlate these two import-adjusted HHIs and find they are highly correlated.[21]

[20] Consider there are N global firms supplying to the US markets with x domestic firms, Gutiérrez and Philippon's import-adjusted HHI (HHI^{IA}) equals: $HHI^{IA} = HHI^{US} \times x/N$. Due to data limitation, Gutiérrez and Philippon use US outputs and total imports as the proxies of x and N-x respectively.

[21] We compute import-adjusted HHI proxies using methods suggested by Feenstra and Weinstein (2017) and Gutiérrez and Philippon (2017) respectively, both based on Feenstra and Weinstein's (2017) online data from 1992 to 2005. The industries are classified by 4-digit Harmonized Commodity Description and Coding System (HS). We find the correlation

To explore the influence of imports on domestic concentration, we compare Census HHI with the import-adjusted HHI built with Gutiérrez and Philippon's (2017) method[22]. The sample years are every fifth year from 1982 to 2012 when Census HHI is available. We find strong correlations between Census HHI and this import-adjusted HHI while mean values of the latter (0.012-0.061) are typically 10-20% lower than the former (0.014-0.075) based on multiple industry classifications.[23] Table 2-5 presents Census HHI and the import-adjusted HHI values for each industry classified by 2-digit SIC (Panel A) and by 3-digit NAICS (Panel B). It shows that the overestimation of domestic HHI is most serious in leather and leather product industries, with yearly average import-adjusted HHI (0.004 in both panels) equalling only 28% of the average Census HHI based on 2-digit SIC (0.014) and 33% based on 3-digit NAICS (0.013). By contrast, domestic sales dominate tobacco and printing (printing and food) industries the most, as classified by 2-digit SIC (3-digit NAICS). For these sectors, incorporation of imports reduces their domestic HHIs by less than 5% on average. For most industries, import-adjusted HHI is the closest to Census HHI in 1982 and then the ratio of import-adjusted HHI to Census HHI (I/C ratio) generally goes down, indicating ascending import penetration. Illustratively, for the apparel sector (SIC code: 31), import-adjusted HHI accounts for 80% of Census HHI in 1982 but only 29% in 2012. However, while I/C ratio could vary significantly from one year to another, those industries ranked at the top or bottom in order of I/C ratio generally keep stable in their positions through the sample years.[24]

---

coefficient between the two import-adjusted HHI estimates is 0.96, and the proxy following Feenstra and Weinstein's method has slightly smaller value on average (0.14 vs 0.16) compared to that following Gutiérrez and Philippon's (2017) method.

[22] We multiply Census HHI by the ratio of Census sales excluding exports to the sum of imports and Census sales excluding exports. International trade data are obtained from Schott's (https://faculty.som.yale.edu/peterschott/international-trade-data/) and Feenstra's websites (https://cid.econ.ucdavis.edu/usix.html) respectively for different years. We find similar results with HHI measures adjusted for foreign rivals in various ways.

[23] The correlation coefficients and mean values of the HHI proxies are reported in Appendix A-3.

[24] Our data could be subject to certain problems related to conversion between SIC and NAICS. Also, we match Census sales to international trades based on finer industry classifications (4-digit SIC and 6-digit NAICS), with higher possibility of no-match than that based on broader industry classifications. As such, we try alternative methods to address these problems. Illustratively, we classify industries by 2-digit SIC before 1997 and 3-digit NAICS since 1997 to avoid conversion between SIC and NAICS. Also, we try to match Census sales to international trades based on broader industries (2-digit SIC and 3-digit NAICS) instead of the finer ones. With all the alternative methods, positions of those industries ranked at the top and bottom by I/C ratio do not change much. All the yearly statistics are unreported.

Table 2-5: Industry comparisons of Census HHI and import-adjusted HHI

This table shows the yearly average values of Census HHI, the import-adjusted HHI, and the proportion of import-adjusted HHI to Census HHI. We show HHI proxies based on 2-digit SIC in Panel A, and 3-digit NAICS in Panel B. The original industry classifications for Census HHI are 4-digit SIC before 1997 and 3, 4, 5, 6-digit NAICS since 1997. We follow Ali et al (2009) to construct Census HHI for broader industries by adding up HHIs for the component finer industries weighted by squared sales shares of the finer industries in the broader industry they belong to. The conversion tables between SIC and NAICS are obtained from the US Census website. The import-adjusted HHI is equal to Census HHI multiplied by the ratio of import values to Census outputs excluding exports. The sample years are every fifth year from 1982 to 2012 when Census HHI is available. The international trade data are obtained from Schott's website for Census years between 1992 and 2012, and from Feenstra's website for the years 1982 and 1987.

Panel A: By 2-digit SIC

2-digit SIC	Definition	Census HHI (1)	Import-adjusted HHI (2)	(2)/(1)
31	Leather and leather products	0.014	0.004	28%
23	Apparel and other finished products made from fabrics and similar materials	0.007	0.003	46%
36	Electronic and other electrical equipment and components, except computer equipment	0.009	0.005	58%
39	Miscellaneous manufacturing industries	0.004	0.002	60%
35	Industrial and commercial machinery and computer equipment	0.005	0.003	69%
37	Transportation equipment	0.035	0.026	73%
38	Measuring, analysing, and controlling instruments; photographic, medical and optical goods; watches and clocks	0.015	0.012	82%
33	Primary metal industries	0.008	0.007	83%
28	Chemicals and allied products	0.006	0.005	84%
22	Textile mill products	0.011	0.009	86%
25	Furniture and fixtures	0.011	0.009	87%
30	Rubber and miscellaneous plastics products	0.015	0.014	88%
24	Lumber and wood products, except furniture	0.003	0.002	89%
32	Stone, clay, glass, and concrete products	0.004	0.003	91%
29	Petroleum refining and related industries	0.045	0.041	91%
26	Paper and allied products	0.009	0.008	92%
34	Fabricated metal products, except machinery and transportation equipment	0.002	0.002	92%
20	Food and kindred products	0.004	0.003	94%
27	Printing, publishing, and allied industries	0.004	0.004	97%
21	Tobacco products	0.156	0.154	99%

Panel B:

3-digit NAICS	Definition	Census HHI (1)	Import-adjusted HHI (2)	(2)/(1)
316	Leather and allied product manufacturing	0.013	0.004	33%
334	Computer and electronic product manufacturing	0.008	0.005	59%
315	Apparel manufacturing	0.004	0.003	60%
339	Miscellaneous manufacturing	0.003	0.002	66%
333	Machinery manufacturing	0.003	0.002	75%
335	Electrical equipment, appliance, and component manufacturing	0.005	0.004	77%
336	Transportation equipment manufacturing	0.041	0.032	79%
314	Textile product mills	0.029	0.024	82%
337	Furniture and related product manufacturing	0.004	0.004	84%
331	Primary metal manufacturing	0.010	0.009	85%
325	Chemical manufacturing	0.005	0.005	85%
313	Textile mills	0.007	0.006	87%
321	Wood product manufacturing	0.003	0.002	88%
312	Beverage and tobacco product manufacturing	0.038	0.034	89%
326	Plastics and rubber products manufacturing	0.021	0.019	90%
327	Nonmetallic mineral product manufacturing	0.004	0.003	90%
324	Petroleum and coal products manufacturing	0.044	0.040	91%
322	Paper manufacturing	0.009	0.009	91%
332	Fabricated metal product manufacturing	0.002	0.001	93%
311	Food manufacturing	0.004	0.003	95%
323	Printing and related support activities	0.003	0.003	97%

2.3.2.3 Common ownership adjusted HHI

Previous studies also argue that firms owned by the same shareholders tend to act in the interests of the common owners. These companies' managers are less inclined to adopt competitive policies that could harm their rival companies under common ownership[25]. For instance, Azar et al. (2018) find that acquisition between shareholders of rival companies in the US airline market is associated with reduced competition and higher ticket prices. The potential common ownership's anticompetitive effect also gains regulators' attention, such as the US and EU competition authorities, who consider

[25] Common ownership is also called partial ownership or partial common ownership by some researchers (Salop & O'Brien, 2000, Schmalz, 2018).

common ownership in their recent antitrust investigations and merger analysis (Kennedy et al., 2017; O'Brien & Waehrer, 2017).

To gauge the anti-competitive effect of common ownership, Salop and O'Brien (2000) develop a modified HHI (MHHI) by incorporating shareholders' equity portfolios and controlling rights into HHI computation.[26] MHHI offers a useful method for researchers to measure industry concentration in the presence of common ownership, provided that common ownership affects market power. Azar et al. (2018) point out that the difference between MHHI and the traditional HHI for US domestic airline industry could be 10 times higher than the HHI threshold set in the US merger guidelines. Nevertheless, MHHI may not reflect the actual anti-competitive effect of every common shareholder, such as inattentive shareholders (e.g., Gilje et al., 2020). What's more, some later studies argue that the anticompetitive effect of common owners is unwarranted based on their empirical examinations (Dennis et al., 2022, Koch et al., 2021, Lewellen & Lowry, 2021). To sum, as suggested by Schmalz (2018), there is yet no generally accepted theoretical model or measure for gauging common owners' anticompetitive impacts. Researchers are thus encouraged to further explore into this area.

2.3.2.4 HHI based on Text-based Network Industry Classification

How to accurately classify firms into different industries has long been debated (e.g., Griffith et al., 2005, Hay & Morris, 1991, Hoberg & Phillips, 2010b, 2016). As Hay and Morris (1991) state, the underlying basis of the widely used industry classifications, such as SIC, is the homogeneity on the production side, rather than the substitutability on the demand side. For example, "*plastic buckets are classified as Plastics, metal ones as Metal Working, and wooden ones as Wood Products. But clearly, these all enter the same market (as buckets) and compete with each other*" (p. 207).

Therefore, to better capture competition among firms with products that are substitutable from the users' or consumers' perspective, Hoberg and Phillips (2016) build a Text-based Network Industry Classification (TNIC) based on product similarity. They measure product similarity using

[26] MHHI is first developed by Bresnahan and Salop (1986), and then revised by Salop and O'Brien's (2000) to the following formula: MHHI = HHI + $\sum_j \sum_{k \neq j} \left(\frac{\sum_i \gamma_{ij} \beta_{ik}}{\sum_i \gamma_{ij} \beta_{ij}} \right) s_k s_j$, where β_{ij} is the fraction of firm j's equity owned by owner i, γ_{ij} stands for the weight of controlling rights over firm j by owner i, and s refers to market share.

descriptions in firms' 10-K reports filed with Securities and Exchange Commission (SEC). As 10-K files are reported annually, firms can be reclassified every year. Particularly, TNIC is firm-centric such that every firm is surrounded by its own set of competitors. On that account, those competitors of a specific company are not necessarily competing under TNIC classification. For example, when Firms A and B are both classified as Firm C's competitors, Firms A and B do not necessarily compete with each other as their products might not be similar enough.

Nonetheless, there are issues associated with TNIC HHI. First, as it is sourced from 10-K files, TNIC HHI excludes unlisted firms and includes the performance of public companies' non-US establishments. Second, TNIC is difficult to be mapped with some commonly used industry classifications, such as SIC, NAICS and FF48, and therefore there could be inconsistency when it is used with other industrial variables based on alternative industry classifications.

2.4 Simple HHI measures

2.4.1 Introduction

Ali et al. (2009) find supporting evidence consistent with theoretical predictions that the US Census HHI reflects actual industrial concentration. However, as already mentioned, Census HHI is only available for Census publication years, typically every five years, previously only covers manufacturing industries and just extends to non-manufacturing industries in 2017. Although alternative approximations of Census HHI have been introduced, they are either available infrequently (e.g., Keil, 2017), or are computationally/data demanding (e.g., Bustamante & Donangelo, 2017; Hoberg & Phillips, 2010a). In empirical studies, as highlighted in Section 2.2, researchers often rely on Compustat HHI despite its poor correlation with Census HHI.[27]

In view of the above issues associated with Census HHI and its alternatives, we suggest two simple HHI measures: the Latest Census HHI and the Compustat-growth-adjusted Census HHI (CGA-HHI). They are easily estimated, closely related with Census HHI, and available at a high frequency for a wide range of industries.

---

[27] As there is no way to calculate the actual HHI, we generally discuss alternative HHI measures with respect to Census HHI as this is considered closest to the actual or true HHI (e.g., Hoberg & Phillips, 2010a).

2.4.2 The Latest Census HHI estimation

As mentioned, the Latest Census HHI is already used by some researchers to fill in the gap of non-Census years (e.g., Bustamante, 2014; Giroud & Mueller, 2011; Hartman-Glaser et al., 2019). However, to our best knowledge, how well the Latest Census HHI measures current industry concentration has not been formally tested and justified. We therefore formally test the performance of the Latest Census HHI by analysing its association with concurrent Census HHI in the following sections. Using the Latest Census HHI to approximate the current Census HHI is based on the premise that industry concentration normally changes slowly, and therefore historical Census HHI is representative of current industry concentration. Construction of the Latest Census HHI is shown in equation (2-1):

$$\text{The Latest Census HHI}_{(t+j)} = \text{Census HHI}_{(t)} \qquad (2\text{-}1)$$

where t denotes the Census publication years. The Latest Census HHI equals the current Census HHI in Census publication years when j=0, and it equals the last preceding Census HHI when j>=1 and j<5.

To gauge concentration for non-manufacturing industries, we construct the Latest Census HHI floor since Census HHI floor is highly correlated with Census HHI and is available for non-manufacturing industries. Similarly, the Latest Census HHI floor equals the present (latest) Census HHI floor in the Census (non-Census) years as shown below.

$$\text{The Latest Census HHI floor}_{(t+j)} = \text{Census HHI floor}_{(t)} \qquad (2\text{-}2)$$

However, when industry concentration is rapidly changing, for example, during industrial restructuring, or merger and acquisitions waves, the Latest Census HHI may not always capture current industry concentration. As such, we propose the second proxy, a Compustat-growth-adjusted Census HHI measure (hereafter CGA-HHI), as described in the following Section 2.4.3.

2.4.3 Compustat-growth-adjusted Census HHI (CGA-HHI) estimation

We construct CGA-HHI using the Latest Census HHI times the growth rate of Compustat HHI. This assumes that the difference between Census HHI and Compustat HHI remains stable in relative terms over time,[28] then:

$$\frac{Census\ HHI_{t+j}}{Compustat\ HHI_{t+j}} \cong \frac{Census\ HHI_t}{Compustat\ HHI_t}, \tag{2-3}$$

and it follows that:

$$\text{CGA-HHI}_{t+j} \cong \frac{\text{Compustat HHI}_{t+j}}{\text{Compustat HHI}_t} \times \text{Census HHI}_t. \qquad (2\text{-}4)^{29}$$

Constructed as above, CGA-HHI equals Census HHI in Census years. In non-Census years CGA-HHI is equal to the Latest Census HHI multiplied by the growth of Compustat HHI from the latest Census year to the current year. We anticipate that CGA-HHI should perform better for broader industry classifications, as Compustat HHI based on broader industries tends to better approximate Census HHI than finer industries as mentioned in Section 2.3.1.2.

For non-manufacturing industries, we can also construct Compustat-growth-adjusted proxies based on close approximations of Census HHI such as Census HHI floor. We calculate Compustat-growth-adjusted Census HHI floor (CGA-HHF) as equation (2-5).[30]

$$\text{CGA-HHF}_{t+j} \cong \frac{\text{Compustat HHI floor}_{t+j}}{\text{Compustat HHI floor}_t} \times \text{Census HHI floor}_t \tag{2-5}$$

2.4.4 Correlations of Census HHI measures with the simple HHI measures

2.4.4.1 Testing strategy

In this subsection, we correlate the simple HHI measures with Census HHI only in Census years when Census HHI is available.[31] Here we examine whether the simple HHI measures combining historical

---

[28] This assumption is plausible as we discuss in Appendix A-4.

[29] Its logarithmic version, which measures relative HHI, is:
$\ln \text{CGA-HHI}_{t+j} \cong \ln \text{Compustat HHI}_{t+j} - \ln \text{Compustat HHI}_t + \ln \text{Census HHI}_t$

[30] In unreported tests, we replace Census HHI floor with Census CR4, and another Census CRn based HHI proxy respectively, which are also very highly correlated with Census HHI. Our conclusions are not affected.

[31] It is likely that in many cases, researchers just use HHI proxies as a control variable or their comparison is in relative terms. Hence, it may be enough to show how closely the simple HHI approximations reflect the variation of the underlying HHI. In untabulated tests, we also calculate the mean values of Census HHI (0.039), the Latest Census HHI (0.038), CGA-HHI (0.044), Census HHI floor (0.032), the Latest Census HHI floor (0.030), and CGA-HHF (0.031), based on different industry classifications. Basically, differences between Census measures and their simple approximations are small, and are often smaller in non-manufacturing industries than in manufacturing industries.

Census HHI information well estimate current Census HHI. For this purpose, we do not let the simple HHI measures equal Census HHI in Census years as introduced in Section 2.4.2. Instead, for manufacturing industries, we take Census HHI in last Census year as the Latest Census HHI, and compute CGA-HHI using the 5-year-lagged Census HHI multiplied by the growth rate of Compustat HHI from last Census year to the current Census year. For non-manufacturing industries, we construct the Latest Census HHI floor and CGA-HHF in the same way, except that we replace Census (Compustat) HHI with Census (Compustat) HHI floor.

The Census concentration measures are based on 4-digit SIC before 1997 and 3, 4, 5, 6-digit NAICS since 1997 for both manufacturing and non-manufacturing industries. To have consistent industry classifications, we conduct tests based on NAICS industries for years 2002, 2007 and 2012 for all industries. Additionally, for manufacturing industries, we also follow Ali et al.'s (2009) method to convert HHIs based on different finer industry classifications in different Census years to HHIs based on the same broader industry classifications through all Census years.[32] Hence, for manufacturing industries, we also analyse HHIs based on consistent broader industry classifications, with data, for Census years from 1987 to 2012.

Finally, to examine whether stability of industry concentration matters, we partition the industry-years into two groups with above and below the median absolute growth of Census measures, and then repeat the tests for each group.

2.4.4.2 Testing results

The test results are summarised in Table 2-6. Panel A reports the ranges of the correlation coefficients of Census HHI with the Latest Census HHI and with CGA-HHI for manufacturing industries. As we can see, correlation coefficients of Census HHI with the Latest Census HHI (0.70-0.89) and with CGA-HHI (0.75-0.79) are much higher than that with Compustat HHI (0.038-0.43) and more robust than that with Hoberg and Phillips' Fitted HHI (0.43-0.78), as tabulated in Table 2-4.

---

32 These broader industry classifications are 2-digit SIC, 3-digit NAICS, FF48, and 6-digit GICS, as discussed in Section 2.3.1.2. We only convert HHI for manufacturing industries as Ali et al.'s (2009) method to convert CRn in non-manufacturing industries might produce more noise.

Table 2-6: Ranges of Census indices' correlation coefficients with the simple HHI measures.

This table shows the ranges of Pearson correlation coefficients of Census HHI (floor) with the simple HHI measures.

Panel A reports correlation coefficients of Census HHI with the 5-year-lagged Census HHI and with CGA-HHI in Census publication years for manufacturing industries. CGA-HHI is computed using 5-year-lagged Census HHI multiplied by the growth from 5-year-lagged Compustat HHI to current Compustat HHI. The first two rows show the ranges of Census HHI's correlation coefficients with the Latest Census HHI and with CGA-HHI respectively based on multiple industry classifications. We classify all industry classifications into broader ones (2-digit SIC, 3-digit NAICS, FF48, and 6-digit GICS) and finer ones (4, 5, 6-digit NAICS classifications). The third row presents the ranges of differences between Census HHI's correlation coefficients with the Latest Census HHI and with CGA-HHI. The fourth (fifth) row shows the proportion of tests that show significantly higher correlations of Census HHI with CGA-HHI (the Latest Census HHI) than with the Latest Census HHI (CGA-HHI) out of all tests on different industries classifications. Sample years for tests based on broader industries are 1987, 1992, 1997, 2002, 2007, and 2012. Sample years for tests based on finer industries are 2002, 2007, and 2012. Group 1 (Group 2) demonstrates the subsample with lower (higher) than median value of absolute growth rate of Census HHI.

Panel B reports the ranges of Pearson correlations coefficients of Census HHI floor with 5-year-lagged Census HHI floor, and with CGA-HHF in Census publication years for non-manufacturing industries. For Panel B, only the Census original industry classifications (the 3, 4, 5, 6-digit NAICS industries on and after 1997) are applied and the sample years are 2002, 2007, and 2012.

Panel A: manufacturing industry-HHI

Correlation coefficient with Census HHI	Manufacturing (all)			Manufacturing (broader)			Manufacturing (finer)		
	All sample	Group-1	Group-2	All sample	Group-1	Group-2	All sample	Group-1	Group-2
The Latest Census HHI (A)	(0.697, 0.891)	(0.982, 0.993)	(0.526, 0.779)	(0.697, 0.833)	(0.982, 0.992)	(0.526, 0.657)	(0.863, 0.891)	(0.983, 0.993)	(0.743, 0.779)
CGA-HHI (B)	(0.748, 0.791)	(0.836, 0.909)	(0.641, 0.706)	(0.750, 0.773)	(0.836, 0.909)	(0.680, 0.705)	(0.748, 0.791)	(0.840, 0.896)	(0.641, 0.706)
Difference (D=B-A)	(-0.006, 0.059)	(-0.156, -0.081)	(-0.125, 0.179)	(-0.006, 0.059)	(-0.156, -0.081)	(0.031, 0.179)	(-0.137, -0.083)	(-0.143, -0.097)	(-0.125, -0.073)
Ratio of significant positive D	29%	0%	43%	25%	0%	75%	0%	0%	0%
Ratio of significant negative D	57%	100%	43%	25%	100%	0%	100%	100%	100%

Panel B: non-manufacturing industry-HHI floor

Correlation coefficient with Census HHI floor	Non-manufacturing (all)			Non-manufacturing (broader)			Non-manufacturing (finer)		
	All sample	Group-1	Group-2	All sample	Group-1	Group-2	All sample	Group-1	Group-2
The Latest Census HHI floor (A)	(0.916, 0.937)	(0.989, 0.996)	(0.776, 0.864)	0.937	0.996	0.864	(0.916, 0.931)	(0.989, 0.991)	(0.776, 0.782)
CGA-HHF (B)	(0.905, 0.948)	(0.967, 0.992)	(0.767, 0.894)	0.948	0.992	0.894	(0.905, 0.925)	(0.967, 0.977)	(0.767, 0.815)
Difference (D=B-A)	(-0.011, 0.011)	(-0.024, -0.004)	(-0.009, 0.037)	0.011	-0.004	0.03	(-0.011, -0.006)	(-0.024, -0.012)	(-0.009, 0.037)
Ratio of significant positive D	25%	0%	50%	100%	0%	100%	0%	0%	33%
Ratio of significant negative D	50%	100%	0%	0%	100%	0%	67%	100%	0%

Classifying the sample periods by the absolute growth rate of Census HHI, we find that correlations of the simple HHI measures with Census HHI are significantly stronger when the absolute growth of Census HHI is below the median (0.836-0.993) than above the median (0.526-0.779). Specifically, when Census HHI experiences smaller changes from the previous Census year, the Latest Census HHI is more highly correlated with Census HHI than CGA-HHI. On the contrary, when Census HH is less stable, CGA-HHI tends to outperform the Latest Census HHI and have higher correlations with Census HHI, based on the broader industry classifications.[33]

Panel B demonstrates ranges of the correlation coefficients of Census HHI floor with the Latest Census HHI floor and with CGA-HHF for non-manufacturing industries. The results are qualitatively similar as that in Panel A. First, both the Latest Census HHI floor and CGA-HHF are highly correlated with Census HHI floor (0.92-0.95). Second, the Latest Census HHI floor and CGA-HHF generally have lower (higher) correlations with Census HHI floor at 0.77-0.90 (0.97-1.00) for the subsamples characterised of greater (smaller) absolute growth of Census HHI floor. Third, when Census HHI floor is more (less) stable, Census HHI floor is often more highly correlated with the Latest Census HHI floor (CGA-HHF) than with CGA-HHF (the Latest Census HHI floor).

2.4.5 Other straightforward Census HHI proxies

There are other straightforward Census HHI estimates. For example, some researchers employ the previous or subsequent Census HHI, whichever is closer to the current non-Census years (Ali et al., 2009; Campello, 2006; Fresard, 2010; Klasa et al., 2009; etc). To analyse proxies containing information of both historical and future Census HHI, we calculate interpolated Census HHI in a particular Census year by averaging the previous and the next Census HHIs, and we find the interpolated Census HHI is also highly correlated with the concurrent Census HHI.[34]

Nevertheless, usage of the leading value of Census HHI is restricted when the next Census HHI is absent. For example, before the publication of 2017 Census HHI (published towards the end of 2021), the interpolated or leading Census HHI is not available for the post-2012 years. Further, using future HHI information may be inappropriate for performing predictions or when forward looking bias is a

[33] We also sort the full sample into 3-5 groups by Census HHI fluctuation, the conclusions are unchanged.

[34] The results are not tabulated but available upon request.

concern.

2.5 Conclusion

In this chapter we suggest two simple HHI measures: the Latest Census HHI and CGA-HHI. They improve on the major defects of the commonly used Census HHI and Compustat HHI. Compared with Census HHI, the simple HHI measures are available at a higher frequency and can be extended to a wider range of industries. Compared with Compustat HHI, the simple HHI measures are more strongly correlated with Census HHI and better approximate the actual industry concentration.

Our analysis further reveals that the correlation between the latest and the current Census HHI is influenced by the stability of Census HHI. Reasonably, correlation between the latest and the current Census HHI is weaker (stronger) when Census HHI experiences greater (smaller) changes from the previous Census year. Meanwhile, using CGA-HHI that incorporates the Latest Census HHI and the growth of Compustat HHI mitigates this problem to some extent. When Census HHI is less stable, CGA-HHI could be more highly correlated with Census HHI than the Latest Census HHI based on broader industry classifications.

We also provide an overview of the generally used and recently introduced HHI estimation methods in this chapter, analysing their benefits and shortcomings. We examine the correlations of Census HHI with Compustat HHI, Fitted HHIs, and the import-adjusted HHI. We show that Compustat HHI is by far the most popular HHI proxies in literature even after Ali et al. (2009) cast doubt on its measuring capability. We then highlight the needs for introducing simple and robust HHI proxies for reaching credible analytical conclusions.

3 Chapter Three (Essay Two): Do relations with simple HHI measures approximate relations with Census HHI empirically?

Abstract:

This chapter examines and finds that the simple HHI measures (the Latest Census HHI and a new Compustat-growth-adjusted Census HHI) discussed in Chapter Two approximate Census HHI in (1) association with important firm characteristics and (2) producing similar results in empirical tests. Comparatively, the commonly used Compustat HHI often fails to closely approximate Census HHI in the examinations. As Census HHI is normally considered the closest to the true HHI, analyses in this chapter provide further evidence that the simple HHI measures should be appropriate options for gauging industrial concentration while Compustat HHI may lead to invalid conclusions in empirical explorations.

Keywords:

Industry concentration; Herfindahl-Hirschman index (HHI); Census HHI; Compustat HHI; firm characteristics; stock return.

JEL classification:

G12; G30; L11; L16

3.1 Introduction

Measuring concentration is notoriously difficult. Academics normally choose Herfindahl-Hirschman Index (HHI) that captures the distribution of all firms' market shares as the concentration proxy, based on theoretical grounds (Adelman, 1969; Marfels, 1971; Schmalensee, 1977).[35] However, acquiring the required information of all firms is often impractical. As a common practice, researchers compute Compustat HHI with accounting data from Compustat database, which is available at high frequencies for a wide range of industries. Nevertheless, Compustat HHI could be misleading because Compustat firms account for only a small proportion of US firm population and Compustat sales include considerable non-US sales (Ali et al., 2009; Keil, 2017). Ali et al. (2009) therefore recommend using Census HHI published by US Census Bureau as it contains the most complete market share information of all firms in US market. However, Census HHI, although regarded as the closest to the true HHI (Hoberg and Phillips, 2010a), is published every five years and covers only manufacturing industries before 2017, which essentially limits its usage in empirical studies.

To overcome the limitations of low frequency and narrow industry coverage for Census HHI, Hoberg and Phillips (2010a) develop a Fitted HHI that is available every year for manufacturing and non-manufacturing industries. However, a presumption for this Fitted HHI is that the relation between concentration and employment is the same across different industries, which is not formally verified. Also, Hoberg and Phillips' Fitted HHI requires extra data and relatively complex computation. It shows only moderate correlation with Census HHI (0.54). Alternatively, Keil (2017) introduces Census HHI floor and a different Fitted HHI constructed with Census concentration ratios. Both indicators are accessible for wider than manufacturing industries and are highly correlated with Census HHI (above 0.90), while the major drawback is that they still have low frequency, available only in Census years. There are other proxies such as the Characteristics-Based Concentration index introduced by Bustamante and Donangelo (2017). This index is correlated with the logarithm of Census HHI at 0.39 and is also complex in construction. Despite all the efforts on improving

[35] There are many alternative concentration measures. For example, the concentration ratio (CRn), defined as the market share of the largest n firms within an industry, is widely adopted but less informative than HHI. The other concentration measures, such as the indices introduced by Hall and Tideman (1967), Hannah and Kay (1977), etc., are less widely used in literature compared to HHI.

concentration measurement in recent studies, the newly developed HHI proxies are just occasionally employed in literature. Meanwhile, Compustat HHI remains the most frequently used proxy even after Ali et al. (2009) warn that it leads to dubious conclusions.[36,37] This may indicate a preference for simple HHI proxies in empirical tests, especially when industry concentration is not the variable of major interest. Therefore, it is important to have easily applicable and appropriate concentration proxies to improve the reliability of relevant studies.

We suggest two simple HHI measures in Chapter Two, the Latest Census HHI and a new Compustat-growth-adjusted Census HHI (CGA-HHI). The Latest Census HHI refers to Census HHI published for the most recent Census year, and CGA-HHI equals the Latest Census HHI multiplied by the growth of Compustat HHI from the latest Census year to the current year.[38] The simple HHI measures can fill the gaps in non-Census years based on the conjecture that, in general, industry concentration changes slowly and the difference between Census HHI and Compustat HHI remains stable in a relatively short horizon. We have shown (in Chapter Two) that the simple HHI measures are more highly correlated with Census HHI (0.70-0.89) than Compustat HHI is (0.04-0.43). Further, they can be applied to wider than manufacturing industries based on close approximates of Census HHI, such as the Census HHI floor.

In this chapter, we use Census HHI as the benchmark and conduct further empirical tests to verify whether Compustat HHI and the simple HHI measures can be appropriate industrial concentration measures leading to valid conclusions. First, we examine how Compustat HHI and the simple HHI

---

[36] Ali et al. (2009) show that Pearson and Spearman correlation coefficients between Census HHI and Compustat HHI are vanishingly low at 0.13 and 0.11.

[37] Our survey (detailed in Essay One) shows that Compustat HHI is adopted in more than 80% papers up to 2019, Census HHI in 16% papers, Hoberg and Phillips' Fitted HHI in 5% papers. The usage of Compustat HHI does not drop after Ali et al.'s (2009) paper is published.

[38] As introduced in Sections 2.4.2-2.4.3 of Chapter Two, the simple HHI measures should equal Census HHI for Census years. However, for the examinations in this Chapter, we investigate whether the simple HHI measures containing historical Census HHI information appropriately approximate the current Census HHI. For this purpose, we follow Section 2.4.4 and only include Census years in the sample period. We take Census HHI in the last Census year as the Latest Census HHI, and we compute CGA-HHI using the 5-year-lagged Census HHI multiplied by the growth rate of Compustat HHI from the last Census year to the current Census year. Our test results concerning the close relationships between Census HHI and the simple HHI measures should be conservative because the simple HHI measures for non-Census years contain historical information of Census HHI shorter than five years ago.

measures estimate Census HHI in association with a number of commonly used firm characteristics. The firm variables and proxies we choose are most frequently used in the literature as suggested by Mitton (2022). Second, we reproduce the empirical tests conducted by Hou and Robinson (2006) and examined by Ali et al. (2009), and we compare the outcomes using Census HHI, Compustat HHI and the simple HHI measures. All our test results support that Compustat HHI generally differ significantly from Census HHI, while the simple HHI measures often closely approximate Census HHI, in relation with other firm characteristics and in generating empirical outcomes.

This chapter contributes to a set of studies that aim to improve concentration measurement (Ali et al., 2009; Hoberg & Phillips, 2010a; Keil, 2017). In an influential work, Ali et al. (2009) highlight the problems of Compustat HHI and recommend Census HHI to be a better concentration proxy. However, they do not provide credible solutions to the problem that Census HHI is only available in the quinquennial Census years. In their empirical tests, Ali et al. (2009) fill in non-Census years with Census HHI in the nearest Census year without formally testing its performance and justifying this usage. Our study formally tests and suggests that the simple HHI measures we suggest are good approximates of Census HHI in association with key corporate characteristics and in examining previous empirical tests. In this way, we encourage the application of the comprehensive Census indices and their close approximates in normal empirical studies that require yearly data. We also provide empirical support to previous studies that already employ the Latest Census HHI as the concentration measure without formally testing its performance (e.g., Bustamante, 2014; Giroud & Mueller, 2011; Hartman-Glaser et al., 2019).

Our empirical results also provide evidence that Compustat HHI may lead to different results from Census HHI in relation with other firm characteristics. This study is supplementary to Keil's (2017) study that shows correlations of several prominent firm variables with Compustat HHI are dissimilar as those with Census HHI. In our investigation, we conduct comprehensive examinations including sorting tests, correlations, and bootstrapped correlations. We analyse the relations between HHI proxies and firm variables at both the industry and firm levels. In addition, most of the proxies we choose are not covered by Keil's (2017) selection of firm variables.

The rest of this chapter is constructed as follows. Section 3.2 investigates the relation of commonly studied firm characteristics with industry concentration proxied by Census HHI, Compustat HHI, and the simple HHI measures. Section 3.3 examines the previous study done by Hou and Robinson (2006) who use Compustat HHI as the major concentration measure. We replace Compustat HHI with Census HHI and the simple HHI measures and check whether and how the original results should change. Section 3.4 concludes.

3.2 Industry concentration and important firm characteristics.

In this section we examine how Compustat HHI and the simple HHI measures approximate Census HHI in association with important firm variables. Section 3.2.1 introduces all the variables and data sources.

3.2.1 Variables and data

Census HHI is published by US Census Bureau, calculated by summing the squared market shares of the 50 largest companies (or all companies when there are fewer than 50 firms) in an industry. We obtain Census HHI from Census website, with the first publication for 1982, then every five years until 2017. The industry classifications for Census HHI are 4-digit SIC before 1997 and 3, 4, 5, 6-digit NAICS since 1997. In accordance, we analyse industries classified by 4-digit SIC before 1997 and a comparable 6-digit NAICS since 1997 throughout this chapter. The other HHI proxies are also calculated in Census years. Compustat HHI is computed as the sum of squared sales share of all firms in an industry based on merged datasets of Compustat and the Centre for Research in Security Prices (CRSP).[39] As mentioned above, here the Latest Census HHI equals the 5-year-lagged Census HHI, and CGA-HHI is the 5-year-lagged Census HHI multiplied by the growth rate from 5-year-lagged Compustat HHI to current Compustat HHI. Because both the simple HHI measures require antecedent based on the same industry classification, years 1982 and 1997 are excluded as the starting years based on SIC and NAICS respectively. The main tests in this chapter cover manufacturing industries

[39] The industries for computing Compustat HHI are classified by CRSP historical 4-digit SIC before 1997 and Compustat historical 6-digit NAICS since 1997. Using CRSP SIC codes follows Ali et al. (2009), who cite the standpoint from Kahle and Walkling (1996) that using historical CRSP SIC codes is advantageous to the Compustat SIC codes over long sample periods. However, as the NAICS code is often missing in CRSP data, we use Compustat NAICS codes for the years on and after 1997.

where Census HHI is available. In Appendix B-2, we also show the conclusions are qualitatively similar for non-manufacturing industries using Census HHI floor in the place of Census HHI.

The firm characteristics we analyse include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability. To estimate these firm characteristics, we employ the most frequently used measurements as suggested by Mitton (2022). To specify, we measure Firm Value with total assets (Compustat mnemonics: AT) less total common equity (CEQ) plus Compustat market equity, all divided by total assets. Market equity is the Compustat stock price times shares outstanding at the end of the fiscal years. Leverage is the long-term debt (DLTT) plus debt in current liabilities (DLC), then divided by total assets. Investment equals total annual capital expenditures (CAPX) divided by total assets. Innovation is calculated with total annual research and development costs (XRD) divided by total assets. Cash Holding is the cash and short-term investments (CHE) divided by total assets. Payout refers to the annual common/ordinary dividends (DVC) divided by total assets. Profitability is the annual earnings before interest, taxes, depreciation, and amortization (EBITDA) divided by total assets. Summary statistics of these firm variables shown by Mitton (2022) can be closely replicated during his sample period (1963-2018).[40] The industry values of the firm variables are the simple average of individual firms within each industry.

Table 3-1 reports the summary statistics of the HHI proxies and industry average value of the discussed firm variables in six Census years (year 1987, 1992, 2002, 2007, 2012, and 2017). We have 1387 industry-years in total, with non-missing value of Census HHI, Compustat HHI, and the simple HHI measures. For all variables, we show the mean and standard deviation, as well as the critical values ranked at the proportion of 25%, 50%, and 75%. All variables are winsorised at 1% and 99%, which is the most popular method for researchers to exclude outliers according to Mitton (2022). For all types of statistics shown in Table 3-1, the simple HHI measures are much closer to Census HHI than Compustat HHI is. For example, the mean value of Compustat HHI (0.669) is more than nine times that of Census HHI (0.073), while the Latest Census HHI (0.070) and CGA-HHI (0.081) differ from Census HHI by less than 11%.

[40] Definitions for all the variables and the data cleaning methods are summarised in Appendix B-1, Replication of Mitton's (2022) summary statistics for the involved firm variables are reported in Appendix B-3.

Table 3-1: Summary statistics

This table reports the mean value, the critical values at 25%, 50%, 75%, and the standard deviation of multiple HHI measures, important firm characteristics, and stock return. The firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability, with details described in Appendix B-1. HHI proxies and the firm variables are at yearly frequency while stock returns are monthly. The sample covers manufacturing industries matched by both Census and the merged datasets of Compustat and CRSP files in Census years 1987, 1992, 2002, 2007, 2012, and 2017. Industries are classified by 4-digit SIC (6-digit NAICS) before (after) 1997. All variables are averaged at industry level and winsorised at 1% and 99%.

Variable	Measure	Obs	Mean	25%	50%	75%	SD
Industry Concentration	Census HHI	1,387	0.073	0.027	0.054	0.103	0.062
	Compustat HHI	1,387	0.669	0.420	0.654	1.000	0.289
	Latest Census HHI	1,387	0.070	0.025	0.053	0.097	0.060
	CGA-HHI	1,387	0.081	0.028	0.056	0.108	0.078
Firm Value	(TA-BE+ME)/TA	1,387	1.647	1.164	1.418	1.909	0.768
Leverage	Total debt/TA	1,387	0.243	0.150	0.226	0.322	0.140
Investment	CAPX/TA	1,382	0.046	0.027	0.040	0.056	0.028
Innovation	R&D/TA	1,387	0.025	0.001	0.010	0.030	0.042
Cash Holding	(Cash + Equivalents)/TA	1,387	0.121	0.044	0.091	0.166	0.111
Payout	Dividends/TA	1,387	0.014	0.002	0.009	0.019	0.017
Profitability	EBITDA/TA	1,386	0.102	0.073	0.118	0.155	0.109
Stock return	Holding period return (%)	15,332	1.085	-4.225	1.018	6.304	11.647

3.2.2 Sorts

In this subsection, we analyse how the important corporate characteristics vary across industry portfolios sorted by alternative HHI proxies. If an HHI proxy closely approximates Census HHI, the firm characteristics should vary across industry portfolios ranked by Census HHI and by this HHI proxy in a similar manner. To verify, we categorise all the industries to three groups by the level of Census HHI, Compustat HHI, and the simple HHI measures respectively. Groups 1 and 3 contain 30% industries with the lowest and highest HHI levels respectively in each year, Group 2 contains the middle 40%. We present the average value of the firm variables across different HHI groups with pillars in Figure 3-1. The edge of a pillar with (without) a dark rhombus represents average value of a specific firm variable for Group 3 (1) sorted by an HHI proxy and the hollow round represents the value for Group 2. The border of a pillar is in solid line (dash line, dots) if the spread between Groups 3 and 1 is significantly different from zero at the level of 1% (5%, 10%). All variables are standardised to be comparable.[41]

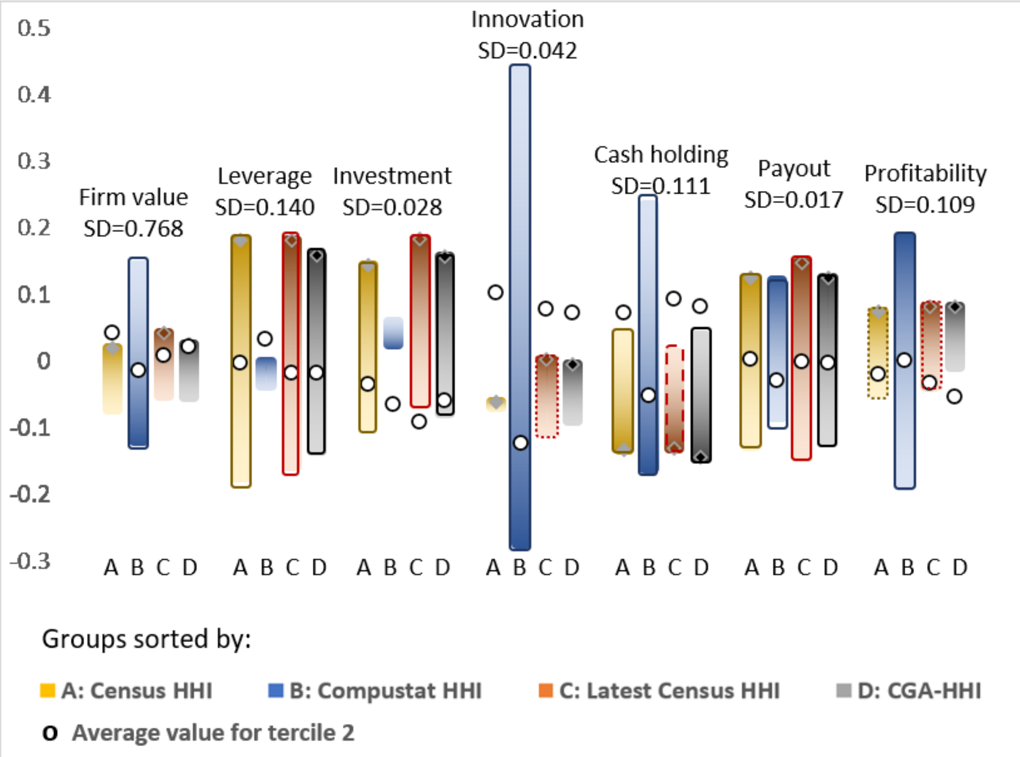
Figure 3-1 shows that most of the firm characteristics vary with Census HHI and with Compustat HHI in different manners. Specifically, Firm Value, Investment, and Innovation increase with Census HHI, but decrease with Compustat HHI. Leverage rises significantly from low to high Census HHI groups but insignificantly through the Compustat HHI groups. Profitability goes up from the bottom to the top groups sorted by both Census HHI and Compustat HHI, but at different significance levels. Cash Holding decreases through Groups 1 to 3 sorted by Census HHI and Compustat HHI at the same level of significance, but with rather dissimilar spreads. Only Payout shows the same variation between groups sorted by Census HHI and Compustat HHI at the same significance level and with comparable spreads. Comparatively, all the firm variables we analyse vary from low to high groups sorted by Census HHI in the same directions as that sorted by the simple HHI measures and with similar spreads in value. The spreads between high and low groups sorted by the simple HHI measures are typically significant or insignificant similarly as that sorted by Census HHI. In sum, Figure 3-1 shows that variations of the commonly studied firm variables across industries ranked by Census HHI are basically mimicked by their variation across industries ranked by the simple HHI measures but not by Compustat HHI.

Figure 3-1: Sorts

Figure 3-1 shows mean values of commonly studied firm variables across manufacturing industries sorted by Census HHI, Compustat HHI, the Latest Census HHI and CGA-HHI. Group 1 (3) contains 30% industries with the smallest (greatest) value of an HHI proxy in each year, Group 2 contains the middle 40%. The edge of a pillar with (without) a dark rhombus represents value for Group 3 (1) and the hollow round for Group 2. The firm variables include Firm Value, Leverage, Investment,

[41] Details are reported in Appendix B-4.

Innovation, Cash Holding, Payout, and Profitability, defined in Appendix B-1. All the variables are standardised. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The pillars are surrounded with full lines (dash lines, dotted lines) when the differences between Groups 3 and 1 are significant at level of 1% (5%, 10%).



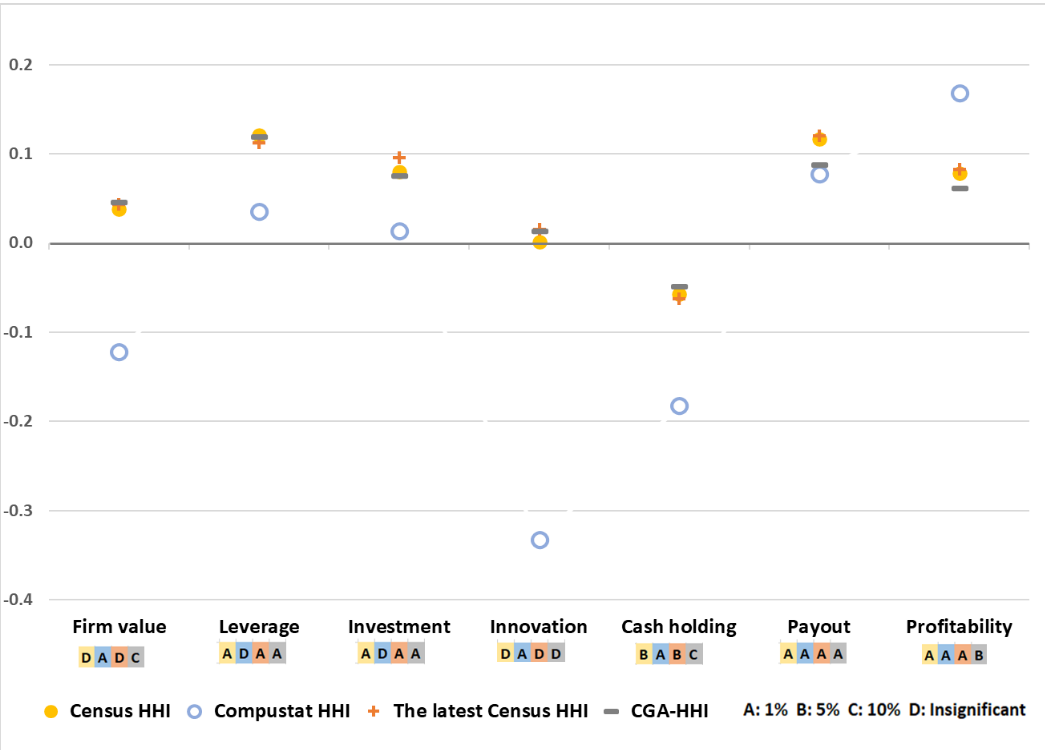
3.2.3 Correlation

We have already shown in Chapter Two that Compustat HHI is weakly correlated with Census HHI, whereas the simple HHI measures are much more highly correlated with Census HHI. In this subsection we correlate different HHI proxies with commonly studied firm variables and show how Census HHI is approximated by the other HHI proxies in correlation with these firm variables.

In Figure 3-2, we denote firm variables on the horizontal line and plot their correlation coefficients with different HHI measures on the vertical axis. We can see, for Firm Value and Innovation, the correlations with Compustat HHI are in opposite directions to that with Census HHI. For Leverage and Investment, the correlations are significantly positive with Census HHI but insignificantly positive with Compustat HHI. For Cash Holding, correlations with the two HHI measures are at dissimilar significance levels. Comparatively, Census HHI has the same signs as the simple HHI measures in correlation with all the firm variables. Census HHI and the Latest Census HHI (CGA-HHI) are at the same significance levels, or both insignificantly correlated with all (most of) the firm variables. In addition, differences between Census HHI and the simple HHI measures in correlation with the firm variables (-0.016 to 0.030) are much smaller in absolute value than that between Census HHI and Compustat HHI (-0.090 to 0.334). To recap, correlations of important firm characteristics with Census HHI can be largely estimated by that with the simple HHI measures but usually misrepresented by that with Compustat HHI in our tests.

Figure 3-2: Correlations of HHI measures with firm characteristics (industry level)

This figure shows correlations of HHI measures with firm characteristics at industry level for manufacturing industries. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The HHI measures include Census HHI, Compustat HHI, the Latest Census HHI, and CGA-HHI. The firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability, defined in Appendix B-1.



3.2.4 Bootstrapping analysis

The correlation coefficients estimated in the above Section 3.2.3 are based on a sample of the US firm population. In this subsection, we bootstrap the sample to show the distribution of numerous estimates of the correlation coefficients that could be more comprehensive than that from one single sample.

To bootstrap the sample, we randomly select half of the observations with replacement from every Census year and every group sorted by Census HHI[42] 10,000 times to form 10,000 new samples. For each new sample we correlate the firm variables with different HHI measures in each Census year. In total, we have 6 (Census years) * 10,000 correlation coefficients between each firm variable with each HHI measure. We then analyse these bootstrapped correlations as the following.

First, we plot correlation coefficients of the firm variables with Census HHI on the horizontal axis against that with Compustat HHI (the Latest Census HHI, CGA-HHI) on the vertical axis in the first (second, third) scatter graph. The points denoting more frequently obtained correlation coefficients are in darker colour.[43] In each scatter graph, we also plot the Census HHI correlation coefficients at both coordinates (called the Census line hereafter). Census line is a straight line through the origin at 45 degrees. Reasonably, the closer the points approach to the Census line, the better that particular HHI measure estimates Census HHI in correlation with the firm variables. All the graphs are summarised in Figure 3-3, with Panels A-G presenting correlations with different firm variables discussed in Section 3.2.1. We can see that among the three scatter graphs in each panel, the first one (Census HHI vs Compustat HHI) presents the largest area of points departing from the Census line. Further, the second and third graphs (Census HHI vs the simple HHI measures) show obviously darker points adjoining the Census line, indicating tighter dispersion for more correlation coefficients. These imply that compared to Compustat HHI, the simple HHI measures estimate Census HHI more closely in correlation with firm variables.

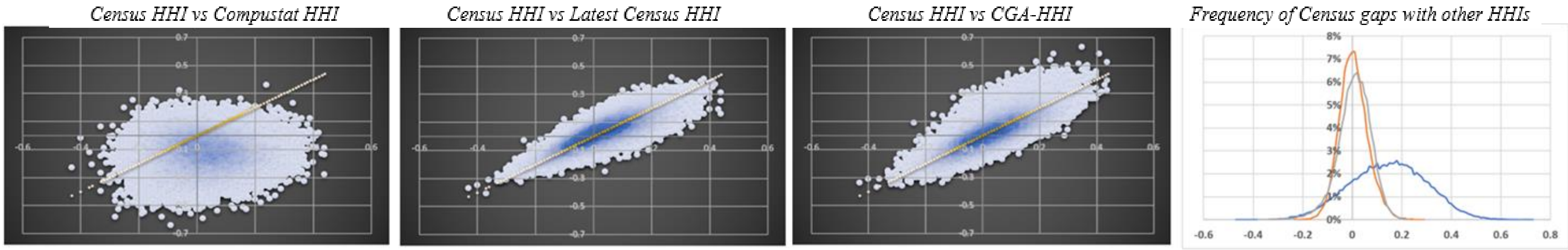
[42] Firms are partitioned into three groups ranked by Census HHI as described in Section 3.2.2 for the sorting tests.

[43] We round the regression coefficients to two decimals and multiple regression coefficients may share the same value.

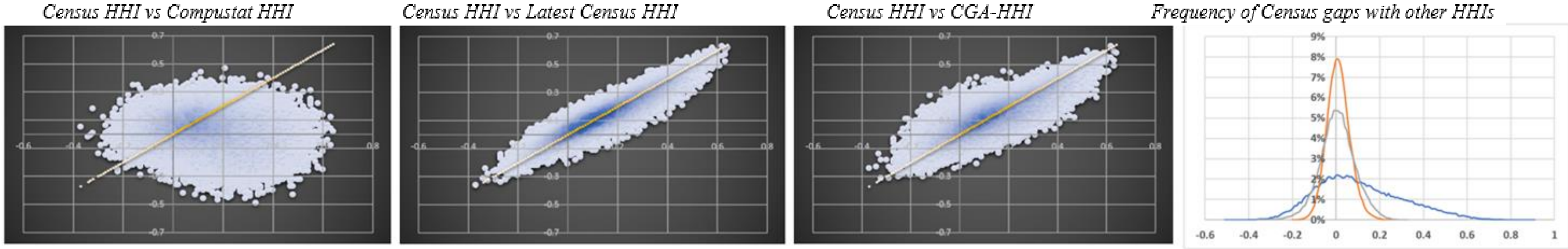
Figure 3-3: Bootstrapped correlation coefficients of firm characteristics with HHI measures

Figure 3-3 shows bootstrapped correlations of firm characteristics with Census HHI, Compustat HHI, the Latest Census HHI, and CGA-HHI for manufacturing industries. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability, all defined in Appendix B-1. Panels A-G present correlation coefficients of different firm characteristics. For each panel, we randomly select half observations as described in Section 3.2.4 and correlate the firm variable with different HHI measures in every Census year. We repeat the process 10,000 times, obtaining 6 (years) * 10,000 correlation coefficients of each firm variable with each HHI. For all the panels, a point in the first (second, third) scatter graph shows the correlation with Census HHI on the horizontal axis against that with Compustat HHI (the Latest Census HHI, CGA-HHI) on the vertical axis, based on the same set of randomly selected observations. The more (less) frequently appearing points are painted with darker (lighter) colour. In each scatter graph, we also draw a set of linear dots at 45 degrees through the origin, representing Census HHI at both coordinates (the Census line). The fourth chart shows the distribution of the gaps between correlations of firm variables with Census HHI and that with Compustat HHI (blue line), with the Latest Census HHI (orange line), and with CGA-HHI (grey line). The horizontal axis denotes the magnitudes of the gaps and the vertical axis denotes the frequencies for the gaps to appear.

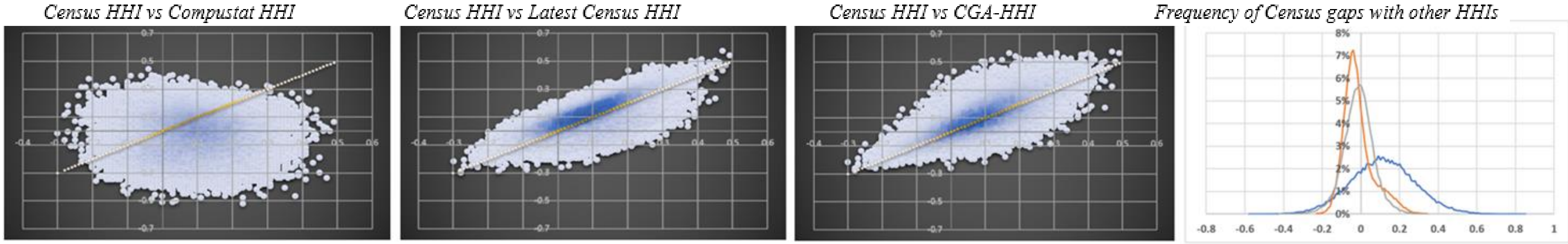
Panel A: Correlation coefficients of Firm Value with multiple HHI measures.



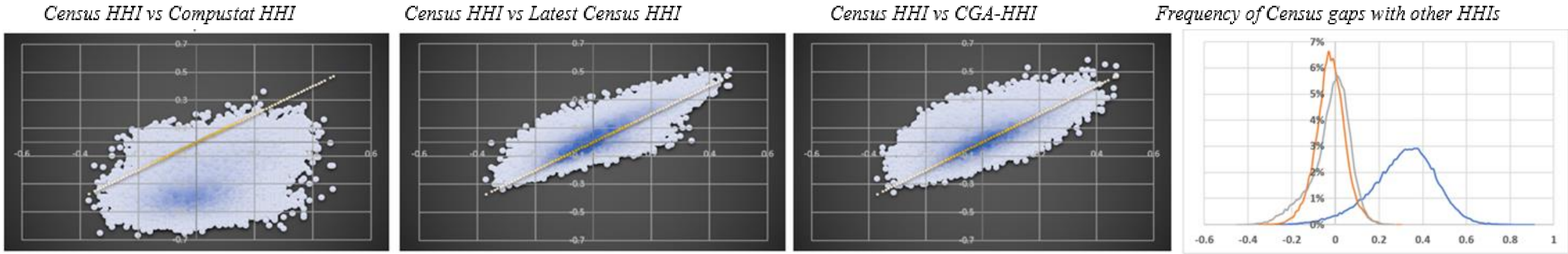
Panel B: Correlation coefficients of Leverage with multiple HHI measures.



Panel C: Correlation coefficients of Investment with multiple HHI measures.

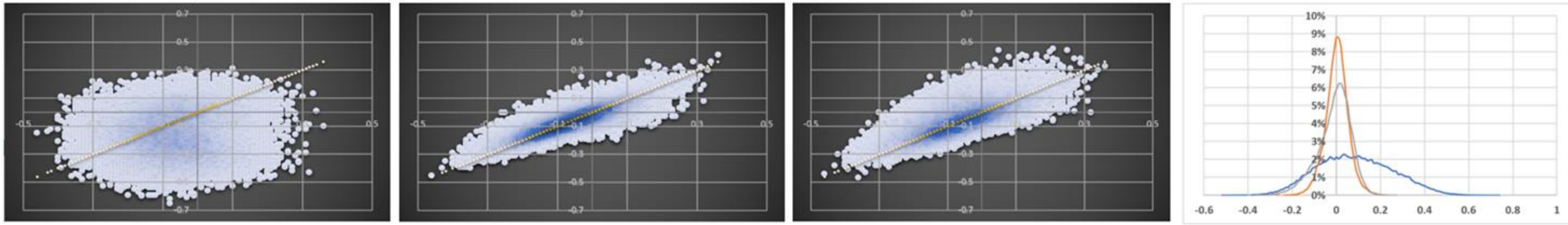


Panel D: Correlation coefficients of Innovation with multiple HHI measures.



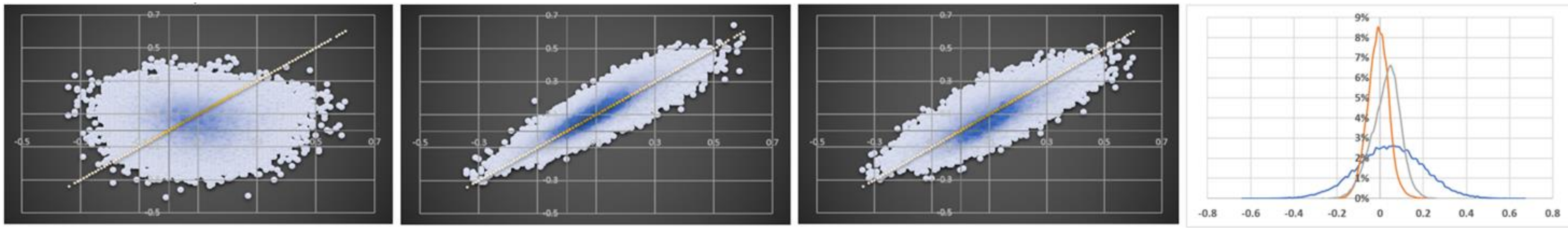
Panel E: Correlation coefficients of Cash Holding with multiple HHI measures.

Census HHI vs Compustat HHI | *Census HHI vs Latest Census HHI* | *Census HHI vs CGA-HHI* | *Frequency of Census gaps with other HHIs*



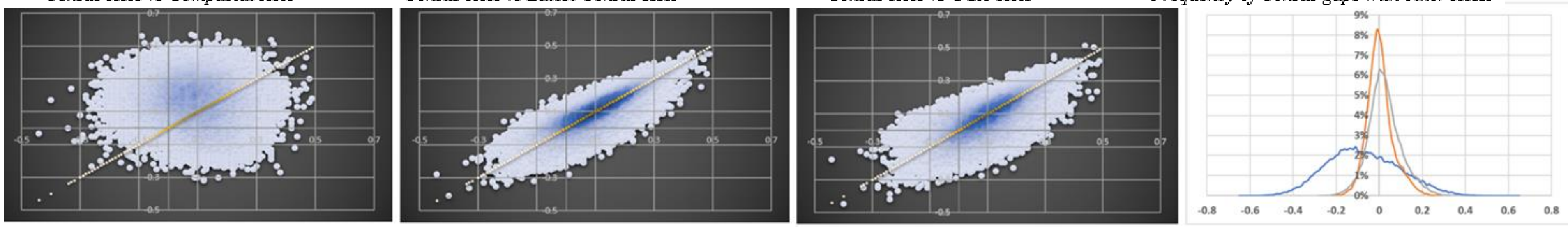
Panel F: Correlation coefficients of Payout with multiple HHI measures.

Census HHI vs Compustat HHI | *Census HHI vs Latest Census HHI* | *Census HHI vs CGA-HHI* | *Frequency of Census gaps with other HHIs*



Panel G: Correlation coefficients of Profitability with multiple HHI measures.

Census HHI vs Compustat HHI | *Census HHI vs Latest Census HHI* | *Census HHI vs CGA-HHI* | *Frequency of Census gaps with other HHIs*



Second, we present distributions of the gaps between Census HHI and the other HHIs in correlation with firm variables. As depicted in the fourth chart of Figure 3-3, the horizontal axis exhibits the magnitudes of the gaps, and the vertical axis denotes the frequencies for the gaps to appear. Reasonably, with smaller gaps appearing more frequently, the corresponding HHI measure better represents Census HHI in correlation with the firm variables. The result shows that the correlation gaps between Census HHI vs simple HHI measures usually peak near zero, while the correlation gaps between Census HHI vs Compustat HHI are farther away from zero and exhibit flatter shapes and longer tails. This further confirms that the correlation of firm characteristics with Census HHI can be better represented by that with the simple HHI measures than with Compustat HHI.

Finally, we regress Census HHI's correlation coefficients on the other HHI proxies' correlation coefficients (with firm variables) to check their substitutability. Plausibly, with the regression coefficients closer to one, the alternative HHI proxy should better estimate Census HHI in correlation with the firm variables. The regression coefficients are plotted in Figure 3-4. It shows that the regression coefficients of Census HHI's correlation coefficients on Compustat HHI's correlation coefficients are all much lower than one. For Firm Value, Leverage, Payout, and Profitability, regressions of Census HHI's correlation coefficients on Compustat HHI's are negative, ranging from -0.02 to -0.20. For the rest, the regression coefficients are 0.00 for Investment, 0.02 for Cash Holding, and 0.31 for Innovation. These low regression coefficients indicate that Compustat HHI poorly approximates Census HHI in correlation with the firm variables. On the other side, regression coefficients of Census HHI's correlation coefficients on the simple HHI measures' correlation coefficients range from 0.64 to 1.01, indicating strong substitutability between Census HHI and the simple HHI measures in correlation with the firm variables.[44]

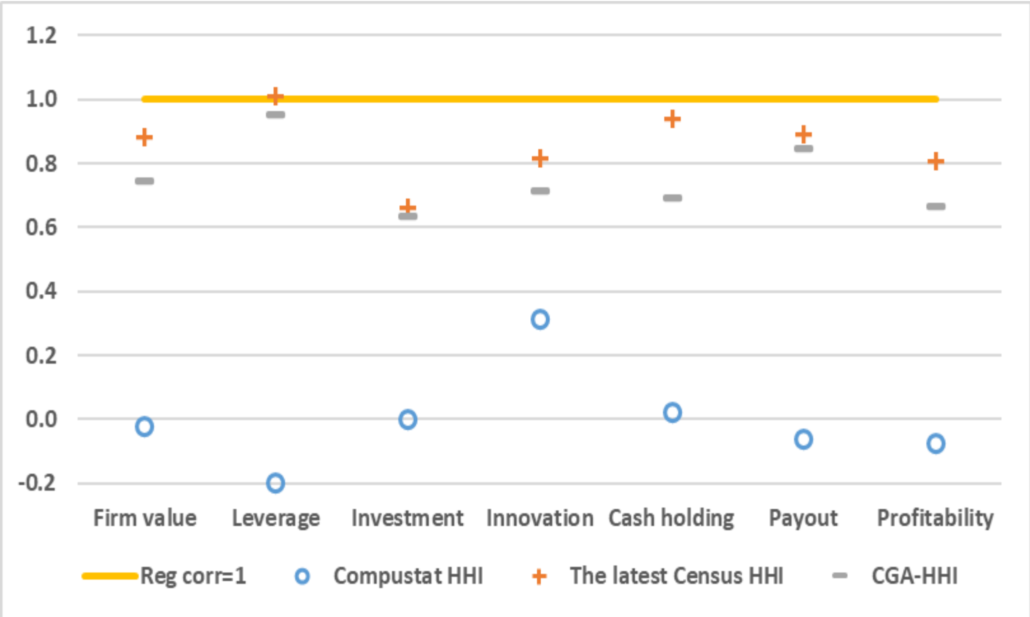
To sum, our tests provide evidence in favour of the simple HHI measures as closer proxies of Census HHI in association with key firm characteristics, compared to Compustat HHI.

---

[44] As variables in the regressions are correlation coefficients which are bounded between -1 and +1. In unreported tests, we conduct Tobit regression and set censoring limits on the outcome variable. The regression coefficients are almost the same.

Figure 3-4: Regression of Census HHI's correlation coefficients on the other HHI measures' correlation coefficients

In Figure 3-4 we regress the bootstrapped correlation coefficients of a firm variable with Census HHI for manufacturing industries on the bootstrapped correlation coefficients of this firm variable with Compustat HHI and with the simple HHI measures. The firm variables include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability, all defined in Appendix B-1. Industry classifications for the HHI proxies are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017.



3.2.5 Firm-level analysis

In addition to industry analysis, HHI is often also employed in cross-sectional firm analysis in the literature. Although we have established that the simple HHI measures perform better than Compustat HHI in the above industry analysis, it is not clear whether this is the case for firm-level analysis. As large firms matter the most in HHI computation and publicly listed companies are usually large, Compustat HHI for industries with more public companies might be more informative. It is thus possible that Compustat HHI better approximates Census HHI in firm analysis where industries containing more public companies are allocated with more weights. To verify whether the simple HHI measures are also superior to Compustat HHI in firm analysis, we allocate HHI values for each industry to all firms in this industry and correlate HHI proxies with firm variables at firm level.[45]

We find that Compustat HHI has the same sign as Census HHI in correlation with all the analysed corporate characteristics at firm level. This is an improvement from the industry-level analysis, in which correlations of Firm Value and Innovation with Compustat HHI are in opposite directions to that with Census HHI. Still, correlation coefficients of firm variables with Compustat HHI are the farthest from that with Census HHI at firm level, with differences (-0.190, 0.342) usually much greater in absolute value than that between Census HHI vs the simple HHI measures (-0.034, 0.024).[46] We also notice that the Latest Census HHI deviates more from Census HHI at firm level than at industry level. For industry analysis, the Latest Census HHI is in the same direction as Census HHI in correlation for all the discussed firm variables, with the same significance status. For firm analysis, the Latest Census HHI has opposite signs to Census HHI correlated with Innovation and Cash Holding, and has different significance levels correlated with Leverage, Payout, and Profitability. Meanwhile, CGA-HHI shares the same signs as Census HHI in correlation for all the firm variables in both firm and industry analyses, with correlation coefficients often at the same significance levels.

In summary, our empirical results generally support the simple HHI measures as more accurate estimates of Census HHI than Compustat HHI in association with key corporate characteristics.[47] Noted that the Latest Census HHI better approximates Census HHI in correlation with the firm variables at industry level than at firm level, while CGA-HHI more robustly approximates Census HHI for both industry and firm level analyses.

3.3 Application of simple HHI measures in empirical tests

3.3.1 Introduction

Hou and Robinson (2006) suggest that firms in less concentrated industries have higher stock returns because these firms are exposed to higher innovation risk and/or higher distress risk. Empirically,

---

[45] Correlation coefficients are plotted in Appendix B-5.

[46] The exception is Investment. Correlation of investment with Census HHI (0.097) is a little closer to Compustat HHI (0.082) than the CGA-HHI (0.076).

[47] In Appendix B-6, we correlate HHI proxies with firm characteristics constructed with alternative methods next to the most frequently used ones as stated by Mitton (2022), with basically unaffected conclusions.

they find negative relation between stock return and concentration measured by Compustat HHI[48]. However, Ali et al. (2009) show that the negative relation between stock return and concentration does not hold using a mixture of Census HHI in Census years and the nearest Census HHI in non-Census years.[49] In this section we re-examine Hou and Robinson's (2006) main empirical tests[50], and replace the original Compustat HHI with Census HHI and the simple HHI measures respectively to compare the results. Our analyses differ from that of Ali et al. (2009) in the following aspects. First, Ali et al. (2009) fill non-Census years with the Census HHI in the nearest Census years, so that their empirical results are driven by a combination of Census HHI and estimates of Census HHI. In our tests, we use Census HHI and the other HHI proxies separately so that we have test results only led by Census HHI and only led by the other HHI proxies respectively, which are easily comparable. Second, our main purpose differs from Ali et al.'s (2009). We not only examine whether Census HHI leads to different results from Compustat HHI, but also verify whether the simple HHI measures lead to more similar results as Census HHI does, compared to Compustat HHI. Finally, Ali et al.'s (2009) estimate of Census HHI in non-Census years involves future information of Census HHI, whereas our measures only use past information. Only using past information is important if future information of concentration causes future bias. Also, future Census HHI could be unavailable for more recent years after the latest publication of Census HHI.

3.3.2 Methods and results

In their study, Hou and Robinson (2006) conduct Fama–MacBeth (FM) regression of monthly stock return on the level of industry concentration and find significantly negative regression coefficients. For our purpose, we repeat the tests only for Census years from 1987 to 2017 (1997 excluded), using Compustat HHI, Census HHI, the Latest Census HHI, and CGA-HHI as concentration measures at

[48] Hou and Robinson (2006) mention that in unreported tests, the negative relation between stock return and concentration holds "*on the much smaller sample of observations for which Census of Manufactures definitions of industry concentration are available.*" (p. 1940) But this is contradicted to Ali et al.'s (2009) empirical results. Our empirical results are consistent to the latter.

[49] There are also other studies using Compustat HHI as the concentration measure (Falato & Liang, 2016; Lang & Stulz, 1992; Harris, 1998; Spiegel & Tookes, 2013) examined by Ali et al. (2009) and Keil (2017) who reach different results using Census concentration indices.

[50] We choose to re-examine Hou and Robinson's regressions because the required data are just from Compustat CRSP merged dataset and we can manually merge Compustat and CRSP datasets for testing.

industry level. We classify industries by 4-digit SIC before 1997 and 6-digit NAICS after 1997, similar as that in Section 3.2.[51] All the variables are standardised to make the regression coefficients comparable. We report the average value of the monthly FM regression coefficients and the time-series t-statistics in Table 3-2.

Table 3-2 shows that FM regression of stock return is significantly negative on Compustat HHI (r=-0.016, t=-2.00) while insignificantly positive on Census HHI (r=0.005, t=0.75), the Latest Census HHI (r=0.007, t=1.12), and CGA-HHI (r=0.007, t=0.94). Our test results present a significantly negative relation between stock return and Compustat HHI, which is consistent with Hou and Robinson (2006) although in different sample period. However, we find insignificantly positive relation between stock return and industry concentration measured with Census HHI, the Latest Census HHI and CGA-HHI. In other words, we show that using Compustat HHI leads to conclusions different from using Census HHI, while using the simple HHI measures come to the same conclusions as using Census HHI in the re-examination of Hou and Robinson's (2006) empirical tests.[52]

Table 3-2: Fama–MacBeth cross-sectional regressions of industry-level stock returns

This table presents Fama–MacBeth cross-sectional regression results at industry-level. Following Hou and Robinson (2006), we regress stock return on Compustat HHI, controlling for industry average value of ln(size), ln(B/M), market leverage, market beta, and the past 1-year return on the industry portfolio (Momentum). Definition of the variables are in Appendix B-1. We then repeat the regression using Census HHI, Compustat HHI, and the simple HHI measures as concentration measures respectively. The values of all the independent variables are at the end of $year_t$ ($year_t$=1987, 1992, 2002, 2007, 2012, 2017). The dependant variable is CRSP monthly return between July of $year_{t+1}$ and June of $year_{t+2}$. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The time series mean value of regression coefficients are reported with t-statistics in

[51] Our industry classifications are consistent with publication of Census HHI but different from Hou & Robinson (2006) who rely on 3-digit SIC. Converting NAICS industries to SIC inevitably involve certain problems so that we still use the original industry classifications from Census publications. In Appendix B-7, we try to replicate the regression results reported by Ali et al.'s (2009) and Hou and Robinson (2006) following their sample period, industry classifications and the other conditions. The regression coefficients on Compustat HHI are all significantly negative. Our replication results are generally similar with that reported by Ali et al. (2009).

[52] In unreported tests, we use value-weighted industry return and the conclusions are unaffected.

parenthesis below.

Dependent variable: industry return	1	2	3	4
Intercept	-0.075 (-1.23)	-0.075 (-1.24)	-0.081 (-1.30)	-0.081 (-1.31)
Compustat HHI	-0.016 (-2.00)			
Census HHI		0.005 (0.75)		
Lagged Census HHI			0.007 (1.12)	
CGA-HHI				0.007 (0.94)
Ln(Market capitalisation)	0.010 (0.68)	0.011 (0.71)	0.011 (0.70)	0.011 (0.73)
Ln(B/M)	0.009 (0.69)	0.009 (0.71)	0.010 (0.75)	0.009 (0.79)
Momentum	-0.021 (-0.64)	-0.020 (-0.61)	-0.020 (-0.63)	-0.021 (-0.54)
Beta	-0.007 (-0.56)	-0.004 (-0.34)	-0.004 (-0.30)	-0.007 (-0.51)
Leverage	-0.008 (-0.54)	-0.010 (-0.70)	-0.011 (-0.73)	-0.008 (-0.49)
Average adjusted R-square	0.054	0.053	0.053	0.054
Average monthly observations	213	213	213	213

3.4 Conclusion

In this essay, we suggest that the simple HHI measures better estimate concentration than Compustat HHI because the former approximate Census HHI, which is the nearest to “true” HHI, more closely than the latter. We show that a set of important firm characteristics vary across industries sorted by the simple HHI measures in similar manners as they vary across industries sorted by Census HHI. In addition, correlations of these firm characteristics with the simple HHI measures usually have the same signs and significance status as their correlations with Census HHI. On the other hand, Compustat HHI and Census HHI often present different relationship with firm characteristics compared to Census HHI. To check the robustness of our results, we analyse distributions of these correlation coefficients based on bootstrapped samples. We also conduct correlations at firm level in addition to the industry level analysis, and we find the superiority of the simple HHI measures to

Compustat HHI remains. Specifically, the Latest Census HHI generally performs better for industry analysis than firm analysis, while CGA-HHI is more robust in both industry and firm analyses.

To further build the confidence of using the simple HHI measures, we re-examine Hou and Robinson's (2006) empirical tests, replacing the originally used Compustat HHI with Census HHI and the simple HHI measures respectively. We show that the simple HHI measures and Census HHI lead to the same conclusions which are different from the original conclusions based on Compustat HHI. These results lend additional support to using the simple HHI measures as appropriate proxies of industry concentration in empirical analyses.

4 Chapter Four (Essay Three): Investor sentiment, market power, and stock returns

Abstract:

This essay reveals a negative relation between market power and sentiment sensitivity. We show return spreads between stocks with high and low market power are significantly higher after optimistic sentiment than after pessimistic sentiment, which is more pronounced when sentiment is more severe, and when sentiment later weakens or transits than strengthens. Our baseline regression results also present significantly positive relation of the return spreads between high and low market power firms with the preceding sentiment levels, which survives a number of robustness tests. The empirical results all indicate that firms with strong market power are less likely to be misvalued in irrational sentiment, and thus disclose a linkage between market power and capital market efficiency. Our findings are consistent with the argument that market power protects profits from idiosyncratic negative shocks and reduces performance uncertainties.

Key words:

Investor sentiment; market power; stock return; valuation

JEL codes:

G12; G41; L41

4.1 Introduction

Asset valuation is a key topic in financial studies. Traditional capital pricing theories establish that an asset's intrinsic value can be estimated by analysing fundamental factors (Damodaran, 2007, 2012; Sharpe, 1964)[53], and that the market is efficient with asset prices always moving towards their intrinsic values (Fama, 1965, 1970). However, as highlighted by the Dutch tulip bubble in the 17th century, absurd price deviation away from fundamental values can occur. In accordance, Keynes (1936) suggests that waves of optimistic or pessimistic sentiment can stimulate a general overestimation or underestimation of stock values, usually followed by a subsequent reverse in asset prices. Later studies find further evidence that fundamental factors are unable to completely explain stock valuation, whereas investor sentiment considerably drives prices away from intrinsic stock values (Baker & Wurgler, 2007; De Long et al., 1990; Kumar & Lee, 2006; Shiller, 2000, 2002, 2003).

While Keynes describes investor sentiment as a pervasive emotion that has effect on stock prices in aggregate, researchers also find that misvaluation driven by sentiment varies across stocks. The seminal works of Baker and Wurgler (2006, 2007) reveal that stocks that are harder to value are more vulnerable to sentiment.[54] They suggest several important characteristics that may affect stock sensitivity to sentiment, including firm size, firm age[55], volatility of stock return, cash dividend (dividend payers vs non-payers), sales growth, etc. In addition to these characteristics which are often analysed in sentiment research (e.g., Brown and Cliff, 2004; Glushkov, 2006), later studies also explore other factors that could influence sentiment sensitivity, such as the return beta, return sensitivity to macro indices, etc. (Antoniou et al., 2016; Shen et al., 2017) However, to our best knowledge, little attention is paid to firm market power that may exert influence on sentiment-driven

[53] Damodaran (2007, 2012) suggests that discounted cash flows valuation is best theoretically backed up, and the method of relative valuation is most frequently adopted in practice, both are determined by fundamental factors. The former proposes that the present value of assets equals the sum of all the discounted future cash flows or earnings, the latter convert prices of comparable assets into multiples of their key variables (e.g. earnings, sales, etc), and then adjust the multiples on fundamental differences between the target and the comparable assets.

[54] This is because, first, harder-to-value stocks depend more on subjective evaluation with a wider dispersion of estimated values, which "allows investment bankers (or swindlers) to further argue for the high end of valuations" during bubble periods (Baker & Wurgler, 2006, p. 1648). Second, harder-to-value stocks are also harder to arbitrage, as in practice these stocks are riskier and costlier to buy or sell short.

[55] Baker and Wurgler (2006) compute firm age as the period starting from a firm's first appearance in CRSP dataset till the nearest month.

misvaluation.

Market power has key influences on firm performance. In the real world, perfect competition in a market with numerous firms that provide the same products without any entry barriers or information asymmetry may not exist. Firms normally enjoy certain market power that helps avoid intense competition. Market power can be formed through multiple channels such as product differentiation, geographical separation, entry barriers, or the advantages in operating efficiency (Hall, 1986). Theoretically, as long as firms face inelastic residual demand curves, they enjoy certain market power to price above marginal costs (Syverson, 2019). Specifically, market power is known as a natural hedge of profitability by facilitating firms to transfer idiosyncratic negative shocks to customers (Abdoh & Varela, 2017; Gaspar & Massa, 2006; Irvine & Pontiff, 2009; Peress, 2010). Under this logic, firms enjoying strong market power should be easier to value with more ascertained future performance and are thus less sensitive to investor sentiment compared to firms in fierce competition.

However, on the other side of the coin, it is also possible that firms with strong market power could be more responsive to sentiment as they may allow for additional risk-taking activities for such firms compared to those under intense competition pressure. For instance, Kubick et al. (2015) find that firms with market power are more likely to conduct tax avoidance strategies which are inherently risky. Contrarily, firms facing fierce competition threats may choose more prudent policies such as holding more cash to mitigate high predatory risks (Hoberg et al., 2014). These contrasting strategies potentially add to valuation difficulties and thereby increasing sentiment sensitivity for strong market power firms compared to weak market power firms.

Since the “profit-insulating” and the “risk-taking” effects counteract each other, it is an empirical question whether, and how, market power affects firm sensitivity to sentiment, which motivates our study. To investigate this question, we examine cross-sectional returns conditional on sentiment stages for firms characterised by different levels of market power. If firms with high or low market power are more vulnerable to sentiment, they should be more overestimated (underestimated) in optimistic (pessimistic) sentiment, followed by lower (higher) returns as the misvaluation is corrected.

Our empirical test results suggest that firms enjoying market power generally have less return sensitivity to sentiment, consistent with the profit-insulating story. Firms with weak market power are more likely to be overestimated (underestimated) than those enjoying strong market power when investors are in high (low) sentiment. This sentiment-mitigating effect of market power is both statistically and economically significant. When sentiment is one standard deviation above the mean, longing deciles with the strongest market power and shorting those with the weakest market power yields annual returns of 6.94%-17.64% based on different market power proxies. On the contrary, when sentiment is one standard deviation below the mean, going long with the weakest market power deciles and short with the strongest market power deciles earns 6.96%-11.65% returns on average in the next year.[56,57] Our empirical results also show that the sentiment-driven return spread is notable between deciles with the median vs the lowest market power, but ambiguous between deciles with the highest vs the median market power.

Our study contributes to the following strands of literature. First, we add to the studies that explore sentiment sensitivity across firms with different characteristics. (Antoniou et al., 2016; Baker & Wurgler, 2006, 2007; Brown and Cliff, 2004; Glushkov, 2006; Mian & Sankaraguruswamy, 2012; Shen et al., 2017; etc.) To the best of our knowledge, our paper is the first to reveal that firm market power is a key factor influencing misvaluation driven by investor sentiment, which is consistent with the "profit-insulating" argument.[58]

Second, we contribute to the market power literature by revealing a linkage between market power and valuation efficiency in capital market. As average firm market power grows rapidly in recent

[56] Here we show returns before trading costs. Estimating transaction costs for short sales is notoriously difficult and beyond the scope of our study. If we assume all transaction costs are normal, say 7-86 basis points for a single-trip trade (De Groot et al., 2012), the trading costs should be rather small for our investment strategy which is based on annual financial report and annual sentiment index.

[57] We follow Baker and Wurgler (2006) to forecast long-short portfolio returns conditional on preceding sentiment. However, it should be noted that this strategy is more theoretical than practical as BW indices are constructed on time series of the component variables, and thus not real-time indices (Antoniou et al., 2016).

[58] Our analysis is different from Baker and Wurgler's (2006) finding that non-profitable firms are more sensitive to sentiment than profitable firms. We suppose Baker and Wurgler's argument is based on the intuition that current negative earnings are unsustainable and can hardly provide useful information for estimating future cash flows, which is not about market power. Their empirical methods are also different from ours in the following aspects. First, Baker and Wurgler (2006) categorise firms by individual earnings, while we sort firms by industry-adjusted excess Markup that mainly captures within-industry difference in market power. We also use sales share as market power proxy in the regressions, with qualitatively similar results. Second, our results are not, or not only caused by the non-profitable firms. In the robustness tests, we drop the non-profitable firms and our conclusions still hold.

decades, the impact of market power has raised much concern by economists and authorities (Aghion et al., 2005; De Loecker & Eeckhout, 2018; De Loecker et al., 2020; etc.). For example, De Loecker et al. (2020) claim that market power harms customer well-being, reduces the total output, discourages the business dynamics, and brings down the welfare of the economy. Meanwhile, whether and how market power influences the capital market is underexplored. In this study, we provide evidence that firms enjoying market power may have higher valuation efficiency in capital market during periods of irrational sentiment.

Our analysis builds on Baker and Wurgler's (2006) seminal work that harder-to-value firms are more sensitive to investor sentiment. In a relevant stream of studies, researchers also investigate sentiment effects on industries, but the relation between common industry factors and sentiment sensitivity is often found inconclusive or inconsistent (e.g., Kumar, 2009; Molchanov & Stangl, 2018).[59] Specifically, Molchanov and Stangl (2018) suggest sentiment sensitivity is not related to industry characteristic that may act as proxies of valuation uncertainties, such as industry average capitalisation, industry return volatility, and industry concentration, etc. Possibly, this is because firm characteristics can vary considerably even within the same industry, such as firm market power (De Loecker et al., 2018), and that emotional investors mainly focus on firms that are significantly difficult to value.[60] As industry characteristics denote the mean or median level for all firms in the industry, they may not reflect the distribution of those firms with disproportionately high sentiment sensitivity. These conjectures are out of the scope of this study and can be explored in future research.

4.2 Data and measures

Our sample consists of firms with ordinary common shares (share code 10-11) in the Centre for Research in Security Prices (CRSP) dataset and with accounting data recorded in Compustat annual files. We keep firms with non-missing values of the market power proxies and non-missing returns.

---

[59] One exception is Kaplanski and Levy (2010), who claim that unstable industries are more affected by low sentiment but they only study pessimism derived from a single disastrous event type (air crashes).

[60] This conjecture is consistent with our examination results as we show that firms with the weakest market power are significantly more vulnerable to sentiment than the other firms, while firms with median and the strongest market power present unclear differences in sentiment sensitivity.

Firms with negative sales are excluded. The sample period is from 1966 to 2019, when BW indices are available at the end of the previous year.[61]

4.2.1 Sentiment measurement

We use Baker and Wurgler's (2006) orthogonalised index (BWO) as our major sentiment measure and their raw index (BWR) as a supplement. BWR is a compound proxy based on five market indices[62], while BWO is based on the same market indices that are orthogonalised to a set of macroeconomic factors[63]. Baker and Wurgler (2006) show that BW indices coincide with anecdotal extreme sentiment events during 1962-2001, and forecast returns of harder-to-value stocks. However, while BW indices are widely applied in the literature, they are not perfect. They are specifically criticised for containing common factors in the component indices that are irrelevant to sentiment and weaken the predictive power on stock returns (Huang et al., 2015; Zhou, 2018). We therefore also use Huang et al.'s (2015) index constructed on Baker and Wurgler's orthogonalised component indicators by means of the Partial Least Squares approach (PLS index). According to Huang et al. (2015), the PLS index is designed to capture the return-predicting power from each component proxy and better estimates aggregate market return than the BW index.[64] Another critique is that BW indices, as well as PLS index, are based on market data which could have endogeneity issues. To address this concern, we employ the Consumer Confidence Index (CCI) compiled by the Conference Board[65] to check the robustness of our results. CCI is a survey-based measure which directly presents participants' beliefs of business conditions, job availability, and family income. It is widely adopted in literature as a measure of investor sentiment (Lemmon and Portniaguina, 2006; Ludvigson, 2004; Jiang et al., 2019), and dates back to the year 1967, covering most of our sample period. We

---

[61] BW index is recently updated to the mid-2022. Extending our sample period to 2021 does not change our main conclusions.

[62] BW indices available online (http://people.stern.nyu.edu/jwurgler) are based on five component factors, which are the closed-end fund discount, the number of IPOs, the average first-day returns on IPOs, the share of equity issues in total equity and debt issues, and dividend premium. In the 2006 paper, Baker and Wurgler also include NYSE share turnover as one of the component indices. They discard this index later as share turnover is later considerably affected by the increasing institutional trading.

[63] BWO is constructed with component variables orthogonalized to the growth of the industrial production index, consumer durables, nondurables, services consumption, employment, and to a dummy variable for NBER recessions.

[64] We obtain PLS index from Huang's website (https://dashanhuang.weebly.com/).

[65] We obtain CCI from Datastream.

orthogonalise CCI to the same macroeconomic factors used by Baker and Wurgler (2006) and Huang et al. (2015). As BW indices and PLS index are standardised by the authors, we also standardise the orthogonalised CCI to be comparable in the empirical tests.

4.2.2 Market power measurement

Following Gaspar and Massa (2006), we use industry-adjusted excess Markup as the measure of market power of individual firms. The Markup derives from price-cost margin, or the Lerner index. It is calculated using sales (Compustat mnemonics: SALE) minus cost of goods sold (COGS) and the selling, general, administrative expense (XSGA), then divided by sales.[66] The industry-adjusted excess Markup is computed using individual Markup minus industry Markup to reduce the influence of structural differences across industries that are unrelated to market power, and thus better captures the intra-industry differences in firms' pricing ability (Gaspar & Massa, 2006). In our tests, VWEMK (EWEMK) denotes industry-adjusted excess Markup that is constructed using firm Markup minus value-weighted (equal-weighted) industry Markup.[67] The industries are classified by 3-digit Standard Industrial Classification (SIC3) before 1997 and 4-digit North American Industry Classification System (NAICS4) since 1997.[68]

A concern of using industry-adjusted excess Markup is that the industry Markup based on Compustat-CRSP merged dataset is biased to public companies.[69] To alleviate this concern, we also sort firms by individual Markup within each industry without requiring the information of industry average

[66] When cost items are not available, we use operating income after depreciation (OIADP) scaled by sales instead, following Gaspar and Massa (2006). Alternatively, using operating income before depreciation (OIBDP) does not change our conclusions.

[67] Alternatively, researchers may calculate the industry-adjusted excess Markup using individual Markup minus sales-weighted industry Markup (such as Peress, 2010). Our conclusions are basically unchanged using this alternative excess Markup indicator.

[68] SIC has been widely used since it is developed in the 1930s. It is not revised after 1987 and then replaced by NAICS since 1997 in US Census publications. Consistently, we use SIC3 before 1997 and NAICS4 since 1997 to reflect better contemporaneous grouping of firms. To verify whether our conclusions are dependent on the chosen industry classifications, we repeat the baseline regressions using broader (SIC2&NAICS3) and finer industries (SIC4&NAICS6), or using SIC2, SIC3 and SIC4 respectively throughout the sample period. No substantial changes are made to our conclusions.

[69] Bustamante and Donangelo (2017) show correlation at 0.64 between industry Markup based on Census and Compustat data for manufacturing industries.

Markup. Another potential concern is that we use average cost as a substitute of marginal cost to calculate Markup based on a usual assumption in the literature that average cost is a meaningful proxy of marginal cost. De Loecker et al. (2020) challenge the assumption arguing that Markup computed with average cost may be problematic as there could be different output elasticities of the variable inputs for different firms. However, they also assume firms in the same industry share similar production technology and thus similar output elasticities for variable inputs. Under this supposition, using average costs should be less problematic for ranking market power within the same industry in our tests.

Table 4-1 reports summary statistics of firm Markup, industry-adjusted excess Markup (VWEMK and EWEMK), and M-dummy which equals 1 for positive Markups and 0 for non-positive Markups. To be comparable to the previous studies, we also report Markup_r and Markup_p. Markup_r equals firm Markup with positive values and equals zero for non-positive Markup values. Markup_p refers to firm Markup with positive values only.[70] All variables are winsorised at 0.5% and 99.5%, following Baker and Wurgler (2006). Variable descriptions are summarised in Appendix C-1.

Table 4-1: Summary statistics

Table 4-1 shows the summary statistics of firm Markup proxies, industry-adjusted excess Markup (VWEMK and EWEMK), and stock returns during the period from 1966 to 2019. Markup is calculated using sales minus cost of goods sold (COGS) and selling, general, administrative expense (XSGA), and then divided by sales. When the calculation of the numerator is not possible, we use the operating income after depreciation (OIADP) instead. M-dummy equals 1 if Markup is positive, and zero if otherwise. Markup_r equals Markup with negative value replaced with zero. Markup_p refers to Markup with positive values. VWEMK (EWEMK) equals firm Markup minus value-weight (equal-weighted) industry Markup. All the variables are in percentage values and winsorised at the

[70] We report Markup_r and Markup_p to be comparable with existing studies. Normally, positive Markups do not exceed one by construction, while negative Markups could be very low with small sales (without lower bounds) and exert disproportionate influences on the average Markup. To exclude influences of non-profitable firms, some researchers report zero profit for non-profitable firms (such as Baker & Wurgler, 2006), some exclude firms with negative profits (such as Kubick et al., 2015). Our baseline results are not materially affected when we replace negative Markup with zero or when we only include observations with positive Markups.

0.5% and 99.5% levels, except for the dummy variable.

Variables	Obs	Mean	SD	25%	50%	75%
Markup (%)	2,297,982	-26.81	303.89	4.76	11.24	20.23
Markup_r (%, negative as 0)	2,297,982	14.40	13.43	4.76	11.24	20.23
Markup_p (%, positive)	1,962,821	16.86	13.02	7.78	13.19	22.33
M-dummy	2,297,982	0.85	0.35	1.00	1.00	1.00
VWEMK(%)	2,297,982	-40.61	298.60	-10.71	-3.00	2.49
EWEMK(%)	2,297,982	0.00	278.23	-2.77	2.98	19.28
RET(%)	2,297,982	1.08	15.09	-6.38	0.00	7.19

Table 4-1 shows that the mean value of firm Markup equals -26.81%, which is much lower than the median Markup (11.24%). M-dummy averages 0.85, indicating 85% of the firms in our sample are profitable and 15% are not profitable. The mean (median) values of Markup_r and Markup_p are 14.40% (11.24%) and 16.86% (13.19%), higher than the average Markup (8.5%) and median Markup (9.5%) reported by Gaspar and Massa (2006) during 1962-2001.[71] This is consistent with the increasing trend of average Markup in recent decades as suggested in recent studies (such as De Loecker et al., 2018, 2020). The average VWEMK is -40.61%, equivalent to the difference between average firm Markup (-26.81%) and average value-weighted industry Markup (13.80%). The EWEMK has zero mean because the differences between firm Markup and the simple average of firm Markup in an industry add up to zero.

4.3 Main tests on the relation between market power and stock sensitivity to sentiment

4.3.1 Sorts

In this subsection, we investigate how stock returns vary across deciles ranked by market power proxies, conditional on the preceding investor sentiment. Stock returns are winsorised at 0.5% and 99.5% and the summary statistics are reported in Table 4-1.[72] We match market power proxies for

[71] In unreported tests, we also calculate Markup_r, Markup_p, and the according industry-adjusted excess Markups during Gaspar and Massa's (2006) sample period and for their industry coverage. The mean and median values of these variables are comparatively similar as that reported by Gaspar and Massa (2006).

[72] The mean and median values of stock return (not winsorised) are similar as that reported by Baker and Wurgler (2006) during their sample period.

fiscal year ending in calendar $year_{t-1}$ to stock returns from July $year_t$ through June $year_{t+1}$, because it takes time for the annual report to be publicised by companies and absorbed by investors (Fama & French, 1992). We then sort firms to deciles based on the ranking of market power proxies at the beginning of each month. Following Baker and Wurgler (2006), we analyse stock returns during a year (from January to December) corresponding to the sentiment level at the end of the previous year.[73] If optimistic (pessimistic) sentiment drives overestimation (underestimation) of stock value, stock return should be lower (higher) in the following year when the mispricing is corrected. Under this logic, if firms with weak market power are more vulnerable to sentiment than those with strong market power, return spreads between high and low market power portfolios should be higher after optimistic sentiment than after pessimistic sentiment, and vice versa.

The sorting results are depicted in Figure 4-1. The horizontal axis denotes deciles from one to ten ranked by VWEMK, EWEMK, and firm Markup within each industry. Decile 1 contains firms with the weakest market power[74] and Decile 10 contains the strongest. The vertical axis denotes average monthly returns for each decile. The solid lines stand for monthly returns across deciles in the year after optimistic sentiment and the dash lines for returns after pessimistic sentiment. The bars denote differences between returns after optimistic and pessimistic sentiment across the deciles. In Panel A of Figure 4-1, we denote sentiment to be optimistic when the BWO index is above the mean (BWO>0), and pessimistic when BWO is below the mean (BWO<0). We have 28 (26) years immediately following the optimistic (pessimistic) sentiment. In Panel B, we exhibit returns following more extreme sentiment when BWO is one standard deviation above, or below the mean. We have seven years after high sentiment (BWO>1) and nine years after low sentiment (BWO<-1).

Panel A shows that the average returns are lower after optimistic sentiment (-0.70% to 0.90%) than after pessimistic sentiment (1.44% to 1.87%) for all market power portfolios, mostly with significant return differences. This supports the argument that optimistic (pessimistic) sentiment leads to overestimated (underestimated) valuation and forecast lower (higher) returns. Cross-sectionally,

---

[73] There are other studies analysing stock return in a year following the previous yearend sentiment, such as Antoniou et al. (2016), Baker et al. (2012), Yu and Yuan (2011), etc. Alternatively, matching returns to BW index in the previous month leads to qualitatively similar results (not tabulated).

[74] Not all firms in Decile 1 have negative Markups. In our sample, Decile 1 sorted by VWEMK (EWEMK, within-industry Markup) contains 17% (29%, 51%) profitable firms.

differences between returns after optimistic and pessimistic sentiment are the most negative for Decile 1 with the least market power (-2.58% to -1.73%) and less negative for Decile 10 with the strongest market power (-1.49% to -0.80%). This implies that firms with weaker market power have returns more dependent on sentiment. Also, return spreads between Decile 10 and 1 ("10-1" spread) are all significantly positive (0.78% to 1.19%) after optimistic sentiment, but insignificant (-0.22% to 0.14%) after pessimistic sentiment. The differences between the "10-1" spreads after optimistic and pessimistic sentiment range between 0.78% and 1.41%, which are significantly different from zero at the level of 1%. It suggests that firms with weak market power are significantly more overvalued than those with strong market power in optimistic sentiment but not in pessimistic sentiment.[75]

Panel B shows more negative differences (-3.55% to -0.95%) between returns after high and low sentiment that is more extreme (BWO>1 vs BWO<-1). Similar as Panel A, deciles with the lowest market power measured by multiple proxies, present the most negative differences (-3.55%, -2.52%) between returns after high and low sentiment. Return spreads between the highest and the lowest market power portfolios ("10-1" spread) are positive (0.58% to 1.28%) after high sentiment and negative (-0.77% to -0.58%) after low sentiment, all significantly different from zero at the 10% level or above. The differences between the "10-1" spread after high and low sentiment range from 1.16% to 2.06%, significant at the level of 5% or above. Besides, in both Panels A and B, the differences between portfolio returns after optimistic and pessimistic sentiment increase rapidly for the lower half of deciles, whereas the changes from the middle to higher deciles are less clear.

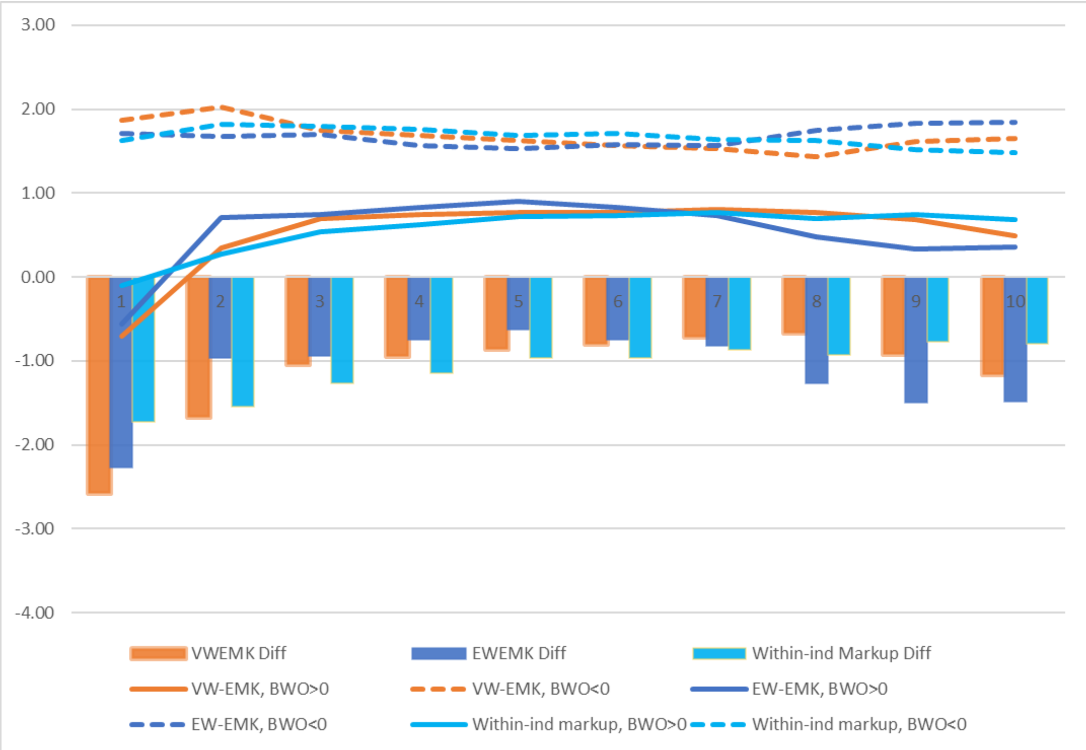
Figure 4-1: Returns of market power deciles following different investor sentiment

Figure 4-1 displays returns of deciles sorted by market power proxies after optimistic and pessimistic sentiment. The horizontal axis denotes Deciles from one to ten ranked by VWEMK (red line), EWEMK (blue line), and firm Markup within the industry (light blue line). Decile 1 contains firms with the weakest market power and Decile 10 contains the strongest. The vertical axis denotes average monthly returns for each decile. The solid lines stand for returns across deciles during the years after optimistic sentiment and the dash lines for those after pessimistic sentiment. The bars denote

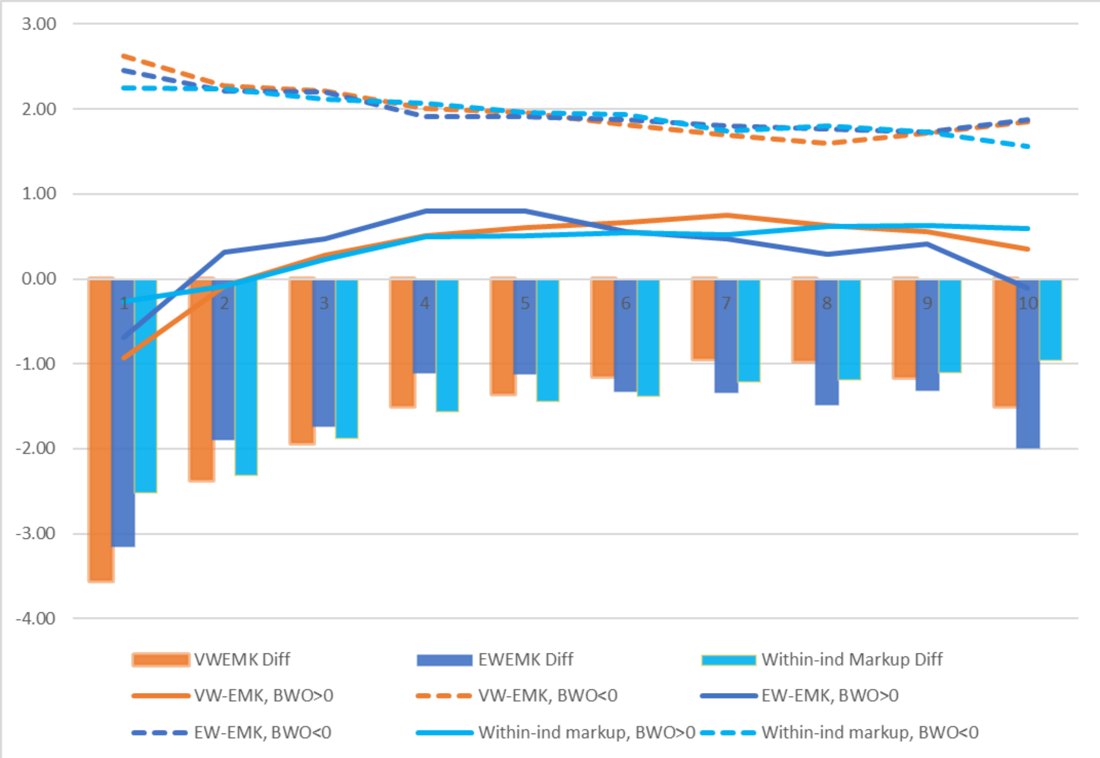
[75] This result is also consistent with previous studies that suggest sentiment-driven misvaluation is more pronounced in optimistic sentiment than in pessimistic sentiment due to short-sale impediments. (Sheng et al., 2017; Stambaugh et al, 2012).

differences between returns after optimistic and pessimistic sentiment. In Panel A, we denote sentiment to be optimistic when the BWO index is above the mean (BWO>0), and pessimistic when BWO is below the mean (BWO<0). There are 28 (26) years immediately following the optimistic (pessimistic) sentiment. Panel B depicts returns following more extreme sentiment when BWO is one standard deviation above or below the mean. There are seven years after high sentiment (BWO>1) and nine years after low sentiment (BWO<-1). Stock returns are in percentage values.

Panel A: Returns after optimistic and pessimistic sentiment (BWO>0 vs BWO<0)



Panel B: Returns after extremely optimistic and pessimistic sentiment (BWO>1 vs BWO<-1)



To sum, Figure 4-1 shows that firms tend to be overestimated (underestimated) in optimistic (pessimistic) sentiment, followed by lower (higher) returns. The influence of sentiment on future stock return is most evident for firms with the weakest market power. Return spreads between high and low market power portfolios after optimistic sentiment are significantly higher than that after pessimistic sentiment. In addition, the relation between preceding sentiment and subsequent return is more pronounced when sentiment is more extreme.[76,77]

4.3.2 Long-short portfolio regression

4.3.2.1 Approaches

Next, we investigate how the return spreads between portfolios with high and low market power vary with the continual level of investor sentiment. Following Baker and Wurgler (2006), we conduct

[76] In unreported tests, we repeat sorting tests during a more comparable sample period as that in Baker and Wurgler's (2006) paper. We find that the differences between the high-minus-low return spreads after optimistic and pessimistic periods are in comparable magnitudes when firms are sorted by market power proxies and by other firm characteristics analysed by Baker and Wurgler (2006).

[77] Replacing BWO with BWR, we get similar results. All the relevant statistics are summarised in Appendix C-2.

regression as below:

$$R_{MPOWER=high, t} - R_{MPOWER=low, t} = a + b*SENT_{t-1} + \varepsilon_t \qquad (4\text{-}1)$$

The left side of the equation denotes the return gained by longing portfolios with the highest market power and shorting those with the lowest. We sort firms into groups at the beginning of each month by the level of VWEMK, EWEMK, and firm Markup within each industry. Group 1 (Group 3) contains 30% lowest (highest) market power firms in each month and Group 2 contains the middle 40%. The dependent variable is the return spread between Groups 3 and 1. SENT denotes sentiment level measured by BWO or BWR at the end of the previous year, *b* stands for the regression coefficient on sentiment, *a* for the intercept, and ε for the disturbance item. If market power mitigates (exacerbates) sentiment-driven misvaluation, then the return spreads between high and low market power portfolios should increase (decrease) with the preceding sentiment level, returning a positive (negative) *b*.

In equation (4-2), we further control factors commonly considered to be associated with stock returns.

$$R_{MPOWER=high, t} - R_{MPOWER=low, t} = a + b*SENT_{t-1} + \beta*V + \varepsilon_t \qquad (4\text{-}2)$$

Following Baker and Wurgler (2006), we apply Fama and French (1993) three factors and the momentum factor proposed by Jegadeesh and Titman (1993) as the control variables. The Fama and French three factors include the market excess return (MKT), the return of small minus big (SMB), and the return of high minus low book-to-market ratio (HML). The momentum factor (MOM) refers to return spread between portfolios with high and low momentum. In more recent studies, Fama and French (2015, 2016) discuss the return of robust minus weak profitability (RMW) and the return of conservative minus aggressive investment (CMA) as additional factors affecting stock return. We include all five Fama and French factors plus MOM as an alternative set of control variables. The above-mentioned asset pricing factors are all obtained from French's website. In unreported tests, we also control for Q factors[78] suggested by Hou et al. (2015), and our conclusions remain unchanged.

[78] The Q factors include four factors: the market excess return (MKT), the return difference between small and big (r_{ME}), low and high investment ($r_{I/A}$), and high and low profitability(r_{ROE}). The Q factors and the relevant Fama and French's

4.3.2.2 Results

Table 4-2, Panel A reports the regression coefficients and the corresponding p-values from equations (4-1) and (4-2). Employing BWO as the predictor, all nine regressions show positive relation between the long-short return and the preceding sentiment. Regression coefficients range between 0.18 and 0.44 for the univariate regressions (Column 1), and 0.19-0.26 when the Fama and French three or five factors and MOM are controlled for (Column 2 and 3). All regression coefficients on BWO are significantly different from zero at the 5% level, or above (with eight at the 1% level). The result indicates that when sentiment increases by one standard deviation at the yearend, the annual raw (abnormal) returns earned by going long on the highest market power portfolios and shorting the lowest are 2.16%-5.28% (2.28%-3.12%) higher in the next year.[79] The results are similar using BWR instead of BWO, with regression coefficients between 0.12-0.40 for the univariate regressions (Column 4) and 0.16-0.21 after controlling for the asset pricing factors (Column 5 and 6). The regressions with BWR as sentiment measure are typically significant at the 5% level, or above. These regression coefficients are comparable to those reported in Baker and Wurgler's (2006) regressions with firms categorised by other characteristics.

The sorting tests in Section 4.3.1 reveal that while market power generally mitigates the sentiment sensitivity, the effect is more evident when we compare portfolios with median versus low than high versus median market power. In this subsection, we further investigate this return pattern. In Panel B, we break the high-minus-low market power returns into high-minus-median and median-minus-low returns and repeat the regressions. The results show that regression coefficients of the median-minus-low return on sentiment are all significantly positive and in comparable magnitude as that of the high-minus-low return. On the other hand, while the regression coefficients for the high-minus-median returns on sentiment are often positive, they are often not significantly different from zero.

---

factors are closely correlated as we examine. Correlation coefficients are 1 between the two MKTs, 0.97 between r_{ME} and SMB, 0.91 between $r_{I/A}$ and CMA, and 0.65 between r_{ROE} and RMW.

[79] As the univariate (multivariate) regression coefficients range from 0.18 to 0.44 (0.19 to 0.26), the monthly long-short returns in current year increase 0.18% to 0.44% (0.19 to 0.26%) for one standard deviation increase of investor sentiment at the end of previous year. Multiplied by twelve, we get yearly return ranging from 2.16% to 5.28% (2.28% to 3.12%).

Table 4-2: Long-short return regression on sentiment

Table 4-2 reports regression coefficient *b* and the p value (shown in *italics*) based on equation (4-1) in Columns 1 and 4, and equation (4-2) in Columns 2, 3, 5, and 6. The equations are as below:

$R_{MPOWER=high,\,t} - R_{MPOWER=low,\,t} = a + b*SENT_{t-1} + \varepsilon_t$ (4-1)

$R_{MPOWER=high,\,t} - R_{MPOWER=low,\,t} = a + b*SENT_{t-1} + \beta*V + \varepsilon_t$ (4-2)

We sort firms to groups by market power proxies including VWEMK, EWEMK, and firm Markup within each industry. Group 1 (3) contains 30% firm-months with the lowest (highest) market power and Group 2 contains the middle 40%. In Panel A the dependent variable is the return spread between Group 3 with the strongest market power and Group 1 with the weakest. In Panel B, we regress return spreads between Groups 3 and 2 and that between Groups 2 and 1 respectively on sentiment. The returns cover the period from 1966 to 2019. We measure sentiment using BWO and BWR at the end of the previous year, both downloaded from Wurgler's website. The control variables in equation (4-2) include Fama and French (1993) three factors (Market excess return, SMB, and HML) and the momentum factor (MOM) suggested by Jegadeesh and Titman (1993). Alternatively, we also include factors of RMV and CMA from Fama and French (2015) as additional control variables.

Panel A:

Partition variable for the portfolio/Sentiment		**BWO** 1	2	3	**BWR** 4	5	6
VWEMK	High-Minus-Low	0.44	0.26	0.21	0.40	0.21	0.17
		0.000	*0.001*	*0.002*	*0.000*	*0.007*	*0.013*
EWEMK		0.18	0.20	0.23	0.12	0.16	0.19
		0.025	*0.002*	*0.000*	*0.147*	*0.011*	*0.002*
Within-industry Markup		0.37	0.23	0.19	0.33	0.20	0.16
		0.000	*0.002*	*0.005*	*0.000*	*0.007*	*0.017*
Control for FF 3 factors and MOM		No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA		No	No	Yes	No	No	Yes

Panel B:

Partition variable for the portfolio/Sentiment		BWO 1	2	3	BWR 4	5	6
VWEMK	High-Minus-Median	0.03	0.01	0.04	0.01	0.01	0.04
		0.501	*0.672*	*0.190*	*0.729*	*0.723*	*0.267*
EWEMK		-0.13	-0.02	0.05	-0.16	-0.02	0.04
		0.153	*0.745*	*0.351*	*0.080*	*0.786*	*0.433*
Within-industry Markup		0.10	0.04	0.03	0.09	0.04	0.03
		0.009	*0.156*	*0.277*	*0.020*	*0.244*	*0.375*
VWEMK	Median-Minus-Low	0.41	0.25	0.17	0.38	0.20	0.14
		0.000	*0.002*	*0.009*	*0.000*	*0.015*	*0.038*
EWEMK		0.31	0.22	0.19	0.27	0.18	0.15
		0.000	*0.000*	*0.000*	*0.000*	*0.001*	*0.004*
Within-industry Markup		0.27	0.19	0.16	0.25	0.16	0.14
		0.000	*0.000*	*0.002*	*0.000*	*0.003*	*0.006*
Control for FF 3 factors and MOM		No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA		No	No	Yes	No	No	Yes

To sum, the regression results show that return differences between firms with high/median and low market power are positively related to the preceding investor sentiment, usually significant above the level of 5%. It implies that firms with market power are more alleviated from sentiment-driven misvaluation than firms with weak market power.

4.3.3 Concerns of persistent predictors

Sentiment is time series with high persistence, and therefore using sentiment as predictors in the regression may lead to spurious regression results (Ferson et al., 2003). Here we employ two methods to address this problem.

First, while the autoregression coefficients for monthly BW indices reach 0.99[80], they are only around 0.60 for the values at the end of each year. Similarly, when we look at BW indices values in a specific month (Jan, Feb, March…Dec) from each year, the autoregression coefficients are all around 0.60 due to the 12-month intervals. As such, we match monthly returns to BW indices in the previous month and separate the sample period to twelve subsamples, each containing a specific month from every sample year. We repeat regressions based on equation (4-1) and (4-2) for each subsample and

[80] In our framework, we match stock returns during 12 months in a year to BW indices at the end of the previous year. The autoregression coefficients of the matched BW indices through all the months could exceed 0.99.

report the mean of the regression coefficients and the corresponding p-value in Table 4-3. In this way, we reduce the persistence of the predictors, and we also control for potential seasonal factors that may play a role in the regressions. The results show that regression coefficients in Table 4-3 are all positive and typically significant at the level of 1%.

Table 4-3: Regressions in twelve subsamples, each containing a specific month from every year (Fama-Macbeth regression)

We separate the sample period to twelve subsamples, each containing a specific month from a year (January, February, March…). We conduct regressions based on equation (4-1) and (4-2) for each subsample and report the mean of the regression coefficients and the corresponding p value (shown in *italics*). The equations are as below:

$$R_{MPOWER=high,\ t} - R_{MPOWER=low,\ t} = a + b*SENT_{t-1} + \varepsilon_t \qquad (4\text{-}1)$$

$$R_{MPOWER=high,\ t} - R_{MPOWER=low,\ t} = a + b*SENT_{t-1} + \beta*V + \varepsilon_t \qquad (4\text{-}2)$$

We sort firms to groups by market power proxies including VWEMK, EWEMK, and firm Markup within each industry. Group 1 (3) contains 30% firm-months with the lowest (highest) market power and Group 2 contains the middle 40%. The dependent variable is the return spread between Group 3 with the strongest market power and Group 1 with the weakest. The returns cover the period from 1966 to 2019. We measure sentiment using BWO and BWR at the end of previous month. The control variables in equation (4-2) include Fama and French (1993) three factors (Market excess return, SMB, and HML) and the momentum factor (MOM) suggested by Jegadeesh and Titman (1993). Alternatively, we further include factors of RMV and CMA from Fama and French (2015) as additional control variables.

Partition variable for the portfolio/Sentiment		BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	High-Minus-Low	0.45	0.31	0.23	0.44	0.29	0.21
		0.001	*0.000*	*0.004*	*0.000*	*0.000*	*0.004*
EWEMK	High-Minus-Low	0.11	0.16	0.24	0.07	0.15	0.22
		0.375	*0.063*	*0.006*	*0.560*	*0.102*	*0.009*
Within-industry Markup	High-Minus-Low	0.42	0.33	0.26	0.41	0.32	0.25
		0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Control for FF 3 factors and MOM		No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA		No	No	Yes	No	No	Yes

Second, we follow Shen et al. (2017) and employ a simple Monte Calo simulation to clarify whether our results are driven by persistent predictors. We construct an artificial index with the same persistence level and the same AR(1) residual distribution as that of BWO. We use the artificial index in the place of BWO and conduct the 9 regressions reported in Table 4-2. We repeat the process 1,000 times and report the distribution of regression coefficients on the artificial index for each regression and the distribution of the mean value of the nine regression coefficients in Table 4-4.

The results show that the mean of the nine regression coefficients from Table 4-2 is higher than 99% of that from the simulations. When we look at each regression coefficient in Table 4-2 separately, four are higher than 99% of the simulated regression coefficients, two are above 97.5%, another two above 95%, and one is close to 90%. To sum, the simulation indicates that the chance of obtaining regression coefficients as high as that in Table 4-2 is very rare by just applying a highly persistent predictor. Therefore, our regression results should not be driven by the persistency of the predictors.

Another potential concern related to our baseline regression is the small-sample bias first proposed by Stambaugh (1999). Researchers often try to investigate whether and how a variable predicts future stock return, and for this purpose, they normally regress stock returns on a lagged predictive variable. However, if the disturbance in the autoregression of the predictive variable is correlated with the disturbance in the stock return regression, the OLS estimator could deviate substantially from the standard regression settings based on finite samples (Stambaugh, 1999). This bias is more severe with smaller sample sizes, so that it is also noted as the small-sample bias. In our case, if the innovation of sentiment is correlated with the innovation of the predictive regression, the regression coefficient could be biased. However, there are several well-cited studies in which the authors also regress stock returns on preceding BW index or other sentiment variables related to BW index (Baker & Wurgler, 2000; Huang et al., 2015; Kolev & Karapandza, 2017). They calculate regression coefficients that are adjusted for Stambaugh's (1999) bias and show that the bias-adjusted regression coefficients are actually very similar as the OLS regression coefficients. These studies suggest that Stambaugh's (1999) bias associated with BW index is less of a concern.

Table 4-4: Simulation of the long-short return regression on artificial sentiment index

We construct an artificial sentiment index with the same autoregression coefficient as that of the BWO at the yearend and match it to the stock returns in the following 12 months. The equation is as the following:

$$S_t = \rho * S_{t-1} + v_t$$

where S_{t-1}=0, ρ=0.648. We also let the residual v~(0.035, 0.764), which is the same distribution as that of the yearend BWO's AR(1) residuals. We then conduct the same nine regressions as in Table 4-2, using the artificial index instead of the BW indices as the sentiment measure. We repeat this process 1,000 times and for each regression we get 1,000 regression coefficients on the artificial sentiment index. We rank the regression coefficients by value and report the key values at 1% 2.5%, 5%, 10%, 50%, 90%, 95%, 97.5% and 99%. In the last column, we show the regression coefficients on BWO in Table 4-2 for comparison.

Partition variable for the portfolio		**Regression on artificial index**										**Regression coefficient on BWO**
		Mean	p1	p2.5	p5	p10	p50	p90	p95	p97.5	p99	
Univariate	VWEMK	0.00	-0.31	-0.27	-0.23	-0.18	-0.01	0.18	0.23	0.28	0.34	0.44
	EWEMK	0.00	-0.30	-0.27	-0.23	-0.17	0.00	0.18	0.24	0.28	0.31	0.18
	Within-industry Markup	0.00	-0.30	-0.25	-0.21	-0.16	0.00	0.17	0.22	0.26	0.30	0.37
Control FF 3 factors and MOM	VWEMK	0.00	-0.25	-0.21	-0.18	-0.14	0.00	0.14	0.18	0.22	0.26	0.26
	EWEMK	0.00	-0.27	-0.22	-0.18	-0.14	0.00	0.15	0.19	0.22	0.25	0.20
	Within-industry Markup	0.00	-0.23	-0.19	-0.15	-0.12	0.00	0.12	0.15	0.19	0.23	0.23
Control FF 5 factors and MOM	VWEMK	0.00	-0.20	-0.17	-0.14	-0.11	0.00	0.11	0.15	0.18	0.22	0.30
	EWEMK	0.00	-0.28	-0.23	-0.20	-0.15	0.00	0.15	0.19	0.23	0.26	0.21
	Within-industry Markup	0.00	-0.18	-0.15	-0.13	-0.10	0.00	0.11	0.13	0.15	0.19	0.19
Mean of the above 9 regression coefficients		**0.00**	**-0.21**	**-0.18**	**-0.15**	**-0.12**	**0.00**	**0.11**	**0.15**	**0.19**	**0.22**	**0.26**

4.4 Robustness tests

To verify that the relation between market power and sentiment sensitivity documented in Section 4.3 is independent of the specific choices made in the tests, we repeat the baseline regressions discussed in Section 4.3.2 using alternative methods. To specify, we regress the long-short returns sorted by market power on the previous sentiment using alternative sentiment measures, alternative market power proxies, and other different methods.

4.4.1 Alternative sentiment measures

In the baseline tests, we follow Baker and Wurgler (2006) to analyse monthly stock returns in a year corresponding to the sentiment level at the end of the previous year. Alternatively, researchers also analyse sentiment effect at a shorter horizon, using monthly BW index to predict stock returns in the following month (e.g., Chen et al., 2021; Eisdorfer et al., 2019). We reach similar conclusions when we regress return spreads between high and low market power portfolios on the BW indices in the previous month. Details are reported in Table 4-5, Panel A.

Besides, sentiment is difficult to gauge. To verify that our conclusions are not specific to BW indices, we also use PLS index and CCI index in our regressions as alternative sentiment proxies, as mentioned in Section 4.2.1. Both PLS and CCI indices are orthogonalised on the same macroeconomic factors that Baker and Wurgler (2006) use to construct BWO. As reported in Table 4-5, Panel B, using yearend PLS and CCI as sentiment indices, the regression coefficients based on portfolios sorted by VWEMK and within-industry Markup are all significantly positive (0.13, 0.35) above the level of 5%. For portfolios sorted by EWEMK, while the univariate regressions are sometimes insignificant, the multivariate regressions controlling for Fama and French five factors and Momentum are significantly positive at the level of 10%. In Panel C, we also analyse the short-term impact of monthly PLS and CCI on returns in the subsequent month. Regression coefficients are mostly positive and significant on monthly PLS index, but often insignificant, although all positive on monthly CCI index. This coincides with Brown and Cliff's (2005) empirical results, which show significant influences of survey-based sentiment on the subsequent stock returns at longer horizons (from 12 months up to 36 months), but insignificant at a shorter horizon (6 months).

Table 4-5: Long-short return regression on sentiment: alternative sentiment measures

Table 4-5 reports the regression coefficients and p values (shown in *italics*) for regressions in Table 4-2 using alternative sentiment proxies. More details for each panel are described in Section 4.4.1.

Panel A: BW indices (monthly value)

Partition variable for the portfolio/Sentiment		**BWO**			**BWR**	
	1	2	3	4	5	6
VWEMK	0.45	0.29	0.18	0.44	0.27	0.17
	0.000	*0.000*	*0.005*	*0.000*	*0.000*	*0.009*
EWEMK	0.11	0.15	0.22	0.07	0.14	0.20
	0.334	*0.052*	*0.003*	*0.518*	*0.078*	*0.006*
Within-industry Markup	0.41	0.29	0.20	0.40	0.28	0.20
	0.000	*0.001*	*0.009*	*0.000*	*0.000*	*0.008*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel B: PLS and CCL as sentiment measures (yearend value)

Partition variable for the portfolio/Sentiment		**PLS**			**CCI**	
	1	2	3	4	5	6
VWEMK	0.28	0.16	0.14	0.30	0.21	0.18
	0.003	*0.024*	*0.019*	*0.015*	*0.017*	*0.016*
EWEMK	-0.03	0.11	0.13	0.16	0.10	0.11
	0.756	*0.185*	*0.087*	*0.085*	*0.161*	*0.093*
Within-industry Markup	0.21	0.16	0.13	0.35	0.25	0.23
	0.017	*0.017*	*0.034*	*0.002*	*0.006*	*0.004*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel C: PLS and CCL as sentiment measures (monthly value)

Partition variable for the portfolio/Sentiment		**PLS**			**CCI**	
	1	2	3	4	5	6
VWEMK	0.29	0.15	0.09	0.21	0.08	0.07
	0.002	*0.036*	*0.152*	*0.064*	*0.335*	*0.339*
EWEMK	-0.04	0.12	0.16	0.06	0.05	0.06
	0.639	*0.095*	*0.024*	*0.527*	*0.434*	*0.386*
Within-industry Markup	0.29	0.22	0.17	0.23	0.11	0.11
	0.005	*0.009*	*0.029*	*0.029*	*0.152*	*0.138*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

4.4.2 Alternative market power measures

In the main tests, we follow Gaspar and Massa (2006) to calculate Markup using sales minus cost of goods sold (COGS) and the overheads (XSGA), then divided by sales. While the overheads are not all distributable to specific products and not strictly "variable", excluding overheads from gross profits avoids the critique that changes in profit margin might just reflect the changes in overheads instead of market power (De Loecker et al., 2020; Traina, 2018). However, it is also common for researchers (e.g., Bustamante & Donangelo, 2017) to include the overheads in profit margin as some of the expenses are more "fixed" than "variable", especially in a short horizon. To test the robustness of our results, we calculate Markup using sales deducted by COGS only and then scale the difference by sales. We repeat the regressions based on this Markup proxy and the according industry-adjusted excess Markups, and we get regression coefficients (0.05-0.18) lower than that in Table 4-2, but still mostly significant at, or above the 5% level. Details are reported in Table 4-6, Panel A.

Second, to ease the influence of abnormal Markup or potential errors in accounting data in a specific year, we use the average Markup over the past 3 years to construct an alternative market power proxy. When Markup values are not available in all three consecutive years, we use the available 2-year or 1-year Markup as the supplement.[81] We then repeat the baseline regressions using the alternative market power proxies and report the regression statistics in Table 4-6, Panel B. In unreported tests, we also try the 5-year average Markup. Our results are not sensitive to these alternative Markup calculations.

Third, our sample includes firms with negative Markup as it is reasonable to assume that such firms lack market power to protect profits from negative events. However, non-positive Markup is unsustainable over the long run. Also, it could be caused by factors other than disadvantages in market power. To ensure that our findings are not, or not only driven by non-positive Markups, we drop observations with non-positive Markups before constructing market power portfolios and then re-conduct the regressions. The results are reported in Table 4-6, Panel C. It shows that regression coefficients are reduced to 0.09-0.25, but mostly remain significantly different from zero above the level of 10% (10 of 18 are at the level of 5%).

---

[81] Results are similar if we drop the observations when Markup values for three consecutive years are not available.

Finally, our market power proxies are based on the Lerner index and on the presumption that firms tend to maximise current profits. However, profit maximisation may not always be the chosen strategy. Firms may sacrifice short-term profitability for market share expansion (e.g., Shepherd, 1972), especially in the short run. As established in many theories (such as the Cournot Model), market share is directly related to market power (De Loecker et al., 2020). Therefore, we also apply firm's market share as an alternative market power proxy to repeat the baseline regressions. Our conclusions are robust as reported in Table 4-6, Panel D.

Table 4-6: Long-short return regression on sentiment: alternative market power proxies

Table 4-6 reports the regression coefficients and p values (shown in *italics*) for regressions in Table 4-2 using alternative market power proxies. More details for each panel are described in Section 4.4.2.

Panel A: Profit margin computed using sales minus COGS only

Partition variable for the		**BWO**			**BWR**	
portfolio/Sentiment	1	2	3	4	5	6
VWEMK	0.17	0.10	0.10	0.15	0.08	0.08
	0.001	*0.014*	*0.019*	*0.004*	*0.043*	*0.056*
EWEMK	0.08	0.11	0.16	0.05	0.10	0.14
	0.308	*0.098*	*0.008*	*0.549*	*0.115*	*0.019*
Within-industry Markup	0.18	0.12	0.12	0.16	0.11	0.11
	0.001	*0.004*	*0.011*	*0.003*	*0.007*	*0.017*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel B: 3-year average of Markup

Partition variable for the		**BWO**			**BWR**	
portfolio/Sentiment	1	2	3	4	5	6
VWEMK	0.42	0.27	0.25	0.39	0.23	0.21
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
EWEMK	0.17	0.20	0.24	0.11	0.17	0.20
	0.101	*0.021*	*0.001*	*0.273*	*0.048*	*0.006*
Within-industry Markup	0.39	0.26	0.21	0.36	0.22	0.18
	0.000	*0.000*	*0.001*	*0.000*	*0.001*	*0.003*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel C: non-positive Markup excluded

Partition variable for the portfolio/Sentiment	BWO 1	2	3	BWR 4	5	6
VWEMK	0.25	0.13	0.10	0.24	0.11	0.09
	0.001	*0.022*	*0.067*	*0.002*	*0.053*	*0.113*
EWEMK	0.25	0.14	0.12	0.23	0.13	0.11
	0.002	*0.022*	*0.056*	*0.003*	*0.026*	*0.067*
Within-industry Markup	0.21	0.12	0.10	0.20	0.11	0.09
	0.009	*0.039*	*0.086*	*0.012*	*0.067*	*0.099*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel D: Market (sales) share as market power proxy

Partition variable for the portfolio/Sentiment	BWO 1	2	3	BWR 4	5	6
Sales share	0.44	0.31	0.27	0.41	0.27	0.24
	0.000	*0.000*	*0.001*	*0.000*	*0.001*	*0.004*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

4.4.3 Other robustness tests

In this subsection, we check the robustness of our conclusions by analysing returns in different sentiment regimes with subsequently strengthened or weakened sentiment, employing alternative industry classifications, and making a series of other different empirical choices.

Baker and Wurgler (2006) presume that the potential misvaluation driven by sentiment is subsequently corrected without taking account of the following development of sentiment. In fact, sentiment could be prolonged or strengthened in the following period when misvaluation is less likely to be corrected. In a relevant study, Cheema and Nartea (2017) show that sentiment acts as a contrarian predictor of returns only when sentiment levels subsequently transits from high to low or from low to high, especially in the short run. By the same token, if return differences across market power portfolios are related to firms' uneven sensitivity to preceding sentiment, then the return differences should be greater when sentiment later weakens (or transits) than strengthens. To verify, we examine and find that the sentiment-return relation is indeed stronger when sentiment weakens (or transits) than strengthens in the following period. This supports our argument that firms with weak (strong) market power are misvalued more (less) in irrational sentiment, followed by greater (smaller) return reversals when the misvaluation is corrected. Details are reported in Appendix C-3 for interested readers.

Besides, the choice of industry classification may influence the values and/or rankings of our market power proxies, and thus influence the test results. To ensure that our empirical results are not restricted to the adopted industry classifications, we use alternative industry classifications to check the robustness of our results. In detail, our market power proxies are based on industries classified by SIC3 before 1997 and NAICS4 since 1997 for the baseline regressions. We repeat the regressions based on broader industry classifications (SIC2 before 1997 & NAICS3 since 1997) and finer ones (SIC4 before 1997 & NAICS6 since 1997). We also employ 2-, 3-, 4-digit SIC respectively throughout the sample periods.[82] Our findings are robust to the applications of alternative industry classifications (see results in Table 4-7, Panels A-E).

Additionally, to ensure that our results do not depend on other specific choices in the empirical tests, we perform a battery of other robustness examinations using alternative methods. To specify, we exclude the utility and financial industries, as these industries are subject to very different regulatory requirements that may distort the relation between Markup and market power; we rank firms by market power proxies using NYSE breakpoints; we drop industries with only one firm, or with less than five firms; we do not winsorise the returns, or we winsorise returns at 1% and 99% instead of the 0.5% and 99.5% in the main tests; finally, we calculate value-weighted portfolio returns instead of equal-weighted portfolio returns. Through all these robustness tests, our major conclusions hold. We report these regression results in Table 4-7, Panels F-L.

Combining all, we document a negative relation between firm market power and sentiment sensitivity, which is robust to the application of different sentiment or market power measures and to a set of alternative empirical choices. It is also consistent with the intuitive conjecture that sentiment-driven misvaluation is more likely to be corrected when sentiment is later reverted than strengthened.

[82] We do not use NAICS industries throughout the sample period because the historical NAICS codes are often absent in the Compustat or CRSP dataset prior to 1997.

Table 4-7: Long-short return regression on sentiment: other empirical choices

Table 4-7 reports the regression coefficients and p values (shown in *italics*) for regressions in Table 4-2 based on other empirical choices. More details for each panel are described in Section 4.4.3.

Panel A: Alternative industry classifications: SIC2&NAICS3

Partition variable for the portfolio/Sentiment	**BWO**			**BWR**		
	1	2	3	4	5	6
VWEMK	0.42	0.26	0.21	0.38	0.20	0.16
	0.000	*0.000*	*0.000*	*0.000*	*0.005*	*0.011*
EWEMK	0.21	0.20	0.22	0.15	0.16	0.18
	0.024	*0.009*	*0.003*	*0.103*	*0.040*	*0.014*
Within-industry Markup	0.41	0.25	0.20	0.38	0.21	0.17
	0.000	*0.001*	*0.005*	*0.000*	*0.004*	*0.017*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel B: Alternative industry classifications: SIC4&NAICS6

Partition variable for the portfolio/Sentiment	**BWO**			**BWR**		
	1	2	3	4	5	6
VWEMK	0.38	0.23	0.18	0.34	0.18	0.14
	0.000	*0.001*	*0.001*	*0.000*	*0.004*	*0.011*
EWEMK	0.16	0.17	0.20	0.12	0.15	0.11
	0.083	*0.016*	*0.003*	*0.200*	*0.044*	*0.057*
Within-industry Markup	0.32	0.20	0.18	0.28	0.17	0.15
	0.000	*0.002*	*0.008*	*0.001*	*0.004*	*0.018*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel C: Alternative industry classifications: SIC2 during all sample years

Partition variable for the portfolio/Sentiment	**BWO**			**BWR**		
	1	2	3	4	5	6
VWEMK	0.44	0.28	0.23	0.39	0.22	0.18
	0.000	*0.000*	*0.000*	*0.000*	*0.001*	*0.004*
EWEMK	0.22	0.20	0.22	0.16	0.16	0.18
	0.013	*0.004*	*0.002*	*0.082*	*0.024*	*0.019*
Within-industry Markup	0.41	0.25	0.19	0.37	0.20	0.15
	0.000	*0.003*	*0.010*	*0.000*	*0.009*	*0.025*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel D: Alternative industry classifications: SIC3 during all sample years

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.43	0.27	0.22	0.40	0.22	0.18
	0.000	*0.000*	*0.000*	*0.000*	*0.002*	*0.004*
EWEMK	0.19	0.20	0.22	0.13	0.16	0.18
	0.038	*0.007*	*0.001*	*0.139*	*0.029*	*0.009*
Within-industry Markup	0.36	0.23	0.18	0.33	0.19	0.16
	0.000	*0.001*	*0.005*	*0.000*	*0.005*	*0.017*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel E: Alternative industry classifications: SIC4 during all sample years

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.36	0.24	0.21	0.33	0.21	0.18
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.001*
EWEMK	0.15	0.17	0.20	0.10	0.15	0.18
	0.087	*0.021*	*0.002*	*0.241*	*0.045*	*0.011*
Within-industry Markup	0.31	0.20	0.18	0.28	0.17	0.15
	0.000	*0.001*	*0.003*	*0.000*	*0.004*	*0.010*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel F: Utility and financial industries excluded

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.43	0.25	0.19	0.38	0.19	0.14
	0.000	*0.001*	*0.004*	*0.000*	*0.015*	*0.035*
EWEMK	0.23	0.22	0.26	0.16	0.18	0.21
	0.018	*0.003*	*0.001*	*0.092*	*0.025*	*0.006*
Within-industry Markup	0.38	0.25	0.20	0.34	0.21	0.16
	0.000	*0.001*	*0.006*	*0.000*	*0.009*	*0.031*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel G: NYSE breakpoints

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.33	0.20	0.18	0.30	0.17	0.15
	0.000	*0.000*	*0.001*	*0.000*	*0.002*	*0.005*
EWEMK	0.16	0.18	0.21	0.11	0.15	0.18
	0.065	*0.011*	*0.002*	*0.223*	*0.038*	*0.008*
Within-industry Markup	0.29	0.16	0.12	0.28	0.14	0.11
	0.000	*0.007*	*0.028*	*0.000*	*0.017*	*0.044*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel H: Excluding industries with only one firm

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.43	0.26	0.21	0.40	0.21	0.17
	0.000	*0.000*	*0.001*	*0.000*	*0.003*	*0.008*
EWEMK	0.18	0.20	0.23	0.12	0.16	0.19
	0.062	*0.011*	*0.002*	*0.211*	*0.042*	*0.012*
Within-industry Markup	0.38	0.24	0.20	0.35	0.21	0.17
	0.000	*0.001*	*0.006*	*0.000*	*0.003*	*2.542*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel I: Excluding industries with less than five firms

Partition variable for the portfolio/Sentiment	BWO 1	BWO 2	BWO 3	BWR 4	BWR 5	BWR 6
VWEMK	0.46	0.28	0.23	0.42	0.22	0.18
	0.000	*0.000*	*0.001*	*0.000*	*0.003*	*0.007*
EWEMK	0.19	0.21	0.24	0.12	0.17	0.20
	0.051	*0.014*	*0.001*	*0.224*	*0.041*	*0.011*
Within-industry Markup	0.40	0.26	0.21	0.36	0.22	0.18
	0.000	*0.000*	*0.002*	*0.000*	*0.003*	*0.009*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel J: No winsorisation

Partition variable for the portfolio/Sentiment	BWO			BWR		
	1	2	3	4	5	6
VWEMK	0.39	0.22	0.17	0.35	0.16	0.12
	0.000	*0.007*	*0.022*	*0.001*	*0.054*	*0.092*
EWEMK	0.20	0.24	0.29	0.14	0.21	0.25
	0.071	*0.012*	*0.001*	*0.203*	*0.029*	*0.004*
Within-industry Markup	0.35	0.19	0.13	0.32	0.15	0.11
	0.001	*0.020*	*0.092*	*0.002*	*0.066*	*0.194*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel K: Winsorise at 1% & 99%

Partition variable for the portfolio/Sentiment	BWO			BWR		
	1	2	3	4	5	6
VWEMK	0.43	0.26	0.22	0.39	0.22	0.18
	0.000	*0.000*	*0.000*	*0.000*	*0.001*	*0.004*
EWEMK	0.19	0.19	0.22	0.13	0.16	0.17
	0.021	*0.005*	*0.001*	*0.118*	*0.027*	*0.010*
Within-industry Markup	0.36	0.24	0.20	0.33	0.21	0.18
	0.000	*0.000*	*0.001*	*0.000*	*0.001*	*0.007*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

Panel L: Using value-weighted portfolio returns

Partition variable for the portfolio/Sentiment	BWO			BWR		
	1	2	3	4	5	6
VWEMK	0.40	0.22	0.14	0.39	0.20	0.13
	0.000	*0.001*	*0.006*	*0.000*	*0.004*	*0.011*
EWEMK	0.20	0.17	0.21	0.18	0.18	0.22
	0.050	*0.030*	*0.005*	*0.069*	*0.020*	*0.004*
Within-industry Markup	0.32	0.20	0.16	0.30	0.19	0.16
	0.000	*0.005*	*0.017*	*0.000*	*0.006*	*0.015*
Control for FF 3 factors and MOM	No	Yes	Yes	No	Yes	Yes
Control for RMW and CMA factors	No	No	Yes	No	No	Yes

4.5 Conclusion

This study provides evidence that firms with weak market power could be more vulnerable to investor sentiment than firms enjoying strong market power. The relation between market power and return sensitivity to sentiment is robust using different sentiment and market power measures, or making a number of different choices in empirical tests. Our examination results show that the significant regression coefficients obtained from our baseline regressions are not driven by the persistency of sentiment time series. We also show that return reversals after irrational sentiment are more pronounced when sentiment subsequently weakens, than when it strengthens. Our study is built on Baker and Wurgler's (2006) seminal work that harder-to-value firms are more sensitive to sentiment. Our findings are consistent with the argument that firms with market power can more easily transfer the undesired shocks to their customers and have lower performance uncertainties.

There are questions and concerns left unsolved in this essay. First, our empirical results show that the sentiment-mitigating effect is unclear between portfolios with median and the strongest market power. Possibly, this is simply because all firms at or above median level of market power can transfer most negative shocks to customers, and therefore firms with the strongest and median market power both have low performance uncertainties that attract little attention of the irrational investors. Alternatively, it is also possible that firms with the strongest market power tend to undertake more risky activities which balance the profit-insulating effect to a certain extent. These conjectures can be verified in future research.

Second, it is possible that our sentiment and market power proxies include endogenous and/or irrelevant factors that drive the results. To address this concern, we can involve exogenous shocks that stimulate investor sentiment, and shocks that change market power for a specific group of firms (treat group) but are immune to other firms (contrast group). If market power is related to sentiment sensitivity, then the relation between market power and stock return should differ systematically between periods before and after sentiment shocks. Also, sentiment sensitivity for the treat group should change in a different way compared to the contrast group after the market power shocks (difference in difference method).

5 Chapter Five: Thesis conclusion

Our study is motived by the increasing concerns on product market competition which is a critical factor in the economic system. In the thesis, we investigate competition measurement to improve the credibility of relevant conclusions, and we explore the influence of market power on stock valuation when the market is dominated by various investor sentiment. This chapter concludes all the findings, implications, concerns, and suggestions on further investigations.

5.1 Findings and implications

Essay One (Chapter Two) of this thesis analyses HHI as a concentration index that is widely used to gauge industry competition based on the industrial organisation theories. As HHI requires all market participants' market share information which is often inaccessible, researchers can only approximate the true HHI with information available. As inappropriate proxies of HHI could lead to flawed conclusions, we provide an overview of the HHI proxies adopted or suggested in existing studies for researchers to better understand each proxy's benefits and drawbacks. We show that the convenient, but problematic Compustat HHI is much more frequently adopted in literature than the better performing Census HHI and other newly developed HHI indicators, even after Ali et al. (2009) point out that Compustat HHI poorly estimates industry concentration. The preference of Compustat HHI is likely to undermine the reliability of relevant conclusions, and it calls for convenient competition measures that gauge competition more accurately.

To improve the credibility of analyses involving competition, we suggest two simple HHI measures, the Latest Census HHI and the CGA-HHI. They are highly correlated with Census HHI, which is generally considered the closest to the true HHI in the US, they are available at high frequencies, and can be extended to wide range of industries. The simple HHI measures require no additional information from data sources other than Census and Compustat datasets, and are rather easily computed. Plausibly, as the simple HHI measures contain historical Census HHI information, their measuring accuracy could be lower during periods when industry concentration substantially changes. Our empirical tests indeed show that the correlations between the simple HHI measures and Census HHI are stronger when the changes of Census HHI from the previous Census year are smaller than the median, and weaker but still at least moderate when the changes of Census HHI are above the median. Comparatively, CGA-HHI tends to be more robust than the Latest Census HHI in correlation with Census HHI, when Census HHI is less stable, especially based on broader industry classifications.

Essay Two (Chapter Three) provides further evidence that the simple HHI measures suggested in Essay One perform better than the most frequently used Compustat HHI. In our examinations, we set Census HHI as the benchmark because Census HHI contains the most complete information of all US firms. We compare the simple HHI measures and Compustat HHI with Census HHI in association with a set of commonly studied firm characteristics, and we examine empirical results in previous studies using different HHI proxies.

We show that the analysed firm characteristics often vary in a similar (different) way across industries or firms ranked by the simple HHI measures (Compustat HHI) as that ranked by Census HHI. The correlation coefficients for the simple HHI measures and for Census HHI are normally in the same sign, at similar significance levels, and with comparable magnitudes for both industry level and firm level analyses. However, this is rarely the case comparing correlations for Compustat HHI and for Census HHI. We also find that the Latest Census HHI often better approximates Census HHI at the industry level than at the firm level, while CGA-HHI more robustly estimates Census HHI in both industry and firm level analyses. Additionally, we analyse distribution of numerous correlations between each firm variable and multiple HHI proxies using bootstrapped samples and we get similar conclusions. Further, in the examination of Hou and Robinson's (2006) empirical tests, we show that the original conclusions using Compustat HHI as concentration measure is rejected by using Census HHI instead. This confirms the conclusions made by Ali et al. (2009) who examine Hou and Robinson's (2006) results using a combination of Census HHI and an approximate of Census HHI. Most importantly, we show that using the simple HHI measures also rejects Hou and Robinson's (2006) conclusions and gets to similar conclusions as using Census HHI.

Essay Three (Chapter Four) investigates the relation between market power and firm sensitivity to investor sentiment. We contribute to a stream of sentiment studies that challenge traditional asset pricing theories and highlight cross-sectionally different influences of investor sentiment on firm valuation (Antoniou et al., 2016; Baker & Wurgler, 2006, 2007; Brown and Cliff, 2004; Glushkov, 2006; Shen et al., 2017). The previous studies suggest a number of important firm characteristics as related to sentiment sensitivity, while we are the first to propose that firms with different market power may have different vulnerability to investor sentiment. In this way, our study also adds to the market power literature by exploring relation between market power and stock valuation efficiency in capital markets. Different from the existing studies on the association between firm market power and stock intrinsic values (such as De Loecker et al., 2020; Garlappi & Song, 2017; Liu et al., 2022; Sullivan, 1978), we suggest that market power is related to the valuation sensitivity to investor sentiment and predicts stock returns conditional on various sentiment stages.

Our study discloses a negative relationship between market power and stock sensitivity to investor sentiment. This is consistent with the argument that market power facilitates firms to transfer undesired shocks to customers and insulates firm profits from uncertainties (Abdoh & Varela, 2017; Gaspar & Massa, 2006; Irvine & Pontiff, 2009; Peress, 2010). We show that firms with market power earn significantly higher returns than firms lacking market power after optimistic sentiment, while this positive return spreads diminish to potentially negative after pessimistic sentiment. The high-minus-low return spreads sorted by firm market power are often significantly positively related to the preceding sentiment with or without controlling for common asset pricing factors, which gets through a set of robustness tests. We clarify that our regression results are not driven by the persistency of the sentiment series as the independent variables, and not likely to contain the small-sample bias first proposed by Stambaugh (1999). Besides, the sentiment-return relationship is more pronounced when sentiment is more extreme than mild, or when sentiment subsequently weakens than strengthens. This affirms the conjecture that we actually document return reversals during the periods when the sentiment-driven misvaluation is corrected.

5.2 Concerns and future research

There are concerns and questions left unsolved in this thesis. First, although concentration indices are most widely used to gauge competition as mentioned in Essay One and Essay Two, concentration proxies also have limitations as competition measures. For example, De Loecker et al. (2020) point out that concentration indices are sensitive to the market definition, and are problematic across industries and through long time periods with changed market definitions. They also show that firms in the same industry could be highly heterogeneous and face different competition pressures. Further, Boone (2008), Syverson (2019), and the others suggest that the relation between concentration and competition could be either negative or positive in different circumstances or based on different models. Therefore, in existing studies, researchers also turn to proxies other than industry concentration to estimate competition intensity, such as the Lerner index, Tobin's Q, relative profit, etc. (e.g., De Loecker et al., 2020; Griffith et al., 2005; Keeley, 1990; Lindenberg & Ross, 1981). These indices provide more dimensions in estimation of competition intensity and should be considered for the specific requirements of different analyses.

Second, in Essay Three, we find insignificant sentiment-relevant return differences between firms with the highest market power and the median. Possible explanations could be that firms with higher than median market power can transfer most of the negative shocks to the customers and therefore they all have relatively low performance uncertainties without attracting much attention of the emotional investors. It is also possible that the market leaders with the strongest market power tend to undertake more risky activities as they can afford the costs of failure. Therefore, the risk-taking effect offsets the profit-insulating effect to a certain extent. Such conjectures are outside the scope of this thesis but could be investigated in future research.

Third, as sentiment and market power are both intangible and elusive, there are always concerns with how accurate the employed measures can gauge sentiment and market power. Involving shocks of sentiment and market power should effectively address this concern and further verify the conclusions in Essay Three. To specify, if market power tends to curtail irrational stock valuations, we should find more responsive returns to sentiment shocks for firms short of market power than firms enjoying market power. On the other hand, when there are shocks that significantly enhance (reduce) market power for a specific group of firms, these firms should have decreased (increased) sentiment sensitivity compared to other firms after the market power shocks. We leave these analyses for future research.

References

Abdoh, H., & Varela, O. (2017). Product market competition, idiosyncratic and systematic volatility. *Journal of Corporate Finance,* 43, 500–513.

Adelman, M.A. (1969). Comment on the 'H' concentration measure as a numbers-equivalent. *Review of Economics and Statistics, 51*(1), 99–101.

Aggarwal, R. K., & Samwick, A. A. (1999). Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *Journal of Finance, 54*(6), 1999–2043.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The quarterly journal of economics, 120*(2), 701–728.

Aghion, P., Harris, C., Howitt, P., & Vickers, J. (2001). Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies, 68*(3), 467–492.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica, 60*(2), 323–351.

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research, 30*(1), 94–108.

Ali, A., Klasa, S., & Yeung, E. (2009). The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. *Review of Financial Studies, 22*(10). 3839–3871.

Antoniou, C., Doukas, J. A., & Subrahmanyam, A. (2016). Investor sentiment, beta, and the cost of equity capital. *Management Science, 62*(2), 347–367.

Arya, A., & Mittendorf, B. (2007). The interaction among disclosure, competition between firms, and analyst following. *Journal of Accounting and Economics, 43*(2–3), 321–339.

Asker, J., Farre-Mensa, J., & Ljungqvist, A. (2011). What do private firms look like? *Available at*

SSRN 1659926.

Azar, J., Schmalz, M. C., & Tecu, I. (2018). Anticompetitive effects of common ownership. *Journal of Finance, 73*(4), 1513–1565.

Baines, A., & Langfield-Smith, K. (2003). Antecedents to management accounting change: a structural equation approach. *Accounting, Organizations and Society, 28*(7–8), 675–698.

Baker, M., & Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. *Journal of Finance, 55*(5), 2219–2257.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

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

Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics, 104*(2), 272–287.

Berger, A. N., & Bouwman, C. H. (2009). Bank liquidity creation. *Review of Financial Studies, 22*(9), 3779–3837.

Berger, A. N., & Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics, 109*(1), 146–176.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., & Stein, J. C. (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial economics, 76*(2), 237–269.

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics, 50*(2–3), 296–343.

Bikker, J. A., & Haaf, K. (2002). Measures of competition and concentration in the banking industry: A review of the literature. *Economic & Financial Modelling, 9*(2), 53–98.

Birt, J. L., Bilson, C. M., Smith, T., & Whaley, R. E. (2006). Ownership, competition, and financial disclosure. *Australian Journal of Management, 31*(2), 235–263.

Bhojraj, S., Lee, C.M., & Oler, D.K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research, 41*(5), 745–774.

Boone, J. (2008). Competition: Theoretical parameterizations and empirical measures. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 587–611.

Boone, A. L., Field, L. C., Karpoff, J. M., & Raheja, C. G. (2007). The determinants of corporate board size and composition: An empirical analysis. *Journal of financial Economics, 85*(1), 66–101.

Boudoukh, J., Israel, R., & Richardson, M. (2022). Biases in long-horizon predictive regressions. *Journal of Financial Economics, 145*(3), 937–969.

Bresnahan, T. F., & Salop, S. C. (1986). Quantifying the competitive effects of production joint ventures. *International Journal of Industrial Organization, 4*(2), 155–175.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance, 11*(1), 1–27.

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *Journal of Business, 78*(2), 405–440.

Bushman, R. M., & Smith, A. J. (2003). Transparency, financial accounting information, and corporate governance. Financial accounting information, and corporate governance. *Economic Policy Review, 9*(1).

Bustamante, M. C. (2014). Strategic investment and industry risk dynamics. *Review of Financial Studies, 28*(2), 297–341.

Bustamante, M.C., & Donangelo, A. (2017). Product market competition and industry return. *Review* *of Financial Studies, 30*(12), 4216–4266.

Campello, M. (2006). Debt financing: Does it boost or hurt firm performance in product markets? *Journal of Financial Economics, 82*(1), 135–172.

Cheema, M. A., & Nartea, G. (2017). Investor sentiment dynamics, the cross-section of stock returns and the MAX effect. Available at SSRN 3036761.

Chen, Y., Han, B., & Pan, J. (2021). Sentiment trading and hedge fund returns. *Journal of Finance, 76*(4), 2001–2033.

Chevalier, J. A. (1995). Do LBO supermarkets charge more? An empirical analysis of the effects of LBOs on supermarket pricing. *Journal of Finance, 50*(4), 1095–1112.

Cremers, M., Ferreira, M. A., Matos, P., & Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics, 120*(3), 539–560.

Curry, B., & George, K. D. (1983). Industrial concentration: A survey. *Journal of Industrial Economics*, 203–255.

Damodaran, A. (2007). Valuation approaches and metrics: a survey of the theory and evidence. *Foundations and Trends® in Finance, 1*(8), 693–784.

Damodaran, A. (2012). Investment valuation: Tools and techniques for determining the value of any asset (Vol. 666). John Wiley & Sons.

De Groot, W., Huij, J., & Zhou, W. (2012). Another look at trading costs and short-term reversal profits. *Journal of Banking & Finance, 36*(2), 371–382.

De Loecker, J., & Eeckhout, J. (2018). Global market power (No. w24768). *National Bureau of Economic Research*.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics, 135*(2), 561–644.

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.

Dennis, P., Gerardi, K., & Schenone, C. (2022). Common ownership does not have anticompetitive effects in the airline industry. *Journal of Finance, 77*(5), 2765–2798.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review, 67*(3), 297–308.

Eaton, J., & Grossman, G. M. (1986). Optimal trade and industrial policy under oligopoly. *Quarterly Journal of Economics, 101*(2), 383–406.

Eisdorfer, A., Goyal, A., & Zhdanov, A. (2019). Equity misvaluation and default options. *Journal of Finance, 74*(2), 845–898.

Falato, A., & Liang, N. (2016). Do creditor rights increase employment risk? Evidence from loan covenants. *The Journal of Finance, 71*(6), 2545–2590.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business, 38*(1), 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417.

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

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

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

Fama, E. F., & French, K. R. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies, 29*(1), 69–103.

Fee, C. E., & Thomas, S. (2004). Sources of gains in horizontal mergers: Evidence from customer, supplier, and rival firms. *Journal of Financial Economics, 74*(3), 423–460.

Feenstra, R. C., Weinstein, D. E. (2017). Globalization, markups, and US welfare. *Journal of Political Economy, 125*(4), 1040–1074.

Feldman, D., Saxena, K., & Xu, J. (2020). Is the active fund management industry concentrated enough? *Journal of Financial Economics, 136*(1), 23–43.

Ferson, W. E., Sarkissian, S., & Simin, T. T. (2003). Spurious regressions in financial economics? *Journal of Finance, 58*(4), 1393–1413.

Fresard, L. (2010). Financial strength and product market behavior: The real effects of corporate cash holdings. *Journal of Finance, 65*(3), 1097–1122.

Garlappi, L., & Song, Z. (2017). Capital utilization, market power, and the pricing of investment shocks. *Journal of Financial Economics, 126*(3), 447–470.

Gaspar, J. M., & Massa, M. (2006). Idiosyncratic volatility and product market competition. *Journal of Business*, *79*(6), 3125–3152.

Gilje, E. P., Gormley, T. A., & Levit, D. (2020). Who's paying attention? Measuring common ownership and its impact on managerial incentives. *Journal of Financial Economics, 137*(1), 152–178.

Giroud, X., & Mueller, H. M. (2010). Does corporate governance matter in competitive industries? *Journal of Financial Economics, 95*(3), 312–331.

Giroud, X., & Mueller, H.M. (2011). Corporate governance, product market competition, and equity prices. *Journal of Finance, 66*(2), 563–600.

Glushkov, D. (2006). Sentiment beta. Available at SSRN 862444.

Griffith, R., Boone, J., & Harrison, R. (2005). Measuring competition. *Advanced Institute of Management Research Paper,* (022).

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

Gutiérrez, G., & Philippon, T. (2017). Declining competition and investment in the US (No. w23583). *National Bureau of Economic Research*.

Hall, R. E., Blanchard, O. J., & Hubbard, R. G. (1986). Market structure and macroeconomic fluctuations. *Brookings Papers on Economic Activity, 1986*(2), 285–338.

Hall, M., & Tideman, N. (1967). Measures of concentration. *Journal of the American Statistical Association, 62*(317), 162–168.

Hannah, L., & Kay, J. A. (1977). *Concentration in modern industry: Theory, measurement and the UK experience*. Springer.

Harris, M. S. (1998). The association between competition and managers' business segment reporting decisions. *Journal of Accounting Research, 36*(1), 111–128.

Hartman-Glaser, B., Lustig, H., & Xiaolan, M. Z. (2019). Capital share dynamics when firms insure workers. *Journal of Finance, 74*(4), 1707–1751.

Haw, I. M., Hu, B., & Lee, J. J. (2015). Product market competition and analyst forecasting activity: International evidence. *Journal of Banking & Finance, 56*, 48–60.

Hay, D.a., Morris, D.J. (1991). Industrial economics and organization: Theory and evidence. Washingto, US: Oxford University Press.

Herfindahl, O. C. (1950). Concentration in the steel industry. *Unpublished dissertation. Columbia University*.

Hirschman, A. O. (1945) 1980 expanded. *National power and the structure of foreign trade* (Vol. 105). University of California Press.

Hoberg, G., & Phillips, G. (2010a). Real and financial industry booms and busts. *Journal of Finance, 65*(1), 45–86.

Hoberg, G., & Phillips, G. (2010b). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies, 23*(10), 3773–3811.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy, 124*(5), 1423–1465.

Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product market threats, payouts, and financial flexibility. *Journal of Finance, 69*(1), 293–324.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies, 20*(4), 1113–1138.

Hou, K., & Robinson, D.T. (2006). Industry concentration and average stock returns, *Journal of Finance*, 61(4), 1927–1956.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies, 28*(3), 650–705.

Houston, J. F., Lin, C., Lin, P., & Ma, Y. (2010). Creditor rights, information sharing, and bank risk taking. *Journal of Financial Economics, 96*(3), 485–512.

Hsu, C., Li, X., Ma, Z., & Phillips, G. M. (2017). Does product market competition influence analyst coverage and analyst career success? Tuck School of Business Working Paper, (2698331).

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies, 28*(3), 791–837.

Irvine, P. J., & Pontiff, J. (2009). Idiosyncratic return volatility, cash flows, and product market competition. *Review of Financial Studies, 22*(3), 1149–1177.

Ittner, C. D., & Larcker, D. F. (2001). Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics, 32*(1–3), 349–410.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance, 48*(1), 65–91.

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics, 132*(1), 126–149.

Kahle, K. M., & Walkling, R. A. (1996). The impact of industry classifications on financial research. *Journal of financial and quantitative analysis, 31*(3), 309–335.

Kaplanski, G., & Levy, H. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics, 95*(2), 174–201.

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *The American economic review*, 1183–1200.

Keil, J. (2017). The trouble with approximating industry concentration from Compustat. *Journal of Corporate Finance, 45*, 467–479.

Kennedy, P., O'Brien, D. P., Song, M., & Waehrer, K. (2017). The competitive effects of common ownership: Economic foundations and empirical evidence. *Available at SSRN 3008331*.

Keynes, J. M. (1936). The General theory of employment, interest and money (London, 1936).

Khan, L. M. (2017). Amazon’s antitrust paradox.

Khanna, N., & Tice, S. (2001). The bright side of internal capital markets. *Journal of Finance, 56*(4), 1489–1528.

Khanna, N., & Tice, S. (2005). Pricing, exit, and location decisions of firms: Evidence on the role of debt and operating efficiency. *Journal of Financial Economics, 75*(2), 397–427.

Kim, J. B., Li, Y., & Zhang, L. (2011). CFOs versus CEOs: Equity incentives and crashes. *Journal of Financial Economics, 101*(3), 713–730.

Klasa, S., Maxwell, W. F., & Ortiz-Molina, H. (2009). The strategic use of corporate cash holdings in collective bargaining with labor unions. *Journal of Financial Economics, 92*(3), 421–442.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies, 62*(4), 515–539.

Koch, A., Panayides, M., & Thomas, S. (2021). Common ownership and competition in product markets. *Journal of Financial Economics, 139*(1), 109–137.

Kolev, G. I., & Karapandza, R. (2017). Out-of-sample equity premium predictability and sample split–invariant inference. *Journal of Banking & Finance, 84,* 188–201.

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics, 9*(4), 469–479.

Kubick, T. R., Lynch, D. P., Mayberry, M. A., & Omer, T. C. (2015). Product market power and tax avoidance: Market leaders, mimicking strategies, and stock returns. *The Accounting Review, 90*(2), 675–702.

Kumar, A. (2009). Dynamic style preferences of individual investors and stock returns. *Journal of Financial and Quantitative Analysis, 44*(3), 607–640.

Kumar, A., & Lee, C. M. (2006). Retail investor sentiment and return comovements. *Journal of Finance*, *61*(5), 2451–2486.

Lang, L.H.P., & Stulz, R.M. (1992). Contagion and competitive intra-industry effects of bankruptcy announcements. *Journal of Financial Economics, 32*(1), 45–60.

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, *19*(4), 1499–1529.

Lewellen, K., & Lowry, M. (2021). Does common ownership really increase firm coordination? *Journal of Financial Economics, 141*(1), 322–344.

Li, X. (2010). The impacts of product market competition on the quantity and quality of voluntary disclosures. *Review of Accounting Studies, 15*(3), 663–711.

Lindenberg, E. B., & Ross, S. A. (1981). Tobin's q ratio and industrial organization. *Journal of*

Business, 1–32.

Liu, E., Mian, A., & Sufi, A. (2022). Low interest rates, market power, and productivity growth. *Econometrica, 90*(1), 193–221.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic perspectives*, *18*(2), 29–50.

Lyons, B. (1981) Price-cost margins, market structure and international trade, in: D. Currie, D. Peel & W. Peters, (Eds) *Microeconomic Analysis* (London, Croom Helm).

MacKay, P., & Phillips, G. M. (2005). How does industry affect firm financial structure? *Review of Financial Studies, 18*(4), 1433–1466.

Marfels, C. (1971). A guide to the literature on the measurement of industrial concentration in the post-war period. *Zeitschrift für Nationalökonomie, 31*(3), 483–506.

Masulis, R. W., Wang, C., & Xie, F. (2007). Corporate governance and acquirer returns. *Journal of Finance, 62*(4), 1851–1889.

Mian, G. M., & Sankaraguruswamy, S. (2012). Investor sentiment and stock market response to earnings news. *The Accounting Review, 87*(4), 1357–1384.

Mitton, T. (2022). Methodological variation in empirical corporate finance. *The Review of Financial Studies, 35*(2), 527–575.

Molchanov, A., & Stangl, J. (2018). Investor sentiment and industry returns. *International Journal of Finance & Economics, 23*(4), 546–570.

Moreno, D., Rodriguez, R., & Zambrana, R. (2018). Management sub-advising in the mutual fund industry. *Journal of Financial Economics, 127*(3), 567–587.

O'Brien, D. P., & Waehrer, K. (2017). The competitive effects of common ownership: We know less than we think. *Antitrust Law Journal*, *81(3)*, 729-776.

Peress, J. (2010). Product market competition, insider trading, and stock market efficiency. *Journal of Finance, 65*(1), 1–43.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy, 98*(5, Part 2), S71–S102.

Salop, S. C. (1979). A model of the natural rate of unemployment. *American Economic Review, 69*(1), 117–125.

Salop, S.C., & O'Brien, D.P. (2000). Competitive effects of partial ownership: Financial interest and corporate control. *Antitrust Law Journal, 67*(3), 559–614 (2000).

Sapienza, P. (2004). The effects of government ownership on bank lending. *Journal of Financial Economics, 72*(2), 357–384.

Schmalensee, R. (1977). Using the H-index of concentration with published data. *Review of Economics and Statistics*, 186–193.

Schmalz, M. C. (2018). Common ownership, concentration and corporate conduct. *Annual Review of Financial Economics, 10*, 413–448.

Schmidt, K. M. (1997). Managerial incentives and product market competition. *Review of Economic Studies, 64*(2), 191–213.

Scott, A. J. (1999). Regions and the world economy: The coming shape of global production, competition, and political order. OUP Catalogue.

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

Shen, J., Yu, J., & Zhao, S. (2017). Investor sentiment and economic forces. *Journal of Monetary Economics, 86,* 1–21.

Shepherd, W. G. (1972). The elements of market structure. *Review of Economics and Statistics*, 25–37.

Shiller, R. J. (2000). Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets, 1*(1), 49–60.

Shiller, R. J. (2002). Bubbles, human judgment, and expert opinion. *Financial Analysts Journal, 58*(3), 18–26.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives, 17*(1), 83–104.

Slovin, M. B., Sushka, M. E., & Hudson, C. D. (1991). Deregulation, contestability, and airline acquisitions. *Journal of Financial Economics, 30*(2), 231–251.

Spiegel, M., & Tookes, H. (2013). Dynamic competition, valuation, and merger activity. *The Journal of Finance, 68*(1), 125–172.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of financial economics, 54*(3), 375–421.

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

Sullivan, T. G. (1978). The cost of capital and the market power of firms. *Review of Economics and Statistics*, 209–217.

Syverson, C. (2019). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives, 33*(3), 23–43.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Traina, J. (2018). Is aggregate market power increasing? Production trends using financial statements.

White, T. K., Reiter, J. P., & Petrin, A. (2012). *Plant-level productivity and imputation of missing data in us census manufacturing data* (No. w17816). National Bureau of Economic Research.

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics, 100*(2), 367–381.

Zhou, G. (2018). Measuring investor sentiment. *Annual Review of Financial Economics, 10*, 239–259.

Appendix A for Chapter Two (Essay One)

Appendix A-1: List of articles from Journal of Finance, Review of Financial Studies, and Journal of Financial Economics applying HHI.

No	Paper
1	Aggarwal, R. K., & Samwick, A. A. (1999). Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *Journal of Finance, 54*(6), 1999–2043.
2	Aktas, N., De Bodt, E., & Roll, R. (2013). Learning from repetitive acquisitions: Evidence from the time between deals. *Journal of Financial Economics, 108*(1), 99–117.
3	Ali, A., Klasa, S., & Yeung, E. (2009). The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. *Review of Financial Studies, 22*(10), 3839–3871.
4	Amore, M. D., Schneider, C., & Žaldokas, A. (2013). Credit supply and corporate innovation. *Journal of Financial Economics, 109*(3), 835–855.
5	Aslan, H., & Kumar, P. (2016). The product market effects of hedge fund activism. *Journal of Financial Economics, 119*(1), 226–248.
6	Atanassov, J. (2013). Do hostile takeovers stifle innovation? Evidence from antitakeover legislation and corporate patenting. *Journal of Finance, 68*(3), 1097–1131.
7	Azar, J., Schmalz, M. C., & Tecu, I. (2018). Anticompetitive effects of common ownership. *Journal of Finance, 73*(4), 1513–1565.
8	Babenko, I. (2009). Share repurchases and pay-performance sensitivity of employee compensation contracts. *Journal of Finance, 64*(1), 117–150.
9	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.
10	Bae, K. H., Kang, J. K., & Wang, J. (2011). Employee treatment and firm leverage: A test of the stakeholder theory of capital structure. *Journal of Financial Economics, 100*(1), 130–153.
11	Bao, J., & Edmans, A. (2011). Do investment banks matter for M&A returns? *Review of Financial Studies, 24*(7), 2286–2315.
12	Baranchuk, N., Kieschnick, R., & Moussawi, R. (2014). Motivating innovation in newly public firms. *Journal of Financial Economics, 111*(3), 578–588.
13	Bebchuk, L. A., Cohen, A., & Wang, C. C. (2013). Learning and the disappearing association between governance and returns. *Journal of Financial Economics, 108*(2), 323–348.
14	Bebchuk, L. A., Cremers, K. M., & Peyer, U. C. (2011). The CEO pay slice. *Journal of Financial Economics, 102*(1), 199–221.

15 Bhagwat, V., Dam, R., & Harford, J. (2016). The real effects of uncertainty on merger activity. *Review of Financial Studies, 29*(11), 3000–3034.

16 Bharath, S. T., & Hertzel, M. (2019). External governance and debt structure. *Review of Financial Studies, 32*(9), 3335–3365.

17 Bhattacharyya, S., & Nain, A. (2011). Horizontal acquisitions and buying power: A product market analysis. *Journal of Financial Economics, 99*(1), 97–115.

18 Binsbergen, J. H. V. (2016). Good-specific habit formation and the cross-section of expected returns. *Journal of Finance, 71*(4), 1699–1732.

19 Bodnaruk, A., Massa, M., & Simonov, A. (2009). Investment banks as insiders and the market for corporate control. *Review of Financial Studies, 22*(12), 4989–5026.

20 Bodnaruk, A., Massa, M., & Simonov, A. (2013). Alliances and corporate governance. *Journal of Financial Economics, 107*(3), 671–693.

21 Bonaime, A., Gulen, H., & Ion, M. (2018). Does policy uncertainty affect mergers and acquisitions? *Journal of Financial Economics, 129*(3), 531–558.

22 Boone, A. L., Field, L. C., Karpoff, J. M., & Raheja, C. G. (2007). The determinants of corporate board size and composition: An empirical analysis. *Journal of financial Economics, 85*(1), 66–101.

23 Borisov, A., Goldman, E., & Gupta, N. (2015). The corporate value of (corrupt) lobbying. *Review of Financial Studies, 29*(4), 1039–1071.

24 Boutin, X., Cestone, G., Fumagalli, C., Pica, G., & Serrano-Velarde, N. (2013). The deep-pocket effect of internal capital markets. *Journal of Financial Economics, 109*(1), 122–145.

25 Boyson, N. M., & Pichler, P. (2018). Hostile resistance to hedge fund activism. *Review of Financial Studies, 32*(2), 771–817.

26 Boyson, N. M., Gantchev, N., & Shivdasani, A. (2017). Activism mergers. *Journal of Financial Economics, 126*(1), 54–73.

27 Bris, A., Goetzmann, W. N., & Zhu, N. (2007). Efficiency and the bear: Short sales and markets around the world. *Journal of Finance, 62*(3), 1029–1079.

28 Bustamante, M. C. (2014). Strategic investment and industry risk dynamics. *Review of Financial Studies, 28*(2), 297–341.

29 Bustamante, M. C., & Donangelo, A. (2017). Product market competition and industry returns. *Review of Financial Studies, 30*(12), 4216–4266.

30 Campello, M. (2006). Debt financing: Does it boost or hurt firm performance in product markets? *Journal of Financial Economics, 82*(1), 135–172.

31 Campello, M., & Gao, J. (2017). Customer concentration and loan contract terms. *Journal of Financial Economics, 123*(1), 108–136.

32 Carvalho, D. (2015). Financing constraints and the amplification of aggregate downturns. *Review of Financial Studies, 28*(9), 2463–2501.

33 Chang, X., Chen, Y., Wang, S. Q., Zhang, K., & Zhang, W. (2019). Credit default swaps and corporate innovation. *Journal of Financial Economics, 134*(2), 474–500.

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

35 Chang, X., Fu, K., Low, A., & Zhang, W. (2015). Non-executive employee stock options and corporate innovation. *Journal of Financial Economics, 115*(1), 168–188.

36 Chari, A., & Gupta, N. (2008). Incumbents and protectionism: The political economy of foreign entry liberalization. *Journal of Financial Economics, 88*(3), 633–656.

37 Chava, S., & Purnanandam, A. (2007). Determinants of the floating-to-fixed rate debt structure of firms. *Journal of Financial Economics, 85*(3), 755–786.

38 Chemmanur, T. J., & He, J. (2011). IPO waves, product market competition, and the going public decision: Theory and evidence. *Journal of Financial Economics, 101*(2), 382–412.

39 Chemmanur, T. J., Krishnan, K., & Nandy, D. K. (2011). How does venture capital financing improve efficiency in private firms? A look beneath the surface. *Review of Financial Studies, 24*(12), 4037–4090.

40 Chemmanur, T. J., Loutskina, E., & Tian, X. (2014). Corporate venture capital, value creation, and innovation. *Review of Financial Studies, 27*(8), 2434–2473.

41 Chen, H., Cohen, L., & Lou, D. (2016). Industry window dressing. *Review of Financial Studies, 29*(12), 3354–3393.

42 Chen, S. S., & Wang, Y. (2012). Financial constraints and share repurchases. *Journal of Financial Economics, 105*(2), 311–331.

43 Chen, T., Harford, J., & Lin, C. (2015). Do analysts matter for governance? Evidence from natural experiments. *Journal of Financial Economics, 115*(2), 383–410.

44 Chod, J., Lyandres, E., & Yang, S. A. (2019). Trade credit and supplier competition. *Journal of Financial Economics, 131*(2), 484–505.

45 Chun, H., Kim, J. W., Morck, R., & Yeung, B. (2008). Creative destruction and firm-specific performance heterogeneity. *Journal of Financial Economics, 89*(1), 109–135.

46 Cohn, J. B., Mills, L. F., & Towery, E. M. (2014). The evolution of capital structure and operating performance after leveraged buyouts: Evidence from US corporate tax returns. *Journal of Financial Economics, 111*(2), 469–494.

47 Cornaggia, J., Mao, Y., Tian, X., & Wolfe, B. (2015). Does banking competition affect innovation? *Journal of Financial Economics, 115*(1), 189–209.

48 Cremers, M., & Ferrell, A. (2014). Thirty years of shareholder rights and firm value. *Journal of Finance, 69*(3), 1167–1196.

49 Custódio, C., & Metzger, D. (2013). How do CEOs matter? The effect of industry expertise on acquisition returns. *Review of Financial Studies, 26*(8), 2008–2047.

50 Custódio, C., Ferreira, M. A., & Matos, P. (2013). Generalists versus specialists: Lifetime work experience and chief executive officer pay. *Journal of Financial Economics, 108*(2), 471–492.

51 D'Acunto, F., Liu, R., Pflueger, C., & Weber, M. (2018). Flexible prices and leverage. *Journal of Financial Economics, 129*(1), 46–68.

52 Dambra, M., Field, L. C., & Gustafson, M. T. (2015). The JOBS Act and IPO volume: Evidence that disclosure costs affect the IPO decision. *Journal of Financial Economics, 116*(1), 121–143.

53 Danis, A., Rettl, D. A., & Whited, T. M. (2014). Refinancing, profitability, and capital structure. *Journal of Financial Economics, 114*(3), 424–443.

54 Dasgupta, S., & Žaldokas, A. (2018). Anticollusion enforcement: Justice for consumers and equity for firms. *Review of Financial Studies, 32*(7), 2587–2624.

55 Dasgupta, S., Li, X., & Wang, A. Y. (2017). Product market competition shocks, firm performance, and forced CEO turnover. *Review of Financial Studies, 31*(11), 4187–4231.

56 Dass, N., Kini, O., Nanda, V., Onal, B., & Wang, J. (2013). Board expertise: Do directors from related industries help bridge the information gap? *Review of Financial Studies, 27*(5), 1533–1592.

57 Deng, X., Kang, J. K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics, 110*(1), 87–109.

58 Denis, D. J., Denis, D. K., & Walker, M. D. (2015). CEO assessment and the structure of newly formed boards. *Review of Financial Studies, 28*(12), 3338–3366.

59 Dessaint, O., Foucault, T., Frésard, L., & Matray, A. (2018). Noisy stock prices and corporate investment. *Review of Financial Studies, 32*(7), 2625–2672.

60 Dimson, E., Karakaş, O., & Li, X. (2015). Active ownership. *Review of Financial Studies, 28*(12), 3225–3268.

61 Donangelo, A. (2014). Labor mobility: Implications for asset pricing. *Journal of Finance, 69*(3), 1321–1346.

62 Durnev, A., Morck, R., & Yeung, B. (2004). Value enhancing capital budgeting and firm-specific stock returns variation. *Journal of Finance, 59*(1), 2013–611.

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

64 Eckbo, B. E. (1992). Mergers and the value of antitrust deterrence. *Journal of Finance, 47*(3), 1005–1029.

65 Eckbo, B. E., Makaew, T., & Thorburn, K. S. (2018). Are stock-financed takeovers opportunistic? *Journal of Financial Economics, 128*(3), 443–465.

66 Edmans, A., Goldstein, I., & Jiang, W. (2012). The real effects of financial markets: The impact of prices on takeovers. *Journal of Finance, 67*(3), 933–971.

67 Falato, A., & Liang, N. (2016). Do creditor rights increase employment risk? Evidence from loan covenants. *Journal of Finance, 71*(6), 2545–2590.

68 Falato, A., Kadyrzhanova, D., & Lel, U. (2014). Distracted directors: Does board busyness hurt shareholder value? *Journal of Financial Economics, 113*(3), 404–426.

69 Fang, V. W., Tian, X., & Tice, S. (2014). Does stock liquidity enhance or impede firm innovation? *Journal of Finance, 69*(5), 2085–2125.

70 Fee, C. E., & Thomas, S. (2004). Sources of gains in horizontal mergers: evidence from customer, supplier, and rival firms. *Journal of Financial Economics, 74*(3), 423–460.

71 Fee, C. E., Hadlock, C. J., & Pierce, J. R. (2008). Investment, financing constraints, and internal capital markets: Evidence from the advertising expenditures of multinational firms. *Review of Financial Studies, 22*(6), 2361–2392.

72 Ferreira, D., Ferreira, M. A., & Raposo, C. C. (2011). Board structure and price informativeness. *Journal of Financial Economics, 99*(3), 523–545.

73 Fich, E. M., Cai, J., & Tran, A. L. (2011). Stock option grants to target CEOs during private merger negotiations. *Journal of Financial Economics, 101*(2), 413–430.

74 Fich, E. M., Harford, J., & Tran, A. L. (2015). Motivated monitors: The importance of institutional investors′ portfolio weights. *Journal of Financial Economics, 118*(1), 21–48.

75 Francis, B. B., Hasan, I., & Hunter, D. M. (2008). Can hedging tell the full story? Reconciling differences in United States aggregate-and industry-level exchange rate risk premium. *Journal of Financial Economics, 90*(2), 169–196.

76 Fresard, L. (2010). Financial strength and product market behavior: The real effects of corporate cash holdings. *Journal of Finance, 65*(3), 1097–1122.

77 Gao, M., & Huang, J. (2016). Capitalizing on Capitol Hill: Informed trading by hedge fund managers. *Journal of Financial Economics, 121*(3), 521–545.

78 Garlappi, L., & Yan, H. (2011). Financial distress and the cross-section of equity returns. *Journal of Finance, 66*(3), 789–822.

79 Garlappi, L., Shu, T., & Yan, H. (2006). Default risk, shareholder advantage, and stock returns. *Review of Financial Studies, 21*(6), 2743–2778.

80 Giroud, X., & Mueller, H. M. (2010). Does corporate governance matter in competitive industries? *Journal of Financial Economics, 95*(3), 312–331.

81 Giroud, X., & Mueller, H. M. (2011). Corporate governance, product market competition, and equity prices. *Journal of Finance, 66*(2), 563–600.

82 Gopalan, R., & Xie, K. (2011). Conglomerates and industry distress. *Review of Financial Studies, 24*(11), 3642–3687.

83 Gormley, T. A., & Matsa, D. A. (2016). Playing it safe? Managerial preferences, risk, and agency conflicts. *Journal of Financial Economics, 122*(3), 431–455.

84 Gorton, G., Kahl, M., & Rosen, R. J. (2009). Eat or be eaten: A theory of mergers and firm size. *Journal of Finance, 64*(3), 1291–1344.

85 Graham, J. R. (2000). How big are the tax benefits of debt? *Journal of Finance, 55*(5), 1901–1941.

86 Graham, J. R., Hanlon, M., Shevlin, T., & Shroff, N. (2017). Tax rates and corporate decision-making. *Review of Financial Studies, 30*(9), 3128–3175.

87 Grennan, J. (2019). Dividend payments as a response to peer influence. *Journal of Financial Economics, 131*(3), 549–570.

88 Gu, L. (2016). Product market competition, R&D investment, and stock returns. *Journal of Financial Economics, 119*(2), 441–455.

89 Guo, B., Pérez-Castrillo, D., & Toldrà-Simats, A. (2019). Firms' innovation strategy under the shadow of analyst coverage. *Journal of Financial Economics, 131*(2), 456–483.

90 Haddad, V., Loualiche, E., & Plosser, M. (2017). Buyout activity: The impact of aggregate discount rates. *Journal of Finance, 72*(1), 371–414.

91 Harford, J., & Uysal, V. B. (2014). Bond market access and investment. *Journal of Financial Economics, 112*(2), 147–163.

92 Harford, J., Humphery-Jenner, M., & Powell, R. (2012). The sources of value destruction in acquisitions by entrenched managers. *Journal of Financial Economics, 106*(2), 247–261.

93 Hartman-Glaser, B., Lustig, H., & Xiaolan, M. Z. (2019). Capital share dynamics when firms insure workers. *Journal of Finance, 74*(4), 1707–1751.

94 Haushalter, D., Klasa, S., & Maxwell, W. F. (2007). The influence of product market dynamics on a firm's cash holdings and hedging behavior. *Journal of Financial Economics, 84*(3), 797–825.

95 He, J. J., & Huang, J. (2017). Product market competition in a world of cross-ownership: Evidence from institutional blockholdings. *Review of Financial Studies, 30*(8), 2674–2718.

96 He, J. J., & Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics, 109*(3), 856–878.

97 Hertzel, M. G., & Officer, M. S. (2012). Industry contagion in loan spreads. *Journal of Financial Economics, 103*(3), 493–506.

98 Hertzel, M. G., Li, Z., Officer, M. S., & Rodgers, K. J. (2008). Inter-firm linkages and the wealth effects of financial distress along the supply chain. *Journal of Financial Economics, 87*(2), 374–387.

99 Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies, 23*(10), 3773–3811.

100 Hoberg, G., & Phillips, G. (2010). Real and financial industry booms and busts. *Journal of Finance, 65*(1), 45–86.

101 Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product market threats, payouts, and financial flexibility. *Journal of Finance, 69*(1), 293–324.

102 Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies, 20*(4), 1113–1138.

103 Hou, K., & Robinson, D. T. (2006). Industry concentration and average stock returns. *Journal of Finance, 61*(4), 1927–1956.

104 Hsu, H. C., Reed, A. V., & Rocholl, J. (2010). The new game in town: Competitive effects of IPOs. *Journal of Finance, 65*(2), 495–528.

105 Huang, P., Huang, H. Y., & Zhang, Y. (2019). Do firms hedge with foreign currency derivatives for employees? *Journal of Financial Economics, 133*(2), 418–440.

106 Humphery-Jenner, M., Lisic, L. L., Nanda, V., & Silveri, S. D. (2016). Executive overconfidence and compensation structure. *Journal of Financial Economics, 119*(3), 533–558.

107 Ivashina, V., & Sun, Z. (2007). Institutional investors and loan market information spillover. *Journal of Financial Economics, 100*, 284–303.

108 Jagannathan, R., Matsa, D. A., Meier, I., & Tarhan, V. (2016). Why do firms use high discount rates? *Journal of Financial Economics, 120*(3), 445–463.

109 Jain, B. A., Kini, O., & Shenoy, J. (2011). Vertical divestitures through equity carve-outs and spin-offs: A product markets perspective. *Journal of Financial Economics, 100*(3), 594–615.

110 Johnson, W. C., Karpoff, J. M., & Yi, S. (2015). The bonding hypothesis of takeover defenses: Evidence from IPO firms. *Journal of Financial Economics, 117*(2), 307–332.

111 Jorion, P., & Zhang, G. (2007). Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics, 84*(3), 860–883.

112 Jung, H. W. H., & Subramanian, A. (2017). CEO talent, CEO compensation, and product market competition. *Journal of Financial Economics, 125*(1), 48–71.

113 Kadyrzhanova, D., & Rhodes-Kropf, M. (2011). Concentrating on governance. *Journal of Finance, 66*(5), 1649–1685.

114 Kale, J. R., & Shahrur, H. (2007). Corporate capital structure and the characteristics of suppliers and customers. *Journal of Financial Economics, 83*(2), 321–365.

115 Karpoff, J. M., Schonlau, R. J., & Wehrly, E. W. (2017). Do takeover defense indices measure takeover deterrence? *Review of Financial Studies, 30*(7), 2359–2412.

116 Kerr, W. R., & Nanda, R. (2009). Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship. *Journal of Financial Economics, 94*(1), 124–149.

117 Kim, E. H., & Lu, Y. (2011). CEO ownership, external governance, and risk-taking. *Journal of Financial Economics, 102*(2), 272–292.

118 Kim, J. B., Li, Y., & Zhang, L. (2011). CFOs versus CEOs: Equity incentives and crashes. *Journal of Financial Economics, 101*(3), 713–730.

119 Kini, O., Shenoy, J., & Subramaniam, V. (2016). Impact of financial leverage on the incidence and severity of product failures: Evidence from product recalls. *Review of Financial Studies, 30*(5), 1790–1829.

120 Klasa, S., Maxwell, W. F., & Ortiz-Molina, H. (2009). The strategic use of corporate cash holdings in collective bargaining with labor unions. *Journal of Financial Economics, 92*(3), 421–442.

121 Krishnan, K., Nandy, D. K., & Puri, M. (2014). Does financing spur small business productivity? Evidence from a natural experiment. *Review of Financial Studies, 28*(6), 1768–1809.

122 Kumar, A., Page, J. K., & Spalt, O. G. (2011). Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics, 102*(3), 671–708.

123 Lang, L. H., & Stulz, R. (1992). Contagion and competitive intra-industry effects of bankruptcy announcements: An empirical analysis. *Journal of Financial Economics, 32*(1), 45–60.

124 Larkin, Y. (2013). Brand perception, cash flow stability, and financial policy. *Journal of Financial Economics, 110*(1), 232–253.

125 Lewellen, K. (2006). Financing decisions when managers are risk averse. *Journal of Financial Economics, 82*(3), 551–589.

126 Li, Z., & Wang, L. (2016). Executive compensation incentives contingent on long-term accounting performance. *Review of Financial Studies, 29*(6), 1586–1633.

127 LILIENFELD-TOAL, U. V., & Ruenzi, S. (2014). CEO ownership, stock market performance, and managerial discretion. *Journal of Finance, 69*(3), 1013–1050.

128 Lin, C., Officer, M. S., & Zou, H. (2011). Directors' and officers' liability insurance and acquisition outcomes. *Journal of Financial Economics, 102*(3), 507–525.

129 Lin, L., Mihov, A., Sanz, L., & Stoyanova, D. (2019). Property rights institutions, foreign investment, and the valuation of multinational firms. *Journal of Financial Economics, 134*(1), 214–235.

130 Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies, 31*(7), 2606–2649.

131 MacKay, P., & Phillips, G. M. (2005). How does industry affect firm financial structure? *Review of Financial Studies, 18*(4), 1433–1466.

132 Maksimovic, V., Phillips, G., & Yang, L. (2013). Private and public merger waves. *Journal of Finance, 68*(5), 2177–2217.

133 Massa, M., Rehman, Z., & Vermaelen, T. (2007). Mimicking repurchases. *Journal of Financial Economics, 84*(3), 624–666.

134 Masulis, R. W., & Zhang, E. J. (2019). How valuable are independent directors? Evidence from external distractions. *Journal of Financial Economics, 132*(3), 226–256.

135 Masulis, R. W., Wang, C., & Xie, F. (2007). Corporate governance and acquirer returns. *Journal of Finance, 62*(4), 1851–1889.

136 McNeil, C., Niehaus, G., & Powers, E. (2004). Management turnover in subsidiaries of conglomerates versus stand-alone firms. *Journal of Financial Economics, 72*(1), 63–96.

137 Morellec, E., & Zhdanov, A. (2019). Product market competition and option prices. *Review of Financial Studies, 32*(11), 4343–4386.

138 Moshirian, F., Tian, X., Zhang, B., & Zhang, W. (2019). Stock market liberalization and innovation. *Journal of Financial Economics*, Forthcoming.

139 Mukherjee, A., Singh, M., & Žaldokas, A. (2017). Do corporate taxes hinder innovation? *Journal of Financial Economics, 124*(1), 195–221.

140 Ouimet, P. P. (2012). What motivates minority acquisitions? The trade-offs between a partial equity stake and complete integration. *Review of Financial Studies, 26*(4), 1021–1047.

141 Pan, Y., Wang, T. Y., & Weisbach, M. S. (2015). Learning about CEO ability and stock return volatility. *Review of Financial Studies, 28*(6), 1623–1666.

142 Panousi, V., & Papanikolaou, D. (2012). Investment, idiosyncratic risk, and ownership. *Journal of Finance, 67*(3), 1113–1148.

143 Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics, 123*(2), 251–272.

144 Qiu, J., & Wan, C. (2015). Technology spillovers and corporate cash holdings. *Journal of Financial Economics, 115*(3), 558–573.

145 Qiu, J., & Yu, F. (2009). The market for corporate control and the cost of debt. *Journal of Financial Economics, 93*(3), 505–524.

146 Rajgopal, S., Shevlin, T., & Zamora, V. (2006). CEOs' outside employment opportunities and the lack of relative performance evaluation in compensation contracts. *Journal of Finance, 61*(4), 1813–1844.

147 Roosenboom, P., Schlingemann, F. P., & Vasconcelos, M. (2013). Does stock liquidity affect incentives to monitor? Evidence from corporate takeovers. *Review of Financial Studies, 27*(8), 2392–2433.

148 Santalo, J., & Becerra, M. (2008). Competition from specialized firms and the diversification–performance linkage. *Journal of Finance, 63*(2), 851–883.

149 Seru, A. (2014). Firm boundaries matter: Evidence from conglomerates and R&D activity. *Journal of Financial Economics, 111*(2), 381–405.

150 Shahrur, H. (2005). Industry structure and horizontal takeovers: Analysis of wealth effects on rivals, suppliers, and corporate customers. *Journal of Financial Economics, 76*(1), 61–98.

151 Smith, J. D. (2016). US political corruption and firm financial policies. *Journal of Financial Economics, 121*(2), 350–367.

152 Song, M. H., & Walkling, R. A. (2000). Abnormal returns to rivals of acquisition targets: A test of the ‘acquisition probability hypothesis’. *Journal of Financial Economics, 55*(2), 143–171.

153 Spiegel, M., & Tookes, H. (2013). Dynamic competition, valuation, and merger activity. *Journal of Finance, 68*(1), 125–172.

154 Stefanescu, I., Wang, Y., Xie, K., & Yang, J. (2018). Pay me now (and later): Pension benefit manipulation before plan freezes and executive retirement. *Journal of Financial Economics, 127*(1), 152–173.

155 Sunder, J., Sunder, S. V., & Zhang, J. (2017). Pilot CEOs and corporate innovation. *Journal of Financial Economics, 123*(1), 209–224.

156 Tahoun, A. (2014). The role of stock ownership by US members of Congress on the market for political favors. *Journal of Financial Economics, 111*(1), 86–110.

157 Tate, G., & Yang, L. (2015). Female leadership and gender equity: Evidence from plant closure. *Journal of Financial Economics, 117*(1), 77–97.

158 Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics, 99*(1), 1–10.

159 Tian, X., & Wang, T. Y. (2011). Tolerance for failure and corporate innovation. *Review of Financial Studies, 27*(1), 211–255.

160 Tookes, H. E. (2008). Information, trading, and product market interactions: Cross-sectional implications of informed trading. *Journal of Finance, 63*(1), 379–413.

161 Uysal, V. B. (2011). Deviation from the target capital structure and acquisition choices. *Journal of Financial Economics, 102*(3), 602–620.

162 Valta, P. (2012). Competition and the cost of debt. *Journal of Financial Economics, 105*(3), 661–682.

163 Xu, J. (2012). Profitability and capital structure: Evidence from import penetration. *Journal of Financial Economics, 106*(2), 427–446.

164 Yang, L. (2008). The real determinants of asset sales. *Journal of Finance, 63*(5), 2231–2262.

Appendix A-2: Proportion of Compustat to Census Data (firm number, sales value)

This table shows proportion of Compustat firm number (sales value) to Census firm number (sales value) for Census years from 1997 to 2012. The industry classification is 3-digit NAICS.

NAICS	Definition	Ratio of Firm Number (Compustat/Census)				Ratio of Sales Value (Compustat/Census)			
		1997	2002	2007	2012	1997	2002	2007	2012
311	Food manufacturing	0.79%	0.57%	0.55%	0.54%	109.41%	112.51%	119.41%	93.65%
312	Beverage and tobacco product manufacturing	3.35%	2.60%	1.67%	1.26%	200.91%	214.70%	221.46%	253.61%
313	Textile mills	0.98%	0.67%	0.39%	0.34%	30.26%	16.74%	13.02%	10.10%
314	Textile product mills	0.19%	0.13%	0.06%	0.05%	38.43%	39.38%	31.50%	31.61%
315	Apparel manufacturing	0.63%	0.54%	0.69%	0.73%	67.93%	146.84%	361.87%	891.98%
316	Leather and allied product manufacturing	2.14%	1.92%	1.75%	1.48%	225.25%	439.82%	581.19%	828.76%
321	Wood product manufacturing	0.35%	0.27%	0.22%	0.23%	56.97%	63.28%	60.77%	33.12%
322	Paper manufacturing	2.86%	1.89%	1.91%	1.75%	131.05%	124.72%	127.10%	116.11%
323	Printing and related support activities	0.14%	0.12%	0.08%	0.07%	34.19%	41.83%	42.08%	31.59%
324	Petroleum and coal products manufacturing	5.15%	4.43%	4.28%	5.88%	589.48%	624.03%	460.28%	430.22%
325	Chemical manufacturing	6.46%	6.74%	6.82%	6.89%	159.96%	188.68%	180.75%	219.91%

326	Plastics and rubber products manufacturing	0.70%	0.74%	0.54%	0.48%	25.85%	42.49%	42.78%	40.63%
327	Nonmetallic mineral product manufacturing	0.59%	0.43%	0.31%	0.35%	54.85%	75.47%	85.16%	90.97%
331	Primary metal manufacturing	3.24%	2.43%	2.00%	2.12%	114.11%	130.30%	139.75%	155.39%
332	Fabricated metal product manufacturing	0.23%	0.19%	0.14%	0.14%	32.83%	32.56%	32.92%	31.98%
333	Machinery manufacturing	1.38%	1.21%	0.98%	0.99%	113.58%	110.83%	121.71%	147.78%
334	Computer and electronic product manufacturing	7.43%	7.25%	6.65%	5.54%	243.54%	300.43%	387.92%	490.65%
335	Electrical equipment, appliance, and component manufacturing	2.38%	2.15%	2.15%	1.93%	209.56%	112.21%	183.46%	254.66%
336	Transportation equipment manufacturing	1.88%	1.71%	1.49%	1.47%	178.86%	216.76%	258.68%	263.68%
337	Furniture and related product manufacturing	0.27%	0.20%	0.14%	0.17%	48.96%	48.94%	33.33%	40.73%
339	Miscellaneous manufacturing	0.90%	0.70%	0.53%	0.58%	58.82%	61.80%	74.90%	77.48%
	Weighted Average	**1.23%**	**1.10%**	**1.00%**	**1.00%**	**152.28%**	**171.91%**	**197.68%**	**214.73%**

Appendix A-3：Descriptive statistics and correlation coefficients for Census HHI and import-adjusted HHI.

This table shows descriptive statistics of, and correlation coefficients between Census HHI and import-adjusted HHI. Pearson correlation coefficients are reported in the bottom left matrix and Spearman in the top right matrix. *, ** and *** indicate the significance levels at 10%, 5%, and 1% respectively. The sample consists of manufacturing industries with non-missing value of both HHI proxies in Census years from 1982 to 2012 (i.e., 1982, 1987, 1992, 1997, 2002, 2007, 2012). Industry classifications are 2-digit SIC (Panel A), 3-digit NAICS (Panel B), FF48 (Panel C), 6-digit GICS (Panel D) for all Census publication years, and a mixture of 4-digit SIC before 1997 and 6-digit NAICS since 1997 (Panel E).

As the original industry classifications for Census HHI published by Census Bureau are 4-digit SIC before 1997 and 3-, 4-, 5-, 6-digit NAICS on/after 1997, we follow Ali et al (2009) to construct HHIs for boarder classifications. To specify, we calculate HHI based on a broader industry by adding up Census HHIs of all the component industries weighted by the squared sales share of the component industry to the broader industry. The conversion tables between SIC and NAICS are obtained from the US Census website. The allocation of SIC industries to FF48 sectors follows the rules from French's website. The transition methods from NAICS to GICS is available on Alison Weingarden's website.

The import-adjusted HHI equals Census HHI multiplied by the ratio of import values to Census outputs excluding exports. International trade data are obtained from Feenstra's website (https://cid.econ.ucdavis.edu/usix.html) before 1992, and from Schott's website (https://faculty.som.yale.edu/peterschott/international-trade-data/) since 1992. In the robustness tests not tabulated, correlations between Census HHI and import-adjusted HHI are not materially affected whether the exports are deduct from the Census outputs or not.

Panels on industry classification	Descriptive statistics								Correlation	
	Obs	**Mean**	**Median**	**SD**	**10%**	**25%**	**75%**	**90%**	**Census HHI**	**Import-adjusted HHI**
Panel A: By 2-Digit SIC										
Census HHI	150	0.020	0.008	0.034	0.002	0.004	0.023	0.050	1	0.898***
Import-adjusted HHI	150	0.016	0.006	0.029	0.002	0.003	0.014	0.036	0.882***	1
Panel B: By 3-Digit NAICS										
Census HHI	150	0.014	0.007	0.022	0.003	0.004	0.015	0.036	1	0.920***
Import-adjusted HHI	150	0.012	0.006	0.020	0.002	0.003	0.012	0.032	0.985***	1
Panel C: By FF48										
Census HHI	170	0.041	0.017	0.062	0.004	0.007	0.041	0.133	1	0.984***
Import-adjusted HHI	170	0.036	0.013	0.059	0.004	0.006	0.033	0.124	0.992***	1
Panel D: By 6-Digit GICS										
Census HHI	178	0.031	0.016	0.040	0.004	0.007	0.045	0.064	1	0.962***
Import-adjusted HHI	178	0.023	0.014	0.029	0.003	0.005	0.032	0.055	0.893***	1
Panel E: By 4-Digit SIC (1982-1992) & 6-Digit NAICS (1997-2012)										
Census HHI	2889	0.075	0.054	0.066	0.012	0.025	0.105	0.171	1	0.920***
Import-adjusted HHI	2889	0.061	0.042	0.058	0.008	0.019	0.082	0.145	0.912***	1

Appendix A-4: CGA-HHI construction reasoning

This appendix addresses the reasons that we assume that differences between Census HHI and Compustat HHI may remain similar in relative terms over relatively short horizons.

First, since large firms have disproportionately higher weights in HHI computation, Compustat firms should be the most important components that influence both Census HHI and Compustat HHI values because these firms are generally much larger than the private firms. To compare the sales scale of Compustat firms with that of private firms, we calculate the number of Compustat companies and private firms (from the US largest private firm list published by Forbes) with sales above \$100, \$10, and \$2 billion respectively for the year 2017. The results show that, out of the 45 firms with earnings above \$100 billion, only two (4%) are privately owned. While the proportion of private firms increases to 5% (12%) out of the 619 (1840) firms with revenues above \$10 (\$2) billion, most large firms are still the public companies. This is consistent with the findings of prior studies that collect private firm information from other sources such as the Sageworks dataset (Asker et al., 2011).[83]

Second, the omissions and "errors" in Compustat data as discussed in Section 2.3.1.2 could remain similar in relative terms over time. Keil (2017) suggests that a major reason for deviations between Compustat HHI and Census HHI is that Compustat revenues include global sales while Census shipments only contain US sales. However, we suppose that US sales for Compustat companies are likely to be stable in proportion to their global sales, as it is unlikely that large mature firms will frequently make dramatic changes on their basic plans for different countries. To verify this supposition, we compare the growth rate of Compustat HHI_F base on total sales from fundamental file and Compustat HHI_S based on US sales from Segment file. We find the differences between the average growth rates of HHI_F and HHI_S based on multiple industry classifications are very small, and the correlations between the two growth rates are very high. In addition, most of the similarity tests cannot reject the assumption at 5% level that the growth rates of HHI_F and HHI_S equal each

[83] Besides the private firms, there are US subsidiaries of multinational companies missed in Compustat dataset as well. However, it is less likely that these firms could lead the US market. For example, the Hyundai motor, the top seller out of all non-Compustat international automakers in the US market (Keil, 2017), only ranks 8th among all US automakers. Its influence on US HHI could be trivial compared to the Compustat top sellers.

other.[84] Details are in the following chart.

Industry classification	Obs	Yearly growth rate (mean)		Difference [(1)-(2)]	Correlation coefficient [(1) vs (2)]	Similarity test: p value (Ho: HHI_F=HHI_S)		
		HHI_F (1)	HHI_S (2)			T-test	Sign-test	Sign-rank
2-digit SIC	2599	109.4%	109.7%	-0.3%	0.976	0.682	0.241	0.258
3-digit NAICS	3125	109.1%	109.0%	0.1%	0.940	0.884	0.053	0.025
FF48	1507	106.8%	107.4%	-0.6%	0.914	0.556	0.551	0.318
6-digit GICS	2069	107.7%	108.7%	-1.0%	0.925	0.177	0.012	0.011

Besides, other Compustat data problems highlighted in Section 2.3.1.2 may also persist through time, such as the double counting of sales volumes from the listed subsidiaries and holding companies, the misclassification of minor product lines, etc. Consequently, we propose that there could be a steady proportional gap between Census and Compustat HHIs over relatively short horizons, as follows:

$$\frac{Census\ HHI_{t+j}}{Compustat\ HHI_{t+j}} \cong \frac{Census\ HHI_{t}}{Compustat\ HHI_{t}} \qquad \text{(A-1)}$$

To test this assumption, we correlate the left part of equation (A-1) with the right part, and we find they are significantly correlated, with correlation coefficient around 0.80. It then follows that:

$$\text{CGA-HHI}_{t+j} \cong \frac{\text{Compustat HHI}_{t+j}}{\text{Compustat HHI}_{t}} \times \text{Census HHI}_{t} \qquad \text{(A-2)}$$

Its logarithmic version, which measures relative HHI, is:

$$\text{Ln CGA-HHI}_{t+j} \cong \ln \text{Compustat HHI}_{t+j} - \ln \text{Compustat HHI}_{t} + \ln \text{Census HHI}_{t} \qquad \text{(A-3)}$$

84 In unreported analysis, we limit the firm-year to merged Compustat and CRSP samples, which is usually adopted by researchers for US analysis, the result still holds. We also manually look into financial reports for some top companies to analyse their total sales and segment sales, such as Nestle and Nike company. We find their US relevant segment proportion in total sales often remain quite stable across consecutive fiscal years.

Appendix B for Chapter Three (Essay Two)

Appendix B-1: Variable definitions

All the variables employed in the empirical tests are defined in below table. Firms with nonpositive or missing value of SALE are dropped. Definitions of the firm characteristics and relevant data cleaning methods follow Mitton (2022).[85]

Variable	Measure	Definition	Notes
Industry Concentration	Census HHI	Census HHI is calculated by summing the squared market shares of the 50 largest companies (or all companies when there are fewer than 50 firms) of each industry. Census HHI is published by the US Census Bureau every 5th year since 1982.	Main variables
	Compustat HHI	Compustat HHI is computed as the sum of squared market share of all firms in the same industry based on Compustat sales.	
	Latest Census HHI	The Latest Census HHI equals Census HHI published for the most recent Census year.	
	CGA-HHI	CGA-HHI equals the Latest Census HHI multiplied by the growth rate of Compustat HHI from the last Census year to the current year.	
	CRn	Concentration ratio (CRn) is usually defined as the proportion of the largest "n" firms in the total market share.	
	HHI floor	The equation for HHI floor is: HHI floor=$4*[CR4/4]^2+(8-4)*[(CR8-CR4)/(8-4)]^2+(20-8)*[(CR20-CR8)/(20-8)]^2+(50-20)*[(CR50-CR20)/(50-20)]^2$. Census HHI floor and Compustat HHI floor are based on Census and Compustat concentration ratios respectively. Replacing HHI with HHI floor, the Latest Census HHI Floor and CGA-HHF are calculated in the same way as the Latest Census HHI and CGA-HHI.	

[85] Nonpositive values of AT, CSHO, and PRCC_C from Compustat file are deleted. Negative values of CAPX, CH, CHE, DLTT, DLC, DVC, PPENT, TXDB, and XRD are also deleted. Missing values of XRD are replaced with zero.

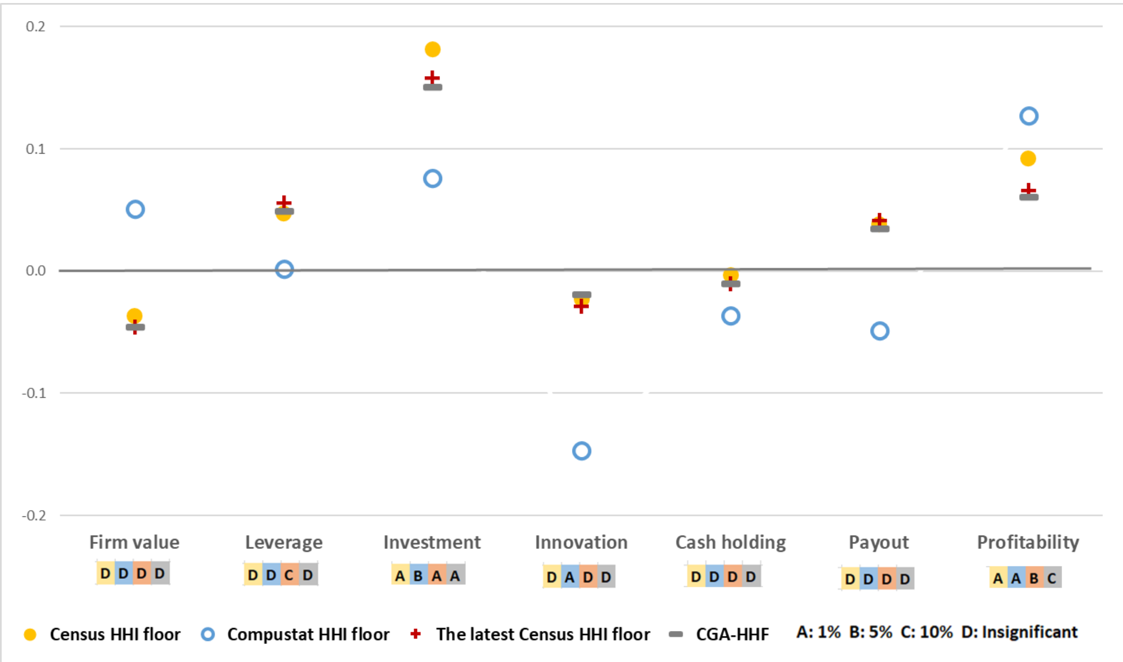
Firm Value	(AT-CEQ+ME)/AT	Firm Value equals total assets (AT) less total common equity (CEQ) plus Compustat market equity (ME), all divided by total assets. ME is Compustat stock price times shares outstanding at fiscal yearend.	Firm characteristics
Leverage	Total debt/AT	Leverage equals the long-term debt (DLTT) plus debt in current liabilities (DLC), all divided by total assets (AT).	
Investment	CAPX/AT	Investment is calculated using total annual capital expenditures (CAPX) divided by total assets (AT).	
Innovation	R&D/AT	Innovation is computed using total annual research and development (XRD) divided by total assets (AT).	
Cash Holding	(Cash + Equivalents)/AT	Cash Holding equals cash and short-term investments (CHE) divided by total assets (AT).	
Payout	Dividends/AT	Payout equals annual common/ordinary dividends (DVC) divided by total assets (AT).	
Profitability	EBITDA/AT	Profitability is calculated using annual earnings before interest, taxes, depreciation, and amortization (EBITDA) divided by total assets (AT).	
Stock return	RET (%)	Stock return refers to the holding period return from CRSP monthly file, multiplied by 100. Following Hou and Robinson (2006), we match CRSP stock return data from July of $year_t$ to June of $year_{t+1}$ with accounting information for fiscal year ending in calendar $year_{t-1}$.	Other regression variables
Market capitalisation	CRSP ME	Following Hou and Robinson (2006), we use CRSP market equity for June of $year_t$ in the replication.	
Book to market ratio	BE/ME	Book to market ratio equals book equity (BE) divided by Compustat market equity (ME). BE is stockholder’s equity plus balance sheet deferred taxes and investment tax credits minus the book value of preferred stock and postretirement assets.	
Momemtum	MOM	Momentum for each month refers to the past one-year stock return.	

Market beta	β	Ali et al. (2009) and Hou and Robinson (2006) calculate market beta in different ways. To our knowledge, Ali et al. (2009) estimate market beta by regressing stock returns on equally weighted CRSP index returns for the prior thirty-six months. While Hou and Robinson (2006) compute beta for portfolios sorted by size and pre-ranking beta and then assign the portfolio betas to stocks in those portfolios. The pre-ranking beta is estimated as the sum of the coefficients of regressions of individual monthly stock returns on contemporaneous and lagged market returns over the past three years.
Market leverage	(AT-BE)/(ME+AT-BE)	Market leverage is calculated using the difference between total assets and book equity, divided by market value (market equity plus total assets minus book equity).

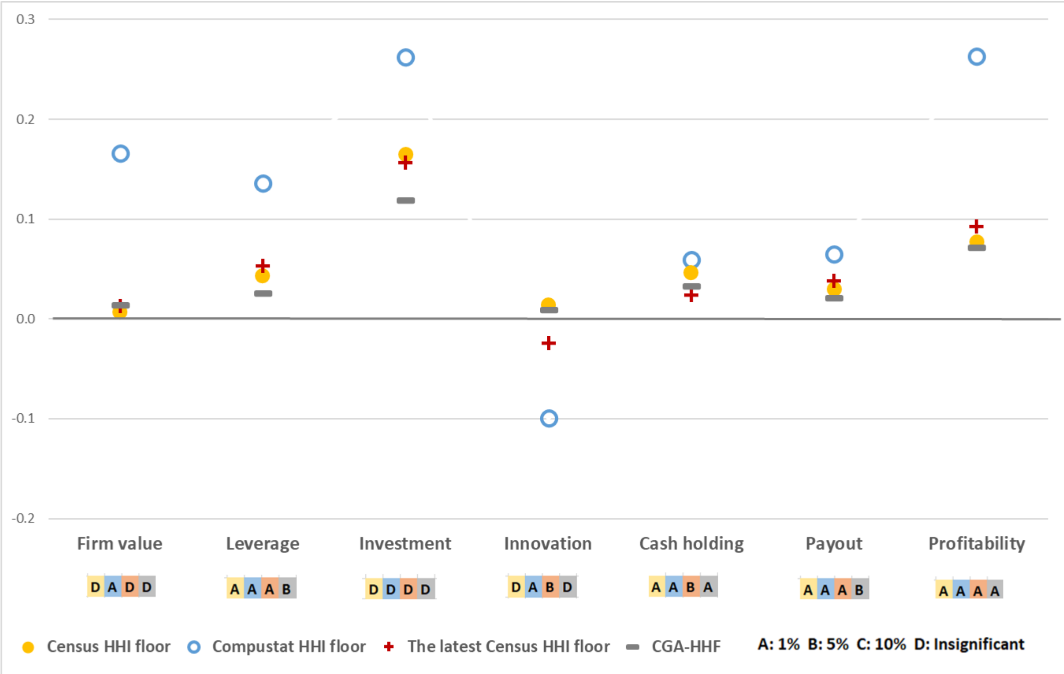
Appendix B-2: Correlations of firm variables with alternative HHI proxies for non-manufacturing industries

Appendix B-2 shows correlations of HHI measures with firm characteristics at industry level (Panel A) and at firm level (Panel B), all for nonmanufacturing industries. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The HHI measures include Census HHI floor, Compustat HHI floor, the Latest Census HHI floor, and CGA-HHF. The firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability. All variables are defined in Appendix B-1.

Panel A: Industry-level correlation



Panel B: Firm-level correlation



Appendix B-3: Replication of Mitton's (2022) summary statistics

This appendix shows the replication of Mitton's (2022) summary statistics for the seven firm variables listed below. The sample covers all firms from Compustat annual file between 1963 and 2018. Description of the variables and the data cleaning methods follow Mitton's (2022) Tables A.1 and A.2.

	Variable	Obs	Mean	Min	50%	Max	SD
Mitton (2022)	Firm Value	348,857	2.58	0.47	1.25	46.32	5.48
	Leverage	399,463	0.30	0.00	0.22	3.01	0.40
	Investment	377,106	0.06	0.00	0.04	0.45	0.08
	Innovation	403,735	0.04	0.00	0.00	0.78	0.11
	Cash Holding	400,665	0.16	0.00	0.07	0.95	0.21
	Payout	400,257	0.01	0.00	0.00	0.15	0.02
	Profitability	394,538	-0.05	-4.85	0.09	0.43	0.64
Replication	**Variable**	**Obs**	**Mean**	**Min**	**50%**	**Max**	**SD**
	Firm Value	358,071	2.67	0.47	1.25	50.43	5.94
	Leverage	427,326	0.30	0.00	0.22	3.04	0.41
	Investment	397,920	0.07	0.00	0.04	0.47	0.08
	Innovation	428,984	0.04	0.00	0.00	0.84	0.12
	Cash Holding	425,582	0.16	0.00	0.07	0.96	0.22
	Payout	423,868	0.01	0.00	0.00	0.16	0.02
	Profitability	417,975	-0.06	-5.09	0.09	0.44	0.67
Difference in Proportion	**Variable**	**Obs**	**Mean**	**Min**	**50%**	**Max**	**SD**
	Firm Value	2.64%	3.53%	0.74%	0.25%	8.87%	8.47%
	Leverage	6.98%	1.03%	0.00%	-0.74%	0.91%	1.66%
	Investment	5.52%	9.50%	0.00%	-2.29%	5.05%	4.42%
	Innovation	6.25%	1.64%	0.00%	0.00%	8.04%	10.74%
	Cash Holding	6.22%	0.65%	0.00%	-0.64%	0.97%	3.04%
	Payout	5.90%	7.96%	0.00%	0.00%	5.23%	19.51%
	Profitability	5.94%	10.92%	5.04%	-1.67%	2.60%	4.81%

Appendix B-4: Average values of firm variables across portfolios sorted by HHI proxies

This appendix reports the mean value of commonly studied firm variables across groups sorted by Census HHI, Compustat HHI, the Latest Census HHI and CGA-HHI for manufacturing industries. Groups 1 and 3 contain 30% industries with the smallest and greatest values respectively in each year, Group 2 contains the middle 40%. The firm variables include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability, defined in Appendix B-1. All the variables are standardised. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The mean values are plotted in Figure 3-1.

Variable	Census HHI				Compustat HHI				The Latest Census HHI				CGA-HHI			
	Group1	Group2	Group3	**Group 3-1**	Group1	Group2	Group3	**Group 3-1**	Group1	Group2	Group3	**Group 3-1**	Group1	Group2	Group3	**Group 3-1**
Firm Value	-0.074	0.042	0.018	0.093	0.142	-0.014	-0.121	-0.264	-0.051	0.007	0.042	0.093	-0.053	0.021	0.026	0.079
Leverage	-0.176	-0.004	0.180	0.356	-0.038	0.032	-0.003	0.036	-0.158	-0.018	0.182	0.340	-0.135	-0.018	0.158	0.293
Investment	-0.097	-0.036	0.144	0.241	0.059	-0.065	0.024	-0.035	-0.064	-0.091	0.183	0.247	-0.079	-0.060	0.157	0.236
Innovation	-0.070	0.102	-0.063	0.008	0.440	-0.123	-0.274	-0.714	-0.104	0.077	0.003	0.107	-0.089	0.073	-0.006	0.083
Cash Holding	0.040	0.072	-0.134	-0.173	0.233	-0.053	-0.161	-0.394	0.011	0.092	-0.132	-0.143	0.038	0.082	-0.145	-0.183
Payout	-0.128	0.002	0.125	0.253	-0.085	-0.031	0.121	0.207	-0.145	-0.002	0.147	0.292	-0.122	-0.003	0.125	0.248
Profitability	-0.047	-0.020	0.073	0.120	-0.188	0.000	0.184	0.372	-0.039	-0.033	0.082	0.122	-0.009	-0.055	0.081	0.090

Appendix B-5: Correlations of firm characteristics with HHI measures (firm level)

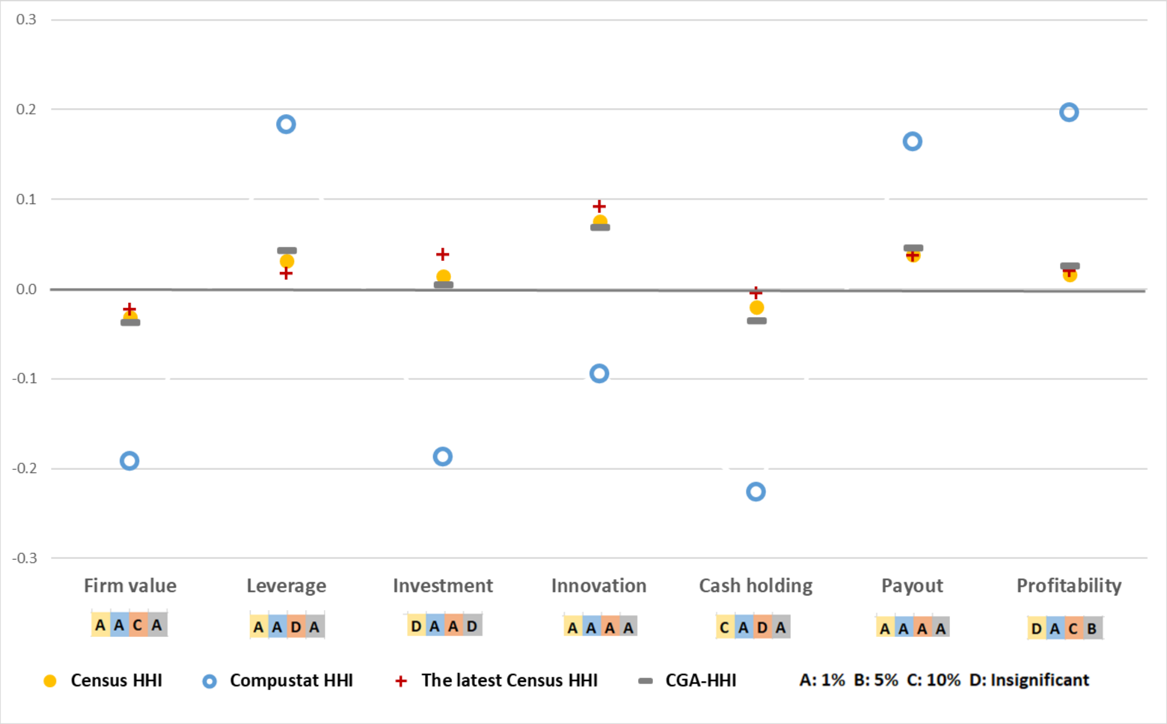
The below chart shows correlations of HHI measures with firm characteristics at firm level for manufacturing industries. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The HHI measures include Census HHI, Compustat HHI, the Latest Census HHI, and CGA-HHI. The firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability. All variables are defined in Appendix B-1.

Appendix B-6: Correlations of firm characteristics with HHI measures (alternative methods for calculating the firm characteristics)

The below chart shows correlations of HHI measures with firm characteristics at industry and firm level for manufacturing industries. The sample years include 1987, 1992, 2002, 2007, 2012, and 2017. Industry classifications are 4-digit SIC before 1997 and 6-digit NAICS after 1997. The HHI measures include Census HHI, Compustat HHI, the Latest Census HHI, and CGA-HHI. The firm characteristics include Firm Value, Leverage, Investment, Innovation, Cash Holding, Payout, and Profitability. Calculation of the firm characteristics are the second most frequently used as surveyed by Mitton (2022).

Panel A: Industry level

Panel B: Firm level



Appendix B-7: Replication of empirical results reported by Ali et al. (2009) and Hou and Robinson (2006)

In this appendix, we follow Ali et al. (2009) and Hou and Robinson (2006) to conduct FM regression of stock return on industry concentration and other control variables. Regression coefficients and the t-statistics in parentheses are reported. All variables are defined in Appendix B-1. In Panel A, we follow Ali et al. (2009) to regress individual stock return on Compustat HHI (CR4) and Census HHI (CR4) during 1980-2001 (1963-2001) for manufacturing industries classified by 3-digit SIC (4-digit SIC). Compustat HHI and Compustat CR4 are averaged over the last three years. Market leverage is chipped at 1% and 99% as described. In Panel B, we follow Hou and Robinson (2006) to regress stock return on the 3-year-average Compustat HHI during 1963-2001 for non-regulatory industries classified by 3-digit SIC. Hou and Robinson (2006) conduct regressions both at firm level and at industry level. They do not mention how they deal with the outliers, so we winsorise all variables at 1% and 99%, which is most frequently adopted method to exclude outliers in empirical tests.

Panel A: Replicate Ali et al. (2009)

Variables	Ali et al. (2009)				Replication			
	1980-2001 (SIC3)		1963-2001 (SIC4)		1980-2001 (SIC3)		1963-2001 (SIC4)	
Intercept	1.408	1.518	2.170	1.840	1.700	1.567	2.130	1.746
	(3.70)	(4.45)	(6.67)	(6.92)	(4.85)	(4.75)	(6.53)	(6.69)
Compustat HHI	-0.342				-0.474			
(3-year-average)	(-1.93)				(-2.41)			
Census HHI		-0.115				-0.124		
		(-0.24)				(-0.16)		
Compustat CR4			-0.342				-0.409	
(3-year-average)			(-2.00)				(-2.30)	
Census CR4				0.054				0.052
				(0.47)				(0.41)
Ln(Market	0.014	-0.016	-0.113	-0.111	-0.031	-0.030	-0.108	-0.109
capitalisation)	(0.24)	(-0.29)	(-2.65)	(-2.61)	(-0.57)	(-0.56)	(-2.59)	(-2.60)
Ln(B/M)	0.258	0.299	0.22	0.218	0.227	0.227	0.188	0.178
	(2.57)	(3.22)	(2.99)	(2.95)	(2.79)	(2.78)	(2.75)	(2.59)
Momentum	0.660	0.644	0.573	0.570	0.657	0.672	0.707	0.723
	(3.26)	(3.59)	(3.28)	(3.27)	(3.95)	(4.03)	(3.92)	(4.01)
Beta	-0.028	-0.098	-0.128	-0.124	-0.163	-0.139	-0.171	-0.174
	(-0.16)	(-0.61)	(-0.97)	(-0.94)	(-0.86)	(-0.71)	(-1.15)	(-1.16)
Leverage	-0.025	-0.059	0.072	0.053	-0.043	-0.090	0.085	0.050
	(-0.09)	(-0.23)	(0.35)	(0.26)	(-0.16)	(-0.33)	(0.42)	(0.25)
Average adjusted R-square	0.032	0.032	0.049	0.049	0.039	0.038	0.052	0.052
Average number of observations	991	991	936	936	859	859	730	730

Panel B: Replicate Hou and Robinson (2006)

Variables	Hou and Robinson (2006)				Replication			
	Industry-level		Firm-level		Industry-level		Firm-level	
	1963-2001 (SIC3)				1963-2001 (SIC3)			
Compustat HHI	-0.30	-0.31	-0.35	-0.42	-0.37	-0.36	-0.33	-0.31
(3-year-average)	(-2.41)	(-2.85)	(-2.41)	(-3.42)	(-4.11)	(-3.91)	(-3.30)	(-3.07)
Ln (Market		-0.25		-0.18		-0.03		-0.05
capitalisation)		(-2.98)		(-3.81)		(-0.58)		(-1.15)
Ln (B/M)		0.27		0.39		0.25		0.35
		(2.68)		(6.62)		(2.99)		(6.78)
Momentum		0.94		0.59		1.65		1.03
		(4.36)		(3.78)		(8.55)		(6.74)
Beta		-1.00		-0.41		-0.50		-0.22
		(-2.73)		(-1.95)		(-1.88)		(-1.10)
Leverage		0.04		-0.33		-0.15		-0.16
		(0.12)		(-1.7)		(-0.71)		(-0.94)

Appendix C for Chapter Four (Essay Three)

Appendix C-1: Variable definitions

Variable	Measure	Definition
Stock return	RET	RET refers to the holding period return for each stock, recorded in CRSP monthly dataset.
Market power	Markup	Markup is calculated using sales minus cost of goods sold (COGS) and the selling, general, administrative expense (XSGA), then divided by sales. When the cost items are not available, we use operating income after depreciation (OIADP) scaled by sales instead.
	VWEMK	VWEMK equals firm Markup minus the value-weighted industry Markup.
	EWEMK	EWEMK equals firm Markup minus the equal-weighted industry Markup.
Sentiment	BWR	BW raw index (BWR) is a compound proxy based on the first principal component of five market variables, as introduced by Baker and Wurgler (2006). The five variables are the closed-end fund discount, the number of IPOs, the average first-day returns on IPOs, the share of equity issues in total equity and debt issues, and dividend premium. BWR is standardised by the authors.
	BWO	BW orthogonalised index (BWO) is a compound proxy based on the first principal component of five orthogonalised market variables, as introduced by Baker and Wurgler (2006). The five variables are the same component variables for the BWR index. All the five variables are othogonalised to the growth in the industrial production index, growth in consumer durables, nondurables, and services consumption, the growth in employment, and a dummy variable for NBER recessions. BWO is standardised by the authors.
	PLS	The PLS index introduced by Huang et al. (2015) is constructed on BW's component variables by means of partial least squares approach. PLS index is orthogonalised in the same way as BWO, and is standardised by the authors.

CCI	The Consumer Confidence Index collected by the Conference Board (CCI) is a survey-based measure which directly presents participants' beliefs in the business condition, job availability, and family income. We orthogonalise CCI to the past and concurrent value of the macroeconomic factor employed by Baker and Wurgler (2006). The orthogonalised CCI is then standardised to be comparable with the other sentiment index.

Appendix C-2: Returns of market power deciles in different sentiment regimes

Appendix C-2 reports returns of deciles sorted by VWEMK, EWEMK, and firm Markup within the industry after optimistic and pessimistic sentiment. The returns are in percentage values. Decile 1 contains firms with the weakest market power and Decile 10 contains the strongest. In Panel A, we denote sentiment to be optimistic when the BWO or BWR index is above the mean (BWO or BWR>0), and pessimistic when BWO or BWR is below the mean (BWO or BWR<0). Panel B report the returns following more extreme sentiment when BWO or BWR is one standard deviation above or below the mean (BWO or BWR>1 vs. BWO or BWR<-1).

Panel A: BWO>0 vs BWO<0

Market power proxy, Sentiment stage	1	2	3	4	5	6	7	8	9	10	10-1	10-1 p-value
VWEMK, BWO>0	-0.70	0.35	0.70	0.75	0.77	0.77	0.81	0.77	0.69	0.49	1.19	0.00
VWEMK, BWO<0	1.87	2.02	1.75	1.69	1.63	1.57	1.52	1.44	1.61	1.65	-0.22	0.30
Diff	-2.58	-1.67	-1.05	-0.94	-0.86	-0.80	-0.71	-0.67	-0.92	-1.17	1.41	0.00
EWEMK, BWO>0	-0.56	0.71	0.75	0.83	0.90	0.83	0.74	0.48	0.34	0.36	0.92	0.00
EWEMK, BWO<0	1.71	1.68	1.70	1.57	1.53	1.58	1.56	1.75	1.83	1.85	0.14	0.41
Diff	-2.27	-0.96	-0.95	-0.74	-0.63	-0.75	-0.83	-1.27	-1.49	-1.49	0.78	0.00
Within-ind Markup, BWO>0	-0.10	0.27	0.54	0.62	0.73	0.74	0.77	0.69	0.74	0.68	0.78	0.00
Within-ind Markup, BWO<0	1.63	1.82	1.80	1.76	1.69	1.70	1.64	1.62	1.52	1.48	-0.15	0.37
Diff	-1.73	-1.55	-1.26	-1.14	-0.97	-0.97	-0.87	-0.93	-0.78	-0.80	0.93	0.00
VWEMK, BWR>0	-0.79	0.32	0.68	0.72	0.74	0.74	0.79	0.76	0.67	0.42	1.21	0.00
VWEMK, BWR<0	1.68	1.87	1.65	1.62	1.57	1.51	1.47	1.37	1.53	1.59	-0.09	0.65
Diff	-2.47	-1.55	-0.97	-0.90	-0.83	-0.76	-0.68	-0.62	-0.85	-1.17	1.30	0.00
EWEMK, BWR>0	-0.62	0.67	0.71	0.81	0.88	0.83	0.71	0.46	0.31	0.28	0.90	0.00
EWEMK, BWR<0	1.53	1.61	1.63	1.50	1.48	1.50	1.50	1.63	1.70	1.77	0.24	0.13

Diff	-2.15	-0.94	-0.92	-0.69	-0.60	-0.67	-0.79	-1.17	-1.39	-1.49	0.66	0.00
Within-ind Markup, BWR>0	-0.15	0.22	0.49	0.59	0.69	0.72	0.75	0.67	0.72	0.66	0.81	0.00
Within-ind Markup, BWR<0	1.49	1.70	1.71	1.67	1.62	1.62	1.57	1.55	1.45	1.41	-0.08	0.61
Diff	-1.64	-1.49	-1.22	-1.08	-0.93	-0.90	-0.82	-0.88	-0.74	-0.75	0.89	0.00

Panel B: BWO>1 vs BWO<-1

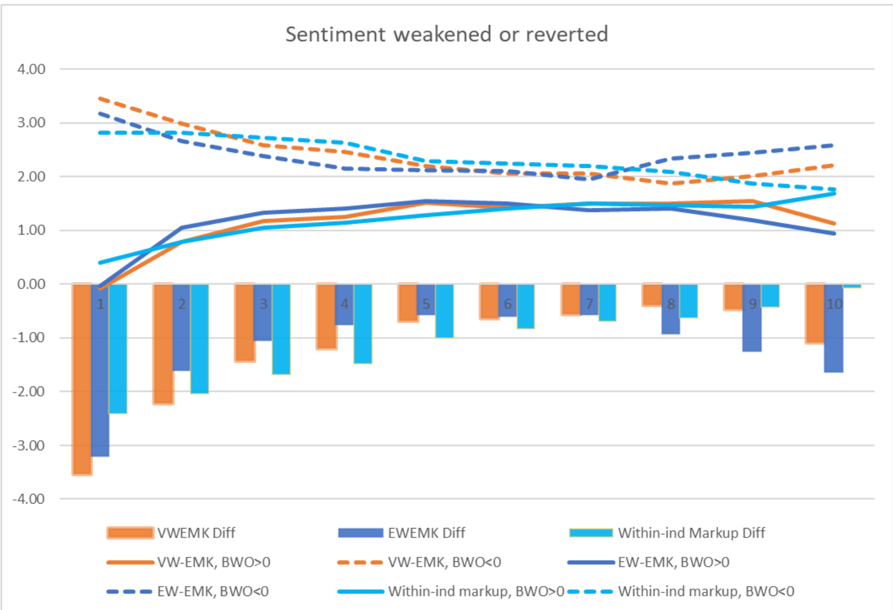
Market power proxy, Sentiment stage	1	2	3	4	5	6	7	8	9	10	10-1	10-1 p-value
VWEMK, BWO>1	-0.93	-0.09	0.28	0.51	0.61	0.66	0.75	0.63	0.56	0.35	1.28	0.00
VWEMK, BWO<-1	2.63	2.28	2.22	2.01	1.96	1.81	1.69	1.59	1.72	1.85	-0.77	0.02
Diff	-3.55	-2.37	-1.93	-1.50	-1.35	-1.15	-0.95	-0.96	-1.17	-1.50	2.06	0.00
EWEMK, BWO>1	-0.69	0.31	0.47	0.80	0.80	0.56	0.47	0.29	0.41	-0.11	0.58	0.08
EWEMK, BWO<-1	2.46	2.21	2.20	1.91	1.92	1.88	1.80	1.77	1.73	1.88	-0.58	0.07
Diff	-3.15	-1.90	-1.73	-1.11	-1.12	-1.32	-1.33	-1.48	-1.32	-1.99	1.16	0.01
Within-ind Markup, BWO>1	-0.27	-0.09	0.23	0.50	0.51	0.55	0.52	0.61	0.63	0.60	0.87	0.01
Within-ind Markup, BWO<-1	2.25	2.23	2.11	2.07	1.96	1.93	1.74	1.80	1.73	1.56	-0.70	0.02
Diff	-2.52	-2.32	-1.88	-1.57	-1.45	-1.38	-1.22	-1.19	-1.11	-0.96	1.57	0.00
VWEMK, BWR>1	-1.32	-0.27	0.23	0.49	0.54	0.57	0.64	0.49	0.37	0.15	1.47	0.00
VWEMK, BWR<-1	2.93	2.49	2.39	2.22	2.17	2.02	1.88	1.71	1.79	1.96	-0.97	0.01
Diff	-4.25	-2.76	-2.16	-1.73	-1.63	-1.45	-1.24	-1.22	-1.43	-1.81	2.44	0.00
EWEMK, BWR>1	-0.98	0.35	0.47	0.75	0.72	0.48	0.24	0.03	0.14	-0.35	0.63	0.04
EWEMK, BWR<-1	2.76	2.39	2.42	2.12	2.16	2.05	1.97	1.88	1.83	1.96	-0.80	0.01
Diff	-3.74	-2.03	-1.95	-1.37	-1.44	-1.57	-1.72	-1.85	-1.70	-2.31	1.43	0.00
Within-ind Markup, BWR>1	-0.45	-0.23	0.09	0.40	0.36	0.44	0.38	0.46	0.45	0.38	0.83	0.01
Within-ind Markup, BWR<-1	2.54	2.44	2.32	2.28	2.13	2.11	1.88	1.90	1.83	1.59	-0.94	0.00
Diff	-2.98	-2.67	-2.23	-1.88	-1.76	-1.68	-1.50	-1.44	-1.38	-1.21	1.77	0.00

Appendix C-3: Relation between sentiment level and the following return, conditional on the following change of sentiment.

In this appendix we investigate whether the predictive power of sentiment level on future stock return is influenced by the subsequent sentiment changes. Reasonably, if the relation between preceding sentiment and the following return shown in our tests is due to correction of misvaluation, then the relation should be stronger when the sentiment is weakened than strengthened in the following period. To verify this conjecture, we repeat the sorting tests in Section 4.3.1 in two sub-periods with subsequently weakened sentiment (Panel A) and strengthened sentiment (Panel B) respectively, as shown in Figure C. We sort firms to deciles each month by VWEMK, EWEMK, and within-industry Markup respectively. Then we display returns after optimistic sentiment (BWO>0, solid lines), returns after pessimistic sentiment (BWO<0, dash lines), and the differences between returns after optimistic and pessimistic sentiment for each decile (solid bars). In Panel A (B), we include periods when positive BWO decreases (increases) from the previous month to current month by more than 0.1 standard deviation, or negative BWO increases (decreases) by 0.1 standard deviation or more.[86] Details are as below.

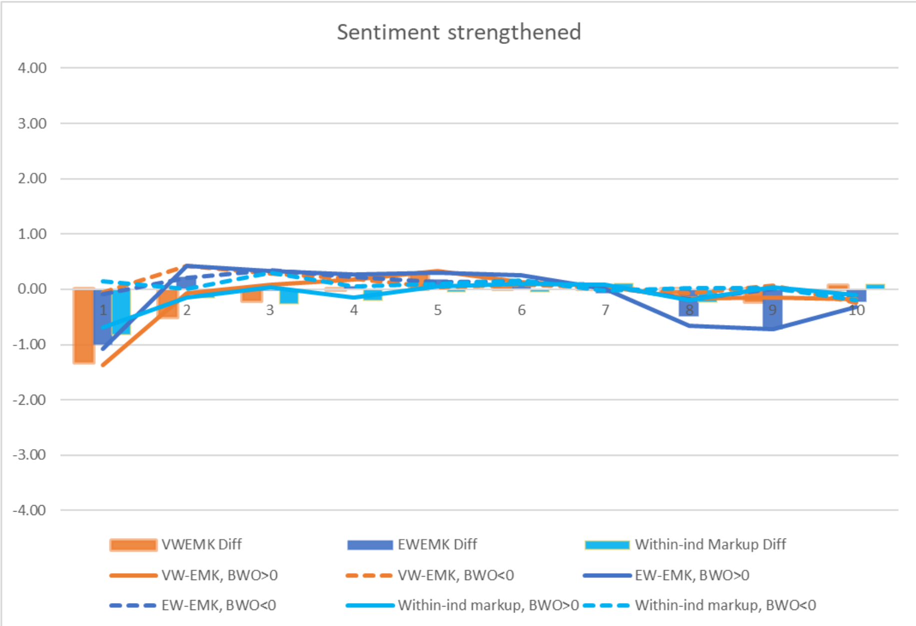
Figure C: Sentiment-return relation conditional on later sentiment changes.

Panel A: Sentiment weakens/reverts



[86] We follow Cheema and Nartea (2017) to use monthly change of sentiment in this analysis.

Panel B: Sentiment strengthens



We can see from Figure C that the differences between portfolio returns after optimistic sentiment and pessimistic sentiment vary across deciles in dissimilar ways for different subsamples. In Panel A (sentiment weakened), differences between returns after optimistic and pessimistic sentiment generally decreases from low to high deciles. Although the differences expand for top deciles sorted by VWEMK and EWEMK, the differences between “10-1” spreads after optimistic and pessimistic sentiment (1.57%-2.45%) are all positive and statistically significant at the level of 1%. Comparatively in Panel B (sentiment strengthened), it is less clear how the return differences between optimistic and pessimistic sentiment vary across the deciles. Although the differences between “10-1” spreads after optimistic and pessimistic sentiment are all positive, they are smaller (0.77%-1.38%) and insignificant.

To sum, our tests show that the negative relation between market power and return vulnerability to the preceding investor sentiment is more pronounced when sentiment is subsequently weakened than strengthened. To check the robustness of the results, we repeat the tests using abnormal returns accounting for Fama and French 5 factors and momentum factor. Our conclusion is basically unaffected. All the relevant statistics are available upon request.

Appendix D: Statement of contribution

Appendix D-1: Statement of contribution (Chapter Two)

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Suxiang Zang
Name and title of main supervisor:	Professor Hamish Anderson
In which chapter is the manuscript/published work?	Chapter Two

Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1]

The student: Zang, Suxiang: 70% contribution.
Members of the supervisory team: Anderson, Hamish: 15% contribution; Chang, Yuk Ying: 15% contribution.
Suxiang Zang gathered the data, conducted analyses, completed the draft of the chapter. The supervisory team provided guidance on research design, method and writing.

Please select one of the following three options:

- ○ **The manuscript/published work is published or in press**
  Please provide the full reference of the research output:
- ○ **The manuscript is currently under review for publication**
  Please provide the name of the journal:
- ◉ **It is intended that the manuscript will be published, but it has not yet been submitted to a journal**

Student's signature:	Suxiang Zang — Digitally signed by Suxiang Zang Date: 2023.06.08 10:20:49 +12'00'	Main supervisor's signature:	Hamish Anderson — Digitally signed by Hamish Anderson DN: cn=Hamish Anderson, c=NZ, o=Massey University, ou=School of Economics & Finance, email=h.d.anderson@massey.ac.nz Date: 2023.06.08 09:57:58 +12'00'

This form should be placed at the beginning of each relevant thesis chapter.

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Doctoral Research Committee
May 2023

Appendix D-2: Statement of contribution (Chapter Three)

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.			
Student name:	Suxiang Zang		
Name and title of main supervisor:	Professor Hamish Anderson		
In which chapter is the manuscript/published work?		Chapter Three	
Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1] The student: Zang, Suxiang: 70% contribution. Members of the supervisory team: Anderson, Hamish: 15% contribution; Chang, Yuk Ying: 15% contribution. Suxiang Zang gathered the data, conducted analyses, completed the draft of the chapter. The supervisory team provided guidance on research design, method and writing.			
Please select one of the following three options:			
○	**The manuscript/published work is published or in press** Please provide the full reference of the research output:		
○	**The manuscript is currently under review for publication** Please provide the name of the journal:		
●	**It is intended that the manuscript will be published, but it has not yet been submitted to a journal**		
Student's signature:	Suxiang Zang — Digitally signed by Suxiang Zang Date: 2023.06.08 10:17:33 +12'00'	Main supervisor's signature:	Hamish Anderson — Digitally signed by Hamish Anderson DN: cn=Hamish Anderson, c=NZ, o=Massey University, ou=School of Economics & Finance, email=h.d.anderson@massey.ac.nz Date: 2023.06.08 09:58:22 +12'00'
This form should be placed at the beginning of each relevant thesis chapter.			

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Doctoral Research Committee
May 2023

Appendix D-3: Statement of contribution (Chapter Four)

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Suxiang Zang
Name and title of main supervisor:	Professor Hamish Anderson
In which chapter is the manuscript/published work?	Chapter Four

Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1]

The student: Zang, Suxiang: 70% contribution.
Members of the supervisory team: Anderson, Hamish: 15% contribution; Chang, Yuk Ying: 15% contribution.
Suxiang Zang gathered the data, conducted analyses, completed the draft of the chapter. The supervisory team provided guidance on research design, method and writing.

Please select one of the following three options:

○ **The manuscript/published work is published or in press**

Please provide the full reference of the research output:

○ **The manuscript is currently under review for publication**

Please provide the name of the journal:

◉ **It is intended that the manuscript will be published, but it has not yet been submitted to a journal**

Student's signature:	Suxiang Zang — Digitally signed by Suxiang Zang Date: 2023.06.08 10:21:21 +12'00'	Main supervisor's signature:	Hamish Anderson — Digitally signed by Hamish Anderson DN: cn=Hamish Anderson, c=NZ, o=Massey University, ou=School of Economics & Finance, email=h.d.anderson@massey.ac.nz Date: 2023.06.08 09:57:06 +12'00'

This form should be placed at the beginning of each relevant thesis chapter.

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Doctoral Research Committee
May 2023