THREE ESSAYS IN FINANCIAL ECONOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Dawoon Kim

August 2019

THREE ESSAYS IN FINANCIAL ECONOMICS

Dawoon Kim, Ph. D

Cornell University 2019

In this dissertation, I explore three topics in financial economics focusing on how financial decision making has evolved under the various changes that surround the capital market. First, I study the potential value of big data analysis in the product market and document evidence that it may affect investors' decision making in the capital market. Second, I focus on how changes in firms operating performance result in changes in corporate decision making and firm composition around the world. Third, I focus on how changes in managerial labor market affect corporate policies.

In the first chapter, I measure the likelihood of a product review being spam using machine learning techniques and show that this likelihood contains valuable information for investors. I also document evidence that this likelihood is negatively correlated with levels of independent institutional holdings. This suggests that there exist sophisticated investors with superior information processing capabilities who can filter out noises in consumer opinions and benefit from it. However, the above significant investment value in filtering out spam reviews suggests that it is not enough to fully incorporate the information into price.

In the second chapter, I document that the number of public firms with persistent operating losses has increased significantly around the world and that such firms usually hold high levels of cash. I find evidence that cash holdings in firms with operating losses are positively correlated with investment, R&D, and firm value but negatively correlated with the probability of bankruptcy and delisting from public exchanges. This

finding is against the view that excessive amount of cash holdings is against investors' interest. Rather, this means high levels of cash holdings in firms with operating losses is in investors' interest.

In the third chapter, I document the movement of executives to new jobs across industries and show that it has grown stronger in recent decades. Then, I show evidence that shift in labor market trends are strongly correlated with changes in various corporate policies. When CEOs see better future labor market opportunities, they have more incentive to perform better in their current employment. I thus argue that an active managerial labor market mitigates incentive problems under dynamic agency models.

BIOGRAPHICAL SKETCH

The author was born in Seoul, South Korea in June 1983. He attended Seoul National University and graduated with a Bachelor of Arts degree in Business Administration and Economics in 2010, and also graduated with a Master of Science degree in Finance. He began his doctoral studies in management at Cornell University in 2013. He pursued his career research in Finance under the direction of Andrew Karolyi.

I dedicate this dissertation to my beloved family. Especially to my wife, Shi Eun who has given me her sincere support from the very beginning of my time in Ithaca. I must also thank my advisor Andrew Karolyi for his everlasting patience and encouragements. His guidance helped me get through the difficult yet enjoyable times during my times at Cornell.

ACKNOWLEDGMENTS

I am very grateful to my advisor Andrew Karolyi and the committee members, Murillo Campello, Scott Yonker and Hyunseob Kim, for their continuous guidance, support and encouragement. I am also indebted to Warren Bailey, David Ng, Justin Murfin, Matthew Baron, Adrian Aycan Corum, Jawad Addoum, Fang Liu, Byoung-Hyoun Hwang, Pamela Moulton, Gaurav Kankanhalli, and Alan Kwan for their insightful comments and suggestions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This dissertation is an original intellectual product of the author, Dawoon Kim. Chapters one and two are solo work of the author and the Chapter three is a co-work by the author and his coauthors, Hyunseob Kim at Cornell University and John Graham from Duke University.

I could not have achieved these chapters without a strong support from my dissertation committee, Andrew Karolyi (Chair), Murillo Campello, Scott Yonker, and Hyunseob Kim, each of whom has provided invaluable advice and guidance throughout the research process. I am indebted to their unwavering support.

CHAPTER 1: NOT ALL INVESTORS KNOW: THE INVESTMENT VALUE OF SPAMMING CONSUMER OPNIONS

1.1 Introduction

The growth of internet shopping has given rise to new tools that firms and consumers use to interact with each other. One such tool is the consumer product review. Not only do consumer opinions contained in product reviews convey information about products to potential consumers, they can also help firms and their investors make investment decisions based on the market's perception of product quality. While there may be considerable informational value in consumer opinions, however, there may also be an incentive to influence those opinions by means of spam reviews,[1] also known as opinion-spamming ("spamming" hereafter). Detecting spam reviews is a non-trivial exercise. In this paper, I measure the probability that a review is spam using machine-learning techniques adapted from computer science and find evidence that a significant amount of valuable information is hidden in consumer spamming.

As internet usage has grown (seemingly exponentially) so has e-commerce.[2] E-commerce helps firms streamline their distribution channels and find consumers. Consumers also benefit from e-commerce due to the ease of transactions and enhanced search capabilities. What may be equally, if not more, important is communication between firms and consumers online. The presence of information asymmetry between consumers and firms causes firms to rely heavily on advertising to signal their quality of their products (e.g., Milgrom and Roberts, 1986). This information asymmetry is even more severe for e-commerce as physical quality inspection before making a

[1] In this paper, I define spam reviews broadly as unsolicited product reviews with commercial intent designed to influence consumer behavior. This includes fake reviews and other forms of biased reviews sent with commercial intent.

[2] For instance, U.S. manufacturing firms' aggregate shipment value from e-commerce grew from $1 trillion in 2004 to $3.6 trillion in 2014 (a 260% increase) and it is about 61.4% of their total shipment value in 2014 (an 89.1% increase from 2004). See multi-sector e-commerce data from the United States Census Bureau available https://www.census.gov/data/tables.html.

purchase is practically impossible.[3] In this context, consumer opinions provide critical means for e-commerce vendors to reduce information asymmetry by providing product quality information to potential consumers.[4]

Consumer opinions may be as important to investors as they are to firms insofar as they are associated with firms' future cash flows. For example, recent studies by Fornell, Morgeson, and Hult (2016) and Huang (2018a) find evidence that consumer opinions contain information that is not fully captured by the stock market. The underlying reason that publicly available consumer opinions are not fully reflected in financial markets is yet to be investigated. In this paper, I hypothesize that the noise diluting consumer opinions that arises from spam reviews is an important factor. If spamming is conducted to dilute unfavorable opinions, their presence will convey information about product quality that is not apparent from observed review scores. Given that spamming tends to distort consumer opinions, spam detection will provide more accurate signals of product quality that may benefit investors who can exploit such information.

Traditional signaling models have focused more intently on the cost of transmitting information from firms to investors than on the cost of gathering and processing data as the primary source of information about a firm. Because managers gather the relevant information they need to operate a firm, the marginal cost of gathering such information by investors has been deemed low (Merton, 1987).[5] In light of ever-advancing information technology and tools for big data analysis, however, the cost of gathering and processing information from a myriad of databases may be at least as important (Dugast and Foucault, 2018).

---

[3] Not surprisingly, internet advertising has become the fastest growing sector in the advertising industry. For example, the aggregate revenue from internet advertising in the U.S. grew from $2.7 billion in 2004 to $49.5 billion in 2014 (a 1,733% increase), an amount that is about 52.8% of the total advertising revenue in 2014 (a 1,190% increase from 2004). See PWC's internet advertising revenue report, available https://www.iab.com/insights/.

[4] Consequently, the success of e-commerce platforms has grown in tandem with the volume of consumer opinion supplied. For instance, Amazon.com's share of the online sales market in the U.S. has risen to 25.8% and, while it attracted 3.74 reviews per product in 2004, this figure increased to 5.35 reviews per product in 2014 (a 43% increase). Also, Yelp.com has now secured more than 70 million reviews of various products and services; its figure of 1.52 reviews per business in 2004 and had risen to 11 reviews per business in 2014 (a 622.9% increase). See Huang (2018b).

[5] Thus, those models focused on conveying the relevant information to investors because attention is a scarce resource. Hence, firms engage in activities designed to attract increased attention from investors and to reduce information asymmetry between firms and investors. See, for example, Karolyi, Kim and Liao (2018).

The cost of gathering and processing information will be amplified if it is accompanied by a significant amount of noise, rendering it difficult to extract relevant information cleanly even from publicly available data.[6] Therefore, investment value might exist in information derived from the product market (such as consumer opinions) when spamming is non-negligible. Consequently, perfect informational efficiency can be attained only if sophisticated investors with superior information-processing capability extract the costly information and reveal it to the market. This process will be incomplete, though, as long as such activities are costly (Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981; Verrecchia, 1982; Kyle, 1985).

The cost of detecting spam reviews comes from several sources. First, there is no clear way to credibly identify a set of spam reviews to train a machine-learning algorithm or to verify the external validity of any estimates (e.g., Ott et al., 2011; Mukherjee, Liu, and Glance, 2012; Mukherjee et al., 2013). Second, spam detection is multi-faceted insofar as it requires a vast amount of data-processing to understand reviewers' behaviors and manufacturers' spamming incentives (e.g., Jindahl and Liu, 2008; Lim et al, 2011; Xie et al., 2012). Third, it is unclear how much value there is in detecting possible spam reviews. This last point is the focus of this paper.

Using almost 45 million product reviews ("reviews" hereafter) from Amazon.com ("Amazon" hereafter) that were posted online from 2004 through 2014, I first measure the probability that a review is fake/spam ("spam" hereafter) via supervised machine learning. To construct the training sample where so-called positive examples (generally accepted spam reviews) are labeled, I identify 435,456 nearly identical reviews following the approach of Jindhal and Liu (2007, 2008). These reviews are suspected of being spam because spam reviewers write the same or similar reviews on multiple products sometimes using separate accounts.

The estimated average probability that a review in my sample is a spam review (or what I call "spamicity") during the sample period is 24.6%, and ranges between 17% and 30% across

[6] There is considerable evidence in economics and computer science pertaining to information noise such as spam reviews (see, among others, Mayzlin, Dover and Chevalier, 2014; and Luca and Zervas, 2016) as well as misleading advertisements, fake news, and phishing emails (for example, Böhme and Holz, 2006; Clarke et al., 2018; Kogan, Moskowitz, and Stein, 2018).

time and product categories. The estimated spamicity of a review is negatively correlated with its abnormal review score ("abnormal score" hereafter) and is positively correlated with abnormal scores of peer products. These findings imply that: (1) spamming decisions are correlated with levels of consumer opinions; (2) large portions of spam reviews are produced to counteract unfavorable opinions; and (3) counteracting unfavorable opinions about a firm's own products is more likely to succeed if competing products receive positive opinions. I also find that spamicity is positively correlated with advertising expenses and a host of firm-level governance quality indices; spamming appears to be a deliberate tactic aimed at improving product-market performance rather than a manifestation of poor firm governance.

In the second part of the paper I show evidence that estimated spamicity provides significant investment value among Amazon product reviews that are associated with public companies. Using 469 unique Compustat firms whose products are sold on Amazon in the U.S., I form tercile portfolios in each month based on spamicity, following Huang (2018a), and show that a spread portfolio that goes long on stocks with low spamicity and that goes short on those with high spamicity yields a Fama-French-Carhart four-factor alpha (Fama and French, 1993; Carhart, 1997) of around 60 basis points per month.[7] This evidence suggests that spamicity contains information about product quality that may be associated with firms' future cash flows.

To further identify the informational value in spamicity, I form the abovementioned spread portfolios conditional on each abnormal score tercile. I find that spamicity is most informative in the mid-tercile of abnormal scores where signals from monthly abnormal scores are mixed to the greatest extent, yielding a Fama-French-Carhart four-factor alpha of around 98.9–112.2 basis points per month. In addition, spamicity predicts future returns. Fama-MacBeth regression results show that spamicity is negatively correlated with one-month-ahead excess returns. A one-standard-deviation increase in spamicity is associated with a decrease of 26.4 basis points in one-

[7] I also replicate Huang (2018a) by forming a spread portfolio that goes long on stocks with high abnormal scores and that goes short on those with low abnormal scores and find a Fama-French-Carhart four-factor alpha of around 70 basis points per month. Huang (2018a) finds a Fama-French-Carhart four-factor alpha of 55.7–73.0 basis points per month in his sample for a period running from 2004 through 2015.

month-ahead excess returns. This effect is amplified by abnormal scores.

I then form a trading strategy based both on scores and spamicity to verify the potential investment value of spam detection. A spread portfolio that goes long on stocks with high abnormal scores *and* low spamicity and that goes short on those with low abnormal scores *and* high spamicity yields a monthly Fama-French-Carhart four-factor alpha of between 1.17% and 1.23% per month.[8] The observed abnormal returns are economically and statistically significant among firms with: (1) high e-commerce sales proxied by average product-level e-commerce sales, (2) large numbers of new products sold on Amazon, and (3) high average product prices. The monthly Fama-French-Carhart model-based alphas among such firms are 1.41%, 1.45%, and 1.34%, respectively. These results imply that the investment value of spam detection is concentrated among firms that are affected to a greater extent by consumer opinions and thus are more likely to suffer from, or have incentives to commit, spamming.

Next, I show evidence that sophisticated investors can discern the credibility of consumer opinions. For instance, a one-standard-deviation increase in spamicity lowers institutional ownership by 46 basis points, which accounts for 1.77% of the standard deviation of institutional ownership in the sample. This effect is concentrated among independent institutional investors such as mutual fund managers and investment advisors, who are more likely to collect various sources of information as opposed to "grey" institutions (e.g., pension funds and endowments), per the definitions given in Ferreira and Matos (2008). Given the large abnormal returns from the spread portfolios on abnormal scores *and* spamicity, the small economic magnitude suggests that not enough institutions take advantage of the potential information harbored in spamicity.

I then offer new evidence that the positive correlation between earnings surprises and abnormal scores found by Huang (2018a) is substantially weaker if a firm exhibits high levels of spamicity in its product reviews. A one-standard-deviation increase in spamicity decreases the positive correlation between abnormal scores and earnings surprises by 50%. This particular effect

[8] I find no evidence of return reversal beyond a one-month holding period. See Table 1.A2 for results pertaining to abnormal returns over longer holding periods.

is concentrated among firms with below-median analyst coverage, suggesting that some analysts may be deriving at least some information from the product market. Consistent with these findings, abnormal returns from spread portfolios on abnormal scores *and* spamicity are concentrated among firms with lower percentages of independent institutional ownership and analyst coverage.

In one final experiment, I investigate the consequences of a new policy adopted by Amazon that altered firms' incentive to engage in spamming. In 2012, Amazon launched its "Global Selling Program," an initiative that allows many Chinese manufacturers to list and sell their products directly on Amazon in the U.S.[9] As a result, the number of Chinese manufacturers/sellers on Amazon increased substantially (by a factor of 13 from 2012 to 2015).[10] The sudden influx of products from China is found in my sample as well. I use universal barcodes such as UPCs (Universal Product Codes) and an EANs (European Article Numbers) where the first few digits indicate the country where the product was barcoded. The ratio of the products barcoded in China to those in the U.S. increased by more than 158.2% between 2011 and 2013.[11]

The influx of such products appears to have increased competition and thus in turn increased incentives to engage in spamming. There is a disproportionate increase in spamicity among firms whose products included above-median Chinese products in their product categories. I find evidence that only independent institutional investors that held shares of firms with higher exposure to the policy change adjust their holdings against spamicity after 2012. This finding is consistent with the hypothesis that only a few sophisticated investors are willing and able to invest in analyzing the information embedded in consumer opinions reflected in spamicity. Less sophisticated investors are unlikely to invest resources to detect spamming and therefore cannot fully infer such information from consumer opinions. Thus, there remains investment value in

[9] Previously, most Chinese manufacturers had to sell their products abroad through importers. The program significantly lowered the bar for Chinese manufacturers to list their products on Amazon websites around the world. This, however, differentiates them from products manufactured in China by U.S. manufacturers via original equipment manufacturing (OEM).

[10] See a news article on the increase in Chinese manufacturers (sellers as well) after using the Amazon program http://m.chinadaily.com.cn/en/2015-12/04/content_22626509.htm.

[11] Many products in the sample lack product identifiers such as UPCs or EANs. As the program allowed many generic products to be sold on Amazon, this estimate is expected to be the lower bound.

spam detection.

To the best of my knowledge, this paper is the first to investigate the implications of spamming on investment returns. The remainder of the paper is organized as follows. In Section 1.2, I summarize the related literature on consumer opinions and spamming. In Section 1.3, I describe the data and empirical methods. In Section 1.4, I explain the spam-detection model used to estimate the spamicity of reviews. In Section 1.5, I present the empirical results and in Section 1.6, I conclude the chapter.

1.2 Literature Review

1.2.1 Consumer Opinions

This paper contributes to the literature on the importance of consumer opinions to firms and investors. The role that positive consumer opinions play in product sales and marketing strategy is well-documented in the marketing and strategy literatures (see, among other studies, Bickart and Schindler, 2001; Godes and Mayzlin, 2004; Chevalier and Mayzlin, 2006; Liu, 2006; Chen and Xie, 2008; Luca, 2011; and Archak, Ghose, and Ipeirotis, 2011). For instance, Chevalier and Mayzlin (2006) find that improved book reviews on Amazon and Barnes & Noble are positively correlated with increases in book sales. Also, using review data from CNET and Amazon, Chen and Xie (2008) find evidence that consumer reviews provide a new tool that sellers use to adjust their marketing strategies in response to consumer reviews. Finally, Archak, Ghose, and Ipeirotis (2011) find that textual contents in consumer reviews can be used to understand consumers' preferences and to forecast changes in product sales.

Just as consumer opinions are related to firm performance, so they also matter to investors. The role of consumer opinions in investment and corporate decision-making is also well-documented (see, among other studies, Ittner and Larker, 1998; Larkin, 2013; Fornell, Morgeson, and Hult, 2016; Huang, 2018a; Huang 2018b). For example, a recent study by Huang (2018a) finds evidence that consumer opinions contain information that is not fully recognized by the market. Using review data from Amazon, he finds that a spread portfolio that goes long on stocks with

high abnormal scores and short on those with low abnormal scores earns monthly abnormal returns of between 55.7 and 73.0 basis points. Meanwhile, Huang (2018b) finds evidence that positive consumer opinions ease firms' financial constraints. Using review data from Yelp, he shows that a one-half-star rating increase is associated with a 40% higher probability of securing Small Business Administration loans.

The investment value of consumer opinions is also related to the literature on the limited attention and information-processing capabilities of investors (see, among other studies, Hong and Stein, 1999; Hong, Lim, and Stein, 2000; Hirshleifer and Teoh, 2003; Hou and Moskowitz, 2005; Peng and Xiong, 2006; Cohen and Frazzini, 2008; DellaVigna and Pollet, 2009; and Dugast and Foucault, 2018). For instance, Cohen and Frazzini (2008) find evidence that the stock market does not fully incorporate news about economically linked firms, and DellaVigna and Pollet (2009) find that earnings announcements issued on Fridays, when investors are more likely to be inattentive, receive a 70% higher delayed response from the market. This paper contributes to this literature by providing a possible channel through which information-processing costs prevent the market from fully incorporating the relevant information. Spamming, along with other sources of fake or misleading information, induces another layer of information asymmetry between producers and consumers of information. The findings I report in this paper suggest that, as a result of such asymmetry, information from the product market may be incorporated in the capital market on a delayed basis.

1.2.2 Opinion Spamming

I adapt the machine-learning technique used in this paper from the literature on spam detection in computer science (see, for example, Jindahl and Liu, 2007, 2008; Lim et al, 2011; Ott et al. 2011; Xie et al. 2012; Mukherjee, Liu, and Glance, 2012; and Mukherjee et al., 2013). The spam detection problem was first brought to light by Jindahl and Liu (2008). Using review data from Amazon, they find that near-duplicate reviews are likely to be spam, and thus use this finding as a positive training example for spam detection. They show that spamming is a widespread

problem and incorporating reviewer- and product-level characteristics improves the performance of their model. Subsequent work focuses on behavioral patterns of reviewers and shows that spam-detection model performance improves greatly if such patterns are considered (Lim et al., 2011; Xie et al. ,2012).

The prevalence of spam reviews is also recognized by both economists and practitioners.[12] For example, by comparing a given hotel's reviews on Expedia, where only customers can post reviews, with reviews of that hotel on TripAdvisor, where anyone can post reviews, Mayzlin, Dover, and Chevalier (2014) find evidence that hotels with strong incentives to post fake reviews attract more positive reviews while their neighbors attract more negative reviews. Luca and Zervas (2016) also find that restaurants attracting few reviews or recent bad reviews are more likely to commit review fraud as measured by filtered reviews (approximately 16% in their sample) on Yelp. This paper contributes to this literature by providing additional large-scale analyses of determinants of spamming and its economic consequences. Finally, this paper provides a new insight into the economic value of big-data processing (Lee, Ma, and Wang, 2015; Huang, 2018b; Froot et al., 2017; and Zhu, 2018).

1.3 Data and Hypothesis Development

1.3.1 Data

I obtain Amazon review data from the Stanford Network Analysis Project (SNAP) for January 2004 through July 2014.[13] SNAP is widely used in computer science, information technology, and marketing for network and sentiment analyses. The raw data contain more than 140 million product reviews covering almost all the reviews posted within the sample period. As Amazon merges reviews of nearly identical products, duplicate reviews of similar products by the

[12] Find examples https://www.washingtonpost.com/business/economy/how-merchants-secretly-use-facebook-to-flood-amazon-with-fake-reviews/2018/04/23/5dad1e30-4392-11e8-8569-26fda6b404c7_story.html.
[13] See McAuley et al. (2015), McAuley, Pandey, and Leskovec (2015) and McAuley and Alex (2016) for details on the data. I thank Julian McAuley for sharing the data.

same reviewer are removed from the sample.[14] Also, I exclude reviews in product categories such as "Books," "Movies and TV," "Kindle Store," "Digital Music," and "Amazon Instant Video," as reviews of these products are expected to provide less information about the manufacturers. The resulting review sample contains 44,699,245 reviews. Each review contains information on 1) the product identifier (ASIN: Amazon's product identification number), 2) reviewer ID, 3) the score given in the review, 4) the title of the review, 5) the contents of the review, 6) the review helpfulness indicator, and 7) the date of the review.

Additionally, I obtain metadata on Amazon products from SNAP. These metadata provide a detailed product description and a snapshot of each product's price.[15] Also, the data include information on each product's related products (e.g., "Also Bought," "Also Viewed," "Bought Together," and "Buy After Viewing").[16] As a complement to each product's review score, I track down the review scores of related products as well.[17]

Also, as the review data do not provide information on manufacturers, I obtain additional metadata on Amazon products via ASINtool.com, where one can acquire detailed product information including manufacturers and brand names, using ASINs. I then manually match those names to firms in Compustat. Following Huang (2018a), I exclude firms that receive fewer than ten total reviews in each month. The resulting sample comprises 469 unique U.S. firms used in the main analyses.[18] One interesting feature of these metadata is that they contain detailed information on each product's registration numbers, such as UPCs, EANs, MPNs (Manufacturer Part Numbers), and SKUs (Stock-Keeping Units). Although the last two identifiers are not universal among manufacturers, UPCs and EANs are standardized (barcode numbers) where the first few

[14] For example, reviews of product A in the color red can be merged into reviews of the same product B in the color blue and vice versa.
[15] This is the price of a product when the review data were collected.
[16] According to Amazon, this relatedness is determined by Amazon's recommendation algorithms, which measure cosine similarities between customers. See Amazon's industry report https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf.
[17] Although the metadata do not indicate historical changes in product relatedness, such relatedness is not expected to change much in the sample as the cost of measuring such relatedness among all products is exorbitant. I assume that this relatedness is constant throughout the sample period and track related products' review scores as they appear in the sample.
[18] This excludes financial firms (SIC one-digit "6") and utility firms (SIC two-digit "49").

digits of each indicate the country where the product was barcoded.

Next, I obtain data on firm-level characteristics, stock returns, analyst forecasts, and institutional holdings from Compustat, CRSP, IBES, and Factset Lionshares. For monthly analyses, I impute quarterly or annual firm-level observations based on data availability. Additionally, I obtain quarterly firm-level governance and ESG indices from MSCI ESG GovernanceMetrics. Lastly, I obtain sentiment lexicons (a list of words and their sentiment scores) from Hamilton et al. (2016) to measure the sentiment expressed by each review's content from textual analysis.[19]

[Insert Table 1.1 Here]

In Table 1.1, I report review, reviewer, and product information. The review sample (Full Review Sample) contains 44,699,245 reviews of 4,977,718 products posted by 26,482,938 reviewers, and the Compustat matched sample (Main Sample)[20] contains 3,827,524 reviews of 250,554 products posted by 3,100,134 reviewers.[21] Panel A of Table 1.1 displays a year-by-year summary of average scores and the number of reviews, reviewers, and products for each year from 2004 through 2014. All the numbers, except for average scores, grew rapidly in each year, especially in 2013. The per-product number of reviews also peaks in both samples (6.32 and 9.64 reviews per product, respectively).

In Panel B of Table 1.1 I report the average review scores (or "scores" interchangeably) and the number of products and reviewers by Amazon product category. Average scores are clustered around 4.00 in both samples (with averages of 4.07 and 4.08, respectively). Among the 18 product categories, "Apps for Android" and "Video Games" attracted the highest number of per-product reviews (43.0 and 24.3 reviews per product, respectively) followed by "Electronics"

[19] I thank William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky for sharing the lexicon data, which are available https://nlp.stanford.edu/projects/socialsent/.
[20] Manufacturer and brand information is not available for all products offered on Amazon. Products for which no manufacturer or brand names are given on the website are not included in the main sample.
[21] These numbers do not reflect reviews written by the same reviewers on the same product, which were excluded from the sample. This reflects either the merging of reviews of nearly similar products by Amazon or technical errors that occur when reviewers (apparently inadvertently) hit the submit button multiple times.

and "Baby" (16.3 and 14.0 reviews per product, respectively). This pattern may occur because these are the types of experience goods that need the highest level of indirect quality inspection by consumers. Meanwhile, the numbers of reviews and reviewers are highly correlated in both samples (ρ = 0.993 and 0.999, respectively) while the correlation is much weaker between the number of reviews and the number of products (ρ = 0.734 and 0.642, respectively). This suggests that the supply of consumer opinions is not a mere function of the number of products consumers can review.

This also reflects a possible determinant of the supply of consumer opinions. For a given score shown to consumers before they purchase, those consumers who perceive that the quality of a product they might purchase is close to the score have little reason to write a review (Li and Hitt, 2008). If there is a significant discrepancy between the score and the perceived quality,[22] however, consumers are more likely to express their opinions until the score is sufficiently close to the perceived quality. Given the skewed high average scores in the sample, low-quality products are expected to accumulate more reviews than high-quality products, as such a discrepancy will be significant for low-quality products. Consistent with this hypothesis, the number of reviews per product and the average scores across product categories shown in Panel A are negatively correlated in both samples (ρ = -0.450 and -0.221, respectively).

The process by which the wisdom of the crowd ensures that indicators of quality such as review scores converge on the true quality in the product market bears some resemblance to how the capital market attains informational efficiency via arbitrageurs, but there exist several differences. First, unlike arbitrageurs who make riskless profits by eliminating inefficiency, reviewers receive little or no reward for their efforts.[23] Second, unlike the market price, which holds across venues, quality indicators can vary in kind. As a result, the process by which efficiency in quality indicators in the product market is ensured may not play out as quickly or as

[22] Here, quality is rather broadly defined to indicate the capacity of a product to satisfy consumers' expectations given its price and other product features.

[23] Some reviewers may receive benefits from firms and platforms when they write reviews for professional purposes (e.g., Paid reviews on Amazon). However, this type of review is more closely related to review spamming as it functions more like advertising than impartial expressions of opinion.

efficaciously as the arbitrage mechanism in the capital market. This may enable spamming to have a more lasting impact on consumer behaviors.

1.3.2 Hypothesis Development

When information such as consumer opinions that derives from the product market is free of noise, it can be efficiently incorporated into the market price. Thus, there will be no investment value in deriving such information from the product market. This process may not be as efficient as possible, however, if 1) there is a significant amount of noise in the information from the product market and 2) detecting such noise is costly. Subsequently, such information from the product market will continue to provide investment value that is not fully recognized by the capital market.

The premise of this paper stems from evidence that there is value in information derived from the product market (Fornell, Morgeson, and Hult, 2016; Huang, 2018a; Hoberg and Philips, 2018). By narrowing the focus on information conveyed in publicly observable product reviews, I hypothesize that the value of such information comes from the noise that lies within it. To verify this hypothesis, I first replicate Huang (2018a) to show that the information in the sample indeed has significant investment value.

Next, in the presence of noise that occurs in reviews via spamming, such noise will increase the cost of processing information from the product market. There is strong evidence that spamming is a widespread problem on multiple e-commerce platforms (e.g. Jindahl and Liu, 2007, 2008; Luca and Zervas, 2016).[24] Spam detection is likely to be costly, as it requires substantial computation to conduct review-, reviewer-, and other product-related analysis. Additionally, it is difficult to ensure the external validity of any detection algorithm as it is nearly impossible to confirm that certain reviews are spam. I analyze Amazon reviews and measure their spamicity to show evidence of the prevalence of spamming and the potential cost of spam detection.[25]

[24] See, for example, https://digiday.com/marketing/vendors-ask-go-around-policy-confessions-top-ranked-amazon-review-writer/.

[25] The existence of spam reviews, such as near-duplicate reviews, may also imply that the cost of spam detection is consequential. As it is critical for e-commerce platforms to establish the credibility of their review data, they may have the greatest incentive to remove spam reviews. However, this incentive may not be sufficient if such activities

Despite the noise that dilutes information derived from the product market and the cost of spam detection, the capital market may attain informational efficiency if there are sophisticated investors with superior information-processing capabilities. With the abundance of big data and increased computing power, there is evidence that big data generates efficiency gains in the capital market (Zhu, 2018). I therefore hypothesize that the presence of such sophisticated investors helps reduce the noise that dilutes information derived from the product market. To test this hypothesis, I subdivide the firms in my sample based on the prevalence of such investors.

The presence of sophisticated investors with superior information-processing capabilities may not in itself, however, be sufficient to fully incorporate information derived from the product market into the market price. If the cost of processing information, such as that involved in spam detection, is non-negligible, those investors are unlikely to reveal all their information (Grossman and Stiglitz, 1980). Thus, the information reflected in spamming is expected to be incorporated into the capital market over time, albeit with delays.

1.4 Review Spam Detection

1.4.1 Near-Duplicate Reviews

I measure spamicity in consumer opinions to test the above hypotheses. Following Jindhal and Liu (2007, 2008), I define spam reviews as those that are intentionally crafted to induce consumers to act against their own best interests. One empirical challenge in spam detection is that it is almost impossible for humans to detect spam reviews (Ott, Choi, Cardie, and Hancock 2011). Thus, measuring spamicity usually relies either on 1) finding or generating a small subsample (training sample) of reviews that are likely to be spam and developing a classification mechanism that assigns probability estimates on reviews based on that training sample, or 2) examining the distributional irregularities in the entire sample. The former method is related to supervised

are expected to harm online sales or review activities. This also will add to the cost to investors of processing consumer opinions.

machine learning while the latter is related to unsupervised machine learning.[26] Although each method has its merits, the latter is difficult to apply to big data because of the substantial computational burden. As such, in this study I rely on supervised machine learning to measure spamicity.

I create a training sample using near-duplicate reviews identified in the review sample by closely following Jindahl and Liu (2007, 2008).[27] Near-duplicate reviews have identical or nearly identical content in the review texts written 1) by different reviewers and/or 2) of different products.[28] Using these near-duplicate reviews as positive examples of review spamming, I develop a classification model and estimate the model using logistic regression to assign spamicity estimates.

Near-duplicate reviews are highly suspected of being spam, for three reasons. First, because spam reviewers are likely to deal with multiple spamming tasks, in some cases repeatedly in relation to multiple products, it is likely to be difficult for them to create completely nonidentical reviews every time. Second, because it may be difficult for manufacturers and vendors to come up with spam reviews frequently, it is widely suspected that they often hire anonymous spam reviewers.[29] To ensure that spam review content is consistent and favorable, those writers are expected to receive similar guidelines that result in their producing similar reviews. Third, unless consumers or e-commerce platforms compare all reviews written at different times possibly by different reviewers, it will be hard to detect such spam reviews, especially when spam reviewers change the content slightly.[30]

---

[26] When conducting supervised machine learning, one should create a training sample by labeling positive examples (with the value of 1) and negative examples (with the value of 0) for regression analysis.
[27] They find that approximately 1% of the reviews in their sample from Amazon have nearly identical review pairs.
[28] Near-duplicate reviews written by the same reviewers and on the same products are discarded from the analysis as they can occur if a reviewer accidently posts the same review multiple times.
[29] Also see, for example, https://blogs.wsj.com/wallet/2009/07/09/delonghis-strange-brew-tracking-down-fake-amazon-raves/.
[30] It is possible for duplicates or near-duplicates of reviews to occur independently. For example, short phrases such as "This is fantastic" can appear multiple times without being posted through spamming. Also, consumers may copy and paste from their own previous reviews for multiple purchases. To control for these instances, I exclude from the training sample very short reviews (of fewer than 10 words and/or fewer than 50 characters) and exclude reviews written by the same reviewers on the same product. Although there remain some honest reviews among them, most of the near-duplicates in the training sample are likely to be spam. Jindahl and Liu (2008) find that 97% of their

[Insert Table 1.2 Here]

To identify near-duplicate reviews, I estimate the 2-gram Jaccard distance between review pairs and choose any pairs with similarity scores above 0.9 where the 1 indicates the distance between identical reviews (Jindahl and Liu 2008).[31, 32] There are 455,467 near-duplicate reviews out of 44,699,245 total reviews in the sample (1.02% to the total).[33] Table 1.2 displays examples of near-duplicate reviews, indicating that some reviews are written very quickly and can be written several months apart. Moreover, as they are written for separate and often dissimilar products, it is difficult for consumers to detect such reviews unless they systematically compare reviews of separate products. A review helpfulness indicator (*Helpful* in Table 1.2) seems not to be particularly informative in this regard.

Since near-duplicate reviews are likely be the easiest to detect, the existence of such reviews indicates that the cost of spam detection may be extremely high. To detect and remove such reviews, each review should be compared with all the other reviews in the database. This means that for the raw number of approximately 140 million reviews in Amazon posted from 2004 through 2014, including any duplicate reviews, there are 140 million x (140 million – 1)/2 review pairs to compare. Assuming there can be other types of hidden spam reviews in the sample that are even harder to detect, spam detection will be costly, and we expect to see significant spamming activity.

Figure 1.1 displays the average scores and the proportion of near-duplicate reviews in the review sample from 2004 through 2014. When a series of spam review scandals on Amazon

---

near-duplicate reviews are of products with slightly different characteristics, implying that they are highly likely to be spam.

[31] To measure the 2-gram Jaccard distance between two texts, one should first dissect each text into a set of adjacent two-word combinations. For example, for the text "I like this product," the set will be {"I-like," "like-this," and "this-product"}. Jaccard distance, then, is the number of elements in the intersection of the two sets over the number of elements in the union of the two sets while treating identical two-word combinations within a set independently.

[32] Since the dataset contains a huge number of reviews, estimating all n * (n – 1)/2 review pairs would be extremely difficult computationally. To ease the computational burden, I first sort reviews lexicographically and measure Jaccard distance among blocks of 1,000 neighboring reviews.

[33] See Figure 1.A1 for the distribution of the Jaccard distances between review pairs in the review sample.

occurred in 2004,[34] the company promised to resolve the problem. Consistent with this effort, the proportion of near-duplicate reviews in the sample decreased over time. Nevertheless, approximately 1% of the total number of reviews in the sample are near-duplicate reviews. Interestingly, the correlation between average scores and the proportion of near-duplicate reviews is -0.904, suggesting that spamming is more prevalent for products with high review scores.

[Insert Figure 1.1 Here]

1.4.2 Spam Classification Model

Using these 435,456 near-duplicate reviews as positive examples of spamming, I construct a training sample for supervised machine learning. Insofar as I use spam detection in this paper to measure the strength of the intention to affect quality information and thus to alter consumer behavior, negative examples of spamming are chosen from among the reviews that are less likely to be spam reviews. As such, I identify reviews that do not deviate from the average scores by +/−0.1 when those reviews were written.[35] I choose the range of +/−0.1 because the average review scores displayed for consumers reflect the use of a 0.1-point scale. The overall empirical results are robust across various range choices.[36] To further eliminate possible selection bias, I randomly choose 10% of such reviews and use them as negative examples of spamming, resulting in 446,842 such reviews.

Based on the training sample of 882,298 reviews (comprising 435,456 positive and 446,842 negative examples), I estimate the following classification model using logistic regression, following Jindhal and Liu (2008) in using only the information available in the review sample without reference to firm-level characteristics:

$$\textit{Near-Duplicate Review Dummy}_{i,j,k,t} = \alpha + \beta\, \textit{Review}_{i,t} + \gamma\, \textit{Reviewer}_{j,t} + \delta\, \textit{Product}_{k,t} \qquad [1]$$

where $\textit{Near-Duplicate Review Dummy}_{i,j,k,t}$ is a binary variable that takes the value of 1 for near-duplicate reviews and 0 otherwise for each review i posted by reviewer j of product k on date t;

[34] See for example, https://www.nytimes.com/2004/02/14/us/amazon-glitch-unmasks-war-of-reviewers.html.
[35] The average review score calculated in the sample may differ from the actual review score displayed when the reviewer posts the review.
[36] See Table 1.A4.

$Review_{i,t}$, $Reviewer_{i,t}$, $Product_{k,t}$ indicate review-, reviewer- and product-centric features on date t when review i was written, respectively. Detailed definitions of all variables are presented in the appendix I.

[Insert Table 1.3 Here]

Model [1] is measured by the following ten-fold cross-validation method (Jindahl and Liu, 2008). First, the training sample is randomly divided into ten groups. Second, I select one of the ten groups and then use the nine remaining groups to estimate model [1] using logistic regression (model training). Third, the estimated coefficients from the logistic regression are used to extrapolate the probability estimate (spamicity) for the initially selected group. Fourth, I carry out the same process nine more times by selecting another one of the ten groups each time. This method is widely used to avoid overfitting of a model in a training sample and to check the external validity of supervised machine learning. As a result, each group is used nine times for model training and one time for spam estimation. I use the average coefficients from those ten logistic regressions to estimate out-of-the-training sample spamicity. In Table 1.3, I present one of the ten logistic regressions.[37]

Spam reviewers may write spam reviews repeatedly, and therefore the reviewer- and product-level features reported in Table 1.3 might be the most informative data that can be derived from the reviews. Table 1.3 displays that reviews written by reviewers who tend to write more first or only reviews of multiple products (Percent of First Reviews and Percent of Only Reviews, respectively), more high-score and fewer dispersed reviews (Avg. Score [Reviewer] and Stddev. of Scores [Reviewer], respectively), and more reviews that deviate from the average scores (% of Negative Dev. and % of Positive Dev.), exhibit higher spamicity. Also, in Table 1.3, I list reviews of products that receive higher and more volatile scores (Avg. Score [Product] and Stddev. of Scores [Product], respectively) as well as products that attract fewer reviews and exhibit wider

[37] The coefficients in each regression are similar to each other in both magnitude and statistical significance, and thus only one of the ten results is reported.

fluctuations in the number of reviews (Logged # of Product Reviews and Logged Stddev. of # of Reviews, respectively), which tend to result in higher spamicity.

The role of classification model [1] is to identify features that distinguish positive examples from negative examples. The more important question is whether applying model [1] to the training sample makes it possible to predict positive examples that are posted out of the sample. One benefit of the ten-fold cross-validation method is that it enables me to evaluate the out-of-sample predictability of a model. In each logistic regression, where only nine of the ten groups are used, one can verify whether the extrapolation from the estimates can be used to distinguish the positive and negative examples in the remaining group.

The most widely accepted metric for evaluating out-of-sample predictability in supervised machine learning is the Area Under Curve (AUC) of a Receiver Operator Characteristic (ROC) curve. An ROC curve measures the proportion of true-positive rates and false-positive rates from a classification under a given probability threshold. The AUC measures the out-of-sample predictive power of a classification model. Statistically, it indicates the probability that a randomly sampled positive example will be ranked higher in probability than a randomly sampled negative example. An AUC of 1 indicates that a classification model can perfectly predict out-of-sample positive examples without type 1 errors and an AUC of 0.5 indicates that a classification model's performance is no better than that of a random assignment. The average AUC of the out-of-sample prediction is 0.855. This shows that when I apply model [1] to the training sample it demonstrates out-of-sample predictive capacity.[38]

To further establish the external validity of the spamicity estimates derived from model [1], I conduct a lift-curve analysis. The AUC of 0.855 shows that the model can predict near-duplicate reviews in the training sample. However, as there are no such reviews outside of the

[38] See Figures 1.A2 and 1.A3 for detailed depictions of ROC, AUC, and the model's performance. In medical or cybersecurity studies, an AUC of 0.99 or higher is desired as the accuracy of a model is critical for predictive diagnosis. However, because the goal of classification model [1] is not about detecting near-duplicate reviews but about detecting reviews that share characteristics of near-duplicate reviews in the training sample, the AUC of 0.855 is acceptable. The AUC of the similar classification model reported in Jindahl and Liu (2008) is 0.78. The AUCs of model [1] when using only either review-centric, reviewer-centric, or product-centric features are 0.7862, 0.7255, and 0.7212, respectively, indicating that all three features improve model performance (see Figure 1.A3).

training sample, it is unclear whether the spamicity estimates for the reviews outside of the training sample are reasonable. The lift curve, in this regard, shows whether the out-of-the-training-sample spamicity estimates are associated with spam reviews by showing how many reviews with high spamicity estimates are in a category that is expected to be associated with spamming. The results are shown in Figure 1.2.

[Insert Figure 1.2 Here]

To obtain the results I report in Figure 1.2, I randomly divide the out-of-the-training-sample reviews into ten groups and count the number of reviews that are within the top #% of spamicity in each group and fall into either one of four categories: [+] *Deviation (Good Product)*, [–] *Deviation (Good Product)*, [+] *Deviation (Bad Product)*, or [–] *Deviation (Bad Product)*. [+]/[–] *Deviation* indicates reviews that have scores higher/lower than an average score when the review was posted. *Good/(Bad) Product* indicates products that have average scores of 4 or higher/(2 or lower) when each review was posted. Therefore, the lift curves displayed in Figure 1.2 show the proportion of reviews with the top #% of spamicity estimates in each category. The reference curve indicates random assignments.

For example, in Figure 1.2 we can see that about 38.2% of reviews that received scores higher than the average scores that are also lower than 2 constitute the top 10% of the reviews with the highest rates of spamicity. Insofar as spamming is intended to alter consumer opinions when quality information about a product is unfavorable, reviews that assign high scores to bad products ([+] *Deviation (Bad Product)*) and reviews that assign low scores to good products ([–] *Deviation (Good Product)*) are more likely to be related to spamming. Consistent with this hypothesis, in Figure 1.2 we see that such reviews exhibit high spamicity rates derived from the out-of-the-training-sample estimates of model [1].

[Insert Figure 1.3 Here]

1.5 Empirical Results

1.5.1 Prevalence of Opinion Spamming

Having established that the spamicity estimates derived from model [1] are reasonable and do reflect the review spamming, I analyze the economics behind spamming. Figure 1.3 displays the time trend for average spamicity in the sample. Overall average spamicity in the sample is 24.6%. We can see that, before 2007, average spamicity in the sample decreases along with the proportion of near-duplicate reviews. Since 2008, however, average spamicity has been increasing despite the downward trend in the proportion of near-duplicate reviews. Also, spamicity is higher among reviews assigning high scores (4 or higher) than those assigning low scores (2 or below). One possible explanation of this trend is that spamming is conducted primarily to conceal the poor quality of one's own products rather than attacking others' (competitors') products.

The time trend for spamicity shown in Figure 1.3 also indicates that spam reviewers may evolve over time. Even though the prevalence of apparent spam reviews (i.e. near-duplicate reviews) decreases throughout the sample period, possibly as a result of efforts to eliminate spam reviews, average spamicity increases slowly. Because it is more difficult to detect evolving spam reviews, the cost of detecting such reviews is expected to have risen as well.

[Insert Figure 1.4 Here]

Figure 1.4 shows average spamicity by product category. Together with Figure 1.3, it shows that spamming is prevalent across time and product categories.[39] Notably, the four product categories exhibiting the highest spamicity in Figure 1.4 ("Electronics," "Video Games," "Apps for Android," and "Baby") are the categories receiving the most positive per-product reviews (Table 1.1). Consumers are more likely to post reviews when there is a significant discrepancy between the score and perceived quality. Such discrepancies are expected to be the highest among the four abovementioned product categories. This also creates an incentive to spam as a means of counteracting unfavorable consumer opinions. This evidence suggests that spamming may undermine the role of the wisdom of the crowd, which ensures that accurate quality information is

[39] Some sources argue that less than 1% of reviews posted on Amazon are inauthentic. Given the number of near-duplicate reviews alone, however, this number seems to be significantly underestimated. See also an example https://www.nbcsandiego.com/news/local/Beware-of-Fake-Online-Product-Reviews-484770801.html.

reflected in the product market.

[Insert Table 1.4 Here]

1.5.2 Determinants of Opinion Spamming

Next, I analyze firm-level determinants of spamming. When there is information asymmetry between firms, consumers, and investors, firms that produce high-quality products will try to signal that quality to their consumers and investors (Milgrom and Roberts, 1986). It is unclear, however, how firms react when they know they are selling products of poor quality. It is reasonable to consider spamming as a tactic such firms might undertake, and if so it will increase the level of information asymmetry. To investigate this channel, I estimate the correlation between firm-level average spamicity and other characteristics, including scores.

In Table 1.4, I report average firm-level spamicity along with other firm-level characteristics used in the analyses and compare between firms exhibiting above-median spamicity with those exhibiting below-median spamicity. These results indicate that review characteristics in the high-spamicity and low-spamicity groups differ across average scores and abnormal scores, measured, following Huang (2018a), by each month's average scores subtracted by the previous-12-month average scores. Consistent with the hypothesis that firms engage in spamming to conceal unfavorable opinions, the high-spamicity group receives lower (negative) average abnormal scores than the other group. In addition, consistent with the evidence reported in Table 1.1 and Figure 1.4 that spamming should be more prevalent when consumers post more reviews as a reflection of quality discrepancy, the high-spamicity group has a higher per product review average (24.8= 6,903/278.3) than the other group (8.1= 667 / 82.9). Other than the review characteristics, the two groups have similar average overall firm characteristics.

[Insert Table 1.5 Here]

To investigate the determinants of spamming, I estimate the following panel regression model:

$$Spamicity_{k,t} = \alpha + \beta\ Review_{k,t} + \gamma\ Firm_{k,t} + \eta_k + \theta_t \qquad [2]$$

where $Spamicity_{k,t}$ indicates the firm-level spamicity of new reviews for each firm k in month t; $Review_{k,t}$ indicates each firm k's monthly review characteristics, such as abnormal scores, standard deviations of scores, and abnormal peer scores, which are monthly average scores for products that are included in the "Also Viewed"[40] category on Amazon subtracted by their previous-12-month average scores; $Firm_{k,t}$ indicates, closely following Huang (2018a), each firm k's characteristics in month t; *Advertising* is defined by advertising expenses divided by total sales, *R&D*, logged market capitalization (*Logged Market Cap.*), *M/B*, *Gross Profitability* (Novy-Marx, 2013), *F-score* (Piotroski, 2000), logged dollar trading volume (*Logged Dollar Volume*: Brennan et al., 1998), the coefficient of the variation in dollar trading volume over the previous 12 months (*CV of Dollar Volume*: Chordia et al., 2001), monthly stock returns (*Stock Return*), the governance index (*GOV*) and the ESG index (*ESG*); η_k and θ_t indicate firm and time fixed effects, respectively.

In Table 1.5, I report the estimation results for model [2]. Throughout the remainder of the paper, I use abnormal scores as a quality indicator to capture new information embedded in review scores in each month (as in Huang, 2018a). In columns (1) through (4), I report the results of investigating the correlation between firm-level spamicity and review characteristics. As can be seen in column (1), abnormal reviews are negatively correlated with spamicity, indicating that new consumer opinions that negatively impact average scores are likely to increase the incentive to spam. A one-standard-deviation decrease in abnormal scores is associated with a 1.1% (= 0.014 x 0.78) increase in spamicity, a sizable increase given that it is 4.8% of the average spamicity (22.8%). Also, spamicity is positively correlated with score volatility (as seen in column (2)), abnormal peer scores (as seen in column (3)), and past spamicity (as seen in column (4)).

To obtain the results reported in column (5), I include firm characteristics as I investigate whether a firm's operational and financial incentives affect spamming behavior. These results imply that advertising expenses and size are positively and significantly correlated with spamicity

[40] Multiple products marketed by a single manufacturer are excluded when measuring the potential competitors' review scores. In this regard, "Also bought" and "Bought Together" products are discarded as they can be complementary goods. "Buy After Viewing" is not used either, as the number of reviews with this information in the review sample is significantly low.

while M/B and Gross profitability are negatively and significantly correlated with spamicity. The results pertaining to Advertising expenses and Gross profitability reported in columns (6) through (9), where review characteristics are included as well, are significant. A one-standard-deviation increase in advertising expenses is associated with a 6-basis-point (= 0.012 x 0.05) increase in spamicity, and a one-standard-deviation decrease in Gross Profitability is associated with a 60-basis-point (= 0.028 x 0.20) increase in spamicity. This suggests that firms experiencing poor performance in the product market may have a stronger incentive to engage in spamming.

To show further evidence that spamming decisions are determined largely by unfavorable opinions, I use the abnormal peer score as a possible instrument for the abnormal score of a product. Abnormally positive opinions of competing products will negatively impact positive opinions of the product of interest (inclusion restriction) and they are expected to be orthogonal to any observable firm characteristics (exclusion restriction). The results reported in column (7) indicate that, using abnormal peer scores as an instrument for abnormal scores, a one-standard-deviation decrease in abnormal scores increases firm spamicity by 47 basis points (= 0.006 x 0.78), which is approximately 2.1% of the average spamicity (22.8%).[41]

Additionally, I investigate whether spamicity decisions are correlated with firm governance. Firms that suffer from poor governance may not allocate their quality control resources optimally and thus may not be able to produce high-quality products. The results reported in columns (8) and (9) show, however, that the correlations between firm-level governance indices and spamicity are positive and significant. These results suggest that spamicity has few implications for poor firm governance, and that spamming decisions are more likely to be made with the product market in mind.

[Insert Table 1.6 Here]

1.5.3 Investment Value of Opinion Spam Detection

[41] See Table 1.A1 for first-stage results. The f-statistic from the first stage is 84.51. Consistent with the inclusion restriction, the coefficient of abnormal peer scores on abnormal scores is 0.74. This indicates that when a given product's peers receive abnormally positive scores, the product receives relatively low scores (sensitivity below 1).

The results reported in Table 1.5 suggest that spamming is correlated with consumer opinions. As spamming is likely to be conducted to compensate for negative opinions, the rate of spamming may covey information on poor product quality and add noise to review scores. This suggests the presence of significant investment value in spam detection. To show such evidence, I follow Huang (2018a) and form tercile portfolios for each month using equal weights or the weight of the number of total reviews (review weight) and investigate their performance using the Fama-French-Carhart four-factor model. In Table 1.6, I report the performance of the tercile portfolios based on either firm-level abnormal scores or spamicity.

To ensure the credibility of the main sample and to verify the results reported in Huang (2018a), I first replicate the main results of Huang (2018a) in Panel A. Consistent with those results, I find that a spread portfolio that goes long on stocks with high abnormal scores and short on stocks with low abnormal scores earns abnormal returns of 69.8—72.6 basis points per month. This shows that, although the scores are publicly available and relatively easy to process using big data analysis, there is significant investment value in consumer opinions.

I then repeat this analysis using spamicity. Instead of labeling each review as spam or non-spam using subjective thresholds, I use the relative magnitude of spamicity for portfolio construction to show the investment value of spam detection. The results I report in Panel B of Table 1.6 show that there are comparable abnormal returns when portfolios are constructed based on spamicity. As in Panel A, here a spread portfolio that goes long on stocks with low spamicity and short on those with high spamicity earns abnormal returns of 59.1—63.9 basis points per month. These returns derive from the high-spamicity portfolio where the abnormal returns are significantly negative. The overall evidence suggests that 1) high spamicity is negatively correlated with stock returns and 2) the information content of abnormal scores and spamicity may overlap, which is consistent with the negative correlation between the two shown in Table 1.5.

[Insert Table 1.7 Here]

The equivalent performance of spread portfolios based on abnormal scores and those based

on spamicity suggests that the information in each may overlap significantly. To identify the incremental information conveyed by spamicity, I form tercile spamicity portfolios conditional on each abnormal score tercile (Novy-Marx, 2013). In Table 1.7, I report Fama-French-Carhart four-factor-model abnormal returns of the tercile spamicity portfolios sorted within each abnormal score tercile. The monotone decrease in abnormal returns across spamicity terciles in each abnormal score tercile seen in Table 1.7 implies that spamicity contains additional information that goes beyond the information embedded in abnormal scores. The difference in abnormal returns between the low- and high-spamicity terciles is the highest in the mid-tercile of abnormal scores, where the signal from the abnormal scores is mixed to the greatest extent. Given the magnitude of the abnormal returns on long/short portfolios based on spamicity, the results reported in Table 1.7 suggest that there is significant incremental information embedded in spamicity.

[Insert Table 1.8 Here]

To further verify the incremental information embedded in spamicity and its impact on the information embedded in review scores, I run the following Fama-MacBeth regressions of one-month-ahead excess returns on abnormal scores and spamicity:

$$\textit{Excess Return}_{k,t+1} = \alpha + \beta_1 \textit{ Abnormal Score}_{k,t} + \beta_2 \textit{ Spamicity}_{k,t}$$

$$+ \beta_3 \textit{ Abnormal Score}_{k,t} \text{ x } \textit{Spamicity}_{k,t} + \gamma \textit{ Firm}_{k,t} + \eta_k + \theta_t \qquad [3]$$

where $\textit{Excess Return}_{k,t+1}$ indicates one-month-ahead excess stock returns on firm k between month t *and* $t+1$; $\textit{Abnormal Score}_{k,t}$ and $\textit{Spamicity}_{k,t}$ indicate firm-level monthly average abnormal scores and spamicity of the new reviews in each month, respectively. $\textit{Firm}_{k,t}$ indicates firm-level characteristics as in model [2]. η_k and θ_t indicate firm and time fixed effects, respectively. If spamicity has a predictive power for stock returns, I expect to find significant and negative β_2 and β_3 in [3].

The results reported in Table 1.8 imply that spamicity does have predictive power for stock returns. The results reported in columns (2) and (4) imply that spamicity is negatively correlated with excess stock returns. For instance, the one-standard-deviation increase in spamicity seen in

column (5) is associated with a 26.4-basis-point (= -2.640 x 0.1) decrease in one-month-ahead excess stock returns. This effect is amplified if abnormal scores are correlated with spamicity. In untabulated results, I find no significant reversals of the results reported in column (5) for excess returns at $t+2$, $t+3$, and $t+4$.

As can be seen in columns (3) and (6) of Table 1.8, β_3 in model [3] is statistically and economically significant, suggesting that spamicity adds noise to signals from review scores. A one-standard-deviation increase in spamicity for a given abnormal score is associated with a -1.628% change in one-month-ahead excess stock returns. Additionally, I find suggestive evidence that the return predictability of spamicity indicated in the results reported in column (6) is associated with the amount of information that firms release. Comparing the results reported in columns (7) and (8), we see that the return predictability of spamicity is much lower in magnitude and significance in a month when a firm announces its earnings. This finding is consistent with the hypothesis that information from the product market signals firms' performance in a way that is not fully captured by the capital market. The overall findings suggest that spamicity contains incremental information that can benefit investors.

[Insert Table 1.9 Here]

To demonstrate the investment value of spam detection in greater detail, I construct double-sorted portfolios on both abnormal scores and spamicity and form a plausible trading strategy using both abnormal scores and spamicity. In Table 1.9, I report the performance of the double-sorted portfolios. The results reported in Panel A of Table 1.9 imply that the performance of the spread portfolio reported in Panel A of Table 1.6 is worst with respect to stocks with low spamicity. Meanwhile, the performance of the spread portfolio reported in Panel B of Table 1.6 is worst with respect to stocks with low abnormal scores.

To reflect the fact that high investment value comes from high abnormal scores and low-spamicity portfolios and low investment value comes from low abnormal scores and high-spamicity portfolios, I form a corner spread portfolio that goes long on stocks with high abnormal

scores and low spamicity and short on those with low abnormal scores and high spamicity as a possible trading strategy based on both scores and spamicity. I report the results in Panel B. The abnormal returns on the abovementioned portfolios range from 1.173–1.234% per month, which is significant both statistically and economically.

In addition, I form a spread portfolio that goes long on stocks with high abnormal scores and spamicity and short on those with low abnormal scores and spamicity to show the value of spam detection. This is a possible worst-case scenario in which an investor cannot filter out spam reviews. This spread portfolio earns abnormal returns of between 18.3 and 22.6 basis points that are insignificant. The trading strategies associated with Tables 1.6, 1.7, and 1.9 do not incorporate trading costs, so the actual abnormal returns in this worst-case scenario are expected to be non-existent. The performance of this spread portfolio, as seen in the results reported in Panel B, shows that a trading strategy based on scores alone may not earn the significant abnormal returns reported in Panel A of Table 1.6 unless the noise in those scores reflected in spamicity is detected.

To shed additional light on the importance of information noise reflected in spamicity, I subdivide the firms associated with Table 1.9 based on the importance of consumer opinions, and thus on spamming incentives. As consumer opinions are valuable for consumers who purchase products online, firms that rely to a greater extent on e-commerce will value consumer opinions to a greater extent and thus will have stronger incentives to engage in spamming. Similarly, firms that introduce new products online more frequently will have stronger incentives to engage in spamming. Moreover, firms whose products bring higher prices will pay more attention to consumer opinions because consumers will be more cautious about making purchases; those firms will also have stronger incentives to engage in spamming.

If the source of the high abnormal returns generated by the corner spread portfolios whose performance is represented in Table 1.9 is associated with noise such as spamming, high abnormal returns will be concentrated among firms with stronger incentives to produce such noise. Firms that are willing to produce spam reviews may also become involved in other types of spamming such as phishing emails and misleading advertising. Spamicity, in this case, is assumed to reflect

the level of such noise even if it may not be the only source. The value of spam detection will come from partially, if not entirely, eliminating noise in consumer opinions.

[Insert Table 1.10 Here]

In Table 1.10, I report the abnormal returns on the corner spread portfolios whose performance is represented in Table 1.9, which I subdivide into the following groups: *E-commerce Sales*, *Number of New Products*, and *Average Product Price*. To calculate the extent of reliance on e-commerce, I first merge aggregate product-level e-commerce sales data from the United States Census Bureau ("Estimated Annual U.S. Retail Trade Sales–Total and E-commerce") into each firm's product based on their product categories. I then calculate the average of those amounts at the firm level (*E-commerce Sales*). *Average Product Price* is based on price information in the metadata. Consistent with the hypothesis that the source of high abnormal returns is noise in consumer opinions, the results reported in Table 1.10 show that abnormal returns are statistically significant only among firms that are expected to rely to a greater extent on e-commerce, introduce new products more frequently on Amazon, and whose average product price is high.

[Insert Table 1.11 Here]

1.5.4 Institutional Ownership, Earnings Surprises, and Opinion Spamming

In the final set of tests for this study, I investigate whether there are sophisticated investors with superior information-processing capabilities who can benefit from the investment value embedded in consumer opinions after removing noise. To do so, I estimate the following panel regression model:

$$\begin{aligned} \textit{Institutional Ownership}_{k,t} = \alpha + \beta_1\, \textit{Abnormal Score}_{k,t} + \beta_2\, \textit{Spamicity}_{k,t} \\ + \beta_3\, \textit{Abnormal Score}_{k,t} \times \textit{Spamicity}_{k,t} + \gamma\, \textit{Firm}_{k,t} + \eta_k + \theta_t \end{aligned} \qquad [4]$$

where $\textit{Institutional Ownership}_{k,t}$ indicates institutional ownership of firm k in quarter t; $\textit{Abnormal Score}_{k,t}$ and $\textit{Spamicity}_{k,t}$ indicate the firm-level quarterly average abnormal scores and spamicity of new reviews in each quarter, respectively. $\textit{Firm}_{k,t}$ indicates firm-level characteristics, as in

model [2]. η_k and θ_t indicate firm and time fixed effects, respectively. If institutional investors, as opposed to retail investors, have superior information processing capabilities and expertise in the field, I expect to find significant and negative β_3 in [4]. I further distinguish institutional investors by type into independent institutions and grey institutions, following Ferreira and Matos (2008).

In Table 1.11, I report the estimation results derived from model [4]. In columns (1) through (3) I show evidence that only β_3 has a negative and significant coefficient, suggesting that institutional investors benefit from the information embedded in consumer opinions in light of the prevalence of spamicity. Considering $Abnormal\ Score_{k,t}$ x $Spamicity_{k,t}$ in model [4] as a new metric for product quality, in column (3) I show that certain institutional investors can distinguish the noise from product market information. A one-standard-deviation increase in spamicity given an abnormal score will lower institutional ownership by 46 basis points (= -0.023 x 0.1–0.023 x 0.1) as seen in column (3) and lower the number of institutional investors by 3.15 (= +3.269 x 0.1–34.745 x 0.1) as seen in column (4). Given the economically sizable abnormal returns reported in Tables 1.6 and 1.9, however, this effect does not seem to be economically significant. This suggests that the cost of processing consumer opinions is non-negligible and that the benefit may not outweigh the cost for many investors.

In columns (5) and (6) I present further evidence of the existence of sophisticated investors with information-processing capabilities. As independent institutions are more likely to collect information pertinent to firm performance and product quality, β_3 is significant only among independent investors. In untabulated results, I find that foreign institutional ownership and domestic institutional ownership do not generate significant β_3 in combination with grey institutional ownership. Therefore, the results presented in column (3) suggest that not all investors adjust their portfolios based on noise they detect in consumer opinions.[42]

[Insert Table 1.12 Here]

[42] I also test additional implications of spamicity for firm governance using activist ownership based on institutions that file Form 13D. I find no evidence that lagged institutional ownership, including activist ownership, changes spamicity. Rather, I find evidence that activist institutions that hold more than 1% of a firm's shares also have negative and significant β_3.

I also examine whether analysts utilize this information noise when they forecast firm performance. To estimate the forecasting performance of analysts, I calculate the level of earnings surprises (*SUE*) by measuring the difference between the reported earnings per share (EPS) and the average of the most recent EPS forecasts issued by analysts before each earnings announcement. I estimate model [4] using firms' earnings surprises instead of institutional ownership as the dependent variable. I find evidence that analysts also partially distinguish the information noise embedded in abnormal scores and spamicity, which I report in Table 1.12. As can be seen in column (3) of Table 1.12, a one-standard-deviation increase in spamicity reduces the effects on earnings surprises of abnormal scores by more than 50% (= (0.032–0.162 x 0.1)/0.32). Also, the results reported in columns (4) and (5) show that those reported in column (3) are concentrated among firms with below-median analyst coverage.

[Insert Table 1.13 Here]

Taking the results reported in Tables 1.11 and 1.12 together, the evidence is consistent with the hypothesis that there are sophisticated investors who use the information embedded in consumer opinions after filtering out noise. I show further evidence that the abnormal returns of the corner spread portfolios reported in Table 1.9 are concentrated only among firms held by sophisticated investors. As such, I subdivide these firms into two groups based on institutional holdings or analyst coverage and repeat the tests that inform Table 1.10 to produce the results reported in Table 1.13. These results suggest that, without the presence of sophisticated investors, the cost of processing information to remove noise such as spamming is non-trivial, and thus there remains significant investment value in consumer opinions.

1.5.5 The Chinese Product Shock of 2012

The evidence reported in the previous sections implies that there are sophisticated investors with superior information-processing capabilities. This does not necessarily mean, however, that these investors acquire information from consumer opinions, let alone detect spamming, as the

information embedded in scores and spamicity may also be attained elsewhere. To shed light on whether observed abnormal returns and institutional holdings are associated with consumer opinions and spamming, I utilize a shock that disproportionately influenced firms' incentive to engage in spamming.

[Insert Figure 1.5 Here]

In the first quarter of 2012, Amazon launched what it called the "Global Selling Program" in China, allowing Chinese manufacturers to list their products on Amazon and sell their products directly to consumers without the use of importers.[43] This does not mean that these manufacturers could not sell their products in the U.S. Rather, the program allowed local Chinese manufacturers to easily list and sell their products globally on Amazon using their own brand names. As a result, the number of Chinese sellers (Chinese manufacturers) increased dramatically in the U.S.[44]

This effect is also captured in the review sample. The sample includes each product's universal barcodes, such as UPC or EAN information, where (as noted above) the first few digits of the codes indicate where the product was initially barcoded. I find that the ratio of Amazon products sold in the U.S. but barcoded in China to those barcoded in the U.S. increased by more than 111.4% (158.2%) between 2011 and 2012 (2013). Although there is no guarantee that the barcode location is the same as the manufacturing location, it is likely that products barcoded in China are also manufactured in China given the strong dominance of Chinese manufacturers in the global economy. It is also more likely that products barcoded in the U.S. are manufactured in China than vice versa. Thus, the above-reported estimates arguably form the lower bound of the true proportion of products manufactured in China. In Figure 1.5, I compare products barcoded in China and other countries with products barcoded in the U.S. around 2012, before and after the program was launched.

[Insert Figure 1.6 Here]

[43] See Amazon's IR website https://www.amazon.cn/gp/press/home/2012.
[44] See also https://alltechasia.com/amazon-china-report-chinese-overseas-sellers-have-grown-13-times-since-2013/.

The influx of such products increased competition on Amazon as Chinese manufacturers could produce cheaper substitute products using generic brand names regardless of their quality.[45] After dividing products into two categories based on the number of new products barcoded in China over those barcoded in the U.S. after 2012 ("China exposure"), I find that the average scores for products in categories that were subject to above-median exposure after the program launched experienced significant drops (see Figure 1.6).

[Insert Figure 1.7 Here]

The increased level of competition also increased the incentive to engage in spamming. There is a disproportionate increase in spamicity among firms whose products experienced above-median China exposure (Figure 1.7). I hypothesize that only sophisticated investors who invest in firms with higher exposure to the shock actively adjust their holdings against spamicity after 2012 as the noise in consumer opinions became more prominent. This is consistent with the hypothesis that only sophisticated investors are capable of analyzing the information embedded in consumer opinions reflected in spamicity in light of the cost involved. Because in this case the shock is Amazon-specific, the differential impact on institutional holdings across distinct types of investors indicates that spam detection has unique investment value for certain sophisticated investors.

To test this hypothesis, I estimate the following triple-difference model:

$$\begin{aligned} \textit{Institutional Ownership}_{k,t} = \alpha &+ \beta_1 \textit{ Abnormal Score}_{k,t} + \beta_2 \textit{ Spamicity}_{k,t} + \beta_3 \textit{ High Exposure}_{k,t} \\ &+ \beta_4 \textit{ Abnormal Score}_{k,t} \text{ x } \textit{Spamicity}_{k,t} \\ &+ \beta_5 \textit{ Abnormal Score}_{k,t} \text{ x } \textit{High Exposure}_{k,t} \\ &+ \beta_6 \textit{ Spamicity}_{k,t} \text{ x } \textit{High Exposure}_{k,t} \\ &+ \beta_7 \textit{ Abnormal Score}_{k,t} \text{ x } \textit{Spamicity}_{k,t} \text{ x } \textit{High Exposure}_{k,t} \\ &+ \gamma \textit{ Firm}_{k,t} + \eta_k + \theta_t \end{aligned} \qquad [5]$$

where *Institutional Ownership*$_{k,t}$ indicates firm *k*'s institutional ownership in quarter *t; High*

[45] Another potential problem is counterfeit products.

$Exposure_{k,t}$ indicates a binary variable for firm-level China exposure that takes the value of 1 if a firm's products generally fall into above-median product categories in terms of the number of products barcoded in China and 0 otherwise. $Firm_{k,t}$ indicates firm-level characteristics similar to [3]. η_k and θ_t indicate firm and time fixed effects, respectively. If sophisticated investors value consumer opinions and spam filtering, I expect to find significant and negative β_7, especially for sophisticated investors.

[Insert Table 1.14 Here]

In Table 1.14, I report the results derived from estimating model [5]. Regarding the results reported in columns (1) through (4), only those listed in (2) show a negative and significant β_7. This is consistent with the hypothesis that sophisticated investors are more responsive to China exposure. As can be seen in column (2), a one-standard-deviation additional increase in spamicity decreases the holdings of independent institutions by 57 basis points (= -0.057 x 0.01) after 2012. I further estimate the impact of the program on institutional ownership by estimating the coefficients included in [5] by year. Figure 1.8 displays the year-by-year estimates of β_7 in [5] using various types of institutional ownership. The results show that only independent institutions after the event year show significantly negative β_7.

[Insert Figure 1.8 Here]

1.5.6 Robustness

The overall empirical analyses are robust under varying specifications. To verify whether the observed abnormal returns from spread portfolios based both on abnormal scores and spamicity result from relevant information about a given product, I analyze the performance of the trading strategies associated with Table 1.9 beyond a month to check for possible return reversal.[46] The evidence shows that there is no clear sign of reversal although most of the abnormal returns disappear one month after portfolio formation. This result provides suggestive evidence that the

[46] See Table 1.A2.

results reported in Tables 1.6 and 1.9 tell us more about information than about temporary investor attention (Lou, 2014).

Next, I use multiple factor models to check the robustness of the abnormal returns reported in Tables 1.6 and 1.9. I find comparable abnormal returns using Fama-French 3- and 5-factor models and find that the overall results do not change significantly. For instance, abnormal returns from a spread portfolio that goes long on stocks with low spamicity and short on those with high spamicity yields a Fama-French-Carhart six-factor alpha (Carhart, 1997; Fama and French, 2015) of between 1.056% and 1.167% per month.[47]

Meanwhile, I check the overall robustness of applying classification model [1] to multiple training samples. Instead of randomly choosing negative examples of spam reviews that do not deviate from the average scores by +/−0.1, I choose several such ranges: reviews that do not deviate from the average 1) by +/−0.01, 2) by +/−0.05, or 3) by +/−0.5; I then include all the reviews. Overall, separate training samples based on randomly chosen negative examples as described above show consistent predictability except for the sample, where all the reviews that are not near-duplicate reviews are used as negative examples.[48] I find abnormal returns comparable to those reported in Tables 1.6 and 1.9 in the various training samples.[49]

Finally, I investigate whether there are differences between spam reviews with high scores (Positive spam) and those with low scores (Negative spam). Positive spam is designed to conceal unfavorable opinions or to proactively improve average scores. Meanwhile, negative spam that harms a given product's score presumably is posted by competitors. In light of the negative correlation between abnormal scores and spamicity that I report in Table 1.5, I expect a large portion of spam reviews to be positive. This potentially reflects the fact that producing spam reviews is costly. We see evidence of this also in Figure 1.3, which indicates that reviews with higher scores have higher average spamicity throughout the entire sample period. A firm having

[47] See Table 1.A3 for 6-factor model estimates.
[48] See Figure 1.A4 for AUCs from each classification model estimation
[49] See Table 1.A4.

limited resources with which to engage in spamming may find it preferable to write positive reviews for one's own goods than to write negative reviews about multiple competitors' goods.

Consistent with this hypothesis, I find evidence that the significant abnormal returns reported in Table 1.9 are concentrated among firms with high positive spamicity (spamicity in reviews with scores of 4 or above) *and* low negative spamicity (spamicity in reviews with scores of 2 or below) rather than among those with low positive spamicity *and* high negative spamicity.[50] This suggests that the return predictability of spamicity indicated by the results reported in Tables 1.7 and 1.8 is associated with firms' concealing unfavorable opinions, suggesting that spamming conveys information mostly about the negative quality perceptions of consumers.

Interestingly, I also find evidence that negative spamicity contains information about abnormal returns. When a firm's competing products receive high abnormal scores, a spread portfolio that goes long on stocks with high abnormal scores and spamicity and short on those with low abnormal scores and spamicity yields a Fama-French-Carhart six-factor alpha of -1.457% per month, indicating that negative spamicity also contains some information that is relevant to investors.[51] My findings show that spam review type conveys information that goes beyond that embedded in review scores.

1.6 Conclusion

Consumer opinions play a crucial role in the ever-growing world of e-commerce. They reduce information asymmetry between firms and consumers by providing indirect purchase experience. To the extent that consumer opinions influence purchase decisions, they are valuable for firms and their investors. When the wisdom of the crowd ensures that quality information is correctly represented in the product market, there is little cost in processing information from that market. Thus, information transmitted from the product market will be fully recognized by the capital market, assuming the usual conditions for capital market efficiency.

[50] See Panel A of Table 1.A5.
[51] See Panel B of Table 1.A5.

There is also, however, a motivation for firms to intervene in this process when consumer opinions are unfavorable. Such motives incentivize firms to introduce noise through various methods, including misleading advertising, fake news, and review spamming. Because it is difficult to detect these activities and eliminate them from the product market, capital market efficiency will be attained only if sophisticated investors with superior information-processing capabilities reveal this information via their activities. The cost of processing big data that reflects multiple dimensions is, however, non-trivial. Thus, there remains significant investment value in consumer opinions that are accompanied by noise.

In this paper, I first show the prevalence of spamming using supervised machine learning. The non-negligible existence of spam reviews alone strongly indicates that the cost of spam detection is very high. Then, I show that there is significant value in consumer opinions, possibly resulting from noise such as spamming. Even where there are sophisticated investors in the product market, as long as spam detection is costly the transmission of information from that market into the capital market may not be fully efficient. A potentially important direction that future research might take involves seeking ways to internalize this negative externality from spamming in both the product market and the capital market, which may yield valuable policy implications.

1.7 References

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns", *Journal of Financial Economics* 49 (3), 345-373.

Bickart, Barbara and Robert M. Schindler, 2001, "Internet forums as influential sources of consumer information", *Journal of Interactive Marketing* 15 (3), 31-40.

Böhme, Rainer and Thorston Holz, "The Effect of Stock Spam on Financial Markets", *working paper*, Available at: https://ssrn.com/abstract=897431

Archak, Nikolay, Anindya Ghose, Panagiotis G. Ipeirotis, 2011, "Deriving the Pricing Power of Product Features by Mining Consumer Reviews", *Management Science* 57 (8), 1485-1509.

Carhart, Mark M., 1997, "On Persistence in Mutual Fund Performance", *Journal of Finance* 52 (1), 57-82.

Chen, Yubo and Jinhong Xie, 2008, "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix", *Management Science* 54 (3), 477-491.

Chevalier, Judith A. and Dina Mayzlin, 2006, "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research* 43 (3), 345-354.

Clarke, Jonathan, Hailiang Chen, Ding Du, and Yu Jeffrey Hu, 2018, "Fake News, Investor Attention, and Market Reaction", *working paper*, Available at: https://ssrn.com/abstract=3213024

Cohen, Lauren and Andrea Frazzini, 2008, "Economic Links and Predictable Returns", *Journal of Finance* 63 (4), 1977-2011.

Chordia, Tarun, Avanidhar Subrahmanyam, V. Ravi Anshuman, 2001, "Trading activity and expected stock returns", *Journal of Financial Economics* 59 (1), 3-32.

DellaVigna, Stefano and Joshua M. Pollet, 2009, "Investor Inattention and Friday Earnings Announcements", *Journal of Finance* 64 (2), 709-749.

Diamond, Douglas W. and Robert E. Verrecchia, 1981, "Information aggregation in a noisy rational expectations economy", *Journal of Financial Economics* 9 (3), 221-235.

Dugast, Jérôme and Thierry Foucault, 2018, "Data abundance and asset price informativeness", *Journal of Financial Economics* 130 (2), 367-391.

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

Fama, Eugene F. and Kenneth R. French, 2015, "A five-factor asset pricing model", *Journal of Financial Economics* 116 (1), 1-22.

Ferreira, Miguel A. and Pedro Matos, 2008, "The colors of investors' money: The role of institutional investors around the world", *Journal of Financial Economics* 88 (3), 499-533.

Fornell, Claes, Forrest V. Morgeson III, and G. Tomas M. Hult, "Stock Returns on Customer Satisfaction Do Beat the Market: Gauging the Effect of a Marketing Intangible", *Journal of Marketing* 80 (5), 92-107.

Froot, Kenneth, Namho Kang, Gideon Ozik, and Ronnie Sadka, 2017, "What do measures of real-time corporate sales say about earnings surprises and post-announcement returns?", *Journal of Financial Economics* 125 (1), 143-162.

Godes, David and Dina Mayzlin, 2004, "Using Online Conversations to Study Word-of-Mouth Communication", Marketing Science 23 (4), 545-560.

Grossman, Sanford J. and Joseph Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets", *American Economic Review* 70 (3), 393-408.

Grullon, Gustavo, George Kanatas, and James P. Weston, 2004, "Advertising, Breadth of Ownership, and Liquidity", *Review of Financial Studies* 17 (2), 439-461.

Hamilton, William L., Kevin Clark, Jure Leskovec, and Dan Jurafsky, 2016, "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora", *Conference on Empirical Methods in Natural Language Processing*, 595–605.

Hirshleifer, David and Siew Hong Teoh, 2003, "Limited attention, information disclosure, and financial reporting", *Journal of Accounting and Economics* 36 (1-3), 337-386.

Hoberg, Gerard and Gordon M. Philips, 2018, "Text-Based Industry Momentum", *Journal of Financial and Quantitative Analysis, forthcoming*.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies", *Journal of Finance* 55 (1), 265-295.

Hong, Harrison and Jeremy C. Stein, 2002, "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets", *Journal of Finance* 54 (6), 2143-2184.

Hou, Kewei and Tobias J. Moskowitz, 2005, "Market Frictions, Price Delay, and the Cross-Section of Expected Returns", *Review of Financial Studies* 18 (3), 981-1020.

Huang, Jiekun, 2018a, "The customer knows best: The investment value of consumer opinions", *Journal of Financial Economics* 128 (1), 164-182.

Huang, Ruidi, 2018b, "The Financial Consequences of Customer Satisfaction: Evidence from Yelp Ratings and SBA Loans", *working paper*.

Ittner, Christopher D. and David F. Larcker, 1998, "Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction", *Journal of Accounting Research* 36 (1), 1-35.

Jindahl, Nitin and Bing Liu, 2007, "Review Spam Detection", *Proceedings of the 16th International Conference on World Wide Web*, 1189-1190.

Jindahl, Nitin and Bing Liu, 2008, "Opinion Spam and Analysis", *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 219-230.

Karolyi, G. Andrew, Dawoon Kim, and Rose Liao, 2018, "The Theory and Practice of Investor Relations: A Global Perspective", *working paper*.

Kogan, Shimon, Tobias J. Moskowitz, and Marina Niessner, "Fake News: Evidence from Financial Markets", *working paper*, Available at: https://ssrn.com/abstract=3237763

Kyle, Albert S., "Continuous Auctions and Insider Trading", *Econometrica* 53 (6), 1315-1335.

Larkin, Yelena, 2013, "Brand perception, cash flow stability, and financial policy", *Journal of Financial Economics* 110 (1), 232-253.

Lee, Charles M. C., Paul Ma, Charles C. Y. Wang, 2015, "Search-based peer firms: Aggregating investor perceptions through internet co-searches", *Journal of Financial Economics* 116 (2), 410-431.

Li, Xinxin and Lorin M. Hitt, 2008, "Self-Selection and Information Role of Online Product Reviews", *Information Systems Research* 19 (4), 456-474.

Lim, Ee-Peng, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady W. Lauw, "Detecting Product Review Spammers using Rating Behaviors", *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 939-948.

Liu, Yong, 2006, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue", *Journal of Marketing* 70 (3), 74-89.

Lou, Dong, 2014, "Attracting Investor Attention through Advertising", Review of Financial Studies 27 (6), 1797-1829.

Luca, Michael, 2011, "Reviews, reputation, and revenue: The case of Yelp.Com", *Harvard Business School NOM Unit Working Paper* 12-016, 1-40.

Luca, Michael and Georgios Zervas, 2016, "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud", *Management Science* 62 (12), 3412-3427.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier, 2014, "Promotional Reviews: An Empirical Investigation of Online Review Manipulation", *American Economic Review* 104 (8), 2421-2455.

McAuley, Julian, Christopher Targett, Qinfeng Shi, and Anton van den Hengel, 2015, "Image-Based Recommendations on Styles and Substitutes", *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 43-52.

McAuley, Julian, Rahul Pandey, and Jure Leskovec, 2015, "Inferring Networks of Substitutable and Complementary Products", *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

McAuley, Julian and Alex Yang, 2016, "Addressing Complex and Subjective Product-Related Queries with Customer Reviews", *Proceedings of the 25th International Conference on World Wide Web*, 625-635.

Merton, Robert C., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information", *Journal of Finance* 42 (3), 483-510.

Milgrom, Paul and John Roberts, 1986, "Price and Advertising Signals of Product Quality", *Journal of Political Economy* 94 (4), 796-821.

Mukherjee, Arjun, Bing Liu, and Natalie Glance, 2012, "Spotting Fake Reviewer Groups in Consumer Reviews", *Proceedings of the 21st international conference on World Wide Web*, 191-200.

Mukherjee, Arjun, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, Riddhiman Ghosh, 2013, *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 632-640.

Novy-Marx, Robert, 2013, "The other side of value: The gross profitability premium", *Journal of Financial Economics* 108 (1), 1-28.

Ott, Myle, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock, 2011, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination", *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 309-319.

Peng, Lin and Wei Xiong, 2006, "Investor attention, overconfidence and category learning", *Journal of Financial Economics* 80 (3), 563-602.

Piotroski, Joseph D., 2000, "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers", *Journal of Accounting Research* 38. 1-41.

Verrecchia, Robert E., 1982, "Information Acquisition in a Noisy Rational Expectations Economy", *Econometrica* 50 (6), 1415-1430.

Xie, Sihong, Guan Wang, Shuyang, and Philip S. Yu, 2012, "Review spam detection via time series pattern discovery", *Proceedings of the 21st International Conference on World Wide Web*, 635-636.

Zhu, Christina, 2018, "Big Data as a Governance Mechanism", *working paper*.

1.8 Figures and Tables

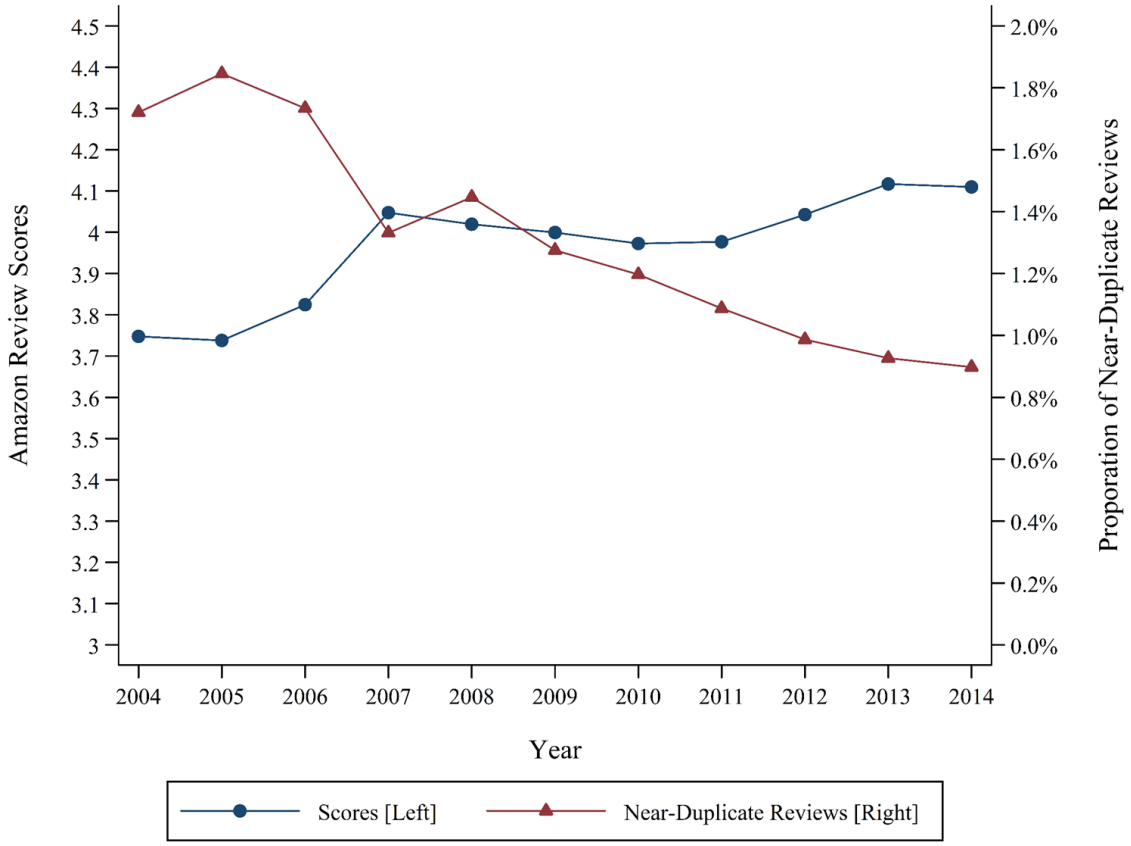


Figure 1.1. Average Review Scores and Near-Duplicate Reviews
This figure depicts the average Amazon review scores and the proportion of near-duplicate reviews from 2004 through 2014. *Near-Duplicate Review* indicates reviews with nearly identical review pairs in the sample (with similarity scores above 0.9 using 2-gram Jaccard distance) posted by distinct reviewers and/or of distinct products. *Proportion of Near-Duplicate Reviews* indicates the number of near-duplicate reviews divided by the total number of reviews in each year.

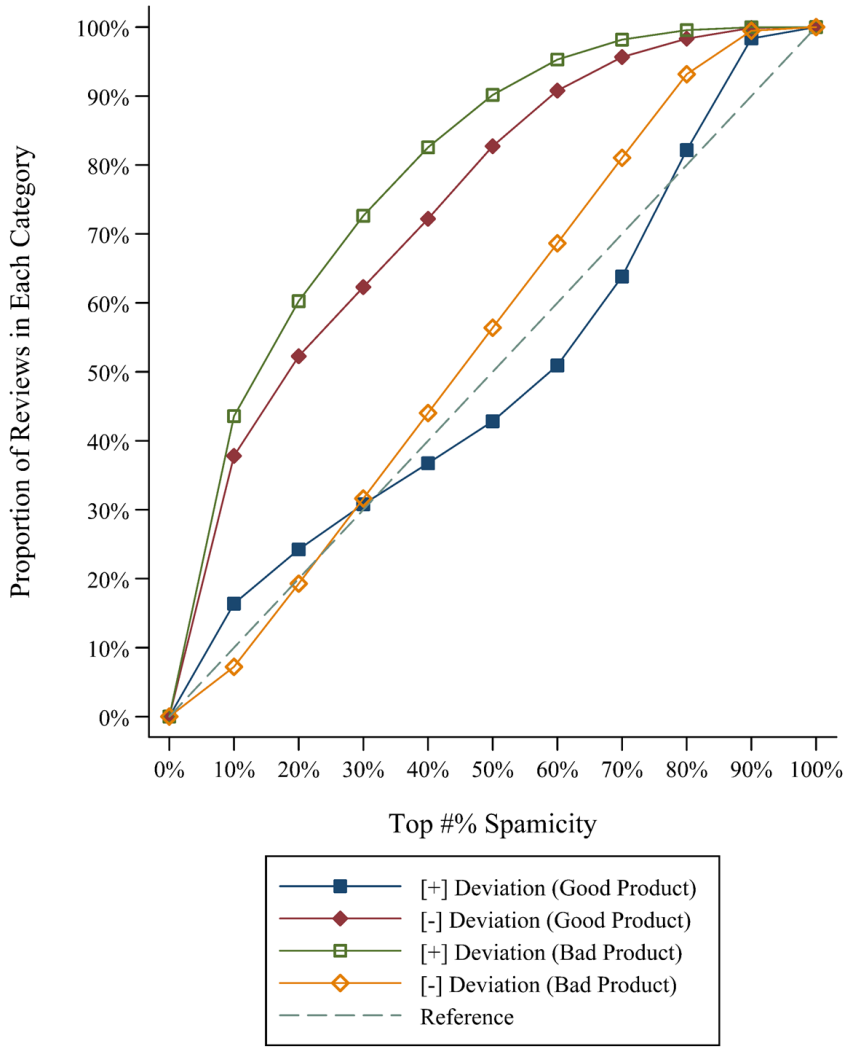


Figure 1.2. Lift Curves

This figure depicts the lift curves of the out-of-the-training-sample reviews that are not used to estimate spamicity in model [1]. The-out-of-the-training-sample reviews are first randomly divided into ten groups. *Top #% Spamicity* indicates reviews that are among the top #% in spamicity in each group. *Proportion of Reviews in Each Category* indicates the proportion of reviews in each group that fall into one of the following categories: [+] *Deviation (Good Product)*, [–] *Deviation (Good Product)*, [+] *Deviation (Bad Product)*, and [–] *Deviation (Bad Product)*. [+]/[–] *Deviation* indicates reviews that receive scores higher/lower than the average score when each review was posted. *Good/(Bad) Product* indicates products that receive an average score of 4 or higher/(2 or lower) when each review was posted.

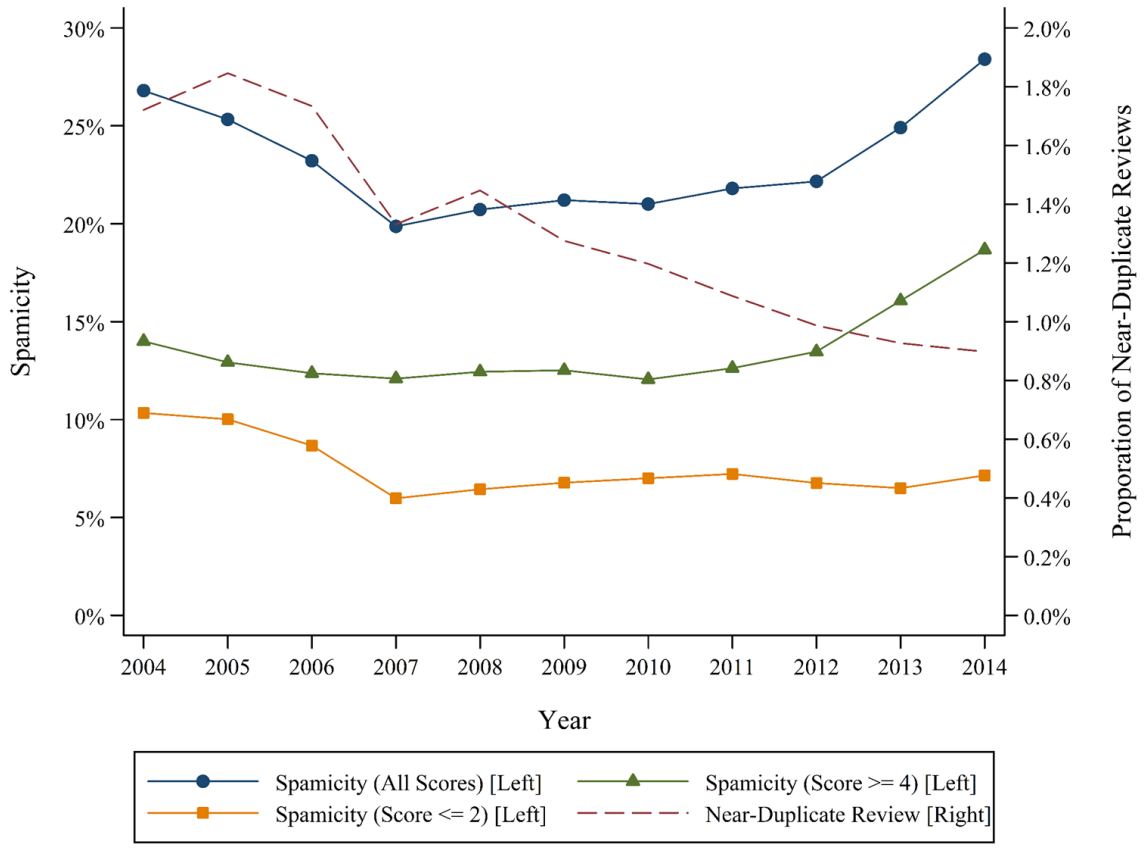


Figure 1.3. **Average Spamicity across Time**

This figure displays the estimated average spamicity (Spam Probability) of reviews from 2004 through 2014. Spamicity is estimated using *Near-Duplicate Reviews* as positive examples in the training sample. *All Scores* indicates average spamicity among all reviews. *Score >= 4* indicates average spamicity among reviews receiving review scores of 4 or higher. *Score <= 2* indicates average spamicity among reviews receiving review scores of 2 or lower.

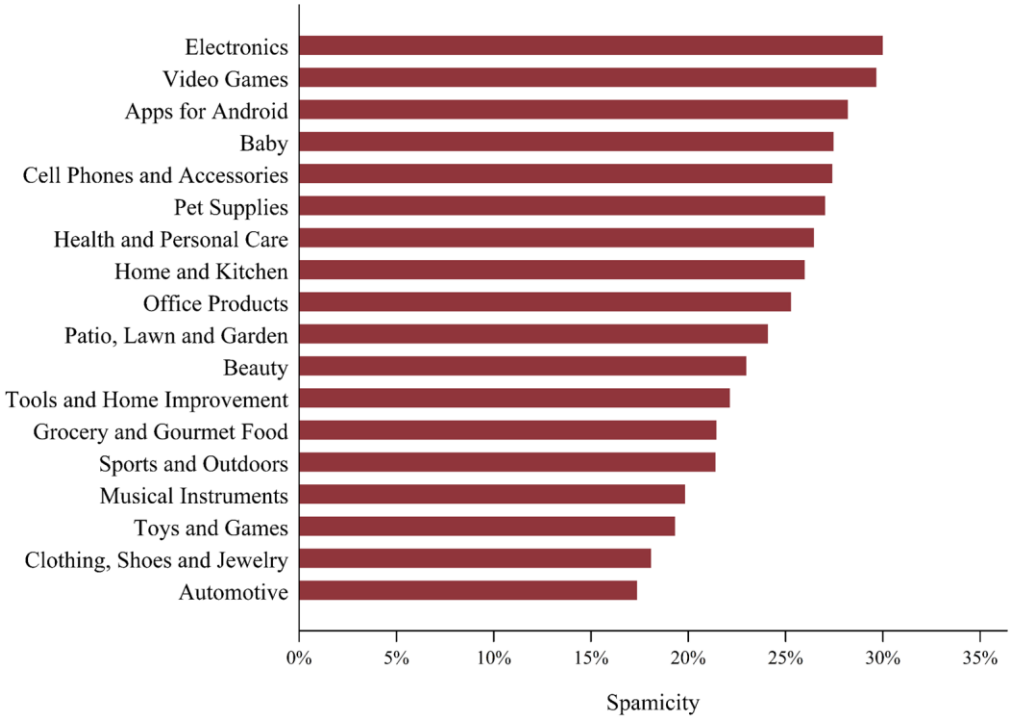


Figure 1.4. **Average Spamicity across Product Categories**
This figure shows the estimated average spamicity (Spam Probability) of sample reviews from 2004 through 2014 across Amazon product categories. Spamicity is estimated using *Near-Duplicate Reviews* as positive examples in the training sample.

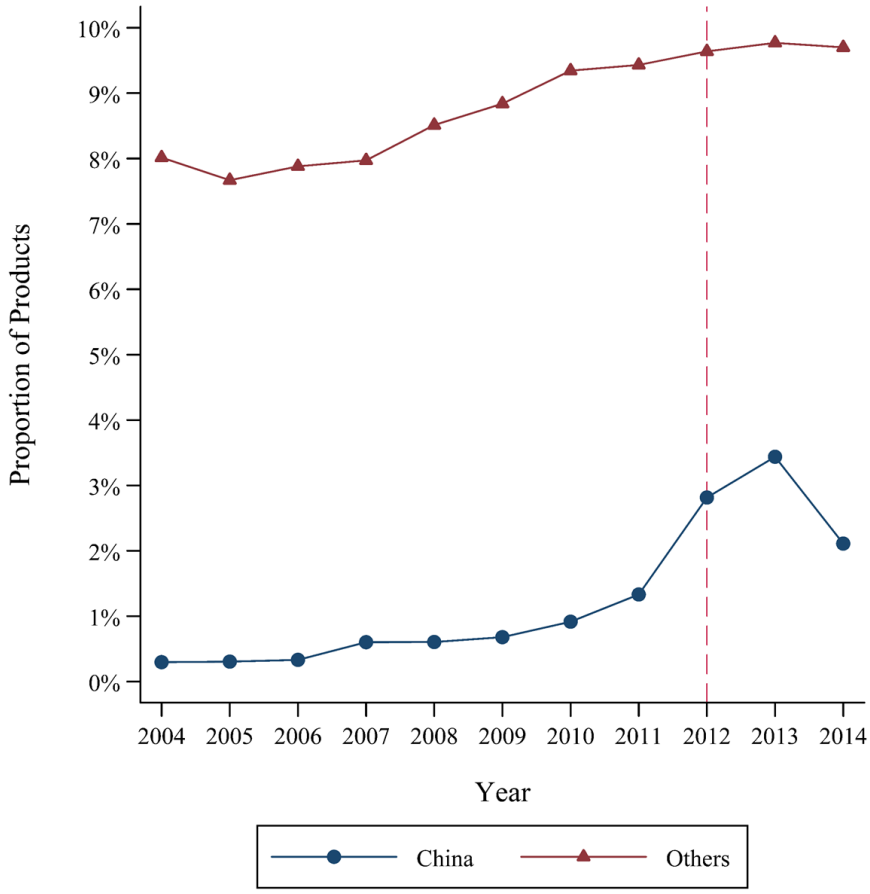


Figure 1.5. Proportion of Foreign Products on Amazon

This figure shows the proportion of new products available on Amazon that are barcoded outside of the U.S. *China* indicates the ratio of new products on Amazon that are barcoded in China to new products on Amazon that are barcoded in the U.S. as percentages. *Others* indicates new products on Amazon that are barcoded in countries other than China and the U.S. over new products on Amazon that are barcoded in the U.S. as percentages. The vertical dotted line indicates when the "Amazon Global Selling Program" was launched.

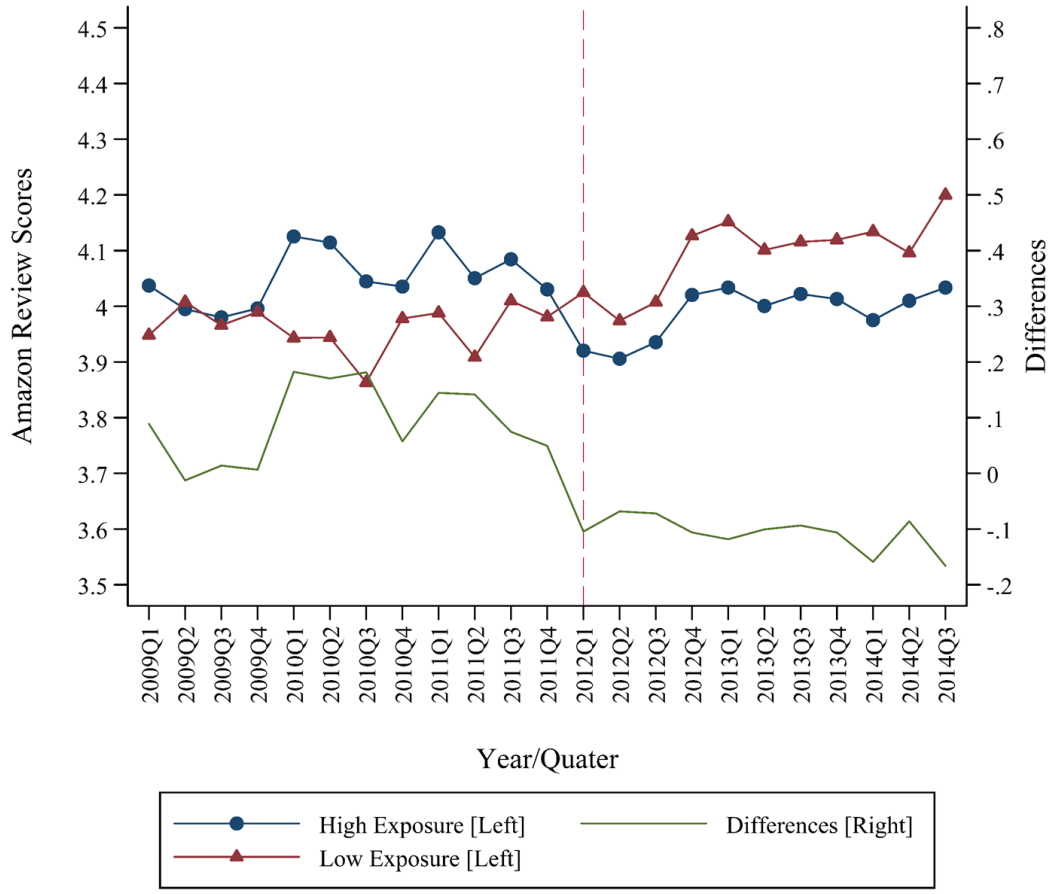


Figure 1.6. Average Review Scores around the Influx of Foreign Products

This figure shows average review scores across product categories with high and low exposure from the influx of products that are barcoded in China. *High (Low) exposure* indicates product categories with above- (below-)median numbers of new products barcoded in China after 2012. *Differences* indicates differences in scores between the *High* and *Low Exposure* groups. The vertical dotted line indicates when the "Global Selling Program" was launched.

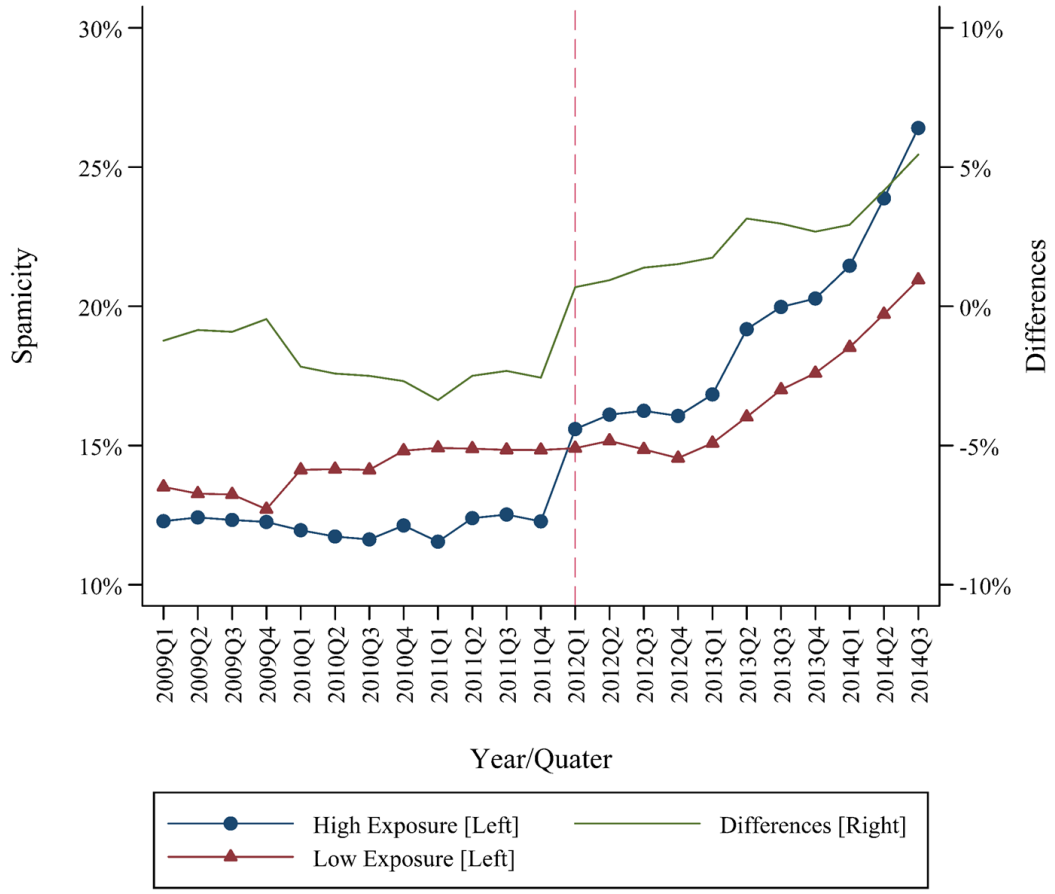


Figure 1.7. Average Spamicity around the Influx of Foreign Products

This figure shows average spamicity for firms whose products are sold in product categories with high and low exposure from the influx of products that are barcoded in China. *High (Low) exposure* indicates product categories with above- (below-)median numbers of new products barcoded in China after 2012. *Differences* indicates differences in spamicity between products in the *High* and *Low Exposure* groups. The vertical dotted line indicates when the "Global Selling Program" was launched.

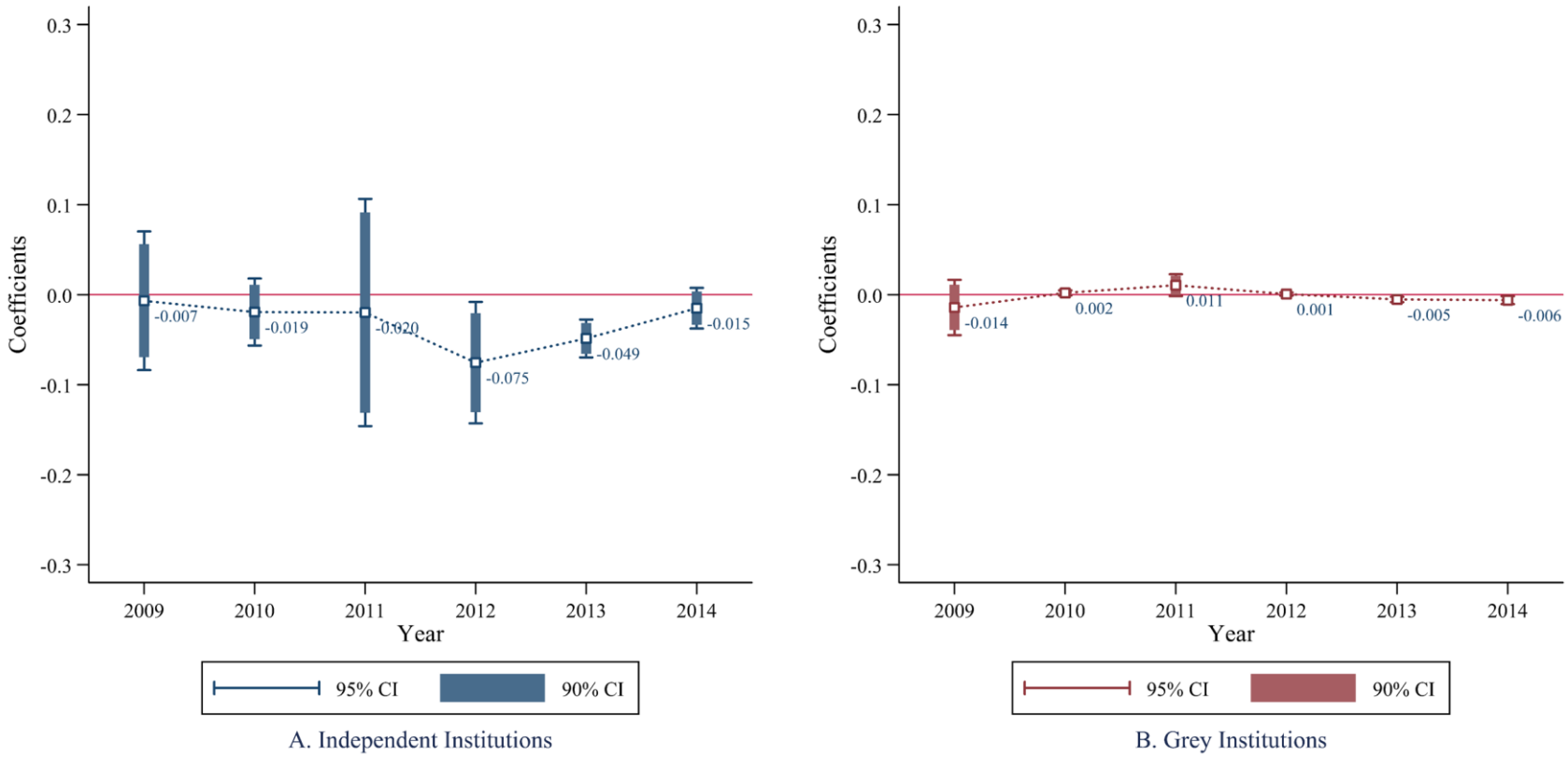


Figure 1.8. Proportion of Foreign Products available on Amazon

This figure shows year-by-year coefficient estimates of model [5]. *Coefficients* indicates coefficient estimates of β_7 in model [5]. The dependent variable in model [5] in A (B) is independent (grey) institutional ownership as defined by Ferreira and Matos (2008). "95% CI" and "90% CI" indicate 95% and 90% confidence intervals, respectively. The "Amazon Global Selling Program" was launched in 2012.

Table 1.1. Sample Products and Reviews

In this table, I report review characteristics by product category and year from 2004 through 2014. *Full Review sample* indicates the review sample used to estimate model [1] (44,699,245 reviews). *Main Sample* indicates products produced by Compustat firms included in the main analyses (3,827,524 reviews). *Category* indicates product category as defined by Amazon. # of Reviews equals the number of observations. Identical reviews by the same reviewer of the same products are excluded from the sample.

Panel A. By Year

Year	*Full Review sample*				*Main sample* (Compustat)				
	# of Reviews	Avg. Score	# of Products	# of Reviewers	# of Reviews	Avg. Score	# of Products	# of Reviewers	# of Firms
2004	113,515	3.75	30,356	82,974	30,964	3.70	6,158	25,323	158
2005	199,665	3.74	51,464	145,495	46,270	3.70	8,386	38,244	183
2006	281,739	3.82	81,232	208,063	58,356	3.81	10,614	49,831	209
2007	661,623	4.05	187,900	441,847	124,255	4.08	18,360	103,436	246
2008	875,679	4.02	254,261	561,313	148,292	4.01	23,091	122,467	254
2009	1,232,291	4.00	365,313	766,724	183,910	4.03	30,594	151,841	267
2010	1,921,431	3.97	558,099	1,142,758	238,491	3.98	40,385	195,472	276
2011	3,518,431	3.98	896,057	1,907,186	356,767	3.95	56,646	290,843	285
2012	6,795,520	4.04	1,463,547	3,249,284	552,374	4.04	78,927	438,104	294
2013	16,820,953	4.12	2,662,523	6,272,291	1,250,592	4.16	129,714	908,157	301
2014	12,278,398	4.11	2,295,506	4,975,458	837,253	4.16	109,029	631,109	301

Panel B. By Item Category

Category	*Full Review sample*				*Main sample* (Compustat)				
	# of Reviews	Avg. Score	# of Products	# of Reviewers	# of Reviews	Avg. Score	# of Products	# of Reviewers	# of Firms
Apps for Android	2,636,018	4.00	61,249	1,323,483	147,669	4.00	283	132,626	55
Automotive	1,369,462	4.19	319,670	849,811	54,934	4.27	13,051	45,170	115
Baby	902,334	4.12	64,308	524,599	86,906	4.09	4,140	75,318	56
Beauty	2,017,209	4.15	249,067	1,208,262	93,452	4.13	10,447	74,264	108
Cell Phones and Accessories	3,429,238	3.81	319,147	2,254,397	192,856	3.73	11,152	164,575	114
Clothing, Shoes and Jewelry	5,740,430	4.14	1,135,315	3,114,168	398,248	4.29	61,838	342,086	153
Electronics	7,653,271	4.01	470,933	4,146,052	1,320,925	4.10	42,428	1,032,383	188
Grocery and Gourmet Food	1,293,591	4.26	165,815	767,510	39,021	4.24	4,817	33,095	85
Health and Personal Care	2,966,935	4.11	251,883	1,845,694	149,240	4.17	8,521	120,523	164
Home and Kitchen	4,208,457	4.10	408,679	2,493,623	178,924	4.03	11,564	157,939	172
Musical Instruments	487,951	4.24	81,142	333,503	9,172	4.28	1,063	8,857	53
Office Products	1,222,650	3.98	129,238	897,129	279,192	3.90	12,384	214,625	141
Patio, Lawn and Garden	984,337	4.01	105,599	709,950	53,898	4.10	3,661	49,276	81
Pet Supplies	1,229,525	4.11	103,178	738,883	14,265	4.09	1,092	12,795	64
Sports and Outdoors	3,249,098	4.18	478,105	1,983,325	84,020	4.15	11,215	78,849	174
Tools and Home Improvement	1,901,547	4.13	259,417	1,202,443	116,605	4.17	10,647	101,206	154
Toys and Games	2,220,341	4.15	326,235	1,325,509	241,049	4.12	33,251	179,368	135
Video Games	1,186,851	3.98	48,738	764,597	367,148	3.98	9,000	277,179	53

Table 1.2. Near-Duplicate Review Examples

In this table, I report examples of near-duplicate reviews used as positive examples in model [1]. Near-duplicate reviews are defined as reviews that receive another review where the textual contents of the two reviews receive above 0.90 similarity scores calculated using 2-gram Jaccard distance. The training sample includes 435,456 near-duplicate reviews. *ReviewerID* indicates each review's ID and *ASIN* indicates the product identifier. *Helpful* [*A*, *B*] indicates how many, out of *B* consumer feedback responses to a given review, *A* consumers found the review helpful.

ReviewerID†	ASIN†	Helpful	Review	Date	Score
The Same Reviewer on Different Products					
A328AQ********	B004G*****	[1, 1]	I swear I never liked any of my shoes like this one - it feels so comfortable and delivers an excellent fashion statement. Great great buy. I used to be a loyal Nike fan, but this one made me change my party. My next shoe purchase, **very high chances will be another Puma**	3/20/2012	4
A328AQ********	B002F*****	[0, 0]	I swear I never liked any of my shoes like this one - it feels so comfortable and delivers an excellent fashion statement. Great great buy. I used to be a loyal Nike fan, but this one made me change my party. My next shoe purchase, **very high chances will be another adidas**	3/20/2012	4
Different Reviewers on the Same Product					
A1MS73********	B0047*****	[0, 0]	As other buyer says this is a great case, the materials are good quality and fits perfectly on the phone. The pen is very useful, works very well on the phone, maybe other buyers hasn't used it properly so they say that is useless. **Very recommended!**	5/31/2011	5
A3M6MI********	B0047*****	[0, 0]	As other buyer says this is a great case, the materials are good quality and fits perfectly on the phone. The pen is very useful, works very well on the phone, maybe other buyers hasn't used it properly so they say that is useless.	9/3/2011	5
Different Reviewers on Different products					
A1O7CE********	B00F2*****	[19, 25]	After using different weight-loss diets/plans with poor results in the past. I'm glad that I gave this a try. It is truly a 5 star review. I only need to take it twice in a day 30 minutes before a meal. I not only have lost 8 pounds already, but I have more energy than I had before.	3/26/2014	5
A3TFFM********	B00HD*****	[3, 3]	After using different weight-loss diets/plans with poor results in the past. I'm glad that I gave this a try. It is truly a 5 star review. I only need to take it twice in a day 30 minutes before a meal. I not only have lost 8 pounds already, but I have more energy than I had before.	6/5/2014	5
AL9A8W********	B00FS*****	[0, 0]	After using different weight-loss diets/plans with poor results in the past. I'm glad that i gave this a try. It is truly **worth** a 5 star review. I only need to take it twice in a day 30 minutes before a meal. I not only have lost 8 pounds already, but I have more energy than I had before.	4/16/2014	5
A2GT59********	B00G1*****	[0, 1]	After using different weight-loss diets/plans with poor results in the past. I'm glad that I gave this a try. It is truly a 5 star review. I only need to take it twice in a day 30 minutes before a meal. I not only have lost 8 pounds already, but I have more energy than I had before.	7/2/2014	5

† Not fully disclosed (concealed by asterisk)

Table 1.3. Spam Classification

In this table, I report the results of one of the ten logistic regressions carried out to estimate model [1] using ten-fold cross-validation. The dependent variable is *Near-Duplicate Review Dummy*, which takes the value of 1 for near-duplicate reviews and 0 otherwise. *Feedback* indicates feedback given in response to a review. *Helpful Feedback* indicates feedback that labels a review as helpful. *Title* indicates the title of a review. *Position* indicates the relative order of a review for a given product. *First (Only) Review Dummy* takes the value of 1 if a review is the first (only) review for a product and 0 otherwise. *Sent.* indicates aggregate individual review-level sentiment measured by sentiment lexicons as introduced by Hamilton et al. (2016). *Positive (Negative) Sent. Score* indicates the sum of the sentiment scores for words with positive (negative) sentiment. *% of Positive (Negative) Sent.* indicates the number of words with positive (negative) sentiment over the total number of words for a given review. *Score Dummy* takes the value of 1 if the score for a review is 4 or higher, -1 if the score is 2 or lower, and 0 otherwise. *Good and Bad Scores* takes the value of 1 if a reviewer has given both good scores (>= 4) and bad scores (<= 2) and 0 otherwise. *Good and Avg. Scores* takes the value of 1 if a reviewer has given both good scores (>= 4) and the middle score (= 3) and 0 otherwise. *Bad and Avg. Scores* takes the value of 1 if a reviewer has given both bad scores (<= 2) and the middle score (= 3) and 0 otherwise. *All Scores* takes the value of 1 if the reviewer has given scores of every kind and 0 otherwise. *% of Negative (Positive) Dev.* indicates the number of reviews that receive scores lower (higher) than the average scores when the review was posted divided by the total number of reviews posted by a reviewer. Standard errors are clustered at each product level. Numbers in the parentheses indicate t-statistics.

Dependent Variable: *Near-Duplicate Review Dummy*			
Review-centric Features			
Logged # of Feedback	-0.113***	Review Score	1.503***
	(-3.63)		(7.34)
Logged # of Helpful Feedback	0.287***	Score Dummy	-3.153***
	(11.13)		(-7.99)
% of Helpful Feedback	-0.236***	*Reviewer-centric Features*	
	(-7.72)	Percent of First Reviews	3.743***
Logged Length of the Review [Title]	0.148***		(59.14)
	(3.71)	Percent of Only Reviews	0.346***
Logged Length of the Review	-0.215***		(7.24)
	(-2.64)	Avg. Score [Reviewer]	0.102***
Logged Position of the Review [Ascending]	0.501***		(4.55)
	(15.60)	Stddev. of Scores [Reviewer]	-0.236***
Logged Position of the Review [Descending]	-0.075***		(-10.68)
	(-8.57)	Score Dummy [Reviewer]	-0.162***
First Review Dummy	-2.213***		(-7.34)
	(-19.94)	Good and Bad Scores	0.122***
Only Review Dummy	1.075***		(3.12)
	(5.65)	Good and Avg. Scores	-0.161***
% of Positive Sent.	0.258		(-6.63)
	(1.48)	Bad and Avg. Scores	-0.266***
Logged Positive Sent. Score	0.341***		(-2.75)
	(3.80)	All Scores	0.018
% of Negative Sent.	0.585*		(0.50)
	(1.93)	% of Negative Dev.	4.282***
Logged Negative Sent. Score	0.107***		(42.90)
	(7.75)	% of Positive Dev.	4.413***
Logged Positive Sent. Score [Title]	-0.103**		(44.80)
	(-2.57)	*Product-centric Reviews*	
Logged Negative Sent. Score [Title]	0.028***	Avg. Score [Product]	0.319***
	(5.05)		(4.84)
% of Numerals	4.855***	Stddev. of Scores [Product]	0.923***
	(8.17)		(6.15)
% of Capital Letters	0.358***	Logged # of Product Reviews	-0.119***
	(4.20)		(-8.27)
% of Comparatives	-5.955**	Logged Stddev. of # of Reviews	-0.177***
	(-2.57)		(-5.11)
Pseudo R^2		0.237	
N		794,068	

*** p<0.01, ** p<0.05, * p<0.1

Table 1.4. Summary Statistics

In this table, I report firm-level summary statistics for firms in the main sample. *Avg. Score* indicates average monthly review scores. *Abnormal Score* indicates average monthly review scores minus the previous-12-month average review scores, following Huang (2018a). *Avg. Peer Score* indicates average monthly review scores for products that are marked "Also Viewed" by Amazon, excluding products manufactured by the same firm. *Abnormal Peer Score* indicates average peer scores minus the previous 12-month average peer score. *Spamicity* indicates average firm-level spamicity of new reviews in a given month estimated with model [1]. *High (Low) Spamicity* indicates firms with above (below) median spamicity. *Stock Return*$_{m\ (q)}$ indicates monthly (quarterly) stock returns. Advertising indicates advertising expenses divided by total sales. *Gross Profitability* indicates the ratio of income before extraordinary items to the book value of assets. *F-score* indicates financial performance as signaled by Piotroski (2000). *Dollar Volume* is the dollar trading volume during the second-to-last month of the previous 12 months, following Huang (2018a). *CV of Dollar Volume* is the coefficient of variations in dollar trading volume calculated over the previous 12 months, beginning in the second-to-last month, following Huang (2018a). *GOV* indicates quarterly averages on the MSCI governance index. *ESG* indicates quarterly averages on the MSCI ESG index. *Independent (Grey)* indicates independent (grey) institutions as defined by Ferreira and Matos (2008). *SUE* indicates the difference between reported quarterly EPS and the median of the most recent EPS forecasts by all analysts prior to earnings announcements divided by stock prices. *Forecast Dispersion* is the standard deviation of the most recent EPS forecasts of all analysts prior to earnings announcements. Review and firm-level characteristics (imputed from quarterly observations) are monthly observations and governance indices; institutional ownership and analyst coverage are quarterly observations. All variables are winsorized at the 1% and 99% levels.

	All Sample				High Spamicity				Low Spamicity			
	N	Mean	Median	Stddev.	N	Mean	Median	Stddev.	N	Mean	Median	Stddev.
Review characteristics												
Avg. Score	18,081	4.02	4.10	0.52	9,041	3.88	3.96	0.53	9,040	4.16	4.22	0.46
Abnormal Score	18,081	0.11	0.03	0.78	9,041	-0.02	-0.01	0.57	9,040	0.24	0.08	0.93
Avg. Peer Score	18,081	4.16	4.21	0.40	9,041	4.10	4.16	0.39	9,040	4.22	4.27	0.41
Abnormal Peer Score	18,081	0.22	0.14	0.75	9,041	0.19	0.14	0.56	9,040	0.26	0.13	0.90
# of Products	18,081	180.6	29.0	584.8	9,041	278.3	48.0	766.6	9,040	82.9	21.0	277.8
# of Reviews	18,081	3,785	86	17,164	9,041	6,903	275	23,304	9,040	667	44	5,163
Spamicity	18,081	0.23	0.22	0.10	9,041	0.31	0.30	0.06	9,040	0.15	0.16	0.05
Firm-level characteristics												
Stock Return$_m$ (%)	18,081	1.14	1.18	9.42	9,041	1.02	1.10	9.51	9,040	1.25	1.24	9.33
Advertising	18,081	0.03	0.01	0.05	9,041	0.03	0.01	0.04	9,040	0.04	0.01	0.05
R&D/TA	18,081	0.04	0.02	0.07	9,041	0.06	0.03	0.08	9,040	0.02	0.01	0.05
Logged Market Cap.	18,081	7.97	7.94	2.31	9,041	8.06	8.02	2.43	9,040	7.88	7.90	2.18
M/B	18,081	2.06	1.69	1.36	9,041	2.04	1.66	1.28	9,040	2.00	1.69	1.30
Gross Profitability	18,081	0.04	0.06	0.20	9,041	0.03	0.06	0.22	9,040	0.05	0.07	0.18
F-Score	18,081	5.23	5.00	1.52	9,041	5.10	5.00	1.54	9,040	5.37	5.00	1.49
Logged Dollar Volume	18,081	14.83	15.10	2.73	9,041	14.93	15.20	2.86	9,040	14.73	14.98	2.58
CV of Dollar Volume	18,081	0.40	0.32	0.25	9,041	0.41	0.32	0.26	9,040	0.38	0.31	0.24
Governance Indices												
GOV	9,207	5.95	5.68	2.51	4,604	5.93	5.60	2.52	4,603	5.97	5.75	2.49
ESG	9,207	4.98	4.92	1.52	4,604	4.95	4.90	1.47	4,603	5.01	4.98	1.58
Institutional Ownership												
# of Institutions	6,475	457	230	535.5	3,238	481.2	237	570	3,237	429.9	220.5	492.5
Institutional Ownership	6,475	0.71	0.77	0.26	3,238	0.70	0.77	0.27	3,237	0.72	0.77	0.24

	All Sample				High Spamicity				Low Spamicity			
	N	Mean	Median	Stddev.	N	Mean	Median	Stddev.	N	Mean	Median	Stddev.
Independent Ownership	6,475	0.68	0.73	0.25	3,238	0.67	0.73	0.25	3,237	0.69	0.73	0.23
Grey Ownership	6,475	0.03	0.03	0.02	3,238	0.03	0.03	0.02	3,237	0.03	0.03	0.02
Analyst Coverage												
# of Analysts	6,465	18.1	14.5	15.7	3,233	19.9	16.0	17.8	3,232	16.3	13.0	12.9
SUE (%)	6,465	0.083	0.046	0.563	3,233	0.106	0.048	0.603	3,232	0.067	0.045	0.559
Forecast Dispersion	6,465	0.019	0.006	0.123	3,233	0.021	0.007	0.137	3,232	0.018	0.004	0.105

Table 1.5. Determinants of Spam

In this table, I report panel and 2SLS regression results derived from model [2]. Below, results derived from Models (1)–(6), (8), and (9) are panel regression estimates while those derived from Model (8) are 2SLS estimates where the instrumental variable is *Average Peer Score* and the instrumented variable is *Abnormal Score.* First-stage estimates are presented in the appendix I. Standard errors are double-clustered by firm and year-month. Numbers in parentheses indicate t-statistics.

Dependent Variable: Spamicity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Abnormal Score	-0.014***	-0.009***	-0.024***	-0.012***		-0.013***	-0.006***	-0.019***	-0.018***
	(-14.15)	(-9.74)	(-15.59)	(-14.74)		(-8.52)	(-6.64)	(-6.62)	(-6.63)
Stddev. of Score		0.031***				0.027***	0.033***	0.025***	0.025***
		(13.24)				(12.93)	(16.44)	(7.08)	(7.22)
Abnormal Peer Score			0.012***			0.005***		0.012***	0.011***
			(6.83)			(3.83)		(4.44)	(4.33)
Lagged Dependent Variable				0.373***		0.355***	0.356***	0.310***	0.316***
				(18.46)		(17.57)	(17.76)	(10.05)	(10.13)
Advertising					0.012*	0.010**	0.010**	0.019**	0.020**
					(1.90)	(2.21)	(2.26)	(2.39)	(2.47)
R&D/TA					-0.063	-0.036	-0.036	0.017	0.009
					(-1.25)	(-1.13)	(-1.13)	(0.22)	(0.11)
Logged Market Cap.					0.009*	0.007**	0.007**	0.010	0.010
					(1.80)	(2.00)	(2.26)	(1.14)	(1.13)
M/B					-0.005**	-0.004**	-0.004***	-0.004	-0.004
					(-2.23)	(-2.59)	(-2.87)	(-1.14)	(-1.14)
Gross Profitability					-0.028**	-0.019**	-0.018**	-0.025*	-0.027*
					(-2.41)	(-2.44)	(-2.47)	(-1.71)	(-1.83)
F-Score					-0.000	0.000	0.000	0.000	0.000
					(-0.39)	(0.35)	(0.45)	(0.35)	(0.47)
Logged Dollar Volume					-0.001	-0.000	-0.000	-0.001	-0.001
					(-0.74)	(-0.56)	(-0.67)	(-0.65)	(-0.54)
CV of Dollar Volume					0.004	0.003	0.003	0.009	0.009
					(0.84)	(0.85)	(0.99)	(1.41)	(1.46)
Stock Return$_m$					-0.000	-0.000	-0.000	-0.000	0.000
					(-1.07)	(-0.73)	(-1.00)	(-0.00)	(0.08)
GOV								0.001*	
								(1.69)	
ESG									0.003**
									(2.01)
Adjusted R^2	0.740	0.749	0.753	0.779	0.733	0.798	0.434	0.804	0.803
Firm fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	18,081	18,081	18,081	18,081	18,081	18,081	18,081	9,207	9,207

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6. Calendar-time Score and Spamicity Portfolio Returns

In this table, I report calendar-time tercile portfolio regression estimates. For each month, sample stocks are sorted into tercile portfolios based on abnormal scores (*Score*) or spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month. *Alpha* indicates abnormal portfolio returns after considering a Fama-French-Carhart four-factor model. *Long/Short* indicates a spread portfolio that buys the top tercile portfolio and sells the bottom tercile portfolio for Panel A and a spread portfolio that buys the bottom tercile portfolio and sells the top tercile portfolio for Panel B. Numbers in parentheses indicate t-statistics.

Panel A. Score

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
T1 (Low Score)	-0.335%	0.965	0.563	0.143	-0.126	0.728	127
	(-1.21)	(12.38)***	(4.07)***	(1.16)	(-1.94)*		
T2	-0.045%	1.014	0.326	0.076	0.001	0.808	127
	(-0.22)	(17.33)***	(3.14)***	(0.82)	(0.02)		
T3 (High Score)	0.363%	1.084	0.068	0.166	-0.097	0.634	127
	(1.10)	(11.70)***	(0.42)	(1.13)	(-1.25)		
Long/Short (High - Low)	0.698%	0.118	-0.495	0.023	0.030	0.071	127
	(2.57)**	(0.23)	(-1.75)*	(0.89)	(2.16)**		
Equal Weighting							
T1 (Low Score)	-0.319%	0.982	0.543	0.172	-0.129	0.701	127
	(-1.07)	(11.72)***	(3.66)***	(1.29)	(-1.84)*		
T2	-0.136%	0.966	0.387	0.033	0.012	0.787	127
	(-0.62)	(15.77)***	(3.56)***	(0.34)	(0.24)		
T3 (High Score)	0.407%	1.082	0.086	0.254	-0.116	0.605	127
	(1.39)	(11.00)***	(0.49)	(1.63)	(-1.42)		
Long/Short (High - Low)	0.726%	0.101	-0.457	0.082	0.013	0.059	127
	(2.55)**	(0.14)	(-1.71)*	(1.15)	(1.76)*		

Panel B. Spamicity

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
T1 (Low Spamicity)	0.223%	1.088	0.419	0.152	-0.053	0.842	127
	(1.10)	(19.11)***	(4.13)***	(1.68)*	(-1.11)		
T2	0.176%	1.103	0.422	0.077	0.055	0.750	127
	(0.94)	(14.64)***	(3.14)***	(0.64)	(0.87)		
T3 (High Spamicity)	-0.416%	1.021	0.513	0.154	-0.049	0.675	127
	(-1.75)*	(11.56)***	(3.26)***	(1.09)	(-0.66)		
Long/Short (High - Low)	0.639%	0.067	-0.094	-0.002	-0.004	0.004	127
	(2.55)**	(0.60)	(-0.54)	(-0.06)	(-0.00)		
Equal Weighting							
T1 (Low Spamicity)	0.197%	1.069	0.415	0.116	-0.062	0.853	127
	(1.02)	(19.79)***	(4.31)***	(1.35)	(-1.36)		
T2	0.148%	1.069	0.426	0.035	0.041	0.764	127
	(0.12)	(14.98)***	(3.35)***	(0.31)	(0.69)		
T3 (High Spamicity)	-0.394%	0.995	0.580	0.141	-0.015	0.631	127
	(-1.76)*	(10.33)***	(3.38)***	(0.92)	(-0.18)		
Long/Short (High - Low)	0.591%	0.073	-0.166	-0.025	-0.047	0.010	127
	(2.54)**	(0.47)	(-0.79)	(-0.16)	(-0.56)		

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7. Calendar-time Spamicity Portfolio Returns Conditional on Score

In this table, I report calendar-time 3 x 3 conditional tercile portfolio regression estimates. For each month, sample stocks are first sorted into tercile portfolios based on abnormal scores (*Score*) and then conditionally sorted into tercile portfolios based on spamicity (*Spamicity*) within each score-tercile portfolio. The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month. *Alpha* indicates abnormal portfolio returns after considering a Fama-French-Carhart four-factor model. In Panel A I report alpha estimates only. *Long/Short* indicates a spread portfolio that buys the bottom tercile spamicity portfolio and sells the top tercile spamicity portfolio within each score tercile. Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

	T1 (Low Spamicity)	T2	T3 (High Spamicity)	Long/Short (L. Spamicity - H. Spamicity)
Review Weighting (Score/Spamicity)				
T1 (Low Score)	-0.065%	-0.107%	-0.834%	0.769%
	(-0.47)	(-0.41)	(-4.35)***	(2.22)*
T2	0.571%	-0.154%	-0.551%	1.122%
	(1.29)	(-1.46)	(-1.87)*	(3.30)***
T3 (High Score)	0.682%	0.487%	-0.081%	0.763%
	(1.62)	(0.27)	(-0.17)	(1.79)
Equal Weighting (Score/Spamicity)				
T1 (Low Score)	-0.056%	-0.089%	-0.813%	0.757%
	(-0.50)	(-0.53)	(-3.87)***	(2.46)**
T2	0.376%	-0.172%	-0.613%	0.989%
	(1.44)	(-1.64)	(-1.97)*	(2.80)**
T3 (High Score)	0.694%	0.423%	0.103%	0.591%
	(1.61)	(1.41)	(-0.23)	(1.70)

Table 1.8. Score, Spamicity and Return Predictability

In this table, I report Fama-MacBeth regression estimates of one-month-ahead excess stock returns on abnormal scores and spamicity using model [3]. Coefficients from monthly cross-sectional regression estimates are averaged across time and reported below. *Month of Earnings Announcement* indicates the month of a year when a firm reports its quarterly earnings and other financial outcomes. Numbers in parentheses indicate Fama-MacBeth t-statistics using Newey-West standard errors corrected for twelve lags.

Dependent Variable: Excess Return$_{t+1}$ (%)	(1)	(2)	(3)	(4)	(5)	(6)	*Month of Earnings Announcement*	
							NO	YES
							(7)	(8)
Score	1.114**		4.649**	1.142*		3.852*	4.909	0.345
	(2.44)		(2.49)	(1.67)		(1.69)	(1.57)	(0.28)
Spamicity		-4.413*	-4.078**		-2.640*	-2.538**	-10.956**	-5.997*
		(-1.67)	(-2.01)		(-1.67)	(-2.01)	(-2.00)	(-1.89)
Score x Spamicity			-16.738**			-13.751*	-34.115**	-7.124
			(-2.20)			(-1.85)	(-2.38)	(-1.05)
Advertising				-0.491	-0.492	-0.670*	-0.455	-1.141
				(-1.39)	(-1.52)	(-1.78)	(-0.52)	(-1.33)
R&D/TA				-5.988	-4.274	-5.241	12.204	-4.379
				(-1.64)	(-1.17)	(-1.31)	-1.01	(-0.59)
Logged Market Cap				0.010	-0.005	0.049	-0.033	0.098
				(0.06)	(-0.03)	(0.34)	(-0.11)	-0.14
M/B				0.747***	0.709***	0.727***	0.291	0.684
				(4.02)	(4.04)	(3.95)	-0.65	-1.25
Gross Profitability				6.416*	6.233**	6.698**	7.586	1.679
				(1.97)	(2.05)	(2.10)	(1.61)	(0.43)
F-Score				0.509***	0.495***	0.446***	0.504	-0.208
				(3.29)	(3.61)	(3.33)	-1.47	(-0.46)
Logged Dollar Volume				-0.262*	-0.276*	-0.272*	-0.115	-0.266
				(-1.81)	(-1.76)	(-1.72)	(-0.45)	(-0.42)
CV of Dollar Volume				4.656***	4.460**	4.573**	7.511	7.469**
				(2.83)	(2.45)	(2.55)	-1.46	-2.07
Stock Return$_m$				-0.036	-0.042**	-0.042**	-0.167	-0.273
				(-1.61)	(-2.05)	(-2.25)	(-1.58)	(-1.62)
Average R^2	0.02	0.014	0.052	0.201	0.184	0.218	0.418	0.677
N	18,081	18,081	18,081	18,081	18,081	18,081	11,967	6,114

*** p<0.01, ** p<0.05, * p<0.1

Table 1.9. Calendar-time Portfolio Returns Double Sorted on Score and Spamicity

In this table, I report calendar-time 3 x 3 unconditional tercile portfolio regression estimates. For each month, sample stocks are double-sorted into 3 x 3 tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews in each month. *Alpha* indicates abnormal portfolio returns after considering a Fama-French-Carhart four-factor model. In Panel A I report alpha estimates only. *Long/Short* indicates a spread portfolio that buys the top tercile score portfolio and sells the bottom tercile score portfolio for (A), a spread portfolio that buys the bottom tercile spamicity portfolio and sells the top spamicity tercile portfolio for (B), a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (C), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (D). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

	T1 (Low Spamicity)	T2	T3 (High Spamicity)	(B) Long/Short (L. Spamicity - H. Spamicity)
Review Weighting (Score/Spamicity)				
T1 (Low Score)	-0.185%	-0.234%	-0.586%	0.401%
	(-0.37)	(-1.05)	(-2.11)**	(1.33)
T2	0.267%	0.107%	-0.509%	0.776%
	(0.09)	(1.24)	(1.98)*	(-2.13)**
T3 (High Score)	0.587%	0.656%	-0.154%	0.741%
	(1.74)*	(1.88)*	(-0.11)	(1.75)*
(A) Long/Short (H. Score - L. Score)	0.772%	0.890%	0.432%	
	(1.94)*	(2.21)**	(0.81)	
Equal Weighting (Score/Spamicity)				
T1 (Low Score)	-0.075%	-0.227%	-0.655%	0.580%
	(-0.43)	(-1.17)	(-2.21)**	(1.42)
T2	0.087%	0.135%	-0.630%	0.717%
	(0.28)	(1.18)	(-2.17)**	(2.18)**
T3 (High Score)	0.579%	0.537%	0.104%	0.475%
	(2.12)**	(1.64)	(0.91)	(1.56)
(A) Long/Short (H. Score - L. Score)	0.654%	0.764%	0.759%	
	(1.81)*	(2.17)**	(2.16)**	

Panel B. Score x Spamicity Corner Portfolios

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.173%	0.066	-0.316	-0.030	-0.279	0.071	127
- L. Score & H. Spamicity)	(2.50)**	(0.16)	(-1.14)	(-0.11)	(-2.57)**		
(D) Long/Short (H. Score & H. Spamicity	0.226%	-0.054	-0.289	-0.119	0.192	0.019	127
- L. Score & L. Spamicity)	(0.06)	(-0.22)	(-0.72)	(-0.39)	(1.07)		
Equal Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.234%	0.081	-0.320	0.020	-0.405	0.090	127
- L. Score & H. Spamicity)	(2.53)**	(0.41)	(-1.24)	(0.02)	(-2.91)***		
(D) Long/Short (H. Score & H. Spamicity	0.183%	-0.037	-0.230	0.025	0.148	0.016	127
- L. Score & L. Spamicity)	(0.06)	(-0.18)	(-0.57)	(0.09)	(0.89)		

*** p<0.01, ** p<0.05, * p<0.1

Table 1.10. Calendar-time Portfolio based on E-commerce Dependence

In this table, I report calendar-time 3 x 3 x 2 tercile portfolio regression estimates. For each month, sample stocks are first divided into high- and low-characteristic groups (*E-commerce Sales, Number of New Products, and Average Product Price*) and then are double-sorted into tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). *E-commerce Sales* indicates the firm-level average of e-commerce sales based on each product line's e-commerce sales from the U.S. Census Bureau. *Number of New Products* indicates firm-level monthly average number of new products that first appear in the sample. *Average Product Price* indicates firm-level average prices of a firm's products based on the July 2014 price. The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month (only review-weighted Fama-French-Carhart four-factor model alpha estimates are reported below). *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (A), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (B). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

Panel A. High and Low E-commerce Sales

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	0.245%	1.414%
	(0.77)	(2.15)**
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	0.224%	-0.148%
	(0.30)	(-0.18)
E-commerce Sales	Low	High

Panel B. High and Low Number of New Products

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	0.595%	1.454%
	(0.86)	(1.98)*
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	-0.271%	-0.900%
	(-0.35)	(-1.09)
Number of New Products	Low	High

Panel C. High and Low Average Product Price

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	0.430%	1.337%
	(0.89)	(1.86)*
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	-0.400%	0.631%
	(-0.43)	(0.70)
Average Product Price	Low	High

*** p<0.01, ** p<0.05, * p<0.1

Table 1.11. Score, Spamicity and Institutional Ownership

In this table, I report panel regression estimates of institutional ownership on abnormal scores (*Score*) and spamicity (*Spamicity*) using model [4]. All variables are calculated at the quarterly level. *Indep. (Grey)* indicates independent (grey) institutions as defined by Ferreira and Matos (2008). Standard errors are double-clustered by firm and by year-quarter. Numbers in parentheses indicate t-statistics.

Dependent Variables:	*Institutional Ownership*			*# of Institutions*	*Indep.*	*Grey*
	(1)	(2)	(3)	(4)	(5)	(6)
Score	0.000		0.003	5.283**	0.004	-0.000
	(0.24)		(1.31)	(2.57)	(1.56)	(-1.18)
Spamicity		-0.022	-0.023	3.269	-0.022	-0.004
		(-0.52)	(-0.53)	(0.07)	(-0.51)	(-0.86)
Score x *Spamicity*			-0.023**	-34.745***	-0.024**	0.001
			(-2.03)	(-2.97)	(-2.05)	(0.34)
Advertising	0.016	0.017	0.017	31.434	0.015	0.003
	(0.47)	(0.50)	(0.52)	(1.44)	(0.45)	(1.11)
R&D/TA	0.089	0.090	0.090	25.220	0.077	0.005
	(1.32)	(1.33)	(1.33)	(0.38)	(1.13)	(0.74)
Logged Market Cap.	0.035**	0.035**	0.035**	141.438***	0.033**	0.002**
	(2.57)	(2.59)	(2.61)	(3.78)	(2.39)	(2.17)
M/B	-0.012**	-0.012**	-0.012**	-10.302	-0.011*	-0.002***
	(-2.13)	(-2.14)	(-2.16)	(-1.25)	(-1.90)	(-4.34)
Gross Profitability	0.008	0.008	0.010	125.072	0.014	0.000
	(0.08)	(0.07)	(0.09)	(1.16)	(0.13)	(0.04)
F-Score	-0.000	-0.000	-0.000	-1.298	-0.000	-0.000
	(-0.35)	(-0.36)	(-0.39)	(-1.33)	(-0.35)	(-0.45)
Logged Dollar Volume	0.030***	0.030***	0.030***	-9.577	0.029***	0.001*
	(3.52)	(3.53)	(3.66)	(-1.33)	(3.45)	(1.69)
CV of Dollar Volume	-0.051***	-0.051***	-0.051***	11.334	-0.050***	-0.003**
	(-4.96)	(-4.97)	(-5.00)	(0.93)	(-4.93)	(-2.05)
Stock Return$_q$	-0.001	-0.001	0.001	-45.973	0.002	-0.001
	(-0.05)	(-0.03)	(0.05)	(-1.43)	(0.11)	(-0.33)
Adjusted R^2	0.905	0.905	0.905	0.980	0.890	0.796
Firm fixed effect	YES	YES	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES
N	6,475	6,475	6,475	6,475	6,475	6,475

*** p<0.01, ** p<0.05, * p<0.1

Table 1.12. Score, Spamicity, and Earning Surprises

In this table, I report panel regression estimates of the effects of abnormal scores (*Score*) and spamicity (*Spamicity*) on earnings surprises using model [4]. All variables are calculated at the quarterly level. The dependent variable is unexpected earnings (*SUE*). *Analyst Coverage* indicates the previous quarter's number of analysts. *High (Low)* indicates above- (below-)median scores in each category. Standard errors are double-clustered by firm and year-quarter. Numbers in parentheses indicate t-statistics.

				Analyst Coverage	
Dependent Variable: SUE	(1)	(2)	(3)	Low	High
				(4)	(5)
Score	0.016*		0.032**	0.056**	0.006
	(1.88)		(2.34)	(2.33)	(1.63)
Spamicity		-0.094	0.000	-0.024	0.107
		(-0.54)	(0.00)	(-0.08)	(1.60)
Score x *Spamicity*			-0.162**	-0.325**	-0.032*
			(-2.09)	(-2.14)	(-1.74)
Forecast Dispersion	-2.834***	-2.795***	-2.795***	-2.650***	4.636
	(-4.26)	(-3.99)	(-3.99)	(-3.76)	(0.68)
Advertising	0.112	0.219	0.215	0.338	0.006
	(0.65)	(1.09)	(1.08)	(0.69)	(0.20)
R&D/TA	3.573***	3.802***	3.796***	6.794**	0.312
	(3.90)	(3.91)	(3.90)	(2.67)	(1.22)
Logged Market Cap	-0.097**	-0.086**	-0.086**	-0.106	-0.026
	(-2.57)	(-2.17)	(-2.17)	(-1.16)	(-1.37)
M/B	0.045***	0.039***	0.040***	0.072***	-0.001
	(3.25)	(2.95)	(3.02)	(2.72)	(-0.18)
Gross Profitability	0.471	0.779	0.780	2.614*	0.003
	(0.95)	(1.46)	(1.45)	(2.02)	(0.02)
F-Score	0.018*	0.018*	0.018*	0.026	0.002
	(1.98)	(1.84)	(1.86)	(1.53)	(0.66)
Logged Dollar Volume	0.018	0.012	0.012	-0.000	0.021*
	(0.80)	(0.55)	(0.52)	(-0.00)	(1.78)
CV of Dollar Volume	-0.052	-0.058	-0.056	-0.083	-0.005
	(-0.89)	(-0.93)	(-0.91)	(-0.73)	(-0.32)
Stock Return$_{q-1}$	0.009	0.011	0.011	0.000	0.004
	(0.82)	(0.93)	(0.93)	(0.04)	(0.67)
Adjusted R^2	0.174	0.174	0.174	0.195	0.594
Firm fixed effect	YES	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES	YES
N	6,465	6,465	6,465	3,232	3,233

*** p<0.01, ** p<0.05, * p<0.1

Table 1.13. Calendar-time Portfolio based on Institutional Ownership

In this table, I report calendar-time 3 x 3 x 2 tercile portfolio regression estimates. For each month, sample stocks are first divided into high- and low-characteristic groups (*Independent Institutional Ownership/Grey Institutional Ownership/Analyst Coverage*) and then are double-sorted into tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). *Independent (Grey) Institutional Ownership* indicates the previous quarter's independent (grey) institutional ownership. *Analyst Coverage* indicates the previous quarter's number of analysts. The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month (only review-weighted Fama-French-Carhart four-factor model alpha estimates are reported below). *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (A), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (B). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

Panel A. By Independent Ownership

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	1.201%	-0.228%
	(2.24)**	(-0.34)
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	-0.609%	-0.125%
	(-0.81)	(-0.15)
Independent Institutional Ownership	Low	High

Panel B. By Grey Institutional Ownership

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	0.672%	0.280%
	(1.30)	(0.48)
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	-0.031%	-0.864%
	(-0.04)	(-1.14)
Grey Institutional Ownership	Low	High

Panel C. By Analyst Coverage

	Alpha	
(A) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	1.298%	0.238%
	(1.80)*	(0.36)
(B) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	0.862%	-1.197%
	(1.02)	(-1.34)
Analyst Coverage	Low	High

*** p<0.01, ** p<0.05, * p<0.1

Table 1.14. Shock in Spamming and Institutional Ownership

In this table, I report panel regression estimates of the effects of institutional ownership on abnormal scores (*Score*) and spamicity (*Spamicity*) derived from model [5]. All variables are calculated at the quarterly level. *Indep. (Grey)* indicates independent (grey) institutions as defined by Ferreira and Matos (2008). *Post 2012* takes the value of 1 if the year of the observation is 2012 or later and 0 otherwise. *High (Low) Exposure* indicates that a firm's products are sold in item categories with above- (below-)median proportions of products produced by Chinese manufacturers after 2012. Control variables are the same as those for which results are reported in Table 1.10. Standard errors are double-clustered by firm and year-quarter. Numbers in parentheses indicate t-statistics.

Dependent Variable: *Institutional Ownership*	*Indep.*		*Grey*	
	(1)	(2)	(3)	(4)
Score	0.004	0.001	-0.000	-0.000
	(1.06)	(0.15)	(-0.05)	(-0.59)
Spamicity	0.031	0.069	-0.008	-0.006*
	(1.03)	(0.69)	(-1.45)	(-1.91)
Score x *Spamicity*	-0.018	-0.001	-0.001	-0.002
	(-0.93)	(-0.04)	(-0.35)	(-0.60)
Spamicity x Post 2012	0.047	0.001	-0.004	0.018***
	(0.62)	(0.02)	(-0.37)	(6.21)
Score x Post 2012	-0.003	0.010	0.000	0.000
	(-0.82)	(1.44)	(1.20)	(0.81)
Score x *Spamicity* x Post 2012	-0.006	-0.057**	-0.000	-0.002
	(-0.17)	(-2.98)	(-0.03)	(-0.56)
Adjusted R^2	0.942	0.906	0.866	0.905
Firm fixed effect	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Exposure	Low	High	Low	High
N	2,160	2,772	2,160	2,772

*** p<0.01, ** p<0.05, * p<0.1

Appendix I. Variable Definitions

In this table, I present detailed definitions of the *Review-*, *Reviewer-*, and *Product-centric features* used in model [1]. The review data are taken from SNAP and sentiments are estimated using the sentiment lexicons (lists of words' sentiment scores) published in Hamilton et al. (2016). *Review-centric features* represent individual-review-level information (full-review -sample level). *Reviewer-centric features* are also defined at the full-review-sample level to capture the hidden overall behavioral patterns of reviewers more comprehensively. *Product-centric features* are monthly product-level scores and review characteristics.

Variable	Definition
Review-centric Features	
Logged # representing Feedback	Logged number representing instances of feedback given to a review.
Logged # representing Helpful Feedback	Logged number representing instances of feedback given to a review that indicates that the review is helpful.
% represent ting Helpful Feedback	The proportion of Helpful feedback of the total number of instances of feedback [0 if there are no instances of Helpful feedback].
Logged Length of a Review [Title]	Logged number of characters used in a review title.
Logged Length of a Review	Logged number of characters used in the main text of a review.
Logged Position of a Review [Ascending]	Logged ascending chronological order of a review.
Logged Position of a Review [Descending]	Logged descending chronological order of a review.
First Review Dummy	Dummy variable that takes the value of 1 if a review is the first review for a given product in the sample and 0 otherwise.
Only Review Dummy	Dummy variable that takes the value of 1 if a review is the only review for a given product in the sample and 0 otherwise.
% of Positive Sent.	The number of words with positive sentiment scores over the total number of words in the main text of a review.
Logged Positive Sent. Score	Logged sum of the scores of the words in the main text of a review with positive sentiment.
% of Negative Sent.	The number of words with negative sentiment scores over the total number of words in the main text of a review.
Logged Negative Sent. Score	Logged sum of the scores of the words in the main text of a review with negative sentiment.
Logged Positive Sent. Score [Title]	Logged sum of the scores of the words in the title of a review with positive sentiment.
Logged Negative Sent. Score [Title]	Logged sum of the scores of the words in the title of a review with negative sentiment.
% of Numerals	The number of numeral characters in the main text of a review over the total number of characters in the main text of a review.
% of Capital Letters	The number of capital characters in the main text of a review over the total number of characters in the main text of a review.
% of Comparatives	The number of comparatives in the main text of a review over the total number of words in the main text of a review.
Review Score	The score of a review [1, 2, 3, 4, or 5]
Score Dummy	Indicator variable that takes the value of 1 for scores of 4 or above, -1 for scores 2 or below, and 0 otherwise.
Reviewer-centric Features	
Percent of First Reviews	The number of reviews that are the first review for a given product over the total number of reviews posted by a reviewer.
Percent of Only Reviews	The number of reviews that are the only review for a given product over the total number of reviews posted by a reviewer.
Avg. Score [Reviewer]	The average score of the reviews posted by a reviewer.
Stddev. of Scores [Reviewer]	The standard deviation of the scores of the reviews posted by a reviewer.
Score Dummy [Reviewer]	Indicator variable that takes the value of 1 if the average score of a reviewer is 4 or above, -1 if it is 2 or below, and 0 otherwise.

Variable	Definition
Good and Bad Scores	Dummy variable that takes the value of 1 if a reviewer has given both scores of 4 or above and scores of 2 or below and 0 otherwise.
Good and Avg. Scores	Dummy variable that takes the value of 1 if a reviewer has given scores of 3 or above and 0 otherwise.
Bad and Avg. Scores	Dummy variable that takes the value of 1 if a reviewer has given both scores of 3 or below and 0 otherwise.
All Scores	Dummy variable that takes the value of 1 if a reviewer has given scores of all levels and 0 otherwise.
% of Negative Dev.	The number of reviews that receive scores higher than the average score of a product when the review was posted over the total number of reviews posted by a reviewer.
% of Positive Dev.	The number of reviews that receive scores lower than the average score for a product when the review was posted over the total number of reviews posted by a reviewer.
Product-centric Reviews	
Avg. Score [Product]	The average review score for a product in each month.
Stddev. of Scores [Product]	The standard deviation of the review scores of a product in each month.
Logged # of Product Reviews	Logged total number of reviews of a product in each month.
Logged Stddev. of # of Reviews	Logged standard deviation of the total number of reviews of a product in each month.

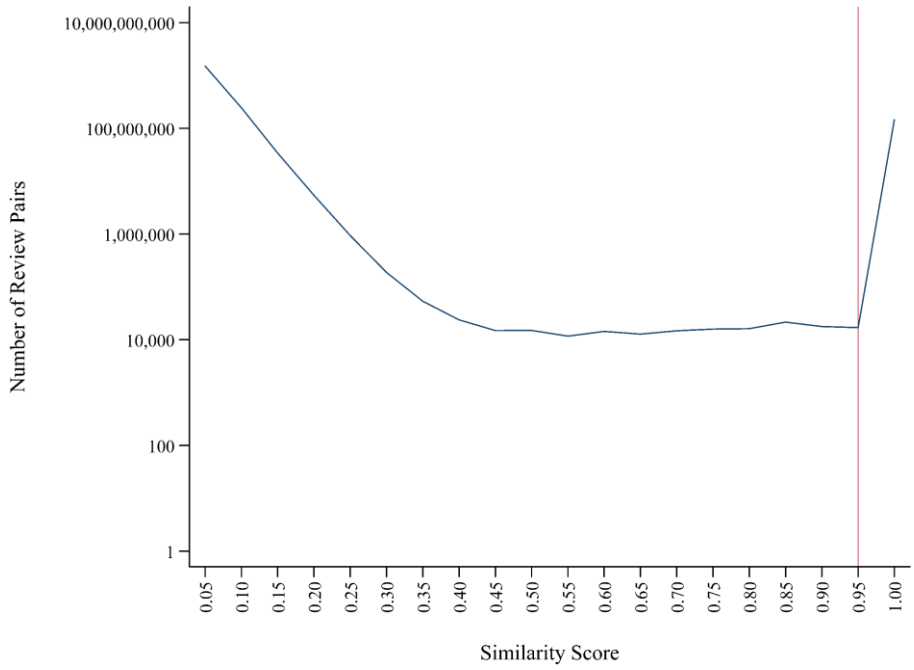


Figure 1.A1. Similarity Scores of Review Pairs

This figure shows how many review pairs in the sample receive the similarity scores represented on the horizontal axis. *Similarity Score* indicates 2-gram Jaccard distances between reviews. There are potentially 44.6 million x (44.6 million - 1)/2 review pairs in the sample. To ease the computational burden, I lexicographically arrange reviews and form 1,000 review-groups and then measure 2-gram Jaccard distances within each group. As a result, there are approximately 44.7 billion review pairs between which to measure the Jaccard distance. The number of review pairs and the number of unique reviews are not equal as there are multiple reviews with the same or similar content in the sample.

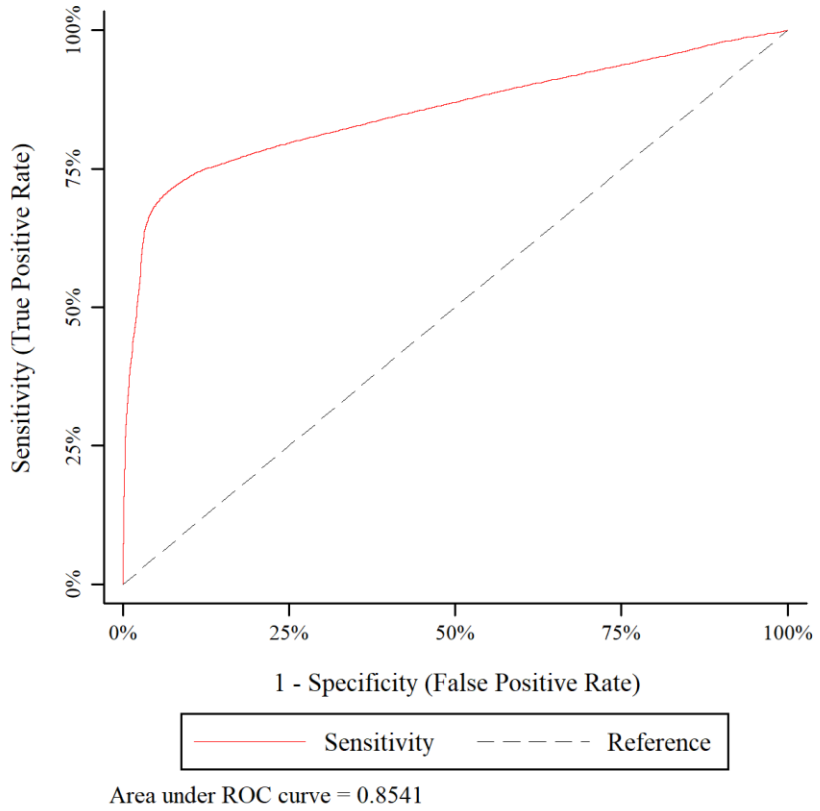


Figure 1.A2. Receiver Operating Characteristic (ROC) Curve

This figure shows one of the ten ROC curves attained after estimating model [1]. *Sensitivity* indicates the proportion of positive examples (near-duplicate reviews) in the training sample that are correctly identified as spam by model [1] under a given probability threshold (1 - x-axis). *1- Specificity* indicates the proportion of negative examples (non-near-duplicate reviews) in the training sample that are falsely identified as spam by model [1] under a given probability threshold (1 - y-axis). Using a ten-fold cross-validation method, the training sample is randomly divided into ten groups and thus yields ten out-of-sample ROC curves. The average Area Under Curve (AUC) of the ten ROC curves is 0.8551.

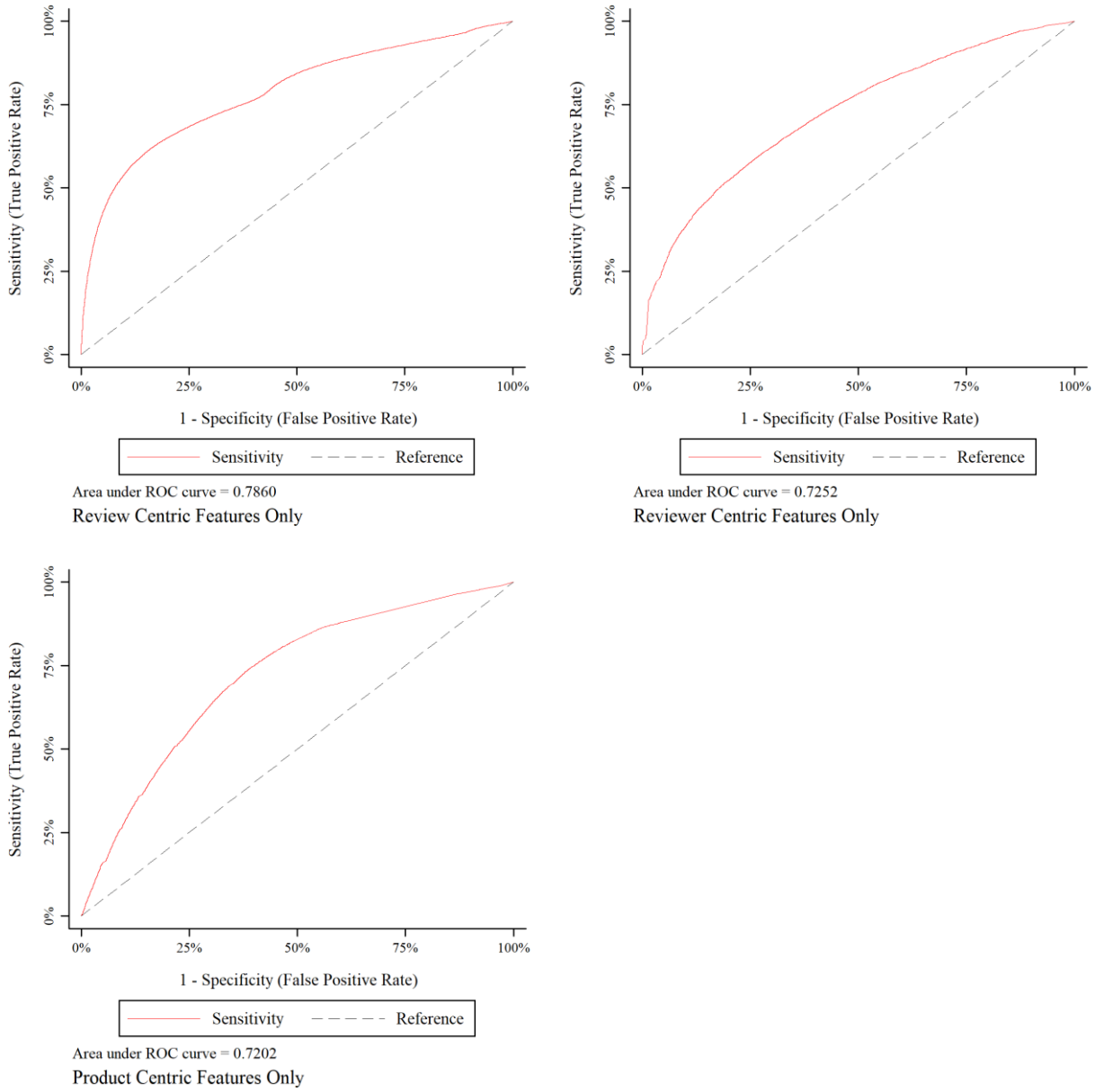


Figure 1.A3. Comparison of Spam Predictability between Review-, Reviewer-, and Product-centric Features

Each graph in this figure shows one of the ten ROC curves in each category after estimating model [1] using one of the review-, reviewer- or product-centric features exclusively. The average Areas Under Curve (AUCs) of the ten groups for review-centric, reviewer-centric, and product-centric features are 0.7862, 0.7255, and 0.7212, respectively.

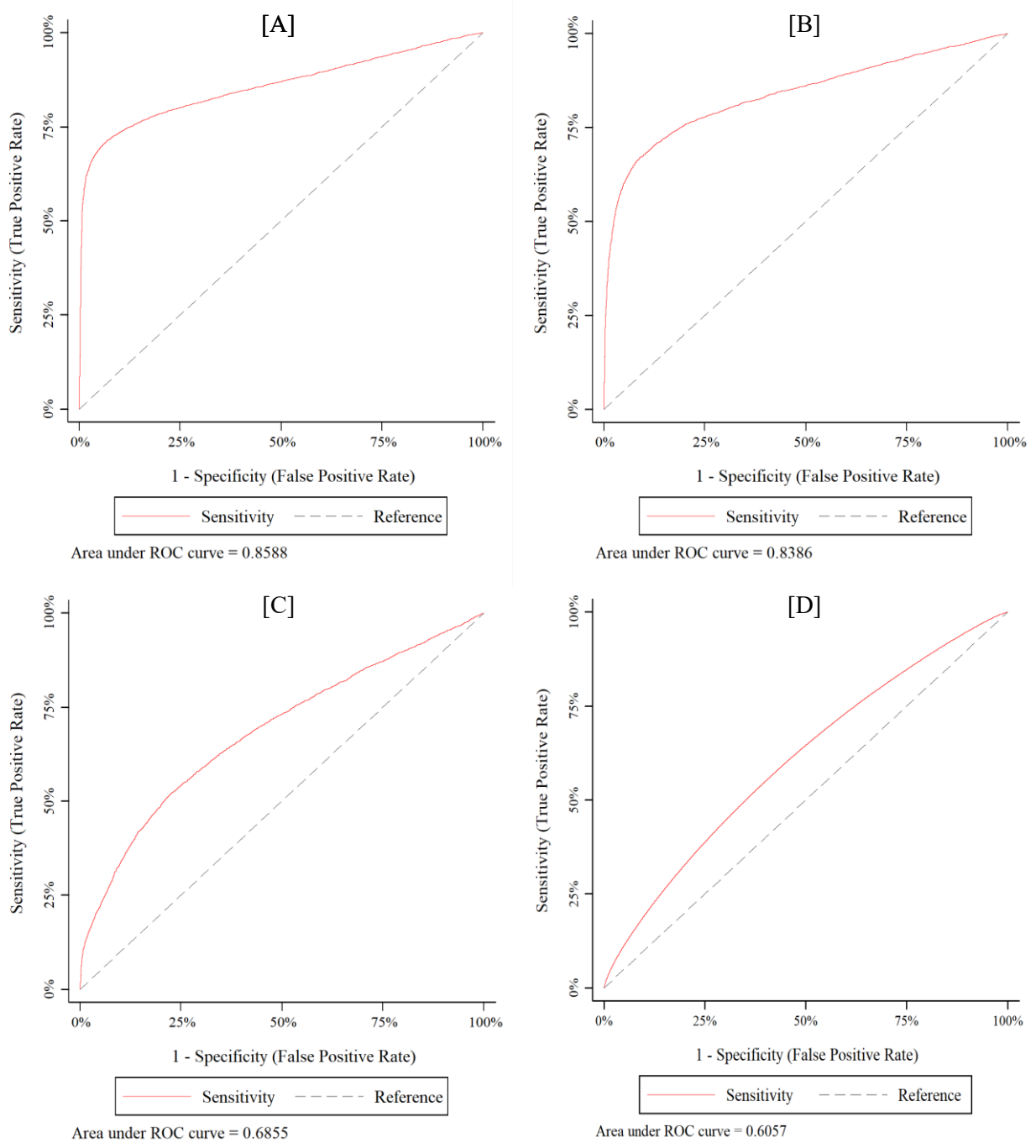


Figure 1.A4. Spam Predictability for separate training samples

Each of the above graphs shows one of the ten ROC curves after estimating model [1] using different negative examples: [A] review scores that do not deviate more than +/- 0.01 from the average, [B] review scores that do not deviate more than +/– 0.05 from the average, [C] review scores that do not deviate more than +/– 0.5 from the average, and [D] all of the non-near-duplicate reviews. The average Area Under Curves (AUCs) of the ten groups are 0.8581, 0.8392, 0.6854, and 0.6055 for [A], [B], [C], and [D], respectively. The replication of results reported in table 1.9 using separate training samples above are presented in Table 1.A5.

Table 1.A1. First-Stage Estimates

In this table, I report the first-stage estimations of the results reported in column (7) of Table 1.5. The instrumental variable is *Average Peer Score* and the instrumented variable is *Abnormal Score.* Standard errors are double clustered by firm and year-month. Numbers in parentheses indicate t-statistics.

Dependent Variable: Abnormal score	
Abnormal Peer Score	0.740***
	(27.55)
Standard Deviation of Score	-0.794***
	(-42.49)
Lagged Dependent Variable	-0.097
	(-0.90)
Advertising	0.031
	(0.98)
R&D/TA	-0.073
	(-0.38)
Logged Market Cap	0.007
	(0.42)
M/B	0.000
	(0.02)
Gross Profitability	-0.112*
	(-1.88)
F-Score	0.001
	(0.32)
Logged Dollar Volume	-0.002
	(-0.30)
CV of Dollar Volume	-0.034
	(-1.05)
Return	0.000
	(0.11)
Adjusted R^2	0.716
Firm FE	Yes
Year FE	Yes
F-stat	84.51
N	18,081

*** p<0.01, ** p<0.05, * p<0.1

Table 1.A2. Longer-period Score and Spamicity Portfolio Returns

In this table, I report longer-period calendar-time 3 x 3 tercile portfolio regression estimates. For each month, sample stocks are double-sorted into tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month (only Fama-French-Carhart four-factor model alpha estimates are reported below). *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (C), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (D). Months [*a* ,*b*] indicates a spread portfolio alpha from month *a* to month *b* after portfolio formation, excluding the first-month returns after the formation month. Each spread portfolio is rebalanced after each month. Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

Panel A. (C) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)

	Review Weighting		Equal Weighting	
Months [2, 4]	0.30%	(0.23)	0.31%	(0.23)
Months [2, 7]	0.04%	(0.02)	0.06%	(0.03)
Months [2, 10]	0.12%	(0.53)	0.01%	(0.60)
Months [2, 13]	0.03%	(1.17)	0.03%	(1.08)

Panel B. (D) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)

	Review Weighting		Equal Weighting	
Months [2, 4]	-0.31%	(-0.18)	0.21%	(0.13)
Months [2, 7]	-0.03%	(-1.06)	-0.03%	(-0.12)
Months [2, 10]	-0.06%	(-0.89)	-0.02%	(-0.77)
Months [2, 13]	-0.09%	(-1.06)	-0.05%	(-1.08)

*** p<0.01, ** p<0.05, * p<0.1

Table 1.A3. Calendar-time Score and Spamicity Spread Portfolio Returns - Alternative Factors

In this table, I report calendar-time 3 x 3 tercile portfolio regression estimates. For each month, sample stocks are double-sorted into tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month. *Alpha* indicates abnormal portfolio returns after considering a Fama-French-Carhart four-factor model adding two additional factors from Fama-French (2015). *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (C), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (D). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

	Alpha	Market	SMB	HML	MOM	RMW	CMA	R^2	N
Review Weighting									
(C) Long/Short (H. Score & L. Spamicity	1.056%	0.041	-0.237	0.302	-0.390	0.543	0.315	0.101	127
- L. Score & H. Spamicity)	(2.60)**	(0.25)	(-0.88)	(1.15)	(-3.11)***	(1.42)	(0.71)		
(D) Long/Short (H. Score & H. Spamicity	-0.051%	-0.102	-0.026	-0.507	0.025	-0.115	-0.028	0.039	127
- L. Score & L. Spamicity)	(-0.09)	(-0.63)	(-0.10)	(-1.91)*	(0.19)	(-0.30)	(-0.06)		
Equal Weighting									
(C) Long/Short (H. Score & L. Spamicity	1.167%	0.058	-0.204	0.332	-0.368	0.584	0.168	0.093	127
- L. Score & H. Spamicity)	(2.21)**	(0.36)	(-0.75)	(1.25)	(-2.90)***	(1.51)	(0.38)		
(D) Long/Short (H. Score & H. Spamicity	-0.086%	-0.106	0.043	-0.463	-0.091	-0.243	-0.040	0.051	127
- L. Score & L. Spamicity)	(-0.15)	(-0.63)	(0.15)	(-1.70)*	(-0.70)	(-0.61)	(-0.09)		

*** p<0.01, ** p<0.05, * p<0.1

Table 1.A4. Alternative Training Samples

In this table, I report calendar-time 3 x 3 tercile portfolio regression estimates using varying spamicity estimates. In the training sample, negative examples are reviews that do not deviate from the average review score by +/−0.01 (*Range = +/−0.01*), +/−0.05 (*Range = +/−0.05*), and +/−0.5 (*Range = +/−0.5*). Average Spamicity in each Range are 0.51 (*Range = +/−0.01*), 0.34 (*Range = +/−0.05*), and 0.05 (*Range = +/−0.5*). For each month, sample stocks are double-sorted into tercile portfolios based on abnormal scores (*Score*) and spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month. *Alpha* indicates abnormal portfolio returns after considering a Fama-French-Carhart four-factor model. *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (C), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (D). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

Panel A. Abnormal Score x Spamicity (Range = +/−0.01)

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.322%	0.028	-0.469	0.348	-0.210	0.054	127
- L. Score & H. Spamicity)	(2.31)**	(0.07)	(-1.56)	(1.48)	(-1.16)		
(D) Long/Short (H. Score & H. Spamicity	-0.632%	-0.015	-0.008	0.302	-0.029	0.009	127
- L. Score & L. Spamicity)	(-1.58)	(-0.09)	(-0.94)	(0.35)	(-0.19)		
Equal Weighting							
(C) Long/Short (H. Score & L. Spamicity	0.801%	-0.143	0.009	0.323	-0.095	0.033	127
- L. Score & H. Spamicity)	(2.14)**	(-0.88)	(0.48)	(1.13)	(-0.98)		
(D) Long/Short (H. Score & H. Spamicity	-0.181%	-0.063	-0.048	0.150	-0.178	0.021	127
- L. Score & L. Spamicity)	(-1.19)	(-0.03)	(-0.47)	(0.37)	(-1.45)		

Panel B. Abnormal Score x Spamicity (Range = +/−0.05)

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.395%	0.027	0.851	0.059	0.056	0.068	127
- L. Score & H. Spamicity)	(2.10)**	(0.14)	(2.58)**	(0.20)	(0.36)		
(D) Long/Short (H. Score & H. Spamicity	0.516%	0.076	-0.688	0.599	-0.138	0.076	127
- L. Score & L. Spamicity)	(0.74)	(0.39)	(-1.97)*	(1.92)*	(-0.84)		
Equal Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.147%	-0.133	0.614	0.303	-0.157	0.038	127
- L. Score & H. Spamicity)	(1.77)*	(-0.73)	(1.90)*	(1.05)	(-1.03)		
(D) Long/Short (H. Score & H. Spamicity	-0.086%	0.007	-0.372	0.711	-0.272	0.083	127
- L. Score & L. Spamicity)	(-0.14)	(0.04)	(-1.18)	(2.52)**	(-1.83)*		

Panel C. Abnormal Score x Spamicity (Range = +/−0.5)

	Alpha	Market	SMB	HML	MOM	R^2	N
Review Weighting							
(C) Long/Short (H. Score & L. Spamicity	1.228%	-0.178	0.695	0.048	-0.270	0.062	127
- L. Score & H. Spamicity)	(1.90)*	(-0.98)	(2.16)**	(0.17)	(-1.78)*		
(D) Long/Short (H. Score & H. Spamicity	0.467%	0.166	-0.239	0.654	-0.198	0.039	127
- L. Score & L. Spamicity)	(0.61)	(0.77)	(-0.63)	(1.93)*	(-1.11)		
Equal Weighting							
(C) Long/Short (H. Score & L. Spamicity	0.904%	-0.203	0.645	0.395	-0.246	0.073	127
- L. Score & H. Spamicity)	(1.67)*	(-1.33)	(2.39)**	(1.64)	(-1.94)*		
(D) Long/Short (H. Score & H. Spamicity	0.237%	-0.032	-0.393	0.574	-0.287	0.071	127
- L. Score & L. Spamicity)	(0.37)	(-0.18)	(-1.23)	(2.01)**	(-1.90)*		

*** p<0.01, ** p<0.05, * p<0.1

Table 1.A5. Score and Spamicity Portfolio Returns - Subsample Analysis

In this table, I report calendar-time 3 x 3 x 2 tercile portfolio regression estimates. *Positive Spamicity* indicates the spamicity estimate of a review if it receives scores of 4 or above and 0 otherwise. *Negative Spamicity* indicates the spamicity estimate of a review that takes the value of 1 if it receives scores of 2 or below and) otherwise. For each month, sample stocks are first divided into high- and low-characteristic groups (*Negative Spamicity/Positive Spamicity and Abnormal Peer Score*) and then are double-sorted into tercile portfolios based on abnormal scores (*Score*) and positive or negative spamicity (*Spamicity*). The dependent variable in each regression is excess portfolio returns in the following month. Following Huang (2018a), stocks in each portfolio are weighted equally or weighted by the number of reviews posted in each month (only review-weighted Fama-French-Carhart four-factor model alpha estimates are reported below). *Long/Short* indicates a spread portfolio that buys the top tercile score and bottom tercile spamicity portfolio and sells the bottom tercile score and top tercile spamicity portfolio for (C), and a spread portfolio that buys the top tercile score and spamicity portfolio and sells the bottom tercile score and spamicity portfolio for (D). Numbers in parentheses indicate t-statistics. (H. = High, L. = Low)

Panel A. High and Low Positive/Negative Spamicity

	Positive Spamicity		Negative Spamicity	
(C) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	1.301% (1.94)*	0.658% (0.77)	0.051% (0.74)	0.413% (0.69)
(D) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	0.067% (0.09)	0.368% (0.44)	-0.140% (-0.34)	0.495% (0.63)
Negative Spam† / Positive Spam‡	Low†	High†	Low‡	High‡

Panel B. High and Low Peer Scores

	Positive Spamicity		Negative Spamicity	
(C) Long/Short (H. Score & L. Spamicity - L. Score & H. Spamicity)	0.213% (1.00)	0.787% (1.05)	0.243% (0.73)	0.640% (1.41)
(D) Long/Short (H. Score & H. Spamicity - L. Score & L. Spamicity)	-0.127% (-0.20)	0.026% (0.03)	-0.018% (-0.46)	-1.457% (-2.31)**
Abnormal Peer Score	Low	High	Low	High

CHAPTER 2: PERSISTENT OPERATING LOSSES, CASH HOLDINGS, AND AGENCY PROBLEMS: THE GLOBAL EVIDENCE

2.1 Introduction

Considerable public attention has focused on ever-increasing corporate cash holdings in the U.S. Theories abound to explain why firms maintain excessively high levels of cash holdings. There is little evidence, however, that pertains to the underlying reasons for the ever-increasing trend in cash holdings, either in the U.S. or the rest of the world. A recent study by Denis and McKeon (2018) provides new evidence of compositional changes in U.S. public firms, finding that there has been a substantial increase in the number of firms experiencing persistent operating losses ("OL" or "OL firms" hereafter). They find that OL became more persistent especially after the late 1990s and these losses are concentrated among R&D-intensive firms. Their findings suggest that the immediate need for operating cash may be the most prominent driver among U.S. firms that continue to maintain high cash holdings. Little is known, however, about how this new evidence can expand our understanding of corporate cash holdings policy and how unique the observed trend in OL is in the U.S. In this paper, I investigate whether the increasing number of firms experiencing OL is a global phenomenon and explore how these compositional changes in the number of OL firms can affect the dynamics of corporate cash holdings and other corporate policies globally.

Using cash-flow data from 47 countries, I find that the increasing number of OL firms is indeed a global phenomenon. Between 1980 and 2015, the proportion of OL firms in the pooled sample increased from 3.42% to 21.70% (a 635% increase). Excluding the U.S., the proportion increased from 3.87% to 20.64% (a 533% increase). This growing trend in the proportion of OL firms seemingly jumps around 1997-1998, when large-scale economic crises struck Asia and Russia. Along with the growth in the number of OL firms, increases in industry-level cash flow volatility and R&D expenditures show seemingly similar increasing trends, especially after the late 1990s. I also find an asymmetric relationship between the level of cash holdings and operating

cash flows that is similar to that found in Denis and McKeon (2018). The magnitude of the negative relationship between operating cash flow and the level of cash holdings when a firm suffers OL is 2–3 times higher than the positive relationship between these two variables when a firm does not suffer OL.

What are the implications of OL for corporate cash holdings? Since OL became more persistent, firms that suffer from them need to maintain large amounts of cash holdings for timely and immediate liquidity injection (Denis and McKeon 2018). Such cash holdings ostensibly ensure firms' ongoing operations and therefore function to protect firms and, indirectly, their investors' capital. Poorly governed firms that have large stockpiles of cash are, however, vulnerable to agency conflicts that often result in private benefits for managers (Jensen and Meckling 1976). Therefore, given a similar level of demand for short-term cash injections, firms that are well-governed may be able to maintain large stockpiles of cash holdings (Harford, Mansi and Maxwell 2008; Gao, Harford and Li 2013). I hypothesize that OL firms with better investor protection can maintain high levels of cash holdings. As there is a significant compositional change in OL firms around the world, I expect to find a positive correlation between OL firms' cash holdings and the quality of firm- and country-level corporate governance (Investor Protection).

To test this hypothesis, I examine the correlation between corporate governance measures and the cash holdings of OL firms. Although there is an underlying negative correlation between indicators of effective corporate governance and the cash holdings of firms in general (Dittmar, Mahrt-Smith and Servaes 2003), I find evidence that there is a positive correlation between governance and the cash holdings of OL firms across various measures of corporate governance. A one-standard-deviation increase in the Anti-Self-Dealing Index (Djankov et al. 2008) is associated with a 51.49-basis-point increase in cash over total assets, which is equivalent to 3.1% of the average global cash holdings between 1980 and 2015. I also find evidence that this positive relationship strengthens in the recent sample period, which is consistent with reported findings that the negative correlation between the level of governance and cash holdings weakens if not

reverses.[52]

The positive correlation between governance and cash holdings among OL firms is consistent with the proposition that cash holdings of such firms may be closely related to investor protection and relatively less subject to free cashflow concerns (Jensen 1986). Agency theories of cash holdings imply that it is optimal for investors to require increased dividend payouts when firms hold too much cash.[53] Thus, the fact that OL firms with better governance maintain higher levels of cash holdings compared with poorly governed firms implies that cash holdings are treated very differently in OL and non-OL firms. OL firms may maintain a high level of cash holdings for reasons discussed below.

First, OL firms may hold cash to ensure short-term solvency. The ongoing persistence of OL challenges firms in tougher business environments where debt capacity alone cannot protect them or their investors from short-term solvency problems (Lins, Servaes and Tufano 2010; Almeida et al. 2014). In such environments, it can be optimal for OL firms to maintain a high level of cash holdings to ensure timely liquidity injection (Denis and McKeon 2018). Second, investors may require OL firms to maintain a certain level of cash holdings. Firms may suffer from risk-shifting when they expect persistent OL in the future (Myers and Majluf 1984). Therefore, an optimal contract between managers and investors or lenders requires firms to maintain certain levels of cash, especially when they suffer from significant losses (DeMarzo and Sannikov 2006).[54] Third, OL firms may hold cash for future investment as a precaution (Opler et al. 1999). Due to significant and persistent risk associated with their projects, it may be optimal for OL firms to hoard cash to attain the optimal level of investment.

I investigate the above inter-related hypotheses using three sets of closely related tests. I start by looking at the correlation between cash holdings in OL firms and the probability of

[52] See Pinkowitz, Stulz and Williamson (2016), for example.
[53] In an unreported table, I find evidence that OL firms in countries with robust investor protections decrease dividends, especially in the recent sample period.
[54] Also, a recent paper by Jarrow, Krishenik and Minca (2017) builds a model in which pessimistic borrowers with heterogeneous beliefs require firms to hold cash reserves as a condition for lending. Cash holdings therefore increase the debt capacity of these firms in their model.

bankruptcy and delisting from public exchanges. I find that, although OL firms face a 60-basis-point and 1-percentage-point higher probability of bankruptcy and delisting, respectively, than non-OL firms, this discrepancy decreases significantly if OL firms maintain a sufficient level of cash holdings. A one-standard-deviation increase in cash over total assets is associated with an 18.1-basis-point greater decrease in the probability of bankruptcy and a 36.1-basis-point greater decrease in the probability of delisting compared with non-OL firms. This evidence supports the proposition that OL firms hold cash to ensure solvency.

Then, I test whether OL firms' cash holdings are positively correlated with greater reliance on borrowing and investment from creditors and investors. The results show that increases in cash holdings in OL firms are associated with next-period equity, debt, and syndicated loan issuance. I find no evidence that cash holdings in non-OL firms are similarly correlated with capital issuance. This finding is consistent with the hypothesis that investors may require firms to maintain certain levels of cash holdings as a condition for lending or investing in these firms. Also, I find evidence that increased cash holdings in OL firms is associated with an increase in investment and the dividend payout ratio. This finding supports the hypotheses that OL firms may also maintain high levels of cash holdings for timely financing needs. Overall, the empirical findings suggest that cash holdings in OL firms are strongly associated with investor protection that ensures that firm operations remain viable.

I use three identification strategies to investigate the relationship between OL and the level of cash holdings. First, I use changes in nominal corporate tax rates for sample firms as an instrumental variable for OL because exogenous changes in nominal corporate taxes can affect firms' operating cash flow. I find evidence that shocks in corporate tax rate changes channeled into firms OL increases cash holdings for OL firms in strong-governance countries.

Second, I compare cash holdings of firms in strong- and weak-governance countries around the financial crisis of 2008 as a potential shock to OL firms. I find that high levels of cash holdings in OL firms decrease sharply around 2008 while non-OL firms' cash holdings are steady around that period. I also find that, after 2009, cash holdings in OL firms increase only in strong-

governance countries. This supports my hypothesis that cash holdings in OL firms are closely related to investor protection.

Third, I use the average proportion of OL firms held by institutional owners as a source of exogenous variation in OL. I expect institutional investors that are actively improving firm governance is a possible channel for explaining how OL firms survive and sustain their businesses. If such institutions are willing to take risk and invest in OL firms that are suffering from a stretch of poor performance, firms that are in countries with more of these institutions and firms held by such institutions will be able to hold high levels of cash holdings. I find consistent evidence that OL firms in strong-governance countries where there are many independent institutional investors holding OL firms hold more cash on average. Moreover, the positive relationship between cash holdings and OL is stronger for OL firms that are held by hedge funds and activist institutions, especially in weak-governance countries. The overall evidence suggests that the presence of active institutional investors is an important channel explaining the cash holdings of OL firms.

For my final set of tests I investigate the effects of compositional changes in the proportion of OL firms on corporate cash holdings. I revisit similar tests by Dittmar, Mahrt-Smith, and Servaes (2003) and find that there is a positive correlation between the level of cash holdings in OL firms and the quality of country-level governance, especially since the late 1990s. After removing OL firms from the sample, there is still a negative correlation between cash-holding levels and country-level governance. Next, motivated by Pinkowitz, Stulz, and Williamson (2006) and Kalcheva and Lins (2007), I find evidence that cash holdings in OL firms are more valuable than those in non-OL firms, especially in good-governance countries. I find little evidence that non-OL firms that increase their cash holdings increase in value. Finally, I revisit Pinkowitz, Stulz, and Williamson (2016) to examine whether OL provides additional insight into the level of U.S. cash holdings compared with other countries. I find evidence that OL explains the high level of cash holdings in U.S. firms. After controlling for high levels of R&D expenditures and OL, U.S. firms hold less cash than firms in other countries.

My empirical findings contribute to the literature on precautionary motives underlying cash

holdings (Lins, Servaes and Tufano 2010; Bates, Kahle and Stulz 2009; Palazzo 2012; Acharya, Davydenko and Strebulaev 2012; Falato, Kadyrzhanova and Sim 2013; Pinkowitz, Stulz and Williamson 2016; Begenau and Palazzo 2018; Kahle and Stulz 2018; Graham and Leary 2018), as I provide evidence that such precautionary motives and R&D expenditures are strongly associated with OL firms that operate in countries with strong investor protections. The findings also contribute to the literature on investor protection and corporate cash holdings (Dittmar, Mahrt-Smith and Servaes 2003; Dittmar and Mahrt-Smith 2007; Kalcheva and Lins 2007; Chen, Chen, Schipper, Xu and Xue 2012; Gao, Harford and Li 2013; Nikolov and Whited 2014) insofar as I provide new evidence pertaining to the growing importance of corporate cash holdings for investor protection.

The rest of the paper proceeds as follows. In the next section, I develop the empirical hypotheses. In Section 2.3, I describe the data and variables and I present the empirical results in Section 2.4. I conclude the chapter in Section 2.5.

2.2 Corporate Cash Holdings and Investor Protection

2.2.1 Cash Holdings and the Agency Problem

Regardless of investment returns, choosing to hold cash itself is an important corporate decision in terms of firms' capital structure choices. The economic benefits of corporate cash holdings are such that firms can offset carrying costs when raising cash is costly or the opportunity cost of missing timely investments is particularly high (Opler et al. 1999). This assumes that transactional demand for cash is based mostly on future demand for cash after offsetting the costs associated with it. The precautionary motive associated with cash holdings is consistent with this view that firms hold more cash when there are strong growth opportunities with riskier cash flows.

Firms with cash holdings motivated by the abovementioned needs are often subject to agency concerns since it is unclear what the optimal level of precautionary cash holdings is. Excessive cash holdings can easily be expropriated by agents seeking to maximize their own wealth. Moreover, even if precautionary cash holdings are set at optimal levels, it is unclear how

the remaining cash is distributed once there is no longer a reason to be cautious with it. Thus, the premise that corporate cash holdings are set by managers who maximize shareholder wealth has come under scrutiny.

It is quite puzzling, however, that some scholars view increases in corporate cash holdings in the U.S., especially after the 1990s, as a sign of increased agency problems. Like firms in many other countries, American firms have strengthened investor protections via increased board monitoring and risk management. Additionally, it may not suffice to explain constantly increasing corporate cash holdings in the U.S. as a consequence of increased precautionary demand for cash. If cash holdings in the U.S. reflect future demand for cash, American firms will rely to a lesser extent on capital issuance and will benefit from better investment opportunities. Recent studies by Denis and McKeon (2018) and Kahle and Stulz (2018) show some contrary evidence that firms with high levels of cash holdings are net equity issuers in the U.S. and there are signs that American corporations suffer from low profitability.

2.2.2 OL and Investor Protection

Denis and McKeon (2018) provide evidence that corporate cash holdings in the U.S. are associated primarily with firms that suffer OL. They show that the proportion of OL firms has constantly increased and the relationship between OL and cash holdings cannot be fully explained by the precautionary motive associated with cash holdings or by R&D expenditures. This can provide a new angle on reconciling the discrepancy between observed increases in corporate cash holdings and agency concerns. If a firm suffers OL such that it requires immediate cash injections to maintain daily operations, increases in cash holdings will have little to do with precautions or excessive holdings since the transactional need for cash is paramount. Also, such firms will be less subject to agency concerns because such cash holdings will be relatively free of managerial discretion.

I hypothesize that changes in the composition of OL firms around the world also alter the degree of agency concerns associated with cash holdings. For firms with positive operating cash

flows, it is optimal for shareholder value maximization to distribute any idle cash to shareholders as dividends and to minimize precautionary cash holdings. However, if a firm suffers persistent OL, ensuring that operations continue without interruption protects shareholder value. In this case, increases in cash holdings along with reduced dividends will serve as investor protection against the conventional wisdom that increased cash holdings may be signs of poor investor protection.

I test this hypothesis that cash holdings and the dividend policies of OL firms are less subject to agency concerns than those of non-OL firms using country- and firm-level investor-protection measures. This hypothesis is not about firms with excessive cash holdings becoming less prone to agency concerns. It is about changes in the composition of firms with unequal levels of cash needs that bring about changes in corporate cash holdings and dividend policies. Therefore, I hypothesize that, for firms with better investor protection or firms in countries with robust investor protection (strong governance countries), OL will be strongly associated with more cash holdings as a better way of protecting investors.

This finding differs from findings reported in Dittmar, Marht-Smith, and Servaes (2003) that firms in strong-governance countries hold less cash. This difference may be due to the growing number of OL firms around the world, but it can be reconciled if global cash holdings and dividend policies change as the composition of OL firms changes. Therefore, if OL firms increase in number, I expect that the positive relationship between cash holdings and investor protection will strengthen along with a stronger negative relationship between dividend payouts and investor protection.

I further test whether cash holdings in firms that suffer OL indeed differ from the conventional wisdom on agency concerns and investor protection. I expect to find that OL firms with more cash holdings tend to issue more capital, invest more capital, and are less likely to default or be delisted.

2.3 Data and Variable Construction

I collect firm-level data in 47 WorldScope countries for 1980–2015. I select countries

based on data availability such that each country has observations before 1995. The 47 countries are Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Denmark, Egypt, Finland, France, Germany, Greece, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Malaysia, Mexico, Netherlands, New Zealand, Norway, Pakistan, Peru, the Philippines, Poland, Portugal, Russia, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Turkey, the United Kingdom, and the United States.

I exclude financial and utility firms (SIC code 6000-6999 and 4900-4999) and firms whose total asset value is less than 5 million dollars in 2015-dollar values. My main variable of interest is operating cash flows over total assets, where cash flows are calculated by adding non-cash charges and credits to net income. Firm-level control variables include firm size, previous ten-year Fama-French 48 industry-level operating cash-flow volatility, an R&D expenditure dummy[55] that takes the value of 1 if a firm's R&D expenditures[56] over total assets exceeds 0.02, the market-to-book ratio, capex over total assets, and the leverage ratio following Denis and McKeon (2018). Country attribute variables for capital market development measures are obtained from the World Bank. Country-level control variables, motivated by Dittmar, Mahrt-Smith, and Servaes (2003), include the total number of public firms over the population in each country, total market capitalization over GDP, and total domestic private-sector credit over GDP. Equity, corporate debt, and syndicated loan issuance data are from SDC Platinum. I aggregate proceeds from each issuance at the firm level annually.[57]

Finally, I collect country- and firm-level investor protection data from various sources. Country-level Anti-Self-Dealing index data are from Djankov et al. (2008). The anti-self-dealing index is used to measure the degree of legal protection of minority shareholders against expropriation by corporate insiders.[58] The Country-level Corporate Opacity Index and Legal Protection Indices are from Karolyi (2015). These indices are constructed using principal

[55] Using R&D/TA instead of the dummy does not change the overall empirical results.
[56] I follow Koh and Reeb (2016) in treating missing R&D figures as zeros.
[57] I include 1980s issuance data as well for the analyses. My results are robust, however, if I restrict the sample period to begin after 1990 (Kim and Weisbach 2008).
[58] My results are similar when I use an anti-director rights index as in La Porta et al. (1998).

component analysis by incorporating various research outcomes into corporate opacity and legal protection. Selection of common law countries is based on La Porta et al. (2000). Lastly, 10-scale firm-level MSCI Governance indices are obtained from MSCI, and dual-class shares, closely held shares, and ADR information are obtained from WorldScope. Detailed descriptions of the variables and their sources can be found in the Appendix II.

[Insert Table 2.1 HERE]

2.4 Empirical Results

2.4.1 Persistent and Increasing OL and Cash Holdings Around the World

In Table 2.1 I report country-by-country summary statistics for firm and country attributes in the sample for 1980–2015. I divide 47 countries in the sample into strong- and weak-shareholder-rights groups based on the median value of anti-self-dealing. Despite the stark difference in the number of firms, average firm-level attributes are quite comparable between the two groups. The average cash holdings over total assets ratio is 0.18 for strong-shareholder-rights countries while it is 0.13 for weak-shareholder-rights country. This result differs from findings reported in Dittmar, Mahrt-Smith, and Servaes (2003), who find the opposite results using 1998 data for 45 countries. The Table 2.1 statistics also show that, although the U.S. experiences high levels of cash holdings (Cash/TA) and R&D expenditures, there are countries with comparable cash holdings and R&D expenditures. The only item that stands out for the U.S. may be the number of firms in the sample. This provides an additional reason to consider non-U.S. analyses to determine whether my results are driven entirely by U.S. firms.

In the fourth column of Table 2.1 I include the dummy variable I(OL) for firms with negative operating cash flows, which takes the value of 1 if a firm's operating cash flow in a given year is negative. The country-level average thus indicates the proportion of firm-year OL in the sample. On average, 19% of the sample firms show OL. Figures 2.1 and 2.2 show the country-by-country average number of OL firms and corporate cash holdings in the 1980s, 1990s, and 2000s, respectively.

[Insert Figure 2.1 HERE]

As seen in Figure 2.1, the increasing number of OL firms in the U.S (Denis and McKeon 2018) indicates that this is not a U.S.-only phenomenon. Except for Brazil, China, Colombia, Korea, Luxembourg, Mexico, Peru, and Thailand, all other countries show sizable increases in the number of OL firms. Moreover, among these eight countries, only Korea shows a 0.3% increase in average operating cash flows between the 1980s and the 2000s. The remaining seven countries show 1.5%–7.3% decreases in average operating cash flows between the two periods even though the number of OL firms did not increase. On average, there are 3.82 times more OL firms in the 2000s compared with those in the 1980s.

[Insert Figure 2.2 HERE]

As is the case with increases in OL firms, it seems that the U.S. is not the only country that has witnessed a major increase in corporate cash holdings. The monotonal increase shown in Figure 2.1 is less obvious. However, there are eight countries that have higher average cash holdings in the 1980s than in the 2000s and there are nine countries that have higher average cash holdings in the 1990s than in the 2000s. Thus, it seems that increasing corporate cash holdings and OL are prevalent around the world and U.S. firms may represent this trend more conspicuously due to its high level of capital market development. The aggregate trends in the number of OL firms and corporate cash holdings are shown in Figures 2.2 and 2.3.

[Insert Figure 2.3 HERE]

In Figure 2.2 we see that OL has become more prevalent and persistent since the late 1990s. On average, between 2000 and 2015 nearly 21% of the publicly traded firms in the WorldScope universe are OL firms, constituting a 40% increase from the 1990–1999 period. As the number of firms suffering from OL increases, so does the number of years of continuous OL (persistency). Such persistency is also not an American-only phenomenon. On average, firms suffer from 2.54 years of persistent OL between 2000 and 2015 whereas the comparable figure was 1.70 years for

the 1990–1999 period (a 49.4% increase). The proportion of OL firms shows a steep increase between 1996 and 2000, during which the Asian financial crisis and Russian default occurred. Another noticeable jump occurs during the 2008 global financial crisis. I also find that, as the proportion of OL firms increases, so does industry-level operating cash-flow volatility, especially after 1995-1996 (unreported).

[Insert Figure 2.4 and 2.5 HERE]

As shown in Figure 2.3, the average global cash holdings also increased by 26.6% between 2000 and 2015, an increase over the 1990–1999 period (to 17.6% in the later period from 13.9%). The results reported in Figures 2.2 and 2.3 suggest that global corporate cash holdings and the proportion of OL firms are moving in tandem. Figures 2.4 and 2.5 display several examples of the evolution of cash flows and cash holdings among OL and non-OL firms.

[Insert Table 2.2 HERE]

2.4.2 Cash Holdings Among OL Firms

Increases in corporate cash holdings along with increases in the number of OL firms are global phenomena. Firms around the world now face more industry-level operating cash-flow fluctuation and suffer OL for longer periods of time. My hypothesis is that increases in cash holdings around the world are concentrated among OL firms. Therefore, I follow Denis and McKeon (2018) to test whether there is asymmetric concentration of increases in cash holdings among OL firms, using the following model:

$$Cash/TA_{i,t} = \alpha + \beta\ Cashflow_{i,t} + \gamma\ I(OL)_{i,t} + \delta\ Cashflow_{i,t} * I(OL)_{i,t} + \theta\ Controls_{i,t} + \varepsilon_{i,t} \qquad [1]$$

For each firm i in year t, $Cashflow_{i,t}$ indicates operating cash flows over total assets and $I(OL)_{i,t}$ indicates a dummy variable that takes the value of 1 if the firm suffers OL.[59] Firm-level control variables include firm size, previous-ten-year Fama-French 48 industry-level operating

[59] I(OL) takes the value of 1 for contemporaneous OL, following Denis and McKeon (2018). However, the overall empirical evidence is quantitively similar if I(OL) takes the value of 1 for up to three-year negative cash-flow persistence.

cash-flow volatility, an R&D expenditure dummy that takes the value of 1 if a firm's R&D expenditures over total assets exceeds 0.02, the market-to-book ratio, capex over total assets, and the leverage ratio, following Denis and McKeon (2018). Country-level control variables include the total number of public firms over the population in each country, total market capitalization over GDP, and total domestic private sector credit over GDP, motivated by Dittmar, Mahrt-Smith, and Servaes (2003). I expect to find δ to be negative and that OL firms accumulate significantly more cash than other firms. Country-level summary statistics for the variables are shown in Table 2.2.

[Insert Table 2.3 HERE]

In Table 2.3, I report the estimation results for equation [1]. With models (2), (4), (6) and (8) there is an asymmetric relationship between negative operating cash flows and positive operating cash flows, with cash holdings explaining mixed results obtained from models (1), (2), (5) and (7). On average, a one-standard-deviation increase in operating cash flows is associated with a 2.79-percentage-point increase in cash holdings over total assets for firms with positive operating cash flows, while it is associated with an 8.61-percentage-point decrease for firms with negative operating cash flows.

Thus, it is clear that the observed correlation between cash holdings and operating cash flows is due mostly to OL firms. Models (9), (10), (11), and (12) compare these results between U.S. and non-U.S. firms. Despite the differing results in models (7) and (9), models (8) and (10) show that asymmetric cash holdings among OL firms are not a U.S.-only phenomenon. A one-standard-deviation decrease in operating cash flows is associated with a 9.27-percentage-point increase for U.S. firms while the decrease is 8.00 percentage points for non-U.S. firms. Notably, the positive correlation between R&D expenditure over total assets and cash holdings is significant only among non-U.S. firms.[60]

[60] In an unreported table, I find that the overall results are similar in both countries with high R&D expenditures and countries with low R&D expenditures. This finding suggests that, despite differences in magnitude, changes in cash holdings along with the proportion of OL firms are conspicuous in both high- and low-R&D countries. Between 1980 and 2015, average cash holdings for high-R&D countries increased by 341% while those in low-R&D

Models (13), (14), (15) and (16) compare the results between developed and emerging economies. Even though capital markets in developed economies are arguably more highly developed than those in emerging economies, the results obtained from models (14) and (17) show that it is OL firms in developed economies that hold higher levels of cash. This result is consistent with my hypothesis that the observed negative correlation between OL and cash holdings is closely related to investor protection as opposed to agency conflicts or capital market development.

2.4.3 Investor Protection, OL, and Cash Holdings

Next, I test whether cash holdings in OL firms are associated with agency problems such that OL and cash holdings reflect weak investor protections. I estimate the following equation:

$$Cash/TA_t = \alpha + \beta\ Governance_t + \gamma\ I(OL)_t + \delta\ Governance_t * I(OL)_t + \theta\ Controls_t + \varepsilon_t \quad [2]$$

In Table 2.4, I report the results of estimating equation [2]. In Panel A, we can see that the results obtained from models (1), (3), (5), and (7) are consistent with the finding reported in Dittmar, Mahrt-Smith, and Servaes (2003) that country-level investor protection is negatively correlated with cash holdings around the world. The results obtained from models (2), (4), (6), and (8) show, however, that the relationship is reversed for OL firms such that strong-governance countries show higher cash holdings among OL firms. On average, OL firms have 1–2-percentage-point higher cash holdings when country-level investor protection measures increase by one standard deviation, while cash holdings are 0.5–1 percentage points lower for non-OL firms.

[Insert Table 2.4 HERE]

In Panel B of Table 2.4, I report estimation results obtained from equation [2] using firm-level measures of investor protection. A dual-class share dummy and closely held shares are used as a proxy for poor investor protection, while ADR indicators are used as a proxy for better investor

countries increased by 208%. During the same period, the proportion of OL firms increased by 776% and 595% for high- and low- R&D countries, respectively.

protection [papers on Dual class, closely held shares, and ADR]. The results obtained from models (1) and (2) are similar to those shown in Panel A with country-level investor protection measures. Using model (7), I find that the ADR dummy is positively correlated with cash holdings as it may be related more closely to capital increases than investor protection.

[Insert Figure 2.6 HERE]

Overall, the results I have reported are consistent with my hypothesis that cash holdings in OL firms are maintained more to protect investors than as a source of expropriation. If increased OL firms are the main drivers of the global increases in cash holdings, I would expect to find that such increases are concentrated in strong-investor-protection countries. Figure 2.6 shows that this conjecture is highly likely to be true. In Figure 2.6, I divide the sample into strong- and weak-investor-protection countries using the median anti-self-dealing index. This finding shows that substantial increases in cash holdings are indeed concentrated among OL firms in strong-investor-protection countries. The difference between cash holdings in strong- and weak-investor-protection countries among OL firms is nearly twofold while there are no seemingly diverging patterns of cash holdings among non-OL firms.

So far, I have shown evidence of globally increasing OL firms and concurrent increases in cash holdings mostly in countries with strong investor protections. If these changes in the number of OL firms shifted prominent corporate cash holdings policies around the world, I would expect to find such a shift when OL firms significantly increase in number. As seen in Figure 2.3, there seems to be a significant change in the number of OL firms around the mid-1990s. I test whether changes in the number of OL firms are closely related to the results shown in Table 2.5.

[Insert Table 2.5 HERE]

In Table 2.5, I report that the observed relationships shown in Table 2.4 are much weaker before the mid-1990s. The results obtained using models (1), (3), (5), and (7) that I report in Panel A show that the negative relationship between country-level investor protection and corporate cash holdings changes after the mid-1990s. These are consistent with my hypothesis that OL firms'

cash holdings are maintained largely to protect investors and increases in OL firms are closely correlated with observed cash holding patterns around the world. Unlike country-level investor protection measures, I find little evidence from firm-level measures.[61]

These findings are consistent with changes in cash holdings shown in Figure 2.7, where the differences in cash holdings between strong- and weak-investor-protection countries are smaller before the mid-1990s. There even appears to be a reversal among non-OL firms. This reverse relationship between strong- and weak-investor-protection countries suggests that firms with positive operating cash flows hold less cash in countries with stronger investor protection. This finding is consistent with that reported in Dittmar, Mahrt-Smith, and Servaes (2003) that excessive cash holdings are sources of agency problems. As there are more OL firms, this relationship is reversed, as expected.

In an untabulated report, I find a similar shift in dividend policies after the mid-1990s among OL firms, as shown in Table 2.5. If cash holdings in OL firms provide better protection for investors while the opposite is true for non-OL firms, I expect to find that such firms decrease dividend payouts and that this trend will change along with their cash holdings. Consistent with this hypothesis, OL firms in strong-investor-protection countries decrease dividend payouts while holding more cash. If cash holdings are maintained to harm investor value, firms in strong-investor-protection countries will show increases in dividend payouts among their firms. However, the relationship is the opposite. This finding is consistent with the proposition that cash holdings in OL firms are likely maintained for the sake of investor protection.

[Insert Table 2.6 HERE]

2.4.4 The Role of Cash Holdings among OL firms

2.4.4.1 Investor Protection

Possible evidence that cash holdings in OL firms are maintained for solvency and for the

[61] There are two possible explanations for the lack of significance of firm-level measures. One is an insufficient firm-level measure in the early sample and the other is the importance of country-level investor protection (Doidge, Karolyi, and Stulz 2007).

sake of investor protection is presented in Table 2.6. Panel A shows the negative correlation between the probability that a firm goes bankrupt or becomes liquidated in a subsequent year and the amount of cash holdings OL firms maintain. This finding shows with respect to model (1) that OL firms that accumulate enough cash are less likely to file for bankruptcy. A one-standard-deviation increase in cash holdings is associated with an 18-basis-point lower chance of going bankrupt, although OL firms are more likely to declare bankruptcy unconditionally.

Interestingly, models (2) and (4) show that, for firms with positive operating cash flows, increases in cash holdings increase the chance of filing for bankruptcy in the early sample period, while the opposite is the case in the later sample, such that firms with less cash are more likely to file for bankruptcy. I find results similar to those shown in Panel B, which is evidence of a relationship between the probability of being delisted from stock exchanges in the subsequent year and cash holdings. As seen in Panel C, I find weak evidence that high levels of cash holdings on the part of OL firms in countries with robust investor protections are positively associated with the probability of becoming a target in the subsequent year M&A.

[Insert Table 2.7 HERE]

2.4.4.2 Capital Issuance

I then test whether there is any further evidence pertaining to the relationship between cash holdings and investor protection among OL firms. In Table 2.7, I report evidence that OL firms in strong-investor-protection countries actively engage in capital issuance. In Panels A, B, and C, I report the results obtained from models (1) through (4), which show that OL firms from strong-investor-protection countries issue more equity, debt, and syndicated loans. For example, a one-standard-deviation increase in the anti-self-dealing index among OL firms is associated with 3.65, 18.18, and 3.84 percentage-point higher proceeds from equity, debt, and syndicated loan issuance, respectively.[62]

[62] These results are robust without U.S. firms. U.S. firms in the sample report capital issuance of approximately 28% for equity, 43% for debt, and 45% for syndicated loans in the sample. Moreover, results from excluding 1980s' issuance data are robust. They consist of less than 3% of my issuance sample.

Compared with non-OL firms, OL firms in the sample rely to a greater extent on equity issuance. In untabulated results, I find that 26.78% of OL firms issued equity while 13.18% of non-OL firms issued equity. In contrast, 3.10% and 3.32% of OL firms engaged in debt and syndicated loan issuance, respectively, while 7.68% and 10.20% of non-OL firms in the sample engaged in debt and syndicated loan issuance, respectively.

These findings indicate that OL firms in strong-investor-protection countries raise more capital regardless of the above three markets. Moreover, these findings also imply that OL firms not only hoard cash from their operations but also actively raise cash. Thus, another possible explanation for the observed cash holdings is that firms hoard cash for future investment opportunities. I test whether increased cash holdings among OL firms is associated with increased investment and R&D expenditures.

[Insert Table 2.8 HERE]

2.4.4.3 Corporate Investment and R&D

Consistent with the abovementioned hypothesis, I find evidence that, regardless of the degree of investor protection, OL firms with more cash holdings increase investment and R&D expenditures to a greater extent, as shown in Table 2.8, where the magnitude of the increase is lower for investment in models (1) through (4) such that OL firms cut investment to a lesser extent than non-OL firms when they hoard cash. On average, a one-standard-deviation increase in cash holdings among OL firms is associated with a 36-basis-point higher Capex/TA and 2.21 percentage points more in R&D expenditures.

[Insert Table 2.9 HERE]

2.4.5 Corporate Income Tax, the Financial Crisis of 2008, and Institutional Investors

2.4.5.1 Corporate Income Tax Rate Changes

To identify the role of OL in corporate cash holdings policies, I use changes in nominal

corporate tax rates from 28 OECD countries as a possible shock to firms' operating cash flows.[63] In Table 2.9, I report Two-Stage Least Square estimation results obtained from equation [1] without an interaction term. In Panel A I report the results of the first path regressions of cash holdings on changes in nominal corporate income tax rates among OECD countries. In models (1) and (2), I find that corporate income tax changes are negatively correlated with operating cash flows and are positively correlated with the dummy variable for OL. This finding is not surprising given that operating cash flows will be directly affected if there is an income tax rate change.

Compared with the results obtained with models (3) and (4), the signs of corporate income tax change coefficients reverse directions in models (5) and (6). This finding may reflect pro-cyclical corporate income tax rate policies among these countries. Consequently, corporate income tax changes in these countries are positively correlated with operating cash flows, making causal inferences in empirical analysis difficult.

In Panel B I report the second-stage regression estimates. I find suggestive evidence that OL is associated with increases in cash holdings. In models (7) and (8) there is a negative correlation between operating cash flows and cash holdings, as reported in Tables 2.3 and 2.4. On average, a one-standard-deviation increase in operating cash flows is associated with a decrease in cash holdings of 14.21 percentage points. The greater economic significance comes from restricting the sample countries to 28. The difference between strong-investor-protection countries and weak-investor-protection countries is consistent with my hypothesis that cash holdings among OL firms are maintained more for investor protection than for other purposes.

[Insert Figure 2.7 HERE]

2.4.5.2 The Financial Crisis of 2008

Next, because OL must be closely related to macro-economic conditions, large-scale crises may provide an additional set of opportunities to identify the impact of OL on cash holdings. In Figure 2.7, I show evidence that, around the financial crisis of 2008, OL firms dramatically

[63] I assume that nominal changes in corporate income tax rates represent an exogenous shock compared with firms' effective tax rates, since firms may choose optimal tax brackets.

changed their cash holdings. This financial crisis provides an opportunity to see differential changes in cash holdings among OL firms as it is a relatively isolated shock compared with other global financial and economic crises clustered around the mid-1990s and early 2000s, when several crises played out over longer periods of time.

Before the end of 2007, there were seemingly conspicuous parallel trends in cash holdings for both OL firms and their counterparts. At the end of 2007, OL firms maintained higher cash holdings. The average difference in cash holdings was about 12.50%. During the crisis that struck the global economy severely in the years 2008 and 2009, OL firms consumed much of their cash holdings. Subsequently, I find that they slowly hoarded cash and the gap in cash holdings between OL firms in strong- and weak-governance countries widenened again. This pattern is non-existent, however, among non-OL firms. The gap in cash holdings between non-OL firms in strong- and weak-governance countries rather slightly narrows after 2008. I formally test these results and find that the observed widening gap in cash holdings among OL firms is statistically and economically significant.[64]

[Insert Table 2.10 HERE]

2.4.5.3 Institutional Investor Channel

One channel that might drive the observed negative relationship between OL and corporate cash holdings policy is the presence of informed institutional investors who are willing to inject sufficient capital into OL firms to ensure improved governance in such firms. The presence of such investors will help OL firms raise additional capital and maintain optimum levels of investment and operations. To test whether the presence of institutions affects the cash holdings of OL firms, I collect institutional holdings data from Factset Lionshares and measure the ratio of the number of OL firms to the number of OL firms held by institutional investors in each country. If an OL firm is located in countries where there are relatively more institutional investors who are willing to invest in OL firms and improve governance, it has a higher chance of acquiring the necessary

[64] See Table 2.A1.

capital to sustain its business. As a result, there will be a positive correlation between the intensity of institutional investors investing in OL firms and negative cash flows in those firms (an OL dummy).

In Table 2.10 I report Two-Stage Least Squares estimation results that are similar to those reported in Table 2.9 using the ratio of the number of OL firms held by several types of institutions to the number of non-OL firms held by institutions of the same types in each country at *t* as a possible instrument for the negative relationship between OL and cash holdings policy.[65] In models (1) thru (4), as seen in Panel A, the proposed instrument has the expected positive correlation with the OL dummy. The stark difference between independent institutions and grey institutions in models (2), (3), (5), and (6) is consistent with the hypothesis that institutions that more actively affect firms' policies and governance may be the source of the observed negative relationship between OL and cash holdings policy.

As shown in Panel B, in countries with more such institutions the relationship between OL and cash holdings is economically and statistically significant. On average, OL firms located in strong-governance countries where there are many independent institutions investing in OL firms maintain 15.2% higher cash holdings than non-OL firms while the figure is 10.7% higher for those in weak-governance countries (a statistically insignificant result).

[Insert Table 2.11 HERE]

To identify the role of institutional investors in the corporate cash holdings policies of OL firms more definitively, I measure firm-level exposure to institutions that are willing to invest in and possibly affect the governance of OL firms. To create the measure, I first calculate the number of OL firms over the total number of firms held by each institutional investor as a proxy for their willingness to invest in OL firms. I then average this proxy at the firm level and use it as a possible instrument for the negative relationship between OL and cash holdings policy.[66] Compared with

[65] Instrumenting cash flows using the same instruments yields results that are quantitatively equivalent to those shown in Table 2.9 (unreported).
[66] Instrumenting cash flows using the same instruments yields results that are quantitatively equivalent to those reported in Table 2.9 (unreported).

the instrument used in Table 2.10, this instrument is expected to measure each OL firm's exposure to influence from institutional investors.

In addition, to investigate the role of institutions that are more active in terms of improving firm governance, I focus on institutions that identified as hedge funds and also on activist institutions that have filed Form 13D in the U.S. If institutional investors are rather active in terms of affecting OL firms' cash holdings policies and other corporate decisions, as seen in Tables 2.7 and 2.8, I expect to find stronger results among OL firms that are held by such institutions.

In Table 2.11 I report 2SLS estimation results derived from equation [1] without the interaction term, as is the case in Tables 2.9 and 10. Consistent with the above hypothesis, I find quantitatively stronger results than those reported in Table 2.10. The firm-level instrument is positively and significantly correlated with the OL dummy in strong- and weak-governance countries, as seen in Panel A. In Panel B, the reported estimated coefficients are higher than those reported in Table 2.10. Notably, models (11) and (12) imply that OL firms in weak-governance countries held by hedge funds (activist) institutions that hold many other OL firms in their portfolios hold 29.9% (68.8%) more cash holdings than non-OL firms. Moreover, I find that the observed coefficients are even higher if the instrument is calculated with institutions that hold more than 1% of a firm's shares.[67] The results reported in Table 2.11 are consistent with the hypothesis that institutional investors who actively influence corporate policies and governance play an important role in the observed negative relationship between OL and cash holdings polices.

[Insert Table 2.12 HERE]

I further test whether institutional investors play a role in the link between OL and cash holdings by measuring differences in levels of investor protection between institutional investors' domicile countries and firms' domicile countries. Institutions from stronger-governance countries would require higher levels of protection when they invest in firms in weaker-governance countries. Results reported in Table 2.10 and 2.11 suggest that the governance spillover from

[67] The results are stronger if I use the 5% threshold instead of the 1% (unreported).

institutions into firms' investor protections is also an important force in cash holdings among OL firms.

The results reported in Table 2.12 indicate that differences in governance levels between firms and their institutional investors increase the cash holdings of OL firms.[68] Using models (1) thru (5), I find that, after controlling for country-level governance, the governance level of a firm's institutional investors is positively correlated with cash holdings, especially among OL firms. Overall, a one-standard-deviation increase in the governance difference between institutions and firms is associated with a 1.43-percentage-point increase in cash holdings. The effect is amplified among OL firms such that there is a 3.23-percentage-point increase in cash holdings.

The results derived with models (1) through (5) are comparable to those reported in Table 2.4. To verify whether the influence of institutions on the cash holdings of OL firms reflects possible governance spillover, I further interact the firm–institution governance difference with country-level governance were firms are located. If the observed effects derived with models (1) through (5) capture hidden aspects of firm-level governance that is not captured by country-level governance, as in Table 2.4, I expect to find positive coefficients of interactions between the firm–institution governance difference and firm-level governance, especially among firms in strong-governance countries.

Results derived from models (7) through (9) show that the effects of the governance difference between institutions and OL firms is stronger in countries with weak investor protections, indicating a possible governance spillover effect. This effect is, however, reversed among firms in strong-governance countries, indicating possible substitution between the two measures. Overall, the results reported in Tables 2.10, 2.11, and 2.12 are consistent with the hypothesis that institutions from stronger-governance countries may seek additional protection that is reflected in cash holdings from firms in weak-governance countries.

[68] Most institutional investors are from the U.S. As a result, a U.S.-firms-only model does not allow for enough variation for higher-level interactions and thus are omitted from Table 2.10.

2.4.6 Consequences of Compositional Changes between OL and Non-OL Firms

In the final set of tests I run, I investigate the implications of compositional changes in the proportion of OL firms for previously documented findings on corporate cash holdings by closely following their empirical tests. I revisit Dittmar, Mahrt-Smith, and Servaes (2003: DMS2003) and find that there is a positive correlation between the level of cash holdings in OL firms and the quality of country-level governance, especially after the late 1990s. After removing OL firms from the sample, there is a stronger negative correlation between cash-holding levels and country-level governance.[69]

Next, motivated by Pinkowitz, Stulz, and Williamson (2006: PSW2006) and Kalcheva and Lins (2007: KL2007), I find evidence that cash holdings in OL firms are more valuable than those in non-OL firms, especially in strong-governance countries. As I report in Panel A of Table 2.A3, OL firms' cash holdings are more valuable than those of non-OL firms, especially in the later sample period. For instance, a one-standard-increase in the Anti-Self-Dealing index (0.165) is associated with an increase in OL firms' Tobin's Q of 18.5 (= 1.112 x 0.165) percentage points in model (11), while it is 11.9 (= 0.720 x 0.165) for non-OL firms. As can be seen in Panel B of Table 2.A3, I find little evidence that non-OL firms that increase their cash holdings increase in value in the later sample period while there is a sizable increase in firm value associated with OL firms in strong-governance countries.

Finally, I revisit Pinkowitz, Stulz, and Williamson (2016: PSW2016) to examine whether OL provides additional insights into the level of U.S. cash holdings compared with those in other countries. As can be seen in Panel A of Table 2.A4, I find consistent evidence that U.S. firms with low R&D expenditures hold less cash than non-U.S. firms and this effect is more pronounced if they engage in R&D spending. A similar correlation is observed in Panel B where we can see that high cash holdings in U.S. firms are concentrated among OL firms. Taking the results reported in Panel C together, OL firms in the U.S. hold significantly more cash even among firms with high

[69] See Table 2.A2.

R&D expenditures. The overall empirical findings suggest that changes in the composition of OL firms around the world are either changing the previously documented findings on corporate policies not by altering the underlying economics but by infusing heterogeneity in the sample firms that are governed by different dynamics or simply driving the findings.

2.4.7 Robustness

I further test whether my results are driven primarily by expropriation tax effects and foreign cash, as in Foley et al. (2007) and Faulkender, Hankins, and Petersen (2017). To do so, I subdivide the sample firms based on median foreign income tax. In an unpublished table, I present similar results in both high- and low-foreign-income-tax firms, as reported in Table 2.2.

I also test whether the observed changes in cash holdings and OL are driven by bias in the sample period, firms, or countries. First, I re-run my tests using earlier sample periods to verify whether the overall findings are clustered only in the recent sample period. I find that the asymmetric relationship between cashflow and cash holdings also exist in the earlier sample period although it is much weaker in terms of economic and statistical significance.[70] This suggests that it is more about the compositional changes in OL that affects the aforementioned dynamics.

Second, I re-run my results using only U.S. firms that also appear in Compustat. Woldscope coverage changes significantly in recent periods, when it covers smaller firms than are in the Compustat universe. Since U.S. firms dominate the sample, as is the case with the results reported in Table 2.2, I verify whether the overall findings are driven by the sample heterogeneity.[71] I find comparable results across two data sources. Third, I restrict the sample countries to those that also appear in the earlier sample to verify whether the results are driven by countries that appear only in the recent sample period.[72] The overall results do not change when I remove such countries from the sample.

[70] See Table 2.A5.
[71] See Table 2.A6.
[72] See Table 2.A7.

2.5 Conclusion

In this paper, I show evidence that increases in cash holdings and OL are global phenomena. Firms around the world held around 1.35 times more cash in the 2000s than in the 1980s and the number of OL firms increased by 3.5 times between the two periods. These two distinctive phenomena seem to be inter-connected such that firms with OL hoard asymmetrically more cash than non-OL firms. R&D expenditures alone do not seem to explain the observed global changes in cash holdings.

I also find evidence that the positive correlation between cash holdings and OL is heavily concentrated in countries with higher-quality investor protections. The correlation is nearly non-existent in weak-investor-protection countries. Moreover, this positive correlation seems to be observable after the mid-1990s when at least two global financial crises occurred. I find that industry-level operating cash flows increased constantly and OL firms exploded after the mid-1990s along with cash holdings. This finding suggests that there was a significant shift in the composition of firms and their corporate policies. Before the mid-1990s, countries with high-quality investor protections had lower average cash holdings but this relationship reversed after the mid-1990s. Additionally, cash holdings were positively correlated with the probability of default and the probability of being delisted from public exchanges, as these relationships flipped after the mid-1990s.

One possible channel through which OL firms can sustain their businesses is the presence of active institutional investors. The evidence shows that OL firms, even in weak-governance countries, maintain high levels of cash holdings if they are held by hedge funds and active investors. This finding supports the idea that institutional investors play an active role in the overall corporate policies of OL firms. This implies that possible governance spillover from such institutions can be crucial for OL firms' financing and investment activities. Although there is no direct causal evidence that links the increasing number of firms with persistent OL with the role of institutional investors, the increasing number of OL firms signifies the importance of such investors.

My overall findings suggest that cash holdings in OL firms may play a role in protecting investors against the conventional wisdom under agency conflicts. Overall, my findings are consistent with my hypothesis that cash holdings have positive value for investor protection, especially when OL firms are generally well-protected yet suffer from OL. I also find, however, that findings such as those reported in Dittmar, Marht-Smith, and Servaes (2003) remain intact in my sample. Thus, it seems that the observed negative correlation between cash holdings and OL is due mainly to compositional changes in OL firms rather than agency concerns, as excessive cash holdings per se weakened in the later sample.

2.6 References

Acharya, Viral, Sergei A. Davydenko and Ilya A. Strebulaev, 2012, "Cash Holdings and Credit Risk", *The Review of Financial Studies* 25 (12), 3572-3609.

Almeida, Heitor, Murillo Campello, Igor Cunha and Michael S. Weisbach, 2014, "Corporate Liquidity Management: A Conceptual Framework and Survey", *The Annual Review of Financial Economics* 6 (1), 135-162.

Brav, Alon, Wei Jiang, Frank Partnoy and Randall Thomas, 2008, "Hedge Fund Activism, Corporate Governance, and Firm Performance", *The Journal of Finance* 63 (4), 1729-1775.

Bates, Thomas W., Kathleen M. Kahle and René M. Stulz, 2009, "Why Do U.S. Firms Hold So Much More Cash than They Used To?", *The Journal of Finance* 64 (5), 1985-2021.

Begenau, Juliane and Berardino Palazzo, 2018, "Firm Selection and Corporate Cash Holdings", *working paper*. SSRN: https://ssrn.com/abstract=2782095

Chen, Qi, Xiao Chen, Katherine Schipperm Yongxin Xu and Jian Xue, 2012, "The Sensitivity of Corporate Cash Holdings to Corporate Governance", *The Review of Financial Studies* 25 (12), 3610-3644.

DeMarzo, Peter M. and Yuliy Sannikov, 2006, "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model", *The Journal of Finance* 61 (6), 2681-2724.

Denis, David J., Stephen B. Mckeon, 2018, "Persistent Operating Losses and Corporate Financial Policies", *working paper*. SSRN: https://ssrn.com/abstract=2881584

Dittmar, Amy, Jan Mahrt-Smith and Henri Servaes, 2003, "International Corporate Governance and Corporate Cash Holdings, *Journal of Financial and Quantitative Analysis* 38 (1), 111-133.

Dittmar, Amy and Jan Mahrt-Smith, 2007, "Corporate governance and the value of cash holdings", *Journal of Financial Economics* 83 (3), 599-634.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer, 2008, "The law and economics of self-dealing", *Journal of Financial Economics* 88 (3), 430-465.

Doidge, Craig, G. Andrew Karolyi and René Stulz, 2007, "Why do countries matter so much for corporate governance?", *Journal of Financial Economics* 86 (1), 1-39.

Faulkender, Michael W., Kristine W. Hankins and Mitchell A. Petersen, 2017, "Understanding Precautionary Cash at Home and Abroad", *working paper*.

Falato, Antonio, Dalida Kadyrzhanova and Jae Sim, 2013, "Rising Intangible Capital, Shrinking Debt Capacity, and the US Corporate Savings Glut", *working paper*. SSRN: https://ssrn.com/abstract=2350863

Ferreira, Miguel A and Pedro Santos, 2008, "The colors of investors' money: The role of institutional investors around the world", *Journal of Financial Economics* 88 (3), 499-533.

Foley, C. Fritz, Jay C. Hartzell, Sheridan Titman and Garry Twite, 2007, "Why do firms hold so much cash? A tax-based explanation", *Journal of Financial Economics* 86 (3), 579-607.

Gao, Huasheng, Jarrad Harford and Kai Li, 2013, "Determinants of corporate cash policy: Insights from private firms", *Journal of Financial Economics* 109 (3), 623-639.

Graham, John R. and Mark T. Leary, 2018, "The Evolution of Corporate Cash", *The Review of Financial Studies*, forthcoming.

Harford, Jarrad, Sattar A. Mansi and William F. Maxwell, 2008, "Corporate governance and firm cash holdings in the US", *Journal of Financial Economics* 87 (3), 535-555.

Jarrow, Robert, Andrey Krishenik and Andreea Minca, 2017, "Optimal cash holdings under

heterogeneous beliefs", *Mathematical Finance*, 1-36.
Jensen, Michael C. and William H. Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics* 3 (4), 305-360.
Jensen, Michael C., 1986, "Agency Cost of Free Cash Flow, Corporate Finance, and Takeovers", *The American Economic Review* 76 (2), 323-329.
Kahle, Kathleen M. and René M. Stulz, 2018, "Is the American public corporation in trouble?", *The Journal of Economic Perspectives*, forthcoming.
Kalcheva, Ivalina and Karl V. Lins, 2007, "International Evidence on Cash Holdings and Expected Managerial Agency Problems", *The Review of Financial Studies* 20 (4), 1087-1112.
Karolyi, Andrew, 2015, "Cracking the Emerging Markets Enigma", Oxford University Press, New York.
Kim, Woojin and Michael S. Weisbach, 2008, "Motivations for public equity offers: An international perspective", *Journal of Financial Economics* 87 (2), 281-307.
Koh, Ping-Sheng and David M. Reeb, 2015, "Missing R&D", *Journal of Accounting and Economics* 60 (1), 73-94.
Nikolov, Boris and Toni M. Whited, 2014, "Agency Conflicts and Cash: Estimates from a Dynamic Model", *The Journal of Finance* 69 (5), 1883-1921.
La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny, 1998, "Law and Finance", *The Journal of Political Economy* 106 (6), 1113-1155.
La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny, 2000, "Investor Protection and corporate governance", *Journal of Financial Economics* 58 (1-2), 3-27.
Lins, Karl V., Henri Servaes and Peter Tufano, 2010, "What drives corporate liquidity? An international survey of cash holdings and lines of credit", *Journal of Financial Economics* 98 (1), 160-176.
Myers, Stewart C. and Nicholas S. Majluf, 1984, "Corporate Finance and Investment Decisions When Firms Have Information That Investors Do Not Have", *Journal of Financial Economics* 13 (2), 187-221.
Opler, Tim, Lee Pinkowitz, René Stulz and Rohan Williamson, 1999, "The determinants and implications of corporate cash holdings", *Journal of Financial Economics* 52 (1), 3-46.
Palazzo, Berardino, 2012, "Cash holdings, risk, and expected returns", *Journal of Financial Economics* 104 (1), 162-185.
Pinkowitz, Lee, René Stulz and Rohan Williamson, 2006, "Does the Contribution of Corporate Cash Holdings and Dividends to Firm Value Depend on Governance? A Cross-country Analysis", *The Journal of Finance* 61 (6), 2725-2751.
Pinkowitz, Lee, René Stulz and Rohan Williamson, 2016, "Do U.S. Firms Hold More Cash than Foreign Firms Do?", *The Review of Financial Studies* 29 (2), 309-348.

2.7 Figures and Tables

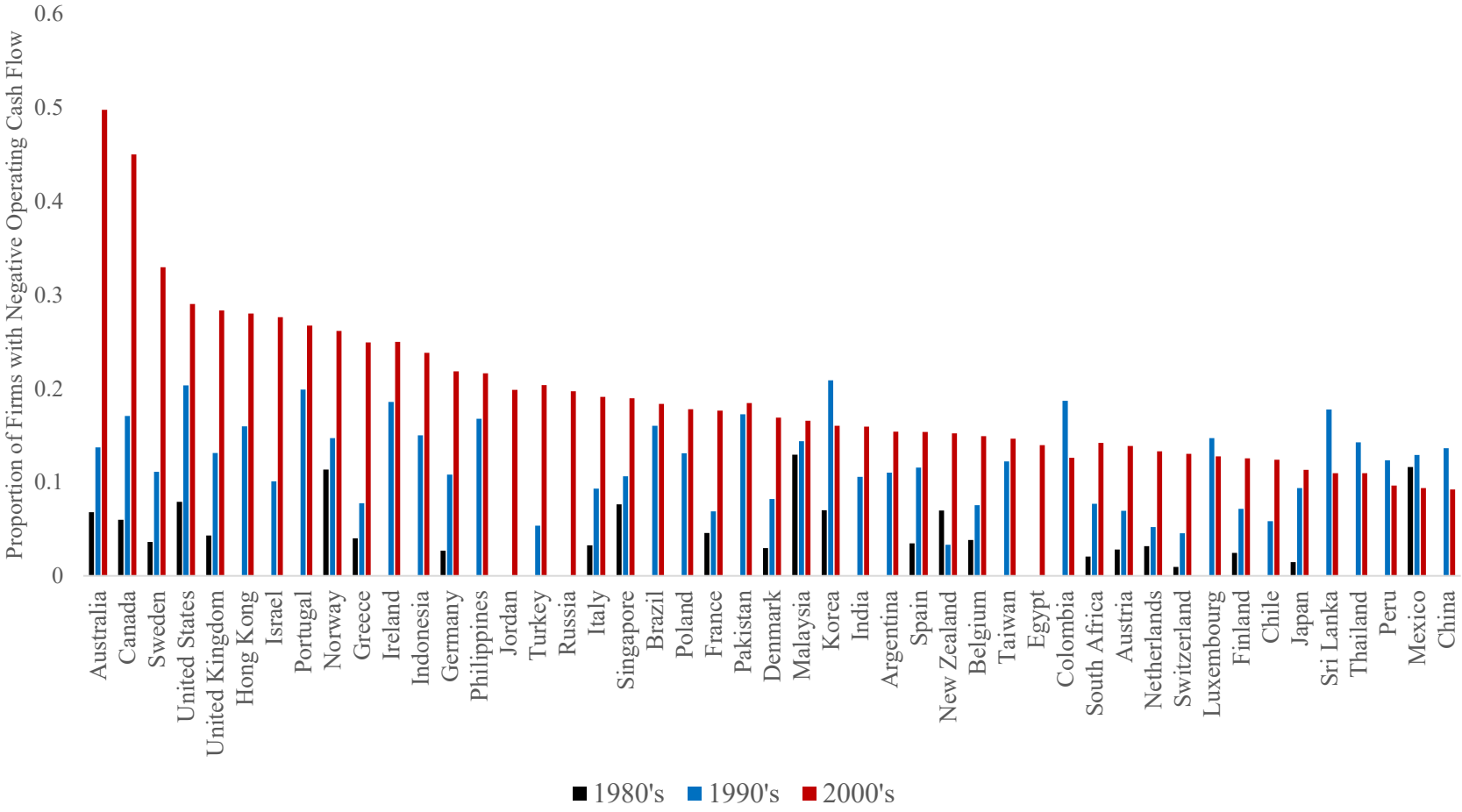


Figure 2.1. The Proportion of Negative Cash Flow Firms Around the World
This figure shows the proportion of firms with negative operating cash flows between 1980 and 2015. 2000s includes the 2000–2015 period.

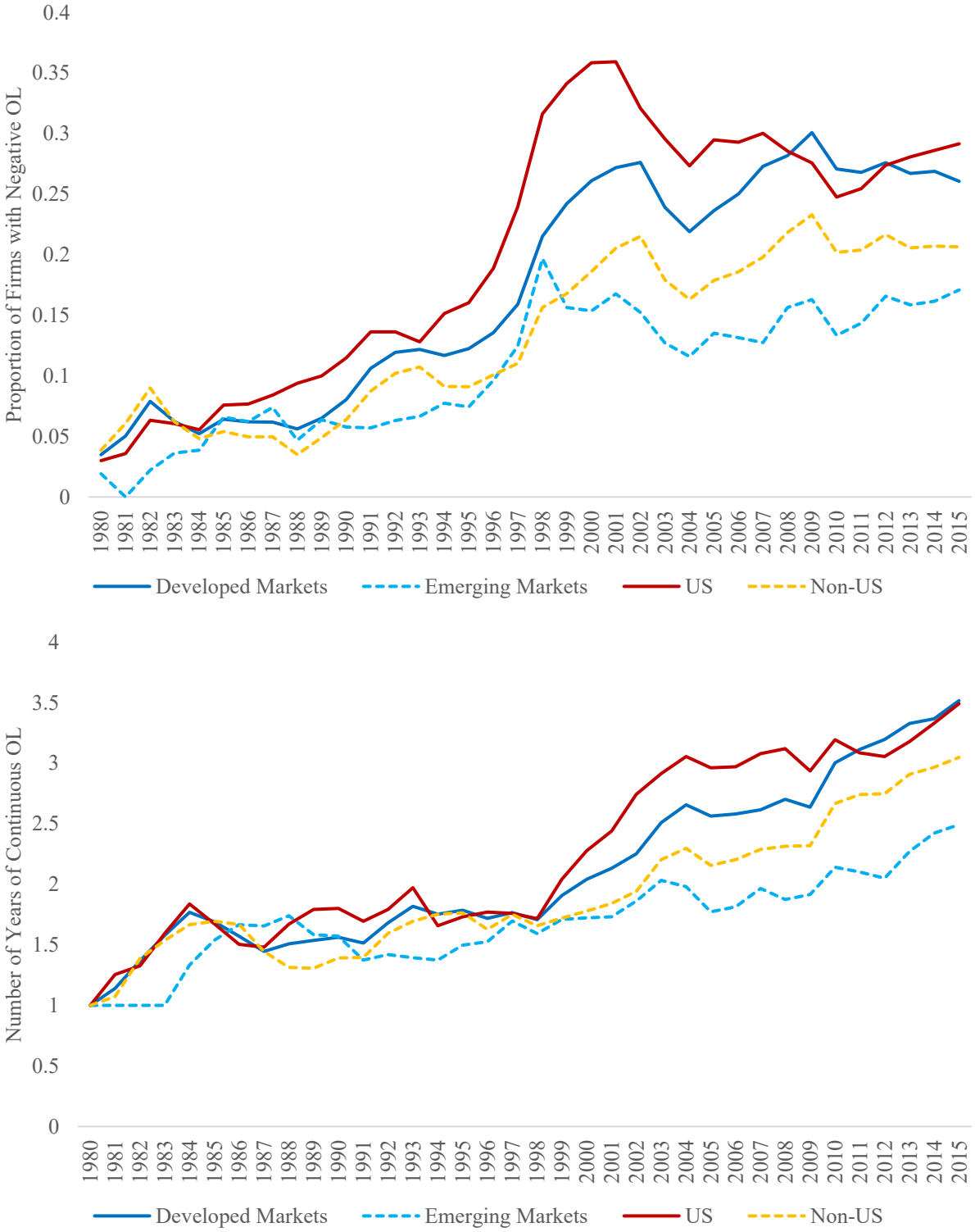


Figure 2.2. Evolution of Persistent OL around the world
This figure shows the proportion of OL firms 1980 and 2015 around the world and the persistence of OL.

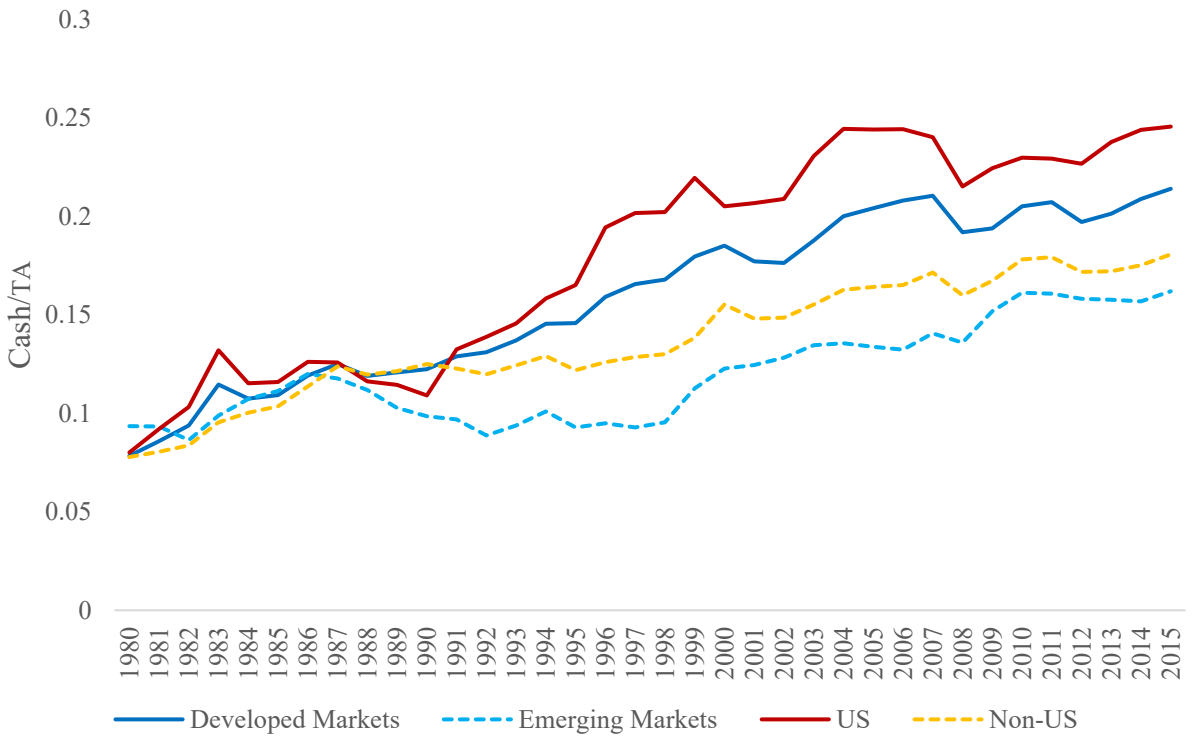


Figure 2.3. Evolution of Cash Holdings around the world
This figure shows the average Cash/TA of the sample firms around the world between 1980 and 2015.

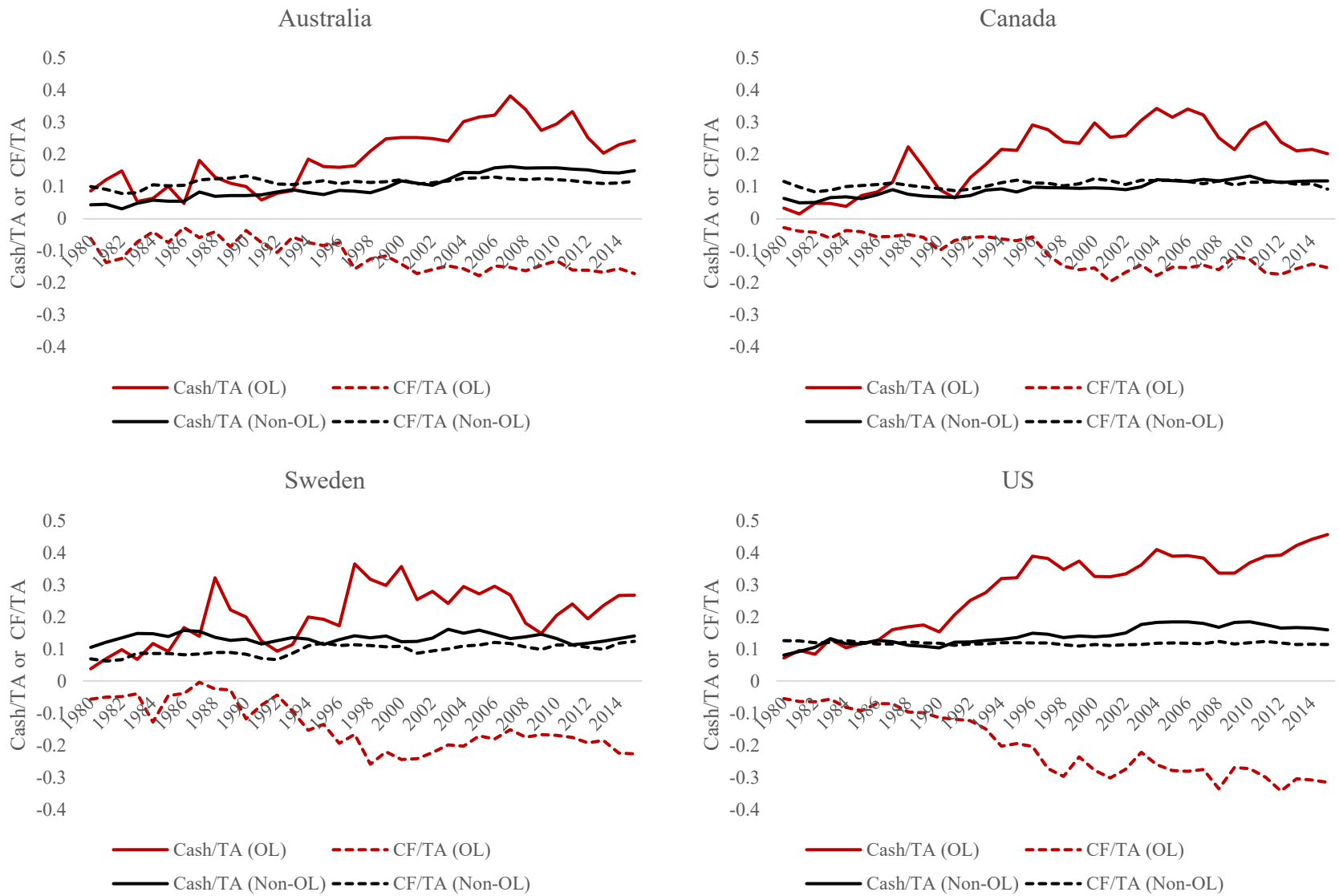


Figure 2.4. Evolution of Cash holdings in high OL Countries

These figures graph the average cash holdings and operating cash flows of top 4 OL countries based on 2000–2015 average proportions of country-level OL.

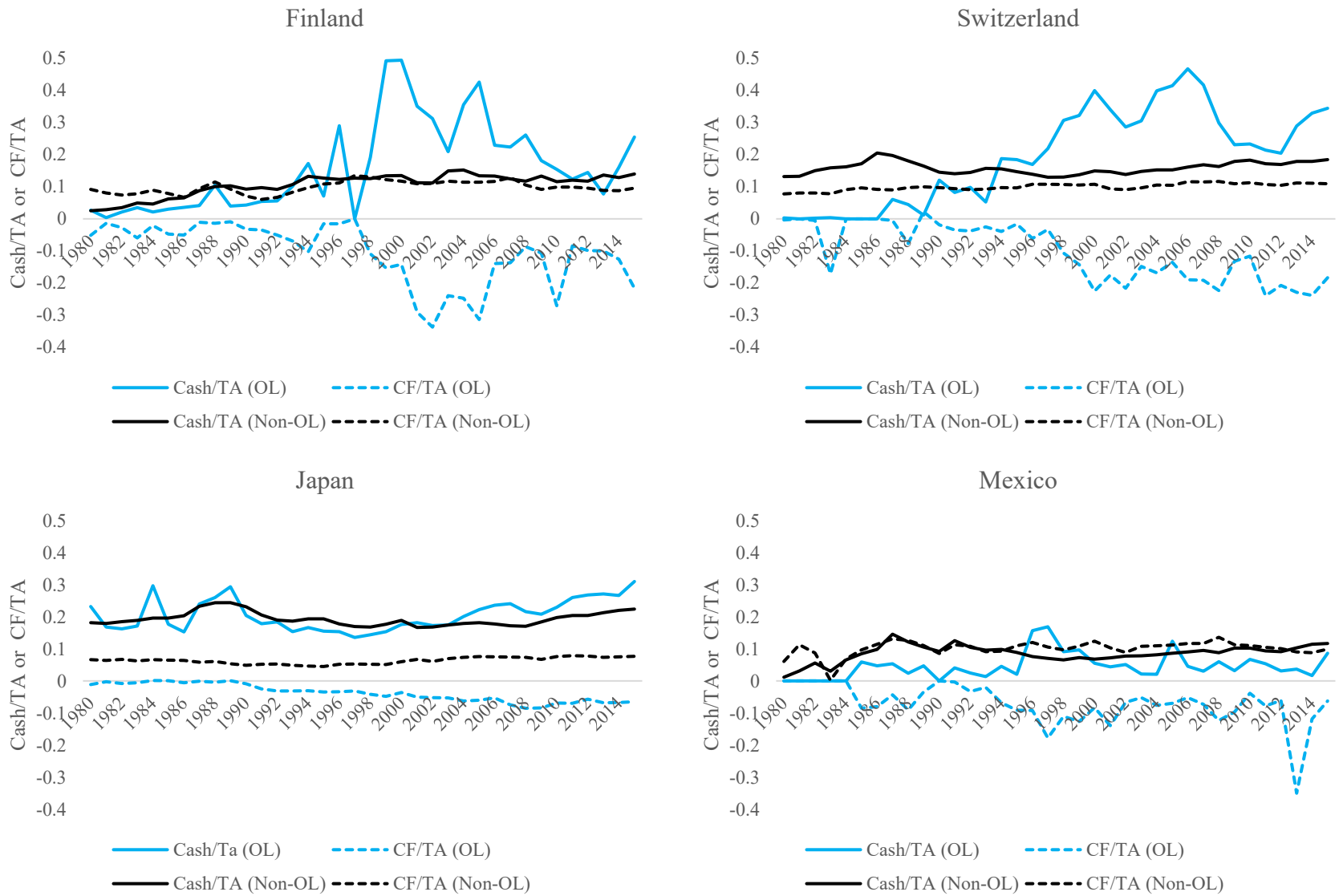


Figure 2.5. Evolution of Cash holdings in low OL Countries

These figures graph the average cash holdings and operating cash flows of the bottom four OL countries based on 2000–2015 average proportions of country-level OL.

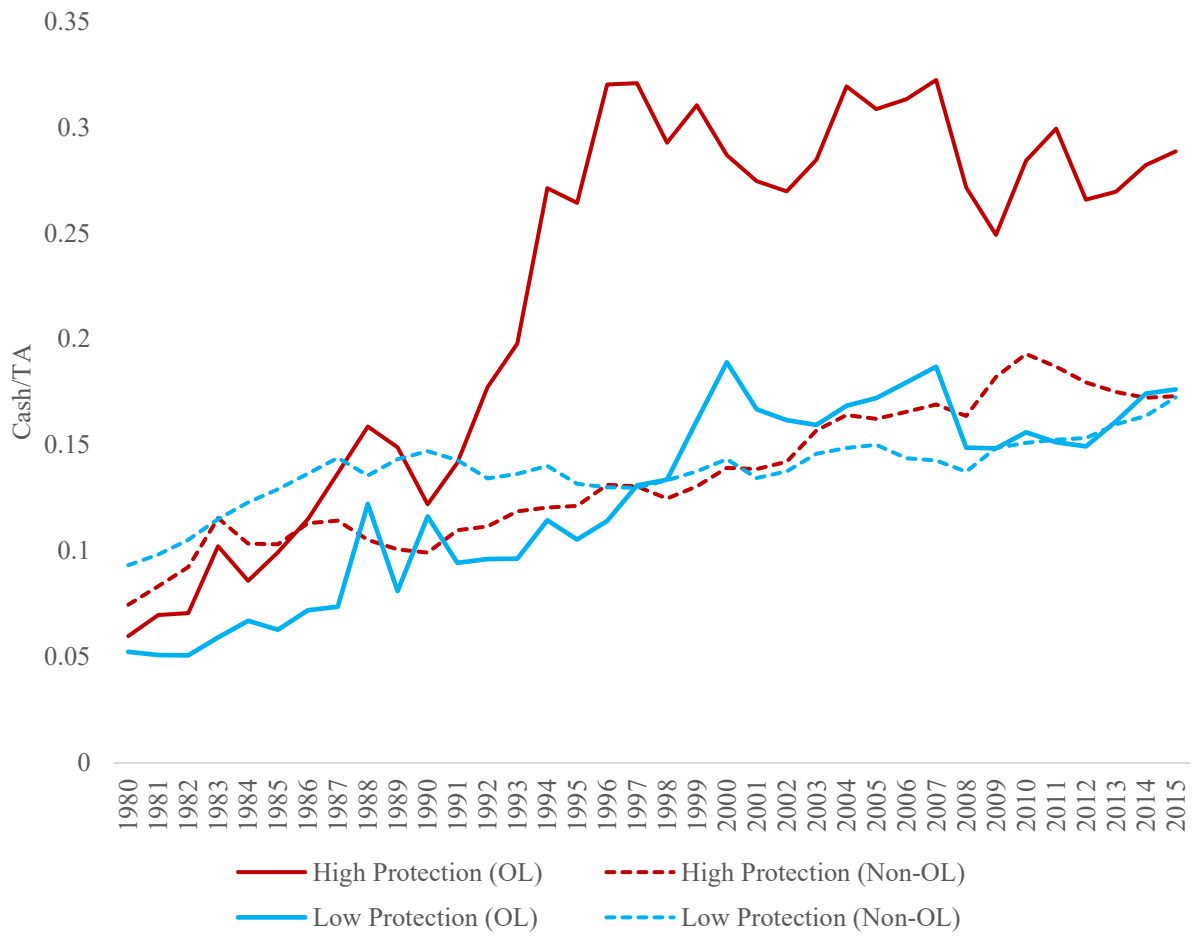


Figure 2.6. Evolution of Cash holdings and Country-Level Investor Protections
This figure shows the average cash holdings of firms based on country-level investor protection and OL. Strong protection indicates countries with above -average Anti-self-dealing index scores.

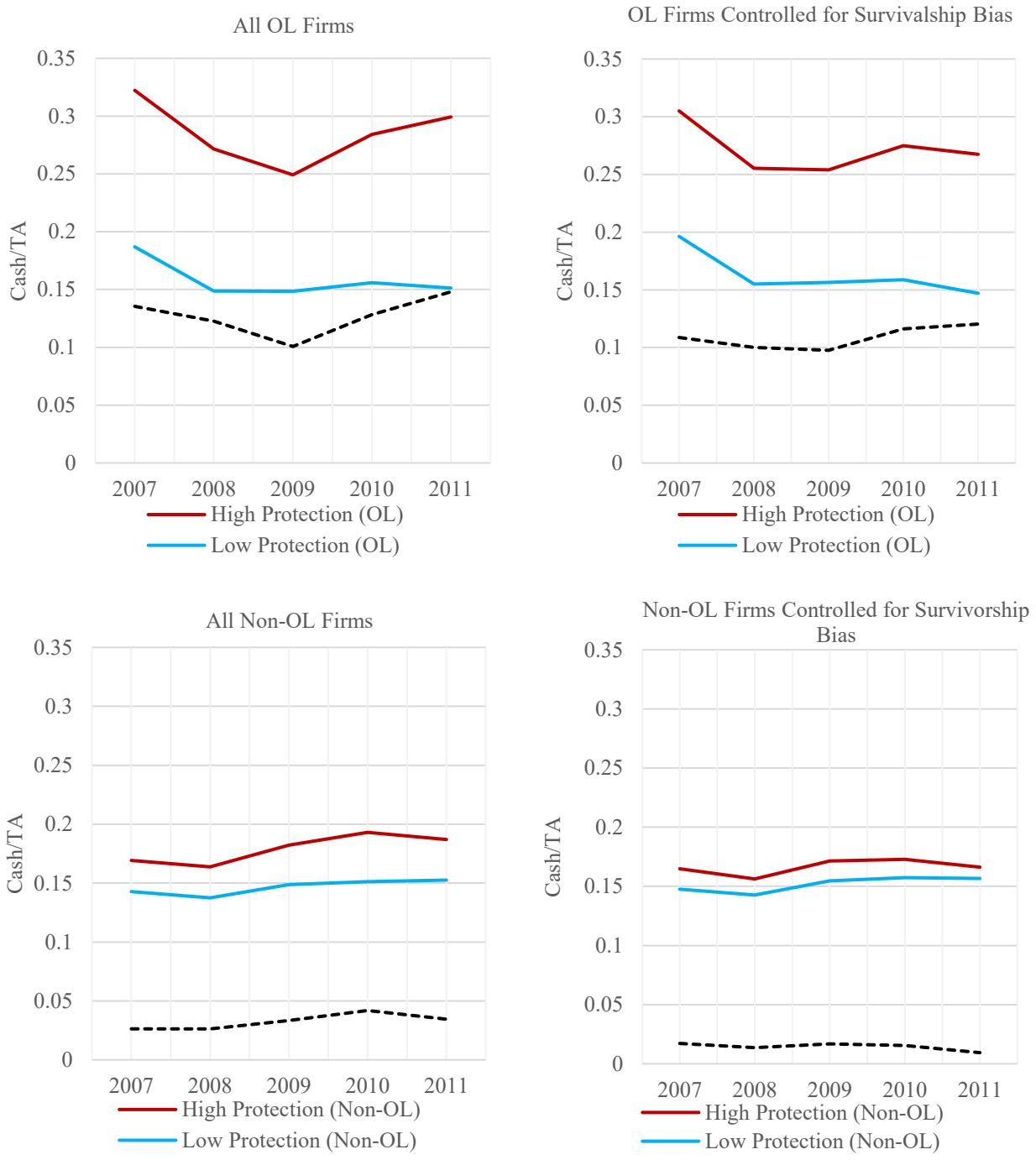


Figure 2.7. Cash Holdings, OL and Investor Protection Around the Financial Crisis
This figure shows the average cash holdings of firms based on country-level investor protections and OL around the 2008 crisis. Strong protection indicates countries with above-average Anti-self-dealing index scores. "Controlled for Survivorship Bias" includes only firms that appear entirely between 2007 and 2011.

Table 2.1. Compositional Changes in OL Firms

In this table, I report the average proportion of OL firms and their operating cash flows between 1980 and 2015 for 47 countries in WorldScope. I(OL) takes the value of 1 if a firm has OL and CF indicates a firm's operating cash flows over total assets. Countries are sorted based on I(OL) in 2000s (2000–2015). CFs are winsorized at the 1% and 99% levels.

Country	N	1980's		1990's		2000's		All Years	
		I(OL)	CF	I(OL)	CF	I(OL)	CF	I(OL)	CF
Australia	19,079	0.068	0.102	0.137	0.087	0.498	-0.017	0.438	0.000
Canada	16,082	0.060	0.095	0.171	0.071	0.450	-0.006	0.339	0.024
Sweden	2,625	0.036	0.082	0.111	0.074	0.330	-0.014	0.166	0.049
United States	116,517	0.079	0.103	0.203	0.047	0.290	0.001	0.238	0.028
United Kingdom	35,144	0.043	0.114	0.131	0.080	0.284	0.025	0.201	0.055
Hong Kong	14,469			0.160	0.065	0.280	0.043	0.266	0.046
Israel	4,620			0.101	0.070	0.276	0.018	0.269	0.020
Portugal	1,080			0.199	0.061	0.267	0.038	0.245	0.045
Norway	3,596	0.114	0.074	0.147	0.063	0.262	0.035	0.222	0.045
Greece	4,777	0.040	0.143	0.077	0.091	0.249	0.040	0.217	0.050
Ireland	1,477			0.186	0.068	0.250	0.043	0.238	0.047
Indonesia	5,131			0.150	0.087	0.238	0.064	0.226	0.067
Germany	12,613	0.027	0.104	0.108	0.086	0.218	0.043	0.175	0.059
Philippines	2,154			0.168	0.062	0.216	0.070	0.208	0.069
Jordan	962					0.199	0.062	0.199	0.062
Turkey	3,628			0.054	0.219	0.204	0.065	0.189	0.080
Russia	2,764					0.197	0.107	0.197	0.107
Italy	4,225	0.033	0.093	0.093	0.069	0.191	0.047	0.149	0.057
Singapore	9,000	0.077	0.062	0.106	0.068	0.190	0.055	0.176	0.057
Brazil	5,116			0.160	0.074	0.184	0.072	0.178	0.073
Poland	3,611			0.131	0.079	0.178	0.056	0.176	0.057
France	14,066	0.046	0.081	0.069	0.078	0.176	0.047	0.136	0.058
Pakistan	2,490			0.173	0.073	0.184	0.086	0.183	0.084
Denmark	2,057	0.030	0.089	0.082	0.080	0.169	0.060	0.101	0.075
Malaysia	14,300	0.129	0.073	0.144	0.066	0.166	0.061	0.162	0.062
Korea	22,693	0.070	0.072	0.209	0.040	0.160	0.066	0.164	0.064
India	23,314			0.106	0.082	0.159	0.066	0.156	0.067
Argentina	1,198			0.110	0.095	0.154	0.184	0.145	0.166
Spain	2,457	0.035	0.105	0.116	0.078	0.154	0.064	0.130	0.072
New Zealand	1,252	0.070	0.074	0.033	0.101	0.152	0.088	0.115	0.090
Belgium	2,157	0.038	0.095	0.076	0.081	0.149	0.068	0.118	0.075
Taiwan	22,222			0.122	0.071	0.146	0.074	0.145	0.074
Egypt	1,283					0.140	0.108	0.140	0.108
Colombia	496			0.187	0.057	0.126	0.071	0.141	0.067
South Africa	6,146	0.021	0.138	0.077	0.103	0.142	0.097	0.115	0.102
Austria	1,548	0.028	0.099	0.069	0.080	0.139	0.061	0.110	0.070
Netherlands	3,873	0.032	0.102	0.052	0.109	0.133	0.068	0.097	0.085
Switzerland	4,754	0.010	0.092	0.046	0.090	0.130	0.070	0.096	0.077
Luxembourg	481			0.147	0.077	0.128	0.076	0.129	0.076
Finland	1,391	0.025	0.079	0.071	0.086	0.126	0.666	0.084	0.294
Chile	2,519	0.000	0.240	0.058	0.105	0.124	0.079	0.110	0.085
Japan	73,146	0.015	0.058	0.094	0.041	0.113	0.055	0.106	0.052
Sri Lanka	1,448			0.178	0.075	0.110	0.084	0.112	0.084
Thailand	7,263			0.143	0.080	0.110	0.095	0.115	0.093
Peru	1,304			0.123	0.099	0.097	0.111	0.100	0.110
Mexico	2,784	0.116	0.094	0.129	0.079	0.094	0.087	0.105	0.085
China	28,731			0.136	0.053	0.092	0.073	0.093	0.072
Mean	514,043	0.059	0.099	0.145	0.062	0.208	0.046	0.187	0.052

Table 2.2. Summary Firm Characteristics

In this table, I report average firm characteristics between 1980 and 2015. The final sample has 514,043 firm-year observations for 47 countries. Size indicates logged total assets denominated in 2015-dollar values. M/B indicates the sum of the market value of equity and the book value of debt over total assets. NWC indicates networking capital less cash and cash equivalents. I(*DIV>0*) takes the value of 1 if a firm pays out dividends. Industry CF Vol indicates Fama-French 48 industry-level cash-flow volatility for the previous ten years (minimum 3 years). R&D is assumed to be 0 if it is missing. All ratio variables are winsorized at the 1% and 99% levels. Details on the variables are presented in the Appendix II.

Country	Number of Firms	Cash /TA	Size	CAPEX /TA	M/B	Leverage	NWC /TA	I(*DIV>0*)	Industry CF Vol	R&D /TA
Australia	2,426	0.20	18.01	0.09	1.77	0.16	-0.01	0.35	0.14	0.01
Canada	2,684	0.16	18.44	0.11	1.86	0.20	0.02	0.29	0.12	0.02
Sweden	423	0.15	19.26	0.06	1.70	0.21	0.08	0.62	0.08	0.02
United States	12,146	0.20	19.08	0.06	2.10	0.24	0.07	0.29	0.13	0.05
United Kingdom	3,254	0.15	18.74	0.06	1.78	0.18	0.01	0.61	0.10	0.02
Hong Kong	1,325	0.22	19.06	0.04	1.51	0.22	0.01	0.48	0.10	0.00
Israel	531	0.25	18.49	0.03	1.56	0.25	0.02	0.32	0.11	0.05
Portugal	105	0.07	19.53	0.04	1.18	0.36	-0.06	0.49	0.09	0.00
Norway	419	0.17	19.21	0.09	1.60	0.30	-0.01	0.39	0.17	0.01
Greece	345	0.09	18.60	0.03	1.48	0.32	0.03	0.49	0.11	0.00
Ireland	169	0.20	19.79	0.04	2.04	0.22	-0.02	0.44	0.09	0.03
Indonesia	459	0.12	18.65	0.06	1.48	0.34	0.00	0.44	0.09	0.00
Germany	1,121	0.14	19.22	0.06	1.52	0.20	0.08	0.48	0.10	0.02
Philippines	182	0.12	18.66	0.06	1.48	0.22	-0.03	0.42	0.07	0.00
Jordan	124	0.09	17.49	0.04	1.29	0.19	0.06	0.43	0.07	0.00
Turkey	321	0.10	18.73	0.06	1.60	0.30	0.05	0.37	0.12	0.00
Russia	460	0.10	20.01	0.07	1.31	0.27	0.02	0.37	0.15	0.00
Italy	314	0.12	20.08	0.04	1.29	0.27	0.01	0.56	0.07	0.01
Singapore	802	0.18	18.64	0.05	1.31	0.21	0.03	0.58	0.12	0.00
Brazil	447	0.12	20.10	0.06	1.44	0.38	-0.04	0.55	0.11	0.00
Poland	472	0.10	18.11	0.06	1.42	0.19	0.05	0.32	0.10	0.00
France	1,390	0.15	19.31	0.04	1.44	0.22	0.03	0.56	0.10	0.02
Pakistan	239	0.08	17.89	0.06	1.30	0.33	-0.02	0.59	0.08	0.00
Denmark	195	0.15	18.77	0.07	1.50	0.27	0.06	0.63	0.07	0.01
Malaysia	1,161	0.13	18.32	0.05	1.28	0.23	0.05	0.57	0.07	0.00
Korea	2,163	0.15	18.92	0.05	1.21	0.27	0.00	0.48	0.09	0.01
India	2,480	0.07	18.22	0.07	1.46	0.33	0.06	0.55	0.10	0.00
Argentina	102	0.08	19.27	0.06	1.39	0.22	0.02	0.44	0.41	0.00
Spain	237	0.10	20.10	0.05	1.45	0.26	0.00	0.58	0.07	0.00
New Zealand	162	0.07	18.65	0.07	1.54	0.25	0.06	0.73	0.08	0.01
Belgium	201	0.13	19.53	0.06	1.44	0.24	0.01	0.57	0.08	0.02
Taiwan	2,106	0.20	18.67	0.05	1.42	0.21	0.05	0.54	0.07	0.03
Egypt	156	0.15	18.60	0.05	1.52	0.17	0.01	0.68	0.13	0.00
Colombia	56	0.07	19.92	0.05	1.07	0.14	0.00	0.69	0.06	0.00
South Africa	662	0.12	19.06	0.07	1.52	0.16	0.04	0.66	0.10	0.00
Austria	152	0.13	19.51	0.06	1.39	0.23	0.05	0.60	0.07	0.02
Netherlands	356	0.11	20.13	0.06	1.53	0.23	0.04	0.64	0.08	0.02
Switzerland	337	0.07	19.47	0.06	1.30	0.23	0.06	0.76	0.08	0.00
Luxembourg	71	0.14	20.91	0.06	1.56	0.25	0.00	0.49	0.08	0.00
Finland	173	0.12	19.44	0.21	1.38	0.30	0.09	0.74	0.20	0.02
Chile	184	0.07	19.47	0.06	1.41	0.23	0.06	0.76	0.08	0.00
Japan	4,731	0.19	19.71	0.03	1.25	0.23	0.00	0.73	0.06	0.01
Sri Lanka	162	0.08	17.39	0.06	1.30	0.24	0.01	0.65	0.07	0.00
Thailand	609	0.10	18.31	0.06	1.38	0.29	0.02	0.63	0.08	0.00
Peru	129	0.07	18.85	0.06	1.42	0.21	0.05	0.55	0.10	0.00
Mexico	243	0.09	20.76	0.05	1.36	0.26	0.02	0.49	0.08	0.00
China	3,774	0.21	19.66	0.06	2.32	0.24	-0.03	0.42	0.08	0.01
Mean	1,080	0.17	19.05	0.06	1.66	0.24	0.03	0.49	0.10	0.02

Table 2.3. OL and Cash holdings around the world

In this table, I report panel regression estimations of cash holdings (Cash and cash equivalents over Total Assets) on firm and country characteristics between 1980 and 2015. Cashflow indicates operating cash flows and I(OL) indicates a dummy variable for OL firms in each year. Size indicates logged total assets denominated in 2015-dollar values. Industry CF Vol indicates Fama-French 48 industry-level cash-flow volatility for the past ten years (minimum 3 years). R&D is assumed to be 0 if it is missing. Firms/Population indicates the total number of listed firms over population (Thousands) in each country and year. External Capital indicates the average level of closely held shares in each country and year. External Credit/GDP indicates the total domestic external credit to the private sector over GDP in each country and year. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. All Country Regressions (Dependent Variable: Cash/TA)

	All Countries							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cashflow	-0.059	0.018	0.005	0.102**	0.002	0.086**	-0.002	0.084**
	(-1.53)	(1.36)	(0.30)	(2.18)	(0.13)	(2.14)	(-0.18)	(2.09)
I(OL)		0.027***		0.024***		0.020***		0.021***
		(4.28)		(3.90)		(3.60)		(3.88)
Cashflow x I(OL)		-0.413***		-0.258***		-0.227***		-0.234***
		(-15.16)		(-3.90)		(-4.22)		(-4.29)
Size			-0.013***	-0.010***	-0.011***	-0.008***	-0.013***	-0.010***
			(-15.86)	(-11.55)	(-15.36)	(-11.23)	(-16.61)	(-12.19)
Industry CF Vol			0.028**	0.013	0.012*	-0.000	0.016**	0.004
			(2.23)	(1.53)	(1.87)	(-0.06)	(2.06)	(0.88)
R&D/TA			0.246**	0.191*	0.184*	0.139*	0.187*	0.143*
			(2.09)	(1.94)	(1.95)	(1.84)	(1.92)	(1.83)
M/B			0.036***	0.033***	0.031***	0.029***	0.032***	0.029***
			(18.55)	(22.72)	(19.88)	(20.91)	(19.91)	(20.95)
Capex/TA			-0.063*	-0.192***	-0.046	-0.160***	-0.037	-0.154**
			(-1.74)	(-2.97)	(-1.46)	(-2.84)	(-1.24)	(-2.73)
Leverage			-0.083**	-0.084**	-0.075**	-0.076**	-0.074**	-0.075**
			(-2.42)	(-2.42)	(-2.34)	(-2.33)	(-2.31)	(-2.31)
NWC/TA			-0.092***	-0.077***	-0.095***	-0.080***	-0.086***	-0.069***
			(-5.51)	(-4.59)	(-5.56)	(-4.76)	(-4.95)	(-4.01)
I(*DIV>0*)			-0.003	0.003	0.005	0.009***	0.005	0.009***
			(-0.82)	(0.85)	(1.44)	(3.09)	(1.63)	(3.02)
Firms/Population			0.200***	0.197***	0.207***	0.201***	0.216***	0.252***
			(4.46)	(5.05)	(5.65)	(5.85)	(2.84)	(3.53)
External Capital			0.010	0.013	0.009	0.012	-0.009	-0.007
			(0.88)	(1.29)	(0.76)	(1.02)	(-1.01)	(-0.84)
External Credit/GDP			0.058***	0.055***	0.047***	0.044***	0.026***	0.023***
			(15.90)	(16.58)	(17.17)	(16.90)	(4.29)	(3.91)
Year, Industry FE	NO	NO	NO	NO	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	YES	YES
Adjusted R^2	0.01	0.074	0.231	0.246	0.276	0.287	0.292	0.303
N	514,043	514,043	514,043	514,043	514,043	514,043	514,043	514,043

*** p<0.01, ** p<0.05, * p<0.1

Table 2.3. OL and Cash holdings around the world (continued)

Panel B. Sub-group Regressions (Dependent Variable: Cash/TA)

	US		Non-US		Developed		Emerging	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Cashflow	-0.096***	0.039	0.034**	0.072*	-0.027	0.154***	0.042	0.032
	(-5.72)	(1.05)	(2.31)	(1.95)	(-1.57)	(21.55)	(1.33)	(1.18)
I(OL)		0.053***		0.007		0.045***		-0.024***
		(8.11)		(1.51)		(11.15)		(-8.53)
Cashflow x I(OL)		-0.101**		-0.200***		-0.282***		-0.012
		(-2.23)		(-4.48)		(-10.30)		(-0.35)
Size	-0.008***	-0.006***	-0.012***	-0.011***	-0.012***	-0.009***	-0.006***	-0.006***
	(-5.34)	(-4.01)	(-14.10)	(-12.17)	(-13.11)	(-10.05)	(-6.24)	(-7.31)
Industry CF Vol	0.315***	0.309***	0.001	-0.004	0.023**	0.013*	-0.006***	-0.004**
	(5.97)	(5.95)	(0.42)	(-1.67)	(2.53)	(2.03)	(-3.80)	(-2.47)
R&D/TA	0.063	0.065	0.582***	0.489***	0.136*	0.105*	0.710***	0.721***
	(1.66)	(1.62)	(22.93)	(17.39)	(1.85)	(1.80)	(11.42)	(11.47)
M/B	0.031***	0.030***	0.028***	0.027***	0.035***	0.032***	0.019***	0.019***
	(24.20)	(22.37)	(16.18)	(15.12)	(27.84)	(33.56)	(7.80)	(7.83)
Capex/TA	-0.373***	-0.379***	-0.074***	-0.126**	0.006	-0.256***	-0.204***	-0.213***
	(-21.42)	(-19.87)	(-3.07)	(-2.38)	(0.20)	(-22.84)	(-14.12)	(-15.02)
Leverage	-0.150***	-0.147***	-0.054*	-0.055*	-0.161***	-0.161***	-0.026	-0.026
	(-3.96)	(-3.94)	(-1.93)	(-1.93)	(-6.83)	(-6.86)	(-1.64)	(-1.64)
NWC/TA	-0.153***	-0.142***	-0.055***	-0.046***	-0.139***	-0.123***	-0.024*	-0.030**
	(-7.72)	(-6.93)	(-3.52)	(-3.01)	(-9.82)	(-8.91)	(-1.92)	(-2.62)
I(*DIV>0*)	-0.029***	-0.028***	0.014***	0.018***	-0.014***	-0.009***	0.034***	0.030***
	(-7.13)	(-7.64)	(4.83)	(7.00)	(-4.47)	(-3.41)	(10.96)	(11.86)
Firms/Population			0.525***	0.528***	0.156**	0.216***	1.083***	1.055***
			(7.23)	(7.39)	(2.12)	(3.26)	(3.66)	(3.55)
External Capital			-0.019**	-0.018**	-0.012	-0.007	-0.009	-0.009
			(-2.25)	(-2.16)	(-0.68)	(-0.40)	(-0.76)	(-0.74)
External Credit/GDP			0.012**	0.011*	0.029***	0.023***	-0.014	-0.013
			(2.12)	(1.88)	(5.60)	(4.59)	(-1.15)	(-1.06)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.425	0.429	0.264	0.27	0.341	0.357	0.268	0.270
N	115,068	115,068	398,975	398,975	340,478	340,478	173,565	173,565

*** p<0.01, ** p<0.05, * p<0.1

Table 2.4. Cash Holdings, OL and Quality of Institutions

In this table, I report panel regression estimation results for cash holdings under various levels of country- and firm-level governance. ASDI indicates the Anti-self-dealing index in Djankov et al. (2008). ADRI indicates the Anti-Director-Rights index in La Porta et al. (1998). Rule of Law takes the value of 1 if a country has English legal origins. Developed takes the value of 1 if a country is classified as developed in all Dow Jones, MSCI, FTSE and S&P classifications. WGI indicates the average of the six dimensions of World Bank governance indices. LP indicates the Legal Protection index in Karolyi (2015). Dual Class takes the value of 1 if a firm has a dual class share. Closely Held Shares indicates the percentage of closely held shares. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. Static Country Level Governance (Dependent Variable: Cash/TA)

	ASDI		ADRI		Rule of Law		Developed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Governance	-0.034***	-0.043***	-0.009***	-0.012***	-0.030***	-0.042***	0.007**	-0.009***
	(-3.78)	(-4.91)	(-11.62)	(-19.47)	(-10.29)	(-16.35)	(2.64)	(-3.59)
I(OL)		-0.014*		-0.056***		-0.016***		-0.036***
		(-1.98)		(-7.34)		(-3.48)		(-7.77)
Governance x I(OL)		0.069***		0.023***		0.078***		0.095***
		(8.67)		(10.59)		(13.58)		(16.04)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO
Firm/Country Controls	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.277	0.281	0.289	0.297	0.281	0.290	0.276	0.287
N	514,043	514,043	474,498	474,498	514,043	514,043	514,043	514,043

Panel B. Time-varying Country/Firm Level Governance (Dependent Variable: Cash/TA)

	WGI		LP		Dual Class		Closely Held Shares	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Governance	-0.005	-0.016***	-0.010***	-0.019***	0.002	0.005**	-0.000	0.010***
	(-1.56)	(-4.56)	(-4.17)	(-8.26)	(0.62)	(2.21)	(-0.12)	(3.72)
I(OL)		-0.026***		-0.004		0.033***		0.046***
		(-5.16)		(-0.88)		(6.89)		(7.88)
Governance x I(OL)		0.062***		0.045***		-0.032***		-0.052***
		(12.57)		(14.92)		(-5.62)		(-4.90)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO
Firm/Country Controls	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.267	0.278	0.265	0.276	0.276	0.279	0.276	0.280
N	385,194	385,194	376,385	376,385	514,043	514,043	514,043	514,043

*** p<0.01, ** p<0.05, * p<0.1

Table 2.5. Cash Holdings, OL and Investor Protection after 1995

In this table, I report panel regression results for cash holdings under various country- and firm-level investor protection measures. After 1995 indicates a dummy variable that takes the value of 1 for the years 1996 and onward. ASDI indicates the Anti-self-dealing index in Djankov et al. (2008). ADRI indicates the Anti-Director-Rights index in La Porta et al. (1998). Rule of Law takes the value of 1 if a country has English legal origins. Dual Class takes the value of 1 if a firm has a dual class share. Closely Held Shares indicates the percentage of closely held shares. Firm- and country-level control variables are the same as those used Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Dependent Variable: Cash/TA

	ASDI		ADRI		Rule of Law		Dual Class		Closely Held Shares	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Governance	-0.101***	-0.099***	-0.013***	-0.014***	-0.054***	-0.055***	0.016***	0.016***	-0.007	-0.002
	(-12.16)	(-12.03)	(-9.18)	(-11.02)	(-12.20)	(-13.88)	(4.35)	(4.49)	(-1.10)	(-0.27)
Governance x After 1995	0.047***	0.034***	0.008***	0.004**	0.029***	0.018***	-0.025***	-0.020***	0.003	0.011
	(4.99)	(3.75)	(4.76)	(2.59)	(6.32)	(4.32)	(-5.97)	(-5.30)	(0.39)	(1.51)
I(OL)		-0.004		-0.070***		-0.026***		0.011		0.015
		(-0.43)		(-7.94)		(-5.58)		(1.35)		(1.46)
Governance x I(OL)		0.022		0.019***		0.052***		-0.016**		-0.022
		(1.35)		(5.82)		(5.50)		(-2.11)		(-1.54)
I(OL) x After 1995		-0.001		0.003		0.020***		0.030***		0.041***
		(-0.10)		(0.27)		(3.89)		(3.40)		(3.68)
Governance x I(OL) x After 1995		0.046**		0.008**		0.020*		-0.014		-0.035*
		(2.46)		(2.12)		(1.97)		(-1.42)		(-1.93)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Firm/Country Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.289	0.294	0.287	0.295	0.290	0.300	0.286	0.290	0.285	0.291
N	432,428	432,428	431,947	431,947	432,428	432,428	432,428	432,428	432,428	432,428

*** p<0.01, ** p<0.05, * p<0.1

Table 2.6. Cash Holdings, OL and Investor Protection

In this table, I report panel regression results for the relationship between cash holdings and bankruptcy, delisting, and M&A target indicators. I(*Bankruptcy*) takes the value of 1 if a firm files for bankruptcy or becomes liquidated in the following year. I(*Delisted*) takes the value of 1 if a firm becomes delisted in the following year. I(*M&A Target*) takes the value of 1 if a firm becomes acquired or merged in the following year. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year level. Numbers in parentheses indicate t-statistics.

Panel A. Dependent Variable: I(*Bankruptcy*)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(1)	(2)	(3)	(4)	(5)
Cash/TA	-0.001*	-0.000	0.000	-0.001	0.000
	(-1.87)	(-0.14)	(0.93)	(-0.69)	(0.05)
I(OL)	0.006***	0.008***	0.001**	0.018***	0.001***
	(4.00)	(4.09)	(2.66)	(4.99)	(2.83)
Cash/TA x I(OL)	-0.009***	-0.012***	-0.001*	-0.027***	-0.002*
	(-4.10)	(-4.07)	(-1.71)	(-4.68)	(-1.87)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.009	0.012	0.001	0.020	0.003
N	514,043	282,366	167,298	115,068	231,677

Panel B. Dependent Variable: I(*Delisted*)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(6)	(7)	(8)	(9)	(10)
Cash/TA	-0.003**	0.000	0.001	-0.006**	-0.006***
	(-2.06)	(0.02)	(0.41)	(-2.52)	(-4.51)
I(OL)	0.010***	0.007***	0.007***	0.008***	0.014***
	(11.61)	(5.43)	(4.73)	(3.61)	(7.58)
Cash/TA x I(OL)	-0.017***	-0.014***	-0.014***	-0.011***	-0.028***
	(-8.39)	(-4.48)	(-3.07)	(-2.79)	(-5.33)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.010	0.012	0.006	0.019	0.010
N	514,043	282,366	167,298	115,068	231,677

Panel C. Dependent Variable: I(*M&A Target*)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(11)	(12)	(13)	(14)	(15)
Cash/TA	0.001	0.000	-0.013***	-0.001	0.001
	(0.51)	(0.02)	(-4.56)	(-0.32)	(0.49)
I(OL)	-0.002**	-0.005***	-0.005***	-0.005**	0.001*
	(-2.27)	(-3.45)	(-3.77)	(-2.35)	(1.86)
Cash/TA x I(OL)	-0.000	0.004	0.010**	0.009*	-0.006*
	(-0.05)	(1.12)	(2.70)	(1.70)	(-1.99)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.019	0.019	0.014	0.014	0.009
N	514,043	282,366	167,298	115,068	231,677

*** p<0.01, ** p<0.05, * p<0.1

Table 2.7. Cash Holdings, OL and Capital Issuance

In this table, I report panel regression results of cash holdings on capital issuances. Each issuance indicates the aggregate proceeds from issuing capital after one year. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. Equity Issuance (Dependent Variable: Equity Issue/TA)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(1)	(2)	(3)	(4)	(5)
Cash/TA	-0.034*	0.057**	0.008	0.092**	-0.049**
	(-1.74)	(2.50)	(0.27)	(2.66)	(-2.30)
I(OL)	0.014*	0.030***	0.051***	-0.032***	0.008
	(1.94)	(3.16)	(4.26)	(-3.33)	(0.94)
Cash/TA x I(OL)	0.332***	0.204***	0.274***	0.169***	0.189***
	(12.91)	(6.92)	(7.59)	(4.65)	(5.19)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.341	0.345	0.276	0.414	0.166
N	62,815	42,670	26,091	16,579	20,145

Panel B. Debt Issuance (Dependent Variable: Debt Issue/TA)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(6)	(7)	(8)	(9)	(10)
Cash/TA	-0.127**	-0.012	-0.199***	0.054	-0.022
	(-2.60)	(-0.18)	(-3.02)	(0.59)	(-0.30)
I(OL)	0.033	0.088**	0.019	0.091**	0.018
	(1.15)	(2.56)	(0.30)	(2.17)	(0.24)
Cash/TA x I(OL)	1.037***	0.744***	1.209***	0.605***	0.335
	(10.47)	(6.32)	(2.86)	(4.54)	(0.77)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.364	0.400	0.195	0.433	0.059
N	33,114	20,657	6,586	14,071	12,457

Panel C. Syndicated Loan Issuance (Dependent Variable: Syndicated Loan Issue/TA)

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(11)	(12)	(13)	(14)	(15)
Cash/TA	-0.345***	-0.240***	-0.260***	-0.185***	-0.110***
	(-12.96)	(-8.98)	(-3.88)	(-6.66)	(-4.25)
I(OL)	-0.015	0.017	0.066**	-0.008	0.014
	(-0.94)	(1.12)	(2.07)	(-0.47)	(1.07)
Cash/TA x I(OL)	0.312***	0.222**	0.539***	0.074	-0.044
	(3.07)	(2.08)	(3.65)	(0.68)	(-0.49)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.196	0.217	0.251	0.235	0.142
N	43,822	26,932	8,096	18,836	16,888

*** p<0.01, ** p<0.05, * p<0.1

Table 2.8. Cash Holdings, OL and Investment and R&D

In this table, I report panel regression results of cash holdings on Investment and R&D. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. Dependent Variable: Capex/TA

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(1)	(2)	(3)	(4)	(5)
Cash/TA	-0.054***	-0.056***	-0.056***	-0.056***	-0.051***
	(-16.48)	(-14.67)	(-10.27)	(-14.36)	(-6.51)
I(OL)	-0.019***	-0.019***	-0.018***	-0.018***	-0.020***
	(-22.37)	(-12.73)	(-8.97)	(-9.04)	(-18.97)
Cash/TA x I(OL)	0.025***	0.025***	0.025***	0.028***	0.045***
	(5.17)	(5.71)	(3.60)	(8.22)	(6.95)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.043	0.199	0.180	0.238	0.025
N	514,043	282,366	167,298	115,068	231,677

Panel B. Dependent Variable: R&D/TA

	All Countries	High Investor Protection			Low Investor Protection
		All High	Non-US	US	All Low
	(6)	(7)	(8)	(9)	(10)
Cash/TA	0.043***	0.047***	0.009**	0.058***	0.026***
	(15.62)	(10.63)	(2.21)	(7.54)	(6.61)
I(OL)	-0.016***	0.004	-0.002	0.019***	-0.031***
	(-5.79)	(1.00)	(-0.34)	(3.33)	(-8.09)
Cash/TA x I(OL)	0.552***	0.523***	0.447***	0.567***	0.481***
	(30.61)	(26.59)	(15.71)	(22.91)	(9.99)
Year, Industry FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.321	0.337	0.221	0.386	0.224
N	514,043	282,366	167,298	115,068	231,677

Table 2.9. Effects of Change in Corporate Income Taxes Around the World on OL and Cash Holdings

In this table, I report 2SLS regression results for the relationship between cash holdings and cash flows and an OL dummy using changes in nominal corporate income taxes in OECD countries as an instrument. *Corporate Income Tax Change* indicates changes in nominal corporate income taxes between year *t* and *t-1*. OECD countries for the variable are Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and United States. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the country and year levels. Numbers in parentheses indicate t-statistics.

Panel A. First-stage Regressions

	All Countries		High Investor Protection		Low Investor Protection	
Dependent Variables:	Cashflow	I(OL)	Cashflow	I(OL)	Cashflow	I(OL)
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Income Tax Change	-0.082**	0.227***	-0.149**	0.369***	0.119***	-0.044
	(-2.04)	(2.83)	(-2.43)	(3.14)	(3.26)	(-0.92)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.775	0.335	0.800	0.353	0.252	0.245
N	361,713	361,713	295,875	295,875	65,838	65,838

Panel B. Second-stage Regressions

	All Countries		High Investor Protection		Low Investor Protection	
Dependent Variables:	Cash/TA		Cash/TA		Cash/TA	
	(7)	(8)	(9)	(10)	(11)	(12)
Cashflow	-0.560		-0.755**		0.147	
	(-1.45)		(-2.07)		(0.63)	
I(OL)		0.203**		0.304***		-0.396
		(1.98)		(2.79)		(-0.49)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
F-stat	4.03	7.77	5.75	9.55	10.32	0.82
N	361,713	361,713	295,875	295,875	65,838	65,838

*** p<0.01, ** p<0.05, * p<0.1

Table 2.10. Effects of Country-Level Institutional Ownership on OL and Cash Holdings

In this table, I report 2SLS regression results for the relationship between cash holdings and an OL dummy using the proportion of the total number of OL firms held by various types of institutions as instruments. *Proportion of OL Institutions* indicates the number of OL firms held by a certain type of institution divided by the number of non-OL firms held by the same type of institution in each country and year, excluding the firm of interest. *Independent* and *Grey* indicate independent and grey institutional ownership defined following Ferreira and Matos (2008). High/Low Investor Protection indicates countries with above/below the mean Anti-self-dealing Index in Djankov et al. (2008). All standard errors are double-clustered at the country and year levels. Numbers in parentheses indicate t-statistics.

Panel A. First-stage Regressions

Dependent Variable: I(OL)	High Investor Protection			Low Investor Protection		
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of OL Institutions	0.158***	0.159***	0.007	0.126*	0.123*	-0.789***
	(5.50)	(5.51)	(0.10)	(1.95)	(1.89)	(-8.01)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES	YES
Institution Type	*All*	*Independent*	*Grey*	*All*	*Independent*	*Grey*
Adjusted R^2	0.352	0.352	0.351	0.223	0.223	0.236
N	282,366	282,366	282,366	231,677	231,677	231,677

Panel B. Second-stage Regressions

Dependent Variable: Cash/Ta	High Investor Protection			Low Investor Protection		
	(7)	(8)	(9)	(10)	(11)	(12)
I(*CF<0*)	0.156***	0.152***	-0.021	0.104	0.107	-0.018
	(4.35)	(4.39)	(-0.97)	(0.46)	(0.46)	(-1.18)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES	YES
Institution Type	*All*	*Independent*	*Grey*	*All*	*Independent*	*Grey*
F-stat	143.568	393.381	174.171	48.700	38.459	12.941
N	282,366	282,366	282,366	231,677	231,677	231,677

*** p<0.01, ** p<0.05, * p<0.1

Table 2.11. Effects of Firm-Level Institutional Ownership on OL and Cash Holdings

In this table, I report 2SLS regression results for the relationship between cash holdings and OL using the average proportion of the number of OL firms among a certain type of institution that holds the firm of interest as instruments. *Average Proportion of OL firms* indicates the number of OL firms divided by the total number of institutions held by an institution averaged across all the institutions that hold the firm of interest in each year (Excluding the firm from the count). *Hedge* indicates institutions that identified their style as a hedge fund and *Activist* indicates institutions that filed form 13D in the U.S. High/Low Investor Protection indicates countries with above/below the mean Anti-self-dealing Index score in Djankov et al. (2008). All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. First-stage Regressions

Dependent Variable: I(OL)	High Investor Protection			Low Investor Protection		
	(1)	(2)	(3)	(4)	(5)	(6)
Average Proportion of OL firms	1.399***	0.946***	0.884***	0.933***	0.523***	0.463***
	(12.10)	(18.54)	(13.52)	(6.99)	(6.54)	(5.43)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES	YES
Institution Type	*All*	*Hedge*	*Activist*	*All*	*Hedge*	*Activist*
Adjusted R^2	0.348	0.367	0.349	0.213	0.185	0.160
N	106,816	56,823	62,214	96,272	25,060	5,738

Panel B. Second-stage Regressions

Dependent Variable: Cash/Ta	High Investor Protection			Low Investor Protection		
	(7)	(8)	(9)	(10)	(11)	(12)
I(*CF<0*)	0.272***	0.461***	0.458***	-0.364***	0.299***	0.688***
	(5.94)	(13.69)	(10.82)	(-3.62)	(3.75)	(4.66)
Year, Industry FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Firm/Country Controls	YES	YES	YES	YES	YES	YES
Institution Type	*All*	*Hedge*	*Activist*	*All*	*Hedge*	*Activist*
F-stat	143.568	393.381	174.171	48.700	38.459	12.941
N	106,816	56,823	62,214	96,272	25,060	5,738

*** p<0.01, ** p<0.05, * p<0.1

Table 2.12. Effects of Governance Spillover from Institutional Investors on OL and Cash Holdings

In this table, I report panel regression results for the relationship between cash holdings and the difference between institutional investors' domicile country governance and firms' domicile country governance (Governance Difference). *Governance Difference* indicates the difference between the average governance score of institutional investors and firms' governance scores. The average governance scores of institutional investors are calculated as a weighted average of Anti-Self-Dealing Index scores using the percentage of ownership as each weight. Each firm's governance score is measured by the ASDI of its domicile country. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Dependent Variable: Cash/TA

	All	High Investor Protection			Low Investor Protection	All	High Investor Protection		Low Investor Protection
	Countries	All High	Non-US	US	All Low	Countries	All High	Non-US	All Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Governance Difference	0.061***	0.049***	0.047***	0.036***	0.154***	0.006	0.124***	-0.098	-0.086**
	(8.06)	(6.47)	(3.46)	(3.72)	(8.15)	(0.40)	(3.47)	(-1.18)	(-2.55)
I(OL)	0.081***	0.216***	0.144***	0.205***	0.019*	-0.092***	0.270***	0.191*	-0.050**
	(9.55)	(20.98)	(9.67)	(18.39)	(1.90)	(-5.12)	(3.44)	(1.77)	(-2.53)
Governance Difference * I(OL)	0.077***	0.237***	0.147***	0.261***	0.043**	0.213***	0.433***	0.462***	0.356***
	(7.06)	(17.27)	(8.12)	(13.57)	(2.28)	(5.40)	(4.99)	(2.96)	(3.60)
ASDI						-0.061***	-0.142***	0.008	0.239***
						(-4.93)	(-3.14)	(0.08)	(6.11)
ASDI * I(OL)						0.537***	-0.050	0.053	0.593***
						(14.95)	(-0.44)	(0.42)	(5.46)
Governance Difference * ASDI						-0.032	-0.157***	0.061	0.470***
						(-1.65)	(-3.10)	(0.62)	(5.85)
Governance Difference * ASDI * I(OL)						0.219***	-0.209*	-0.228	0.286**
						(4.91)	(-1.75)	(-1.38)	(2.16)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	YES	NO	NO	NO	NO
Firm/Country Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.291	0.329	0.271	0.427	0.294	0.288	0.331	0.256	0.276
N	391,944	205,881	135,357	70,524	186,063	391,944	205,881	135,357	186,063

*** p<0.01, ** p<0.05, * p<0.1

Appendix II. Variable Descriptions

This table reports descriptions of variables in the paper and their sources. Items in parentheses indicate WorldScope item numbers.

Variable	Description	Source
Cash Flow and Firm Characteristics		
Cash Flow	Funds from Operations (ITEM4201) / Total Assets (ITEM2999)	WorldScope
I(OL)	1 if Funds from Operations (ITEM4201) is negative and 0 if otherwise	WorldScope
Cash/TA	Cash and Short Term Investments (ITEM2001) / Total Assets(ITEM2999)	WorldScope
CAPEX/TA	Capital Expenditures (ITEM4601) / Total Assets(ITEM2999)	WorldScope
Size	Logged Total Assets in USD (ITEM7230)	WorldScope
Industry CF Vol	Past 10-year Fama-French 48 industry Cash Flow (ITEM4201) volatility (Minimum 3 years of Cash Flow)	WorldScope
M/B	(Market Capitalization in USD (ITEM7210) + (Total Assets in USD (ITEM7230) - Common Equity in USD (ITEM7220)) / Total Assets in USD (ITEM7230)	WorldScope
Leverage	Total Debt (ITEM3255) / Total Assets (ITEM2999)	WorldScope
Country Characteristics		
Firms/Population	Number of Listed Firms / Population in each year	World Bank†
Mktcap/GDP	Market Capitalization / GDP	World Bank†
External Credit/GDP	Domestic credit to private sector / GDP	World Bank†
Capital Issuance		
Equity Issue/TA	Aggregate Annual Proceeds from Issuing Equity / Total Assets (ITEM2999)	SDC Platinum/ WorldScope
Debt Issue/TA	Aggregate Annual Proceeds from Issuing Corporate Debt / Total Assets (ITEM2999)	SDC Platinum/ WorldScope
Syndicated Loan Issue/TA	Aggregate Annual Proceeds from Issuing Syndicated Loan / Total Assets (ITEM2999)	SDC Platinum/ WorldScope
Investor Protection		
Anti-Self-Dealing	Country-Level Anti-Self-Dealing Index	Djankov et al. (2008)
Corporate Opacity	Country-Level Corporate Opacity Index	Karolyi (2015)
Legal Protection	Country-Level Legal Protection Index	Karolyi (2015)
Common	1 if a country has common law system	Djankov et al. (2008)

Variable	Description	Source
MSCI Gov	10-scale firm-level MSCI Governance Index	MSCI
Dual Class	1 if a firm has multiple share classes (ITEM11501)	WorldScope
Closely Held Shares	The proportion of closely held shares (ITEM8021)	WorldScope
ADR	1 if a firm has ADR (ITEM11496)	WorldScope
I(*Bankruptcy*)	1 if a firm's new status footnote for year t+1 has the following keywords: "BANKRUPT," "TERMINATED," "LIQUIDATED," "CHAPTER 11," "CHAPTER 13," or similar expressions and 0 otherwise	WorldScope
I(*Delisting*)	1 if a firm's status footnote for year t+1 has the following keywords: "DELISTED," "PRIVATELY HELD," "WITHDREW," "CEASED," "PRIVATE COMPANY," "STOCK DOES NOT ACTIVELY TRADE," "CANCEL," or similar expressions and 0 otherwise	WorldScope
I(*M&A Target*)	1 if a firm's status footnote for year t+1 has the following keywords: "ACQUIERED," "MERGED," or similar expressions and 0 otherwise	WorldScope
Instrument for Operating Cash Flow Change		
Corporate Income Tax Change	Changes in nominal Corporate Income Tax Rates between *t* and *t-1*	OECD
Proportion of OL Institutions	The ratio between the number of OL firms held by a type of institution and the number of non-OL firms held by the same type of institution in each country at *t*.	FactSet Lionshares
Average Proportion of OL firms	The number of OL firms held by an institution over the total number of firms held by the same institution averaged across all the institutions that hold the firm of interest at *t*.	FactSet Lionshares

†Data for Taiwan were collected from various sources including the Taiwan Stock Exchange, CIA Factbook, and Datastream

Table 2.A1. The Effects of the 2008 Financial Crisis on OL and Cash Holdings

In this table, I report panel regression results for the relationship between cash holdings and the level of investor protection around the 2008-2009 financial crisis dummy between 2005 and 2011. Financial Crisis takes the value of 11 for years after 2009. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. Survivorship Bias Control indicates that only firms that appear in the entire 2007–2011 period are analyzed. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Dependent Variable: Cash/TA

	All Countries				Non-US			
	OL	Non-OL	OL	Non-OL	OL	Non-OL	OL	Non-OL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High Protection	0.018***	0.006	0.020***	0.001	-0.003	0.012**	-0.009	0.003
	(2.94)	(1.39)	(3.51)	(0.18)	(-0.53)	(2.13)	(-1.24)	(0.64)
Financial Crisis	-0.006	0.030***	-0.013***	0.023***	-0.004	0.032***	-0.011***	0.024***
	(-1.63)	(7.19)	(-7.79)	(13.76)	(-1.06)	(7.07)	(-6.91)	(12.18)
High Protection * Financial Crisis	0.010*	0.015***	0.006***	-0.003	0.014**	0.023***	0.012**	-0.001
	(1.70)	(3.09)	(3.82)	(-1.15)	(2.43)	(3.64)	(2.36)	(-0.34)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO
Firm/Country Controls	YES	YES	YES	YES	YES	YES	YES	YES
Survivorship Bias Control	NO	NO	YES	YES	NO	NO	YES	YES
Adjusted R^2	0.325	0.233	0.357	0.231	0.285	0.214	0.269	0.201
N	26,432	101,396	16,307	80,800	21,382	87,641	13,540	69,612

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A2. Cash Holdings, OL and Quality of Institutions under DMS2003

In this table, I report panel regression estimation results for the relationship between cash holdings and country-level governance used in DMS2003. ADRI indicates the Anti-Director-Rights index in La Porta et al. (1998). Rule of Law takes the value of 1 if a country has English legal origins. DMS2003 Countries indicate countries covered by DMS2003. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. Anti-Director-Rights (Dependent Variable: Cash/TA)

	DMS2003 Countries												
	DMS2003 Sample Period			All Years			Before 1996			After 1996			(7) - (10)
	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
ADRI	-0.005***	-0.007***	0.011**	-0.009***	-0.011***	0.012***	-0.010***	-0.011***	0.002	-0.008***	-0.012***	0.017***	
	(-4.71)	(-6.85)	(2.54)	(-10.04)	(-14.60)	(4.00)	(-6.62)	(-8.02)	(0.74)	(-6.84)	(-14.10)	(6.12)	
Sample Period	1998	1998	1998	1980 -2015	1980 -2015	1980 -2015	1980 -1995	1980 -1995	1980 -1995	1996 -2015	1996 -2015	1996 -2015	(-2.63)***
Adjusted R^2	0.328	0.316	0.360	0.239	0.222	0.275	0.302	0.294	0.382	0.231	0.215	0.269	
N	15,498	12,349	3,149	434,552	350,093	84,459	68,349	62,421	5,928	366,203	287,672	78,531	

Panel B. Rule of Law (Dependent Variable: Cash/TA)

	DMS2003 Countries												
	DMS2003 Sample Period			All Years			Before 1996			After 1996			(19) - (22)
	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	
Rule of Law	-0.029***	-0.035***	0.010	-0.032***	-0.041***	0.034***	-0.039***	-0.046***	-0.011	-0.030***	-0.042***	0.044***	
	(-10.51)	(-12.35)	(0.87)	(-12.33)	(-17.86)	(5.15)	(-11.15)	(-11.53)	(-1.45)	(-10.64)	(-19.67)	(6.60)	
Sample Period	1980 -2011	1980 -2011	1980 -1995	1980 -2015	1980 -2015	1980 -2015	1980 -1995	1980 -1995	1980 -1995	1996 -2015	1996 -2015	1996 -2015	(-4.66)***
Adjusted R^2	0.332	0.323	0.360	0.243	0.230	0.276	0.308	0.302	0.382	0.234	0.223	0.270	
N	15,532	12,376	3,156	434,552	350,093	84,459	68,349	62,421	5,928	366,203	287,672	78,531	

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A3. Cash Holdings, OL and Firm Value

In this table I report panel regression estimation results for the relationship between cash holdings and Firm value. ASDI indicates the Anti-Self-Dealing index. PSW2006 & KL2007 Countries indicate countries covered by PSW2006 & KL2007. Firm- and country-level control variables are the same as those used in Table 2.3. In Panel B, all control variables are changes from the previous year. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. Dependent Variable: Tobin's Q

	PSW2006 & KL2007 Countries											
	PSW2006 & KL2007 Sample Period			All Years			Before 1996			After 1996		
	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ASDI	-0.046	-0.022	-0.354*	0.237***	0.309***	-0.238***	-0.089	-0.078	-0.836***	0.244***	0.310***	-0.234***
	(-0.65)	(-0.33)	(-1.96)	(4.47)	(6.36)	(-3.18)	(-1.20)	(-0.95)	(-5.79)	(4.23)	(5.70)	(-2.87)
Cash/TA	1.042***	1.314***	0.469	1.139***	1.060***	0.878**	0.720***	0.973***	-0.837	1.206***	1.127***	0.919**
	(3.85)	(5.69)	(0.57)	(4.98)	(5.27)	(2.73)	(3.16)	(4.19)	(-1.29)	(4.85)	(4.99)	(2.76)
Cash/TA * ASDI	1.291***	0.358	1.995*	1.043***	0.817***	1.165***	1.472***	0.594*	3.595***	0.938***	0.720**	1.112**
	(3.22)	(1.07)	(1.81)	(3.77)	(3.13)	(2.94)	(3.88)	(1.80)	(3.58)	(3.15)	(2.50)	(2.73)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Sample Period	1983 -1998	1983 -1998	1983 -1998	1980 -2015	1980 -2015	1980 -2015	1980 -1995	1980 -1995	1980 -1995	1996 -2015	1996 -2015	1996 -2015
Adjusted R^2	0.266	0.345	0.375	0.215	0.187	0.325	0.284	0.345	0.458	0.219	0.191	0.318
N	115,129	101,856	13,273	420,752	338,948	81,804	75,574	68,721	6,853	345,178	270,227	74,951

Panel B. Dependent Variable: ΔTobin's Q

	PSW2006 & KL2007 Countries											
	PSW2006 & KL2007 Sample Period			All Years			Before 1996			After 1996		
	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL	All Firms	Non-OL	OL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ASDI	0.312***	0.242***	0.579***	0.226***	0.137***	0.306***	0.431***	0.382***	0.436*	0.220***	0.131***	0.309***
	(6.00)	(4.00)	(3.73)	(5.31)	(3.49)	(3.53)	(6.17)	(6.13)	(1.83)	(5.00)	(3.23)	(3.48)
ΔCash/TA	2.082***	2.686***	1.263**	-0.044	-0.030	-0.232*	3.604***	3.128***	4.122***	-0.047	-0.029	-0.248*
	(4.24)	(5.61)	(2.22)	(-0.92)	(-0.59)	(-1.72)	(3.44)	(4.09)	(4.61)	(-0.96)	(-0.59)	(-1.85)
ΔCash/TA * ASDI	-2.698***	-2.658***	-1.926**	0.125	0.098	0.611**	-4.749***	-3.693***	-5.888***	0.131	0.094	0.638**
	(-4.59)	(-4.88)	(-2.78)	(1.34)	(0.99)	(2.72)	(-3.77)	(-4.00)	(-6.66)	(1.35)	(0.98)	(2.83)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Sample Period	1983 -1998	1983 -1998	1983 -1998	1980 -2015	1980 -2015	1980 -2015	1980 -1995	1980 -1995	1980 -1995	1996 -2015	1996 -2015	1996 -2015
Adjusted R^2	0.184	0.187	0.248	0.153	0.163	0.189	0.191	0.187	0.322	0.152	0.161	0.184
N	56,334	49,002	7,332	342,379	273,298	69,081	28,126	25,226	2,900	314,253	248,072	66,181

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A4. Cash Holdings, OL and Intensive R&D on U.S. Cash Holdings

In this table I report panel regression results for the relationship between cash holdings and indicator variables for U.S. firms following PSW2016. β_{US} indicates a coefficient on a dummy variable that takes the value of 1 if a firm is headquartered in the U.S. and zero otherwise. High R&D indicates the top two deciles of R&D/Sales. PSW2016 Countries indicate countries covered in PSW2016. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Panel A. R&D and US Cash Holdings (Dependent Variable: Cash/TA)

	OSW2016 Countries			
	Low R&D	High R&D	Low R&D	High R&D
	(1)	(2)	(3)	(4)
β_{US}	-0.007	0.023***	-0.032***	0.013
	(-1.64)	(3.68)	(-5.94)	(1.14)
Firm Controls	YES	YES	YES	YES
Country Controls	NO	NO	YES	YES
Sample Period	OSW2016	OSW2016	OSW2016	OSW2016
N	246,546	66,951	234,233	64,416

Panel B. OL and US Cash Holdings (Dependent Variable: Cash/TA)

	OSW2016 Countries			
	Non-OL	OL	Non-OL	OL
	(5)	(6)	(7)	(8)
β_{US}	0.004	0.047***	-0.022***	0.056***
	(2.83)	(10.82)	(-2.75)	(6.98)
Firm Controls	YES	YES	YES	YES
Country Controls	NO	NO	YES	YES
Sample Period	OSW2016	OSW2016	OSW2016	OSW2016
N	244,934	68,563	234,631	64,018

Panel C. R&D, OL and US Cash Holdings (Dependent Variable: Cash/TA)

		OSW2016 Countries			
		Low R&D	High R&D	Low R&D	High R&D
		(9)	(10)	(11)	(12)
β_{US}	Non-OL	-0.018***	0.014***	-0.045***	0.012
		(-4.19)	(2.69)	(-6.90)	(1.13)
	OL	0.021***	0.033***	0.023***	0.029**
		(3.31)	(3.50)	(3.03)	(2.35)
Firm Controls		YES	YES	YES	YES
Country Controls		NO	NO	YES	YES
Sample Period		OSW2016	OSW2016	OSW2016	OSW2016
N (Non-OL)		197,261	47,673	188,770	45,861
N (OL)		49,285	19,278	45,463	18,555

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A5. Early Sample Period Cash Holdings and OL

In this table, I report panel regression results for the relationship between cash holdings and firm and country characteristics between 1980 and 1995. High Investor Protection indicates above average Anti-Self-Dealing Index scores. Firm- and country-level control variables are the same as those used in Table 2.3. All standard errors are double-clustered at the firm and year levels. Numbers in brackets indicate t-statistics.

Dependent Variable: Cash/TA

	All Countries		High Investor Protection						Low Investor Protection	
			All High		Non-US		US		All Low	
	(1)	(2)	(1)	(2)	(3)	(4)	(5)	(8)	(9)	(10)
Cashflow	-0.042	0.032	-0.043	-0.002	0.082**	0.130***	-0.083**	-0.086*	0.119***	0.186***
	(-1.25)	(1.07)	(-1.20)	(-0.07)	(2.53)	(4.09)	(-2.66)	(-2.07)	(5.17)	(7.33)
I(OL)		-0.004		-0.005		-0.001		-0.002		0.010*
		(-0.95)		(-1.03)		(-0.15)		(-0.28)		(2.03)
Cashflow x I(OL)		-0.199***		-0.116**		-0.162**		-0.002		-0.200**
		(-4.15)		(-2.40)		(-2.55)		(-0.03)		(-2.59)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO	YES	YES	YES	YES
Adjusted R^2	0.357	0.358	0.391	0.392	0.281	0.282	0.452	0.452	0.320	0.322
N	78,791	78,791	50,121	50,121	20,651	20,651	29,470	29,470	28,670	28,670

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A6. WorldScope and Compustat Comparison for US firms

In this table, I replicate Table 2.3 following Denis and McKeon (2018) using a Compustat sample. Compustat-WorldScope indicates firms in Compustat that also appear in the WorldScope sample using Compustat data. WorldScope-Compustat indicates firms in WorldScope that appear in the Compustat sample using WorldScope data. All standard errors are double-clustered at the firm and year levels. Numbers in parentheses indicate t-statistics.

Dependent Variable: Cash/TA

	Compustat-WorldScope		WorldScope-Compustat		Compustat 1980-2015		Compustat 1970-2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cashflow	-0.042**	0.172***	-0.107***	0.068*	-0.007	0.178***	-0.002	0.163***
	(-2.36)	(6.88)	(-7.64)	(1.88)	(-0.34)	(5.23)	(-0.10)	(5.71)
I(OL)		0.045***		0.055***		0.038***		0.035***
		(8.91)		(8.45)		(7.00)		(7.29)
Cashflow x I(OL)		-0.234***		-0.151***		-0.221***		-0.205***
		(-7.08)		(-3.19)		(-5.77)		(-6.34)
Size	-0.010***	-0.008***	-0.010***	-0.008***	-0.010***	-0.008***	-0.010***	-0.008***
	(-7.41)	(-5.48)	(-7.38)	(-5.59)	(-8.79)	(-7.13)	(-10.20)	(-8.16)
Industry CF Vol	0.371***	0.348***	0.287***	0.279***	0.337***	0.311***	0.369***	0.343***
	(5.72)	(5.54)	(5.69)	(5.69)	(5.42)	(5.19)	(6.60)	(6.36)
R&D/TA	0.154***	0.142***	0.051	0.050	0.192***	0.172***	0.203***	0.182***
	(4.07)	(3.97)	(1.58)	(1.53)	(5.31)	(5.20)	(5.52)	(5.33)
M/B	0.022***	0.020***	0.021***	0.020***	0.022***	0.020***	0.022***	0.020***
	(12.18)	(13.73)	(7.23)	(7.13)	(13.77)	(14.82)	(14.65)	(15.77)
Capex/TA	-0.371***	-0.398***	-0.374***	-0.393***	-0.313***	-0.329***	-0.311***	-0.324***
	(-16.45)	(-18.00)	(-18.95)	(-18.60)	(-19.49)	(-19.53)	(-21.66)	(-22.01)
Leverage	-0.219***	-0.215***	-0.210***	-0.206***	-0.199***	-0.197***	-0.201***	-0.199***
	(-14.46)	(-14.28)	(-13.53)	(-13.29)	(-9.64)	(-9.63)	(-10.26)	(-10.25)
NWC/TA	-0.208***	-0.197***	-0.216***	-0.203***	-0.183***	-0.173***	-0.190***	-0.181***
	(-12.95)	(-12.03)	(-15.50)	(-13.96)	(-15.21)	(-14.28)	(-17.30)	(-16.40)
I(DIV>0)	-0.034***	-0.035***	-0.029***	-0.029***	-0.025***	-0.025***	-0.019***	-0.018***
	(-9.82)	(-10.39)	(-8.08)	(-9.10)	(-6.63)	(-6.22)	(-5.04)	(-4.70)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	NO	NO	NO	NO	NO	NO	NO	NO
Adjusted R^2	0.473	0.478	0.462	0.466	0.426	0.431	0.435	0.441
N	106,748	106,748	103,336	103,336	154,680	154,680	185,835	185,835

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A7. Cash Holdings of 1980's WorldScope Countries

In this table, I report panel regression results for the relationship between cash holdings and firm and country characteristics between 1980 and 2015 using countries that have firms from the 1980s in the sample. This excludes: Argentina, China, Egypt, India, Indonesia, Israel, Jordan, Pakistan, Peru, Philippines, Poland, Russia, and Sri Lanka. Firm- and country-level control variables are the same as those used in Table 2.3. High Investor Protection indicates above-average Anti-Self-Dealing Index scores. All standard errors are double-clustered at the firm and year levels. Numbers in brackets indicate t-statistics.

Dependent Variable: Cash/TA

	All Countries		High Investor Protection						Low Investor Protection	
			All High		Non-US		US		All Low	
	(1)	(2)	(1)	(2)	(3)	(4)	(5)	(8)	(9)	(10)
Cashflow	0.000	0.170***	-0.110***	0.058**	-0.089***	0.092***	-0.096***	0.039	0.129***	0.213***
	(0.01)	(20.37)	(-6.98)	(2.59)	(-7.46)	(5.22)	(-5.72)	(1.05)	(9.18)	(13.93)
I(OL)		0.035***		0.040***		0.026***		0.053***		0.010***
		(9.11)		(8.70)		(5.84)		(8.11)		(4.58)
Cashflow x I(OL)		-0.295***		-0.168***		-0.236***		-0.101**		-0.248***
		(-9.63)		(-4.86)		(-8.64)		(-2.23)		(-10.37)
Year, Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO	YES	YES	YES	YES
Adjusted R^2	0.297	0.313	0.365	0.368	0.302	0.306	0.425	0.429	0.273	0.279
N	432,428	432,428	241,239	241,239	126,171	126,171	115,068	115,068	191,189	191,189

*** p<0.01, ** p<0.05, * p<0.1

CHAPTER 3: EXECUTIVE MOBILITY IN THE UNITED STATES, 1920 TO 2011

3.1 Introduction

Several important trends in the labor market for corporate executives have emerged in recent decades. The level of executive pay and its dispersion have increased significantly since the 1970s (e.g., Murphy 1999; Frydman and Saks 2010). Over a similar period, chief executive officer (CEO) turnover and external-to-the-firm hires have increased as general managerial skills have become more important (e.g., Huson, Parrino, and Starks 2001; Murphy and Zabojnik 2007).[1] In addition, the numbers of firms and occupations that typical executives work in during their careers have increased since 1970 (Frydman 2017).

A common thread across these trends is increasing executive mobility across firms. In this paper, we examine the long-run evolution of across-firm executive mobility and its implications for incentives and corporate decisions by constructing a new dataset of executive movements between US firms over the 1920-2011 period. The dataset contains more than 14,000 executive moves and 315,000 unique executives, involving nearly 18,000 public firms. Our long data sample allows us to document new patterns over the century, and also put recent trends in the context of century-long patterns in executive mobility.

We document several new trends in executive mobility. First, movements of executives to jobs in new firms became much more common in recent decades relative to previous decades. While fewer than 2% of departing CEOs became CEOs of other firms before 1986,[2] nearly 5% of departing CEOs moved to other firms to become CEO during 1986-2001 (see Table 3.2). Second, executives move across an increasingly diverse set of industries over time, consistent with the skill sets of CEOs becoming more general. For example, during 1986-2011, the proportion of new manufacturing CEO hires from non-manufacturing jobs has nearly doubled relative to the 1950-1985 era (see Table 3.1). More generally, the Herfindahl-Hirschman Index (HHI) of the fraction

[1] See also Kaplan and Minton (2012) and Graham, Kim, and Leary (2019) for related evidence.
[2] See also Vancil (1987), who shows a 2.2% fraction using a smaller sample before 1985.

of CEO moves between industry pairs, a measure of concentration of across-industry moves, was 0.187 during 1920-1949, 0.144 during 1950-1985 and decreased to 0.063 during 1986-2011, indicating that moves in the latter period have been across a wider range of industries.

Third, conditional on moving to other firms, executives typically moved to larger, more profitable, higher-paying firms, even more so in recent decades. During 1950-1985, the fraction of CEOs who move to new jobs at larger (more profitable) firms was 67% (58%), which increased to 83% (80%) during 1986-2001. In addition, CEO salary plus bonus increased by 16% when CEOs moved to other firms as executive during 1950-1985, while it increased by 105% upon moving during 1986-2001.

Fourth, another new result in our paper is documenting declining executive mobility since the early 2000s.[3] During 2002-2011 executives moved to larger and more profitable firms less often, and the magnitude of pay increases upon moving decreased relative to 1985-2001. In addition, movements of executives to new jobs in other firms became less common in the 2002-2011 period relative to 1985-2001. Consistent with these patterns, the measures of executive mobility that we construct below show an increasing-then-declining mobility trend, as well as considerable cross-sectional variation in mobility over 1920-2011 (details below).

Given that we document substantial variation in executive mobility across firms and through time, in the second part of our paper we explore the important issue of whether and how labor mobility affects executive incentives, monitoring of executives, and corporate decision-making. Fama (1980) is the first to argue that efficient labor markets for executives can solve agency problems inherent in modern corporations. If executives expect that their current actions affect the marketable value of their human capital, improved mobility would increase work incentives in their current jobs, thereby reducing the need for monitoring and compensation structure to incentivize them.

We empirically explore this general prediction in the context of career concern and

[3] This trend for executives appears to be consistent with a general decline in labor mobility in the US since 2000. See e.g., "Fewer Americans Uproot Themselves for a New Job," *The Wall Street Journal*, August 20, 2018.

dynamic agency models (e.g., Gibbons and Murphy 1992; Holmstrom 1999; DeMarzo and Fishman 2007a,b). In these models, when agents (executives in our context) can shirk or divert cash flows to themselves as private benefits, several forces motivate employees to work: implicit incentives from mobility ("career concerns"), deferred compensation, and the threat of termination. In particular, a key insight from dynamic agency models is that an optimal contract for the agent (such as the CEO) can be implemented (partly) using firms' financial and investment policies.

Thus, when an increase in executive mobility provides additional implicit incentives, incentives from corporate policies will change to offset, as the optimal contract adjusts to a new equilibrium. For example, when mobility increases, corporate debt usage (which improves incentives by increasing the threat of termination), will decrease to maintain the overall incentive (e.g., DeMarzo and Fishman 2007b). Similarly, pay-performance sensitivity will decline as executives become more mobile (Gibbons and Murphy 1992). Improved executive mobility will also substitute for board monitoring (e.g., less independent boards or appointing the CEO as board chair more often). Moreover, reduced agency costs due to improved labor-market incentives will increase the return on investment for shareholders and thus spur investment (e.g., DeMarzo and Fishman 2007a; DeMarzo et al. 2012).

We test these theoretical implications by analyzing how a positive external shock to CEO mobility (through the external labor market) affects the aforementioned corporate policies that, as previously mentioned, are part of the optimal contracting environment. The external shock in our experiment is the death of CEO A in industry A. This shock improves potential job opportunities for CEO B in connected industry B because the executive job market spans these two industries (we define "connected" to mean that CEOs from industry B moved to new jobs in industry A in the past three years; we validate this measure in several different ways below). Thus, following the shock, CEO B experiences improved labor opportunities, which increase implicit labor market incentives on CEO B (Fama 1980; Holmstrom 1999). As a result, potential agency conflicts fall in Firm B, which in turn affects the corporate policies in Firm B given their role in the new optimal

contract. In short, we estimate how the exogenous death of a CEO in industry A affects executive incentives and firm policies in industry B, where A and B are connected via the executive labor market.

We measure executive mobility at the industry level as the number of CEOs who move to other firms (both within and across industries) scaled by the lagged number of CEOs in the industry. This measure increased from the mid-1980s until the early-2000s, and then began to decline, consistent with increasing-then-decreasing mobility during the most recent three decades we discussed above. We instrument an industry's executive mobility using the value-weighted average number of CEO deaths in other, connected industries, scaled by the lagged number of CEOs in the industry. To get a sense of the magnitude of this labor market shock during our main sample period, note that when the current CEO dies in office, a majority of firms hire an external-to-the-firm replacement CEO; and that 84% of these replacements were CEOs in their previous firms, and about half of these emanate from other one-digit SIC industries. Furthermore, there is a 'chain reaction' in that the newly hired CEO also needs to be replaced at her old firm, so within one year a single CEO death results in approximately two external CEO hires across the affected industries.

We first analyze whether the implicit incentives provided by increased labor mobility lead to reduced explicit CEO incentives and/or less internal monitoring by the board. Using the instrumental variable approach, we find that an increase in executive labor market mobility leads to a significant reduction in pay-for-performance sensitivity. We also show that an increase in mobility leads to reduced monitoring of the CEO (a decrease in board independence and an increase in CEO-board chair duality), consistent with the prediction by dynamic agency models of a substitution between monitoring and incentives (e.g., Piskorski and Westerfield 2016).

We next explore the prediction from dynamic agency models that a reduction in agency problems decreases the need for implicit termination threats (such as high leverage or low financial slack) to incentivize the agent (DeMarzo and Fishman 2007b; Piskorski and Westerfield 2016). Consistent with this prediction, we find that an increase in labor market mobility leads to a decrease

in net financial leverage. We also find that firms increase investment and grow faster in response to increased labor market mobility of executives, consistent with the argument in dynamic agency models that a reduction in agency problems increases investors' return on investment (e.g., Quadrini 2004; DeMarzo and Fishman 2007a).

Finally, we investigate whether the decline in labor market mobility that we document during the 2000s alters the patterns in corporate policies just mentioned. We in fact find that the links between executive mobility and these policies become insignificant over 2001-2011 as external labor market incentives presumably declined. Overall, our results suggest that labor mobility provides executives with important incentives, which in turn affect explicit incentives, monitoring, and corporate decisions.

As discussed above, we organize our analysis of corporate policies around predictions from dynamic agency theory. In addition to the contracting theory already mentioned (which interprets an increase in mobility as an increase in the CEO's labor-market incentives), contracting models featuring limited commitment (for the CEO not to leave the firm) would also predict an increase in investment and firm growth, as well as increased pay, to match the CEO's outside option (e.g., Harris and Holmstrom 1982; Ai, Kiku, Li, and Tong 2018). In addition, bargaining models also make predictions that could help explain our results. For example, adding an assumption that CEOs prefer to manage larger firms ("empire building" as in e.g., Jensen 1986), a bargaining model between the CEO and board could generate the predictions regarding investment and firm size. Moreover, bargaining models in which the CEO is risk averse and/or values private benefits of control could generate the prediction that increased mobility will (optimally) lead to lower pay-for-performance sensitivity, board monitoring intensity, and financial leverage (e.g., Hermalin and Weisbach 1998). Thus, while we organize our paper around career concern and dynamic agency models because of the rich set of predictions inherent in these models, we recognize that other models could generate some of the same predictions.

To shed light on mechanisms underlying our results, we examine how firm value, measured by Tobin's q, changes in response to an increase in executive mobility. Using the IV approach

described above, we find that a one-SD increase in mobility leads to a 0.41% increase in Tobin's *q* two years after the shock (significant at the 1% level). While only suggestive, this finding is consistent with the view that improved executive labor market mobility provides additional incentives (Fama 1980), so that firms optimally alter other aspects of their contracts with the CEO, such as pay-for-performance. In contrast, the increased Tobin's *q* is not consistent with a class of bargaining models in which the CEO simply extracts more rent in response to an increase in mobility (and outside option) without a change in incentives.

In summary, while previous research has explored trends in the executive labor market and implications for compensation, ours is the first to study how executive mobility drives corporate decisions using a large sample of firms and executives over 90 years. Our approach has several important advantages. First, prior research uses small, selected samples of firms (often large companies) over long time horizons, or broader sets of firms over relatively short horizons. Relative to these studies, we are able to document patterns of executive mobility over a near-century using a sample of firms representing a wide variety of firm sizes.

Second, thanks to our detailed data on executive movements across firms over a long period, we document new patterns of labor market mobility. For example, we construct "heat maps" of across-industry executive mobility for several periods from 1920 to 2011, uncovering that executives move across an increasingly diverse set of industries over time. In addition, we document for the first time that departing executives' moving to new jobs has become more common over the century. In comparison, earlier work measures executive mobility by focusing on career paths of given individuals or executive turnover rates (e.g., Huson, Parrino, and Starks 2001; Murphy and Zabonjik 2007; Frydman 2017).

Third, we document a reversal in the mobility trend starting in the early 2000s; this is new relative to existing evidence that shows increasing executive mobility through the early-2000s. We also explore implications of this recent change in trends. Fourth, we explore how executive mobility affects a broad set of corporate decisions, adding to existing research that focuses on the relation between executive mobility and compensation (e.g., Custodio, Ferreira, and Matos 2013).

Our paper is also among the first to test predictions of dynamic agency models.

3.2 Data and Measurement

3.2.1 Data Sources and Sample Selection

We construct a comprehensive database of corporate officers, such as the chief executive officer (CEO), chief financial officer (CFO), various corporate vice presidents (VPs), and others, and their movements across US public firms from 1920 to 2011. We combine information from a number of sources. First, we hand-collect names of corporate executives, as well as financial data on their firms, from Moody's Industrial Manuals ('Moody's') from 1920 to 1988, and also the year 1998. Second, we collect names of corporate executives from Compact Disclosure during 1985-2005. Third, we supplement these two primary data sources using Mergent (which took over the Industrial Manual from Moody's; 2002-2009) and Board Analyst (2002-2011) for more recent years. We also gather information on corporate boards of directors from the same sources from 1920 to 2011.

CEO Compensation data are from Jensen and Murphy (1990), Frydman and Saks (2010), and Execucomp covering 1950-2011.[4] Industry-level GDP growth is from the Bureau of Economic Analysis and firm characteristics are from Compustat and Moody's. The appendix III describes definitions and sources of variables for executive, board, and firm characteristics. We exclude firms in the financial (SIC 6000-6999) and utility (4900-4999) sectors and firms whose total assets are less than $5 million in 2011-constant dollars.

Our full sample includes 184,494 firm-year observations for 17,767 unique firms and 315,423 executives (including 37,529 CEOs) from 1920 to 2011. We describe trends in executive mobility using this full sample (see the next section) and perform auxiliary tests. For our analysis of the effect of executive mobility on incentives and corporate decisions, we focus on 1986 to 2011; previous research, as well as our data, suggests that executive mobility is relatively high for

[4] We thank Kevin J. Murphy for sharing his dataset of CEOs collected from Forbes and Carola Frydman for making her dataset available on her website.

the most part (e.g., Murphy and Zabojnik 2007), though we find a reversal of this trend for the last decade of the sample.[5] The post-1986 subsample includes 67,949 firm-year observations for 8,345 unique firms and 206,127 executives (including 13,010 CEOs) from 1986 to 2011.

3.2.2 Trends in Executive Mobility in the Past Century

The breadth of the US executive labor market has changed considerably over the past 90 years. The first three tables explore whether these changes indicate improved mobility for executives, particularly CEOs. Table 3.1 presents the fraction of CEO moves from an "origin" to a "destination" industry defined at the one-digit SIC level. In our empirical analysis below (and as explained more fully in Section 3.2.4), we use the frequency of executive movements between two industries to measure how "connected" the industries are in terms of managerial skills. Panel A (1920-1949) and Panel B (1950-1985) show that before 1986, over 60% of CEO move-to-new-firms, including moves to same- and different-industry firms, occur within the manufacturing sector (SIC = 2 or 3).[6] In comparison, Panel C shows that, from 1986 through 2011, the fraction of moves between manufacturing industries decreases to 31%. This magnitude of reduction (nearly 50%) is greater than that for the general reduction in the fraction of manufacturing firms among public firms during the same period (a 31% reduction, from 71% to 49%). Thus, Panel C demonstrates a more varied set of origin and destination industries for external CEO hires in recent decades.

[Insert Table 3.1 here]

We examine the diversity of industries across which CEOs move more generally using the Herfindahl-Hirschman Index (HHI) of the fraction of moves between industry pairs. This concentration measure of between-industry moves was 0.187 during 1920-1949, 0.144 during 1950-1985, and decreased to 0.063 during 1986-2011, indicating that CEO moves in the later

[5] Our results are robust to alternative sample periods that begin in the 1980s, such as 1980-2011. In addition, we repeat our main analysis separately for the 2002-2011 period (the period during which we find executive mobility declined). See Table 3.12.

[6] 63.3% = 19.1% + 7.4% + 36.8% for 1920-1949; and 61.1% = 14.7% + 6.0% + 7.9% + 32.5% for 1950-1985.

period are more widely dispersed across industries. In addition, we find that "off-diagonal" movements (i.e., across different industries) became more frequent over the century: The fraction of across-different-industry movements was 35.3% during 1920-1949, and increased to 42.9% and 44.2% during 1950-1985 and 1986-2011.[7] The finding that CEOs moved to a more diverse set of industries in recent decades is consistent with the increasing importance of general managerial skills, as opposed to industry- or firm-specific skills, in the executive labor market (Murphy and Zabojnik 2007; Frydman 2017). Overall, the evidence in Table 3.1 suggests enhanced across-industry mobility for CEOs through time (though, as explored below, we find that this trend reversed in the recent decade).

[Insert Table 3.2 here]

In addition to the increased breadth of the managerial labor market, in recent decades executive who leave their current jobs have become increasingly more likely to accept a new job as an executive in another (public) company (versus retiring from executive work). Table 3.2, Panel A shows that among the CEOs who left their jobs during 1920-1949 and 1950-1985, only 3.6% (= 68/1,864) and 5.7% (= 382/6,697) moved to other firms to become executives, respectively. In comparison, during 1986-2001, 9.7% (= 1,222/12,594) of former CEOs became officers at other firms, representing 166% and 56% increases in the "move rate" after vacating a CEO position. A more detailed examination reveals that this increase in the rate at which CEOs move to other firms is driven by those moving from one CEO position to another CEO position. 611 out of 12,594 departing CEOs (4.9%) became CEOs at other firms during 1986-2001, while only 38 out of 1,864 (2.0%) and 115 out of 6,697 departing CEOs (1.7%) moved to other firms as CEO for the periods from 1920 to 1949 and from 1950 to 1985.[8] The fraction of former CEOs who move to become non-CEO officers at their new firms also increased, from 1.6% during 1920-1949 to 4.0% and 4.9% during the 1950-1985 and 1986-2001 periods.

[7] In Appendix Table 3.A5, we find that the fraction moderately decreased to 42.5% during 2002-2011.
[8] These fractions of CEO-to-CEO moves are similar to estimates in Vancil (1987), which show that 2.2% of 1,631 departing CEOs in his dataset become CEOs at other firms before 1985.

Comparing the two right-most columns, however, reveals that this trend has somewhat reversed in the last ten years of the sample horizon. The fraction of former CEOs moving to become executives at other firms declined from 9.7% to 8.0% between 1986-2001 and 2001-2011. The fraction of CEO-to-CEO moves has also declined from 4.9% to 4.0% over the same period.

[Insert Table 3.3 here]

Table 3.3 explores an important question: Do these CEO movements represent improved opportunities for the CEOs through the labor market (e.g., an external promotion)? Panel A shows that CEOs move to firms with different characteristics and compensation. First, over the full sample period (1920-2011), about three-fourths of CEO moves are to larger and more profitable (measured by ROA) firms (74.8% and 70.1%, respectively), and more moves involve a pay increase on average. Second, CEOs who become non-CEO executives (such as CFOs and VPs) at other firms tend to move to larger and more profitable firms more so than for those who become CEOs at other firms. Moving to non-CEO executive positions involves a pay increase that is similar to the increase for moving to new CEO positions. Thus, results for the full sample in Panel A suggest that the majority of CEO moves are likely external promotions in terms of prestige of the employer and compensation.

Importantly, compared to the 1920-1949 and 1950-1985 periods, during 1986-2001 the fraction of CEOs who move to larger firms increased by 39.0% (from 59.5% to 82.7%) and 23.4% (from 67.0% to 82.7%), respectively. Similarly, compared to the earlier two periods, over 1986-2001 the fraction of CEOs who move to more profitable firms increased substantially (from 37.8% and 58.4% to 80.2%). In addition, during 1986-2001 CEO salary and bonus more than doubled on average following a move to another firm, while during 1950-1985, the average pay increase was 15.9%. These trends suggest that a typical executive move in most of the 1980s and 1990s is more likely a promotion than in the preceding decades (in terms of job title, pay, or prestige of the new employer). These findings are consistent with Frydman and Saks (2010) and Frydman (2017), who use samples of select large public firms and show that executive compensation and mobility have risen substantially between the mid-1970s and early-2000s, arguably due to rising importance of

general managerial skills.

The last column in Panel A shows that these patterns have reversed in the last ten years of the sample period (2002-2011). In the most recent decade, the fraction of moving CEOs who join larger and more profitable firms decreased to levels similar to those in the 1950-1985 period. Relative to the 1986-2001 period, the magnitude of pay increases upon moving also declined during the recent ten years (e.g., from 104.5% to 62.9% for the case of CEO-to-CEO moves). These reversed trends, combined with the declining rate of executives moving to other firms (Table 3.2) and lack of changes in firm-imposed incentives and contracts in response to variation in mobility during this period, indicate that executive mobility as measured by external hire opportunities has declined in the recent decade. This pattern is new to the literature, and future research may want to investigate underlying reasons for the decreasing mobility.

The second part of our paper examines what happens to cooperate policies when a shock (i.e., the death of a CEO in a connected industry) improves CEO mobility across firms. Therefore, Panel B of Table 3.3 presents similar information for the subset of cross-industry moves initiated by the death of a CEO in a connected industry. All moves in the panel are CEO-to-CEO by construction. The findings are generally similar to that in Panel A, namely that cross-industry moves due to the death of a CEO are primarily to larger, more profitable firms, with partial reversal in the last ten years of the sample. Overall, these patterns suggest that implicit incentives from mobility from a CEO's current employer became more important over the past century, though they declined in the last decade.

3.2.3 Measuring Executive Mobility

We construct industry-level measures of executive mobility as the number of CEOs who move to other firms as executives (either within or across industries), scaled by the lagged number of CEOs (or separately, CEO turnover events) in a given industry and year. These measures capture the average rate at which sitting CEOs are hired by other firms at the industry level. The resulting measure is defined as follows:

$$Mobility_{i,j,t} = \frac{\#\ CEO\ moves_{-i,j,t}}{\#\ CEOs_{j,t-1}},\ \text{or}\ = \frac{\#\ CEO\ moves_{-i,j,t}}{\#\ CEO\ turnovers_{j,t-1}} \quad [1]$$

where $Mobility_{i,j,t}$ is a measure of mobility for a CEO of firm *i*, employed in one-digit SIC industry *j*, in year *t; # CEO moves*$_{-i,j,t}$ represents the number of moves between years *t-1* and year *t* by CEOs in industry *j* to become officers in another firm in any industry, excluding firm *i*'s own turnover;[9] *# CEOs* (or *# CEO turnovers*$_{j,t-1}$) is the number of CEOs (or their turnovers) in industry *j* in year *t-1*.[10] We use one-digit SIC codes to define industries, given the relatively low frequency of CEO moves. Figure 3.A1 illustrates construction of the *Mobility* measure using an example.

Figure 3.1 shows that the two measures of the executive mobility (whether deflated by the number of CEOs or CEO turnover events) move 'in parallel' from 1920 through 2011, on average (ρ = 0.93). Importantly, the measures generally trend up throughout most of the sample period, suggesting that mobility of US executives increased over most of the century. A new result in our paper is that executive mobility, measured by the frequency of across-firm moves, began to decline in the early 2000s, which is consistent with the post-2002 patterns we documented in the previous section. In Section 3.4.3, we examine the implications of the recent decline in mobility for managerial incentives and firm decisions.

[Insert Figure 3.1 here]

3.2.4 Instrumental Variables: CEO Deaths in Connected Industries

The measures of executive mobility defined above could be correlated with economic and labor market conditions. For example, mobility of executives could be correlated with business cycles, industry- or firm-level performance (e.g., Saks and Wozniak 2011; Jenter and Kanaan 2015; Chodorow-Reich and Wieland 2016).[11] In this case, an association between executive mobility and corporate decisions may not necessarily imply a causal link.

To mitigate this omitted-variable concern, we employ an instrumental variables approach

[9] Whether excluding a firm with its own CEO turnover from the analysis does not significantly affect our results.
[10] In the numerator of equation [1], we exclude cases where there is an M&A between firm *i* and other firms in years *t-2*, *t-1* and *t*.
[11] For example, the time-series correlation between the average measure of mobility based on the number of CEO moves and the US GDP growth rate is -0.21 (significant at the 5% level).

that exploits variation in mobility due to CEO deaths elsewhere in the labor market (i.e., outside a given CEO's own industry). Specifically, we instrument the mobility measures in equation [1] using the one-year lagged weighted average number of CEO deaths across connected industries divided by the number of CEOs or CEO turnovers in a given industry.[12] We measure connectedness (and the associated weight) by the fraction of CEO moves from a given ("origin") industry to each of the other ("destination") industries in the past three years (see Table 3.1).[13] Presumably, a pair of industries that shares executives is likely to share managerial human capital.[14] Thus, a sudden increase in demand for top managers in connected industries would lead to improved across-firm mobility for executives in a given industry. The main identifying assumption underlying our instrument is that executive deaths in other industries affect managerial decisions only through their impact on labor mobility. We exclude CEO deaths in a firm's own industry to avoid potential omitted-variable bias.[15] The resulting instrumental variable is defined as follows:

$$Death_{j,t-1} = \frac{\sum_{k \neq j} w_{j \to k,t-1} \, \# \, Deaths_{k,t-1}}{\# \, CEOs_{j,t-1}}, \quad or \quad \frac{\sum_{k \neq j} w_{j \to k,t-1} \, \# \, Deaths_{k,t-1}}{\# \, CEO \, turnover_{j,t-1}} \qquad [2]$$

where $Death_{j-1,t}$ is an instrumental variable for the mobility measures in equation [1], representing a shock to mobility due to CEO deaths in year *t-1* in industries connected to industry *j*; $\# \, Deaths_{k,t-1}$, represents the number of CEO deaths in industry *k* and year *t-1*; and $w_{j \to k,t-1}$ is the 'connectedness weight' and represents the fraction of CEO hires from industry *j* to industry *k*, among all moves from industry *j*, from year *t-3* to year *t-1*; $\# \, CEOs_{j,t-1}$ and $CEO \, turnovers_{j,t-1}$ are defined as in equation [1]. Figure 3.A2 illustrates construction of the instrument *Death* using an example.

To implement the instrument in equation [2], we collect CEO death events from 1950-2010

[12] We include the financial industry (SIC6) in the calculation to define connected industries.
[13] For example, from 1986 through 1988, 11% of CEO moves from the mining and construction industry (one-digit SIC = 1) are to firms in the light manufacturing industry (one-digit SIC = 2). Thus, for mining and construction, we define light manufacturing as a connected industry for 1988 and use 11% as the weight of the industry in constructing the instrument.
[14] See Tate and Yang (2017) and Kim (2018) for approaches to defining labor markets using worker moves within and across industries based on the US Census Bureau's worker-level micro data.
[15] We obtain qualitatively similar results when including CEO deaths in own industries in constructing the instrument.

from the following sources. We start with CEO death events from Salas (2010), Fee, Hadlock, and Pierce (2013), Quigley, Crossland, and Campbell (2017), and Karolyi (2018).[16] We supplement these data with our own data collection as follows. First, we collect names of CEOs who died as reported in the obituary section of Standard and Poor's Register of Corporations, Directors, and Executives ('S&P Register') from 1950 through 2010. Second, we perform news searches to collect additional CEO changes due to death at public firms from 1950 through 2010. Third, we supplement this set by examining all CEO turnover events in our database from 1950 through 2010 that are not identified above, and determine whether they are due to the death of a CEO by searching for news articles in Factiva and Google. We keep track of which CEOs passed away suddenly (e.g., due to accident, heart attack, etc.).

All total, we match 265 death events from 1950 to 2010 to our database, and use 173 CEO deaths in our main analyses for the 1986-2011 period, 87 of which we classify as sudden deaths.[17] Panel B of Table 3.3 describes characteristics of CEO moves due to the death of a CEO (see Section 3.2.2).

One potential concern with our instrument is whether a CEO death is a sufficiently large shock to the executive labor market. Table 3.4 shows that within one year (two years) following the death of a single CEO, there are on average 1.81 (3.06) external CEOs hires in connected industries that can be directly tied to the initial CEO death; these numbers are greater than one because they capture the 'chain reaction' of replacing the CEO who replaced the deceased CEO, and so on. Within five years after an initial death, 9.13 external CEO hires in connected industries can be traced to the initial death. To put these numbers in perspective, the average industry hires about 14 external CEOs per year, which suggests that the cumulative effect of the death of one CEO on the mobility of other executives is moderate-to-large in magnitude. Furthermore, there likely are unobserved chain reactions to each death event (e.g., for firms not in our sample), which

[16] We thank Charlie Hadlock, Steven Karolyi, Timothy Quigley, and Jesus Salas for sharing their datasets on CEO death events.
[17] We also collect 62 events of CEO turnovers due to health-related reasons. See Table 3.A3 for results that also incorporate CEO turnovers due to serious health issues in constructing the instrument.

will increase true executive mobility even more.

[Insert Table 3.4 here]

3.2.5 Descriptive Statistics for Firm Characteristics

Table 3.5 presents descriptive statistics for characteristics of firm-years in our main analysis sample from 1985 to 2011, including the measures of executive mobility and instruments. For a typical firm in the sample, there are 129.2 departing CEOs (i.e., turnovers) per year in its one-digit SIC industry, 14.0 of whom become executives (including CEOs) in other firms. In addition, the mean and standard deviation of the $Mobility_{Turnover}$ measure are 0.114 and 0.075, respectively, suggesting that there is considerable executive mobility and that the mobility exhibits substantial cross-sectional variation. The average number of CEO deaths in connected industries is 1.23 per year. Given the average number of CEO moves (14.0) at the industry level, this magnitude of death should induce a meaningful shock to executive mobility. Other characteristics of the CEO and board are comparable to those from previous research (e.g., Graham, Kim, and Leary 2019). For example, the average CEO tenure is 5.76 years and the ratio of the number of independent directors to all directors ("independence ratio") is 0.60.

[Insert Table 3.5 here]

3.3 Empirical Results

3.3.1 Conceptual links between labor market incentives, contracts, and corporate policies

Career concern models argue that implicit incentives from the labor market help resolve managerial incentive problems (e.g., Fama 1980; Holmstrolm 1999). In particular, Fama (1980) argues that as the labor market incorporates firm performance to determine an executive's external opportunities (e.g., becoming CEO of another firm), the CEO essentially has a "stake" in the firm's success, which induces efficient behavior. Gibbons and Murphy (1992) argue that the optimal contract balances the combination of this implicit, labor-market incentive and explicit incentives from contracts, such as pay-for-performance. Thus, this class of models predicts that

explicit contracts are more important when implicit incentives are weaker, and vice versa.

In addition, dynamic agency theory argues that firms' capital structure, investment, and monitoring of the agent are important parts of the explicit contract that provides optimal incentives. In this class of models, the agent's (an executive in our context) continuation value (i.e., "stake" in the firm) and the threat of termination or monitoring help align incentives. Because labor market mobility provides an implicit stake in the firm, an increase in mobility reduces the need for these explicit incentive and monitoring mechanisms.

Thus, career concern and dynamic agency theories predict that when executive mobility and associated incentives increase, firms will substitute for explicit incentives from contracts such as high-powered incentive pay and financial leverage, as well as internal monitoring (e.g., Gibbons and Murphy 1992; Piskorski and Westerfield 2016; DeMarzo and Fishman 2007b). In the cross-section, these effects will be more pronounced for executives with many years until retirement or high mobility, for whom the implicit incentives from the labor market are more important.

3.3.2 Executive Mobility, Incentives and Monitoring

3.3.2.1 Pay-for-Performance Sensitivity

We now explore the prediction that strong labor market incentives will substitute for incentives from contracts, starting with pay-for-performance sensitivity. As described in Section 3.2.4, we test this and other predictions by instrumenting executive mobility in a given industry and year using the number of CEO deaths in connected industries scaled by the number of CEOs in the industry in the previous year. Specifically, we estimate the following two-stage least square (2SLS) regressions:

$$Mobility_{i,j,t} = \alpha + \beta Death_{j,t-1} + \gamma X_{i,j,t} + \delta_t + \theta_{i,c} + \varepsilon_{i,j,t}, \quad [3]$$

$$Pay - Perf_{i,j,t} = \mu + \varphi \widehat{Mobility}_{i,j,t} + \rho X_{i,j,t} + \pi_t + \tau_{i,c} + \sigma_{i,j,t}, \quad [4]$$

where $Mobility_{i,j,t}$ represents our measure of executive mobility defined in equation [1];[18] $Death_{j-1,t}$ represents the weighted average number of CEO deaths in connected industries scaled by the number of CEOs in industry *j* in year *t-1*; $Pay\text{-}Perf_{i,j,t}$ is pay-for-performance sensitivity, defined as changes in CEO pay (the sum of salary and bonus) between year *t* and year *t+1* scaled by changes in market value of equity between year *t-1* and year *t* (in percentage points) (e.g., Inderst and Mueller 2010); $X_{i,j,t}$ represents a vector of control variables including one-digit SIC-level industry GDP growth rates and average Tobin's *q* (both of which control for time-varying industry-level economic conditions which may be correlated with the measure of mobility), CEO tenure, a dummy for CEO turnover, firm size (measured by log book assets), ROA, cash flow, cash holdings, leverage, asset tangibility, and market-to-book for firm *i* in industry *j* and year *t*; δ_t and π_t represent year fixed effects; $\theta_{i,c}$ and $\tau_{i,c}$ represent firm-by-CEO fixed effects; and $\varepsilon_{i,j,t}$ and $\sigma_{i,j,t}$ represent random errors double clustered both at the industry and year levels. Given that we include firm-by-CEO fixed effects in equations [3] and [4], we identify the effect of mobility on explicit incentives using within-firm-CEO variation.

[Insert Table 3.6 here]

Table 3.6 presents the estimation results based on a subsample of firm-years for which the variable for pay-for-performance is available (N = 16,386). Panel A shows that both instruments ($Death_{CEO}$ and $Death_{Turnover}$) in the first-stage regression are significant at the 10% level, indicating that executive mobility increases with the death of connected industry CEOs. Consistent with the prediction that increased labor market incentives substitute for explicit incentives, Panel B (second stage) show that the coefficients on *Mobility* are significantly negative. For example, estimates in column (1) show that a one-standard deviation (SD) increase in executive mobility leads to a 33.6-percentage-point decrease in a CEO's pay-for-performance sensitivity, which is 21.2% of its standard deviation (1.59).[19]

[18] To facilitate comparisons across two versions of this variable, we scale each by its standard deviation in all regression analysis.

[19] In unreported analysis, we find that CEO pay level increases modestly (2.8%) when CEO mobility increases by one standard deviation (*t*-stat = 0.62).

The next subsections continue to explore this issue of whether increased executive mobility affects corporate policies in ways that reduce monitoring and explicit incentives, measured by board structure and leverage.

3.3.2.2 Monitoring by the Board

In a dynamic agency framework, improved mobility of the agent (an executive in our context) increases her continuation value (or implicit "stake" in the firm). Thus, total incentives can be maintained with less intensive monitoring or a smaller termination threat, when incentives from the labor market mobility strengthen (e.g., DeMarzo and Fishman 2007a; Piskorski and Westerfield 2016). In this section, we examine this prediction using board structure as a measure of monitoring intensity. In particular, we test whether an increase in executive mobility reduces the board's monitoring intensity reflected in the fraction of independent directors (Hermalin and Weisbach 1998) and whether the CEO is appointed board chair (Graham, Kim, and Leary 2019).

[Insert Table 3.7 here]

Table 3.7, Panel A presents the first-stage estimation results for equation [3] using the full sample of firm-years from 1986 to 2011. It shows that the number of CEO deaths in related industries in the previous year is significantly positively related (at the 1% level) to the mobility of executives in a given industry and year. The F-statistics for testing the relevance of $Death_{CEO}$ and $Death_{Turnover}$ as instruments are 101.12 and 114.99, well over the usual threshold value of ten (e.g., Staiger and Stock, 1997). In the second stage shown in columns (1) and (2) of Panel B, we use the board independence ratio, defined as the ratio of the number of independent directors to total directors, as the dependent variable in equation [4]. We find that executive mobility instrumented by the scaled weighted average number of CEO deaths in connected industries has a significantly negative effect on board independence ratio (at the 1% level). Estimates in column (1), which uses $Death_{CEO}$ as the instrument, indicate that a one-SD increase in mobility leads to a

0.10-percentage-point decline in the independence ratio.[20] This reduced independence of the board is consistent with the CEO being optimally monitored less when the labor market mobility provides her with stronger incentives.

One would also expect that boards chaired by the CEO would be weaker in monitoring the CEO. The results in columns (3) and (4) of Table 3.7, Panel B suggest that CEOs who experience an increase in across-firm mobility are more likely to be board chairs, again implying less monitoring. Estimates in column (3) indicate that a one-SD increase in mobility leads to a 0.20-percentage-point increase in the probability that the CEO is board chair (significant at the 1% level). Taken together, the results in Table 3.7 are consistent with the prediction that increased mobility provides additional labor-market incentives to the CEO, which partially substitute for the board's monitoring role.

3.3.2.3 Capital Structure as a Termination Threat

In addition to explicit pay-for-performance sensitivity and board monitoring, dynamic agency models suggest that the firm's capital structure can also work as an indirect monitoring mechanism (e.g., DeMarzo and Sannikov 2006; DeMarzo and Fishman 2007b). Specifically, the firm can incentivize the CEO by using long-term debt, which increases the threat of termination, as part of an optimal contract. Thus, if heightened mobility provides stronger incentives to executives, the firm will optimally decrease debt usage in its capital structure. In addition, holding (excess) cash helps the agent avoid costly termination (e.g., losing private benefits of control or continuation value). Thus, we consider net leverage (defined as total debt minus cash holdings scaled by total assets, in percentage points) as an outcome to test the implication for termination threat. We explore this link using the instrumental variables regressions in equations [3] and [4].[21]

[Insert Table 3.8 here]

[20] In unreported analysis, we find that this decrease in board independence is due both to an increase in the number of inside directors and a decrease in the number of outside directors, with the decrease in outside directors representing two-thirds of the change in independence.

[21] In this and subsequent analyses, the first stage result is the same as that in Panel A of Table 3.7, which uses the same sample and instrumental variables.

Table 3.8 shows that an increase in executive mobility leads to a decrease in net leverage (significant at the 5% level). The coefficient on $Mobility_{CEO}$ in column (1) suggests that a one-SD increase in CEO mobility leads to a 0.11-percentage-point decrease in the net leverage ratio (mean = 5.0%). These results are consistent with the prediction from dynamic agency models that additional implicit incentives due to enhanced mobility substitutes for termination threats as a mechanism to discipline the CEO.

3.3.3 Executive Mobility, Corporate Investment and Growth

The results so far indicate that an increase in executive mobility leads to lower pay-for-performance, board monitoring intensity, and less debt usage, consistent with career concern and dynamic agency models. In this section, we turn our analysis to the relation between CEO mobility and corporate investment and growth. Specifically, we examine the prediction from dynamic agency models that firms will invest more (and grow assets faster) as the manager's implicit labor market incentives increase (e.g., Quadrini 2004; DeMarzo and Fishman 2007a; DeMarzo et al. 2012).[22] The intuition for this prediction is that the executive's increased stake in the firm due to heightened mobility mitigates agency conflicts, which in turn increases *shareholders'* return on investment. We measure investment (in percentage points) using capital expenditures scaled by total assets and estimate a variant of equation [4] that uses it as the dependent variable.

[Insert Table 3.9 here]

Table 3.9 presents the estimation results. We find that increased executive mobility precedes an increase in corporate investment (significant at the 1% level). Estimates in column (1) suggest that a one-SD increase in the mobility measure leads to a 0.13-percentage-point increase in the investment rate, which represents 2.1% of the average annual investment rate in the sample (6.1%). In unreported analysis, we also find that an increase in executive mobility leads to a

[22] Alternatively, dynamic contracting models with limited commitment (e.g., Ai et al. 2018), in which increased mobility increases the outside option of the agent, can generate the same prediction. See the Introduction for a related discussion.

significant increase in asset growth rate. These results are consistent with a dynamic agency framework in which a reduction in agency problems optimally leads to increases in firm growth and investment (e.g., DeMarzo and Fishman 2007a).

3.4 Heterogeneity in Executive Mobility and Career Concerns

3.4.1 CEO Tenure and Incentive Effect of Executive Mobility

In this section, we examine whether variation in executive mobility and career concerns, both in the cross-section and time-series, shapes the link between implicit incentives from the labor market, explicit contracts and corporate decisions.

We first explore whether the effects of executive mobility on explicit contracts differ conditional on CEO tenure. We hypothesize that the effects are stronger for CEOs with longer careers ahead of them (e.g., CEOs with short tenure) because they have greater career concerns (e.g., Gibbons and Murphy 1992). To examine this hypothesis, we interact our instrument based on deaths of other CEOs with a measure of CEO tenure as follows:[23]

$$Mobility_{i,j,t} = \alpha + \beta_1 Death_{j,t-1} + \beta_2 Short\ tenure_{i,j,t} + \beta_3 Short\ tenure_{i,j,t} \times Death_{j,t-1} + \gamma X_{i,j,t} + \delta_t + \theta_{i,c} + \varepsilon_{i,j,t}, \quad [5]$$

$$Outcome_{i,j,t} = \mu + \varphi_1 \widehat{Mobility}_{i,j,t} + \varphi_2 Short\ tenure_{i,j,t} + \varphi_3 Mobility_{i,j,t} \times \widehat{Short\ tenure}_{i,j,t} + \rho X_{i,j,t} + \pi_t + \tau_{i,c} + \sigma_{i,j,t}, \quad [6]$$

where *Short tenure*$_{i,j,t}$ is an indicator variable equal to one if CEO tenure is less than eight years (median ultimate tenure in the sample); we instrument the interaction term between *Mobility* and the *Short tenure* indicator by interacting the *Death* instrument with the indicator, assuming that this indicator variable is relatively free from omitted variable concerns (see e.g., Angrist and Pischke 2009); *Outcome*$_{i,j,t}$ is either of pay-for-performance sensitivity, board independence, an indicator for CEO-chair duality, net leverage, or capital expenditures scaled by lagged assets; and all other variables are defined in equations [3] and [4]. The main coefficient on interest is φ_3, which

[23] Another sensible proxy for the degree of career concerns is CEO age. However, the variable is not available in Moody's or Mergent; thus, we do not use it as a conditioning variable.

measures the additional effect of executive mobility for CEOs with shorter tenure.

[Insert Table 3.10 here]

Table 3.10 presents estimates of equation [6] across the five outcome variables concerning explicit incentives and corporate policies as contracts. For brevity, starting with this table we show results based on the $Mobility_{CEO}$ measure only (but results are quantitatively similar using the $Mobility_{Turnover}$ measure). Column (1), which uses pay-for-performance sensitivity as the dependent variable, shows that the coefficient on $Mobility_{CEO}$ × *Short tenure* is -66.121 and significant at the 10% level, consistent with a larger negative impact of executive mobility on explicit incentives for CEOs with longer careers ahead. Estimates in columns (2) and (3) show that in response to heightened mobility, CEOs with shorter tenures experience a greater decline in board independence (significant at the 10% level) and increase in CEO-board chair duality (insignificant at a conventional level). This finding is consistent with firms optimally reducing monitoring intensity more when the strengthened labor market incentives affect CEOs more.

Column (4), which uses the net leverage ratio as the dependent variable, shows that the negative effect of executive mobility on leverage, a measure of termination threat (or indirect monitoring), is more pronounced for firms with CEOs with short tenures ($Mobility_{CEO}$ × *Short tenure* = -0.071; *t*-stat = -3.01). In contrast, the effect on leverage is economically and statistically insignificant (-0.017; *t*-stat = -0.60) for firms headed by CEOs with tenure greater than seven, whose incentives are presumably less affected by labor market mobility. Lastly, estimates in column (5) show that the effect of mobility on investment is significantly more pronounced among CEOs with shorter tenures. Taken together, the results that the impact of improved mobility on these outcomes is more pronounced among short-tenure CEOs are consistent with career concern models (e.g., Fama 1980; Gibbons and Murphy 1992; Holmstrom 1999), in which the incentive effect of external labor markets hinges on the prospects of executives' moving to other firms.

3.4.2 State-Level Enforcement of Non-Compete Clauses

Our main measures of mobility (equation [1]) capture mobility of *average* executives in a

given industry and year. However, there could be considerable heterogeneity in mobility among executives within the same industry-year, for example, across firms in different geographical areas. This heterogeneity will lead to variation in the effects on explicit incentives and firm decisions we document. We explore non-competition agreements ('non-competes') as a driver of executive mobility across firms located in different locations (states in particular). Non-competes, which are widely used among US firms particularly for executives, create a significant legal constraint to their moving to other firms (see e.g., Kaplan and Strömberg 2003; Marx, Strumsky, and Fleming 2009; Garmaise 2011; Kini, Williams, and Yin 2018). For example, Garmaise (2011) finds that (i) 70.2% of firms in his sample from 1992 through 2004 use non-compete agreements with their top executives, and that (ii) when a state's enforcement of non-compete clauses is stricter, executives of firms located in that state become less mobile. Thus, we predict that the effect of an increase in industry-level mobility will be more pronounced in states where non-compete agreements are less strictly enforced, which increases labor mobility.

We test this prediction by estimating an IV specification similar to those in equations [5] and [6] that interacts our industry-year-level measure of executive mobility with a measure of non-compete enforceability from Garmaise for a given firm's headquarter state by year.[24] The definition of the enforceability index is in Table 3.A1. We follow Garmaise (2011) and use headquarters location (from Compustat) to determine the level of non-compete enforcement for executives.[25]

[Insert Table 3.11 here]

Table 3.11 presents the estimation results. We find that the effects of managers' mobility on pay-for-performance sensitivity (column (1)), net leverage (column (4)), and investment

[24] Non-competes generally prohibit movements within industries, whereas our instruments are based on the number of CEO deaths in other, connected industries. However, enforcement of non-competes is likely to provide variation in executive mobility in our IV specification, given that a CEO death leads to multiple executive movements, approximately half of which are moves within the same industry (see Section 3.2.4).

[25] See Garmaise (2011, p.15): "…the enforcement of non-competition agreements is governed by employment law, not corporate law, so the relevant jurisdiction is typically the one in which the employee works. Our study analyzes top executives at large firms, who will typically work at headquarters, so it is the headquarters location, not the state of incorporation that we consider."

(column (5)) are significantly weaker for firms located in states where non-compete agreements are more strongly enforced (hence reduced mobility). The estimated coefficients on $Mobility_{CEO}$ × *Non-compete* for the board monitoring variables in columns (2) and (3) have the correct sign (i.e., positive for board independence and negative for CEO-chair duality) but are not significant.

Estimates in column (1) suggest that a one-SD (2.23) increase in the enforceability index would reduce the effect of increased mobility on pay-for-performance by 14% from -38.11 to -33.45 (= -38.113 + 2.093 × 2.23). Estimates in column (5) suggest that the same magnitude of increase in non-compete enforceability would reduce the effect of improved executive mobility on investment by 19%, from 0.20 to 0.16 (= 0.201 + (-0.017 × 2.23)). Thus, the results presented in Table 3.11 are consistent with the interpretation that enforceability of non-compete laws provides an additional dimension of executive mobility, contributing to the heterogeneity in the effects of the industry-level mobility measure on explicit incentives, capital structure, and investment we document above.

3.4.3 The Effect of Market-Wide Variation in Executive Mobility

The trends in executive movements shown in Section 3.2 point toward generally increasing mobility of corporate executives in the US from the 1970s-1980s to the early 2000s.[26] In contrast, we find new evidence that executive mobility has declined since the early 2000s (see Figure 3.1). This reversal of the mobility trend during the last decade in the sample coincides with more prevalent use of non-compete agreements in executive labor contracts (Kini, Williams, and Yin 2018),[27] as well as declining CEO turnover rates and declining external CEO appointments during the 2000s (Graham, Kim, and Leary 2019).[28]

Exploiting this rich time-series variation in executive mobility over nearly a century, we address the following question: Do implicit incentives from the labor market affect explicit

[26] See e.g., Murphy and Zabojnik (2007) and Frydman (2017).
[27] Non-complete agreements are also more widely used in labor contracts in recent years (Krueger and Posner 2018).
[28] See also "More CEO Jobs Go to Insider Candidates," *The Wall Street Journal*, March 9, 2016 for evidence for a decline in the fraction of outside-the-firm CEO hiring at large US public firms over 2004 to 2015.

incentives and contracts when overall executive mobility is low? We hypothesize that low market-wide mobility makes the labor market less efficient in terms of reallocating executives conditional on their performance, thus providing weaker Fama (1980)-like incentives. To test this hypothesis, we examine whether during 1950-1985 and 2002-2011 (periods of low mobility) incentive effects are low relative to the 1986-2001 period.

[Insert Table 3.12 here]

Table 3.12 shows results from estimating IV regressions in equations [3] and [4] separately for three periods: 1950-1985, 1986-2001, and 2002-2011. We find that the effects on pay-for-performance and other measures of explicit contracts and firm decisions are indeed weaker for the two periods when market-wide mobility is lower (1950-1985 and 2002-2011). In particular, Panels A (1950-1985) and C (2002-2011) show that across the columns, the coefficients on *Mobility* are insignificant or opposite from the predictions of career concern and dynamic agency models. In contrast, Panel B shows that the effects of mobility on these incentives and corporate policies are highly significant in the 1985-2001 period, when overall executive labor market mobility is high.

This heterogeneity in the effect of mobility across time periods is consistent with the notion that low market-wide labor mobility offers little incentive to the CEO, small enough apparently that we detect no evidence of substitution for firm-implemented incentives and contracts.

3.5 Robustness Tests

3.5.1 Alternative Explanation: Product and Input-Output Market Links

Our interpretation of the findings above is that the changes in corporate decisions are due to firms substituting between explicit incentives and incentives from labor market mobility. Alternatively, changes in corporate decisions could be driven by changes in product markets or interactions of firms through supply chains. For example, when a firm's CEO dies unexpectedly, another firm that directly competes in the product market could respond by investing more aggressively and growing faster to exploit the temporarily weaker competitor. To help rule out this type of alternative explanation, we construct alternative measures of executive mobility that

exclude CEO moves between firms that produce similar products or between industries with close input-output relations.

We first construct an alternative version of the *Mobility* measure and *Death* instrument in equations [1] and [2] that excludes CEO moves between firm pairs with high product similarity (a proxy for competition). Specifically, we use the Hoberg-Phillips (2010, 2016) TNIC-3 product similarity scores, and exclude firm pairs with similarity above the annual median. This results in eliminating 4.8% of the total moves in the sample. The idea is that the remaining 95.2% of moves represent executive labor market connections above and beyond any that might occur due to close product market relations. In addition, we construct versions of the *Mobility* measure and *Death* instrument that exclude CEO moves between industry pairs with above-the-median input-output flows from the Bureau of Economic Analysis (BEA) Use Table.[29] This results in excluding 35.6% of the total moves from the sample, implying that the remaining 64.4% of moves represent executive labor market connections above and beyond input-output relations between firms.

Panel A of Table 3.13 shows that our findings are essentially unchanged after excluding moves related to product markets. The economic magnitudes remain similar to the baseline results. In unreported analysis, we alternatively use Hoberg-Phillips fixed industry classifications (FIC)-by-year fixed effects, which controls for time-varying shocks for product-market peers, and again find robust results.[30] Panel B presents results with the mobility measure and instrument that exclude CEO moves between industry pairs with high input-output relations. These estimates are similar to our baseline findings in terms of both statistical and economic significance. Thus, results in Table 3.13 are consistent with our empirical results being attributable to executive mobility and its implicit incentive effects on CEOs; we do not find evidence that our results are driven by the effects of product markets or input-output linkages that may be correlated with CEO movements or deaths.

[29] The BEA Use table describes input-output flows at the BEA industry level. We use a concordance table between BEA industry classifications and SIC industry classifications to match the Use table with our dataset.
[30] Specifically, we use FIC25, FIC50, and FIC100.

[Insert Table 3.13 here]

3.5.2 Alternative Measures and Instruments for Mobility

We employ several alternative measures of executive mobility as robustness checks of the baseline results. First, we use the three-year moving average of the baseline measures and find results that are similar to those presented in Tables 3.6 through 3.11 (unreported). Second, we expand our measurement of mobility in equation [1] to include movements of other top executives, such as CFOs and VPs, in addition to CEOs, and find similar results (Table 3.A2). With the exception of pay-for-performance sensitivity, the estimates in Table 3.A2 are comparable with the baseline results in terms of both statistical and economic significance.

Moreover, to check the robustness of our IV estimates, we employ several alternative instruments: i) the number of sudden CEO death events, ii) the number of CEO turnover events due to serious health issues as well as deaths (see Jenter et al. 2017; Graham, Kim, and Leary 2019), and iii) the number of CEOs approaching retirement age (Karolyi 2018). Panel A of Table 3.A3 shows that the results using only sudden death (87 events) in the instrument are quantitatively similar to the main results, both statistically and economically significance remaining similar across the columns. Panel B of Table 3.A3 also shows qualitatively similar results using the number of CEO turnover events due to health-related reasons (additional 62 events) as well as deaths scaled by the number of CEOs in a given industry as an alternative instrument.

Last, we use the proportion of CEOs in connected industries who are close to a retirement age as another alternative instrument for mobility of executives. While not as discrete as CEO deaths, the age measure captures the fraction of CEOs who may potentially retire, which increases *expected* mobility for executives in connected industries. We use the weighted number of CEOs in connected industries who are 63 or older, scaled by the number of CEOs in a given one-digit SIC industry and year, as an alternative instrument (see e.g., Gibbons and Murphy 1992; Jenter and Kanaan 2015).[31] Table 3.A4 shows that using the alternative instrument, the overall results are

[31] We thank Kevin J. Murphy for sharing the CEO age data.

qualitatively similar to the baseline results with the coefficients for pay-for-performance significant at the 1% level, and net leverage and investment at the 10% level. The monitoring variables (board independence and CEO-chair duality) have the correct sign on their estimated coefficients but are not significant. Relative to the death instrument, the reduced statistical significance is expected given that the age-based instrument measures a potential increase in executive mobility, as opposed to an actual increase.

3.6 Executive Mobility and Firm Value

Do changes in labor market-induced incentives affect firm value? On the one hand, enhanced implicit incentives from the labor market reduce agency conflicts, which may benefit shareholders (Fama 1980). On the other hand, we show above that increased mobility-induced incentives appear to be offset by company-implemented incentives (such as high performance pay and leverage) so any valuation effect may be small. Moreover, it is plausible that improved executive mobility increases the CEO's outside option (or bargaining power), which the CEO could use to extract rents from the firm (see a related discussion in the introduction). To shed light on this issue, we explore potential valuation effects by estimating the relation between executive mobility and Tobin's q using an IV approach, as described in equations [3] and [4], up to two years after a shock to mobility.

[Insert Table 3.14 here]

Table 3.14 presents the estimation results. As seen in Panel A, we find that an increase in executive mobility is associated with higher Tobin's q. Coefficients on *Mobility* are positive and significant one to two years after an increase in the executive mobility. In terms of economic magnitude, a one-SD increase in the measure is associated with a 0.007 increase in Tobin's q in two years out (column (3)), which represents a 0.41% increase from the average q in the sample (1.62). Panel B uses a sub-sample of firms that survive throughout the three-year period after the shock, for which changing sample composition is less of concern. The panel again shows that an improvement in executive mobility has a positive effect on firm value up to two years. While only

suggestive, this result is consistent with improved executive mobility modestly increasing firm value, on net, plausibly due to reduced agency problems and subsequent optimal contracting. In contrast, the increase in Tobin's *q* is inconsistent with alternative explanations that emphasize rent extraction by powerful CEOs (e.g., Bebchuk and Fried 2004) or the "dark side" of managerial mobility such as slow revelation of agents' ability (e.g., Acharya, Pagano, and Volpin 2016).

3.7 Conclusion

Researchers have observed notable trends in the market for corporate executives in the US over the past few decades. The level and dispersion of executive pay have increased considerably as the frequency of CEO moves and external-to-the-firm CEO hires have increased. In this paper, we uncover several new long-run trends in mobility of executives by constructing a new dataset of executive movements over the 1920-2011 period. First, movements of executives to new jobs across firms became more common in recent decades (e.g., 1986-2001) relative to previous decades over the century. Second, executives moved across an increasingly diverse set of industries over time. Third, conditional on moving, executives move to larger and more profitable firms more often, and their pay increases more, even more so in recent decades. However, we show for the first time that many of these trends have reversed over the last ten years of the sample period (2002-2011), indicating declining executive mobility in the recent decade. In addition, we find considerable cross-sectional variation in mobility over 1920-2011.

Given the substantial variation in executive mobility that we document, understanding whether and how labor mobility affects executive incentives, monitoring of executives, and corporate decision-making is an important issue. To this end, we construct measures of executive mobility that vary across industry and over time, and capture movements of corporate officers across firms. Motivated by dynamic agency models, in which explicit contracts include company policies, as well as board and compensation structures, we examine whether there is a substitution of the incentives provided by company-implemented policies to offset incentives provided by the external mobility.

Using CEO deaths in connected industries as an instrument for executive mobility, we find that increased mobility leads to lower pay-for-performance sensitivity, net leverage, and monitoring intensity (decrease in board independence; increase in CEO-chair duality). We interpret these findings as being consistent with mobility of executives mitigating agency problems, thereby substituting for explicit incentives and monitoring. Furthermore, consistent with executive mobility providing incentives to managers and thus increasing the return on investment for shareholders, we find that firms increase investment and grow faster in response to these positive shocks to mobility. Overall, our paper is among the first to show that the labor market for corporate executives provides dynamic incentives, which, in turn, affect key incentive compensation, corporate governance and financial decisions of firms.

3.8 References

Chodorow-Reich, Gabriel and Johannes Wieland, 2016, "Secular Labor Reallocation and Business Cycle", *working paper*, Available at SSRN: https://ssrn.com/abstract=2713589.

Custodio, Claudia, Miguel A. Ferreira, and Pedro Matos, 2013, "Generalists versus specialists: Lifetime work experience and chief executive officer pay", *Journal of Financial Economics* 108 (2), 471-492.

DeMarzo, Peter M. and Michael J. Fishman, 2007a, "Agency and Optimal Investment Dynamics", *The Review of Financial Studies* 20 (1), 151-188.

DeMarzo, Peter M. and Michael J. Fishman, 2007b, "Optimal Long-Term Financial Contracting", *The Review of Financial Studies* 20 (6), 2079-2128.

DeMarzo, Peter M., Michael J. Fishman, Zhiguo He, and Neng Wang, 2012, "Dynamic Agency and the q Theory of Investment", *The Journal of Finance* 67 (6), 2295-2340.

Fama, Eugene F., 1980, "Agency Problems and the Theory of the Firm", *Journal of Political Economy* 88 (2), 288-307.

Fee, C. Edward, Charles J. Hadlock, and Joshua R. Pierce, 2013, "Managers with and without Style: Evidence Using Exogenous Variation", *The Review of Financial Studies* 26 (3), 567-601.

Frydman, Carola and Raven E. Saks, 2010, "Executive Compensation: A New View from a Long-Term Perspective, 1936-2005", *The Review of Financial Studies* 23 (5), 2099-2138.

Frydman, Carola, 2017, "Rising Through the Ranks: The Evolution of the Market for Corporate Executives, 1936-2003" Forthcoming, *Management Science*.

Garmaise, Mark J., 2011, "Ties that Truly Bind: Non-competition Agreements, Executive Compensation and Firm Investment", *Journal of Law, Economics, & Organizatio*n 27 (2), 376-425.

Gibbons, Robert and Kevin J. Murphy, 1992, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence", *Journal of Political Economy* 100 (3), 468-505.

Graham, John R., Hyunseob Kim, and Mark Leary, 2019, "CEO-Board Dynamics", *working paper*, Available at SSRN: https://ssrn.com/abstract=2938120.

Hoberg, Gerard and Gordon Phillips, 2010, "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis", *The Review of Financial Studies* 23 (10), 3773-3811.

Hoberg, Gerard and Gordon Phillips, 2016, "Text-Based Network Industries and Endogenous Product Differentiation", *Journal of Political Economy* 124 (5), 1423-1465.

Hermalin, Benjamin and Michael Weisbach, 1998, "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO", *The American Economic Review* 88 (1), 96-118.

Holmstrom, Bengt, 1999, "Managerial Incentive Problem: A Dynamic Perspective", Review of Economic Studies 66 (1), 169-182.

Huson, Mark R., Robert Parrino, and Laura T. Starks, 2001, "Internal Monitoring Mechanisms and CEO Turnover: A Long-Term Perspective", *The Journal of Finance* 56 (6), 2265-2297.

Inderst, Roman and Holger M. Mueller, 2010, "CEO Replacement Under Private Information", *The Review of Financial Studies* 23 (8), 2935-2969.

Jenter, Dirk and Fadi Kanaan, 2015, "CEO Turnover and Relative Performance Evaluation", *The Journal of Finance* 70 (5), 2155-2183.

Jenter, Dirk, Egor Matveyev, and Lukas Roth, 2017, "Good and Bad CEOs", *working paper*.

Kaplan, Steven N. and Per Strömberg, 2003, "Financial Contracting Theory Meets the Real World:

An Empirical Analysis of Venture Capital Contracts", *The Review of Economic Studies* 70 (2), 281-315.

Kaplan, Steven N. and Bernadette A. Minton, 2012, "How has CEO turnover changed?", *International Review of Finance* 12 (1), 57-87.

Karolyi, Stephen, 2018, "Personal lending relationships", *Journal of Finance* 73, 5-49.

Kim, Hyunseob, 2018, "How Does Labor Market Size Affect Firm Capital Structure? Evidence from Large Plant Openings", *working paper*, Available at SSRN: https://ssrn.com/abstract=2147932.

Kini, Omesh, Ryan Williams, and Sirui Yin, 2018, "Restrictions on CEO Mobility, Performance-Turnover Sensitivity, and Compensation: Evidence from Non-compete Agreements", *working paper*, Available at SSRN: https://ssrn.com/abstract=3170804.

Krueger, Alan and Eric Posner, 2018, "A Proposal for Protecting Low-Income Workers from Monopsony and Collusion", *The Hamilton Project*.

Marx, Matt, Deborah Strumsky, and Lee Fleming, 2009, "Mobility, Skills, and the Michigan Non-Compete Experiment", *Management Science* 55 (6), 875-889.

Murphy, Kevin J. and Jan Zabojnik, 2007, "Managerial Capital and the Market for CEOs", *working paper*, Available at SSRN: https://ssrn.com/abstract=984376.

Quadrini, Vincenzo, 2004, "Investment and Liquidation in Renegotiation-proof Contracts with Moral Hazard", *Journal of Monetary Economics* 51 (4), 713-751.

Quigley, Timothy J., Craig Crossland, and Rober J. Campbell, 2017, "Shareholder perceptions of the changing impact of CEOs: Market reactions to unexpected CEO deaths, 1950–2009", *Strategic Management Journal* 38 (4), 939-949.

Saks, Raven E. and Abigail Wozniak, 2011, "Labor Reallocation over the Business Cycle: New Evidence from Internal Migration" *Journal of Labor Economics* 29 (4), 697-739.

Tate, Geoffrey A. and Liu Yang, 2017, "The Human Factor in Acquisitions: Cross-industry Labor Mobility and Corporate Diversification", *working paper,* Available at SSRN: https://ssrn.com/abstract=2578636.

Vancil, Richard F., 1987, "Passing the Baton: Managing the Process of CEO Succession", *Harvard Business School Press*.

3.9 Figures and Tables

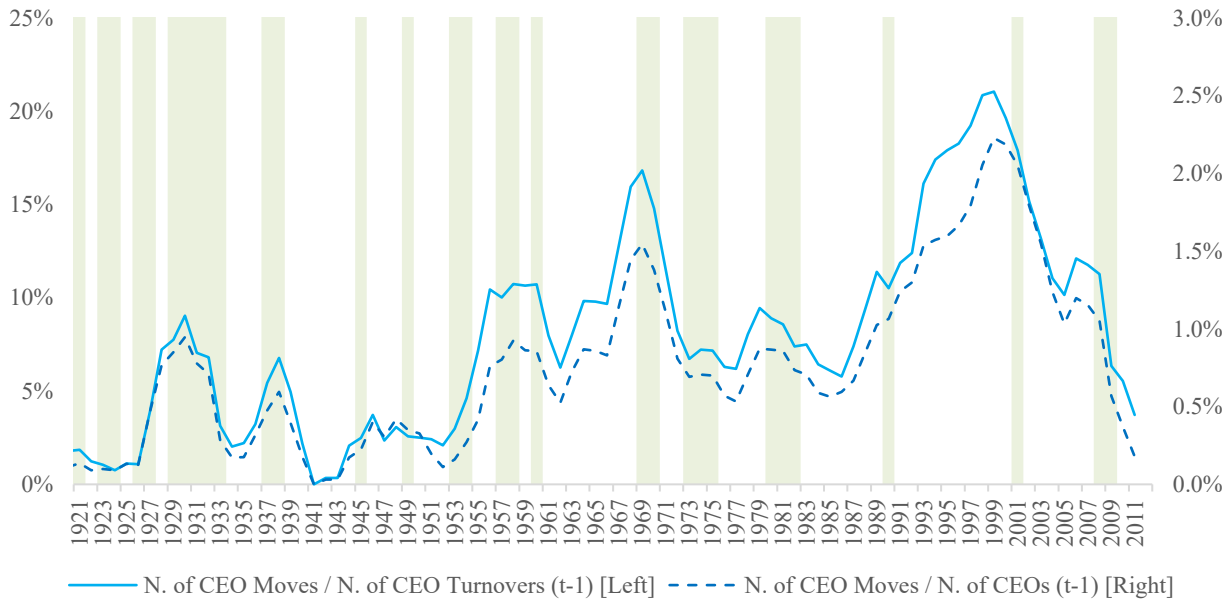


Figure 3.1. Measures of Executive Mobility

This figure plots two measures of executive mobility averaged across one-digit SIC industries from 1920 through 2011 (three-year moving average). *N. of CEO Moves* indicates the number of CEOs who become CEOs or other executive officers at other firms in our sample within 2 years after their departure. *N. of CEO Turnovers* indicates the number CEOs who leave firms in a given year. *N. of CEOs* indicates the total number of CEOs a given year. The correlation between the two measures is 0.928. Shaded area indicates NBER recession periods.

Table 3.1. Frequency of CEO Moves between Industries

This table presents the frequency of CEOs moving from a firm in the origination industry to another firm in the destination industry for three time periods from 1920 through 2011. *SIC-From* indicates the one-digit SIC industry of the firm that a given CEO departs and *SIC-To* indicates the one-digit SIC industry of the firm to which the CEO moves as CEO or another executive position. The mean and Herfindahl-Hirschman Index of the frequency across industries are in the parentheses.

Panel A. 1920-1949: 68 moves (Mean 5.26%, HHI 0.187)

SIC-From \ SIC-To	SIC0 Agric., Forestry, Fishing	SIC1 Mining & Constr.	SIC2 Light Manuf.	SIC3 Heavy Manuf.	SIC4 Transport. & Public Utilities	SIC5 Wholesale & Retail Trade	SIC6 Finance, Insurance, Real state	SIC7 Services	SIC8 Health Services	SIC9 Public Admin.	Total
0.Agriculture, Forestry, Fishing											0.0%
1.Mining & Construction		2.94%	2.94%	2.94%		1.47%					10.3%
2.Light Manufacturing			19.12%		1.47%		2.94%				23.5%
3.Heavy Manufacturing		1.47%	7.35%	36.76%		2.94%					48.5%
4.Transportation & Public Utilities				1.47%	1.47%						2.9%
5.Wholesale & Retail Trade			1.47%	1.47%		4.41%					7.4%
6.Finance, Insurance, Real Estate		2.94%	2.94%								5.9%
7.Services			1.47%								1.5%
8.Health Services											0.0%
9.Public Administration											0.0%
Total	0.0%	7.4%	35.3%	42.6%	2.9%	8.8%	2.9%	0.0%	0.0%	0.0%	100%

Panel B. 1950-1985: 382 moves (Mean 2.17%†, HHI 0.144)

SIC-From \ SIC-To	SIC0 Agric., Forestry, Fishing	SIC1 Mining & Constr.	SIC2 Light Manuf.	SIC3 Heavy Manuf.	SIC4 Transport. & Public Utilities	SIC5 Wholesale & Retail Trade	SIC6 Finance, Insurance, Real state	SIC7 Services	SIC8 Health Services	SIC9 Public Admin.	Total
0.Agriculture, Forestry, Fishing	0.26%										0.3%
1.Mining & Construction	0.26%	3.40%	1.31%	1.31%	0.52%	0.26%	0.26%		0.26%		7.6%
2.Light Manufacturing		0.52%	14.66%	6.02%	0.79%	2.36%		1.31%			25.7%
3.Heavy Manufacturing		1.05%	7.85%	32.46%	0.52%	1.83%	1.57%	0.52%	0.79%		46.6%
4.Transportation & Public Utilities				0.52%		1.05%		0.26%			2.1%
5.Wholesale & Retail Trade			2.09%	3.14%	0.26%	4.97%	0.52%		0.26%		11.3%
6.Finance, Insurance, Real Estate			0.26%	0.52%	0.26%	0.26%	0.26%		0.26%		1.8%
7.Services			0.79%	1.05%		0.79%	0.52%	1.05%			4.2%
8.Health Services			0.26%	0.26%							0.5%
9.Public Administration											0.0%
Total	0.5%	5.0%	27.2%	45.3%	2.6%	11.5%	3.1%	3.1%	1.6%	0.0%	100%

indicates top 10%, top 20%, top 30% and top 50% industry pairs in terms of frequency of moves

† indicates t-stat for Mean difference from the previous period

*** p<0.01, ** p<0.05, * p<0.1

Table 3.1. Frequency of CEO Moves between Industries (Continued)

Panel C. 1986-2011: 1,655 moves (Mean 1.27%†, HHI 0.063)

SIC-From \ SIC-To	SIC0 Agric., Forestry, Fishing	SIC1 Mining & Constr.	SIC2 Light Manuf.	SIC3 Heavy Manuf.	SIC4 Transport. & Public Utilities	SIC5 Wholesale & Retail Trade	SIC6 Finance, Insurance, Real state	SIC7 Services	SIC8 Health Services	SIC9 Public Admin.	Total
0.Agriculture, Forestry, Fishing	0.06%	0.06%		0.12%		0.06%			0.06%		0.4%
1.Mining & Construction		4.95%	0.73%	1.03%	0.54%	0.36%	0.12%	0.36%	0.18%		8.3%
2.Light Manufacturing	0.12%	0.24%	8.82%	2.84%	0.54%	1.45%	0.30%	0.42%	1.03%	0.12%	15.9%
3.Heavy Manufacturing		0.42%	3.02%	16.80%	1.81%	1.51%	0.42%	3.02%	0.85%	0.06%	27.9%
4.Transportation & Public Utilities		0.30%	0.48%	1.09%	6.04%	0.18%	0.18%	1.27%	0.18%		9.7%
5.Wholesale & Retail Trade	0.06%	0.30%	1.15%	1.75%	0.48%	6.04%	0.30%	0.97%	0.66%		11.7%
6.Finance, Insurance, Real Estate		0.12%	0.42%	0.73%	0.24%	0.42%	0.48%	0.73%	0.12%		3.3%
7.Services	0.06%	0.12%	0.60%	2.54%	0.48%	0.85%	0.48%	10.21%	1.09%		16.4%
8.Health Services	0.12%	0.18%	1.03%	0.73%	0.24%	0.48%	0.12%	0.91%	2.36%	0.06%	6.2%
9.Public Administration		0.06%		0.06%		0.06%					0.2%
Total	0.4%	6.8%	16.3%	27.7%	10.4%	11.4%	2.4%	17.9%	6.5%	0.2%	100%

indicates top 10%, top 20%, top 30% and top 50% industry pairs in terms of frequency of moves

† indicates t-stat for Mean difference from the previous period

*** p<0.01, ** p<0.05, * p<0.1

Table 3.2. CEO Departures and New Job Titles

This table presents the number of CEOs who leave office and the number of these CEOs who are hired at new firms in our sample from 1920 through 2011. The new title is the first new job title of an externally hired former CEO, with *Non-CEO* indicating a move to a non-CEO role at the new firm (e.g., CFO). Numbers in parentheses indicate the number of externally hired former CEOs divided by the number of CEO turnovers in each period. Panel B data are only for CEO-to-CEO moves by construction.

Panel A. CEO Departures and New Job Titles

	All	1920-1949	1950-1985	1986-2001	2002-2011
CEO Turnovers	26,559	1,864	6,697	12,594	5,404
- Become officer of new firm (%)	2,105	68	382	1,222	433
	(7.9%)	(3.6%)	(5.7%)	(9.7%)	(8.0%)
- Become CEO of new firm (%)	981	38	115	611	217
	(3.7%)	(2.0%)	(1.7%)	(4.9%)	(4.0%)
- Become Non-CEO officer of new firms (%)	1,124	30	267	611	216
	(4.2%)	(1.6%)	(4.0%)	(4.9%)	(4.0%)

Panel B. New Job Titles of departing CEOs

	All	1950-1985	1986-2001	2002-2011
CEO	46.7%	30.1%	50.0%	50.1%
President/Vice-President	4.8%	5.0%	5.2%	4.4%
CFO	3.3%	1.9%	3.6%	3.9%
Other Executive Job Titles	45.2%	63.0%	46.4%	41.6%

Table 3.3. Firm Size, Profitability and Compensation at Moving CEO's New Firm

For CEOs who move to other firms from 1920 through 2011, this table reports the size and profitability of the new firm and the mover's change in compensation. Not all CEOs are hired as CEO at their new firm, and 'New title' indicates the first new job title of an externally hired former CEO. 'Larger Firms' indicates the proportion of former CEOs whose new firms are larger than their previous firms in terms of total assets. For example, for 63.39% of CEOs who became CEO of a new firm, their new firm is larger than their previous firm in the full sample ("All" column). 'More Profitable Firms' indicates the proportion of former CEOs who are hired by firms with higher ROA. 'Pay Change' indicates the difference between an externally hired CEO's first salary and bonus and the same from previous employment. The bottom row is 'Pay Change' divided by the CEO's most recent salary and bonus from previous employment. Panel B is for the death sample and thus contains only CEO-to-CEO moves by construction.

Panel A. Entire Sample (2,105 moves)

	All	1920-1949	1950-1985	1986-2001	2002-2011
New Firms					
Larger Firms	74.8%	59.5%	67.0%	82.7%	67.1%
New Title: CEO	63.3%	45.5%	56.8%	66.8%	64.0%
New Title: Non-CEO	76.5%	65.4%	71.3%	84.2%	70.2%
More Profitable Firms	70.1%	37.8%	58.4%	80.2%	62.4%
New Title: CEO	54.4%	42.9%	32.4%	63.6%	55.8%
New Title: Non-CEO	71.7%	30.8%	59.9%	82.1%	69.0%
CEO Pay†					
Pay Change ($ Thousands)	481.2	-	45.7	544.8	426.8
New Title: CEO	513.4	-	49.1	562.0	477.3
New Title: Non-CEO	466.7	-	32.5	525.7	417.4
Pay Change / (Salary+Bonus)	88.9%	-	15.9%	104.5%	74.2%
New Title: CEO	101.8%	-	17.0%	116.2%	88.8%
New Title: Non-CEO	71.3%	-	11.8%	80.8%	62.9%

†Data is from Execucomp and Frydman and Saks (2010). There are 12 obs for 1950-1985, 327 obs for 1986-2001 and 286 for 2002-2011.

Panel B. CEO moves due to CEO death (206 moves)

	All	1950-1985	1986-2001	2002-2011
New Firms				
Larger Firms	63.8%	56.1%	73.0%	56.9%
More Profitable Firms	63.3%	60.6%	65.2%	62.7%
CEO Pay‡				
Pay Change ($ Thousands)	422.9	-	568.7	306.4
Pay Change / (Salary+Bonus)	187.4%	-	183.2%	189.2%

‡Data are from Execucomp and Frydman and Saks (2010). There are no observations for 1950-1985, 56 obs for 1986-2001 and 24 obs for 2002-2011.

Table 3.4. Impact of CEO Death Events on Subsequent External Hires

This table reports the number of external CEOs hired in the sample and the number of external CEO hires following pertinent CEO deaths from 1985 through 2010 (*t-1* death).

	N
External CEOs	17,467
External CEO Hires after CEO Deaths	173
"Chain Reaction" Hires	
Number of CEOs (replacing CEOs) hired within 1 year to replace CEOs who died or who left to replace a deceased CEO, and so on	1.81
Number of CEOs (replacing CEOs) hired within 2 years to replace CEOs who died or who left to replace a deceased CEO, and so on	3.06
Number of CEOs (replacing CEOs) hired within 5 years to replace CEOs who died or who left to replace a deceased CEO, and so on	9.13

Table 3.5. Descriptive Statistics

This table reports sample summary statistics from 1986 through 2011. *N. of CEO Moves* indicates the number of CEOs in a one-digit SIC industry who become CEOs or other executive officers at other firms within 2 years of departing their most recent firm. *N. of CEO Turnovers* indicates the total number of one-digit SIC industry CEOs who left firms each year. *N. of CEOs* indicates the total number of one-digit SIC industry CEOs each year. *Mobility* indicates *N. of CEO Moves*$_t$ divided by *N. of CEO Turnovers*$_{t-1}$ or *N. of CEOs*$_{t-1}$. *Connected Industry Death* is the weighted average number of CEOs in connected one-digit SIC industries who pass away in a year. Both weight and connectedness are determined by the external CEO hires over the past three years between one-digit SIC industries. *Death* indicates *Connected Industry Death*$_{t-1}$ divided by *N. of CEO Turnovers*$_{t-1}$ or *N. of CEOs*$_{t-1}$ respectively. *Independence* indicates the number of outside directors who are not affiliated with a CEO based on last names divided by the total number of directors. *CEO-Chair* is a dummy variable that takes the value of 1 if a CEO is also the Chair and 0 otherwise. CEO Turnover takes the value of 1 if the CEO changes and 0 otherwise. Total CEO Tenure is the total number of years a CEO spends at a firm (ultimate tenure). *Pay-Perf* indicates changes in a CEO's salary and bonus from *t* to *t+1* over changes in firm value from *t-1* to *t*. Net Leverage is total debt less cash divided by total assets. Firm and Industry Characteristic variables are defined in the Appendix III.

	N	Mean	Median	Stddev
Executive Mobility Characteristics				
N. of CEO Moves	67,949	14.0	12.0	9.6
N. of CEOs	67,949	995.7	870.0	550.3
N. of CEO Turnovers	67,949	129.2	112.0	76.0
Mobility$_{CEO}$	67,949	0.006	0.006	0.004
Mobility$_{Turnover}$	67,949	0.114	0.113	0.075
Connected Industry Death	67,949	1.23	1.12	0.76
Death$_{CEO}$	67,949	0.08%	0.05%	0.19%
Death$_{Turnover}$	67,949	1.61%	0.92%	4.04%
CEO & BOD Characteristics				
N. of Directors	67,949	7.5	7.0	4.2
Independence	67,949	0.600	0.636	0.217
CEO-Chair = 1	67,949	0.485	0.000	0.500
CEO Turnover = 1	67,949	0.123	0.000	0.329
CEO Tenure	67,949	5.762	4.000	4.946
Total CEO Tenure	67,949	9.795	8.000	6.195
Pay-Perf	16,386	0.015	0.003	1.597
Firm Characteristics				
Leverage	67,949	0.227	0.187	0.224
Net Leverage	67,949	0.050	0.088	0.388
Cash flow	67,949	0.188	0.263	1.866
Size	67,949	5.599	5.442	2.052
ROA	67,949	-0.031	0.034	0.263
PPE/TA	67,949	0.281	0.219	0.224
M/B	67,949	2.882	2.000	2.539
Investment	67,949	0.061	0.041	0.066
CASH/TA	67,949	0.180	0.091	0.215
Tobin's *q*	67,949	1.618	1.185	1.185
Industry Characteristics				
Industry Tobin's *q*	67,949	2.607	2.356	1.182
Industry GDP Growth	67,949	0.044	0.050	0.034

Table 3.6. Executive Mobility and Pay-for-Performance Sensitivity

This table presents 2SLS estimation results for the effect of executive mobility on the CEO's pay-for-performance sensitivity from 1986 through 2011. *Pay-Perf* indicates changes in a CEO's salary and bonus from year *t* to *t+1* divided by changes in firm value from year *t-1* to *t*. The instrumental variable is *Death*, which is calculated as the weighted number of connected industry CEOs who passed away at *t-1* divided by *N. of CEOs*$_{t-1}$ or *N. of CEO Turnovers*$_{t-1}$. Each weight and connectedness is determined by past three-year external CEO hires between one-digit SIC industries. Numbers in parentheses are *t*-statistics based on standard errors clustered at the industry and year levels.

Panel A. First Stage

Dependent Variables:	$Mobility_{CEO}$	$Mobility_{Turnover}$
	(1)	(2)
$Death_{CEO}$	1.864*	
	(2.15)	
$Death_{Turnover}$		0.103*
		(2.23)
Industry GDP Growth	0.007	0.007
	(0.76)	(0.80)
Industry Tobin's *q*	-0.001	-0.001
	(-1.26)	(-1.32)
CEO Turnover	-0.000	-0.000
	(-0.43)	(-0.36)
Lagged CEO Turnover	-0.000	-0.000
	(-1.07)	(-1.10)
CEO Tenure	0.000	0.000
	(0.12)	(0.12)
Size	-0.000	-0.000
	(-0.70)	(-0.69)
ROA	0.000	0.000
	(0.06)	(0.09)
Cash Flow	0.000	0.000
	(0.44)	(0.51)
Cash/TA	0.000	0.000
	(0.51)	(0.46)
Leverage	0.001	0.001
	(0.63)	(0.60)
PPE/TA	-0.002	-0.002
	(-1.72)	(-1.65)
M/B	-0.000	-0.000
	(-0.18)	(-0.17)
N	16,386	16,386
Year Fixed Effect	YES	YES
CEO-Firm Fixed Effect	YES	YES
F-stat	23.80	5.50
R^2	0.613	0.614

*** p<0.01, ** p<0.05, * p<0.1

Panel B. Second Stage

Dependent Variables:	*Pay-Perf*	
	(1)	(2)
$Mobility_{CEO}$	-33.653**	
	(-2.24)	
$Mobility_{Turnover}$		-39.220*
		(-1.85)
Industry GDP Growth	0.378	0.463
	(1.00)	(1.20)
Industry Tobin's *q*	-0.010	-0.013
	(-0.74)	(-0.80)
CEO Turnover	-0.049	-0.048
	(-0.91)	(-0.88)
Lagged CEO Turnover	0.037	0.037
	(0.82)	(0.85)
CEO Tenure	0.007	0.007
	(0.48)	(0.48)
Size	-0.083**	-0.084**
	(-2.52)	(-2.48)
ROA	0.054	0.050
	(0.53)	(0.49)
Cash Flow	0.004	0.004
	(0.27)	(0.30)
Cash/TA	-0.178	-0.176
	(-1.08)	(-1.07)
Leverage	-0.329***	-0.330***
	(-4.33)	(-4.09)
PPE/TA	-0.792**	-0.808**
	(-2.29)	(-2.29)
M/B	0.026***	0.025***
	(3.21)	(3.35)
N	16,386	16,386
IV	$Death_{CEO}$	$Death_{Turnover}$
Year Fixed Effect	YES	YES
CEO-Firm Fixed Effect	YES	YES
R^2	0.02	0.01

*** p<0.01, ** p<0.05, * p<0.1

Table 3.7. Effect of Executive Mobility on Board Monitoring

This table presents 2SLS estimation results for the effect of executive mobility on board monitoring from 1986 through 2011. Monitoring intensity is measured by board independence and whether the CEO is also board chair. *Independence* indicates the number of independent directors divided by the total number of directors. *CEO-Chair* takes the value of 1 if a CEO serves as chair of the board and 0 otherwise. The instrumental variable is *Death*, which is calculated as the weighted number of connected industry CEOs who passed away at *t-1* divided by $N.\ of\ CEOs_{t-1}$ or $N.\ of\ CEO\ Turnovers_{t-1}$. Each weight and connectedness is determined by past three-year external CEO hires between one-digit SIC industries. Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Panel A. First Stage

Dependent Variable:	$Mobility_{CEO}$	$Mobility_{Turnover}$
	(1)	(2)
$Death_{CEO}$	3.726***	
	(10.06)	
$Death_{Turnover}$		0.188***
		(10.72)
Industry GDP Growth	-0.008	-0.013
	(-0.35)	(-0.69)
Industry Tobin's *q*	-0.000	-0.000
	(-0.92)	(-0.43)
CEO Turnover	-0.000	-0.000
	(-0.57)	(-0.57)
Lagged CEO Turnover	-0.000	-0.000
	(-0.98)	(-1.32)
CEO Tenure	-0.000	-0.000
	(-0.24)	(-0.10)
Size	0.000	0.000
	(1.00)	(0.43)
ROA	-0.000	-0.000
	(-1.03)	(-0.47)
Cash Flow	0.000	0.000
	(1.11)	(1.43)
Cash/TA	-0.000	-0.000
	(-0.01)	(-0.48)
Leverage	0.000	0.000
	(0.11)	(0.44)
PPE/TA	0.000	0.000
	(0.20)	(0.13)
M/B	-0.000	-0.000
	(-0.21)	(-0.04)
N	67,949	67,949
Year Fixed Effect	YES	YES
CEO-Firm Fixed Effect	YES	YES
F-stat	101.12	114.99
R^2	0.733	0.757

*** p<0.01, ** p<0.05, * p<0.1

Panel B. Second Stage

Dependent Variables:	*Independence*		*CEO-Chair*	
	(1)	(2)	(3)	(4)
$Mobility_{CEO}$	-0.100***		0.200***	
	(-4.92)		(4.71)	
$Mobility_{Turnover}$		-0.087***		0.158***
		(-4.96)		(9.03)
Industry GDP Growth	-0.096***	-0.096***	-0.050***	-0.049***
	(-6.90)	(-7.05)	(-2.88)	(-2.60)
Industry Tobin's *q*	0.000	0.000	0.000	0.000
	(0.46)	(0.49)	(0.29)	(0.26)
CEO Turnover	-0.001	-0.001	-0.101***	-0.101***
	(-0.86)	(-0.86)	(-14.05)	(-14.06)
Lagged CEO Turnover	-0.002	-0.002	-0.043***	-0.043***
	(-0.91)	(-0.91)	(-12.15)	(-12.16)
CEO Tenure	0.000	0.000	-0.002	-0.002
	(0.57)	(0.57)	(-1.28)	(-1.28)
Size	0.009***	0.009***	0.017**	0.017**
	(9.53)	(9.49)	(2.34)	(2.34)
ROA	0.003	0.003	-0.009	-0.009
	(0.81)	(0.82)	(-1.14)	(-1.14)
Cash Flow	-0.001**	-0.001**	0.003***	0.003***
	(-2.55)	(-2.55)	(19.95)	(20.11)
Cash/TA	0.004	0.004	-0.022**	-0.022**
	(0.32)	(0.32)	(-2.11)	(-2.11)
Leverage	-0.005	-0.005	0.032***	0.032***
	(-0.55)	(-0.55)	(3.02)	(3.02)
PPE/TA	-0.000	-0.000	-0.050**	-0.050**
	(-0.02)	(-0.02)	(-2.25)	(-2.24)
M/B	0.001**	0.001**	0.001**	0.001**
	(2.02)	(2.02)	(2.23)	(2.22)
N	67,949	67,949	67,949	67,949
IV	$Death_{CEO}$	$Death_{Turnover}$	$Death_{CEO}$	$Death_{Turnover}$
Year Fixed Effect	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES
R^2	0.06	0.06	0.11	0.11

*** p<0.01, ** p<0.05, * p<0.1

Table 3.8. Effect of Executive Mobility on Capital Structure

This table presents second stage 2SLS estimation results for the effect of executive mobility on capital structure decisions from 1986 through 2011. *Net Leverage* is total debt less cash divided by total assets. The instrumental variable is $Death_{CEO}$. The first stage results are in Table 3.7, Panel A. Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variable: Net Leverage	(1)	(2)
$Mobility_{CEO}$	-0.109**	
	(-2.12)	
$Mobility_{Turnover}$		-0.098**
		(-2.15)
Industry GDP Growth	-0.103*	-0.103*
	(-1.85)	(-1.85)
Industry Tobin's *q*	-0.002	-0.002
	(-1.51)	(-1.49)
CEO Turnover	-0.018***	-0.018***
	(-5.45)	(-5.45)
Lagged CEO Turnover	-0.013***	-0.013***
	(-6.49)	(-6.49)
CEO Tenure	-0.000	-0.000
	(-0.04)	(-0.04)
Size	0.044***	0.044***
	(9.11)	(9.11)
ROA	-0.212***	-0.212***
	(-3.87)	(-3.87)
Cash Flow	-0.000	-0.000
	(-0.00)	(-0.00)
PPE/TA	0.748***	0.748***
	(13.83)	(13.83)
M/B	0.000	0.000
	(0.42)	(0.42)
N	67,949	67,949
Year Fixed Effect	YES	YES
CEO-Firm Fixed Effect	YES	YES
F-stat	101.12	114.99
R^2	0.14	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.9. Effect of Executive Mobility on Corporate Investment

This table presents second stage 2SLS estimation results for the effect of executive mobility on corporate investment from 1986 through 2011. *Investment* is capital expenditures divided by total assets. The instrumental variable is $Death_{CEO}$. The first stage results are in Table 3.7, Panel A. Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variable: Investment	(1)	(2)
$Mobility_{CEO}$	0.134***	
	(4.03)	
$Mobility_{Turnover}$		0.120***
		(4.11)
Industry GDP Growth	0.023	0.023
	(0.74)	(0.78)
Industry Tobin's *q*	0.001*	0.001*
	(1.92)	(1.86)
CEO Turnover	-0.003***	-0.003***
	(-5.86)	(-5.84)
Lagged CEO Turnover	-0.002***	-0.002***
	(-3.08)	(-3.07)
CEO Tenure	-0.000*	-0.000*
	(-1.76)	(-1.76)
Size	0.004***	0.004***
	(3.55)	(3.56)
ROA	-0.000	-0.000
	(-0.11)	(-0.12)
Cash Flow	-0.000***	-0.000***
	(-3.77)	(-3.75)
Cash/TA	-0.001	-0.001
	(-0.52)	(-0.51)
Leverage	-0.026***	-0.026***
	(-5.30)	(-5.30)
PPE/TA	0.194***	0.194***
	(16.88)	(16.89)
M/B	0.002***	0.002***
	(4.90)	(4.91)
N	67,949	67,949
Year Fixed Effect	YES	YES
CEO-Firm Fixed Effect	YES	YES
F-stat	101.12	114.99
R^2	0.14	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.10. Executive Career Concern and the Effects of Executive Mobility on Incentives and Corporate Decisions

This table presents second stage 2SLS estimation results for the interactive effect of executive mobility measures with executive career concerns on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment from 1986 through 2011. *Short tenure* takes the value of 1 if a CEO's tenure is less than the median CEO's ultimate tenure (8 years), and zero otherwise. Instrumental variables are $Death_{CEO}$ and its interaction term with the dummy variable. We instrument the interaction term between *Mobility* and the *Short tenure* indicator by interacting the *Death* instrument with the indicator. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	38.591	0.096	-0.104	-0.017	0.046
	(1.64)	(1.08)	(-0.33)	(-0.60)	(0.97)
$Mobility_{CEO}$ × *Short tenure*	-66.121*	-0.279*	0.314	-0.071***	0.120**
	(-1.90)	(-1.90)	(0.84)	(-3.01)	(2.05)
Short tenure	0.152	0.004	0.017	0.002**	-0.003**
	(0.40)	(0.92)	(1.59)	(1.99)	(-2.02)
N	16,386	67,949	67,949	67,949	67,949
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	6.72	43.10	43.10	43.10	43.10
R^2	0.02	0.06	0.11	0.14	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.11. Non-compete Clauses and the Effect of Executive Mobility on Incentives and Corporate Decisions

This table presents second stage 2SLS estimation results for the interactive effect of noncompete enforcement with executive mobility on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment from 1986 through 2011. *Non-compete* indicates the state-level index of non-compete clause enforcement as in Garmaise (2011), where a high score indicates stricter enforcement. The instrumental variables are $Death_{CEO}$ and its interaction term with *Non-compete*. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-38.113***	-0.123	0.288	-0.159***	0.201***
	(-5.72)	(-0.68)	(1.49)	(-3.70)	(4.32)
$Mobility_{CEO} \times$ *Non-compete*	2.093***	0.002	-0.027	0.024**	-0.017**
	(3.17)	(0.05)	(-0.63)	(2.31)	(-2.00)
Non-compete	-0.096	-0.005***	0.007**	-0.001	-0.001
	(-1.55)	(-3.20)	(2.41)	(-0.85)	(-1.13)
N	16,386	67,949	67,949	67,949	67,949
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	6.69	23.96	23.96	23.96	23.96
R^2	0.02	0.06	0.11	0.14	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.12. Executive Mobility Effects for Three Eras: 1950-1985, 1986-2001, and 2002-2011

This table presents second stage 2SLS estimation results for the effect of executive mobility on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment from 1950 through 1985 (Panel A), from 1986 through 2001 (Panel B), and from 2002 through 2011 (Panel C). The instrumental variable is $Death_{CEO}$. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Panel A. 1950-1985

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	89.682	0.421	-8.274	2.994	-1.207
	(1.39)	(0.23)	(-0.88)	(1.44)	(-1.09)
N	2,827	20,663	20,663	20,663	20,663
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	1.19	3.41	3.41	3.41	3.41
R^2	0.02	0.07	0.18	0.08	0.05

*** p<0.01, ** p<0.05, * p<0.1

Panel B. 1986-2001

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-4.549***	-0.109***	0.165***	-0.070**	0.124***
	(-7.78)	(-4.34)	(3.68)	(-2.25)	(5.44)
N	6,620	44,280	44,280	44,280	44,280
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
F-stat	16.34	101.68	122.50	122.56	122.50
R^2	0.01	0.04	0.10	0.18	0.12

Panel C. 2002-2011

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-18.422	0.835	-0.066	1.305**	-0.973*
	(-1.34)	(0.95)	(-0.03)	(2.38)	(-1.65)
N	9,766	22,276	22,276	22,276	22,276
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
F-stat	8.53	5.55	5.55	5.55	5.55
R^2	0.02	0.08	0.14	0.14	0.11

Table 3.13. Accounting for Product Market Competition and Input-Output Linkages

This table examines the robustness of the baseline results to adjusting the measures of mobility for product market competition and input-output relations between firms. When constructing $Mobility_{CEO}$ and $Death_{CEO}$, we exclude CEO moves between firms that have above the median *TNIC-3* product similarity scores (Hoberg-Phillips 2010, 2016) or exclude CEO moves between industries that have above median input-output flows (BEA, Panel B). Pre-1996 period *TNIC-3* product similarity scores are imputed based on 1996 scores. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Panel A. Mobility excluding CEO external hires between above median *TNIC-3* firms

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-29.052**	-0.101***	0.195***	-0.126**	0.139***
	(-2.00)	(-4.60)	(3.76)	(-1.96)	(3.86)
N	15,482	67,548	67,548	67,548	67,548
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	11.95	94.65	94.65	94.65	94.65
R^2	0.01	0.06	0.10	0.14	0.14

Panel B. Mobility excluding CEO external hires between above median input-output industries

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-12.014***	-0.096**	0.234**	-0.091***	0.155***
	(-4.35)	(-2.34)	(2.37)	(-4.17)	(2.99)
N	14,855	66,929	66,929	66,929	66,929
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	8.25	51.08	51.08	51.08	51.08
R^2	0.01	0.05	0.10	0.09	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.14. The Effect of the Executive Mobility on Firm Value

This table presents second stage 2SLS estimation results for the effect of executive mobility on firm value from 1986 through 2011. *t*, *t+1*, *t+2* and *t+3* indicate Tobin's *q* in years *t*, *t+1*, *t+2* and *t+3* respectively. The instrumental variable is $Death_{CEO}$. Control variables are the same as those used in Table 3.6, excluding M/B. Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Panel A. Full Sample

Dependent Variable: Tobin's q	*t*	*t+1*	*t+2*
	(1)	(2)	(3)
$Mobility_{CEO}$	0.548	0.908*	0.674***
	(1.08)	(1.72)	(3.04)
N	67,949	63,403	57,224
Controls	YES	YES	YES
Year Fixed Effect	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES
F-stat	85.66	91.07	98.72
R^2	0.12	0.11	0.11

Panel B. Firms with Minimum 3-year Observations in Sample

Dependent Variable: Tobin's q	*t*	*t+1*	*t+2*
	(1)	(2)	(3)
$Mobility_{CEO}$	0.961*	1.009*	0.677***
	(1.65)	(1.72)	(3.31)
N	50,758	50,758	50,758
Controls	YES	YES	YES
Year Fixed Effect	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES
F-stat	95.76	95.76	95.76
R^2	0.11	0.11	0.11

*** p<0.01, ** p<0.05, * p<0.1

Appendix III. Figures and Tables

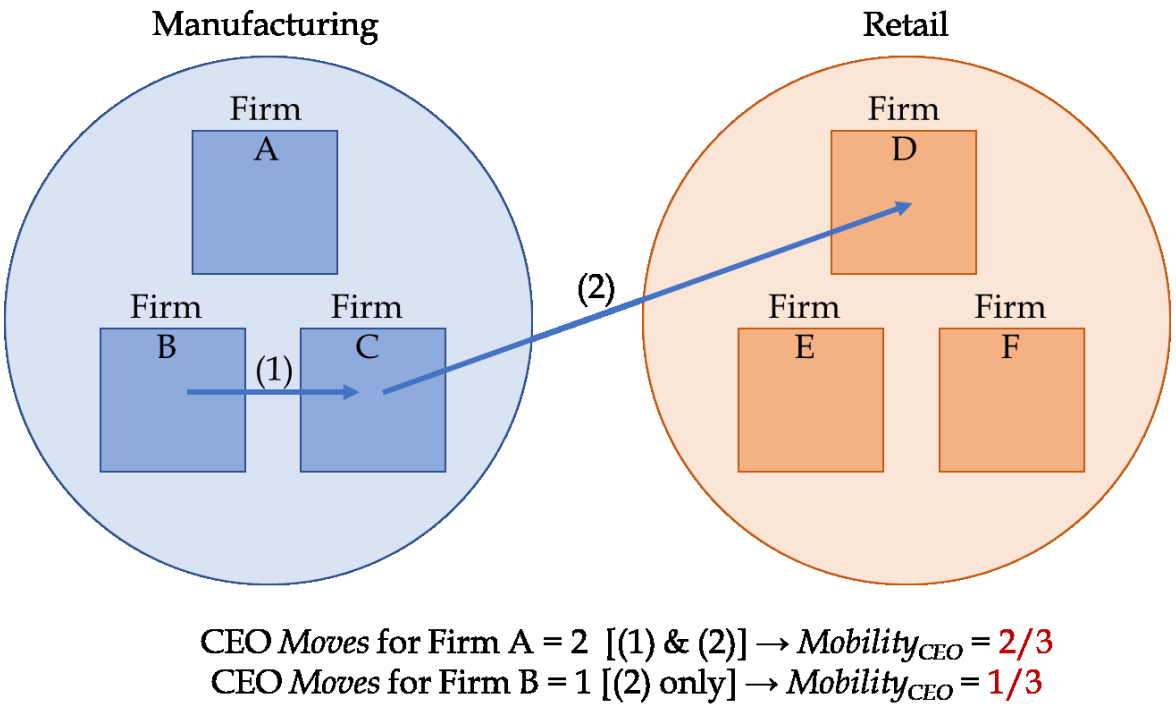
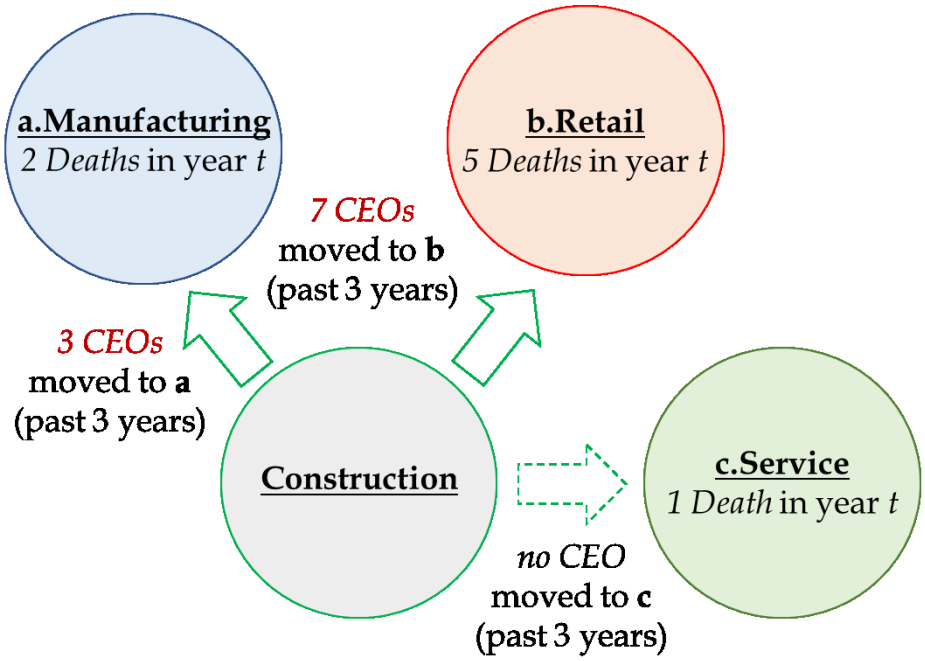


Figure 3.A1. *CEO Moves* and *Mobility*

This figure describes how we define *CEO Moves* and *Mobility* in equation [1]. *CEO Moves* is the number of within- and across-industry CEO moves for a given industry between year *t-1* and *t*, excluding a given firm's own CEO turnover. *Mobility* in equation [1] is the number of CEO moves divided by the number of CEOs at the industry level in year *t-1*. A second measure of mobility deflates by number of CEO turnovers at the industry level in year *t-1*.



CEO deaths in connected industries for Construction in year t
$= 0.3 \times 2 + 0.7 \times 5 = 4.1$

Figure 3.A2. CEO Death in Connected Industries

This figure describes how we define *Death* in equation [2]. Connectedness between two industries is defined using CEO moves between the industries in the past three years. *Death* of an industry is a weighted average number of CEO deaths in *t-1* where the weight of a given industry is determined by the industry pair's connectedness divided by the number of CEOs (or separately, by the number of CEO turnovers) in the industry in *t-1*.

Table 3.A1. Variable Definitions

Item	Method	Source
*Firm & Industry Characteristics**		
Leverage	(DLC + DLTT) / AT	Compustat
Net Leverage	(DLC + DLTT - CHE) / AT	Compustat
Cash flow	(IB + DP) / lagged PPENT	Compustat
Size	logged AT converted to 2011-dollar value	Compustat
ROA	NI / AT	Compustat
PPE/TA	PPENT / AT	Compustat
M/B	(PRCC_F * CSHO + DLC + DLTT) / CEQ	Compustat
Investment	CAPX / AT	Compustat
CASH/TA	CHE / AT	Compustat
Tobin's *q*	Q = (PRCC_F * CSHO DLC + DLTT) / AT	Compustat
Industry Tobin's *q*	One-digit SIC average Q	Compustat
Industry GDP Growth	One-digit SIC GDP growth rate	BEA
Board and CEO characteristics		
CEO Tenure	The number of years a CEO has been CEO at her current firm	Moody's Industrial Manuals/ Compact Disclosure/ Mergent/Board Analyst
Total CEO Tenure	Total number of years served as CEO at a given firm	Moody's Industrial Manuals/ Compact Disclosure/ Mergent/Board Analyst
CEO Turnover	1 if a CEO is replaced in a given year, 0 otherwise	Moody's Industrial Manuals/ Compact Disclosure/ Mergent/Board Analyst
Independence	The number of independent directors divided by the total number of directors	Moody's Industrial Manuals/ Compact Disclosure/ Mergent/Board Analyst
CEO-Chair	1 if a CEO is currently chair of the board, 0 otherwise	Moody's Industrial Manuals/ Compact Disclosure/ Mergent/Board Analyst
Pay-Perf †	$[\{(Salary_{t+1} + Bonus_{t+1}) - (Salary_t + Bonus_t)\} / (Salary_t + Bonus_t)] / \{(PRCC\_F_t - PRCC\_F_{t-1}) / PRCC\_F_{t-1})\}$	Execucomp/ Frydman and Saks (2010)
Non-compete	10 for maximum non-compete clause enforcement, 0 for minimum enforcement	Garmaise (2011)‡

†Winsorized at the 5% and 95% levels, respectively
*Ratio Variables are winsorized at the 1% and 99% levels
‡We collect data on state-level legal reforms to extend and impute the scores in Garmaise (2011) for the 1986-2011 period.

Table 3.A2. Alternative Measures of Executive Mobility

This table presents second stage 2SLS estimation results for the effect of executive mobility on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment using alternative measure from 1986 through 2011. *Executive Mobility (All)* is the number of executive officers (not just CEOs) in a one-digit SIC industry who move to another firm divided by the lagged number of firms in a one-digit SIC industry. The instrumental variable is $Death_{CEO}$. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
Executive Mobility (All)	-8.236	-0.113***	0.222***	-0.078***	0.151***
	(-0.44)	(-9.76)	(6.77)	(-6.67)	(3.81)
N	16,386	67,949	67,949	67,949	67,949
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	23.85	339.58	339.58	339.58	339.58
R^2	0.02	0.06	0.11	0.14	0.14

*** p<0.01, ** p<0.05,
* p<0.1

Table 3.A3. Alternative Instrument: CEO Sudden Death and Health-Related Turnovers

This table reports second stage 2SLS estimation results for the effect of executive mobility on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment using alternative instruments from 1986 through 2011. The instrumental variable is $Death_{CEO}$, which only includes CEO deaths identified as "sudden" (Panel A) or includes all CEO deaths plus CEO turnovers due to health-related reasons (Panel B). Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (3). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Panel A. CEO Sudden Deaths

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-27.424**	-0.111***	0.119**	-0.120*	0.139***
	(-2.17)	(-3.49)	(2.06)	(-1.92)	(4.03)
N	16,386	67,949	67,949	67,949	67,949
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	20.97	85.68	85.68	85.68	85.68
R^2	0.02	0.06	0.11	0.14	0.14

Panel B. CEO Death and Health-related CEO Turnovers

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-24.610**	-0.139***	0.241***	-0.092*	0.133***
	(-1.98)	(-5.42)	(3.17)	(-1.88)	(3.44)
N	16,386	67,949	67,949	67,949	67,949
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	8.04	61.12	61.12	61.12	61.12
R^2	0.02	0.06	0.11	0.14	0.14

*** p<0.01, ** p<0.05, * p<0.1

Table 3.A4. Alternative Instrument: CEO Retirement Age

This table reports second stage 2SLS estimation results for the effect of executive mobility on CEO pay-for-performance sensitivity, corporate monitoring, capital structure, and corporate investment using an alternative instrument based on CEO retirement age from 1986 through 2011. The instrumental variable is the weighted number of CEOs above 63 of age† (Jenter and Kanaan 2015) in connected industries at *t-1* divided by *N. of* $CEOs_{t-1}$. Both weight and connectedness are determined by the external CEO hires over the past three years between one-digit SIC industries. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-6.967***	-0.166	0.160	-0.208*	0.312*
	(-6.94)	(-0.90)	(0.41)	(-1.82)	(1.74)
N	12,360	60,757	60,757	60,757	60,757
Controls	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
F-stat	25.27	21.54	21.54	21.54	21.54
R^2	0.01	0.05	0.08	0.09	0.13

†To calculate the proportion of CEOs above 63 of age, we use Forbes data. There are 27,775 non-missing age information in the data.

*** p<0.01, ** p<0.05, * p<0.1

Table 3.A5. Frequency of CEO Moves between Industries: Pre and Post 2002

This table presents the frequency of CEOs moving from a firm in the origination industry to another firm in the destination industry by time period from 1986 through 2001 (Pre-2002) and from 2002 through 2011 (Post-2002). *SIC-From* indicates the one-digit SIC industry of the firm that a given CEO departs and *SIC-To* indicates the one-digit SIC industry of the firm that the CEO moves as CEO or other officers. The mean and Herfindahl-Hirschman Index of the frequency are in the parentheses.

Panel A. Pre-2002: 1,222 moves (Mean 1.37%, HHI 6.37%)

SIC1-From \ SIC1-To	SIC0 Agric., Forestry, Fishing	SIC1 Mining & Constr.	SIC2 Light Manuf.	SIC3 Heavy Manuf.	SIC4 Transport. & Public Utilities	SIC5 Wholesale & Retail Trade	SIC6 Finance, Insurance, Real state	SIC7 Services	SIC8 Health Services	SIC9 Public Admin.	Total
0.Agriculture, Forestry, Fishing		0.08%		0.16%		0.08%			0.08%		0.4%
1.Mining & Construction		4.99%	0.74%	1.06%	0.74%	0.33%	0.08%	0.49%	0.25%		8.7%
2.Light Manufacturing	0.16%	0.33%	8.27%	2.70%	0.49%	1.55%	0.08%	0.49%	0.74%		14.8%
3.Heavy Manufacturing		0.49%	3.11%	17.18%	1.64%	1.55%	0.41%	3.11%	0.74%		28.2%
4.Transportation & Public Utilities		0.16%	0.25%	1.15%	5.40%	0.08%	0.25%	1.55%	0.25%		9.1%
5.Wholesale & Retail Trade	0.08%	0.25%	0.98%	1.88%	0.25%	6.71%	0.41%	1.15%	0.82%		12.5%
6.Finance, Insurance, Real Estate		0.16%	0.57%	0.98%	0.33%	0.57%		0.82%	0.08%		3.5%
7.Services	0.08%	0.16%	0.57%	2.70%	0.41%	0.90%	0.65%	10.23%	0.82%		16.5%
8.Health Services	0.16%	0.25%	0.90%	0.65%	0.16%	0.49%	0.08%	1.06%	2.29%		6.1%
9.Public Administration				0.08%		0.08%					0.2%
Total	0.5%	6.9%	15.4%	28.6%	9.4%	12.4%	2.0%	18.9%	6.1%	0.0%	100%

Panel A. Post-2002: 433 moves (Mean 1.85%, HHI 6.35%)

SIC1-From \ SIC1-To	SIC0 Agric., Forestry, Fishing	SIC1 Mining & Constr.	SIC2 Light Manuf.	SIC3 Heavy Manuf.	SIC4 Transport. & Public Utilities	SIC5 Wholesale & Retail Trade	SIC6 Finance, Insurance, Real state	SIC7 Services	SIC8 Health Services	SIC9 Public Admin.	Total
0.Agriculture, Forestry, Fishing	0.23%										0.2%
1.Mining & Construction		4.85%	0.69%	0.92%		0.46%	0.23%				7.2%
2.Light Manufacturing			10.39%	3.23%	0.69%	1.15%	0.92%	0.23%	1.85%	0.46%	18.9%
3.Heavy Manufacturing		0.23%	2.77%	15.70%	2.31%	1.39%	0.46%	2.77%	1.15%	0.23%	27.0%
4.Transportation & Public Utilities		0.69%	1.15%	0.92%	7.85%	0.46%		0.46%			11.5%
5.Wholesale & Retail Trade		0.46%	1.62%	1.39%	1.15%	4.16%		0.46%	0.23%		9.5%
6.Finance, Insurance, Real Estate							1.85%	0.46%	0.23%		2.5%
7.Services			0.69%	2.08%	0.69%	0.69%		10.16%	1.85%		16.2%
8.Health Services			1.39%	0.92%	0.46%	0.46%	0.23%	0.46%	2.54%	0.23%	6.7%
9.Public Administration					0.23%						0.2%
Total	0.2%	6.2%	18.7%	25.2%	13.4%	8.8%	3.7%	15.0%	7.9%	0.9%	100%

indicates top 10%, top 20%, top 30% and top 50% industry pairs in terms of frequency of moves

*** p<0.01, ** p<0.05, * p<0.1

Table 3.A6. Executive Mobility from 1986 through 2011 – OLS Estimates

This table presents baseline panel estimation results using ordinary least squares for the effect of executive mobility on CEO pay-for-performance sensitivity, board monitoring, capital structure and corporate investment from 1986 through 2011. Control variables are the same as those used in Table 3.6, excluding Cash/TA and Leverage in (4). Numbers in parentheses are *t*-statistics based on robust standard errors clustered at the industry and year levels.

Dependent Variables:	*Pay-Perf*	*Independence*	*CEO-Chair*	*Net Leverage*	*Investment*
	(1)	(2)	(3)	(4)	(5)
$Mobility_{CEO}$	-6.095**	-0.087	-0.072	-0.062	0.102***
	(-3.40)	(-0.99)	(-0.71)	(-1.06)	(6.66)
N	16,386	67,949	67,949	67,949	67,949
Year Fixed Effect	YES	YES	YES	YES	YES
CEO-Firm Fixed Effect	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Adjusted R^2	0.026	0.535	0.717	0.792	0.631