Frictions in Financial Intermediation

by

Sahil Raina

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2016

Doctoral Committee:

Associate Professor Sugato Bhattacharyya, Chair
Professor Michael Barr
Associate Professor Norman Bishara
Professor Charles Brown
Assistant Professor Ing-Haw Cheng
Professor Amy Dittmar
Assistant Professor Martin Schmalz

ACKNOWLEDGEMENTS

First and foremost, I want to thank Sugato Bhattacharyya, my advisor and committee chair. Sugato's patient and unwavering support, in addition to his prescient advice, were invaluable for me as I was working on this thesis.

I would also like to thank the rest of my dissertation committee: Michael Barr, Norman Bishara, Charles Brown, Amy Dittmar, Ing-Haw Cheng, and Martin Schmalz. The non-finance members of my committee (Michael, Norm, and Charlie) helped push me to think about my research from different perspectives. The finance members (Amy, Ing, and Martin) helped improve the dissertation by providing insightful advice and suggestions for the eventual publication of my work.

Finally, I would like to thank my family members. My parents and brother patiently supported me through my meandering path through higher education that took me from architecture to economics to finance. My son provided endless entertainment when I needed some time away from thinking about research. But most importantly, I have to thank my wife, without whom, this dissertation simply would not have been possible. She provided support, advice, kind words, and objective critiques exactly when I needed them most.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

<u>Table</u>

LIST OF APPENDICES

Appendix

ABSTRACT

Frictions in Financial Intermediation

by

Sahil Raina

Chair: Sugato Bhattacharyya

This dissertation comprises essays examining frictions in three types of financial intermediaries: venture capital firms, investment banks, and commercial banks. The first two essays consider the decision-making of agents within these intermediaries while the third examines the structure of the commercial bank sector. The first essay explores the effect of gender on VC-financed entrepreneurship. I find that women's participation in venture capital-financed entrepreneurship is lower than in other sectors of the economy. Further, the women that do participate lead startups that perform worse than startups led by men. Does interaction with venture capitalists (VCs) contribute to the low participation and performance gap? To answer these questions, I compare the gender gap in successful exits from VC financing between two sets of startups: those initially financed by VCs with only male general partners (GPs) and those initially financed by VCs that include female GPs. Constructing a novel dataset to perform this analysis, I find a large performance gender gap among startups financed by VCs with only male GPs but no such gap among startups financed by VCs that include female GPs. The disparity is solely due to improved performance among female-led startups. This suggests that VC gender composition

has contributed strongly to the performance gap between female- and male-led startups, which could deter women from leading VC-financed projects and lower their participation.

The second essay analyzes whether mid-level managers in securitized finance were aware of a large-scale housing bubble and a looming crisis in 2004-2006 using their personal home transaction data. We find that the average person in our sample neither timed the market nor were cautious in their home transactions, and did not exhibit awareness of problems in overall housing markets. Certain groups of securitization agents were particularly aggressive in increasing their exposure to housing during this period, suggesting the need to expand the incentives-based view of the crisis to incorporate a role for beliefs.

The third essay uses a uniquely extensive branch-level dataset on deposit account interest rates, maintenance fees, and fee thresholds, and document substantial time-series and cross-sectional variation in these prices. We then examine whether variation in bank concentration helps explain the variation in prices. The standard measure of concentration, the HHI, is not correlated with any of the outcome variables. We then construct a generalized HHI (GHHI) that captures both common ownership (the degree to which banks are commonly owned by the same investors) and cross-ownership (the extent to which banks own shares in each other). The GHHI is strongly correlated with all prices. We use the growth of index funds as an arguably exogenous source of cross-sectional variation of county-level common ownership to suggest a causal link from the GHHI to higher prices for banking products.

CHAPTER I

VC financing and the entrepreneurship gender gap

Abstract

Women's participation in venture capital-financed entrepreneurship is lower than in other sectors of the economy. And the women that do participate lead startups that perform worse than startups led by men. Does interaction with venture capitalists (VCs) contribute to the low participation and performance gap? To answer these questions, I compare the gender gap in successful exits from VC financing between two sets of startups: those initially financed by VCs with only male general partners (GPs) and those initially financed by VCs that include female GPs. Constructing a novel dataset to perform this analysis, I find a large performance gender gap among startups financed by VCs with only male GPs but no such gap among startups financed by VCs that include female GPs. The disparity is solely due to improved performance among female-led startups. This suggests that VC gender composition has contributed strongly to the performance gap between female- and male-led startups, which could deter women from leading VC-financed projects and lower their participation.

1.1 Introduction

Anecdotally, Silicon Valley is a difficult environment for female entrepreneurs. In an article published in *The New York Times* in April 2014, the author notes that "sexism exists in many places, but start-up companies have particular qualities that can allow problems to go unchecked." A January 2015 *Newsweek* article describes the venture capital (VC) industry in northern California as a "boys' club" and implies that the industry's actions create "a particularly toxic atmosphere for women in Silicon Valley." Does this translate into worse performance for VC-financed startup led by female entrepreneurs? In this paper, I measure performance using exits from VC financing via IPO or acquisition[1] and find that startups led by one or more female entrepreneurs (hereafter referred to as female-led startups) have a 37% lower rate of exit than startups led by male entrepreneurs (male-led startups), a sizeable performance gap.

What might be the underlying reasons for this gap? Is it that female-led startups are intrinsically less valuable than male-led startups?[2] Or are VC financiers responsible for the performance gap? It could be that some VC financiers are poor at evaluating female-led startups and some may also be poor at advising them. In this paper, I explore whether VC financing contributes to the performance gap among startups. Addressing this question is important both from the perspective of the VC financier and the startup. From the VC perspective, reducing financed startups' potential success means that some VCs are wasting resources invested by their limited partners (LPs). For the startup, financing by the right VC may be the difference

[1]Successful exit from VC financing via IPO or acquisition is a standard measure of performance in the VC literature (for example, Hochberg et al., 2007; Cockburn and MacGarvie, 2009; Puri and Zarutskie, 2012).

[2]Given the nature of VC-financed projects, gender differences in risk aversion, competitiveness, and ability that are described in the literature may contribute to this gap. Risk aversion differences between men and women are documented in Powell and Ansic (1997) as well as Barber and Odean (2001). Croson and Gneezy (2009) shows evidence that women are more averse to competition than men. And Tierney (2010) presents the argument made by Larry Summers that the far right tail of the ability distribution may be more populated by men than women.

between success and failure. If women entrepreneurs believe VC financing may hurt their startups' likelihood of success, they may be less likely to pursue VC-financed entrepreneurial projects.

Reduced female participation would imply that some intrinsically valuable projects in the economy are not undertaken because of the possibility of VC-induced failure. This may be part of the reason for markedly lower female participation in VC-financed entrepreneurship than in other segments of the economy. In 2012, women comprised 47% of the labor force while 36% of small businesses were majority-owned by women (Sewell, 2013; Lichtenstein, 2014). In stark contrast, 15% of VC-financed firms had a woman on the executive team in 2011-2013 and only 2.7% of them had a female CEO, according to a survey by The Diana Project.[3]

To explore the potential role of gender in VC-financed entrepreneurship, I compare the performance gap among startups financed by VC syndicates with and without female partners (hereafter, female VCs and male VCs, respectively), and find that, although the two VC groups finance similar startups, the performance gap is large among startups financed by male VCs but nonexistent among startups financed by female VCs. My findings provide strong initial evidence that VC financing has contributed to the performance gap among startups.

To run my analyses, I construct a novel dataset using CrunchBase, a large, crowd-sourced database on the activities of high-tech startups. I use CrunchBase not only because it includes data on a large number of startups and financing rounds, but also because it includes biographical information on entrepreneurs and financing VCs' general partners (GPs), which is crucial for this study and not available in most public databases on VC financing. While CrunchBase limits the dataset to high-tech startups, the percentage of female-led startups within this dataset is comparable to that reported for all VC-financed firms (14.8% in my sample versus 15% in The Diana

[3]The Diana Project is a non-profit organization focused on female entrepreneurship which releases periodic statistics on female participation in entrepreneurship.

Project data).

My dataset provides a number of interesting insights into VC financing of startups. For instance, 9% of all founders and GPs are female in my dataset, and one-sixth of the startups initially financed by female and male VCs are female-led. The equal representation suggests that the two VC groups do not suffer differentially from any possible evaluation biases against female entrepreneurs. Of course, equal representation could also arise from differences in the intrinsic value of female-led startups seeking initial financing from two (differently biased) VC groups. However, this explanation is inconsistent with another fact revealed by the data: the proportion of portfolio firms that successfully exit VC financing from the two VC groups is the same, approximately 29%.[4] If the intrinsic value of financed startups differed across the two VC groups, it would likely be reflected in their overall exit rates.

Comparing other features of the data across the two syndicate groups, I find that startups that were initially financed by the two VC groups also have similar exit rates via IPO (5%) and acquisition (24%). The duration of VC financing for successful portfolio firms also does not differ substantially across the two groups (3.5 years), nor the number of financing rounds (2.4 rounds). The number of entrepreneurs in each portfolio firm is similar across the two groups as well (2.0 entrepreneurs). As there is a general dearth of female GPs, syndicates with female GPs tend to be larger, have more GPs per syndicate, have more VC firms per syndicate, and have more prior financing experience across all firms in the syndicate.

To explore whether VC financing plays a role in the performance gap between female- and male-led startups, I conduct a logistic regression analysis of startup performance on female presence as a founder and in the financing VCs. In particular, I compare differences in exit rates between female- and male-led startups initially

[4]This exit rate is higher than the exit rates reported earlier for both female- and male-led startups (17% and 27%, respectively) because the two sets of exit rates are calculated using different subsamples based on data availability.

financed by female VCs and male VCs. If inherent differences between female- and male-led startups explain the performance gap, then the gap in exit rates should be the same across startups financed by the two VC groups. If the gap differs, then VC financing has an impact on the performance gap. I find a 30 percentage point gender gap in exit rates among startups initially financed by male VCs, whereas this gap disappears among startups financed by female VCs. This difference arises from female-led startups' exit rates being significantly higher when financed by female VCs. In contrast, male-led startups' exit rates are the same regardless of the gender composition of the initial financing syndicate.

Could the observed difference in the performance gap be driven by differences in entrepreneurial preference for female versus male VCs? For instance, could the gap be explained by female-led startups of high intrinsic value preferring financing from female VCs? If this were true, female VCs would have greater proportions of female-led startups in their portfolios. However, the representation of female-led startups in the portfolios of the two VC groups is the same.[5] Alternatively, could the difference in the performance gap arise due to female-led startups of all types preferentially seeking financing from female VCs? This conjecture is not consistent with the observed difference across the VC groups in the performance gap. Finally, could it be that all high intrinsic value startups preferentially seek financing from female VCs? In such a case, the overall exit rate would be higher for all startups financed by female VCs. However, as mentioned earlier, the overall exit rate of portfolio startups in the two VC groups is approximately the same. Given my empirical setting, I cannot entirely rule out the possibility that entrepreneur financing choice drives the observed difference in the gap. However, by eliminating these most commonly-posed scenarios, I alleviate much of the concern revolving around this alternative explanation of my findings.

[5] My argument against this entrepreneur choice explanation depends on the assumption that VCs are evaluating startups in terms of expected future performance. If VCs provide financing to all startups or evaluate startups on some other characteristics, my argument does not rule out this explanation.

The findings above jointly paint a picture in which VC financing contributes to the performance gap among startups based on founder gender. But *how* do VCs contribute to the gap? Is it due to poor evaluation or poor advising? My extensive data allow me to compare initial financing rounds to subsequent rounds to shed some light on this question. Comparing the difference in the gap across the two VC groups in initial rounds versus later rounds shows that it is much larger in initial rounds, suggesting that female VCs are better at evaluating female-led startups, which helps drive VC contribution to the performance gap.

Can the observed difference in female-led startups' exit rates be attributed to matching female entrepreneurs and GPs or does it arise from some cultural characteristic of VC syndicates that have female GPs? I compare the gap in exit rates between female- and male-led startups initially financed by a single VC to those financed by multiple VCs. Among single-VC financing rounds, each GP is more likely to be directly involved with the financed startup. I find a larger difference in the gap for single VC financing rounds, which suggests that female GP involvement has a direct impact on the performance gap.

This paper has two principal findings. First, it establishes the existence of a gender gap in performance among VC-financed startups. Second, it presents persuasive evidence that the structure of the VC financing industry has contributed to this performance gap. This latter result has three important implications. First, it suggests that some intrinsically valuable firms do not succeed despite getting access to VC financing. Second, VC-induced reduction in success rates means that some VCs waste LP-invested resources. Finally, if women are thus inefficiently dissuaded from entrepreneurship, it implies that some intrinsically valuable projects are not undertaken because of the possibility of VC-induced failure.

1.2 Related literature

This paper adds to a growing body of literature on the interaction between financing and gender among entrepreneurial firms. Alesina et al. (2013) finds that female entrepreneurs seeking bank loans pay more for credit than do male entrepreneurs. Bellucci et al. (2010) finds that female entrepreneurs face tighter credit availability than male entrepreneurs when seeking bank loans. Bellucci et al. (2010) also finds that female loan officers require lower collateral from female entrepreneurs for loans than from male entrepreneurs. These papers look at the impact of entrepreneur gender on specific financing outcomes, whereas I examine impact on firm performance. In the context of crowdfunding, Marom et al. (2015) discovers a preference among female investors for female-led projects. Bengtsson and Hsu (2010) finds a preference for shared identity along ethnicity and educational background in VC financing pairings between entrepreneurs and GPs. My paper adds to these papers by examining the impact of such pairings across entrepreneurs and financiers on firm performance. Like my paper, Gompers et al. (2014) looks at performance impacts of gender, but within VC firms. It finds that, while female GPs' investments perform worse than male GPs' investments, this difference goes away if the VC firm has multiple female GPs.

This paper also adds to the literature on entrepreneur and VC characteristics that affect entrepreneurial firm performance. Hochberg et al. (2007) shows that greater VC firm connectedness is associated with better exit outcomes for financed entrepreneurial firms. Lerner (1994) presents evidence that VC firms' experience helps them better time the exit of financed firms via IPO. Gompers et al. (2010) documents that previous entrepreneur success also predicts entrepreneurial firm success. There is also a large subliterature interested in whether the project or the management team is more important for entrepreneurial firm success (see Kaplan et al., 2009; Gompers and Lerner, 2001; Gladstone and Gladstone, 2002). Another branch of this literature considers the role of VC firms' bargaining power in fund performance (see Hsu, 2004; Kaplan and Schoar, 2005; Hochberg et al., 2010). This paper offers evidence that matching between VC firms and entrepreneurs also impacts the performance of entrepreneurial firms.

Outside of entrepreneurial finance, this paper also relates to a wider literature examining the role of gender pairings on various outcomes. Within finance, Huang and Kisgen (2013) provides evidence that male executives exhibit overconfidence in corporate decision-making relative to female executives, which suggests that the impact of female GPs may come from actions of the female GP herself. Ahern and Dittmar (2012) finds that constraints on the gender composition of corporate boards has an impact on firm value. In a labor setting, Tate and Yang (2014) shows that female workers lose more in wages than male workers when they lose a job but that this difference is narrower if the workers are rehired by a firm with female leadership. In management, Tsui et al. (1989) finds that superior-subordinate dissimilarity is associated with lower effectiveness in corporate settings. In education, Lim and Meer (2015) and Paredes (2014) show that female students paired with female teachers perform better in testing whereas male students do not exhibit any change in performance due to teacher gender. My findings suggest that similar effects of gender pairings may exist in VC financing as well.

This paper draws some techniques and insights from the economics literature on discrimination. In labor economics, there is a great deal of research on discrimination based on gender, ethnic, and racial identities. Goldin and Rouse (2000), for instance, provides evidence of discrimination against females in symphony orchestra auditions. Bertrand and Mullainathan (2004) presents evidence of discrimination by race in employment interview callbacks. While such discrimination is not the principal focus of my study, the underlying frameworks of discrimination pioneered by Becker (1971) and Arrow (1973) help motivate some of the empirical analyses in this paper as well.

1.3 Empirical setting

Because most publicly-available databases on VC financing lack biographical information, I construct a novel dataset that includes information for both the entrepreneurs leading startups and the financiers financing them. In this section, I (briefly) discuss the structure of the VC financing industry, present basic statistics detailing my dataset and outline the new sources I use for it. For information on how I constructed the dataset, see Appendix A.

1.3.1 VC financing process

VC financing is a private form of financing for startups whose businesses exclude financing via debt. VCs form a bridge between three parties: startups, early investors, and later investors. They evaluate potential startups and advise the startups they choose to finance. They interact with large investors (limited parters or LPs) who provide the bulk of the capital for the early-stage financing of the startups. These investors tend to be institutions such as pension funds but can also be wealthy individuals. Finally, VCs also manage the exits from VC financing of successful startups. In this role, they deal with the public equity markets and potential acquirers who provide subsequent financing for the now-matured, successful startups.

The two-sided matching between VCs and startups is highly informal.[6] As this paper focuses on the interaction between VCs and startups, it is important to understand this fact. First, information about startups seeking financing can come from a number of sources: GPs' personal connections, the VC's network of lawyers, investment bankers, accountants, et cetera, and, sometimes, even through the formal channels provided by the VC. Once contact with the startup is made, analysts at the VC study the startup and provide recommendations to the VC's leadership. The GPs then jointly decide on whether to finance the startup. While this is not always

[6]This insight arises from discussions I had with VCs about how they source their portfolios.

the case, the decision to finance a startup often needs to be unanimous. Additionally, while the analysts provide quantitative analysis of the startups, there is no "cutoff" above which a startup is certain to receive financing or below which is it certain to be rejected.

The startups that secure VC financing are provided with capital by its financiers in a series of rounds. At each round, the financiers reassess the performance of the startup.[7] The periodic reassessment of startups is one characteristic of VC financing that helps mitigate some of the problems associated with financing high uncertainty, early-stage businesses (Gompers and Lerner, 2004).

1.3.2 Data description

My dataset contains information on 3,660 entrepreneurial firms (startups). To the extent possible, it includes data on each startup's financing rounds, founders, and whether and how the startup eventually exits VC financing. For the financing rounds, of which there are 10,015, I know the date on which the financing round was announced, which VCs were involved in the round, and the GPs of the involved VCs. For founders and GPs, the dataset includes full name and gender information. And, for exits, I know the type of exit (IPO or acquisition) and the date of exit announcement.

All of the startups I observe are in the high-tech sector, with the vast majority in the computing high-tech sector. As can be observed in Figure 1.1, 89% of my startups are computing high-tech firms, 9% are biotech firms, and the rest are manufacturing high-tech firms. This is reflective of the VC-financed sector as a whole. VCs, with their equity-like contracts and intensive monitoring and advising of startups, are better equipped than other forms of private financing for high tech, high information

[7]This does not imply that VCs do not monitoring and advising startups between disbursements. As Gorman and Sahlman (1989) shows, VCs spend a significant amount of time monitoring and advising their investments between financing rounds.

asymmetry sectors.

While my data include 10,015 financing rounds, many of them lack the information required for my analyses.[8] For instance, in Table 1.1, we see that 70% of financing rounds (65% of initial rounds) have gender data for founders, 57% (61%) have gender data for GPs involved in the financing, and 39% (37%) have gender data for both founders and GPs. While less than 40% of the data are usable in some of the analyses I run, the overall sample sizes are still large, with well over a thousand initial financing rounds with gender data for both founders and financing GPs.

While I have data for financing rounds as far back as 1995, I limit my analyses to startups with initial financing rounds in 2005 or later. I do this because, although it has data for rounds prior to 2005, CrunchBase was established in 2005. As a result, the startups with financings before 2005 reported in CrunchBase are markedly different from the rest of the startups in the data. In particular, they are much more likely to be successful in exiting VC financing. This brings up concerns of backfill bias for the pre-2005 financing startups. To avoid problems associated with this bias, I exclude these startups from my analyses. This does not limit my sample severely as there are far fewer firms in the years before 2005 than following it, as can be seen in Figure 1.2.

Figure 1.2 also shows that the number of financing rounds reported in CrunchBase generally increases over time. This is reflective of the nature of VC financing. With more firms being added through initial financings each year, more startups are likely to have subsequent financings in each year than in the previous one. Unlike all rounds, we see that the number of initial financing rounds reported each year after 2005 stays between 250 and 500, suggesting a steady flow of information to the database. This suggests that, at some point, the number of financing rounds per year should level off, unless duration of financing is increasing or CrunchBase starts increasing its coverage

[8] As I discuss in Section 1.3.3.1 below, this is one of the shortcomings of using crowdsourced data.

further.

In Figure 1.2, we also see a dip in initial financings between 2006 and 2011, which coincides neatly with the economic downturn. There is no similar dip in all financing rounds, but the rate at which all rounds increase falls for that period as well. These features suggest that CrunchBase has a good representation of the overall startup economy in each year after its inception.

1.3.3 Sources

I combine data from multiple sources to generate my dataset. These sources include TechCrunch's CrunchBase, Namepedia, SEC's EDGAR data, Thomson One SDC's M&A database, and Thomson One's VentureXpert. I use VentureXpert primarily to define my sample, as I explain in Appendix **??**. SEC's EDGAR and Thomson One SDC's M&A databases are used to identify VC financing exit events (IPOs and acquisitions) for the startups in my data. In this section, I describe CrunchBase and Namepedia, as these two sources are less familiar to general finance and economics audiences.

1.3.3.1 CrunchBase

The CrunchBase database provides data on high-tech startup activity. In fact, they claim to be the "most comprehensive information source" for such activity (CrunchBase, 2014). A key feature of the database is that it is crowdsourced. This affords CrunchBase three substantial benefits. The greatest benefit is its extensive coverage of VC financing of startups. In my sample of CrunchBase startups financed at least once by top VCs, I have information on 3,660 entrepreneurial firms and 10,015 financing rounds from 3,318 investors. For comparison, the Burgiss database has data on 775 VC funds, which means their data are, at most, based on 775 VCs' data (Harris et al., 2014). Similarly, the Venture Economics database has data on 1,114 VC funds.

Crowdsourcing also limits concerns of bias arising from a limited number of contributors. Most VC databases arise from data provided by a few or even one source. In 2013 alone, over 53,000 sources contributed to CrunchBase (Kaufman, 2013). Having a wider base of contributors reduces the likelihood of a bias tied to single perspective or few perspectives.

Crowdsourcing also mitigates issues tied to voluntary disclosure. Most of the data we have on the industry come from voluntarily disclosed information. These data are more likely to be biased in a manner that favors the data provider than data coming from involuntary disclosure. For instance, in Kaplan and Strömberg (2003), the authors point out that their sample of 119 portfolio companies may be "biased towards more successful investments," given that they find a 25% IPO rate. While this bias does not impact their findings, it highlights the potential issues with voluntary disclosure. CrunchBase data, while voluntarily provided, are not sourced solely from VCs, LPs, or portfolio firms. Rather, the CrunchBase data are sourced from the general public. This sharply mitigates concerns about biases stemming from voluntary disclosure.

Being crowdsourced is also responsible for CrunchBase's primary weakness: a plurality of observations with incomplete information. For instance, I have founder gender data for 65% of initial financing rounds and GP gender data for 61% of initial financing rounds. This incompleteness of my dataset arises primarily because, for a significant number of startups and VCs, personnel information is not available on CrunchBase.[9]

CrunchBase has a number of mechanisms in place to ensure data quality: news article citation for any database alteration, authentication of a data provider's identity, and algorithmic and manual verification of all database changes (CrunchBase,

[9]A much smaller issue that contributes to the high missings rate is that I am unable to match all personnel to their genders.

2014).[10]

There are three ways to access the CrunchBase data: full subscription, monthly tabulation, and API access. I use the API (Application Program Interface) access provided by CrunchBase. I detail this procedure in Appendix A, where I describe the construction of my dataset. This API access allows the user to download all information associated with a single object. As there is no way to use the API access to download information on the entire universe of objects in CrunchBase at once, I serially download all information on each object of interest.

1.3.3.2 Namepedia

Namepedia is the largest information portal for personal names in the world (Namepedia, 2015). I use it to classify by gender those personnel that I cannot categorize myself. Namepedia is partially crowdsourced, much like CrunchBase. Therefore, it possesses many of the same strengths and weaknesses. To improve data quality, Namepedia staff verify crowd-sourced name data. The database also uses national census data and birth statistics to build up its database of name information.

I access Namepedia gender data using a "webscraping" procedure. Webscraping is a method wherein you access a webpage and extract data from the HTML code of that webpage. For my purposes, I request the Namepedia webpage for a first name and then, from the provided webpage, extract the field with the name's associated gender. In Figure 1.8, I show the data that I webscrape from each webpage, using one gender-neutral name and one gender-specific name as examples. Namepedia provides many different gender categories: "Female," "Male," "Neutral," "Unknown," "More female, also male" and "More male, also female." I only use the "Female" and "Male" categorizations to avoid miscategorization problems.

[10] Additionally, recent evidence based on a comparison of Wikipedia to other encyclopedias suggests that the error rate in crowdsourced data may be lower than data gathered otherwise (Giles, 2005; Casebourne et al., 2012).

1.4 Participation

In this section, I explore whether there is any systematic difference in participation within VC-financed entrepreneurship based on gender. First, I examine female and male VC-financed entrepreneurs' participation levels. Next, I consider VC general partners' participation by gender. I find clear evidence of differences in participation by gender in both domains, with a trend towards equality among entrepreneurs which is absent among GPs.

1.4.1 Entrepreneurs

As presented in Table 1.2A, I have founder gender data for 2,372 startups, out of which, 352 (14.8%) are led by one or more female founders. The 352 female-led startups are led by 402 female founders in my data, which comprise 8.8% of the 4,568 entrepreneurs in the data with gender information. The 14.8% female participation figure confirms the results of the Diana Project survey in 2011-2013, which found that 15% of VC-financed firms had a female executive (Brush et al., 2014).

While there are a total of 4,859 entrepreneurs, I have gender information for 4,568 entrepreneurs (94%). Similarly, out of 3,660 startups, I have some entrepreneur gender information for 2,372 (64.8%) of them. As explained in Section 1.3.3.1, the primary reason for 35% of startups having no data on founders is that, for a significant number of startups, founders are not listed in CrunchBase. A secondary reason is that I am unable to categorize an entrepreneur's gender, but, as the 94% match rate implies, this is a much smaller part of the issue than the missing data in CrunchBase.

How do female participation rates in VC-financed entrepreneurship compare to other sectors of the economy? Compared to overall female participation in the US workforce, it is substantially lower. In 2012, 47% of the workforce in the US was female (Sewell, 2013). This is perhaps unsurprising, given that women's participation in STEM fields is known to be lower than in the overall economy and VC-financed startups operate largely in STEM fields. For instance, only 27% of computer science and math positions and only 24% of STEM positions overall were filled by women in 2009 (Beede et al., 2011). But, even compared to females' STEM participation, female participation in VC-financed entrepreneurship is lower, being slightly more than one-third the rate of female STEM workforce participation.

And compared to general entrepreneurship, female participation in VC-financed entrepreneurship is also considerably lower. In 2012, 36% of small businesses in the US were majority-owned by women, according to the Small Business Administration (Lichtenstein, 2014). Of course, the typical firm financed by VCs differs from the typical firm financed by bank loans and other early-stage financiers. But, taken together with the differences between female participation in VC-financed entrepreneurship and in other sectors of the economy, these findings suggest that there is something peculiar about VC-financed entrepreneurship that leads to female participation being lower than in all other comparable settings.

Over time, the difference between female and male participation is declining. As reflected in Figure 1.3A, approximately 8.5% of startups initially financed in 2005 had a female founder, whereas nearly 17% had a female partner in 2014. We find a similar trend for the percent of entrepreneurs that are female. Figure 1.3B shows that approximately 5% of entrepreneurs in initial financing rounds were female in 2005 and nearly 10% were female in 2014. This is a dramatic change over the course of a decade. A possible reason for this trend is the emphasis placed on encouraging women's participation in STEM entrepreneurship in the last decade or so through the InnovateHER Women's Business Challenge run by the Small Business Administration and other such programs (see Council of Economic Advisers, 2015).

There is a greater increase over time in female participation in initial financing rounds than in all rounds. As Figure 1.3A shows, the growth in female-led startups' financing rounds overall has been somewhat lower, increasing from 9% to 14.5%

between 2005 and 2014. This could arise for two reasons: (a) female-led startups succeed more quickly than male-led startups (needing fewer financing rounds in the interim) or (b) they fail more quickly than male-led startups. Given the performance results I present in Section 1.5, it seems likely that the latter is driving the difference in female participation between the two sets of financing rounds.

1.4.2 General partners

Much like the startups they finance, VCs tend to have low female participation rates. As presented in Table 1.2B, out of 5,970 GPs, I have gender information for 5,672, 514 (9.1%) of whom are female. As many GPs in VCs are successful former entrepreneurs themselves, it is not surprising that the rate of female participation is similar in the two groups.

Table 1.2B also shows that, out of the 10,015 financing rounds, I have some GP gender data for 5,682 rounds (56.7%)[11] and about 63% of these rounds are financed by a VC syndicate with one or more female GPs. The percentage of VC syndicates with female GPs is so high because there are multiple VCs in each syndicate, which dramatically increases the proportion of financing rounds where the financing syndicate has at least one VC with a female GP.

Unlike entrepreneurs, GPs do not exhibit a secular increase over time in female participation. As shown in Figure 1.3C, approximately 60% of VC syndicates involved in initial financing rounds had VCs with one or more female GPs in 2005 and 45% had female GPs in 2005, which suggests that female participation actually fell among VCs over time. However, as Figure 1.3D shows, this trend is unique to syndicate-level aggregation; overall female GP participation in initial financing rounds is around 7.5% in 2005 and 2014. But, it seems clear that there is no upward trend in female participation among VCs.

[11] The main reason that 43% of the rounds do not have GP gender information is, as on the entrepreneur side, that the data are missing in CrunchBase.

Also unlike entrepreneurial participation, initial and all round participation rates are similar. Around 60% of syndicates involved in all financing rounds had one or more female GPs and 7.5% of GPs were female across all financing rounds in 2005 and 2014.

1.5 Performance

In this section, I explore the interaction of gender with performance of VC-financed startups. I use VC financing exit, either via IPO or acquisition, as an indicator of good performance. In the first part of this section, I present the overall exit information for the startups in my sample and discuss other measures related to performance. Next, I explore whether founder or GP gender affects performance. I find that there is a large performance gap between female- and male-led startups but no evidence of a difference in the overall performance of the portfolios of female versus all-male VC syndicates.

1.5.1 Exit as performance

I measure VC-financed startups' performance using exit from VC financing via initial public offering (IPO) or acquisition. Table 1.3 presents the exit counts and rates for four subsets of my startups. The second set of columns in the table shows exit statistics for startups initially financed in 2005 or later. These startups have an overall exit rate of 17.1%, with slightly under one-fifth exiting via IPO (3.3%) and the rest exiting via acquisition (13.8%). This four-to-one ratio of acquisition-to-IPO exits is roughly consistent with overall sector exits, as reported by the National Venture Capital Association (NVCA). In its 2016 Yearbook, NVCA reported that there were 2,010 IPOs and 7,515 acquisitions of VC-financed startups between 1995 and 2015 (Haque, 2016), which is quite similar to the ratio I observe.

The first column of Table 1.3 shows the exit statistics for startups initially financed prior to 2005. As can be seen, the rates of exit are higher than in the second column, both overall (38.1%) and via IPO (14.8%) and acquisition (23.3%). The IPO rate is nearly four times higher than in the second column. The higher rates for this sample are evidence of the backfill bias that exists among startups in CrunchBase prior to its establishment in 2005. The pre-2005 rounds filled in later are more likely to be tied to startups that had financial activity after 2005, otherwise they would be unlikely to be reported. This backfill bias is even observable in the ratio of IPOs to acquisitions, which, in the pre-2005 sample, is far higher (two-fifth are IPO exits) than among startups in the rest of the data and as reported by NVCA for the sector as a whole. To avoid issues tied to this backfill bias, I exclude the pre-2005 startups from analyses.

Looking at trends in exits over time, I find evidence that acquisitions are more prevalent than IPOs in the early part of the sample, which coincides with the recession of 2006-2009. Figure 1.4 shows the number of exits, overall and via IPO or acquisition, for each year of exit. It shows that acquisitions make up a much larger portion of overall exits in 2005-2009. However, the greater prevalence of acquisition exits in the early years of the sample is also consistent with acquisition exits generally occurring earlier than IPO exits. This fact is confirmed in Table 1.4, which shows that IPO exits take almost one-and-a-half times longer than acquisitions, which, on average, occur just under 4 years after the initial financing round. The shorter time-to-exit for acquisitions may be exacerbated by the recession in the first half of the sample period. Alternatively, startups may have a greater preference for quicker, acquisition exits during recessions.

I exclude startups initially financed after 2010 from my performance analyses because of the noisiness of exit as a measure of performance. This noisiness can be observed in the difference between the third and fourth columns of firms in Table 1.3, where we clearly observe that the rate of exit for late entrants (startups with initial financing rounds after 2010) is five times lower than that for the other startups. The late entrants have an IPO rate of just 1% and an acquisition rate of 4.4%. As both exits are relatively rare in the late entrant sample, exit is a coarse and noisy measure of performance, as it may not pick up "good" startups that simply require more time to exit VC financing. Anecdotally, we know that both Facebook and Google took six years from their initial financing round to their IPO. Three years out from their initial financing, neither Facebook nor Google would be considered "good" startups.

The noisiness of exit for late entrants can also be confirmed visually in Figure 1.5. The figure depicts the percentage of all startups that have successfully exited VC financing after a given number of years since their initial financing round. Looking at the figure, it is apparent that, even after two years, less than 5% of startups have successfully exited VC financing, whereas, given ten years, over 15% are able to exit.

While exit from VC financing is often used as a measure of performance in the VC literature[12], it cannot distinguish between exits that provide large versus small returns on VC investment. Returns cannot be calculated for startups in the data because of a lack of information about the VC contracts offered to startups in exchange for funding.[13] However, Hochberg et al. (2007) provide some assurance that, at the fund level, exit rates are positively correlated with returns: based on Freedom of Information Act suits, they find a correlation of 0.42 between exit rates (via IPO or acquisition) and funds' IRRs. Given the lack of data necessary to calculate returns at the startup level, exits are the best, albeit an imperfect, measure of startup performance available.

[12] For instance, Hochberg et al. (2007) use portfolio firm exits via IPO or acquisition to measure fund performance. Gompers et al. (2010) use exits via IPO to measure entrepreneur success (and find that results are similar if they include acquisition as a success).

[13] In order to calculate returns for the initial financiers' investment, the empiricist needs to know not only the contract details for the initial financing but also for all intermediate investments in the startup, as each of those investments may dilute the stake of the initial financier in the company. This makes it even harder to calculate returns on investment for the VC financiers of these startups.

1.5.2 Gender effects on performance

Does the gender of a startup's founders play a role in its performance? To examine startup performance by founder gender, I separate the data into two groups: startups with all male founders ("male-led startups") and startups with one or more female founders ("female-led startups"). Comparing the exit rates for these two groups in Table 1.5A, I find that male-led startups have a 27.4% overall rate of exit and female-led startups have a 17.3% rate, a difference of 10.1% which is statistically highly significant. The difference is also economically large, being slightly more than one-third of the exit rate for male-led startups. Acquisitions and IPOs reveal an economically similar difference between the two groups: female-led startups tend to exit one-third less often than male-led startups for both types of exits. The IPO exits difference is not statistically significant primarily because IPOs occur infrequently, making it more difficult to establish statistical significance. On the other hand, acquisitions are more common and we see that the difference between the two groups is statistically significant.

Unlike founder gender, GP gender does not matter for startup performance. As with the founder gender comparison, I split the data into two groups for this analysis: startups initially financed by VC syndicates with all male GPs ("male VCs") and those initially financed by syndicates with one or more female GPs ("female VCs"). As shown in Table 1.5B, I find there to be almost no difference at all between the two groups in terms of exits, overall or via IPO or acquisition.

How should we interpret these starkly different impacts of founder and GP gender on performance? The worse performance of female-led startups is interesting in that it is inconsistent with screening discrimination predictions. If VCs engage in taste-based discrimination against female-led startups, as defined in Becker (1971), the female-led startups they do finance should be of higher quality and, consequently, perform better. In the data, we see the reverse. Assuming that the quality distribution of female- and male-led startups is similar, this finding suggests that taste-based screening discrimination is not the sole contributor to gender differences in participation rates. The lack of difference in exit rates based on GP gender suggests that startups financed and advised by female VCs and male VCs are similar in terms of performance. Following up on this no-difference finding, in the following sections, I explore whether, rather than having an overall effect, VC gender has any effect on *differences* between the participation of female and male entrepreneurs and the performance of female- and male-led startups.

1.6 VC effect on participation gap

In Section 1.4, we observed that females participate far less than males on both sides of the VC financing table. Could this jointly low participation arise from same-gender matching among founders and GPs? That is, could female GPs' preference for female-led startups and female-led startups' preference for female GPs lead to low female participation in VC-financed entrepreneurship, as there are relatively few female entrepreneurs and GPs? As reported in Table 1.6, for female VCs, 15.6% of initially financed startups are female-led and, for male VCs, 17.4% of financed startups are female-led. The 1.7% difference in female-led startup representation is not statistically significant. This suggests that there is no same-gender preference among founders and GPs (or an opposite-gender preference) in terms of financing.

As shown in Figure 1.6, the proportions of female-led startups' initial round financed by female and male VCs remains similar over time. While male VCs have a larger female-led startup representation in 2009 and female VCs have a larger representation in 2013, in general, the percentages of initial financings rise in tandem for the two VC groups, from under 5% in 2005 to just over 10% in 2014. The rise in the interim period of female-led startups' initial financings is consistent with the increase in female participation as entrepreneurs discussed in Section 1.4.

Both the overall similarity and the persistence of the similarity of female-led startup initial financings over time add to the evidence against taste-based discrimination playing a role in the lower participation of female entrepreneurs. Assuming that the VC groups have differing tastes for discriminating against women, we should expect the group with the lesser taste to finance a greater proportion of female-led startups, which we do not observe in Table 1.6. Additionally, given that taste-based discrimination may reduce profits, the VC group with less of a taste for discrimination should "crowd out" the other VC group over time with respect to female-led startup financing.[14] Again, we do not see such a crowding out of either VC group in Figure 1.6. Under the assumption that these VC groups have differing tastes for discrimination, these findings suggest that taste-based discrimination does not drive the lower female participation that we observe.

1.7 VC effect on performance gap

Does the performance gap between female- and male-led startups differ based on who finances them? In Figure 1.7, I present overall exit rates that highlight a stark difference in the performance gap observed for startups financed by the two VC groups. We see that the exit rates for female- and male-led startups are approximately the same for startups financed by female VCs, whereas male-led startups financed by male VCs have an exit rate approximately 25 percentage points higher than their female-led counterparts. Additionally, the difference in the performance gap between the two VC groups is due to better performance of female-led startups initially financed by female VCs. This finding suggests that female VCs are better able to evaluate and/or advise female-led startups. In the following subsections, I examine this hypothesis more rigorously using regression analysis and attempt to separate the above hypothesis from other conjectures consistent with these findings.

[14]This is another implication of Becker (1971).

1.7.1 Regression design

To rigorously examine whether the findings presented in Figure 1.7 imply that female VCs are better at evaluating and/or advising female-led startups, I test whether startups initially financed by female VCs exhibit a different gap in exit rates based on founder gender than startups initially financed by male VCs. The null hypothesis for this test is that there is no difference in the exit rate gender gap between startups financed by the two sets of VC syndicates. If there is a difference in the exit rate gender gap, I reject the null hypothesis.

The null hypothesis, that the performance gap between female- and male-led startups is the same in both VC groups, implies that any difference in exit rates between female- and male-led startups stems from differences between female and male entrepreneurs. There are a number of gender differences that could drive these differences in exit rates between female- and male-led startups. For instance, a number of papers discuss lower risk tolerance among females (see Powell and Ansic, 1997; Barber and Odean, 2001), which could drive female entrepreneurs to lead startups that have a lower likelihood of extreme right tail outcomes.[15] Larry Summers forwards another theory that the far right tail of the ability distribution may be more populated by males than females (see Tierney, 2010). Finally, there is evidence that females are more averse to competition than men (see Croson and Gneezy, 2009), which could explain worse performance in an environment as competitive as VC-financed entrepreneurship.

If the exit rate gender gap is different across the two VC groups, I reject the null hypothesis. There is empirical evidence supporting both a narrowing and a widening of the founder gender-based gap with financing from female VCs. On the narrowing side, the aforementioned Tate and Yang (2014) and Gompers et al. (2014) show that

[15] Note that there is still debate as to whether there truly exists a difference in risk aversion between men and women. For instance, Nelson (2015) states that contextual influences may be driving some of the risk aversion findings in the literature.

female leadership within a group improves the outcomes of other females in the group as well. On the widening side, Gompers et al. (2012) shows that a shared identity among GPs based on ethnic background reduces the likelihood of investment success.

The specification for the above-detailed regression is

$$\Pr(exit_i = 1) = F\left(\alpha + \gamma_1 fem_i^e + \gamma_2 fem_i^v + \beta(fem_i^e \times fem_i^v)\right), \tag{1.1}$$

where $F(z) = e^z/(1 + e^z)$ is the cumulative logistic distribution and $exit_i$ indicates whether firm i exits venture financing successfully, fem_i^v indicates whether firm i has at least one female GP in its initial financing syndicate, and fem_i^e indicates whether firm i has at least one female entrepreneur. As $exit_i$ is a binary outcome variable, I perform a logistic regression analysis. If the marginal effect of the interaction of fem_i^e and fem_i^v is non-zero, I reject the null hypothesis, as the non-zero effect implies that the performance gap is different between female and male VCs.

1.7.2 Regression results

While Equation 1.1 does not list any covariates, I include the following independent variables as controls: (a) initial financing year fixed effects, (b) sector fixed effects, (c) number of GPs in the initial financing VC syndicate, (d) total number of startups financed by syndicate members in the past, (e) interaction of number of GPs with a female-led startup indicator, and (f) interaction of number of past deals with a female-led startup indicator. Initial financing year fixed effects are included to account for differences between startups due to macroeconomic changes over time. Sector fixed effects are included to account for differences in startups due to their operating sectors. Number of GPs in the syndicate is included to account for the effect of VC syndicate size on startup performance. Number of past deals is included to account for the effect of VC experience on startup performance.

I include VC size and experience as controls because I find that they differ significantly between female and male VCs. As Table 1.7 shows, the main differences between the two VC groups are in the average size of the VC syndicate and the aggregate experience of the syndicate.[16] While these differences arise because of how I define the two VC groups[17], I include controls for the level of VC size and experience as well as their interactions with the female-led startup indicator. As I explain below, the findings remain statistically significant and economically meaningful even with inclusion of these controls.

Table 1.8, column (1), presents results for the basic specification from Equation 1.1, showing that the gender gap in exit rates among startups initially financed by female VCs is nearly 31% narrower than among startups financed by male VCs.[18] Including initial financing year fixed effects and operating sector fixed effects does little to alter either the magnitude or statistical significance of the impact, as seen in columns (2) and (3), respectively. In column (4), we see that including the level effects of VC syndicate size and experience also does not greatly alter the impact of female GP presence on the gender gap in exit rates.

In the second-to-last row of Table 1.8, I present the likelihood that exit rates of female- and male-led startups financed by female VCs are the same. This likelihood is high across all specifications in the table, meaning that there is no performance gap among startups financed by female VCs. It also implies that the difference in the gap between startups financed by the two VC groups is comparable in magnitude to the performance gap itself.

The last row of Table 1.8 shows the likelihood that female-led startups financed

[16] There is also a significant difference in the total number of financing rounds for financed startups, it is not economically large.

[17] Assuming that female GPs arrive randomly, the likelihood of a larger VC having a female GP is higher than that of a smaller VC. Therefore, the average size of syndicates with female GPs is mechanically larger than that of syndicates with all male GPs.

[18] Point estimates presented in all regression tables have been transformed to be directly interpretable as average marginal effects (in percentage points) on dependent variables. The marginal effects of interactions are calculated as in Ai and Norton (2003).

by female VCs have the same exit rates as female-led startups financed by male VCs. For columns (1) through (4), this likelihood is below 10%. For the two specifications discussed next, the likelihood is just above 10%. This is fairly strong evidence that female-led startups financed by female VCs have significantly higher exit rates than female-led startups financed by male VCs. On the other hand, based on the second row of results in the table, we see that, across all regressions, male-led startups' exit rates are the same, regardless of female GP presence in the initial financing syndicate. Together, these two findings imply that the narrower gender gap for startups financed by female VCs comes from higher exit rates for female-led startups and not from lower exit rates for male-led startups.

Interestingly, including the interaction between VC syndicate size and the female-led startup indicator in column (5) of Table 1.8 reduces the impact to 26.3%. Similarly, including the interaction of VC experience and female-led firm indicator in column (6) drops the impact to 24.6%. While both of these estimates are somewhat smaller and less statistically significant than the 31% reported for the "raw regression" from column (1), they are statistically significant at the 10% level and a 25% impact on exit rates is a sizeable impact of female GP presence on the performance gap between female- and male-led startups.

1.7.3 Interpretations

I interpret the reduced performance gap among startups financed by female VCs as evidence that female VCs are better able to evaluate and/or advise female-led startups than male VCs. When female-led startups seek VC financing, female VCs are either better able to judge the startup's future performance or better able to advise the female-led startups that they choose to finance. In Section 1.8, I explore, among other questions, whether the difference arises due to better evaluation or advising. In this subsection, I consider alternative interpretations of the finding.

Besides my preferred interpretation, there are other possible interpretations of the main finding. Prime among these is one that posits that male VCs may finance a larger proportion of the female-led startups that approach them for financing, which results in their female-led startups having lower intrinsic quality and performance than those financed by female VCs. This would result in a greater performance gap among startups financed by male VCs. However, it would also result in a greater proportion of female-led startups in the portfolios of male VCs. As discussed previously, Table 1.7 shows that male VCs do not have a greater proportion of female-led startups than female VCs. This means the male VC overfinancing conjecture is inconsistent with the data.

1.7.3.1 Entrepreneur financing choice

There are also a number of alternative interpretations tied to entrepreneurs' choice of financing. I cannot rule out all entrepreneur financing choice conjectures. However, in this section, I rule out the three most commonly-suggested ones. Ruling out these alternatives greatly improves the likelihood that the observed difference in the performance gap arises from female VCs' ability to evaluate or advise female-led startups better.

For each argument in this section, I make two assumptions. First, I assume that my preferred interpretation of the findings is not true and VC abilities and actions are the same across female and male VCs. Second, I assume that the overall supply of male-led startups is weakly greater than that of female-led startups. This assumption is empirically supported in the data, as Table 1.2A shows that only 14.8% of startups are female-led.

High value startups seek financing from female VCs. One conjecture about entrepreneur financing choice is that entrepreneurs with high intrinsic value startups preferentially seek financing from female VCs and, therefore, startups financed by such VCs have better exit rates. However, if *all* high intrinsic value startups prefer financing from female VCs and everything else is the same across the two VC groups, the exits of female- and male-led startups should be the same within each VC group and we should observe the same performance gap in both groups. As we see a larger performance gap in startups financed by male VCs, this conjecture is inconsistent with the data.

Building on the simple conjecture above, if the distribution of intrinsic value is more right-skewed for male-led startups, a greater proportion of male-led startups are likely to exit VC financing generally. As a result, the difference in exit rates among male-led startups financed by the two VC groups is narrower than the difference among female-led startups financed by the two VC groups. This generates the observed narrower performance gap among startups financed by female VCs.

This, more nuanced, conjecture also has implications that are inconsistent with the data. If high value startups preferentially seek financing from female VCs, startups financed by them should have higher overall exit rates than startups financed by male VCs. However, Table 1.7 shows no differences in the exit rates of startups initially financed by the two VC groups. This leads me to rule out this entrepreneur financing choice explanation for the performance gap difference across the VC groups.

Female-led startups seek financing from female VCs. Another conjecture related to entrepreneur financing choice posits that female-led startups preferentially seek financing from female VCs and, therefore, female-led startups financed by female VCs have better exit rates. Unlike the previous conjecture, regardless of the underlying distributions of intrinsic value for female- and male-led startups, this conjecture is inconsistent with the findings of this paper. Such a financing preference among *all* female entrepreneurs has no impact on the performance gap, assuming that the two VC groups are alike in their treatment of startups. As a result, I rule out this explanation of my findings as well.

High value female-led startups seek financing from female VCs. The third entrepreneur financing choice conjecture is a combination of the first two. It posits that high intrinsic value female-led startups preferentially seek financing from female VCs and, therefore, female-led startups financed by female VCs have better exit rates. This conjecture is consistent with the performance gap findings of this paper.

Assuming that VCs carefully scrutinize startups on expected future performance when evaluating them, this conjecture further implies that female VCs have greater proportions of female-led startups in their portfolios.[19] If more high value female-led startups seek financing from female VCs, given the same screening rules across the two VC groups, more female-led startups should pass screening for female VCs and, as a result, female VCs should have a higher proportion of financed female-led startups. This is not what we observe in the data. As Table 1.6 shows, the two VC groups finance equal proportions of female-led startups. This finding leads me to rule out this conjecture as an explanation of my findings.

1.8 Further exploration

In this section, I explore some questions that, while not directly related to the gender and VC financing focus of this paper, arose as a result of the investigation. For instance, I consider whether the difference in the performance gap between startups financed by the two VC groups arises because of better evaluation or better advising by female VCs. I also explore whether the difference is due to the female GPs within female VCs or the culture of female VCs. Finally, I consider whether the performance gender gap I observe is limited to only IPO or only acquisition exits. In general, I find that the difference likely arises because of the better ability of female GPs within

[19]If this assumption does not hold and VCs are passively financing all startups that seek financing from them or evaluating them on some other characteristics, then my counterargument here does not rule out this conjecture.

female VCs to evaluate female-led startups and that the difference is not limited to either form of exit from VC financing.

1.8.1 Evaluation and advising

Which of the two VC roles contributes more to the founder gender-based performance gap? VCs may contribute to VC-financed startups' performance through their evaluation of startups seeking VC financing and their advising of startups they choose to finance.[20] To shed light on which role matters more for the difference in the performance gap, I compare the difference in the gap across initial and subsequent financing rounds. To the extent that evaluation is more important in initial financing rounds than in subsequent rounds, a greater difference in the gap in initial rounds suggests better evaluation by female VCs of female-led startups plays a role in the different performance gaps for startups financed by the two VC groups.

There are differences between initial and subsequent financing rounds that we need to consider. Foremost is that startups in subsequent financing rounds have already interacted with VCs at least once before. Therefore, it is difficult to disentangle (previous round) VC effect on performance from inherent startup quality. Table 1.9 shows that the two sets of rounds are also different in observable ways. Startups in later rounds are financed by larger VCs with more experience.[21] Female-led startups also represent a smaller proportion of later financing rounds, which is consistent with female-led startups' earlier exits documented in Figure 1.3. Their performance is also somewhat different, with initial round startups ultimately getting acquired more often and later round startups going public more often. This is consistent with IPOs requiring more time and financing rounds to occur.

[20] The performance of VC-financed startups is better if VCs are better able to evaluate and screen out the bad startups from receiving VC financing. Startups' performance is also better if VCs are better able to advise them during their VC financing stage towards a successful exit.

[21] While not directly connected to this research, this is an interesting finding in itself, as it suggests that larger and more experienced firms are less likely to interact with higher risk, earlier financing rounds.

Table 1.10 presents the results of the regression specified in Equation 1.1 run on subsequent financing rounds. Because each startup may have more than one financing round in the analysis, I cluster standard errors at the startup level. In all specifications, we see, based on the interaction marginal effects shown in the third row, that there is no difference in the performance gap between VC groups among startups financed in subsequent rounds. From Table 1.8, we know that the analogous estimate for initial financing rounds is 25-30 percentage points. This comparison of the difference for subsequent versus initial financing rounds suggests that female VCs are better at evaluating female-led startups and this better evaluation ability plays a role in the performance gap between female- and male-led startups.[22]

1.8.2 Matching or culture

Is female-led startups' better performance with female VCs due to the female GPs themselves or because of the general culture of VCs that have females in leadership positions? To help distinguish between these alternatives, I compare the difference in the performance gap between the two VC groups for initial financing rounds with only one VC versus rounds with multiple VCs. A single VC financing round has fewer GPs involved, which means that each GP is more likely to be directly involved in the financed startup. If female GPs are directly responsible for the better performance, the effect of female VCs on performance should be greater in the single VC financing rounds, as female GPs in those syndicates are more likely to be directly involved in those rounds.

Before presenting the findings for single VC financing rounds, let us compare the two initial financing round subsamples. Table 1.11 presents the differences in

[22]Given the differences between the initial and subsequent financing rounds discussed above, I hesitate to treat this difference between financing rounds as unequivocal evidence of differential impact of VC firms in initial and subsequent rounds. However, in unreported tests, I find that the difference between initial and subsequent financing rounds are highly statistically significant ($p = 0.004$).

characteristics of single VC and multiple VC financing rounds. The differences in number of VCs, GPs, female GPs, and previous financings are mechanically driven by the definition of the two groups. However, the differences in exits are likely due to lower inherent quality of startups initially financed by just one VC. We see that single VC rounds have significantly lower exit rates (20% of startups initially financed by single VCs exit, as compared to 32% of startups initially financed by multiple VCs). This difference holds for IPO exits and acquisition exits separately as well. This is consistent with an intuition that, if a startup is observably of high quality, a lot of VCs will be interested in backing it financially.[23]

Comparing Table 1.12 to Table 1.8, we see that female-led startups perform even better when initially financed by female VCs in the single VC subsample. The marginal effect of the interaction is nearly twice as large as in the overall sample. This regression is performed on 156 firms, only a quarter of the sample available for the main analysis. Even with the smaller sample, the difference in the performance gap among startups financed by single VCs is large and statistically significant. Keeping in mind the intrinsic differences between the financed startups, these findings suggest that female GPs directly influence the better performance of female-led startups and that the difference in the performance gap between startups financed by female and male VCs is a direct impact of female GPs.[24]

1.8.3 IPO versus acquisition

Do female-led startups perform better with financing from female VCs overall or is the effect limited to IPO exits or acquisitions? In Table 1.13, we observe results of regressions run exclusively on IPO exits and acquisition exits in columns (1) and (2),

[23]VCs face something analogous to the "winner's curse" found in the IPOs of high quality firms.

[24]In unreported tests, I show that the differences in the performance gap between single and multi-VC initial financing rounds are economically large although they are (barely) not statistically significant ($p = 0.109$).

respectively.[25] The first thing we observe is the general lack of statistical significance for the IPO exits regression. This is because thet est in column (1) has less power than the test in column (2). Note that the absolute magnitudes of the regression coefficient estimates is much smaller in the IPO exits regression. Because IPO exits are approximately four times rarer than acquisition exits, exit is a rarer and, therefore, noisier measure of performance in the IPO regression than in the acquisition regression.

However, in both columns of Table 1.13, we observe that the gap among startups financed by male VCs (7.8% and 27.4% for IPO and acquisition exits, respectively) is almost entirely erased among startups financed by female VCs (8.6% and 27.4%). For both IPOs and acquisitions, female-led startups have worse performance with male VCs and being initially financed by female VCs is associated with a sharp improvement in their performance. The similar overall pattern of effects across the two forms of exit suggests that female-led startup performance improves with female VC financing overall, rather than only in terms of IPO or acquisition exits.

1.9 Conclusion

In this paper, I explore the effect of gender on VC-financed entrepreneurship. I find that women's participation in the sector is low both as entrepreneurs and GPs. I also show that there is a large difference by gender in terms of performance: only 17% of female-led startups successfully exit VC financing whereas 27% of male-led firms do so, which is a 37% performance difference between them. To explore the underlying reasons for the performance gap, I delve into whether VC financiers may be responsble for it. To that end, I examine whether the performance gap among female- and male-led startups varies based on whether they are financed by female

[25] To cleanly test IPO and acquisition exits, I exclude the other form of exit from the regression sample in each regression. For example, for the IPO exits regression in column (1), I exclude all firms that exit via acquisition.

VCs or male VCs. I find that startups initially financed by male VCs have a 25 to 30 percentage point performance gap whereas startups financed by female VCs have no performance gap at all. This finding suggests that female VCs are better able to evaluate and advise female-led startups and the difference in VCs' abilities may contribute to the overall performance gap.

There are, of course, alternative interpretations of the VC effect findings. First, it may be that male VCs finance female-led startups in greater proportions, which lowers the average quality of female-led startups in their portfolios and leads to the observed difference in the gap. However, I provide empirical evidence that runs counter to this interpretation. A second set of alternative interpretations revolves around entrepreneur financing choice. While I cannot rule out all conjectures relating to entrepreneur financing choice, I eliminate three commonly-posed such conjectures as their other implications are inconsistent with the data. Ruling out these alternatives improves the likelihood that my performance gap findings are due to differences in VC ability.

To shed light on whether the difference in performance gap between female and male VCs' portfolio startups is due to evaluation or advising, I compare the difference in the gap between initial and non-initial financing rounds. This relies on the reasoning that evaluation is much more important for initial financing rounds than for subsequent ones. I find that female VCs' effect on the performance gap is much smaller in subsequent financing rounds. This suggests that better ability to evaluate female-led startups contributes to the difference in the performance gap between female and male VCs.

By comparing single VC and multiple VC financing rounds, I provide evidence on the interesting question of whether female GPs are directly responsible for improved female-led startup performance. My test relies on the reasoning that a female GP in a single VC round is much more likely to be directly involved with the financed startup. I find evidence that suggests female GPs are directly responsible for improved female-led startup performance.

What do my findings ultimately tell us? First, they confirm that females are underrepresented in VC-financed entrepreneurship, which was already documented to some extent. My performance results show that VC-financed female-led startups perform worse than their male-led counterparts. Additionally, my findings show that this worse performance arises primarily among female-led startups financed and advised by male VCs, suggesting VCs may play a role in the performance gap. While the precise mechanism is hard to pin down, I find some evidence that the performance gap difference between startups backed by female VCs versus male VCs comes from the direct involvement of female VCs in the evaluation of female-led startups.

1.9.1 Implications

My findings and interpretations have a number of important and interesting implications. First, VC contribution to the performance gap means that some intrinsically valuable female-led startups do not succeed because of VC financing, which is important in its own right. Second, VC contribution suggests that the actions of male VCs worsens the performance of their portfolio startups. This is a waste of LP-invested resources by male VCs and implies private inefficiency. If LPs choose to reduce investment in venture capital as a result, this inefficiency may have large negative externalities for the VC financing sector.

The different performance of startups backed by female and male VCs also suggests that GP characteristics impact portfolio firm performance. This is consistent with recent work by Gompers et al. that shows startups advised by female GPs perform worse than startups advised by male GPs but that the performance difference is attenuated by the presence of other female GPs. Both these findings imply that GP composition of a VC impacts the performance of its portfolio firms.

Given that VC contribution to the performance gap is privately inefficient, do we find any evidence of it dissipating or is it an immutable characteristic of VC financing? Figure 1.6 shows that female VCs may be financing more female-led startups after 2011. With female-led startups' better performance with female VCs, this trend suggests that the VC effect on female-led startups' performance, and the associated private inefficiency, may start falling in the near future.

In the broader economic context, given the role of VC financing for launching large and economically important firms, the missing successes of potentially important female-led startups may dampen economic growth, as suggested in Robb et al. (2014). Recent work by Hsieh et al. calculates that the improved allocation of talent in the labor market between 1960 and 2008 for innately talented minorities may have been responsible for 15 to 20% of the growth in aggregate output per worker in that period. Assuming that the observed difference in the gap can be eradicated, a similar improvement in allocation of resources within VC financing may increase success (as measured by exits) by nearly 12%. Given the growth prospects of VC-financed startups, a 12% overall improvement in performance may lead to similar (or even greater) economic gains.

Finally, if some VCs hurt female-led startups' performance, women may be less likely to lead VC-financed projects. This feedback from performance into participation suggests that some intrinsically valuable projects are never undertaken due to the possibility of VC-induced failure. This reduced participation is currently the focus of a serious debate in policy circles. Much of the policy focus is on increasing the appeal of entrepreneurship for women. For instance, the Small Business Administration set up a number of "InnovateHER Women's Business Challenge" events to incentivize greater female entrepreneurship (Council of Economic Advisers, 2015). This paper's findings suggest a complementary strategy to increase female entrepreneurship: focusing on increasing female participation in the VC industry. This strategy may improve not only female participation in entrepreneurship, but also their performance.

Tables and Figures

	All rounds		Initial rounds	
	N	%	N	%
Total rounds	10,015		3,660	
Rounds w/ gender data on				
founders	7,040	70.3%	2,372	64.8%
GPs	5,682	56.7%	2,223	60.7%
founders & GPs	3,944	39.4%	1,348	36.8%

Table 1.1. Gender information availability for entrepreneurs and GPs by financing round. This table reports the number and percent of rounds in the data that have gender information for entrepreneurs, GPs, and both.

Table 1.2. Female participation in VC-financed entrepreneurship.

	N	%
Startups	3,660	
with founder info	2,372	
with female founder(s)	352	14.8%
Founders	4,859	
with gender info	4,568	
female	402	8.8%

(A) Female entrepreneurial participation. This table reports the number of startups in total, with founder gender information, and with female founder(s). It also reports the overall number of entrepreneurs, those with gender information, and those categorized as female.

	N	%
Financing rounds	10,015	
with GP info	5,682	
with female GP(s)	3,583	63.1%
General partners	5,970	
with gender info	5,672	
female	514	9.1%

(B) Female GP participation. This table reports the number of financing rounds in total, with GP gender information, and with female GP(s). It also reports the overall number of GPs, those with gender information, and those categorized as female.

	Initial financing round							
	Pre-2005		2005-2014		2005-2010		2011-2014	
Successful exits	N	%	N	%	N	%	N	%
All	80	38.1	590	17.1	513	25.4	77	5.4
IPOs	31	14.8	114	3.3	100	4.9	14	1.0
Acquisitions	49	23.3	476	13.8	413	20.4	63	4.4

Table 1.3. Successful startup exits from VC financing. This table provides the number and percent of startups that exit generally, exit via IPO, or exit via acquisition from VC financing. These statistics are provided for four samples. The pre-2005 sample is comprised of startups with initial financing rounds before 2005. The 2005-2014 sample is comprised of all startups initially financed in 2005 or later, which coincides with the establishment of CrunchBase. The 2005-2010 sample removes all startups initially financed after 2010. The 2011-2014 sample includes all startups with initial financings after 2010.

	Duration (in years)	
	Mean	Median
All exits	3.85	3.52
IPOs	5.27	5.71
Acquisitions	3.51	3.16

Table 1.4. VC financing duration for successful startups. This table provides mean and median VC financing durations, in years, for startups initially financed in 2005 or later that successfully exit VC financing via IPO or acquisition. Financing duration is measured from date of initial financing round to date of exit.

Table 1.5. Performance by gender.

	Female-led startups	Male-led startups	Diff. [t-stat]
Overall exits	17.3%	27.4%	-10.1% ***
	150	1,080	[2.976]
IPOs	4.0%	6.1%	-2.1%
	150	1,080	[1.197]
Acquisitions	13.3%	21.3%	-8.0% ***
	150	1,080	[2.610]

(A) Performance of female- and male-led startups. This table presents performance measured by overall exit, IPO exit, and acquisition exit for startups led by one or more female founders ("female-led startups") and startups led by all male founders ("male-led startups") as well as the difference in performance between the two groups. All the startups in this sample have initial financing rounds between 2005 and 2010, inclusive.

	Female VC syndicates	Male VC syndicates	Diff. [t-stat]
Overall exits	29.5%	29.4%	0.1%
	689	480	[0.032]
IPOs	5.4%	5.0%	0.4%
	689	480	[0.281]
Acquisitions	24.1%	24.4%	-0.3%
	689	480	[0.111]

(B) Performance of startups initially financed by female and male VC syndicates. This table presents performance measured by overall exit, IPO exit, and acquisition exit for startups initially financed by VC syndicates with one or more female GPs ("female VC syndicates") and startups initially financed by VC syndicates with all male GPs ("male VC syndicates") as well as the difference in performance between the two groups. All the startups in this sample have initial financing rounds between 2005 and 2010, inclusive.

	Female VCs	Male VCs	Diff.
Female-led startups	118	103	
Male-led startups	637	490	
% female-led startups	15.6%	17.4%	-1.7%
N	755	593	[0.852]

Table 1.6. Female and male VCs' financing of startups by founder gender. This table presents a cross-tabulation of the number of female- and male-led startups initially financed by female and male VCs (VC syndicates with and without one or more female GPs, respectively). It also presents the percent of portfolio firms that are female-led startups for the two VC groups as well as the difference in that percentage across the VC groups (and the t-statistic for that difference). The startups in this sample are all initially financed in 2005 or later.

	Female VCs	Male VCs	Difference
Startups			
Number of entrepreneurs	1.973	1.957	0.015
	731	563	[0.263]
VC financing duration, years	3.462	3.643	-0.181
	235	155	[0.779]
Number of financing rounds, all firms	2.204	2.018	0.186***
	1,179	954	[2.890]
successful firms	2.343	2.406	-0.063
	239	160	[0.370]
Years between rounds	1.331	1.266	0.065*
	2,143	993	[1.830]
VC syndicates			
Number of VCs in syndicate	3.390	2.245	1.145***
	1,179	954	[14.062]
Number of GPs	18.088	7.939	10.149***
	1,179	954	[24.244]
Number of female GPs	1.773	0.000	1.773***
	1,179	954	[51.823]
Total number of past financings	271.285	253.116	18.169*
	1,179	953	[1.949]
% of financed startups exited	29.463%	29.375%	0.088%
	689	480	[0.032]
% of financed startups exited via IPO	7.075%	6.612%	0.463%
	523	363	[0.269]
% of financed startups exited via acquisition	25.460%	25.658%	-0.198%
	652	456	[0.074]

Table 1.7. Differences between female and male VC syndicates. The first panel of this table presents differences between startups initially financed by VC syndicates with one or more female GPs ("female VCs") and syndicates with all male GPs ("male VCs"). It includes all startups initially financed in 2005 or later. The second panel presents differences in the characteristics of VC syndicates in the female and male VC groups. It includes all VC syndicates involved in initial financing rounds in 2005 or later. For the rows on exits (the last three rows of the panel), the data are further restricted to financings of startups initially financed before 2011. For each statistic, the table provides the mean for startups financed by the two VC groups in the top row and the sample size for the statistic in the bottom row. The last column shows the difference in the statistic between startups financed by female and male VCs in the top row and the t-statistic for that difference in the bottom row, based on a t-test of means with unequal variances.

	Successful exits					
	(1)	(2)	(3)	(4)	(5)	(6)
Fem-led startup	-0.326***	-0.285**	-0.284**	-0.282**	-0.323**	-0.496
	[-2.83]	[-2.52]	[-2.51]	[-2.50]	[-2.43]	[-1.33]
Fem VC	-0.0509	-0.0469	-0.0454	-0.0505	-0.0486	-0.0478
	[-1.28]	[-1.18]	[-1.15]	[-1.23]	[-1.18]	[-1.16]
Fem-led startup × fem VC	0.309**	0.299**	0.290**	0.288**	0.263*	0.246*
	[2.24]	[2.21]	[2.14]	[2.13]	[1.85]	[1.69]
Num GPs				0.00212	0.00176	0.00175
				[1.12]	[0.89]	[0.88]
Fem-led startup × num GPs					0.00385	0.00396
					[0.60]	[0.62]
Log past deals				-0.0265	-0.0264	-0.0305
				[-0.75]	[-0.75]	[-0.84]
Fem-led startup × log past deals						0.0387
						[0.50]
Observations	621	621	621	621	621	621
Funding year FEs	N	Y	Y	Y	Y	Y
Sector FEs	N	N	Y	Y	Y	Y
R^2	0.0139	0.0390	0.0463	0.0484	0.0488	0.0492
Pr(No gap in fem VCs)	0.818	0.854	0.934	0.936	0.654	0.538
Pr(No diff in fem-led startups)	0.0511	0.0520	0.0593	0.0686	0.119	0.158

t statistics in brackets

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

Table 1.8. Female GP presence impact on female- and male-led startups' exit rates. This table presents the results of the logistic regression analysis detailed in Equation 1.1. The dependent variable in all columns is an indicator of successful exit from VC financing for startups initially financed in 2005-2010. The explanatory variables, in order, are an indicator for a startup with one or more female founders ("female-led startup"), an indicator for initial financing of the startup by a VC syndicate with one or more female GPs ("female VC"), the interaction of the female-led startup and female VC indicators, the number of GPs in the initial syndicate, the interaction of the female-led startup indicator and number of GPs, log of the aggregate number of deals financed by syndicate members prior to this financing round, and the interaction of female-led startup indicator and log of prior deals financed by the syndicate. In column (1), results of the basic regression specification are presented. In columns (2) and (3), initial financing year fixed effects and sector fixed effects are introduced, respectively. In column (4), controls for the level effects of number of GPs in the initial financing syndicate and log of prior deals financed are introduced. In columns (5) and (6), the interactions of the female-led startup indicator with number of GPs in the syndicate and log of prior deals financed are introduced, respectively. The second-to-last row of the table presents the likelihood that there is no gender gap in exit rates between female- and male-led startups initially financed by female VCs. The last row presents the likelihood that there is no difference in the exit rates of female-led startups initially financed by male VCs and those financed by female VCs.

	Initial round	Subsequent rounds	Difference
Number of VCs in syndicate	2.2	2.5	-0.4***
	3,450	5,708	[8.782]
Number of GPs	8.4	9.7	-1.3***
	3,450	5,708	[5.239]
Number of female GPs	0.6	0.7	-0.1***
	3,450	5,708	[5.177]
Aggregate number of previous financings	200.7	289.2	-88.5***
	3,078	4,650	[15.980]
% female-led startups	15.2%	13.2%	2.0% **
	2,246	4,249	[2.228]
% of financed firms exited	25.4%	22.4%	3.0% ***
	2,021	4,336	[2.604]
% of financed firms exited via IPO	6.2%	9.4%	-3.2% ***
	1,608	3,717	[4.187]
% of financed firms exited via acquisition	21.5%	15.5%	6.0% ***
	1,921	3,985	[5.427]

Table 1.9. Differences between initial and subsequent financing rounds. This table presents differences between initial and subsequent financing rounds. It includes all startups initially financed in 2005 or later. For the rows on exits (the last three rows of the table), the data are further restricted to financings of startups initially financed before 2011. For each statistic, the table provides the mean for initial and subsequent financing rounds in the top row and the sample size for the statistic in the bottom row. The last column shows the difference in the statistic between startups financed by female and male VCs in the top row and the *t*-statistic for that difference in the bottom row, based on a *t*-test of means with unequal variances.

	Successful exits, non-initial financing rounds			
	(1)	(2)	(3)	(4)
Fem-led startup	-0.0630	-0.0544	-0.0450	0.0431
	[-0.82]	[-0.76]	[-0.62]	[0.14]
Fem VC	0.00424	0.0282	-0.00740	-0.0127
	[0.14]	[0.96]	[-0.24]	[-0.40]
Fem-led startup × fem VC	-0.0540	-0.0513	-0.0578	0.0155
	[-0.56]	[-0.57]	[-0.63]	[0.14]
Num GPs			0.00272^{**}	0.00317^{***}
			[2.33]	[2.69]
Fem-led startup × num GPs				-0.00590
				[-0.92]
Log past deals			0.0267	0.0276
			[0.98]	[0.94]
Fem-led startup × log past deals				-0.00698
				[-0.11]
Observations	1726	1726	1724	1724
Funding year FEs	N	Y	Y	Y
Sector FEs	N	Y	Y	Y
SE cluster	startup	startup	startup	startup
R^2	0.00494	0.0865	0.0940	0.0957

t statistics in brackets

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

Table 1.10. Female GP presence impact on female- and male-led startups' exit rates, subsequent financing rounds only. This table presents the results of the logistic regression analysis detailed in Equation 1.1 for non-initial financing rounds. The dependent variable in all columns is an indicator of successful exit from VC financing for startups initially financed in 2005-2010. The explanatory variables, in order, are an indicator for a startup with one or more female founders ("female-led startup"), an indicator for initial financing of the startup by a VC syndicate with one or more female GPs ("female VC"), the interaction of the female-led startup and female VC indicators, the number of GPs in the initial syndicate, the interaction of the female-led startup indicator and number of GPs, log of the aggregate number of deals financed by syndicate members prior to this financing round, and the interaction of female-led startup indicator and log of prior deals financed by the syndicate. In column (1), results from the basic regression specification are presented. In column (2), initial financing year fixed effects and sector fixed effects are introduced. In column (3), controls for the level effect of number of GPs in the initial financing syndicate and log of prior deals financed are introduced. In column (4), the interactions of the female-led startup indicator with number of GPs in the syndicate and log of prior deals financed are introduced. For all regressions presented, errors are clustered at the startup level.

	Financing round with		
	One VC	Mult. VCs	Difference
Number of VCs in syndicate	1.0	3.6	-2.6***
	1,916	1,534	[54.497]
Number of GPs	1.8	16.6	-14.8***
	1,916	1,534	[48.667]
Number of female GPs	0.1	1.2	-1.1***
	1,916	1,534	[30.667]
Aggregate number of previous financings	143.6	258.1	-114.5***
	1,544	1,534	[15.307]
% female-led startups	15.2%	15.3%	-0.2%
	1,313	933	[0.111]
% of financed firms exited	19.7%	32.3%	-12.6%***
	1,104	917	[6.460]
% of financed firms exited via IPO	5.0%	7.9%	-2.8%**
	934	674	[2.246]
% of financed firms exited via acquisition	16.1%	28.1%	-12.0%***
	1,057	864	[6.329]

Table 1.11. Differences between single and multiple VC syndicates. This table presents differences between initial financing rounds with one VC and multiple VC syndicates. It includes all startups initially financed in 2005 or later. For the rows on exits (the last three rows of the table), the data are further restricted to financings of startups initially financed before 2011. For each statistic, the table provides the mean for initial and subsequent financing rounds in the top row and the sample size for the statistic in the bottom row. The last column shows the difference in the statistic between startups financed by female and male VCs in the top row and the t-statistic for that difference in the bottom row, based on a t-test of means with unequal variances.

	Successful exits, single VC rounds			
	(1)	(2)	(3)	(4)
Fem-led startup	-0.353*	-0.329*	-0.313*	-0.751
	[-1.87]	[-1.77]	[-1.69]	[-0.81]
Fem VC	-0.122*	-0.102	-0.102	-0.0985
	[-1.65]	[-1.32]	[-1.32]	[-1.26]
Fem-led startup × fem VC	0.578***	0.535**	0.509**	0.466**
	[2.65]	[2.47]	[2.35]	[2.01]
Num GPs			0.00526	0.00470
			[0.97]	[0.82]
Fem-led startup × num GPs				0.00427
				[0.25]
Log past deals			0.00330	-0.00120
			[0.06]	[-0.02]
Fem-led startup × log past deals				0.0870
				[0.46]
Observations	156	156	156	156
Funding year FEs	N	Y	Y	Y
Sector FEs	N	Y	Y	Y
R^2	0.0498	0.0754	0.0805	0.0822

t statistics in brackets

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

Table 1.12. Female GP presence impact on female- and male-led startups' exit rates, single VC syndicates only. This table presents the results of the logistic regression analysis detailed in Equation 1.1 on startups initially financed by single VC syndicates only. The dependent variable in all columns is an indicator of successful exit from VC financing for startups initially financed in 2005-2010. The explanatory variables, in order, are an indicator for a startup with one or more female founders ("female-led startup"), an indicator for initial financing of the startup by a VC syndicate with one or more female GPs ("female VC"), the interaction of the female-led startup and female VC indicators, the number of GPs in the initial syndicate, the interaction of the female-led startup indicator and number of GPs, log of the aggregate number of deals financed by syndicate members prior to this financing round, and the interaction of female-led startup indicator and log of prior deals financed by the syndicate. In column (1), results from the basic regression specification are presented. In column (2), initial financing year fixed effects and sector fixed effects are introduced. In column (3), controls for the level effect of number of GPs in the initial financing syndicate and log of prior deals financed are introduced. In column (4), the interactions of the female-led startup indicator with number of GPs in the syndicate and log of prior deals financed are introduced.

	IPO (1)	Acquired (2)
Fem-led startup	-0.0781	-0.274**
	[-1.01]	[-2.33]
Fem VC	-0.0143	-0.0427
	[-0.54]	[-1.04]
Fem-led startup × fem VC	0.0857	0.274**
	[0.94]	[1.97]
Num GPs	0.00192*	0.000838
	[1.69]	[0.43]
Log past deals	-0.000922	-0.0270
	[-0.04]	[-0.78]
Observations	465	580
Funding year FEs	Y	Y
Sector FEs	Y	Y
R^2	0.194	0.0371

t statistics in brackets

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

Table 1.13. Female GP presence impact on female- and male-led startups' exit rates via IPO and acquisition. This table presents the results of the logistic regression analysis detailed in Equation 1.1. The dependent variable in column (1) is an indicator of successful exit via IPO and in column (2) is an indicator of successful exit via acquisition from VC financing for startups initially financed in 2005-2010. The explanatory variables, in order, are an indicator for a startup with one or more female founders ("female-led startup"), an indicator for initial financing of the startup by a VC syndicate with one or more female GPs ("female VC"), the interaction of the female-led startup and female VC indicators, the number of GPs in the initial syndicate, the interaction of the female-led startup indicator and number of GPs, log of the aggregate number of deals financed by syndicate members prior to this financing round, and the interaction of female-led startup indicator and log of prior deals financed by the syndicate. For each dependent indicator variable, observations for which the values of the other dependent variable are true are omitted from the regression.

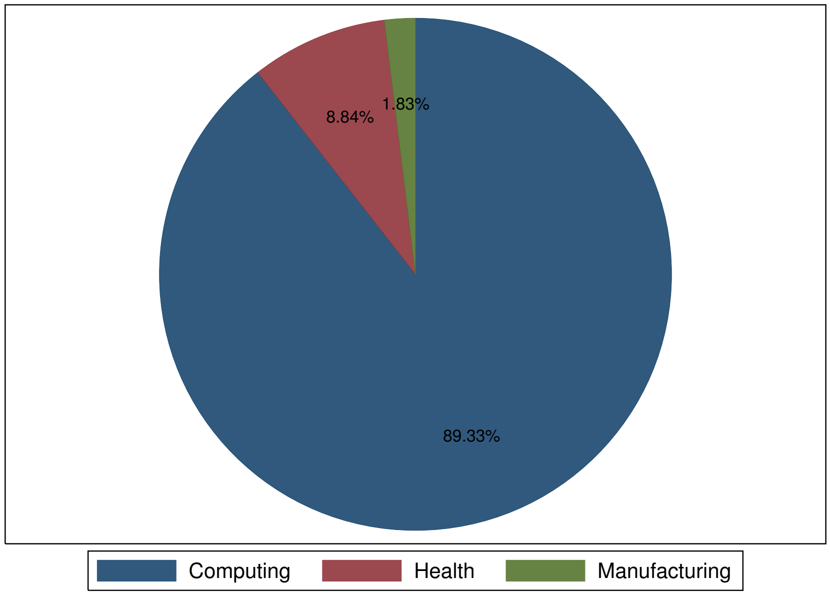


Figure 1.1. Operating sectors of startups. This figure provides a breakdown of the proportion of startups operating in each high-tech subsector. The categorization is based on textual descriptions of operating fields provided by the startups to CrunchBase.

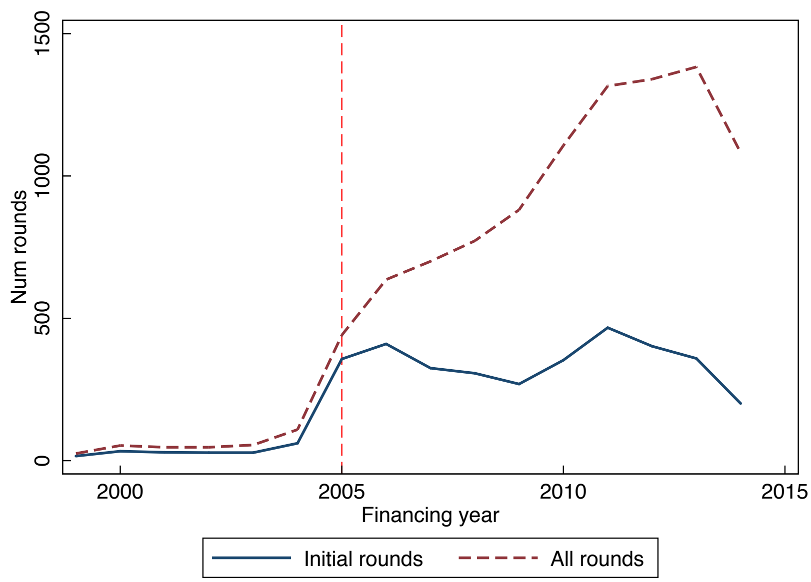
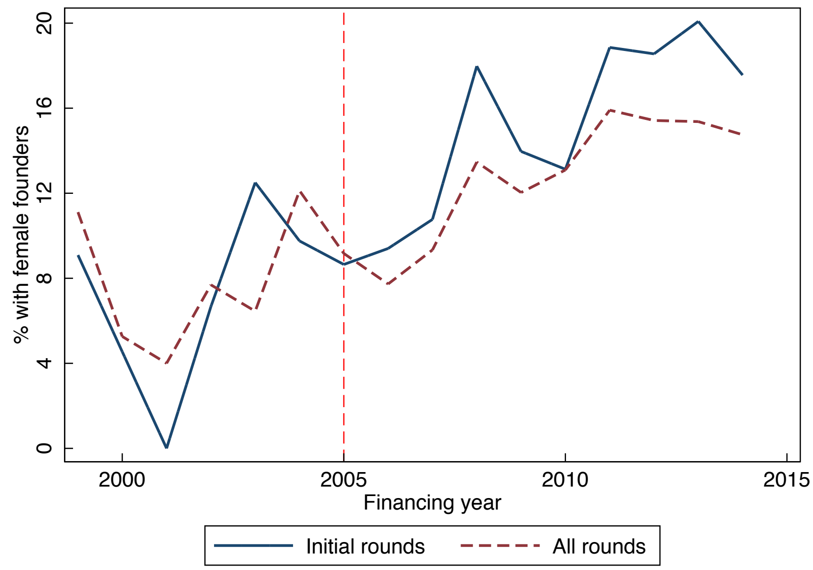
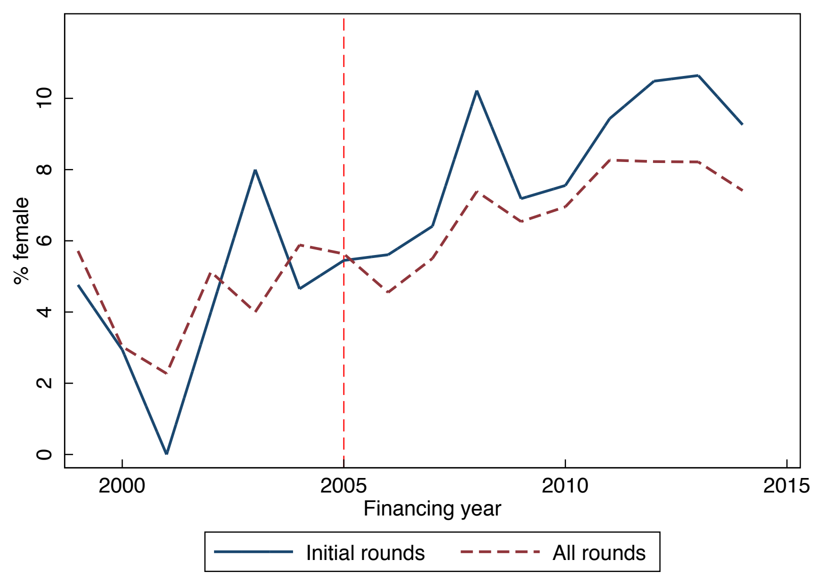


Figure 1.2. Financing rounds by year. This figure presents the annual number of financing rounds in the data from 1999 to 2014. The two plots depict initial financing rounds and all financing rounds (including initial rounds). The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.

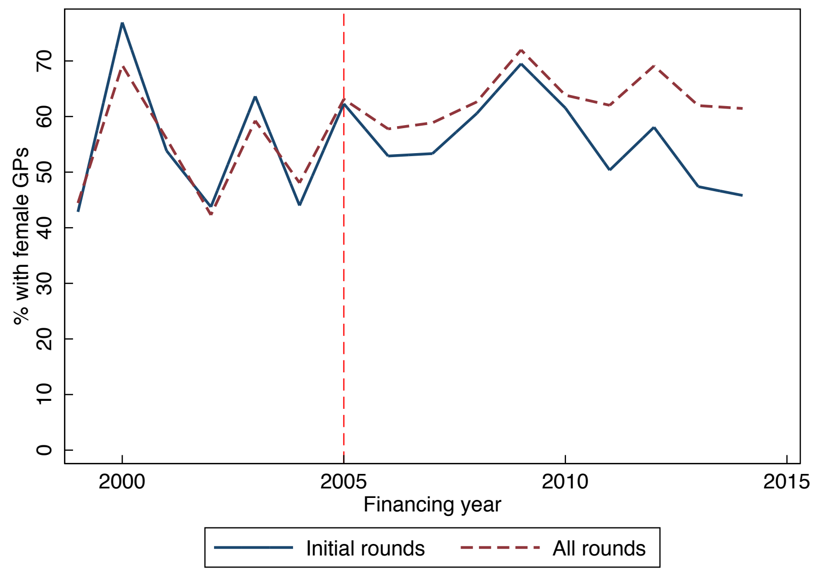
Figure 1.3. Female participation by year.



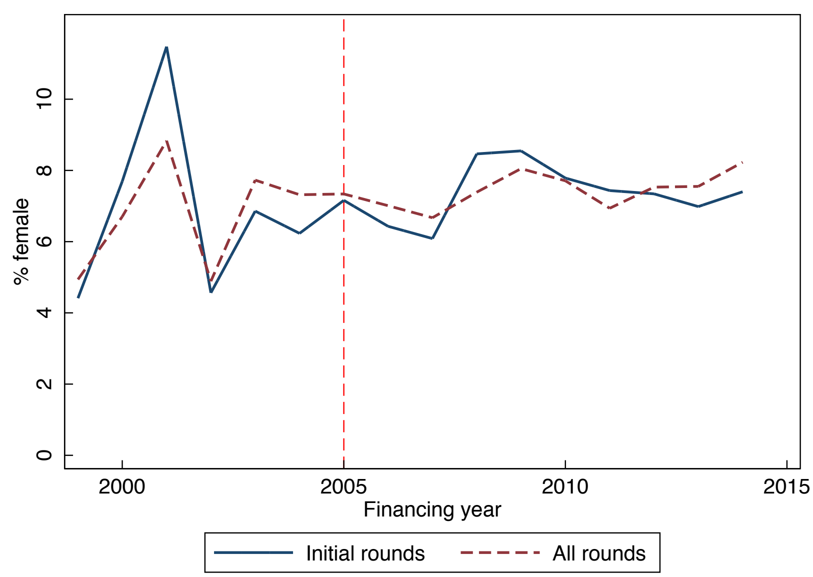
(A) Female-led startups. This figure shows the percent of financed startups with one or more female founders in each year. The plots represent female-led startup participation in initial rounds and all rounds. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.



(B) Female entrepreneurs. This figure shows the percent of financed entrepreneurs in each year that are female. The plots represent female participation in initial rounds and all rounds. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.



(C) Female VCs. This figure shows the percent of VC syndicates with VCs led by one or more female general partners in each year. The plots represent female VC participation in initial rounds and all rounds. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.



(D) Female general partners. This figure shows the percent of general partners financing projects in each year that are female. The plots represent female participation in initial rounds and all rounds. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.

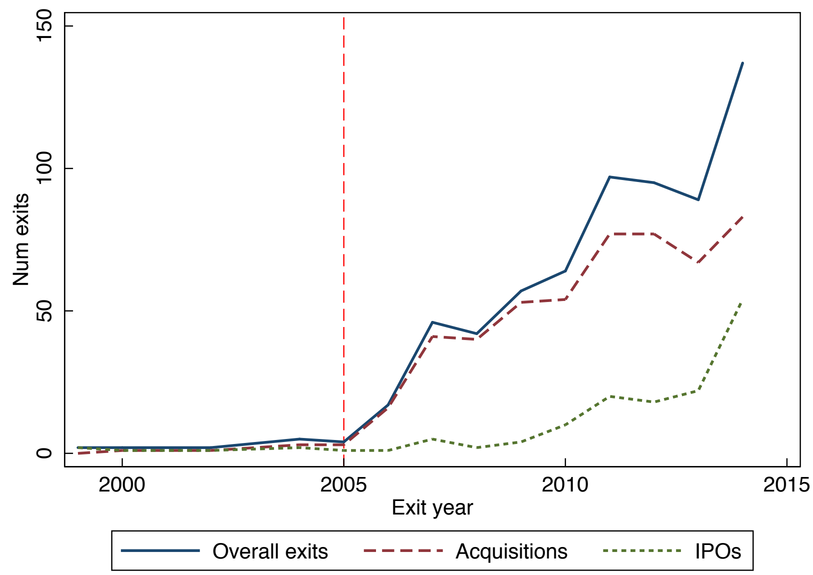


Figure 1.4. Annual number of exits, overall and via IPO and acquisition. This figure depicts the annual number of startups that have exited VC financing, overall and via IPO or acquisition. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.

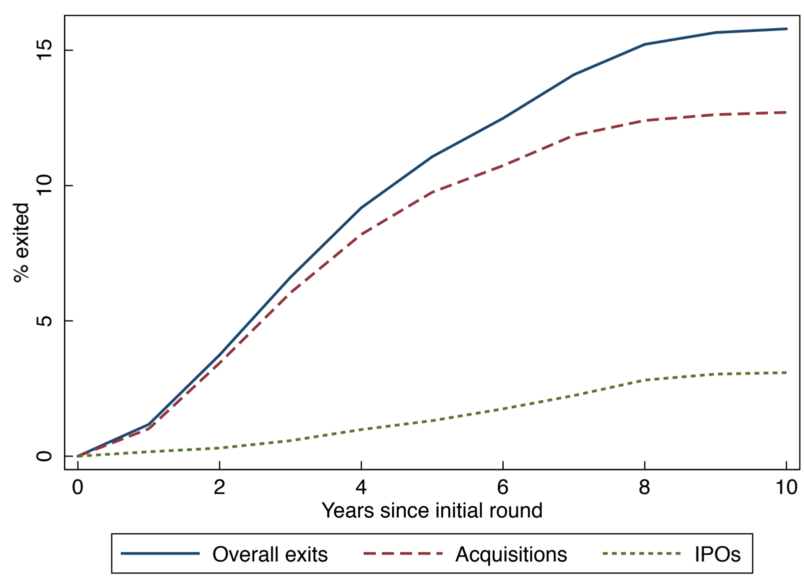


Figure 1.5. Percent of startups exited, by financing duration. This figure depicts the percentage of startups that have successfully exited VC financing (overall and via IPO or acquisition) after a given number of years, from 0 to 10. Only startups initially financed after 2005 are included in this figure. The vertical red dashed line indicates the establishment of the CrunchBase database in 2005.

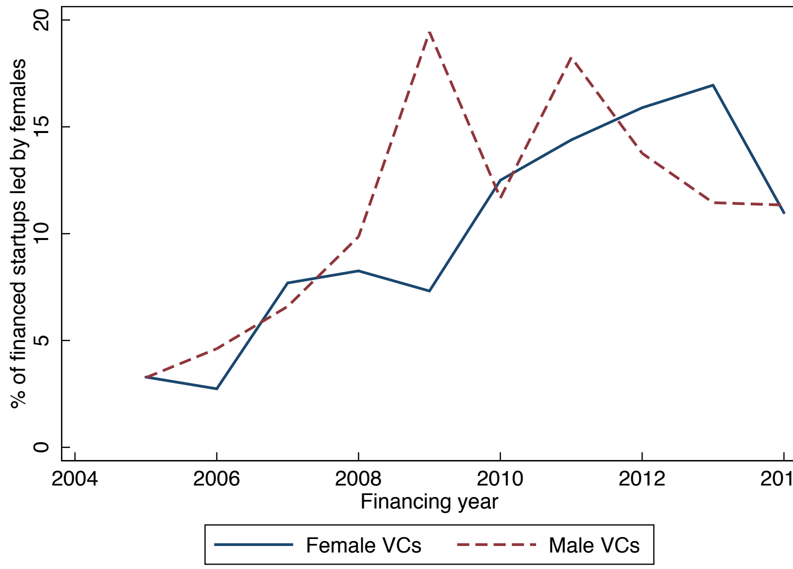


Figure 1.6. Portfolio representation of female-led startups for female and male VCs, by year. This figures plots the percentage of initial financings that are of female-led startups in each year for female and male VCs. Only startups initially financed in 2005 or later are included in the sample.

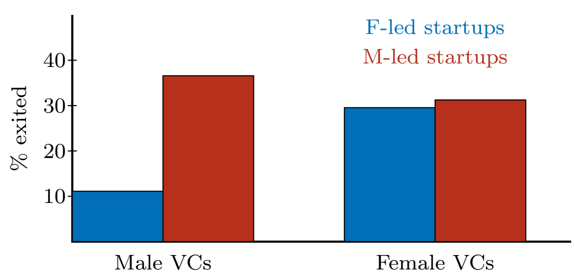


Figure 1.7. Overall exit rates, by founder-GP gender interaction. This figure depicts the overall exit rates for four groups of startups: female- and male-led startups initially financed either by female and male VCs. Only startups initially financed between 2005 and 2010, inclusive, are included in the sample.

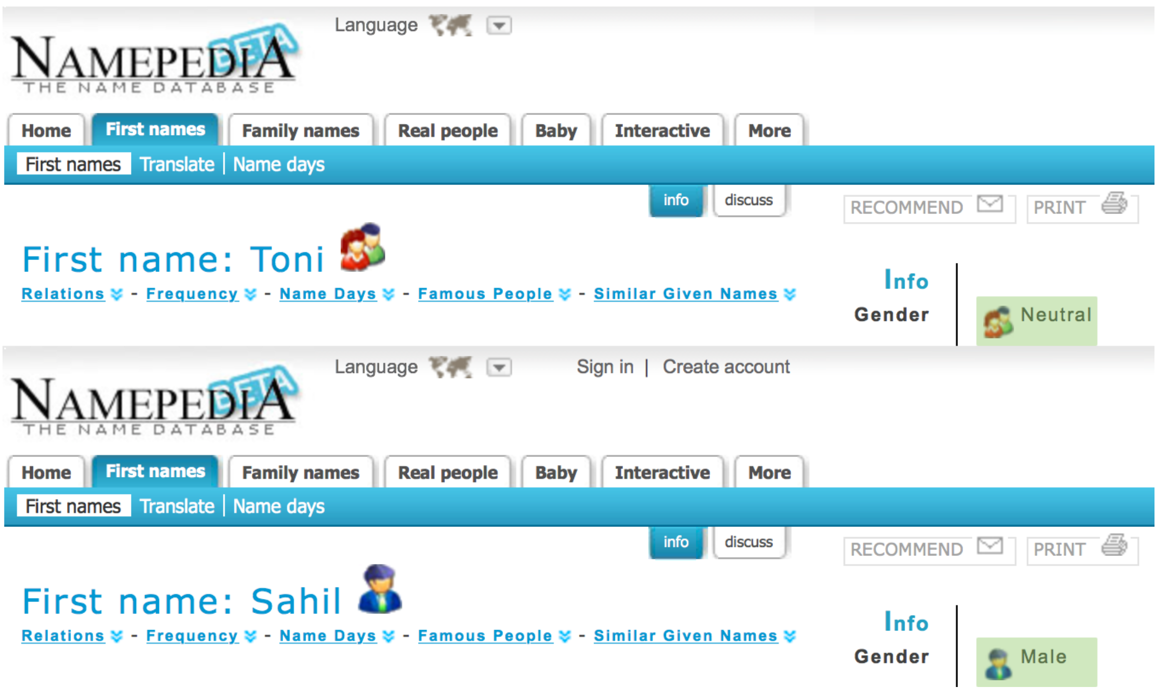


Figure 1.8. Namepedia name-gender webscraping examples. This figure provides two example of Namepedia's web response to first name queries. The top response is for a gender-neutral name and the bottom response is for a name classified as "Male." The webscraping method is used to extract the data highlighted in the top right corner of the webpage.

CHAPTER II

Wall Street and the Housing Bubble

Ing-Haw Cheng	Sahil Raina	Wei Xiong
Dartmouth College	Univ. of Michigan	Princeton University

Abstract

We analyze whether mid-level managers in securitized finance were aware of a large-scale housing bubble and a looming crisis in 2004-2006 using their personal home transaction data. We find that the average person in our sample neither timed the market nor were cautious in their home transactions, and did not exhibit awareness of problems in overall housing markets. Certain groups of securitization agents were particularly aggressive in increasing their exposure to housing during this period, suggesting the need to expand the incentives-based view of the crisis to incorporate a role for beliefs.

Appendices available online.

Did Wall Street foresee the recent crash of the U.S. housing bubble? Given the role played by Wall Street in facilitating the credit expansion that precipitated the housing market boom, understanding this question is important for systematically understanding the causes of the worst financial crisis since the Great Depression. With the benefit of hindsight, many find it hard to imagine that Wall Street missed seeing large-scale problems in housing markets before others. For example, the Financial Crisis Inquiry Commission wrote in its report that, in the years preceding the collapse, "Alarm bells were clanging inside financial institutions" (FCIC, 2011). If Wall Street was aware that the process of securitization was generating a national housing bubble that would lead to a deep financial crisis, yet proceeded to securitize mortgage loans of dubious quality, this would reveal far more severe incentive problems on Wall Street than many have recognized and confirm many of the worst fears underlying outrage from the public and policy-makers. On the other hand, if Wall Street employees involved in securitization systematically missed seeing the housing bubble, despite having better information than others, this raises fundamental questions regarding how Wall Street employees process information and form their beliefs.

In this paper, we examine this issue by studying personal home transactions of Wall Street employees. We test the simple hypothesis that they were fully aware during the boom that a large-scale housing market crisis was likely and imminent, which we term the "full awareness" hypothesis, by examining whether they avoided losses in their own homes. We focus on mid-level employees in the mortgage securitization business, such as traders, as they are a natural focal point for potential awareness of problems in the mortgage market.[1]

Because a home typically exposes its owner to substantial house price risk, mid-level employees in the financial industry, even with relatively high incomes, should have maximum incentives to make informed home-transaction decisions on their own accounts. Individual home transactions thus reveal beliefs regarding their own housing markets in isolation of any distortions arising from job incentives. Although our hypothesis concerns whether these employees were fully aware of an imminent crisis because of their superior information set, our

[1] E-mails unearthed during civil lawsuits deriding securitized mortgage instruments as garbage are rarely from C-suite-level executives, but rather are from those involved in the issuance of collateralized debt obligations (CDOs), whose job is to understand the pricing of these instruments at the center of the crisis (Coval, Jurek and Stafford, 2009). See, for example, e-mails and instant messages documented in China Development Industrial Bank v. Morgan Stanley (2013), Dexia v. Deutsche Bank (2013), Federal Housing Finance Agency v. J.P. Morgan Chase (2011), and People of the State of New York v. J.P. Morgan Chase (2012).

methodology allows us to ask broad questions about their beliefs. Were those involved with mortgage securitization pessimistic about housing markets? Or were they as optimistic, or even more optimistic, than other groups?

Indeed, a growing theoretical literature emphasizes that distortions in beliefs about house prices may have affected the development of the crisis. An environment where households neglect risk and yet demand safe assets may have endogenously fostered the very financial innovation that enabled house prices to rise with credit expansion, subsequently sparking the crisis (Gennaioli, Shleifer and Vishny, 2011, 2012). Once prices started rising, wishful thinking among agents in the financial sector may have led to contagious over-optimism, collective willful blindness, and groupthink (Benabou, 2011), an effect which may be exacerbated by cognitive dissonance (Barberis, 2012). To enhance liquidity, widely-used near riskless debt securities may have been designed to reduce any agent's incentive to acquire information about risk, thus disarming the financial system's self-correction mechanisms (Dang, Gorton and Holmstrom, 2012). During the prime years of the housing boom, the empirical literature surveying the housing market had emphasized the possibility of distorted beliefs influencing house prices (Himmelberg, Mayer, and Sinai, 2005; Mayer, 2006; Shiller, 2006, 2007; Smith and Smith, 2006), potentially arising out of contagious social dynamics (Burnside, Eichenbaum, and Rebelo, 2011) or biases such as money illusion (Brunnermeier and Julliard, 2008; Piazzesi and Schneider, 2008). While anecdotal evidence of biased beliefs has surfaced after the crisis,[2] relatively few studies since the crisis have studied beliefs, with a few exceptions (Chinco and Mayer, 2012; Foote, Gerardi, and Willen, 2012; Gerardi, et al. 2008; Glaeser, 2013; Soo, 2013). In particular, there is scant empirical evidence regarding the role played by the beliefs of Wall Street employees.

We sample a group of securitization investors and issuers from a publicly available list of conference attendees of the 2006 American Securitization Forum, the largest industry conference. These investors and issuers, whom we refer to collectively as securitization agents, comprise vice presidents, senior vice presidents, managing directors, and other non-executives who work at major investment houses and boutique firms. Using the Lexis-Nexis Public Records database,

[2] For example, Lewis (2011, p.89) suggests via anecdotes that prominent Wall Street traders did not believe house prices could fall everywhere in the country at once. In the August 2007 earnings call for American International Group (AIG), Joseph Cassano, who was involved with AIG Financial Products, says "it is hard for us with [sic], and without being flippant, to even see a scenario within any kind of realm of reason that would see us losing $1 in any of those [residential mortgage CDO] transactions." (Seeking Alpha AIG Conference Call Transcript, 2007).

which aggregates information available from public records, such as deed transfers, property tax assessment records, and other public address records, we are able to collect the personal home transaction history of these securitization agents.

We compare how securitization agents fared in housing against control groups who arguably had no private information about housing and securitization markets. We test for two forms of full awareness. First, securitization agents may have attempted to time their own housing market. A necessary condition for this strong form of "market timing" awareness is to observe home-owning securitization agents divest homes before the bust in 2007-2009. Given the difficulties of timing the market, however, awareness of a housing bubble might appear in a weaker, "cautious" form, whereby securitization agents knew enough to avoid increasing their housing exposure during the bubble period of 2004-2006.

We construct two uninformed control groups. The first control group consists of S&P 500 equity analysts who do not cover homebuilding companies. Due to their work outside securitization and housing markets, they were less likely to be informed about the housing bubble than securitization agents, yet are nonetheless a self-selected group of agents who work for a similar set of finance firms. A nuanced issue for our analysis is that securitization agents received large bonuses during the bubble years, which may motivate them to buy houses despite any potential awareness of the housing bubble. By working for similar finance firms, equity analysts arguably also experienced income shocks. Our second control group consists of a random sample of lawyers who did not specialize in real estate law. This control group serves as a benchmark for a wealthy segment of the general population and helps us understand the broader question of whether securitization agents exhibited awareness relative to the public.

Our analysis shows little evidence of securitization agents' awareness of a housing bubble and impending crash in their own home transactions. Securitization agents neither managed to time the market nor exhibited cautiousness in their home transactions. They increased, rather than decreased, their housing exposure during the boom period, particularly through second home purchases and swaps of existing homes into more expensive homes. This difference is not explained by differences in financing terms such as interest rates or financing, and is more pronounced in the relatively bubblier Southern California region compared to the New York metro region. Our securitization agents' overall home portfolio performance was significantly worse than that of control groups. Agents working on the sell-side and for firms which had poor stock price performance through the crisis did particularly poorly themselves.

Housing provides a consumption stream for which there may be poor substitutes in the rental market. Much of our analysis focuses on second home purchases and swaps into more expensive homes on the premise that the timing of their purchases may better capture beliefs about housing markets than the timing of first-home purchases, which may be driven by the life-cycle. Even second home purchases, however, may contain a consumption motive, as homes in general may be "status," or Veblen, goods. We test whether securitization agents perceived their high current income during the housing boom was transitory by examining whether the value of their purchases was conservative relative to their current income, a test premised on the idea that home transactions trace out beliefs about expected permanent income jointly with beliefs about future home prices. Using stated incomes from mortgage application data, we find little evidence that their purchases were more conservative than that of control groups, although this data is noisy. We also find that homes purchased in 2004-2006 were aggressively sold in 2007-2009, relative to both control groups, suggesting that securitization agents overestimated the persistence of their incomes.

Our analysis complements the large literature in the aftermath of the crisis studying whether poorly designed incentives led Wall Street to take excessive risks in the housing market, leading to disastrous consequences.[3] The literature has accumulated ample evidence that the practice of securitizing mortgages in the originate-to-distribute model contributed towards lax screening of subprime borrowers (Agarwal and Ben-David, 2012; Berndt and Gupta, 2009; Demyanyk and Van Hemert, 2011; Jiang, Nelson and Vytlacil, 2011; Keys et al., 2009, 2010, 2012; Mian and Sufi, 2009; Piskorski, Seru and Witkin, 2013; Purnanandam, 2011; Rajan, Seru and Vig, 2012). It is important to note that the roles played by beliefs and incentives are not mutually exclusive (Cole, Kanz, and Klapper, 2012), and in fact are very much related in that distorted beliefs about overall housing markets and bad incentives to lend to unqualified borrowers are two forces which may interact and reinforce each other. For example, any weakened incentives to screen subprime borrowers and securitize mortgage loans pooled across the country would be

[3] The key friction in this narrative is that agents on Wall Street did not have incentives appropriately aligned with outside stakeholders such as shareholders (see, e.g., the debate in Bebchuk, Cohen, and Spamann, 2010; Bhagat and Bolton, 2011; Fahlenbrach and Stulz, 2011), or other stakeholders such as creditors, taxpayers, and society at large (Acharya, et al. 2010; Bolton, Mehran and Shapiro, 2011; Edmans and Liu, 2011; Rajan, 2006, 2010).

exacerbated if lenders and security originators were buoyed by expectations that prices in overall house markets would never fall.

We caution that our analysis does not isolate the beliefs of securitization agents about subprime housing markets, as they are not subprime borrowers themselves. Furthermore, although we employ indirect tests to examine whether securitization agents' home purchases were hedged, we do not observe the entire household balance sheet and thus cannot rule out that they mitigated their overall housing exposure through other means such as shorting housing stocks. Overall, however, our analysis presents evidence that is inconsistent with systematic awareness of broad-based problems in housing among mid-level managers in securitized finance. Our analysis does leave open the possibility that even by rationally processing all available information to securitization agents, one might not have been able to identify the housing bubble. Nevertheless, the aggressiveness of certain groups in increasing their home exposure suggests a role for belief distortions.

2.1 Empirical Hypothesis

The aim of our analysis is to examine whether Wall Street employees anticipated a broad-based housing bubble and crash. Figure 2.1 depicts the Case-Shiller house price indices for the composite-20 metropolitan areas as well as New York, Chicago, and Los Angeles from 2000-2011. Of these areas, Los Angeles had the most dramatic boom and bust cycle, with house prices increasing by over 170% from 2000 to a peak in 2006 and then crashing down by over 40% from the end of 2006 through the end of 2011. New York also experienced a boom/bust cycle, with prices increasing by over 110% from 2000-2006 and then dropping by over 20% through 2011. Over the composite 20 metropolitan areas, prices rose by 100% from 2000-2006 and fell by over 30% through 2011. Despite the differences in magnitudes, the cycles across different regions experienced rapid price expansions in 2004-2006, which we define as a bubble period in our analysis, the beginning of a decline in 2007, followed by steep falls in 2008.

The practice of securitizing mortgages has been widely recognized as one of the important enablers in the development of the housing bubble. As such, we focus on understanding the beliefs of mid-level managers in the securitization business across these boom and bust periods, whom we collectively refer to as securitization agents. In practice, our mid-level managers are mostly Vice Presidents, Senior Vice Presidents, and Managing Directors at investment banks, commercial banks, hedge funds, mortgage lenders, and other financial companies. These managers buy and sell tranches of securitized mortgages and are largely responsible for understanding the pricing of these instruments and the correlation of the underlying securities.

There are several reasons to analyze the beliefs of mid-level managers rather than C-level executives. First, they made many important business decisions for their firms. The 2012 "London Whale" risk-management failure of JP Morgan Chase illustrates that, if anything, CEO Jamie Dimon realized relatively late that traders had accumulated significant exposure to specific CDS positions which subsequently resulted in outsized losses. Second, mid-level managers were very close to the housing markets. There is a growing notion that perhaps those outside of the top-level C-suite – for example, Joseph Cassano of AIG Financial Products, or Fabrice Tourre of Goldman Sachs – knew about the problems in the housing markets even if C-level executives did not.

We use a revealed belief approach based on people's personal home transactions.[4] A home is typically a significant portion of a household's balance sheet. As our data will confirm later, this is likely true even for the mid-level securitization agents in our sample. To the extent that homeowners have thick skin in their homes, they have maximum incentives to acquire information and make informed buying and selling decisions. This is a key feature that allows us to isolate their beliefs, separately from any distortionary effects related to job incentives.[5]

Our general strategy focuses on testing whether securitization agents were more aware of the imminent housing market crash compared to plausibly unaware counterfactual control groups. This strategy relies on the cross-sectional variation in home purchase and sale behavior across these groups during the boom and bust periods. We have four primary tests. We first test for awareness in a strong "market timing" form. Under this strong form, securitization agents knew about the bubble so well that they were able to time the housing markets better than others. This implies that securitization agents who were homeowners anticipated the house price crash in 2007-2009 and reduced their exposures to housing markets by either divesting homes or downsizing homes in the bubble period of 2004-2006.

[4] Home transactions are more informative of individuals' beliefs than buying and selling of their companies' stocks, which is contaminated by potential effects from loyalty (Cohen, 2009).

[5] A subtle issue for our analysis is that poorly designed incentives can distort beliefs among agents (Cole, Kanz, and Klapper, 2012). Our analysis is informative about this hypothesis in the following way. If agents exhibited beliefs consistent with awareness of the bubble, this would be inconsistent with the hypothesis of this interaction, as their beliefs would be aligned with their presumably bad incentives. Evidence of unawareness would be consistent with this interaction, with the cause of unawareness being poorly designed incentives. However, our tests do not distinguish between specific reasons for unawareness.

Market timing is a strong form of awareness for two reasons. First, the cost of moving out of one's home, especially the primary residence, is high, and may prevent securitization agents from actively timing the house price crash. Second, even if securitization agents knew about the presence of a housing bubble, they might not be able to precisely time the crash of house prices. While these caveats reduce the power of using the securitization agents' home divestiture behavior to detect their awareness of the bubble, it is useful to note that the cost of moving out of second homes is relatively low and should not prevent the securitization agents from divesting their second homes.

More importantly, the cost of moving and inability to time the crash should not prevent securitization agents from avoiding home purchases if they were indeed aware of problems in housing. This consideration motivates our second empirical test for a weaker, "cautious" form of awareness, which posits that securitization agents knew enough to avoid increasing their housing exposure during the bubble period of 2004-2006. We focus on purchases of second homes and swaps of existing homes into more expensive homes instead of first home purchases, since the timing of first home purchases is arguably motivated by necessary consumption related to the life-cycle rather than beliefs about housing.

Our third test focuses on the net trading performance to see whether securitization agents' observed transactions as a whole improved or hurt their financial performance. We benchmark their observed strategy against a static buy-and-hold strategy and compare whether securitization agents' portfolios did better against their benchmark than control groups' portfolios during the 2000-2010 period. This test sheds light on whether agents exhibited awareness through more complex strategies. For example, agents could have tried to "ride the bubble" by buying as prices rose and selling near the peak (Brunnermeier and Nagel, 2004). Although our test only spans one boom and bust, the substantial cross-sectional heterogeneity in securitization agents allows us to test whether the group, on average, anticipated this particular housing crash.

Although our analysis of cautiousness removes first home purchases, even purchases of second homes and swaps into expensive homes are plausibly related to a consumption motive, particularly as homes in general may be "status," or Veblen, goods. This motivates our fourth test, which is based on the idea that securitization agents' home transactions jointly trace out beliefs about permanent income and house prices, as their human capital is tied to housing. Awareness of an impending large-scale housing crash would have led either to reduced expectations of permanent income or awareness that bonuses during the boom were transitory, and thus to less expensive purchases relative to their current incomes. We test whether the value-to-income ratio of securitization agents' purchases during the boom fell, relative to unaware control groups, where current income is in the denominator.[6] We also test whether homes purchased by securitization agents during the boom were held for significant periods of time. If fully-aware securitization agents purchased homes during the boom for consumption, these homes should be held for significant periods of time (Sinai and Souleles, 2005), or else a significant discount rate would be required to justify these purchases.

Economic determinants of home transaction behavior other than beliefs could drive cross-sectional differences between securitization agents and potential control groups. First, the level of risk aversion may vary, particularly if the age profile varies across career groups. Second, there may be career selection and life cycle effects. Different careers may have different optimal points of purchasing housing not obtainable in the rental market due to career risk and different life cycle patterns in when to have children. Third, heterogeneity in wealth levels and income shocks may drive home purchase behavior. Less wealthy people may be less likely to purchase a home due to credit constraints, and credit constrained agents may be more likely to purchase a home after a positive income shock.

To address these issues, we construct two uninformed control groups. The first group is a sample of equity analysts covering S&P 500 companies in 2006, excluding major homebuilders. The assumption is that, being a self-selected group of agents who work for similar finance companies, they face similar ex ante career risks and have similar risk aversion and life cycle profiles. They also received some forms of income shocks during the housing boom, as finance companies generally performed very well over this period. We also construct a second control group comprised of lawyers practicing outside of real estate law. Although differences or similarities between these two groups may be less ascribable to beliefs due to heterogeneity, this exercise tests for awareness among securitization agents relative to a benchmark group of wealthy, high-income people in the general population.

[6] For example, large bonuses during the boom may have led securitization agents to purchase homes, particularly if they were previously credit constrained (Yao and Zhang, 2005; Cocco, 2005; Ortalo-Magne and Rady, 2006), although awareness that bonuses were transitory would lead to more conservative purchases relative to their current income.

Another worry is that awareness of problems in housing markets may not have manifested itself as cautious or market-timing behavior if securitization agents were more pessimistic about tail risk probabilities (rather than the conditional expectation of prices) and had laid off most of the tail risk on lenders. This narrative is particularly relevant for the public debate as many CDOs were constructed and priced as if there was little tail risk (Coval, Jurek and Stafford, 2009). We investigate this issue by further examining the loan-to-value of purchases of securitization agents versus those of control groups, as well as whether they purchased with more intensity in non-recourse states to minimize the consequences of any potential defaults. We also examine whether sell-side securitization agents' housing portfolios outperformed that of buy-side agents' to address the role of this narrative in the public debate.

Taken together, we test the following hypothesis regarding whether securitization agents were aware of the housing bubble:

Hypothesis (Full Awareness): Securitization agents exhibited more awareness of a broad-based housing bubble relative to equity analysts and lawyers in four possible forms:

A. *(market timing form) Securitization agents were more likely to divest homes and down-size homes in 2004-2006.*
B. *(cautious form) Securitization agents were less likely to acquire second homes or move into more expensive homes in 2004-2006.*
C. (*performance) Overall, securitization agents had better performance after controlling for their initial holdings of homes at the beginning of 2000.*
D. *(conservative consumption) Relative to their current income, any purchases made by securitization agents during the boom were more conservative.*

We emphasize that securitization agents are likely not subprime borrowers themselves, and thus our analysis does not isolate the beliefs of securitization agents about subprime mortgage markets. A limitation of our analysis is that we do not observe the entire household balance sheet, and thus do not see whether they took other steps to short the housing market. Several considerations are reassuring. First, directly shorting the housing market is notoriously difficult. Shiller (2008) documents that repeated attempts to create markets to hedge house price risk have failed to attract liquidity, pointing out that the "near absence of derivatives markets for real estate…is a striking anomaly." Second, shorting homebuilding stocks and real estate investment trusts also leaves substantial basis risk with any home investments. Third, shorting stocks requires borrowing stocks, which is costly, and exposes short sellers to recall risk (D'Avolio, 2002) and the risk of running out of capital if prices appreciate (Mitchell, Pulvino, and Stafford, 2002).

2.2 Data and Empirical Framework

2.2.1 Data collection

We begin by collecting names of people working in the securitization business as of 2006. To do so, we obtain the list of registrants at the 2006 American Securitization Forum's (ASF) securitization industry conference, hosted that year in Las Vegas, Nevada, from January 29, 2006 through February 1, 2006. This list is publicly available via the ASF website.[7] The ASF is the major industry trade group focusing on securitization. It published an industry journal and has hosted the "ASF 20XX" conference every year since 2004. The conference in 2006 featured 1760 registered attendees and over 30 lead sponsors, ranging from every major US investment bank (e.g., Goldman Sachs) to large commercial banks such as Wells Fargo, to international investment banks such as UBS, to monoline insurance companies such as MBIA.

We construct a sample of 400 securitization agents by randomly sampling names from the conference registration list and collecting their information from our data sources until we have 400 agents with data. We make sure to oversample people at the most prominent institutions associated with the financial crisis by attempting to collect information for all people associated with those firms.[8] We screen out people who work for credit card, student loan, auto, and other finance companies primarily involved in the non-mortgage securitization business, and also use any available information in LinkedIn to screen out people working in non-mortgage securitization segments of diversified financial firms. We also use LinkedIn to collect any background information about each person that will be helpful in locating them within the Lexis/Nexis database. Lexis/Nexis aggregates information available from public records, such as deed transfers, property tax assessment records, and other public address records to person-level reports and provides detailed information about property transactions for each person.

There are a number of reasons that a person we selected from the registration list may not appear in our final sample, as described in Table 2.1, Panel A and also Appendix A in more

[7] As of this writing, this list is no longer available on the web. The authors have copies of the webpages available.
[8] The companies we oversample are AIG, Bank of America, Bear Stearns, Citigroup, Countrywide, JP Morgan Chase, Lehman Brothers, Merrill Lynch, Morgan Stanley, Washington Mutual, Wachovia, Barclays, Deutsche Bank, HSBC, UBS, Credit Suisse, and Mellon Bank.

detail. Chief among these are that they worked in the securitization business but in a non-housing segment such as credit card loans, or that they have a very common name that cannot be uniquely identified in Lexis-Nexis. All told, we sample 613 names to obtain 400 securitization agents in sample.

For each person in our sample, we collect data for all properties ever owned, including the location, the date the property was bought and sold, the transaction price, and mortgage terms, when available.[9] Lexis/Nexis contains records for individuals who never own property, since it also tracks other public records, and we record these individuals as not having ever owned property. We also collect data about any refinances undertaken during the sample period. Our data collection began in May 2011 and we thus have all transactions for all people we collect through this date. Our analysis focuses on the period 2000-2010, the last full year we have data.[10]

Our sample of equity analysts consists of analysts who covered companies during 2006 that were members of the S&P 500 anytime in 2006, excluding homebuilding companies. These people worked in the finance industry but were less directly exposed to housing, where the securitization market was most active. We download the names of analysts covering any company in the S&P 500 during 2006 outside of SIC codes 152, 153 and 154 from I/B/E/S. These SIC codes correspond to homebuilding companies such as Toll Brothers, DR Horton, and Pulte Homes.[11] There are 2,978 analysts, from which we randomly sample 469 names to obtain 400 equity analysts with information in our sample.

To construct our sample of lawyers, we select a set of lawyers for each person in our securitization sample from the *Martindale-Hubbell Law Directory*, an annual national directory of lawyers which has been published since 1868, matched on age and the work location of the lawyer. We provide details in Appendix A. This matching is not available for equity analysts given the information we have available ex ante in our sampling. We have 406 total names that

---

[9] If we do not find a record of a person selling a given property, we verify that the person still owns the property through the property tax assessment records. In cases where the property tax assessment indicates the house has been sold to a new owner, or if the deed record does not contain a transaction price, we use the sale date and sale price from the property tax assessment, when available.

[10] We collect data for all transactions we observe, even if they are after 2010. This mitigates any bias associated with misclassifying the purpose of transactions, as we discuss below. To ease data collection requirements, we skip properties sold well before 2000, as they are never owned during the 2000-2010 period and are thus immaterial for our analysis.

[11] Our references for SIC codes is CRSP, so a company needs to have a valid CRSP-I/B/E/S link.

we search for within Lexis/Nexis to obtain 400 lawyers matched on age and location to our securitization sample.[12]

2.2.2 Classifying home purchases and sales

Our starting point for understanding home purchase behavior is a broad framework to categorize the purpose of a transaction for a given person. We think of person *i* at any time *t* as either being a current homeowner, or not. If she is not a current homeowner, she may purchase a house and become a homeowner (which we refer to generically as "buying a first home"). Note that one may have been a homeowner at some point in history and still "buy a first home" if one is currently not a homeowner. If a person is currently a homeowner, she may do one of the following:

A) Purchase an additional house ("buy a second home"),
B) Sell a house and buy a more expensive house ("swap up"),
C) Sell a house and buy a less expensive house ("swap down"),
D) Divest a home but remain a homeowner ("divest a second home"),
E) Divest a home and not remain a homeowner ("divest last home").

To operationalize this classification of transactions, we define a pair of purchase and sale transactions by the same person within a six month period as a swap, either a swap up or a swap down based on the purchase and sale prices of the properties. If either the purchase or sale price is missing, we classify the swap generically as a "swap with no price information."

The purchases that are not swaps are either non-homeowners buying first homes, or homeowners buying second homes.[13] We use the term "second" to mean any home in addition to the person's existing home(s). Divestitures are classified similarly: among sales that are not involved in swaps, if a person sells a home and still owns at least one home, we say she is divesting a second home; if she has no home remaining, we say the person is divesting her last home.[14]

2.2.3 Transaction intensities

[12] The success rate for collecting information about lawyers is much higher because the Martindale-Hubbell Law Directory provides detailed information about each lawyer, allowing us to pinpoint the name in Lexis-Nexis more easily than other groups.
[13] If a home is on record for an individual, but the home does not have a purchase date, we assume the owner had the home at the beginning of our sample, January 2000. We provide more details of our classification in Appendix A.
[14] When classifying transactions in 2010, we use information collected on purchases and sales in 2011 to avoid over-classifying divestitures and first-home/second-home purchases and underclassifying swaps in the final year of data.

Our main analysis centers on the annual intensity of each transaction type – that is, the number of transactions per person per time period.[15] We focus on an annual frequency to avoid time periods with no transactions. Formally, the intensity of one type of transaction in year t in a sample group is defined as the number of transactions of that type in year t divided by the number of people eligible to make that type of transaction at the beginning of year t:

$$Intensity_t = \frac{\#\, Transactions_t}{\#\, people\ eligible\ for\ the\ transaction_t}.$$

For example, the intensity of buying a first home is determined by the number of first home purchases during the year divided by the number of non-homeowners at the beginning of the year. An important feature of our data is that we observe not only transaction activity but also transaction *inactivity*, due to the comprehensiveness of the public records tracked by Lexis/Nexis. This allows us to test the hypothesis that one group was more cautious (i.e., bought less) than other groups, as we can normalize the number of transactions by the total number of people who could have made that transaction, instead of the number of people who made the transaction.[16]

2.2.4 Income data

We are able to observe income in the purchase year of a home for a subset of people by matching information we observe about the year of their purchase, their mortgage amount, and property location with the information provided in the 2000-2010 Home Mortgage Disclosure Act (HMDA) mortgage application data. The HMDA dataset contains information on the income relied on by the originating institution to underwrite the loan. Although most identifying information – such as the borrower's name, exact date of origination, property address and zip code – is not provided, the data provides the mortgage amount (up to the thousands) as well as the census tract of the property. We match purchases with all originated mortgages in HMDA of the same amount in the purchase year with the same census tract as the property. If we

[15] We focus on the intensity of transactions rather than the probability of an eligible person making a given transaction because the latter discards information about a person making multiple transactions of one type in one year. However, focusing instead on probabilities yields nearly identical results.

[16] A complication in this calculation is that, in a given year, a person may make multiple transactions. As a result, the number of non-homeowners at the beginning of the year does not fully represent the number of people eligible for buying a first home during the year, because, for instance, a homeowner may sell her home in February and then buy another home in September. To account for such possibilities, we define "adjusted non-homeowners," who are eligible for buying a first home during a year, to be the group of non-homeowners at the beginning of the year plus individuals who divest their last homes in the first half of the year. We similarly adjust the number of homeowners and multiple homeowners. Appendix A contains a detailed description of adjustments.

successfully find a match, we take the stated income on the HMDA application as the income of our person at the time the purchase was made.[17]

2.3 Descriptive Statistics

Table 2.1, Panel A presents the number of people in each sample. Our groups of interest each have 400 people by construction. Panel B presents the age distribution for each group. The median ages in 2011 for the securitization agent, equity analyst, and lawyer samples are 45, 44, and 46, respectively. Chi-square tests of homogeneity fail to reject the hypothesis that the distributions presented in Panel B are the same.

Our sample features people from 176 distinct firms, of which we are able to match 65 as publicly traded companies in CRSP. Our sample is tilted towards people working at major firms due to our oversampling of those firms. The most prominent companies in our sample are Wells Fargo (27 people), Washington Mutual (23), Citigroup (16), JP Morgan Chase (14), AIG (12), Countrywide, Deutsche Bank, Merrill Lynch, UBS, and Lehman Brothers (9 each). The most common position titles are Vice President (87), Senior or Executive Vice President (58), and Managing Director (39). In addition to the large firms, a number of regional lenders such as BB&T, smaller mortgage originators such as Fremont General and Thornburg Mortgage, and buy-side investors such as hedge funds and investment firms are present as well. Additional details about the people in our securitization sample are provided in Table B1 in Appendix B.[18]

Table 2.2, Panel A breaks down the number of properties owned over 2000-2010. Our data spans 674 properties owned by securitization agents during the 2000-2010 period, 604 by equity analysts, and 609 by lawyers. Of these, the majority were bought during the same period, while roughly 40% of total properties were sold during this period.[19]

---

[17] One concern is that, even given an exact mortgage amount (e.g., $300K), census tract, and purchase year, there may be multiple matches within HMDA. The average number of matches per purchase is roughly three, and the median match is unique. Given the economically-motivated construction of census tracts, we average income over all matches in HMDA as the income for that purchase. One can repeat the analysis using only unique matches, which reduces our sample by slightly less than half, and obtain qualitatively similar results that are more influenced by a small number of observations at the tail ends of the distribution.

[18] Our reading suggests that many of these agents were involved in forecasting, modeling, and pricing cash flows of mortgage-backed paper. As an example, one person in our group lists their job title in LinkedIn as "Mortgage Backed Securities Trader, Wells Fargo," with job responsibilities including "Head of asset-backed trading group for nonprime mortgage and home equity mortgage products," "Built a team of 3 traders with responsibility for all aspects of secondary marketing of these products, including setting pricing levels, monthly mark-to-market of outstanding pipeline/warehouse, and all asset sales."

[19] There are a substantial number of properties with either no sale date or a sale date after December 31, 2010; these are homes that were still owned as of that date.

Appendix B maps the geographical distribution of properties in our sample. The New York combined statistical area (roughly the NJ-NY-CT tri-state metro area plus Pike County, PA) is the most prominent metro area, followed by Southern California (Los Angeles plus San Diego). Both equity analysts and securitization agents are concentrated in New York, with a slightly higher concentration for equity analysts. Appendix B also contains details about how purchases and sales are distributed through time, and how these purchases and sales were classified.

Table 2.2, Panel B summarizes mortgage information. For the securitization sample, we have mortgage information for 327 purchases out of 437 we observe from 2000-2010. Of these, we are able to match 253 to HMDA, a conditional success rate of 77%; for both the equity analyst and lawyer groups, this rate is 79%. Over the entire 2000-2010 period, the average income at purchase was $350K for the securitization sample, $409K for the equity analyst sample, and $191K for the lawyers. All income figures are reported in real December 2006 dollars adjusted using the Consumer Price Index (CPI) All Items series.

One concern is that these numbers appear a bit too "small" relative to what is commonly perceived as finance industry pay. The income reported in HMDA represents income used by the bank to underwrite the loan, which may often include only taxable income provided by the mortgage applicant and is thus likely downward biased. Forms of compensation not taxable during the year, such as employee stock option grants, may not be included.[20]

Even if this reporting issue were not present, observed income levels are not unbiased representations of the true distribution of underlying income because we only observe income at purchase, and not income in other years (nor for non-purchasers). As a descriptive exercise, however, Table 2.3 breaks down average income observed at purchase into three bins, corresponding to the pre-housing boom (2000-2003), housing boom (2004-2006), and housing bust (2007-2010).[21] Our securitization agents received income shocks from the pre-boom to the boom period, with average income rising by $92K, over 37% of average pre-boom income. Equity analysts also received income shocks, with average income at purchase rising by $57K, although this is a smaller fraction of pre-boom income, 16%. These results are roughly

[20] If the amount of underreporting varies across time, the bias becomes problematic for our analysis comparing average value-to-income ratios at purchase across groups and time. We discuss this in Section 4.4.

[21] Because we are interested in average income per person, we first average within person over purchases to obtain a person-level average income for the period before averaging over people in each period.

consistent with our initial hypothesis that the two finance industry groups received positive income shocks, although securitization agents received a slightly larger average shock.

2.4 Empirical Results

2.4.1 Were securitization agents more aware of the bubble?

We first examine whether securitization agents divested houses in advance of the housing crash. Figure 2.2, Panel A plots the divestitures per person per year for each group through time. The divestiture intensities for the securitization agent sample are, if anything, lower than those of equity analysts and lawyers in years before 2007. Compared to equity analysts, the divestiture intensity for securitization agents is lower every year from 2003-2006, and slightly higher during the bust period, 2007-2009.[22]

To account for heterogeneity in the age and multi-homeownership profiles of each group, we compute regression-adjusted differences in intensities. We do this by constructing a strongly-balanced person-year panel that tracks the number of divestitures each year for each person, including zero if no divestiture was observed. We then estimate the following equation for each pairing of the securitization group with a control group using OLS:

$$E[\#Divestitures_{it}|HO_{it-1} = 1]$$
$$= \alpha_t + \beta_t \times Securitization_i + \sum_{j=1}^{7} \delta_j \, Age_j(i,t) + \lambda \, MultiHO_{it-1}. \quad (1)$$

The variable $\#Divestitures_{it}$ is the number of divestitures for individual i in year *t*, $Securitization_i$ is an indicator for whether individual i is part of our securitization agent sample, $Age_j(i,t)$ is an indicator for whether individual *i* is part of age group *j* in year *t* (where eight age brackets are defined according to Table 2.1, Panel B, and one age group is excluded), $MultiHO_{it-1}$ represents whether individual *i* was also a multi-homeowner at the end of year *t-1*, and HO_{it-1} is an indicator for whether individual i was a homeowner at the end of year *t-1*. We use indicators for age brackets instead of a polynomial specification for age as it makes coefficients easily interpretable as average group effects. In each year *t*, we condition the sample such that only the adjusted homeowners as of the end of year *t-1* (i.e., those who started year *t* as

[22] The raw number of divestitures each year may be read off by multiplying the intensity in a given year from Table 2.4 by the number of homeowners in that year given by Table B5 in Appendix B. For example, in 2008, there were 19 divestitures (0.061 times 313) in the securitization sample. In contrast to our regression-adjusted differences, we do not condition on having age information when reporting these raw intensities.

homeowners or became a homeowner during year t, so that $HO_{it-1} = 1$) are included in the estimation. We cluster standard errors by person. The effective sample size is the number of homeowners during the 2000-2010 period.[23]

The coefficients β_t are the difference in average annual divestitures per person within the homeowner category across samples, adjusted for average group effects captured by age and multi-homeownership indicators, and are our coefficients of interest, with $\beta_t > 0$ during the 2004-2006 period suggesting evidence of market timing.[24] Table 2.4 presents these regression-adjusted differences. Consistent with the raw divestiture intensities, these differences are very small during the boom period; point estimates are negative compared to equity analysts. There is weak evidence that securitization agents had a slightly higher intensity of divestiture in 2007 and 2008. This could be consistent with a form of market timing such as riding the bubble, but also consistent with divestitures related to job losses, a point which we return to in Section 4.2.7. Overall, however, there is little evidence that suggests people in our securitization agent sample sold homes more aggressively prior to the peak of the housing bubble relative to either equity analysts or lawyers.

We next examine whether securitization agents were cautious in purchasing homes in 2004-2006. This cautiousness alternative emphasizes that securitization agents knew about the bubble, but that the optimal response was to avoid purchasing homes given the difficulty in timing the crash. We focus on second home purchases and swap-ups into more expensive houses by homeowners. Results for first-home purchases by non-homeowners are reported in Appendix B

---

[23] The effective sample size (number of people contributing to the variation) of this estimation will be the total number of people who we ever observed as adjusted homeowners during the 2000-2010 period for whom we have age information across these two groups. This may be read off from the last row of Table B5, Panel B. For example, when estimating equation (1) for the securitization sample and the equity analyst sample, the number of people will be 633 (328 plus 305). The number of homeowners contributing to the variation each year may similarly be read off from the same table, which lists the number of homeowners and non-homeowners each year with age information. For example, when estimating (1) for the securitization agent and equity sample, the number of people contributing variation in 2000 is 415 (220 plus 195).

[24] We estimate equation (1) using OLS to maintain the simplest interpretation of β_t. In the absence of covariates, β_t would be equivalent to average marginal effects estimated from non-linear limited dependent variable models (e.g., a Poisson model), because $Securitization_i$ is binary. One alternative method of estimating equation (1) would be to replace the left-hand side variable with an indicator for whether a person divested and interpret β_t as the probability of a person divesting, making equation (1) a linear probability model and the corresponding non-linear model a logit or probit model. Appendix B reports results average marginal effects estimated from such a logit model, analogous to results in Tables 4 and 5. Results are nearly identical. Angrist and Pischke (2009) discuss the relative merits of OLS as a robust approximation (in the minimum mean-squared error sense) to the conditional expectation function versus these non-linear methods.

and do not reveal significant differences; if anything, there are more first home purchases by non-homeowning securitization agents than equity analysts.

Figure 2.2, Panel B plots the raw intensity of second home purchases and swap-ups through time, while Table 2.5 presents regression-adjusted differences. The regression-adjusted differences are computed using a specification analogous to equation (1) where we replace the left-hand side variable with the number of second home purchases plus swap-up transactions for individual *i* during year *t*, again conditioning the sample to adjusted homeowners as of the end of year *t-1*. Contrary to what would be suggested by the full awareness hypothesis, we observe $\beta_t > 0$ consistently throughout the 2004-2006 period, with statistically significant differences with the equity analyst group at the 1% level in the 2005 period. Pooling intensities every other year reveals positive and statistically significant differences in the 2002-2003, 2004-2005, and 2006-2007 periods (reported in Appendix B). Economically, the intensity of second home purchase and swap-up activity was 0.07 homes per person higher in 2005 for securitization agents than equity analysts. This suggests that securitization agents were aggressively increasing, not decreasing, their exposure to housing during this period. We now explore this issue in more detail.

2.4.2 Second home and swap-up purchases

2.4.2.1 Firm-specific effects. We exploit the fact that we observe 78 securitization agents and 136 equity analysts working at a common set of 19 firms to remove company-specific effects. For this test and for other subsample tests, we pool together intensities every other year (2000-2001, 2002-2003, and so forth) to mitigate the concern that our results are driven by spurious differences between a small number of transactions we may observe during a single year when we condition the sample tightly. We estimate the following equation:

$$E[\#BuySecondOrSwapUp_{it}|HO_{it-1} = 1] \tag{2}$$

$$= \gamma_k + \alpha_{s(t)} + \beta_{s(t)} \times Securitization_i + \sum_{j=1}^{7} \delta_j \, Age_j(i,t) + \lambda \, MultiHO_{it-1},$$

where γ_k represents company-specific effects and $s(t) = 0$ if t=2000 or 2001, $s(t) = 1$ if t=2002 or 2003, and so forth. The first column of Table 2.6 reports the results and shows that, within this subsample, purchase intensities for second homes and swap-ups are higher for securitization agents in the 2002-2003 and 2006-2007 periods, even controlling for firm effects.

2.4.2.2 Location effects. Heterogeneity in property locations is a concern, since the magnitude of the housing bubble was very heterogeneous across areas, as shown previously in Figure 2.1. Although our sample of lawyers is location matched with our securitization agents, equity analysts are relatively more concentrated in the New York metro area. If securitization agents lived in areas where it was cheaper or easier to purchase a second home or swap up, this location effect may drive our previous results. To check whether this is the case, we condition the sample of homeowners each year to those who own property in the New York metro region at the end of the previous year, and estimate the following model:

$$E[\#BuySecondOrSwapUp_{it} | HO_{it-1} = 1, PropNYC_{it-1} = 1]$$
$$= \alpha_{s(t)} + \beta_{s(t)} \times Securitization_i + \sum_{j=1}^{7} \delta_j \, Age_j(i,t) + \lambda \, MultiHO_{it-1}, \qquad (3)$$

where $PropNYC_{it-1}$ is an indicator for whether person *i* owns property in the New York combined statistical area at the end of year *t-1*. Results are reported in Columns 2 and 3 of Table 2.6. We find that, even within this smaller subsample, securitization agents were more aggressive with purchases of second homes and swap-ups in 2004-2005 relative to equity analysts, an effect that is statistically significant at the 5% level. In Columns 4 and 5, we repeat this exercise for people who live in Southern California, our second most represented metro region and find similar behavior results, although the sample size is smaller than in the New York metro area.

2.4.2.3 Differences-in-differences across locations. Comparing columns 2 and 4 of Table 2.6, the difference in intensities between securitization agents and equity analysts is larger in Southern California than New York. Given that Southern California had a much larger boom-bust cycle than New York, this suggests that securitization agents were even less aware of the bubble in areas where the bubble was very pronounced relative to areas where the bubble was less pronounced.

To further test this insight, we focus on the relative difference between securitization agents and equity analysts in Southern California with that in New York by estimating:

$$E[\#BuySecondOrSwapUp_{it} | HO_{it-1} = 1, (PropSoCA_{it-1} = 1 \text{ } or \text{ } PropNYC_{it-1} = 1)]$$
$$= \alpha_{s(t)} + \gamma_{s(t)} PropSoCA_{it-1} + \delta_{s(t)} Securitization_i$$
$$+ \beta_{s(t)} (Securitization_i \times PropSoCA_{it-1}) + \sum_{j=1}^{7} \delta_j \, Age_j(i,t) + \lambda \, MultiHO_{it-1}, \qquad (4)$$

where $PropSoCA_{it-1}$ is an indicator for whether person i owns property in the Southern California region at the end of year *t-1*.[25] We perform this exercise both with the number of second home purchases and swap ups on the left hand side (Column 6 of Table 2.6) as well as just second home purchases (Column 7). The thought experiment is the following. Suppose Southern California begins to look bubbly in the 2004-2005 period, relative to New York. Allowing for differences between the New York and Southern California regions (through the $\gamma_{s(t)}$ coefficients) and between securitization agents and equity analysts (through the $\delta_{s(t)}$ coefficients), do securitization agents in Southern California react more or less cautiously compared to those in New York during that time period? Evidence of $\beta_{s(t)} < 0$ during 2004-2005 would suggest that securitization agents living in areas which experienced larger boom/bust cycles were more alerted than their counterparts in regions with more moderate cycles.[26]

In fact, the aggressiveness of securitization agents relative to equity analysts is more pronounced in Southern California than in New York. This suggests that securitization agents living in areas which experienced larger boom/bust cycles were potentially even more optimistic about house prices than otherwise. To mitigate the concern that there are relatively fewer equity analysts in Southern California, and to demonstrate that these results are driven by differences across areas, columns 8 and 9 of Table 2.6 estimate only the single-difference between Southern California and New York within securitization agents and shows results consistent with the difference-in-differences.

2.4.2.4 Financing. One concern is that differences in purchase behavior are driven by differential financing terms. Figure 2.3, Panel A plots the average interest rate at purchase for each year and each group. On average, the interest rates observed at purchase between the two groups are very similar and experienced overall time variation similar to that of national benchmark rates.

A second concern is that securitization agents with knowledge of the bubble and crash may have financed the purchase of their homes very differently. Securitization agents may have been aware of tail risks yet laid off these risks to lenders in their purchases. Figure 2.3, Panel B plots

[25] To conservatively avoid an ex ante classification bias in either direction, we discard a handful of observations where people own property in both New York and Southern California at the end of year *t-1*.

[26] There were insufficient observations in the Arizona/Nevada/Florida regions to conduct this type of test. We chose New York and Southern California both because New York experienced a much more moderate bubble than Southern California, but also because of practical considerations given how many observations we have.

the median loan-to-value (LTV) ratio at purchase – including any second lien mortgages recorded on or within 14 days of the purchase – and shows that it holds steady near the unconditional median of 80% throughout the sample period. The median LTV ratios of the marginal second home and swap-up purchases are also very close to 80% through time. Overall, we see little evidence that securitization agents purchased more homes with different financial exposure than equity analysts.

A third financing-related concern is that securitization agents with knowledge of the bubble and crash may have reduced their house price exposure by refinancing and withdrawing equity after purchase during the boom period. Although this would lower the direct financial exposure to home prices, it would significantly increase leverage and the expected cost of bankruptcy, as second lien mortgages – including home equity lines of credit (HELOCs) – are often loans where lenders have recourse to a borrower's non-home assets and where borrowers face personal liability in the event of default. This can be true even in states which are more generally considered non-recourse states due to the protections afforded to mortgages directly connected with purchase money.[27] Consistent with this, Lee, Mayer and Tracy (2012) find that many borrowers continue to pay their second lien mortgages (including HELOCs) even when their first mortgage is in default. An increase in debt through second lien mortgages and HELOCs could thus also be consistent with significant optimism about home prices and a belief that default is not likely. In Appendix B, we show that the average change in debt resulting from refinancing is similar across groups during our sample period.

2.4.2.5 Default risk. To directly examine beliefs about the likelihood of default, we test whether second home and swap-up purchases among securitization agents were differentially concentrated within non-recourse states rather than recourse states relative to equity analysts in Appendix B. Awareness of tail risks would likely lead agents to purchase in non-recourse states as it reduces the expected cost of default. Ghent and Kudlyak (2011) classify states based on lender friendliness and whether it is practical for lenders to obtain deficiency judgments and find that borrowers are substantially more likely to default on first mortgages in non-recourse states, particularly when equity is negative. Conditional on whether a person already has a home in a non-recourse state, we find little evidence of a higher marginal intensity for securitization agents

[27] The most notable example is California. Ghent and Kudlyak (2011) classify California as a non-recourse state. However, California courts have held in numerous cases such as Union Bank v Wendland (1976) that only loans connected to financing the purchase price of a home are protected from deficiency judgments.

to purchase second homes or swap up into more expensive homes in non-recourse states than equity analysts.

2.4.2.6 Type of property. In Appendix B, we provide evidence that, conditional on a second home purchase, the type of home (single-family or condominium) is significantly more likely to be a condominium for securitization agents relative to equity analysts, even though they are no more likely to be farther away. This suggests that they are potentially condominiums purchased to rent rather than for a pure consumption motive.

2.4.2.7 Job switches. The higher number of divestitures in 2007 and 2008 may suggest market timing, with securitization agents divesting homes earlier than others. On the one hand, this difference is small relative to the difference in intensity of second home and swap-up purchases. For example, between securitization agents and equity analysts, the difference in divestiture intensity is 0.026 per homeowner in 2008 while the difference in second home/swap-up intensity is 0.068 per homeowner in 2005. We explore this issue further by using Bayes' rule to decompose the divestiture intensity into the intensity among those who experience job losses (job-losers), the intensity among those who do not experience job losses (no-job-losers), and the rate of job loss.[28] In Appendix B, we provide evidence which suggests that securitization agent job-losers were more likely than equity analyst job-losers to divest a home, despite significant job losses among both groups. In contrast, there is a smaller difference in divestiture intensities between securitization agent no-job-losers and equity analyst no-job-losers. Since both the initial difference in divestiture intensities and the total absolute number of divestitures are small, one caveat to this result is that this decomposition is over a small sample, so that this holds only qualitatively. On the other hand, results for total sales yield statistically significant differences between the two groups of job-losers, while no differences for no-job-losers. Under the full awareness hypothesis, we should have expected to see differences between securitization agents and equity analysts in both job-loser and no-job-loser groups, rather than only in the job-loser group.

2.4.3 Net trading performance

We next systematically analyze which groups fared better during this episode by comparing their trading performance. Our strategy is to compare their portfolio performance based on the

[28] We examine the LinkedIn profiles of each of our securitization agents and define years in which a person switches jobs as the last year of employment within an employer on a person's resume. We provide details in Appendix A.

relative differences in the location and timing of their sales and purchases from the beginning of our sample onwards to see whether trades subsequent to this date helped or hurt each group on average.

Our thought experiment is the following: if agents follow a self-financing strategy from 2000 onwards, where the available investments are houses in different zip codes and a risk-free asset, how did their observed performance compare with that of a hypothetical buy-and-hold strategy? We sketch the assumptions for this exercise here and provide full details in Appendix B. First, we assume time flows quarterly, and we mark the value of each house up or down every quarter from its actual observed purchase price and date in accordance with quarterly zip-code level home price indices from Case-Shiller when possible. Second, we assume that agents each purchase an initial supply of houses at the beginning of 2000 equal to whichever houses they are observed to own at that time. Third, agents have access to a cash account which earns the risk-free rate, and we endow each agent with enough cash to finance the entirety of their future purchases to abstract away from differences in leverage. This last assumption errs on the side of conservatism in isolating performance differences arising from the timing of home purchases.

We compute both the return from the observed strategy and the return from a counterfactual buy-and-hold strategy, where agents purchase their initial set of houses and then subsequently never trade. We denote the difference between the returns of these two strategies as the performance index for each individual, which captures whether trading subsequent to the initial date helped or hurt the individual relative to a simple buy-and-hold strategy.

We test for value-weighted differences in performance by projecting the performance index onto an indicator for the securitization group and indicators for the age categorizations using ordinary least squares in the cross-section of individuals, with sampling weights equal to their initial wealth. Intuitively, this methodology is a value-weighted "difference-in-difference" where the first difference compares observed performance with buy-and-hold performance and the second difference compares this first difference for securitization agents with that of the control group.

Table 2.7, Panel A presents summary statistics for our exercise, while Panel B tabulates the value-weighted average return, buy-and-hold return, and performance index per person for each group, as well as the regression-adjusted differences. Figure 2.4 illustrates the comparative evolution of the performance indices. What is apparent is that all groups, including securitization agents, were worse off at the end of 2010 relative to a buy-and-hold strategy that began in 2000q1.

In fact, the securitization group's portfolio experienced significantly worse gross returns than the equity analyst group, a difference of 4.5% on a regression-adjusted basis. Although part of this is due to a difference in the buy-and-hold return across the two groups (1.7%), the remaining difference of 2.7% quantifies the net trading underperformance of the securitization group, a difference which is statistically significant at the 5% level.[29] In particular, the gross return during the 2007-2010 bust period for the securitization group was particularly poor. Differences with the lawyer group were more modest, although still negative.[30] In summary, the observed trading behavior of securitization agents hurt their portfolio performance.

We also compare portfolios of groups of agents within our securitization group to further isolate the full awareness hypothesis. One salient view is that those who were selling mortgage-backed securities and CDOs knew that the asset fundamentals were worse than their ratings suggested, which suggests that they may have anticipated problems in the wider housing market earlier than others. Table 2.8, Panel A compares the performance of sell-side agents (issuers) with agents from the buy side (investors). Of the 400 securitization agents, 161 work on the sell side and 239 work on the buy side. Evidently, sell-side analysts' portfolios performed more poorly compared to their buy-side peers, with a performance index 6% lower, a difference that is statistically significant at the 5% level.

Table 2.8, Panel B compares the performance of housing portfolios belonging to people working at firms who performed well during the crisis and those who did not. The idea is to test whether people whose firms did poorly anticipated the wider crisis and were able to escape the broad-based fall in home prices themselves. We hand-match our list of companies to CRSP and sort them into terciles of buy-and-hold stock performance from July 2007 through December

---

[29] In interpreting this magnitude, it is worth recalling that our performance evaluation fully collateralizes all purchases and endows agents with large amount of cash, so that this difference likely significantly understates the true difference in portfolio performance across the two groups. For example, housing forms only a 25% portfolio weight for securitization agents in the first period of our calibration, rising to 54% in the last period, both of which are quite conservative. In Appendix B, we present an alternative calibration where we halve the amount of cash given to each agent, so that the initial portfolio weight on housing is 51%, rising to 111% in the final period. This produces larger magnitudes while not affecting statistical inference.

[30] We have also experimented with different initial dates for the performance evaluation. For starting dates between 2000q1 and 2004q4, results are very similar. Differences between the two groups when using a starting date of 2005q4 and 2006q4 manifest mostly in the gross return, since the bulk of homes had been purchased by then.

2008, the period over which a significant portion of the crisis develops. Poorly-performing companies include Lehman Brothers and Countrywide. Better-performing firms include BB&T, Wells Fargo, and Blackrock. The results suggest that the housing portfolios of people working at poorly-performing firms did worse than those of people working at better-performing firms, although the smaller size is smaller. Overall, if fully aware agents were attempting to "ride the bubble," they missed the peak, leading not only to sharply negative returns, but also worse performance relative to other groups.

2.4.4 Consumption and income shocks

We examine the consumption component of housing and whether securitization agents perceived that their high current income during the housing boom was transitory by testing whether the value-to-income (VTI) ratios of their purchases were more conservative than that of control groups. If securitization agents understood that income shocks during the boom were transitory, we should observe lower value-to-income ratios at purchase for them relative to control groups.

We compute the value-to-income (VTI) ratio for the subsample of purchases where we have both income data from HMDA and an observed purchase price.[31] Table 2.9 tabulates the mean and median VTI for each group in pre-boom, boom and bust periods. From pre-boom to boom, the average VTI for purchasers in the securitization sample increased from 3.2 to 3.4; the median showed a slight decrease from 3.1 to 3.0, suggesting there are some purchasers who purchased homes at a very large VTI ratio, even after trimming out those with very low incomes. The average VTI among equity analyst purchasers increased from 2.9 to 3.1, while the median increased from 2.7 to 2.8. Overall, the evidence does not display any strong pattern consistent with the hypothesis that the securitization agents were more conservative in their VTI ratios when purchasing homes.[32]

One caveat to analyzing the value-to-income ratio is that our measures of income are likely downward biased, as noted in Section 3. Because our analysis focuses on comparing the change

[31] Due to the nature of VTI as a ratio, we require a minimum nominal reported income of $100K in the year of purchase to avoid drawing conclusions based on possible extreme tails overly influencing our analysis.
[32] Including the mark-to-market value of other existing homes at the time of purchase, computed using the method described in Section 4.3, to form a portfolio value-to-income ratio at purchase yields similar results, which we report in Appendix B. We focus on the purchase value-to-income ratio to ensure any results or non-results are being driven by the data rather than the additional assumptions required in computing the mark-to-market value of each house.

in value-to-income across groups, the change in VTI will be mis-measured if the bias in underreporting income itself varies across time. In Appendix B, we also examine whether there were differential patterns of selling during the bust across our groups within the subsample of purchases during the 2004-2006 period. We find that, in the prime crisis years (2007 and 2008), sales of 2004-2006 properties per purchaser were much higher for securitization agents than equity analysts and lawyers, making any consumption stream short-lived. As discussed in Section 4.2.7, differences in divestiture and sale intensities during this period are related to a higher intensity among securitization job-losers relative to equity analyst job-losers. This suggests that securitization agents had based earlier purchase decisions on overoptimistic projections of permanent income relative to equity analysts.

2.5 Conclusion

We find little systematic evidence that the average securitization agent exhibited awareness through their home transactions of problems in overall house markets and anticipated a broad-based crash earlier than others. Although we do not observe each household's entire balance sheet, we believe our results provide a useful starting point for understanding the role of beliefs leading to the recent crisis. Other consumption and investment patterns of Wall Street traders may yield additional useful observations on this role. Understanding how incentives interact with beliefs is also one area which might bear substantial fruit. Our evidence that some groups of agents were particularly aggressive in increasing exposure to housing suggests that job environments that foster groupthink, cognitive dissonance, or other sources of over-optimism are of particular concern. Changing the compensation contracts of Wall Street agents alone, for example through increased restricted stock holdings or more shareholder say on pay, may be insufficient to prevent the next financial market crisis (Bolton, Scheinkman and Xiong, 2006; Cheng, Hong and Scheinkman, 2013). The whole financial system may benefit from having securities that incentivize information acquisition about tail-risk states. Given the crucial role of the financial sector in intermediating capital across the economy, systematic analysis of the macroeconomic implications of the belief dynamics of Wall Street employees is needed. We leave these important questions for future research.

Figure 2.1: Home Price Indices

This figure plots the Case-Shiller non-seasonally-adjusted home price indices from January 2000 through July 2012. Values for January 2000 are normalized to 100.

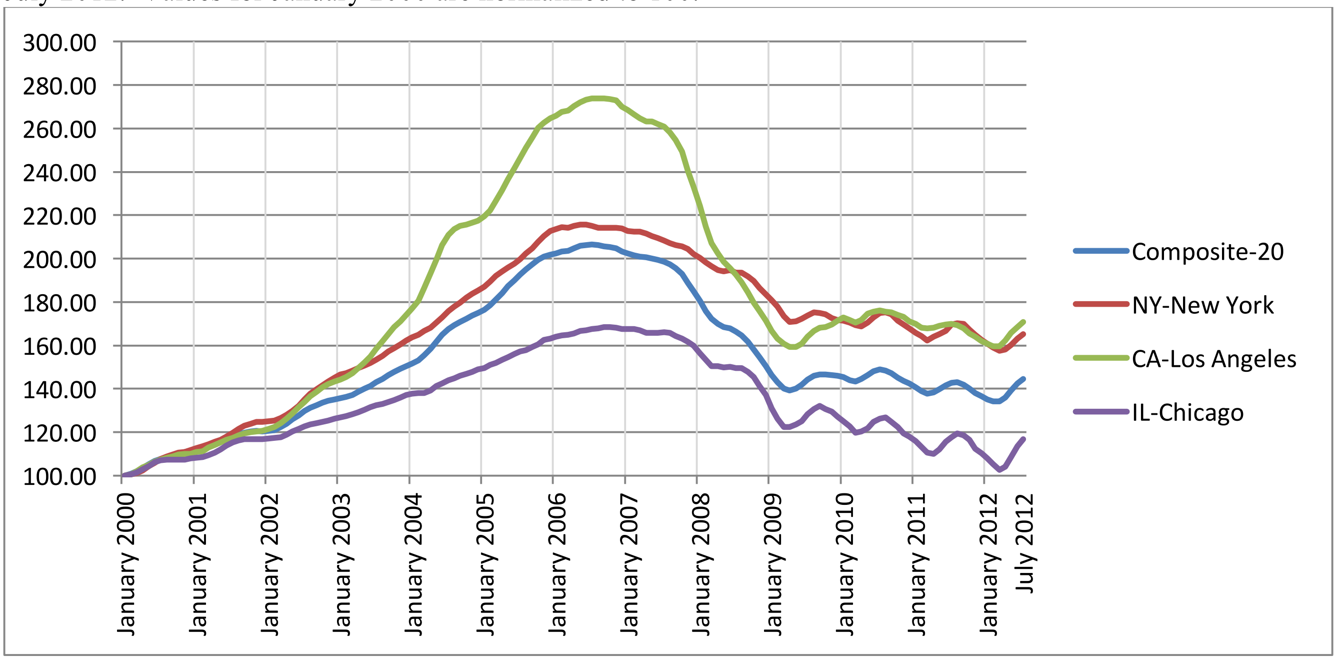


Figure 2.2: Transaction Intensities

Panel A plots the intensity of divestitures through time, defined as the number of divestitures per adjusted homeowner each year, for each group. Panel B plots the intensity of second home purchases and swap-ups.

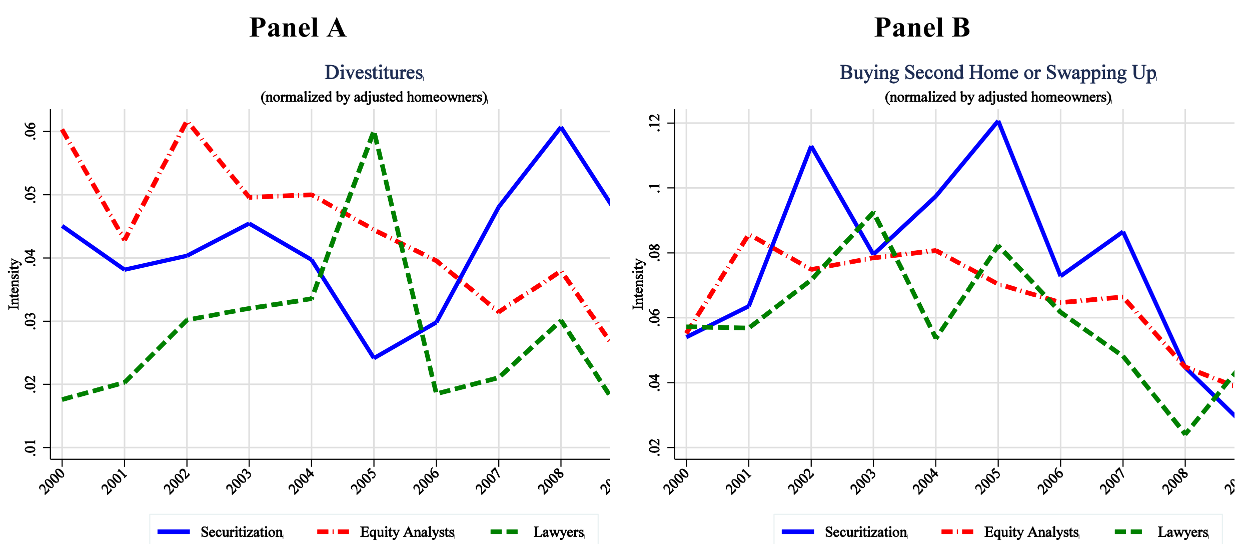
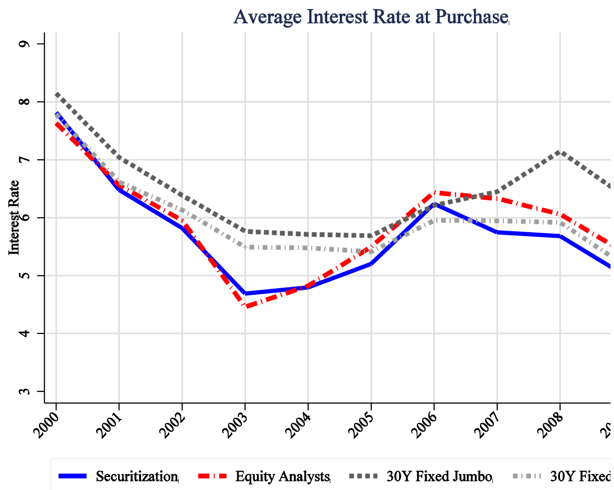


Figure 2.3: Financing

Panel A plots the average interest rate at purchase for securitization and equity analyst groups, as well as average annual national benchmark 30-year jumbo and conforming interest rates from BankRate. Panel B plots the median loan-to-value observed at purchase.

Panel A



Panel B

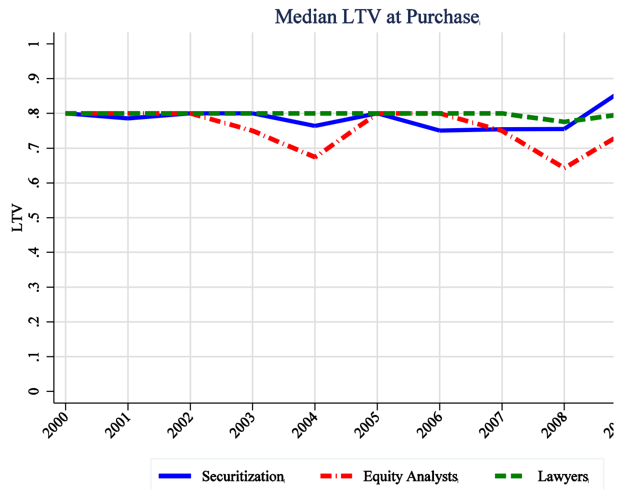


Figure 2.4: Trading Performance Indices

This figure plots the average performance index, defined as the initial-wealth-weighted average difference between the cumulative return on the self-financed trading strategy and the buy-and-hold return of the initial stock of houses, where 2000q1 is taken as the initial quarter, for each group.

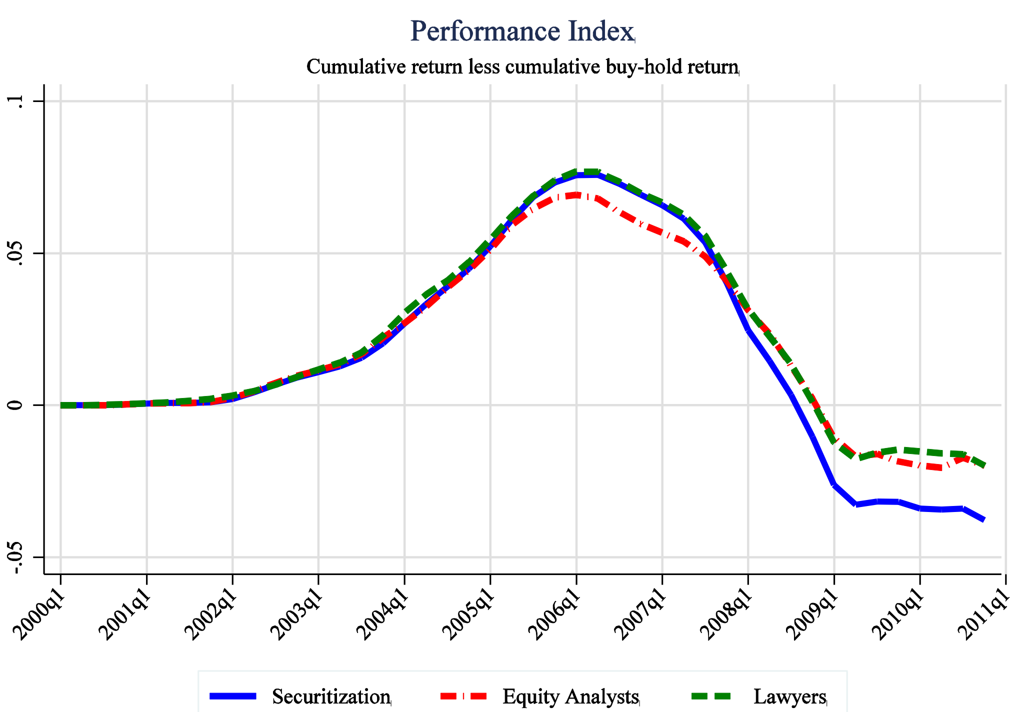


Table 2.1: People

This table lists the number of people for which we gathered information in each of three samples: securitization agents, equity analysts, and lawyers. Panel A tabulates the number of names we searched for and reasons for why a name may not be in our sample. Panel B shows the age distribution of people in our sample.

Panel A: Number of People

Sample	Securitization	Equity Analysts	Lawyers
Number of names	**613**	**469**	**406**
Not mid-level manager	13	N/A	N/A
Not housing	94	N/A	N/A
Not found in public records	29	16	3
Multiple found in public records	50	27	3
International	27	25	0
Deceased	0	1	0
People in sample	**400**	**400**	**400**
Person found, but no homes owned	58	82	42
People who sold all properties before 2000	3	1	0
People who only own homes beginning after 2010	3	4	3
People in sample owning at least one home, 2000-2010	336	313	355
Unconditional rate of homeownership	0.84	0.78	0.89

Panel B: 2011 Age Distribution

Age	Securitization	Equity Analysts	Lawyers
30 and under	0.53%	0.26%	0.26%
31 to 35	6.60%	6.46%	5.37%
36 to 40	16.09%	21.96%	15.86%
41 to 45	27.97%	32.56%	24.04%
46 to 50	23.48%	18.60%	19.69%
51 to 55	13.72%	10.08%	18.16%
56 to 60	6.07%	4.13%	10.74%
Over 60	5.54%	5.94%	5.88%
Total with age data	379	387	391
Missing age data	21	13	9
Chi-Square Test of Homogeneity with Sctzn Sample	N/A	10.92	10.67
Homogeneity Test, p-value	N/A	0.14	0.15
Median age	45	44	46

Table 2.2: Properties

This table provides summary statistics for properties owned anytime over 2000-2010. Dollar amounts are reported in December 2006 CPI-adjusted real thousands. Panel A presents the fraction of people owning more than one address over 2000-2010. Panel B presents summary statistics for our matching process with mortgage applications.

Panel A: Total Properties, Purchases and Sales

	Securitization	Equity Analysts	Lawyers
Total properties ever owned, 2000-2010	674	604	609
Total purchases, 2000-2010	437	368	355
with purchase price	392	318	306
average purchase price	761.67	1032.38	485.62
Number of homes with no purchase date	81	112	101
Total sales, 2000-2010	266	207	171
with sale price	238	172	145
average sale price	633.74	794.76	446.37
Number of homes with no sale date or sold after Dec 31 2010	408	397	438

Panel B: Mortgage Applications

	Securitization	Equity Analysts	Lawyers
Purchases, 2000-2010	437	368	355
with mortgage info	327	247	257
mean, median LTV	0.72 / 0.79	0.71 / 0.75	0.73 / 0.80
with income from HMDA match	253	196	203
income at purchase, property average	350.01	408.74	191.32
People purchasing, 2000-2010	274	242	243
with income from any HMDA match	191	153	167
Average # HMDA mortgage applications per match	2.41	2.62	2.46
Median # HMDA mortgage applications per match	1	1	1

Table 2.3: Income

This table presents average income in three periods for each group. We first average income from purchases observed within each person-period before averaging across people to obtain an average income for each period. Dollar amounts are in December 2006 CPI-adjusted thousands. Row A tests whether the boom minus pre-boom difference in averages was zero by projecting person-level income onto an indicator for the boom period in a two-period unbalanced panel of person-level income. Row B tests whether the difference-in-difference is significant across groups. Standard errors are clustered at the person level. */**/*** denotes significant at the 10%, 5%, and 1% levels, respectively.

		Income		
		Sctzn.	Equity Analysts	Lawyers
Pre-Boom period	Mean	246.4	360.4	170.4
(2000-2003)	Median	180.9	224.7	148.7
	SD	266.8	335.4	114.8
	People	83	72	70
Boom period	Mean	338.8	418.0	174.0
(2004-2006)	Median	210	246.4	131.8
	SD	513.8	501.9	116.4
	People	89	58	68
Bust period	Mean	369.2	476.1	231.9
(2007-2010)	Median	205.8	308.0	151.4
	SD	489.2	433.4	258.6
	People	68	56	54
A) Boom-PreBoom	Point Est.	92.36	57.62	3.68
	t-stat	[1.68]*	[0.76]	[0.19]
	N	172	130	138
	R2	0.012	0.005	0.000
B) DID	Point Est.		34.75	88.68
Sctzn. minus	t-stat		[0.37]	[1.53]
Control	N		302	310
	R2		0.021	0.047

Table 2.4: Divesting Houses

The first three columns tabulate the number of divestitures per homeowner for each group, by year. T-statistics from a two-sample test of differences in means with the securitization sample are reported each group-year for the two control groups. The next two columns report regression-adjusted differences in the number of divestitures per person each year, where we control for the eight age groups defined in Table 2.1 as well as an indicator for whether someone is a multi-homeowner at the start of the year, and the sample period is 2000-2010. The number of people in-sample each year is the number of homeowners at the beginning of each year for the two groups that are compared. T-statistics computed from person-clustered standard errors are reported in brackets below each difference. */**/*** represents statistically significant at the 10%, 5%, and 1% levels, respectively.

	Divestitures per person			Regression-Adjusted Difference	
				Sctzn. minus:	
Year	Securitization	Equity Analysts	Lawyers	Equity Analysts	Lawyers
2000	0.045	0.060	0.018	-0.012	0.026
		[-0.67]	[1.67]*	[-0.50]	[1.59]
2001	0.038	0.043	0.020	0.003	0.019
		[-0.25]	[1.16]	[0.14]	[1.22]
2002	0.040	0.062	0.030	-0.011	0.012
		[-0.94]	[0.62]	[-0.48]	[0.71]
2003	0.045	0.050	0.032	0.001	0.019
		[-0.21]	[0.78]	[0.06]	[1.10]
2004	0.040	0.050	0.034	-0.006	0.004
		[-0.58]	[0.39]	[-0.36]	[0.24]
2005	0.024	0.044	0.060	-0.014	-0.033
		[-1.26]	[-1.80]*	[-0.85]	[-1.69]*
2006	0.030	0.040	0.019	-0.007	0.009
		[-0.61]	[0.92]	[-0.44]	[0.75]
2007	0.048	0.031	0.021	0.023	0.025
		[1.03]	[1.89]*	[1.40]	[1.66]*
2008	0.061	0.038	0.030	0.026	0.025
		[1.28]	[1.88]*	[1.44]	[1.50]
2009	0.045	0.024	0.015	0.029	0.031
		[1.45]	[2.17]**	[1.91]*	[2.09]**
2010	0.029	0.020	0.027	0.012	0.001
		[0.59]	[0.13]	[0.79]	[0.09]
			Multi-homeowner?	0.063	0.066
				[7.71]***	[8.14]***
			Age Indicators?	Y	Y
			N	5739	6149
			R-Squared	0.022	0.026
			People	633	675

Table 2.5: Buying a Second Home or Swapping Up

The first three columns tabulate the number of second home/swap up purchases per homeowner for each group, by year. T-statistics from a two-sample test of differences in means with the securitization sample are reported each group-year other than the securitization group. The next two columns report regression-adjusted differences in the number of second home/swap up purchases per person each year, where we control for the eight age groups defined in Table 2.1 as well as an indicator for whether someone is a multi-homeowner at the start of the year. The number of people in-sample each year is the number of homeowners at the beginning of each year for the two groups that are compared, and the sample period is 2000-2010. T-statistics computed from person-clustered standard errors are reported in brackets below each difference. */**/*** represents statistically significant at the 10%, 5%, and 1% levels, respectively.

	Second home/swap up purchases per person			Regression-Adjusted Difference	
				Sctzn. minus:	
Year	Securitization	Equity Analysts	Lawyers	Equity Analysts	Lawyers
2000	0.054	0.055	0.057	0.014	-0.011
		[-0.05]	[-0.15]	[0.66]	[-0.56]
2001	0.064	0.086	0.057	0.006	0.013
		[-0.86]	[0.31]	[0.27]	[0.65]
2002	0.113	0.075	0.072	0.065	0.047
		[1.38]	[1.62]	[2.65]***	[2.11]**
2003	0.080	0.079	0.093	0.024	-0.008
		[0.04]	[-0.50]	[0.99]	[-0.33]
2004	0.097	0.081	0.054	0.034	0.035
		[0.65]	[1.78]*	[1.45]	[1.54]
2005	0.121	0.070	0.082	0.068	0.037
		[1.94]*	[1.49]	[2.96]***	[1.62]
2006	0.073	0.065	0.062	0.029	0.002
		[0.36]	[0.53]	[1.37]	[0.12]
2007	0.087	0.066	0.048	0.037	0.031
		[0.86]	[1.78]*	[1.72]*	[1.50]
2008	0.045	0.045	0.024	0.017	0.013
		[-0.01]	[1.44]	[1.02]	[0.91]
2009	0.026	0.037	0.048	0.006	-0.023
		[-0.79]	[-1.40]	[0.41]	[-1.48]
2010	0.029	0.044	0.030	-0.002	-0.002
		[-1.02]	[-0.08]	[-0.10]	[-0.14]
			Multi-homeowner?	0.246	0.262
				[19.83]***	[18.04]***
			Age Indicators?	Y	Y
			N	5739	6149
			R-Squared	0.183	0.202
			People	633	675

Table 2.6: Robustness

We report the regression-adjusted differences in the annual intensity of a second home purchase or swap-up, where we pool together intensities every other year in our sample, as in equations (2) through (4). Column 1 compares the intensity of securitization agents versus equity analysts among the sample of people who work at common firms, and includes firm effects. Columns 2-3 report differences where we condition the sample to homeowners in the New York City area. Columns 4-5 report differences where the sample is conditioned to homeowners in the Southern California. Columns 6 and 7 report difference-in-differences estimates of the effect of securitization agents minus equity analysts in Southern California minus New York City. Columns 8 and 9 report differences between securitization agents in Southern California and New York. Standard errors clustered at the person level are reported below in brackets. */**/*** represents statistically significant at the 10%, 5%, and 1% levels, respectively.

	Firm Effects	NYC Homeowners		S.CA Homeowners		Diff in Diff, S.CA-NYC		Within Securitization	
	Sctzn. Minus	Securitization minus:		Securitization minus:		Sctzn-Eq.Analysts, $\beta(s(t))$		S.CA minus NYC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Year	Equity Analysts	Equity Analysts	Lawyers	Equity Analysts	Lawyers	Second Home or Swap Up	Second Home Only	Second Home or Swap Up	Second Home Only
2000-2001	0.006	0.010	-0.016	0.067	0.001	0.017	-0.082	0.055	-0.014
	[0.22]	[0.39]	[-0.61]	[0.73]	[0.02]	[0.17]	[-0.91]	[1.13]	[-0.49]
2002-2003	0.092	0.051	-0.021	0.043	0.024	-0.160	-0.109	0.059	0.060
	[2.81]***	[1.78]*	[-0.64]	[0.35]	[0.39]	[-1.02]	[-0.89]	[0.88]	[0.89]
2004-2005	0.025	0.055	0.019	0.218	0.056	0.113	0.103	0.067	0.084
	[0.76]	[2.04]**	[0.64]	[4.15]***	[1.30]	[2.03]**	[2.16]**	[1.54]	[2.09]**
2006-2007	0.086	0.019	-0.004	0.006	-0.039	-0.076	-0.047	-0.008	-0.002
	[2.46]**	[0.74]	[-0.14]	[0.10]	[-0.93]	[-1.07]	[-0.69]	[-0.24]	[-0.07]
2008-2009	0.034	0.013	-0.017	0.079	-0.103	0.051	0.034	-0.025	-0.016
	[1.47]	[0.70]	[-0.69]	[1.62]	[-2.38]**	[1.20]	[1.04]	[-0.79]	[-0.59]
2010	0.007	-0.040	-0.049	0.026	-0.044	0.126	0.098	0.041	0.042
	[0.25]	[-1.94]*	[-1.76]*	[0.26]	[-0.86]	[2.25]**	[1.97]*	[0.94]	[0.94]
Multi-HO?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Age Indicators?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm Effects?	Y	N	N	N	N	N	N	N	N
N	1876	1868	1478	373	581	2183	2183	999	999
R-Squared	0.179	0.122	0.174	0.215	0.280	0.149	0.098	0.191	0.119
People	214	234	179	52	74	279	279	130	130

Table 2.7: Performance Index

Panel A presents summary statistics for the performance index exercise. Averages per person are reported while standard deviations are reported below in parentheses. Dollar amounts are in nominal thousands. Panel B reports average performance and regression-adjusted differences in performance weighted by the initial portfolio value. Regression-adjusted differences are the coefficient on an indicator for the securitization group in a person-level cross-sectional regression of the dependent variable indicated in first column of the row on a securitization group indicator and indicators for age controls, with samplings weights equal to the initial portfolio value and heteroskedasticity-robust standard errors reported in brackets. */**/*** denotes statistically significant at the 10%, 5% and 1% levels, respectively.

Panel A: Summary Statistics

	Securitization		Equity Analysts		Lawyers	
	2000q1	2010q4	2000q1	2010q4	2000q1	2010q4
Number of properties per person	0.603	1.020	0.590	0.993	0.652	1.095
	(0.693)	(0.766)	(0.799)	(0.809)	(0.727)	(0.817)
Value of properties	236.8	751.2	308.2	992.2	191.1	522.6
	(390.2)	(893.8)	(568.7)	(1210.1)	(282.0)	(522.4)
Cash account	848.0	689.2	1159.7	988.0	470.3	375.1
	(874.7)	(975.4)	(1090.6)	(1005.6)	(461.9)	(529.5)
Portfolio value	1084.8	1440.4	1467.9	1980.2	661.4	897.7
	(1035.9)	(1586.0)	(1214.1)	(1661.3)	(548.9)	(829.1)
Housing portfolio weight	0.256	0.542	0.245	0.506	0.322	0.599
	(0.316)	(0.314)	(0.327)	(0.335)	(0.345)	(0.291)
Number of people	400		400		400	

Panel B: Performance, 2000q1-2010q4

	Means and Std. Devs.			Reg.Adj. Differences Sctzn. minus:	
	Sctzn.	Equity Analysts	Lawyers	Equity Analysts	Lawyers
Return	0.328	0.349	0.357	-0.045	-0.027
	(0.197)	(0.169)	(0.221)	[-2.63]***	[-1.08]
Buy-and-hold return	0.366	0.369	0.377	-0.017	-0.008
	(0.120)	(0.116)	(0.140)	[-1.72]*	[-0.75]
Performance index	-0.0378	-0.0199	-0.0198	-0.027	-0.018
	(0.147)	(0.113)	(0.145)	[-2.19]**	[-1.02]
Return, 2006q4-2010q4	-0.0736	-0.0457	-0.0814	-0.022	0.004
	(0.108)	(0.0936)	(0.115)	[-2.68]***	[0.44]
N	400	400	400	766	770
R-squared on perf. index				0.033	0.034

Table 2.8: Within-Securitization Performance Index

This table reports average performance and regression-adjusted differences in performance within subgroups of the securitization sample, weighted by the initial portfolio value. Regression-adjusted differences are the coefficient on an indicator for the securitization group in a person-level cross-sectional regression of the dependent variable indicated in first column of the row on a securitization group indicator and indicators for age controls, with samplings weights equal to the initial portfolio value and heteroskedasticity-robust standard errors reported in brackets. */**/*** denotes statistically significant at the 10%, 5% and 1% levels, respectively.

Panel A: Sell-side vs. Buy-side

	Means and SDs		Reg.Adj Diff.
	Sell-side	Buy-side	Sell-Buy
Return	0.275	0.361	-0.092
	(0.184)	(0.198)	[-3.01]***
Buy-and-hold return	0.347	0.377	-0.031
	(0.118)	(0.120)	[-2.17]**
Performance index	-0.0727	-0.0162	-0.060
	(0.168)	(0.127)	[-2.44]**
Return, 2006q4-2010q4	-0.0985	-0.0583	-0.039
	(0.118)	(0.0990)	[-2.95]***
N	161	239	379
R-squared on perf. index			0.080

Panel B: Worst and Best Performing Firms

	Means and Std. Devs.		Reg.Adj Diff.
	Worst	Best	Worst-Best
Return	0.269	0.337	-0.057
	(0.159)	(0.193)	[-1.76]*
Buy-and-hold return	0.347	0.350	0.011
	(0.135)	(0.103)	[0.49]
Performance index	-0.0783	-0.0134	-0.068
	(0.158)	(0.138)	[-2.29]**
Return, 2006q4-2010q4	-0.0957	-0.0619	-0.043
	(0.0977)	(0.102)	[-2.61]***
N	103	77	174
R-squared on perf. index			0.102

Table 2.9: Value-to-Income

This table presents average value-to-income (VTI) at purchase in three periods for each group. We first average VTI from purchases observed within each person-period before averaging across people to obtain an average VTI per purchaser for each period. Row A tests whether the boom minus pre-boom difference in averages was zero by projecting person-level income onto an indicator for the boom period in a two-period panel of person-level income. Row B tests whether the difference in difference is significant across groups. Standard errors are clustered at the person level. */**/*** denotes significant at the 10%, 5%, and 1% levels, respectively.

		Sctzn.	Equity Analysts	Lawyers
Pre-Boom period	Mean	3.2	2.9	2.9
(2000-2003)	Median	3.1	2.7	2.5
	SD	1.3	1.5	1.2
	People	65	60	49
Boom period	Mean	3.4	3.1	3.3
(2004-2006)	Median	3.0	2.8	3.2
	SD	2.0	1.7	1.7
	People	73	45	46
Bust period	Mean	3.1	3.1	2.8
(2007-2010)	Median	3.0	3.1	2.8
	SD	1.2	1.4	1.3
	People	55	51	40
A) Boom-PreBoom	Point Est.	0.268	0.175	0.400
	t-stat	[0.94]	[0.57]	[1.37]
	N	138	105	95
	R2	0.006	0.003	0.019
B) DID	Point Est.		0.093	-0.132
Sctzn. minus	t-stat		[0.22]	[-0.32]
Control	N		243	233
	R2		0.015	0.015

CHAPTER III

Ultimate ownership and bank competition

José Azar Sahil Raina Martin Schmalz

IESE Univ. of Michigan Univ. of Michigan

Abstract

We use a uniquely extensive branch-level dataset on deposit account interest rates, maintenance fees, and fee thresholds, and document substantial time-series and cross-sectional variation in these prices. We then examine whether variation in bank concentration helps explain the variation in prices. The standard measure of concentration, the HHI, is not correlated with any of the outcome variables. We then construct a generalized HHI (GHHI) that captures both common ownership (the degree to which banks are commonly owned by the same investors) and cross-ownership (the extent to which banks own shares in each other). The GHHI is strongly correlated with all prices. We use the growth of index funds as an arguably exogenous source of cross-sectional variation of county-level common ownership to suggest a causal link from the GHHI to higher prices for banking products.

3.1 Introduction

Many fees for banking deposit services and the deposit thresholds to avoid the fees are at historic highs (Figure C.1). Of course, many factors contribute to changes over time. But those fees, their avoidance thresholds, and the interest rates paid on the related accounts also vary substantially in the cross section. For example, even in the low-interest environment of 2013, CD rates vary by almost one percentage point across US counties (Figure 3.1C). Do differences in bank competition help explain the variation in prices consumers pay for the privilege of storing their savings? In other words, which measure(s) of bank concentration effectively capture the price variation in this input market?

Measuring bank concentration and its consequences has been a primary interest of financial economists for decades, because it is hugely important for many areas of economics. For example, higher bank concentration (1) is related to increased barriers to entry for firms and latent entrepreneurs – particularly for the poor and for minorities – and can negatively affect economic growth, (2) hamper the transmission of monetary policy, (3) slow down the adoption of new technologies, (4) increase inequality and crime, and (5) adversely affect households, who receive lower rates on their savings and pay more for consumer loans. The degree of bank competition can also affect (6) the fragility of the financial system, (7) the value of lending relationships, (8) lending standards, (9) the propensity of lenders to foreclose on their borrowers, and (10) the allocation of labor to its most productive use.[1] For all these applications, it is crucial to understand the economic forces governing bank competition and to

[1] (1) Black and Strahan (2002); Collender and Shaffer (2003); Beck et al. (2004); Cetorelli (2004); Cetorelli and Strahan (2006); Kerr and Nanda (2009); Canales and Nanda (2012); Chatterji and Seamans (2012); Love and Pería (2014), (2) Hannan and Berger (1991); Neumark and Sharpe (1992); Drechsler et al. (2014); Scharfstein and Sunderam (2014), (3) Allen et al. (2009), (4) Garmaise and Moskowitz (2006); Beck et al. (2010), (5) Kahn et al. (1999, 2005), Célérier and Matray (2014), and others (reviewed below), (6) Beck et al. (2006); Beck (2008); Berger et al. (2009); Martinez-Miera and Repullo (2010); Hakenes and Schnabel (2011); Anginer et al. (2013); Beck et al. (2013); Egan et al. (2014), (7) Petersen and Rajan (1995); Simkovic (2013), (8) Ruckes (2004), (9) Favara and Giannetti (2015); Gormley et al. (2015), (10) Bai et al. (2015).

find effective ways to measure them. However, because of data limitations and the use of conceptually incomplete measures of concentration, our understanding of bank competition remains uncomfortably limited, as the present study helps illustrate.

This paper contributes to the literature (i) new facts from a uniquely extensive branch-level panel data set on various prices of bank deposit products, (ii) the computation of a more general and conceptually complete, more realistic, and empirically more effective measure of bank concentration: the generalized Herfindahl-Hirschman Index of market concentration (GHHI), (iii) first evidence that common ownership and cross-ownership increase monopsony power, and (iv) a new source of cross-sectional variation of ownership structures (and hence market concentration as measured by the GHHI): index fund growth.

We find that (i) fees and thresholds have increased markedly over the last decade, and exhibit large cross-sectional variation. In particular, prices of deposit products are higher in California, New York, and New Jersey than in the midwest (e.g., Kansas or Nebraska). This is perhaps surprising, given that there are a lot more banks in New York, and HHI is *lower*, than in the Midwest (Figure 3.2A). Indeed, we also find that (ii) changes in the HHI do not correlate with changes in either fees, thresholds, or deposit spreads.

One reason why the HHI fails to reliably explain variation in prices is that it assumes that every bank is owned by individuals that hold no stakes in other banks. In other words, HHI ignores the very high and increasing degree of overlapping ownership between banks, illustrated in Table 3.1. The same four institutional investors are among the top 5 shareholders of the nation's five largest banks. The fifth important player is Berkshire Hathaway, which ranks among the top five shareholders of three of the top six banks. In addition to such common ownership links, there are cross-ownership links: many banks have asset management divisions that are shareholders of competitor banks. As a consequence, banking is an industry in which an effective concentration measure has to jointly take into account common ownership and cross-ownership.

We provide such a measure: the GHHI. The GHHI is a generalization of the MHHI of O'Brien and Salop (2000) that accounts not just for common ownership but also for cross-ownership.[2] The market-level GHHI as of 2013 is more than 2,500 points higher than the HHI (Figure 3.6B); we call this difference "GHHI delta". This magnitude compares to regulatory thresholds for merger review of 200 HHI points. Given the large common ownership concentration (GHHI delta), and given the negative correlation between HHI and GHHI delta, it is not surprising that omitting GHHI delta leads to a severe downward bias when estimating the effect of concentration on prices with existing measures (HHI).

By contrast to the HHI, GHHI is strongly and reliably correlated with all fees and thresholds. Indeed, GHHI levels are higher in the high-price areas on the coasts, compared to the middle of America (Figure 3.2B), and changes over time in market-level GHHIs correlate with local price changes. These findings indicate that (iii) depository institutions' monopsony power, generated through common ownership and cross-ownership links, has a strong correlation with prices in one of their input markets: the market for deposits.

As a final contribution, (iv) we use variation from the growth of index funds (as opposed to actively chosen portfolios) to suggest a causal relationship between GHHI-based concentration and prices using two methods. Doings so attenuates several endogeneity concerns present in the panel regression analysis above, including the reverse causality concern that "active" investors choose their portfolios in response

[2]Brito et al. (2015) also develop an index that generalizes MHHI to allow for simulateneous common ownership and cross-ownership (i.e., partial ownership by competitors). The main difference between our and Brito et al. (2015)'s derivation is that "our" GHHI has the appealing property that the ultimate control shares add up to 100%; see Appendix E. The generalization from MHHI to GHHI is important: there are 656 counties (out of about 3000 counties in the contiguous US) where the difference between MHHI and GHHI is greater than 100 HHI points, and 231 counties where the difference is greater than 200 HHI points.

to expected deposit prices. First, we directly instrument our GHHI measure using county-level index fund ownership of banks and find that index fund growth-induced variation in the GHHI is strongly correlated with higher fees, thresholds, and deposit spreads. Under the assumption that aggregate index fund growth is not primarily caused by across-county variation in banking market outcomes, this finding indicates that index fund growth causes higher prices for banking products.

To mitigate the concern implied by the identifying assumption of the instrumental variable analysis described above, we complement it with a difference-in-differences (DiD) analysis. Specifically, we show that deposit prices in 2013 can be predicted using only information about county-level industrial organization of banks, the banks' ownership, and prices from ten years before. The reason is that higher common ownership in 2003 predicts greater increases in common ownership, which in turn predicts greater price changes over the next decade. We mitigate remaining endogeneity concerns by showing that similar results obtain when we use only variation from changes over time in within-bank variation across counties in the prices they charge.[3] As long as index fund ownership in a given county in 2003 is not determined by price changes a decade in the future *over and above what is reflected in the market value of the bank holding companies or any other characteristic of the bank holding company*, these findings imply that greater levels of common ownership cause higher prices for deposit products.

Given these findings, the question arises which corporate governance mechanism implements these outcomes. The first thing to note is that the fact that concentrated ownership is related to higher prices for banking products need not be driven by collusion, i.e., coordinated price setting between banks. Mutual funds' unrecorded "engagement" meetings with their various portfolio firms could in principle be used as

[3]Within-bank variation exists for large banks, but is limited for fees and thresholds (as opposed to rates) As a result, statistical significance of the fee and threshold results is reduced when we use only within-bank variation over time.

such a coordination device; see Azar et al. (2015) for a more comprehensive discussion and literature review. But overlapping ownership interests can cause anti-competitive effects even in a world of competition, in which each firm independently maximizes its shareholders' economic interests – their portfolio profits (O'Brien and Salop, 2000). Such a model is a simple generalization of the traditional Cournot model, which assumes that shareholders' portfolios contain only a single stock; the generalization allows for any portfolio composition. Moreover, Azar (2012) shows that the O'Brien and Salop (2000) equilibrium can be microfounded as the outcome of the battle for corporate control, in which potential managers strive to earn the votes of the industry firms' shareholders. Managers who–through either conscious calculation, intuition, or pure luck–propose broad strategic plans that correctly represent shareholder interests will tend to be selected to run the firms, and managers that fail to propose such strategic plans will tend to be selected out.[4]

Thus, rather than actively encouraging or facilitating collusion, it is possible that the large, diversified mutual funds let portfolio firm managers live a "quiet life" with high margins and low competition (Bertrand and Mullainathan, 2003). This problem becomes more severe when index funds can outvote smaller undiversified activist investors that would otherwise push firms to compete harder (Aslan and Kumar, 2015). Indeed, Ackman (2016) expects that the crowding out of activists by index funds will lead to keiretsu-style corporate governance failures in the US.

In sum, the outcomes we document can be implemented either by active involvement in corporate governance on behalf of the mutual fund companies, or, more simply, by the index funds' failure to push firms to compete hard and prevention of activist campaigns that would pursue that goal.

The most direct policy implication of our findings is that bank regulators should consider taking ownership structures into account when measuring bank concentra-

[4]See Schmalz (2015) for a case study where index funds voted against an activist campaign that arguably would have strengthened product market competition.

tion. A failure to do so may lead to "hidden" increases in bank concentration through partial common- and cross-ownership links that can cause adverse effects on bank competition and the economy at large that go undetected when using the HHI.

Common ownership has been shown to increase monopoly power in a different industry and using different techniques (Azar et al., 2015). The fact that anti-competitive effects of concentrated ownership appear to increase market power in more than one industry, and that the finding is robust to multiple identification strategies justifies that antitrust agencies now devote more resources to investigating the role of a small set of large asset management companies in firms' competitive behavior.

Let us give some perspective on these players. The largest asset manager, Black-Rock, has \$4.7trn assets under management, and is the largest shareholder of more than a fifth of *all* American publicly traded firms. "Some have mistakenly assumed that [these investors'] predominantly passive [investment] management style suggests a passive attitude with respect to corporate governance. Nothing could be further from the truth." (Vanguard Chairman and CEO F. William McNabb)[5] If this misperception continues, further growth and consolidation of these multi-trillion-dollar asset management firms as well as coordination between them with respect to their corporate governance activities (Foley and McLannahan, 2016) could lead to further gradual erosion of competition across the entire economy, with adverse consequences for consumer welfare, economic efficiency, macroeconomic output, and inequality. Elhauge (2016) discusses some of these potential consequences as well as legal implications of our findings.

[5]February 27, 2015. https://about.vanguard.com/vanguard-proxy-voting/CEO_Letter_03_02_ext.pdf. See also: Carleton et al. (1998); Becht et al. (2007); Chen et al. (2007); Appel et al. (forthcoming); McCahery et al. (forthcoming); Mullins (2014); Boone and White (forthcoming); Dimson et al. (forthcoming).

3.2 Data

In this section, we detail the data sources for our analysis, and then present the first main result of the paper: the variation over time and across geographies of fees, thresholds, and deposit interest rates and spreads.[6]

3.2.1 Data sources

We use three main sources of data: RateWatch, FDIC's Summary of Deposits, and Thomson Reuters's SEC 13F filings database. RateWatch provides branch-level data on rates and fees that we use as our outcome variables. FDIC's Summary of Deposits supplies the branch-level deposits data used to calculate market share. Thomson Reuters's database of SEC 13F filings provides data on institutional ownership of public banks, which we use, along with FDIC data, to construct the GHHI. We also use Thomson Reuters's SEC 13F database to measure index fund ownership over time.

3.2.1.1 Data on banking product prices

RateWatch was established to provide their clients – the major US banks – with information on competitors' prices at the branch level. We use their data on fees, fee thresholds, and deposit rates to examine the total price customers pay when depositing savings. We have deposit fees data from over 3,000 banks and deposit rates data from over 9,600 banks.

Our fees data are extensive. For example, in 2013, we have data on money market maintenance fees for at least one branch in the same county as 99.1% of the US population. Overall, we have over 4.5 million fee amount and fee threshold observations

[6] Motivated by theory (detailed in the next section), we define deposit rate spreads as the difference between the ten-year treasury rate and the respective deposit interest rate, normalized by the ten-year treasury rate. This expression most closely corresponds to the theoretical equivalent for margins, and is also more stable over time than spreads in levels.

for both money market and interest-bearing checking accounts.

Our interest rates data coverage is even more extensive than the fees data. For 12-month CDs, in 2013, we have interest rates from at least one branch in the same county as 99.9% of the US population. Overall, for each deposit rate that we explore, we have over 60 million observations.

3.2.1.2 FDIC Summary of Deposits

We use FDIC's Summary of Deposits (SoD) to calculate market share for banks. SoD is based on an annual survey of deposits completed by all FDIC-insured bank branches. SoD is a standard source for measuring bank product market concentration in the extant literature (e.g., Cetorelli and Strahan, 2006; Drechsler et al., 2014).

3.2.1.3 Thomson Reuters SEC 13F data

All institutions that "exercise investment discretion over $100 million or more" must file a Form 13F every quarter with the SEC that provides information on their holdings of US firms' equity. We use the Thomson Reuters collection of this data for two purposes. First, we use it to calculate GHHI indices, our generalized measure of market concentration, for local banking markets. Second, we use it to identify five of the largest index fund groups – iShares, Vanguard, SPDR, Invesco, and Fidelity Spartan. We use their growth as an instrument for exogenous changes in the ownership structure of banks.

3.2.2 Description of fees, fee thresholds, and rates

We begin with a description of the cross-sectional variation in interest-bearing checking and money market account maintenance fees that banks charge their customers. Figure 3.1A shows that money market account fees are higher in areas which feature more banks, such as the coastal areas, and in particular the Northeast. These fees range from under \$8 to \$25. Similarly, Figure 3.1B shows considerable geographic variation in money market account maintenance fee thresholds in 2013, going as high as \$15,000 in some counties and as low as \$50 in others. There is similar geographic variation in money market account fees and thresholds holds in other years and in interest-bearing checking accounts as well.

Figures 3.3A and B present the annual median, 20th, and 80th percentile county average maintenance fees for money market and interest-bearing checking accounts, respectively. These figures confirm that the annual distributions of fees for the two deposit products have considerable variation across counties. Similarly, Figures 3.4A and B present the annual 20th, 50th, and 80th percentile county average maintenance fee thresholds for money market and interest-bearing checking accounts.[7] Again, we see that there is considerable variation in thresholds across counties within each year.[8]

For the analyses of fees and fee thresholds, we take the annual mean of survey responses for each outcome variable for each branch and then winsorize the right side of the distribution at the 1% level to reduce the impact of suspected data errors.[9]

In our analysis on interest rates, we use deposit rates for CDs with 12-, 24-, and 36-month terms, money market accounts, and interest-bearing checking accounts. We begin by describing their cross-sectional variation. In Figures 3.1C and D, we present a map of the average interest rates in each county in 2013 for 12-month CDs and money market accounts, respectively. The variation is substantial, despite the low interest rate environment. Rates are somewhat higher in the central regions of the

[7]The figures discussed in this paragraph present data in constant 2013 USD, adjusted for inflation using CPI.

[8]The threshold dispersion in 2002 appears smaller than in other years, but, in proportion to the mean maintenance fee thresholds in that year, the variation in 2002 is similar to other years. Additionally, there is less data on thresholds in 2002 than later years. We do not use 2002 in our regression analyses.

[9]Some reported fees for some branches within some banks appear to be typos. For example, a bank reported charging a \$213 maintenance fee for certain accounts in some geographies, whereas most other branches of the same bank charged \$13. Such outliers are not part of the data we use. Unrelated, note the aggregation of responses at the annual level within branches is the reason for the smaller sample size reported in our analyses.

US than in coastal regions, with rates ranging from 0.8% for 12-month CDs and 0.5% for money market accounts in some counties to close to 0% for both types of deposit accounts in other counties. There is similar geographic variation in rates for other years in 12-month CDs and money market accounts and in 24- and 36-month CDs and interest-bearing checking accounts as well. In sum, perhaps surprisingly, banks are able to charge higher spreads in the banking markets that feature more natural competitors, such as the east coast and California, than in the markets that feature fewer competitors, such as the midwest.

Figures 3.5A, B, and C plot quarterly median, 20th percentile, and 80th percentile county-level average interest rates for 12-month CDs, money market accounts, and interest-bearing checking accounts, respectively. We observe cross-sectional dispersion in the interest rates for all three products, although there is less dispersion in the 12-month CD rates. As these figures show, the distribution of interest rates and their spread margins expands and contracts enormously over time as interest rates increase or decrease with the business cycle. This can be problematic, for example, because a bank with relatively low rates during a period of high interest rates would see a smaller drop in rates as the overall level of interest rates decline, relative to a bank with higher initial rates, simply because there is less room for its rates to fall. To avoid this "accordion" econometric problem, we run our analyses for interest rates on the within-year ranking of interest rate spread margins

We prepare the interest rates data for analysis by, first, calculating the difference between the 10-year Treasury Constant Maturity average rate for each month and our deposit interest rates as a fraction of the Treasury rate. Next, like for the fees analyses, we take annual means of the reported interest rate spread margins for each branch in the data. After that, we symmetrically winsorize the data at the 2% level. This helps reduce the impact of outliers on our analyses. Finally, we take the within-year percentiles of these rate spread margins to minimize problems tied to the "accordion" problem discussed above.

3.2.3 Banking market concentration

In this paper, we take ownership into account by measuring bank concentration using GHHI, our generalized measure of bank concentration.[10] Figures 3.2A and B present the geographic dispersion of the HHI and GHHI measures of bank concentration in 2013, respectively. Based on the maps, we observe that considering ownership significantly increases county-level banking concentration. This impact is especially clear on the coasts, particularly the New York City and DC metropolitan areas and California.

Figure 3.7 shows the cross-sectional distribution and the growth of county-level ownership-based concentration (GHHI delta) from 2002 to 2013. As most of the points are above the 45-degree line, ownership-based concentration increased in most counties from 2002 to 2013. The figure also shows that there are many counties in 2002 with high ownership-based concentration in 2002 and 2013, implying that ownership adds to county-level banking concentration not just in 2013 but in all years.

3.3 Hypothesis development and basic research design

3.3.1 HHI versus generalized (G)HHI

This paper tests two alternative concentration measures for their effectiveness in capturing differences across markets and over time in the competitiveness of the local banking sector. The standard measure, used by regulators and researchers alike, is the Herfindahl-Hirschman Index (HHI) of market concentration, which is simply the

[10]The economic reasoning for considering ownership when measuring concentration is presented in Section 3.3.

sum across firms j of market shares squared,

$$\mathrm{HHI} = \sum_j s_j^2. \tag{3.1}$$

This measure of market concentration is meaningful if each firm maximizes its own profits, i.e., each firm acts in the financial interest of an investor who has no wealth invested in other firms in the industry (or several investors with such undiversified portfolios). Under that assumption, if firms compete à la Cournot,[11] markups $\frac{P - C_j'(x_j)}{P}$ in a given market will be proportional to the market's HHI,

$$\eta \sum_j s_j \frac{P - C_j'(x_j)}{P} = \mathrm{HHI} = \sum_j s_j^2. \tag{3.2}$$

A corresponding empirical prediction is that markets with high HHI should have higher prices. This prediction assumes that marginal cost is constant across markets. This relatively strong assumption can be weakened by instead correlating changes over time in the HHI with changes in prices. A regression in changes captures the above prediction under the weaker assumption that within a market, marginal costs don't change over time, whereas marginal costs are allowed to differ across markets in ways that are correlated with firm's entry and exit decisions. Adding time fixed effects also allows for changes in marginal costs over time if they are similar across markets. These are the standard regressions the literature has examined.

As reviewed in Section 3.7, existing work finds mixed results on the correlation between the HHI and prices, especially for regressions in changes. One possible interpretation of a missing link between changes in the HHI and changes in prices is that changes in the HHI are accompanied not only by increases in market power, but

[11] Note however that the HHI as a measure of concentration is also applicable in contexts other than Cournot competition. See for example, Moresi et al. (2008).

also by decreases in marginal costs, i.e., efficiency gains. Or perhaps deposit markets simply are not local and do not vary at the county level.

An alternative interpretation of the tenuous link between prices and the HHI is that the regression model corresponding to Equation 3.2 does not fully reflect the economic forces shaping bank competition. We show below that making counterfactual assumptions about ownership leads to such a mismeasurement of economic forces, and leads to an omitted variable bias in the empirical implementation that can lead to a false negative in these regressions.

Specifically, one way in which the HHI model is inconsistent with factual reality is that it assumes that each bank is controlled by undiversified investors who do not own stakes in competitors. We have shown that assumption to be factually wrong in Table 3.1A. A generalized version of the HHI, the GHHI, can adjust the HHI model to reflect these realities.

Similar to the HHI, the GHHI can be derived from a Cournot game between competitors. Also, the assumption that the firm acts in its shareholders' interests is maintained, i.e., does not change relative to the HHI model. The only difference is that the generalized approach does not restrict the competitors to have only undiversified controlling shareholders; instead, any shareholder structure is allowed. In particular, the GHHI allows for simultaneous common ownership and cross-ownership as well. This is an important generalization of existing concentration measures especially for measuring bank concentration. The reason is that many large banks have large asset management divisions, which are major owners of other banks. We explain this point in more detail below.

Allowing for general ownership structures implies that shareholder unanimity may fail. That is, the interests of investors with different portfolios may differ. An assumption has to be made how such conflicts are resolved. We follow O'Brien and Salop (2000) in assuming that each firm maximizes the weighted average of its shareholders economic interests. Denoting shareholder i's share of control rights in firm j as γ_{ij} and her share of cash flow rights in firm k as β_{ik}, firm j's objective function is assumed to be

$$\max_{x_j} \Pi_j = \sum_{i=1}^{M} \gamma_{ij} \sum_{k=1}^{N} \beta_{ik} \pi_k, \tag{3.3}$$

where π_k are firm k's profits, β_{ik} is the ultimate financial interest of shareholder i in firm k, and γ_{ik} is the ultimate control share of shareholder i in firm k. Thus, $\beta_{ik}\pi_k$ are shareholder i's portfolio profits.

That is, we assume that firms primarily focus on the economic incentives of those shareholders with the most control rights in the firm. The outcome is that the firm will put weight not only on its own profits but also on the profits of its competitors – to the extent that its most powerful shareholders also have stakes in those competitors. Indeed, the firm's objective function (3.3) is proportional to

$$\pi_j + \sum_{k \neq j} \frac{\sum_i \gamma_{ij} \beta_{ik}}{\sum_i \gamma_{ij} \beta_{ij}} \pi_k. \tag{3.4}$$

That is, under common ownership, firm j will not compete quite so hard with more commonly owned competitors as it does with competitors that are not part of firm j's largest owners' portfolios, because any increase in own profits would come at the expense of that commonly owned competitor; such a product market strategy would not be in the largest investors' interests. In other words, the assumption is that firms internalize the externalities that come from aggressive product market behavior that they impose on competitors, to an extent that is proportional to the degree to which these competitors are owned by their largest shareholders. Note that the maximization problem in the traditional HHI model is a special case of the one presented here.

If firms represent their (potentially diversified) investors' economic interests and compete à la Cournot, the prediction ensues that markups are proportional to the GHHI index,

$$\eta \sum_{j} s_j \frac{P - C_j'(x_j)}{P} = GHHI = \sum_{j} \sum_{k} s_j s_k \frac{\sum_i \gamma_{ij} \beta_{ik}}{\sum_i \gamma_{ij} \beta_{ij}}. \tag{3.5}$$

As a result, the same regressions as in the traditional literature can be run, with the only change that the HHI index is replaced with its generalized version, the GHHI. In particular, we can examine if changes of ownership and control (e.g. because of Berkshire Hathaway's acquisition of a multi-billion dollar stake in Bank of America's cash flows in addition to the top ownership and control of Wells Fargo, or because of index fund growth) are related to price changes. The main empirical question this paper addresses is which one of these alternative indexes, the HHI or the GHHI, is better able to capture variation in prices of banking products.

3.3.2 Ultimate ownership

A complication arises in the construction of the GHHI in the banking industry. Banks often have asset management divisions, which own substantial stakes in other banks. As a result, many banks are both competitors and non-trivial owners of other banks. In addition, "pure" asset management firms such as BlackRock or Vanguard typically own large stakes in several banks. Hence, the ownership structure combines cross-ownership and common ownership. Existing modified measures of market concentration, such as the MHHI by O'Brien and Salop (2000), cannot be applied directly to this situation. We use a more general index that solves for ultimate ownership, and can simultaneously account for general patterns cross-ownership and common ownership. We describe the construction of this general index of market concentration in Appendix B. When ultimate ownership is the same as direct ownership (as is the case in the study of airline competition, Azar et al. (2015)) the MHHI and the GHHI are the same.

3.3.3 Empirical methodology: panel regressions

To examine the question whether the HHI or the GHHI better captures variation in prices of banking products, we start by examining simple correlations between the two concentration measures and banking prices. The panel regressions we run are of the form

$$R_{ijbt} = \beta \cdot \text{Concentration Index}_{it} + \theta \cdot X_{it} + \xi \cdot Q_{bt} + \nu_j + \zeta_t + \varepsilon_{ijbt}, \qquad (3.6)$$

where R_{ijbt} is an outcome variable (various fees, fee thresholds, and deposit interest rate spreads) assessed by branch j of bank b in county i in period t. Concentration Index_{it} is alternatively the HHI_{it} or GHHI_{it}. As controls X_{it}, we include market characteristics such as log median household income and log population. Q_{bt} is the market capitalization of each bank. ν_j and ζ_t are branch and year fixed effects, respectively. The motivation for market-level controls such as household income and population is to account for differences in the demand for deposit products across markets. Banks' market capitalization is included in regressions as a proxy for differences across banks in the level and and changes over time in variable costs. We include branch fixed effects to capture differing levels of service, product offerings, etc. that might otherwise bias our estimate of β. In our regressions, we estimate the coefficient β not from cross-sectional variation across markets alone, but from changes over time in the cross-sectional differences between markets. We run our panel regressions on all branches in RateWatch from 2003 to 2013 and cluster standard errors at the county level as there may be a shared component in the variation of data across branches in a given county.[12]

3.3.4 Empirical hypotheses

The key question we examine in the following section is whether the HHI or the GHHI are more robustly linked to various prices of banking deposit products. Because the only difference between the HHI and the GHHI is taking ownership structures into account, this question can be restated equivalently as whether ownership of banks matters empirically in important ways or not.

There are several reasons why the anti-competitive incentives arising from common- and cross-ownership might not get implemented: for example, agency conflicts between shareholders and management, informational frictions, or fear of antitrust backlash on behalf of the investors. Corresponding to the idea that these frictions overwhelm any anti-competitive incentives from overlapping ownership, our null hypothesis is that partial ownership links are irrelevant for economic outcomes. In that case, the HHI and the GHHI should be equally effective at capturing variation in prices. (Recall that the HHI is the special case of the GHHI in which common ownership links are irrelevant.)

H0: The HHI and the GHHI are equally effective at capturing variation in prices.

On the other hand, if firms (here: banks) indeed act in their most important shareholders' economic interests, i.e., if economic incentives matter for economic outcomes, the following alternative hypothesis should find support in the data.

H1: The GHHI is a better predictor of prices of banking products than the HHI.

[12] We do not two-way cluster our standard errors using counties and years because our panel is not long enough to justify clustering errors within years.

Formally, an important reason for the prediction that the GHHI is a stronger predictor of prices than the HHI is a classic omitted variable problem. The HHI and the difference between the GHHI and the HHI, called GHHI delta, are negatively correlated. The reason is that deposit markets such as New York City or many areas of California feature a large number of banks (low HHI), but many of them are commonly owned to a large degree (high GHHI delta), whereas banking markets in the midwest often feature only a small number of banks (high HHI), but these banks tend to be independently owned. (HHI-based merger regulation can contribute to generating this pattern.) Whatever the cause for the negative correlation, omitting the GHHI delta from the standard HHI regression (as specified in Equation 3.6) hence leads to a downward bias of the coefficient on the HHI, $E[\hat{\beta}_{HHI} \mid X] = \beta_{HHI} + (X'X)^{-1}X'(\beta_{\text{GHHI delta}} \cdot \text{GHHI delta})$.

This section only laid out the basic research design using panel regressions. We describe our strategies that examine causality in Sections 3.5 and 3.6 for our instrumental variable and difference-in-differences designs.

3.4 Panel regression results

In our panel regressions, we compare the relationship between changes in deposit product prices and changes in the two alternative market concentration measures defined in Section 3.3: the HHI and the GHHI. As dependent variables, we consider fees, fee thresholds, and the (within-year) percentile ranking of interest rate spread margins.

We measure interest rate spreads as the difference between the 10-year Treasury Constant Maturity rate and each interest rate, expressed as a percent of the 10-year treasury yield.[13] The reason for calculating percentage spreads is that we try to

[13]The results are similar when we use 1-year primary mortgage average rates from Freddie Mac

proxy for margins, as given in Equations 3.2 and 3.5. Relatedly, the reason we look at percentile rankings is to avoid the "accordion" econometric problem explained in Section 3.2.

Overall, we find that the relationship between concentration and fee amounts and thresholds is much stronger and more robust when concentration is measured using the GHHI. Similarly, a GHHI-based estimation of the relationship between concentration and CD rate spreads is a lot more effective than an HHI-based estimation. The sensitivity of rate spreads to changes in concentration is insignificant for both the HHI and the GHHI only for checking account accounts, for which banks charge higher fees and thresholds when concentration is higher. However, this non-result could also be due to the "accordion" econometric problem explained above. Generally, we find a positive and, at times, statistically significant relationship between within-year rate spread percentiles and the GHHI, suggesting that banks also adjust rates to concentration (but perhaps not as much as fees).

We now turn to a detailed discussion of the results. In Table 3.3A, we regress within-year CD interest rate spread margin percentiles on banking sector concentration for CDs with 12-, 24-, and 36-month maturities. the HHI has a small, positive, and statistically insignificant correlation with interest rate spreads for all three types of CDs, as shown in columns (1), (3), and (5). By contrast, we see in columns (2), (4), and (6) that the GHHI consistently has a four- to eight-fold larger correlation with rate spreads that is highly statistically significant for all three types of CDs. In terms of economic magnitudes, a one-standard deviation increase in the GHHI is associated with a 2.9 percentile point higher ranking in 12-month CD rate spreads, a 3.4 percentile point higher ranking in 24-month CD rate spreads, and a 3.5 percentile point higher ranking in 36-month CD rate spreads for the average branch. Note, however, that these are equilibrium correlations rather than causal effects (which we

to calculate spread margins. Aside from 10-year and 1-year loan rates for normalization, we ran specifications using raw average rates, also with similar results.

discuss in later sections).

In Table 3.3B, we present the results of regressions with the prices of money market accounts as the dependent variables. Columns (1) and (2) present maintenance fee amount regressions against the HHI and the GHHI, respectively. Columns (3) and (4) present maintenance fee threshold regressions. Columns (5) and (6) present interest rate spread margin percentile regressions. In columns (1), (3), and (5), we observe that concentration measured by the HHI has basically no statistically significant correlation with money market account prices: there is a small, marginally statistically significant relationship of the HHI with maintenance fees but all the other prices are not correlated with the HHI at all (if anything, the other prices are negatively correlated with the HHI). On the other hand, in columns (2), (4), and (6), we see that GHHI-based concentration shows a highly statistically significant, positive correlation with fee amounts, fee thresholds, and rate spreads. To get a sense of the economic magnitude of the coefficients, note that a one-standard deviation increase in the GHHI is associated with a \$0.21 increase in maintenance fees (a 2.1% increase), a \$230 increase in maintenance fee thresholds (a 7.9% increase), and a 1.6 percentile point higher ranking for the average branch's money market account.

Finally, Table 3.3C shows the results of regressing interest-bearing checking account maintenance fee amount, maintenance fee threshold, and interest rate spread on banking sector concentration. First, examining fee amounts (column (1)) and thresholds (column (3)) we find that the HHI has no statistically significant correlation with fee amounts and thresholds. However, when we measure concentration using the GHHI (columns (2) and (4), respectively), we see a large, highly statistically significant, positive correlation of concentration with both dependent variables. For interest rate spreads, we find in column (5), that the HHI has, against the standard HHI model's prediction, a highly statistically significant *negative* relationship with interest rate spread margin percentiles. On the other hand, the GHHI has a positive but statistically insignificant relationship with rate spreads (column (6)). Again, looking at economic magnitudes of these correlations, we find that a one-standard deviation increase in the GHHI is associated with a \$0.56 increase in maintenance fees (a 4.6% increase) and a \$408 increase in maintenance fee thresholds (a 9.5% increase) for the average branch's interest-bearing checking account.

To summarize, the panel regressions provide supportive evidence for the hypothesis that banking concentration as measured by the GHHI more robustly explains the variation in prices of banking products than the HHI.[14] This is true for maintenance fees, maintenance fee thresholds, and CD and money market account interest rate spreads as outcome variables. There does not seem to be a significant association between the GHHI and interest checking rate spreads. On the other hand, the HHI as a measure of concentration shows inconsistent statistical significance and inconsistent signs of the regression coefficient. Overall, the results reject the null hypothesis that the HHI and the GHHI are equally effective, and support the alternative hypothesis that the GHHI is more effective.

While many potentially omitted variables are differenced out already in the results presented above, reverse causality remains a potential concern. The results so far leave open the possibility that investors predict banks' profit margins, buy more stock in those banks, and thus generate the link between the GHHI and prices we documented above. To examine if that is indeed the main driver of the results, in the following two sections, we use variation in the GHHI from "passive" index funds' ownership alone and thus address the question whether there is a causal link between the GHHI and prices.

[14]The R^2 for HHI versus GHHI regressions does not seem to differ in our tables because the covariates and fixed effects differentially absorb the variation remaining in our banking product prices, masking any difference in explanatory power. In unreported regressions where we first regress prices on covariates and fixed effects only and then regress the residuals from this first stage regression, we find that the R^2s for residual regressions against the GHHI are higher than for residual regressions against the HHI.

3.5 Instrumental Variables results using index fund ownership

Thus far, we have documented new facts about variation in fees, fee thresholds, and interest rates across markets and over time, and have shown the variation in fees and fee thresholds to correlate far more strongly and reliably with the GHHI than with the HHI. The GHHI also correlates strongly with CD and money market account rates; the HHI does not. The difference between the HHI and the GHHI is common ownership concentration. This section addresses the question of whether the association between concentration and prices is driven only by the endogenous choice of active fund managers' portfolios, or also by changes in ownership of investors with passive investment strategies such as index funds. We use variation in prices correlated with changes in causation caused by the latter source as evidence of the causal effect of concentration on banking product prices.

3.5.1 Using index fund ownership for variation in common ownership and market concentration (GHHI)

To address the question of whether "passive" ownership of banks is related to higher prices of banking products at the branch level, we use variation in index fund ownership of banks. The idea for this research design is as follows. First, index funds' ownership changes are not driven by fund managers predicting temporary changes in margins in some banking markets versus others. As a result, reverse causality stemming from active fund managers' investment strategies that could be related to branch-level prices should not be a concern for the results we obtain using this strategy.[15]

[15] The reason that despite index funds' "passive" portfolio choice, they nevertheless can have a substantial impact on firm policies is that funds typically make their voting rights available to their fund family's central proxy voting office. These offices also engage with their portfolio firms with the aim of increasing the value of their portfolio firms. (Fund families' revenues are typically a fraction

Index fund ownership can cause cross-sectional differences in the GHHI as follows. Some banks are part of stock indices. Index funds' ownership of these banks grows when the overall fund size grows. Index funds grow when people invest their savings in index funds or when the value of their aggregate holdings goes up. Neither depends on the performance or pricing decisions of an individual bank, let alone bank branch. Hence, to a first order, index fund growth is exogenous to pricing decisions. But how does this cause cross-sectional shocks? Not all banks are part of an index – indeed, there are many privately owned, not publicly traded, banks in our sample. Those banks' ownership structure does not change when index funds grow. And some geographical areas have more banks that are part of an index than others to start with. Index fund growth thus affects the ownership structure of banks, and thus the GHHI, differentially in markets in which all players are publicly traded banks that are part of an index than in markets mainly comprised of privately owned banks or of publicly traded banks that are not part of major stock indices. Employing variation in the GHHI arising from changes in index fund ownership of banks in each market is the basic idea of our instrumental variable (IV) regression analysis.

3.5.2 Implementation

The IV regression analysis we implement is based on the specification presented in Equation 3.6. The difference is that we instrument the variation in the GHHI using index fund ownership of banks in each market, which we measure as

$$\text{Index Fund Ownership}_{it} = \sum_j s_{ijt} \times \text{Pct. Owned by Top Index Funds}_{jt}, \qquad (3.7)$$

of assets under management. Assets under management grow when the firms in the portfolio become more valuable. Firms become more valuable when their profits increase.) In that sense, there is no difference between the anti-competitive threats from common ownership of index funds or from common ownership of Berkshire Hathaway, Warren Buffett's investment firm. See Azar et al. (2015) for a more comprehensive discussion.

where s_{ijt} is the share of deposits in county i owned by bank j in period t and Pct. Owned by Top Index Funds_{jt} is the percent of bank j owned by top index funds in period t. We define top index funds as five index fund groups and ETFs: iShares (currently part of BlackRock, previously managed by Barclays Global Investors), Vanguard's index funds, SPDR (managed by State Street Global Advisers), Invesco's PowerShares, and Fidelity's Spartan index funds.

We use variation in the county-level index fund ownership measure described above to generate "exogenous" variation in the GHHI. The first stage of our IV regression is as follows:

$$GHHI_{ijbt} = \gamma \cdot \text{Index Fund Ownership}_{it} + \Theta \cdot X_{it} + \xi \cdot Q_{bt} + \eta_j + \phi_t + \psi_{ijbt} \quad (3.8)$$

where controls X_{it} and Q_{bt} are defined as in Equation 3.6, η_j is a branch fixed effect, and ϕ_t is a year fixed effect. The second stage of our IV regression is identical to Equation 3.6 with variation in our GHHI concentration index instrumented by index fund ownership from our first stage. We implement our IV analysis on a sample of all bank branches in RateWatch from 2003 to 2013, clustering errors at the county level.

3.5.3 Results

We implement our IV regression on prices for all three types of deposit products explored in Section 3.4: interest rate spread margin percentiles for 12-, 24-, and 36-month CDs and maintenance fee amounts, maintenance fee thresholds, and interest rate spread margin percentiles for money market accounts and interest-bearing checking accounts. We present results for regressions where we examine the direct correlation between prices and index fund ownership and the correlation between prices and the GHHI instrumented by index fund ownership. The results indicate that increases in concentration due to index fund ownership are indeed robustly linked to higher prices of all of these banking products.

Before presenting IV regression results, we should note that the first stage of the IV analyses for all explored depository prices show a large and highly statistically significant positive relationship between the GHHI and Index Fund Ownership, with t-statistics around 20 in all cases. The results of all the first stage regressions can be found in the panels of Table D.1.

For CD interest rate spreads, we find clear evidence that increases in common ownership due to index fund ownership are linked to higher interest rate spreads. In Table 3.4A, we present this evidence. Columns (1), (3), and (5) show that index fund ownership is positively, strongly, and highly statistically significantly correlated with rate spread margin percentiles for 12-, 24-, and 36-month CD rate spreads. In columns (2), (4), and (6), we present results for the second stage of our IV regression and find that the GHHI instrumented by index fund ownership has a positive and highly statistically significant effect on the within-year percentiles of rate spread margins for CDs of all three maturities.

Interpreting the GHHI results causally, we estimate that a one-standard deviation increase in the GHHI due to changes in common ownership causes a 5.7 percentile point higher ranking in 12-month CD rate spreads, an 8.0 percentile point higher ranking in 24-month CD rate spreads, and an 8.5 percentile point higher ranking in 36-month CD rate spreads. These results imply that common ownership has economically large effects as we see that concentration changes due common ownership substantially alter the relative location of branches within the CD rate spread distribution.

The IV regression results of Table 3.4B show that common ownership due to index fund ownership has a positive and highly significant effect on prices for money market accounts. In particular, columns (1), (3), and (5) show that index fund ownership is positively and highly statistically significantly correlated with fee amounts, fee thresholds, and within-year rate spread margin percentiles. The IV results confirm these findings. In columns (2) and (4), we see that the GHHI instrumented by index fund ownership has a positive and highly statistically significant effect on maintenance fee amounts and thresholds of money market accounts. Column (6) shows that rate spread margin percentiles are positively affected by the GHHI instrumented by index fund ownership as well and the effect is highly statistically significant. Interpreting the IV results causally, we estimate that a one-standard deviation increase in the GHHI causes an increase of \$0.31 in fees (a 3.2% increase), an increase of \$490 in thresholds (a 16.9% increase), and a 2.4 percentile point higher ranking in rate spreads for money market accounts. From 2003 to 2013, fees and thresholds for money market accounts grew by \$3.15 and \$1,960, respectively. The effect of a one-standard deviation increase in the GHHI is comparable to 10% of the overall growth of fees and 25% of the overall growth of thresholds in that period, which suggests that these GHHI effects have relatively large economic magnitude. The percentile change in rate spreads is also substantial, suggesting that the GHHI effects for interest rates are similarly important.

We find that the GHHI has a positive and highly significant effect on fees and thresholds for interest-bearing checking accounts but not on the interest rate spreads for the accounts. Table 3.4C presents the results for interest-bearing checking account prices. Columns (1) and (3) show that index fund ownership is positively and highly statistically significantly correlated with fee amounts and thresholds for interest checking olumn (5) shows that rate spread margin percentiles are positively, but statistically insignificantly, correlated with index fund ownership. IV results bear out these reduced form findings. Columns (2) and (4) show positive, highly statistically significant effects of the instrumented GHHI on maintenance fee amounts and thresholds for interest checking accounts. Column (6) shows a positive, but statistically insignificant, effect of the instrumented GHHI on money market rate spread margin percentiles. We estimate that a one-standard deviation increase in the GHHI due to changes in common ownership leads to an increase of \$1.33 in fees (an 11% increase) and an increase of \$719 in thresholds (a 16.8% increase). From 2003 to 2013, fees and thresholds for interest-bearing checking accounts grew by \$6.64 and \$4,100, respectively. Again, this suggests that the GHHI effects we observe on fees are economically large: the effect of a one-standard deviation increase in the GHHI is comparable to 20% of the growth in fees and 17% of the growth in thresholds for interest-bearing accounts in that period.

3.5.4 Remaining identification challenges

The Panel IV identification strategy is of course not perfect. Its merits are that using index fund ownership variation eliminates the reverse causality concern that active fund managers' holdings decisions are endogenous to branch-level variation in prices as we simply do not use that variation here. However, a challenge is that market-level variation in concentration stemming from changes in "passive" ownership does not only come from the aggregate growth of index funds, but could also stem from the inclusion and exclusion of banks in indices, as well as from entry and exit of banks with different levels of index ownership concentration into and out of a particular banking market. To illustrate why that is a concern, consider that the inclusion of a bank in an index could be endogenous to market-level outcomes. This observation does not challenge the primary motivation, which is showing that the variation in ownership from index fund growth is related to prices, but it puts limits on a causal interpretation of the results. In sum, the strategy employed in this section removes endogeneity concerns present in the panel regression design, but not all of them. We therefore offer difference-in-differences (DiD) analyses in the following section that avoids this concern.

Before we turn to the DiD analyses, let us examine the likely importance of the concern with the IV. To do so, we compare the baseline IV results with specifications in which the instrument is lagged by one year. The idea is that predicting the future is harder for longer-term predictions. Hence, the reverse causality concern should be attenuated when lagging the instrument. That is, the coefficients should be smaller on a lagged instrument if reverse causality due to index inclusions is the key driver behind our results. Contrary to that prediction, we find that, in general, coefficients are *higher* when the instrument is lagged. We conclude that reverse causality is less likely to be the driver of the baseline results (but, again, not impossible).

To summarize, across all the deposit products discussed, we find evidence of a robust relationship between increased common ownership and higher prices. We now offer DiD analyses that use different sources of variation from the IV and thus mitigate the endogeneity concerns pointed out above.

3.6 Difference-in-Differences results

In this section, we present our difference-in-differences (DiD) analyses. These analyses help mitigate reverse causality concerns that are not fully addressed by the IV regressions presented in Section 3.5. We show that one can predict price changes for deposit products a decade into the future, using only cross-sectional information about banks' market shares, ownership, and current price levels. Specifically, index ownership of a county's banks in 2003 predicts how much index ownership in a county increases until 2013, which predicts how much deposit prices change from 2004 to 2013.

3.6.1 Implementation

Our DiD analysis takes as given the cross-section of counties and their characteristics in 2003. This information is useful to predict the cross-section of deposit price changes over the next decade. In other words, we compare the difference in the change in deposit product prices from 2004 and 2013 between bank branches in treatment and control counties, where treatment is determined by index fund ownership terciles in 2003. The regression specification for these analyses is:

$$\begin{aligned}\Delta R_{ijb2004-2013} = \beta \cdot \mathbb{1}\left(\text{Index Fund Ownership Tercile}_{i2003} = 1\right) + \gamma \cdot R_{ijb2004} \\ + \Theta \cdot X_{i2004} + \xi \cdot Q_{b2004} + \varepsilon_{ijb}, \\ \forall\, i \text{ s.t. Index Fund Ownership Tercile}_{i2003} \in \{1, 3\}, \end{aligned} \tag{3.9}$$

where $\Delta R_{ijb2004-2013}$ is the change in an outcome variable for branch j of bank b in county i between 2004 and 2013, Index Fund Ownership Tercile$_{i2003}$ is an indicator for the tercile to which county i belongs, based on index fund ownership in 2003, $R_{ijb2004}$ is the outcome variable value for branch j of bank b in county i in 2004, X_{i2004} is a vector of market-level controls (median county income, county population) in 2004, and Q_{b2004} is 2004 market capitalization for bank b. The 2004 market-level controls are included to control for the potential effect of local demand for deposit products on subsequent product price growth and index fund ownership growth. The obvious remaining concern is that treatment may not be exogenous even conditional on these controls because (i) the size of the bank holding company corresponding to branch j may be related to changes in cost of capital or other variable costs between 2003 and 2013, and/or (ii) larger banks may be smarter in market selection and be invested in higher-growth markets in 2003, combined with the fact that larger banks are more likely to be included in index funds. To control for this potential mechanism, we include Q_{b2004}, the banks' market capitalization in 2004, as well. Of course, this strategy does not *rule out* all potential endogeneity concerns, but it mitigates the most obvious ones.

Note that we only include branches in top and bottom tercile counties in this analysis, with top tercile counties forming our treatment group and bottom tercile counties forming our control group. Bank branches in counties for which we do not possess index fund ownership data are included as control branches. Our standard errors for these regressions are clustered at the county level and we run the regressions on all bank branches in RateWatch for which we have data in 2004 and 2013.

Compared to the IV regressions in Section 3.5, these DiD analyses are far less exposed to the aforementioned concerns of reverse causality. The reverse causality concern is that inclusion of a bank in an index is endogenous to the profit margins in the markets in which it chooses to operate. In our DiD analyses, we "instrument" ultimate ownership changes through index fund ownership from up to a decade ago. For the reverse causality concern to be valid for the DiD "instrument" (and our identifying assumption to be invalid), index ownership must depend on performance of banking markets and banks' entry into and exit out of markets up to a decade in the future, *over and above what is reflected in the market value of the bank* and conditional on the other controls.

Finally, in these DiD analyses, we do not employ a concentration measure that we constructed. The terciles of index fund ownership that define our treatment and control groups are based on aggregations of ownership across banks in each county in 2003. Therefore, the findings in this section also help alleviate any concerns that our panel and IV findings arise from our GHHI measure being defined to exaggerate relationships between prices and ultimate ownership. In other words, they also offer a less structural test of the effect of ultimate ownership on prices.

3.6.2 Results

We implement the regression specified in Equation 3.9 for all the outcome variables explored previously: within-year interest rate spread margin percentiles for 12-, 24-, and 36-month CDs and maintenance fee amounts, maintenance fee thresholds, and within-year interest rate spread margin percentiles for money market and interest-bearing checking accounts. Overall, the results indicate that banking product prices increase significantly more from 2004 to 2013 for our treated bank branches than for control group branches. In fact, the difference in growth for fee amounts and thresholds between the treatment and control branches is comparable to the overall growth of these prices in the same period documented in Appendix C.

Before presenting the treatment effect of being in a high index fund ownership county on price growth, we note the clear positive relationship between high ultimate ownership in 2003 and ultimate ownership growth in 2004 through 2013. As we observe in the first row of Table 3.5, a bank branch in a top tercile county, where terciles are based on index fund ownership in 2003, sees 586 points greater growth in the GHHI over the 2004-2013 period than a bank branch in a bottom tercile county. In other words, higher index fund ownership in 2003 indeed predicts increases in common ownership over the next decade. This difference is highly statistically significant. Furthermore, relative to the overall growth in the GHHI over the same period of about 1,200 HHI points, this difference in ultimate ownership growth rates is economically large.

Table 3.6A shows that bank branches in top tercile counties have much higher growth in interest rate spread margin percentiles for CDs than bottom tercile counties. In Table 3.6A, column (1), we see that 12-month CD spreads percentile ranking growth is more than 3.5 percentile points higher for top tercile counties. Column (2) shows that the percentile ranking of 24-month spreads rises by 5.5 percentile points more for top tercile counties and column (3) shows that the percentile ranking of 36-month spreads rises by nearly 5.2 percentile points more for top tercile counties. Furthermore, these differences are highly statistically significant.

Table 3.6B presents evidence that money market prices increase more for branches in top tercile counties than for branches in bottom tercile counties. Column (1) shows that the growth in maintenance fees is $1.16 higher for branches in top tercile counties and column (2) shows that maintenance fee threshold growth is over $900 higher for these branches. The differences between top and bottom tercile county branches are highly and moderately statistically significant for the fees and thresholds, respectively. And, given that average fees and thresholds for money market account maintenance increased by approximately $0.80 and $1,200 from 2004 to 2013, respectively, the differences in growth between the top and bottom terciles are economically quite large. Column (3) shows that rate spread growth is 3.4 percentile points higher in top tercile counties. That difference is highly statistically significant as well.

In Table 3.6C, we observe that overall interest-bearing checking account price growth is greater for branches in top tercile counties. In column (1), we see that interest-bearing checking account maintenance fees grow $1.41 more in top tercile county branches. Column (2) shows that maintenance fee thresholds grow by nearly $2,600 more in top tercile county branches. Both these interest-bearing checking account price growth differences are highly statistically significant. They are also economically meaningful as the overall growth of fees and thresholds for interest-bearing checking accounts in that same time period was approximately $5 and $4,000, respectively. Column (3) shows that there is no statistically significant difference between top and bottom tercile county branches in terms of growth of interest rate spread margin percentiles (although, ignoring statistical significance, top tercile counties seem to have 1.5 percentile point higher growth in that period).

In Table 3.7, we present DiD findings that address questions associated with differences across banks. The general theme is that differences in these characteristics may be driving the cross-sectional variation in both deposit prices charged by banks' branches and banks' choice of markets in which they operate. The findings we present in Table 3.7 incorporate bank fixed effects into our DiD analyses precisely to absorb the effect of bank characteristics that might be driving our results. What we find is that, across all deposit products, the treatment effect is strong and statistically significant for interest rate spreads. For money market and interest-bearing checking accounts, the effect on spreads actually seems to strengthen (Table 3.7B and Table 3.7C, respectively). For fees and thresholds, the treatment effect, while in the correct direction, is weaker. This is likely because of a general lack of within-bank variation in our fees datasets.[16] Nevertheless, these within-bank DiD findings ease concerns about bank-related endogeneity driving our results.

3.7 Related literature

This paper contributes to a large literature on bank competition and to a smaller literature on the anti-competitive effects of common ownership and cross-ownership. Within the literature on bank competition, this is the first paper that studies the relationship between a broad set of fees and competition. Considering fees as part of the price vector is important for an accurate measurement of the effective price of deposit banking, especially in times of low interest rates. Studying the relation between competition and fee *thresholds* is important because it uncovers a previously undiscussed mechanism that can amplify inequality.[17,18] The only paper in the literature we are aware of that examines a relation between fees and competition is Melzer and Morgan (2014). That paper is based on information on checking account overdraft fees for a sample of depository institutions. Berg et al. (2015b) examine

[16]For instance, in 2004 and 2013, the RateWatch money market fees dataset includes data from branches of approximately 1,000 and 1,500 banks, respectively. However, in those years, less than 50 banks (around 3% in each year) have cross-sectional variation in fees. With so little cross-sectional variation, it is unsurprising that we have difficulty identifying statistically significant relationships for fees and thresholds.

[17]While no academic study on deposit fees and fee thresholds exists, there is a public debate, largely based on bank-level revenues from fees versus rates; see, e.g., Wall Street Journal, May 12, 2015, "Overdraft Fees Continue to Weigh on Bank Customers," BloombergView, November 11, 2015, "A checking account is a dangerous thing," USA Today, September 29, 2014, "Survey: ATM, checking overdraft fees surge," or US News & World Report, "Are bank fees set to rise?". The New York Times reports that 8% of 2015 bank profits stem from overdraft fees alone.

[18]Depositors can avoid account maintenance fees by maintaining a balance in excess of some fee avoidance threshold. Naturally, richer households are in a better position to avoid such fees than less affluent depositors. Hence, if lessened competition was associated with higher thresholds, lessened bank competition would contribute to inequality through this channel.

the role of fees in syndicated *loans* between the US and Europe, but without studying competition as a factor. Moreover, for deposits and loan rates, our data has an order of magnitude more banks (over 9,000) and over two orders of magnitude more observations (over 60 million interest rate data points each for 12-month CDs, money market, and interest-bearing checking accounts) than most existing studies.

Within the literature on the anti-competitive effects of common ownership and cross-ownership, the conceptual contribution of the present paper is to jointly study common ownership and cross-ownership. The GHHI index we develop is the most general in the literature, and is thus suitable for future studies of market concentration in all areas of economics. In addition, we contribute a new source of exogenous variation in ownership: index fund growth. Lastly, we offer first evidence for increased monopsony power through common ownership.

We contribute to the more general literature in industrial organization by providing a new explanation – common ownership as an omitted variable – for the failure of the HHI to capture variation in markups. This contribution is important because it suggests a re-interpretation of several results in the literature that have attributed the lack of correlation between the HHI and prices to potential efficiency gains reflected in increased HHIs, or the failure of the Cournot model to capture the economic forces at play. Those are not the reasons for a lack of correlation between the HHI and prices in our data – here, omitting common ownership is the reason for the lack of correlation between the HHI and prices.

A more detailed comparison to the most closely related papers is given below.

Literature on the relation between bank concentration, profits, and prices

The literature on the relation between bank concentration, profits, and prices, reviewed by Northcott et al. (2004); Gilbert and Zaretsky (2003), finds that the HHI captures some cross-sectional differences in the level of competition between banks, to various extents. Specifically, local market HHIs correlate positively with bank profits (Rhoades, 1995; Pilloff and Rhoades, 2002; Akhigbe and McNulty, 2003) and loan rates (Cyrnak and Hannan, 1999; Hannan, 1991; Hannan and Liang, 1995; Berger et al., 2001), and negatively with deposit rates (Sharpe, 1997; Prager and Hannan, 1998; Heitfield and Prager, 2004). CR3, a measure closely related to the HHI, also correlates with higher loan rates (Edwards, 1964) and lower deposit rates (Berger and Hannan, 1989; Calem and Carlino, 1991).

However, the correlation between the HHI and prices is not very robust over time, to the introduction of controls, or to other changes in the econometric specification. Moore et al. (1998) find that the correlation of the HHI with profits declines over time and is only present in the early years of their sample; similarly, Hannan and Prager (2004) also find that the HHI loses its significant influence on deposit rates over time. Note that the disappearance of the HHI effect can be explained with an increased importance of the omitted variable concern. Controlling for market or bank characteristics is sufficient to render the HHI coefficient insignificant in Flechsig (1965); Berger (1995); Hannan (1997); Melzer and Morgan (2014). A redefinition of profitability eliminates the correlation between the HHI and profitability also in Punt and Van Rooij (1999). Also, the literature finds that a correlation between the HHI and prices is more difficult to find in changes, a specification that comes closer to the theoretical idea of the Cournot model.[19] For example, Corvoisier and Gropp (2002) use country-product-level prices and variation in the HHI from bank mergers to examine the concentration-price relationship. They find no robust effect of the HHI changes on prices, and interpret their finding as consistent with efficiency increases from the mergers.[20]

[19]The Cournot model predicts that markups – not prices – correlate with the HHI. Absent measures of markups, research designs in changes thus more closely approximate the model: assuming constant cost, changes in the HHI should relate to changes in prices as costs are differenced out.

[20]Related in a different way is Allen et al. (2014), who point to a reason other than common ownership – namely, search frictions – why market power can be underestimated. Their result is likely to be strictly complementary to ours, because common ownership and search frictions are

Lastly, ours is not the first paper that points out that banks assess fees for services. Greenwood and Scharfstein (2013) present evidence showing that *aggregate nation-wide revenue* from fees has replaced interest revenue from 1997 to 2007. By contrast, we present evidence on *prices* (not revenues) *at the branch level*, and relate changes in prices to changes in competition. Berg et al. (2015a) show that fees are a significant contributor to the cost of corporate borrowing, whereas we analyze fees as a contributor to the cost of depositing money with a bank, and relate the variation in fees to bank concentration.

Literature on anti-competitive effects of common ownership and cross-ownership

There is a long theoretical literature that predicts anti-competitive effects of common ownership and cross-ownership (Rotemberg, 1984; Gordon, 1990; Gilo, 2000; O'Brien and Salop, 2000; Gilo et al., 2006). Also, there are historical precedents of increased profits due to common ownership, such as the "Morganization" of the US railroads in the 19th century, voting trusts, as well as studies of pyramidal structures in the economic history literature (Kandel et al., 2013). However, the literature has thus far offered only one piece of empirical support from one particular industry for a causal link between common ownership and anti-competitive prices (Azar et al., 2015). Relative to Azar et al. (2015), the present paper does not only expand the evidence to a larger and macro-economically important industry, it also offers the conceptual and technical innovation to jointly examine common ownership and cross-ownership to solve for ultimate control and financial interest, inspired by Leontief (1941), Leontief (1966), Ellerman (1991), Gilo et al. (2006), Brito et al. (2013). Moreover, the present paper contributes a new arguably exogenous source of variation

not obviously related. Note also that our results can result from uncoordinated, unilateral anti-competitive effects that arise under competition under common ownership; they need not be ascribed to common ownership fostering collusion, which is the mechanism considered by Knittel and Stango (2003).

in ownership that can be used generically in future research. Lastly, deposit prices are *input* prices for banks, so the present paper studies monopsony power, whereas airline tickets are outputs, i.e., Azar et al. study monopoly power.

3.8 Conclusion

There are two main empirical takeaways from this paper. First, we provide evidence that prices of deposit products are at an all-time high and vary substantially in the cross-section. Variation in bank competition that is due to variation in partial common ownership links helps explain the variation in prices. The inclusion of fees and thresholds is important: because fees and thresholds seem to be as responsive as interest rates to changes in competition, an exclusive focus of researchers and regulators on interest rates can result in an incomplete picture of the competitive outcomes in the banking industry.

Second, the paper provides a more complete picture of the economic forces shaping bank competition, with direct implications for policy. We show that who owns the banks matters for how the banks compete. Specifically, we calculate a new generalized concentration index, the GHHI, which can capture the effect of general ownership structures. Empirically, the GHHI is a more effective and robust predictor of market outcomes than the HHI, the measure traditionally used by researchers and regulators. In addition, we provide analyses that suggest a causal link from the GHHI to prices by using index funds' ownership of banks as a source of variation in bank ownership patterns across geographical markets. Given that concentrated ownership causes higher product prices in other industries (Azar et al., 2015), it appears reasonable that antitrust agencies and Senate allocate considerable resources to understanding the role of institutional investors in product pricing and capacity decisions (Drew, 2015; Dayen, 2016).

Aside from having direct policy implications, these results also challenge the way researchers have thought about corporate finance, industrial organization, and antitrust law for the past few decades. From a corporate finance perspective, our findings indicate that firms, at least to some extent, maximize their shareholders' economic interests, i.e., their portfolio profits. That means that a firm's objective function does not necessarily coincide with its individual profits. The contribution to the industrial organization literature is to point out that not only full mergers, but also partial ownership links matter. Implications for antitrust law are discussed by Elhauge (2016) and responses by Baker (2016) and others.

Researchers could benefit from taking partial ownership stakes into account to reexamine various questions also in other fields of economics. For example, decreased competition due to common ownership concentration has the potential to explain the rising capital share of income, increased inequality, reduced aggregate output, and sluggish macroeconomic growth amid record corporate profits.[21]

Aside from using GHHIs to measure concentration, what should policy makers do about the problem? There are remedies in existing US competition law that may help restore the efficiency loss due to market power that arises under concentrated ownership structures as the ones we document, but additional assumptions are needed to decide whether the enforcement of these laws would increase welfare.

Elhauge (2016) argues that stock acquisitions that substantially lessen competition are illegal under Clayton Act Section 7, *irrespective of whether there was an intent* to lessen competition, and *irrespective of the mechanism* by which the outcome is implemented. Acquisitions by modern-day index funds that lead to higher prices are not exempt from that law. Hence, if existing antitrust laws were to be enforced, it would become a necessity to rethink either the industrial organization of asset management (and, specifically, consider limits to within-industry diversification) or

[21] High margins amid slow growth are a large enough puzzle even for Goldman Sachs to ask "broader questions about the efficacy of capitalism." (Bloomberg 2/3/16: "Goldman Sachs says it may be forced to fundamentally question how capitalism is working.")

the meaning of "good governance" (specifically, consider limits to the voting power of the large institutions that also hold stock in natural competitors).

If such re-thinking were to be done, however, much care would have to be taken to appropriately weigh the benefits and costs of the current structure of the asset management industry. These benefits can be substantial, even if one ignores potential pro-competitive effects of concentrated ownership. Specifically, the benefit to asset owners of large-scale diversified asset management are (i) cheap diversification as well as (ii) improved corporate governance as a result of active involvement by the largest asset managers. These activities serve individual investors' interests in ways the investors could not achieve as independent agents. Indeed, the mutual funds' coordination of corporate governance activities may constitute a partial solution to the free-rider problem that arguably plagued corporate governance in previous decades, when more individuals held stocks directly and there weren't many large shareholders that engaged in monitoring activities.

Unfortunately, the benefits to shareholders from diversification and good governance come at a cost to consumers, and to society at large: efficient capital markets with perfect diversification and "good governance" imply deadweight losses in input and output markets. Examining this tradeoff between three individually desirable goals is a quantitative question we leave for future research.

Tables

Table 3.1. Top 5 owners of the largest six US banks These tables show the top 5 shareholders in the second quarter of 2013 and the first quarter of 2002 of the largest six American banks by deposits in the second quarter of 2013. The data source is Thomson institutional ownership data and proxy statements in the second quarter of 2013.

JP Morgan Chase	*[%]*	*Bank of America*	*[%]*	*Citigroup*	*[%]*
BlackRock	6.4	Berkshire Hathaway*	6.9	BlackRock	6.1
Vanguard	4.7	BlackRock	5.3	Vanguard	4.4
State Street	4.5	Vanguard	4.5	State Street	4.2
Fidelity	2.7	State Street	4.3	Fidelity	3.6
Wellington	2.5	Fidelity	2.1	Capital World Investors	2.4
Wells Fargo	*[%]*	*U.S. Bank*	*[%]*	*PNC Bank*	*[%]*
Berkshire Hathaway	8.8	BlackRock	7.4	Wellington	8.0
BlackRock	5.4	Vanguard	4.5	BlackRock	4.7
Vanguard	4.5	Fidelity	4.4	Vanguard	4.6
State Street	4.0	State Street	4.4	State Street	4.6
Fidelity	3.5	Berkshire Hathaway	4.3	Barrow Hanley	4.0

* These are warrants with no voting rights.

(A) Top 5 owners in 2013q2

JP Morgan Chase	*[%]*	*Bank of America*	*[%]*	*Citigroup*	*[%]*
Capital Research	6.0	AXA	4.2	State Street	4.4
Barclays	3.9	Barclays	4.0	Fidelity	3.9
AXA	3.7	Capital Research	3.6	AXA	3.7
State Street	2.5	Fidelity	3.2	Barclays	3.7
Fidelity	2.3	State Street	2.4	Wellington	1.8
Wells Fargo	*[%]*	*U.S. Bank*	*[%]*	*PNC Bank*	*[%]*
Barclays	3.4	Putnam Investment	7.4	Fidelity	6.8
Fidelity	3.2	Barclays	3.7	Barclays	3.9
Berkshire Hathaway	3.1	U.S. Bank	3.0	Barrow Hanley	3.7
Citigroup	2.9	JP Morgan Chase	2.8	Wellington	2.9
State Street	2.3	State Street	2.5	State Street	2.3

(B) Top 5 owners in 2002q1

Table 3.2. Summary statistics. This table provides annual, branch-level summary statistics that describe our outcome and explanatory variables. The first three variables are maintenance fee amounts, the next three are maintenance fee thresholds. The next six variables are interest rates for each of the deposit products examined. The next two variables are county-level HHI and GHHI, our two concentration measures. Finally, the last two variables are two covariates that we employ in our regressions: log of county-level average income and county population.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Maintenance Fee: Interest Checking	12.126	6.244	0	25	535360
Maintenance Fee: Money Market	9.800	4.251	0	20	536451
Maintenance Fee Threshold: Interest Checking	4291.701	4685.847	0	25000	493563
Maintenance Fee Threshold: Money Market	2904.967	2831.349	0	15000	491310
Interest Rate: 12-Month CD	1.674	1.381	0.042	5.608	951588
Interest Rate: 24-Month CD	1.963	1.343	0.094	6.36	932394
Interest Rate: 36-Month CD	2.201	1.324	0.1	5.758	907310
Interest Rate: Money Market	0.656	0.690	0.01	4.325	914586
Interest Rate: Interest Checking	0.205	0.264	0.01	2.5	917577
HHI	0.184	0.115	0.05	1	1004842
GHHI	0.325	0.148	0.059	1	1004842
Top Index Fund Ownership (Percent)	2.395	1.692	0	13.028	1005055
Log Income	10.799	0.255	9.766	11.706	1002906
Log Population	12.468	1.655	6.084	16.12	1002906

Table 3.3. Panel regressions of deposit prices on HHI and GHHI

(A) Panel regressions of time deposit spread percentiles on HHI and GHHI, respectively. This table shows the effect of market concentration measures on time deposit spread percentiles with 12-, 24-, and 36-month maturities. Percentiles are calculated for each year based on spreads defined as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	12-Month CD Spread (Percentile)		24-Month CD Spread (Percentile)		36-Month CD Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
HHI	4.822		5.115		3.526	
	(5.257)		(4.046)		(3.973)	
GHHI		19.44***		22.75***		23.49***
		(3.577)		(3.548)		(3.520)
Log Income	-27.27***	-24.53***	-16.28***	-13.13***	-12.35**	-9.140*
	(3.556)	(3.498)	(4.360)	(4.300)	(5.220)	(5.141)
Log Population	21.73***	20.48***	23.25***	21.85***	21.30***	19.85***
	(4.331)	(4.168)	(5.446)	(5.164)	(6.244)	(5.931)
Log(1+Market Cap)	0.266***	0.235***	0.368***	0.332***	0.414***	0.376***
	(0.0226)	(0.0228)	(0.0238)	(0.0215)	(0.0310)	(0.0274)
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	947,052	947,052	927,727	927,727	902,540	902,540
R-squared	0.672	0.673	0.659	0.660	0.670	0.672

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

(B) Panel regressions of money market account maintenance fees, thresholds, and spreads on HHI and GHHI, respectively. This table shows the effect of market concentration measures on money market account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Maintenance Fee		Maintenance Fee Threshold		Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
HHI	0.842*		-118.4		-1.945	
	(0.479)		(746.5)		(3.463)	
GHHI		1.418***		1,554***		10.76***
		(0.354)		(495.9)		(3.495)
Log Income	-1.576***	-1.303***	-1,294**	-1,016*	-13.30***	-11.83***
	(0.530)	(0.504)	(611.8)	(586.4)	(3.925)	(3.662)
Log Population	-0.718	-0.803	2,509***	2,439***	-4.800	-5.357
	(0.610)	(0.600)	(804.9)	(783.1)	(4.586)	(4.423)
Log(1+Market Cap)	0.0313***	0.0292***	45.28***	42.88*** 0.0177***	0.339***	0.321***
	(0.00360)	(0.00370)	(4.505)	(4.457) (0.0236)	(0.0233)	
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	533,815	533,815	488,666	488,666	911,361	911,361
R-squared	0.795	0.795	0.529	0.530	0.655	0.655

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

(C) Panel regressions of interest checking account maintenance fees, thresholds, and spreads on HHI and GHHI, respectively. This table shows the effect of market concentration measures on interest checking account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Maintenance Fee		Maintenance Fee Threshold		Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
HHI	1.148		492.8		-7.058***	
	(0.961)		(858.3)		(2.631)	
GHHI		3.797***		2,758***		1.590
		(0.712)		(685.2)		(3.475)
Log Income	-6.565***	-5.880***	-2,168**	-1,656*	-12.64***	-12.52***
	(0.932)	(0.958)	(892.2)	(904.1)	(3.822)	(3.664)
Log Population	6.286***	6.116***	9,024***	8,884***	-23.99***	-24.00***
	(1.553)	(1.470)	(1,365)	(1,363)	(4.493)	(4.559)
Log(1+Market Cap)	0.0242***	0.0181***	36.90***	32.53***	0.327***	0.324***
	(0.00711)	(0.00645)	(6.316)	(6.365)	(0.0207)	(0.0185)
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	532,634	532,634	490,230	490,230	913,328	913,328
R-squared	0.704	0.705	0.582	0.583	0.752	0.752

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

Table 3.4. Panel regressions of deposit prices on index fund ownership and panel IV regressions instrumenting GHHI with index fund ownership.

(A) Panel regressions of time deposit spread percentiles on index fund ownership and panel IV regressions instrumenting GHHI with index fund ownership. This table shows the effect of index fund ownership, and the effect of the GHHI instrumented with index fund ownership, on time deposit spread percentiles with 12-, 24-, and 36-month maturities. Percentiles are calculated for each year based on spreads defined as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	12-Month CD Spread (Percentile)		24-Month CD Spread (Percentile)		36-Month CD Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
Index Fund Ownership	1.830***		2.567***		2.761***	
	(0.204)		(0.239)		(0.237)	
GHHI		38.77***		54.15***		57.72***
		(4.196)		(4.884)		(5.031)
Log Income	-24.22***	-22.54***	-11.53***	-9.191**	-7.706	-5.113
	(3.174)	(3.624)	(3.888)	(4.437)	(4.821)	(5.123)
Log Population	17.00***	19.35***	16.93***	20.18***	15.10***	18.45***
	(4.021)	(4.255)	(4.970)	(5.224)	(5.685)	(5.867)
Log(1+Market Cap)	0.228***	0.202***	0.312***	0.276***	0.354***	0.316***
	(0.0226)	(0.0226)	(0.0223)	(0.0213)	(0.0283)	(0.0262)
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	947,052	947,052	927,727	927,727	902,540	902,540
R-squared	0.674	0.672	0.662	0.658	0.672	0.667

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

(B) Panel regressions of money market account maintenance fees, thresholds, and spreads on index fund ownership and panel IV regressions instrumenting GHHI with index fund ownership. This table shows the effect of index fund ownership, and the effect of the GHHI instrumented with index fund ownership, on money market account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Maintenance Fee		Maintenance Fee Threshold		Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
Index Fund Ownership	0.111***		178.5***		0.765***	
	(0.0303)		(42.03)		(0.236)	
GHHI		2.101***		3,313***		16.20***
		(0.555)		(771.7)		(4.814)
Log Income	-1.345**	-1.178**	-953.5	-699.6	-12.36***	-11.65***
	(0.524)	(0.496)	(604.7)	(606.0)	(3.573)	(3.511)
Log Population	-0.905	-0.829	2,271***	2,355***	-7.090	-6.076
	(0.626)	(0.596)	(782.0)	(764.0)	(4.521)	(4.274)
Log(1+Market Cap)	0.0295***	0.0281***	42.40***	40.17***	0.317***	0.306***
	(0.00371)	(0.00380)	(4.539)	(4.502)	(0.0237)	(0.0243)
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	533,815	533,815	488,666	488,666	911,361	911,361
R-squared	0.795	0.795	0.530	0.529	0.655	0.655

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

(C) Panel regressions of interest checking account maintenance fees, thresholds, and spreads on index fund ownership and panel IV regressions instrumenting GHHI with index fund ownership. This table shows the effect of index fund ownership, and the effect of the GHHI instrumented with index fund ownership, on interest checking account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003 to 2013. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Maintenance Fee		Maintenance Fee Threshold		Spread (Percentile)	
	(1)	(2)	(3)	(4)	(5)	(6)
Index Fund Ownership	0.480***		258.2***		0.359	
	(0.0644)		(67.51)		(0.223)	
GHHI		8.972***		4,858***		7.592
		(1.186)		(1,270)		(4.738)
Log Income	-5.653***	-4.970***	-1,672*	-1,272	-13.50***	-13.18***
	(0.952)	(1.015)	(923.0)	(1,025)	(3.669)	(3.695)
Log Population	5.565***	5.952***	8,618***	8,793***	-26.58***	-26.13***
	(1.486)	(1.427)	(1,386)	(1,396)	(4.821)	(4.695)
Log(1+Market Cap)	0.0160**	0.00965	32.55***	29.16***	0.303***	0.298***
	(0.00692)	(0.00651)	(6.204)	(6.053)	(0.0196)	(0.0193)
Year FE	✓	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓	✓
Observations	532,634	532,634	490,230	490,230	913,328	913,328
R-squared	0.705	0.703	0.583	0.582	0.753	0.753

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

Table 3.5. Regressions of change in GHHI between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in GHHI. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Δ GHHI (1)
Top Tercile Index Fund Ownership in 2003	0.0586***
	(0.0192)
Log Income	0.00968
	(0.0266)
Log Population	0.0226***
	(0.00546)
Log(1+Market Cap)	9.17e-05
	(0.000165)
Constant	-0.329
	(0.259)
Observations	50,684
R-squared	0.198

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

Table 3.6. Regressions of change in deposit prices between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003.

(A) Regressions of change in time deposit spread percentiles between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in time deposit spread percentiles with 12-, 24-, and 36-month maturities. Percentiles are calculated for each year based on spreads defined as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level.

	Δ 12-Month CD Spread (Percentile) (1)	Δ 24-Month CD Spread (Percentile) (2)	Δ 36-Month CD Spread (Percentile) (3)
Top Tercile Index Fund Ownership in 2003	3.597***	5.594***	5.172***
	(1.066)	(0.848)	(1.058)
Log Income_{2004}	8.182***	9.651***	7.721***
	(1.854)	(1.800)	(1.903)
Log Population_{2004}	3.204***	3.932***	4.606***
	(0.510)	(0.412)	(0.462)
Log(1+Market Cap_{2004})	0.726***	0.782***	0.773***
	(0.0366)	(0.0362)	(0.0324)
Spread_{2004}	-0.659***	-0.803***	-0.774***
	(0.0148)	(0.0166)	(0.0188)
Constant	-101.9***	-121.2***	-110.3***
	(17.07)	(17.14)	(17.95)
Observations	50,684	49,429	47,930
R-squared	0.412	0.529	0.531

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

(B) Regressions of change in money market fees, thresholds, and spreads between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in money market account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level.

	Δ Maintenance Fee	Δ Maintenance Fee Threshold	Δ Spread (Percentile)
	(1)	(2)	(3)
Top Tercile Index Fund Ownership in 2003	1.158***	909.4**	3.377***
	(0.301)	(360.0)	(0.965)
Log Income$_{2004}$	-0.993	-178.3	1.774
	(0.708)	(408.1)	(1.683)
Log Population$_{2004}$	0.0149	-4.710	1.372***
	(0.0861)	(92.42)	(0.258)
Log(1+Market Cap$_{2004}$	-0.0448***	-59.67***	0.162***
	(0.00853)	(9.879)	(0.0288)
Maintenance Fee$_{2004}$	-0.417***		
	(0.0290)		
Maintenance Fee Threshold$_{2004}$		-0.422***	
		(0.0407)	
Spread$_{2004}$			-0.538***
			(0.0133)
Constant	14.05*	3,391	-12.62
	(7.412)	(3,805)	(17.00)
Observations	16,818	13,414	46,763
R-squared	0.275	0.174	0.277

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

(C) Regressions of change in interest checking fees, thresholds, and spreads between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in interest checking account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level.

	Δ Maintenance Fee	Δ Maintenance Fee Threshold	Δ Spread (Percentile)
	(1)	(2)	(3)
Top Tercile Index Fund Ownership in 2003	1.410***	2,585***	1.477
	(0.417)	(645.8)	(1.000)
Log Income$_{2004}$	1.557	480.2	0.971
	(1.380)	(929.1)	(1.858)
Log Population$_{2004}$	0.636***	-215.3	2.265***
	(0.149)	(149.0)	(0.353)
Log(Market Cap$_{2004}$)	0.172***	64.88***	0.385***
	(0.0138)	(14.21)	(0.0451)
Maintenance Fee$_{2004}$	-0.652***		
	(0.0245)		
Maintenance Fee Threshold$_{2004}$		-0.446***	
		(0.0388)	
Spread$_{2004}$			-0.523***
			(0.0135)
Constant	-16.53	-399.1	-15.88
	(14.53)	(9,615)	(18.33)
Observations	16,105	10,678	48,004
R-squared	0.350	0.100	0.254

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

Table 3.7. Regressions of change in deposit prices between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003 with bank fixed effects.

(A) Regressions of change in time deposit spread percentiles between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003 with bank fixed effects. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in time deposit spread percentiles with 12-, 24-, and 36-month maturities. Percentiles are calculated for each year based on spreads defined as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level and bank-level fixed effects absorb all across-bank variation.

	Δ 12-Month CD Spread (Percentile) (1)	Δ 24-Month CD Spread (Percentile) (2)	Δ 36-Month CD Spread (Percentile) (3)
Top Tercile Index Fund Ownership in 2003	2.158***	2.317***	2.281***
	(0.426)	(0.437)	(0.352)
Log Income_{2004}	0.580	0.608	0.0209
	(0.386)	(0.416)	(0.331)
Log Population_{2004}	-0.131	0.000697	0.193***
	(0.0815)	(0.0962)	(0.0563)
Spread_{2004}	-0.923***	-0.943***	-0.952***
	(0.00888)	(0.00813)	(0.00743)
Constant	41.83***	41.12***	45.56***
	(4.105)	(4.371)	(3.563)
Bank FE	✓	✓	✓
Observations	50,684	49,429	47,930
R-squared	0.970	0.978	0.980

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

(B) Regressions of change in money market fees, thresholds, and spreads between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003 with bank fixed effects. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in money market account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level and bank-level fixed effects absorb all across-bank variation.

	Δ Maintenance Fee	Δ Maintenance Fee Threshold	Δ Spread (Percentile)
	(1)	(2)	(3)
Top Tercile Index Fund Ownership in 2003	0.238	385.0	0.919**
	(0.178)	(377.9)	(0.461)
Log Income_{2004}	-0.0632	-49.07	1.131**
	(0.0485)	(74.39)	(0.461)
Log Population_{2004}	-0.00816	-14.32	0.309***
	(0.0107)	(20.81)	(0.0734)
Maintenance Fee_{2004}	-0.947***		
	(0.0254)		
Maintenance Fee Threshold_{2004}		-0.994***	
		(0.00305)	
Spread_{2004}			-0.867***
			(0.0127)
Constant	10.17***	3,400***	28.37***
	(0.431)	(689.5)	(4.926)
Bank FE	✓	✓	✓
Observations	16,818	13,414	46,763
R-squared	0.988	0.976	0.949

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

(C) Regressions of change in interest checking fees, thresholds, and spreads between 2004 and 2013 on whether the county is in the top or bottom tercile of index fund ownership in 2003 with bank fixed effects. This table shows the effect of an indicator variable for whether a market's index fund ownership is in the top or the bottom tercile of the distribution of index fund ownership in 2003 on the change over the period 2004-2013 in interest checking account maintenance fees, maintenance fee thresholds, and interest rate spread percentiles. Spread percentiles are calculated as the within-year percentile rank of the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch. Standard errors are clustered at the county level and bank-level fixed effects absorb all across-bank variation.

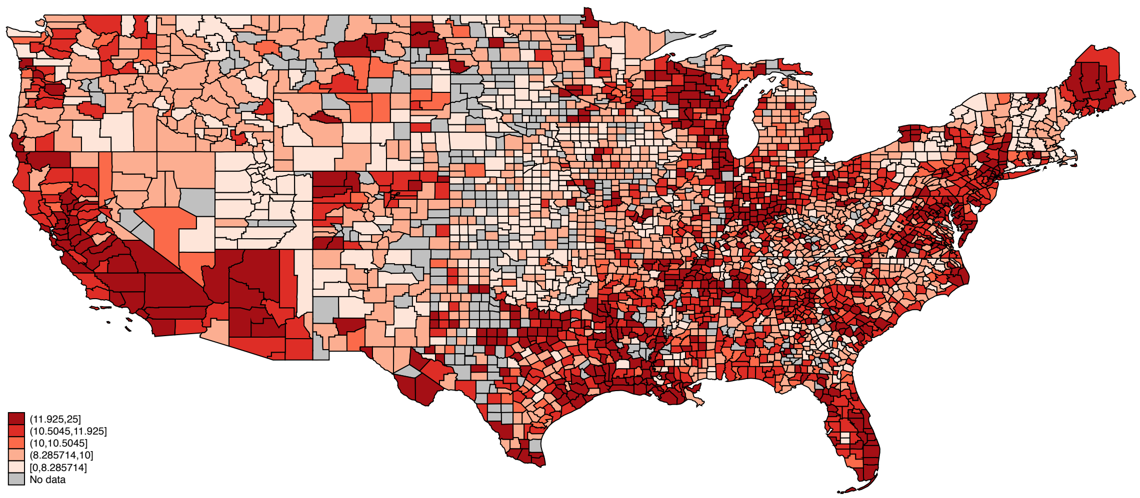
	Δ Maintenance Fee	Δ Maintenance Fee Threshold	Δ Spread (Percentile)
	(1)	(2)	(3)
Top Tercile Index Fund Ownership in 2003	0.110	406.7***	1.247***
	(0.0938)	(149.4)	(0.481)
Log Income$_{2004}$	-0.00920	-270.9	1.128
	(0.0506)	(378.1)	(0.876)
Log Population$_{2004}$	-0.0257**	-197.5**	-0.313*
	(0.0130)	(83.25)	(0.185)
Maintenance Fee$_{2004}$	-0.989***		
	(0.00341)		
Maintenance Fee Threshold$_{2004}$		-0.763***	
		(0.0533)	
Spread$_{2004}$			-0.888***
			(0.0173)
Constant	15.13***	11,019***	37.35***
	(0.506)	(3,807)	(8.665)
Bank FE	✓	✓	✓
Observations	16,105	10,678	48,004
R-squared	0.995	0.911	0.915

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

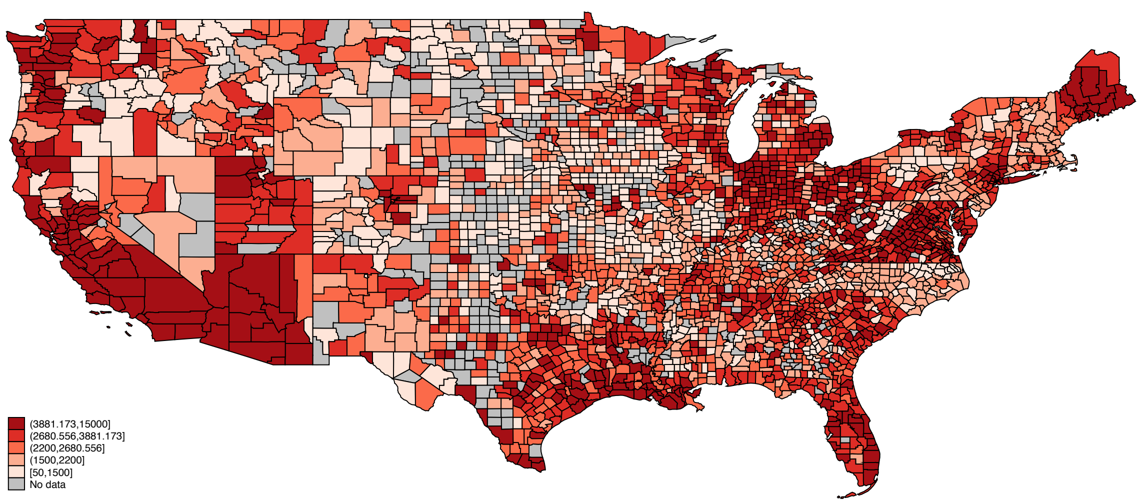
Figures

Figure 3.1. 2013 Average Prices, by County

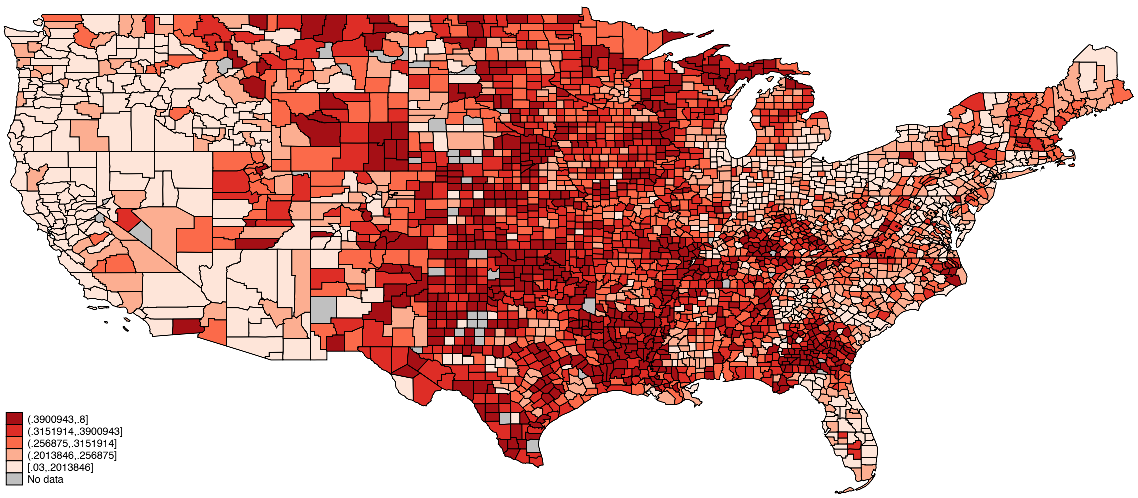
(A) Money market account maintenance fee amounts by county (2013). This figure shows the average money market account maintenance fee amount for each county in 2013. The counties in grey are counties in which RateWatch has no data on any bank branches.



(B) Money market account maintenance fee thresholds by county (2013). This figure shows the average money market account maintenance fee threshold for each county in 2013. The counties in grey are counties in which RateWatch has no data on any bank branches.



(C) 12-month CD rates by county (2013). This figure shows the average 12-month CD interest rate for each county in 2013. The counties in grey are counties in which RateWatch has no data on any bank branches.



(D) Money market account rates by county (2013). This figure shows the average money market account interest rate for each county in 2013. The counties in grey are counties in which RateWatch has no data on any bank branches.

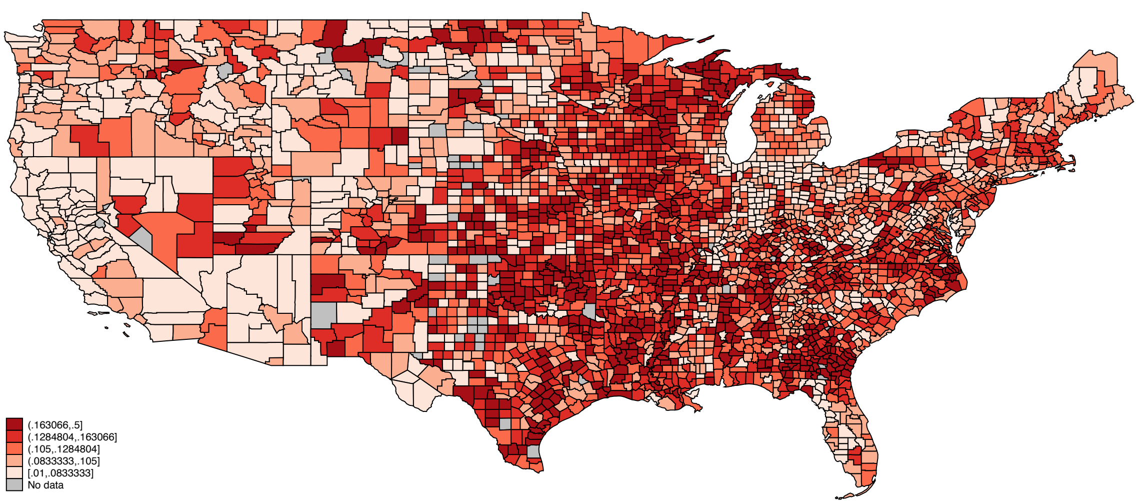
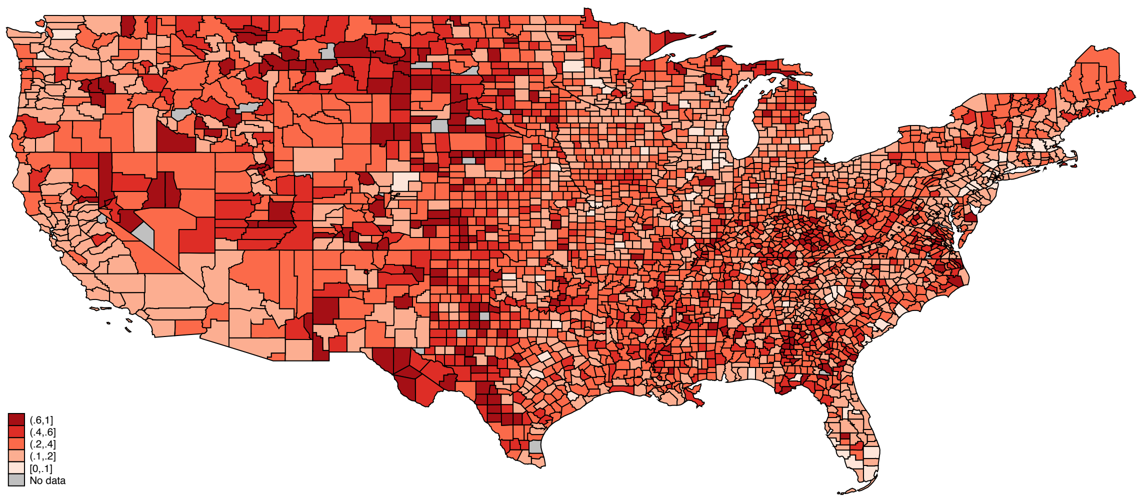
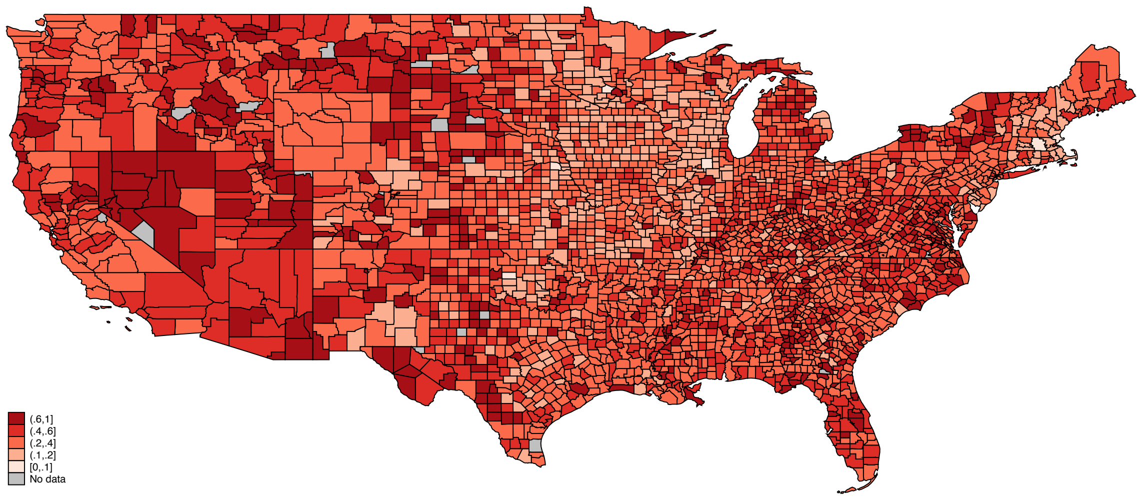


Figure 3.2. Banking market concentration, county-level (2013). This figure shows the county-level banking sector concentration in 2013, as measured using HHI and GHHI.



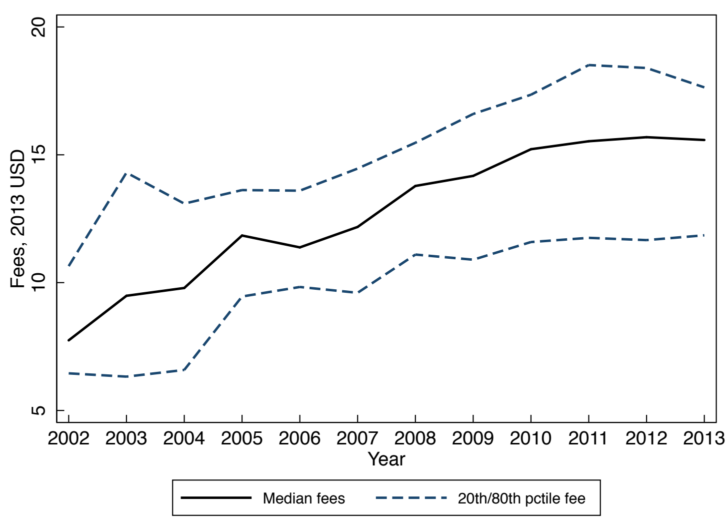
(A) HHI



(B) GHHI

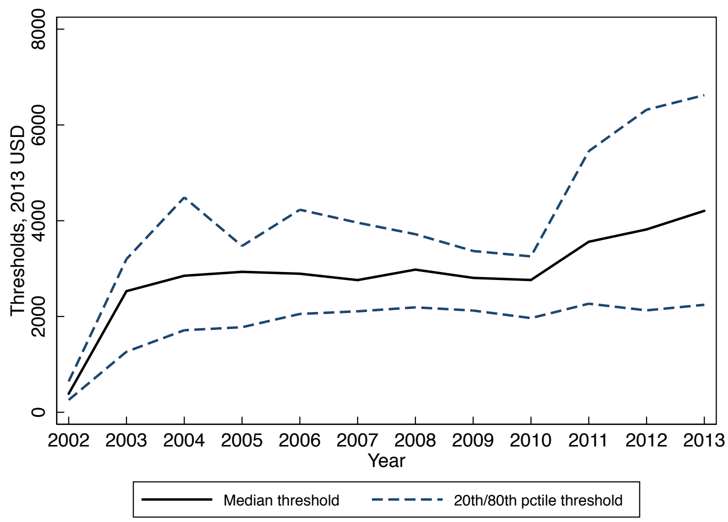
Figure 3.3. Median, 20th percentile, and 80th percentile maintenance fee amounts, 2002-2013. This figure shows the annual median, 20th percentile, and 80th percentile of maintenance fee amounts for money market accounts and interest-bearing checking accounts, for 2002-2013, in 2013 USD (adjusted for inflation using CPI).

(A) Money market accounts

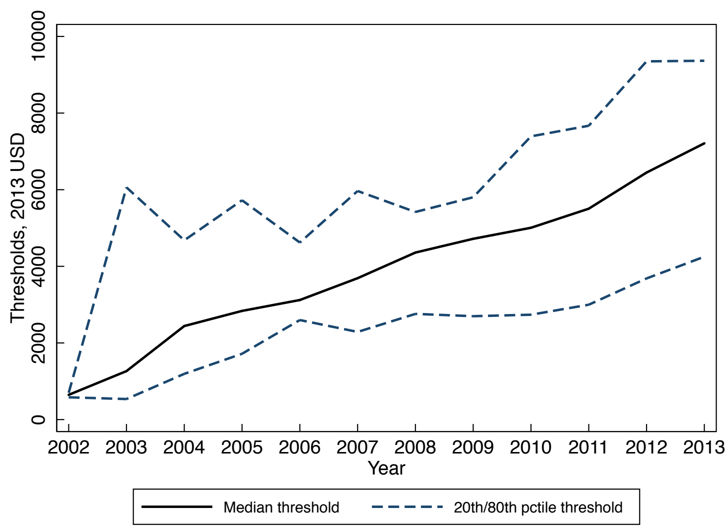


(B) Interest-bearing checking accounts

Figure 3.4. Median, 20th percentile, and 80th percentile maintenance fee thresholds, 2002-2013. This figure shows the annual median, 20th percentile, and 80th percentile of maintenance fee thresholds for money market accounts and interest-bearing checking accounts, for 2002-2013, in 2013 USD (adjusted for inflation using CPI).

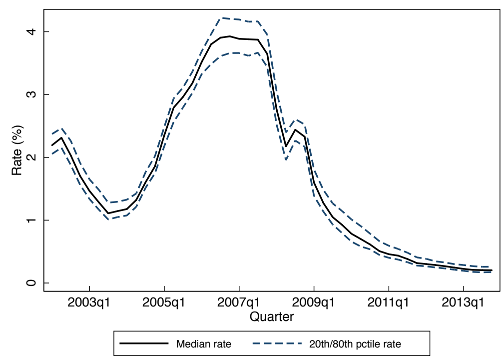


(A) Money market accounts

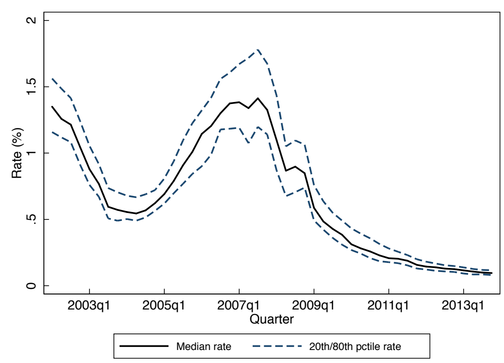


(B) Interest-bearing checking accounts

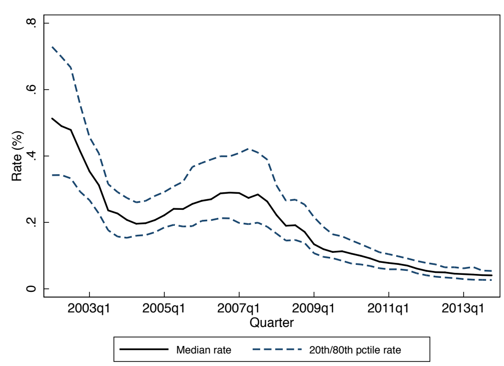
Figure 3.5. Median, 20th percentile, and 80th percentile interest rates, 2002-2013. This figure shows the quarterly median, 20th percentile, and 80th percentile of the interest rate for deposit products offered by banks from 2002 through 2013. The bank interest rates in this figure are for 12-month CDs with $10,000 minimum deposit, money market accounts, and interest-bearing checking accounts.



(A) 12-month CDs

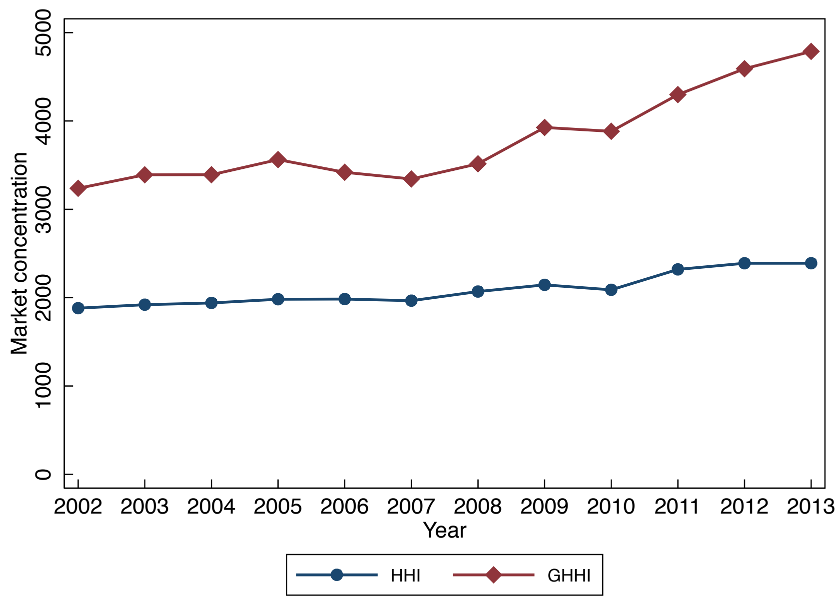


(B) Money market accounts

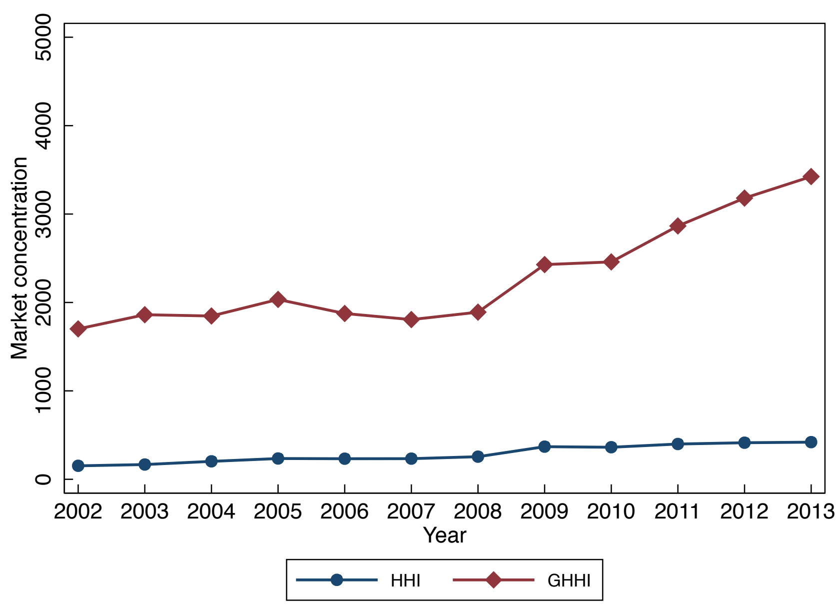


(C) Interest-bearing checking accounts

Figure 3.6. National and County-Level Bank Concentration, 2002-2013. This figure shows the annual bank concentration from 2002 through 2013 taking the entire United States as a unified market, and the deposit-weighted average across counties of bank concentration measures. Bank concentration is measured using the HHI and GHHI.

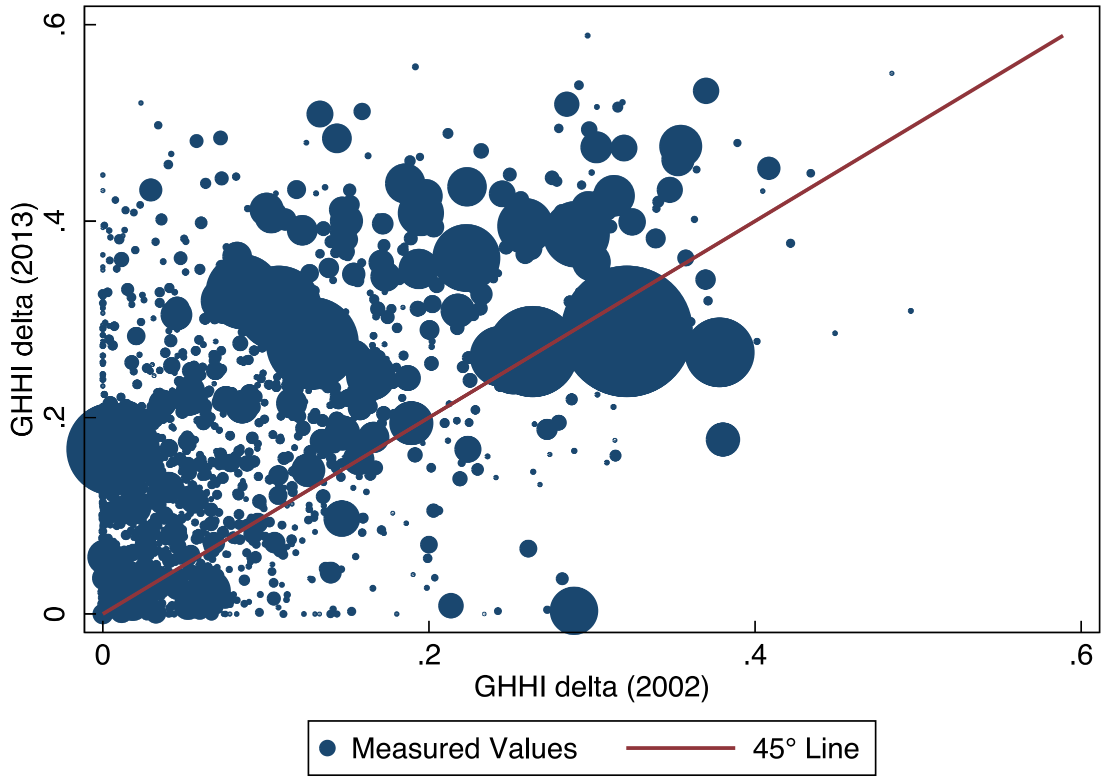


(A) County-level bank concentration



(B) National-level bank concentration

Figure 3.7. GHHI Delta County-Level Scatter, 2002 to 2013. This figure is a scatterplot of county-level GHHI deltas in 2013 against county-level GHHI deltas in 2002. The horizontal axis plots the GHHI delta of counties in 2002 and the vertical axis plots the GHHI delta of counties in 2013. The diagonal red line is a 45°, which is where all counties would lie if there was no change in GHHI delta from 2002 to 2013 in any county. Finally, the size of the plotted point signifies the total average amount of deposits in the county, with more deposits signified by a larger plotted point.



APPENDICES

APPENDIX A

VC financing and the entrepreneurship gender gap: Dataset Construction

In this section, I lay out the process used to build my dataset. As explained in Section **??**, the API access to CrunchBase prevents users from downloading detailed information on all entrepreneurial firms, financing rounds, and VC firms at once. Instead, I use VentureXpert data to find the sixteen VC firms with the greatest number of financings as of late September 2014 and then find all entrepreneurial firms in CrunchBase that ever received financing from these sixteen VC firms. This approach solves two problems at once. First, it reduces the likelihood of including organizations that are not true entrepreneurial firms in my dataset. As the definition of early entrepreneurship is vague, many "firms" in CrunchBase may be nothing more than a hobby of an "entrepreneur." Focusing on firms that receive financing at some point by a well-established VC firm removes such hobbyist projects from the sample. Second, it provides a systematic rule, devoid of subjective biases, that I can reliably use to collect data.

For each of the entrepreneurial firms in my sample, I apply the Python programs described earlier to download all data on the entrepreneurial firm object in CrunchBase and on all associated financing round and VC firm objects in CrunchBase. For instance, to download all information for Cloudera, my programs download all information for the entrepreneurial firm object for Cloudera, all financing round objects associated with Cloudera (including its Series A through F financing rounds), and all VC firm objects associated with it as well (including Accel Partners, Greylock Partners, Ignition Partners, and so on). For an example of the data files provided by CrunchBase, see Figure A.1.

Using the data from entrepreneurial firm and VC firm objects, I determine whether each person associated with a given firm is important for my analysis. For entrepreneurial firms, founders are the important personnel, so I take the list of founders provided for each entrepreneurial firm object as important personnel. For VC firms, general partners are important personnel. I take all founders and associated people whose roles signify that they are GPs as important personnel.[1]

Next, I classify all important personnel by gender. First, I categorize as many personnel as possible using my own knowledge of female and male first names. For names that I cannot categorize, I use Namepedia. Table A.1 provides a breakdown of gender classification for entrepreneurs and GPs separately and together. From that table, we can see that I manually classify 94.4% of entrepreneurs and 93.5% of GPs who are successfully classified into gender groups. About 6% and 5% of entrepreneurs and GPs, respectively, are not successfully classified either by me or by Namepedia. To classify personnel into gender group, I submit a query to Namepedia for each name and "webscrape" the gender from the response. Namepedia successfully classifies many of the names I am unable to categorize. I ignore uncertain Namepedia categorizations such as "More male, also female," "More female, also male," and "Neutral" to ensure that my results are not driven by misclassifications.

[1] I take any personnel whose role descriptions include the phrases "general partner," "principal," or "founder" to be a GP in the VC firm.

I link each entrepreneurial firm to potential exit from VC financing via IPO or acquisition. I use SEC's EDGAR filings to link to IPO exits and Thomson One's SDC M&A database to link to acquisition exits. Given the lack of shared identifiers between CrunchBase data and either SEC or SDC, I must match firms between CrunchBase and the two exit data sources manually by firm name. I do this manual matching in two parts. First, I pair observations between CrunchBase and the exits data that have perfectly matching firm names. Next, I run a Levenshtein lexical distance algorithm on all remaining pairwise combinations of CrunchBase-SEC and CrunchBase-SDC observations. I then manually go through all pairs that fall below a normalized Levenshtein distance threshold of 0.2, and find all Crunchbase-SEC and CrunchBase-SDC pairs that refer to the same company. Finally, for potential IPO exits, I manually check whether the firm filed a subsequent withdrawal (Form RW) and reversed its decision to go public at that point. If so, I remove the CrunchBase-SEC pair. By this method, I find all IPO and acquisition exits for entrepreneurial firms in my dataset.

Appendix Tables and Figures

	Entrepreneurs	GPs	All people
All	4,859	5,970	10,829
... Gender matches	4,568	5,672	10,240
... ... Manual gender match	4,313	5,306	9,619
... ... Namepedia gender match	255	366	621

Table A.1. Gender classification.
This table reports counts of gender classification of entrepreneurs and venture capitalists. Gender is classified manually by the author or, if the author is unable to classify manually, with the aid of Namepedia. The gender classification is provided for entrepreneurs and GPs separately and for all important personnel together.

```
{
 ...
 "data": {
  "uuid": "f267a617e1cefa6eb4409adb8ff09caf",
  "type": "Organization",
  "properties": {
   "role_company": true,
   ...
   "founded_on": "2008-10-13",
   "founded_on_trust_code": 7,
   "permalink": "cloudera",
   ...
   "closed_on": null,
   ...
   "founders": {
    "items": [
     {
      "type": "Person",
      "name": "Mike Olson",
      "path": "person/mike-olson-2",
      "created_at": 1410023922,
      "updated_at": 1424979282
     },
     {
      "type": "Person",
      "name": "Christophe Bisciglia",
      "path": "person/christophe-bisciglia",
      "created_at": 1224039872,
      "updated_at": 1424965495
     },
     {
      "type": "Person",
      "name": "Amr Awadallah",
      "path": "person/amr-awadallah",
      "created_at": 1218178635,
      "updated_at": 1427790738
     }
    ]
   },
   "investors": [
    {
     "type": "Organization",
     "name": "EquityZen",
     "path": "organization/equityzen"
    },
    ...
   ]
  },
  "relationships": {
   "past_team": {
    "items": [
     {
      "first_name": "Eric",
      "last_name": "Sammer",
      "title": "Engineering Manager",
      "started_on": null,
      "started_on_trust_code": null,
      "ended_on": null,
      "ended_on_trust_code": null,
      "path": "person/eric-sammer",
      "created_at": 1402581479,
      "updated_at": 1427886689
     },
     ...
    ]
   },
   "board_members_and_advisors": {
    "items": [
     {
      "first_name": "Frank",
      "last_name": "Artale",
      "title": "Advisor",
      "started_on": "2009-01-01",
      "started_on_trust_code": 4,
      "path": "person/frank-artale",
      "created_at": 1202470740,
      "updated_at": 1424993700
     },
     ...
    ]
   },
   "current_team": {
    "items": [
     {
      "first_name": "Jeff",
      "last_name": "Hammerbacher",
      "title": "Chief Scientist",
      "started_on": null,
      "started_on_trust_code": null,
      "path": "person/jeff-hammerbacher",
      "created_at": 1224039873,
      "updated_at": 1424964507
     },
     {
      "first_name": "Amr",
      "last_name": "Awadallah",
      "title": "Founder & CTO",
      "started_on": "2008-08-01",
      "started_on_trust_code": 6,
      "path": "person/amr-awadallah",
      "created_at": 1218178635,
      "updated_at": 1427790738
     },
     ...
    ]
   },
   ...
   "funding_rounds": {
    "items": [
     {
      "type": "FundingRound",
      "name": "Cloudera Funding Round",
      "path": "funding-round/948066bf45c4ea6984ee01fa54aa9676",
      "created_at": 1415820046,
      "updated_at": 1415820090
     },
     {
      "type": "FundingRound",
      "name": "Cloudera raises 0 in venture round",
      "path": "funding-round/c47e8ad2f90fcd312af1e84d89c95b52",
      "created_at": 1396448533,
      "updated_at": 1427415033
     },
     ...
    ]
   },
  }
 }
}
```


Figure A.1. *Cloudera* (entrepreneurial firm) information on CrunchBase.
This figure provides an example of the JSON file provided by CrunchBase for an entrepreneurial firm query. The data are organized into subparts in the JSON file using brackets and braces. Early stage firm data include entrepreneur and financing round information.

APPENDIX B

Ultimate ownership and bank competition: Construction of RateWatch data

RateWatch provides deposit rates, loan rates, and fees data at regular intervals for each "rate-setter." A "rate-setter" is an entity within a bank that is responsible for setting the rates and fees for one or more branches within the bank. Each bank in the data has at least one rate-setter and there may be more than one rate-setter per bank. Although not always the case, a rate-setter is generally responsible for setting banking prices for all branches within that bank within a geographic region. Additionally, the rate-setter for a given branch may change over time and the same rate-setter may not set all rates and fees for a given branch.

For each of the bank rates and fees that we explore, we convert the rate-setter-level data from RateWatch to branch-level data. We assemble these branch-level data by matching the rate or fee data series provided for each rate-setter to the bank branches for which it sets those rates and fees. This matching is possible with the use of RateWatch data that links bank branches to their rate-setters for each data series.

In the panels in Figure 3.1A, we provide evidence of the geographical coverage and dispersion for some deposit rates, loan rates, and fees. In particular, in Panel A, we show coverage and dispersion for maintenance fees on money market accounts in 2013, in Panel B, for maintenance fee thresholds on money market accounts in 2013, in Panel C, for 12-month CD annual percent yields in 2013, and in Panel D, for money market account interest rates in 2013.

APPENDIX C

Ultimate ownership and bank competition: Changes in prices and concentration over time

While our identification depends on cross-sectional variation in prices and concentration, there are trends in the time series as well. Figure C.1A shows the average maintenance fee charged for money market accounts, interest-bearing checking accounts, and non-interest-bearing checking accounts across all branches in 2002-2013. We see a clear upward trend over time in the cross-sectional average of maintenance fees charged for the three deposit products. Money market maintenance fees rise from just over \$7.50 in 2002 to over \$11 in 2013. Interest-bearing checking shows a remarkable rise in these fees, going from around \$8.50 to nearly \$15 in that same period. Non-interest bearing checking fees also increased over that period, from around \$1.50 in 2002 to nearly \$7 in 2013.[1]

We also observe that the average threshold below which maintenance fees are charged rises over time. In Figure C.1B, we see that all three deposit products' maintenance fee thresholds rise over time. The greatest rise occurs for interest-bearing checking accounts, where the threshold increases more than ten-fold from around \$650

[1]The average fees reported here and average thresholds reported in the next paragraph are in constant 2013 USD, adjusted for inflation using CPI. The fees and thresholds used in the regressions are nominal. Results are quantitatively similar using inflation-adjusted fees and thresholds.

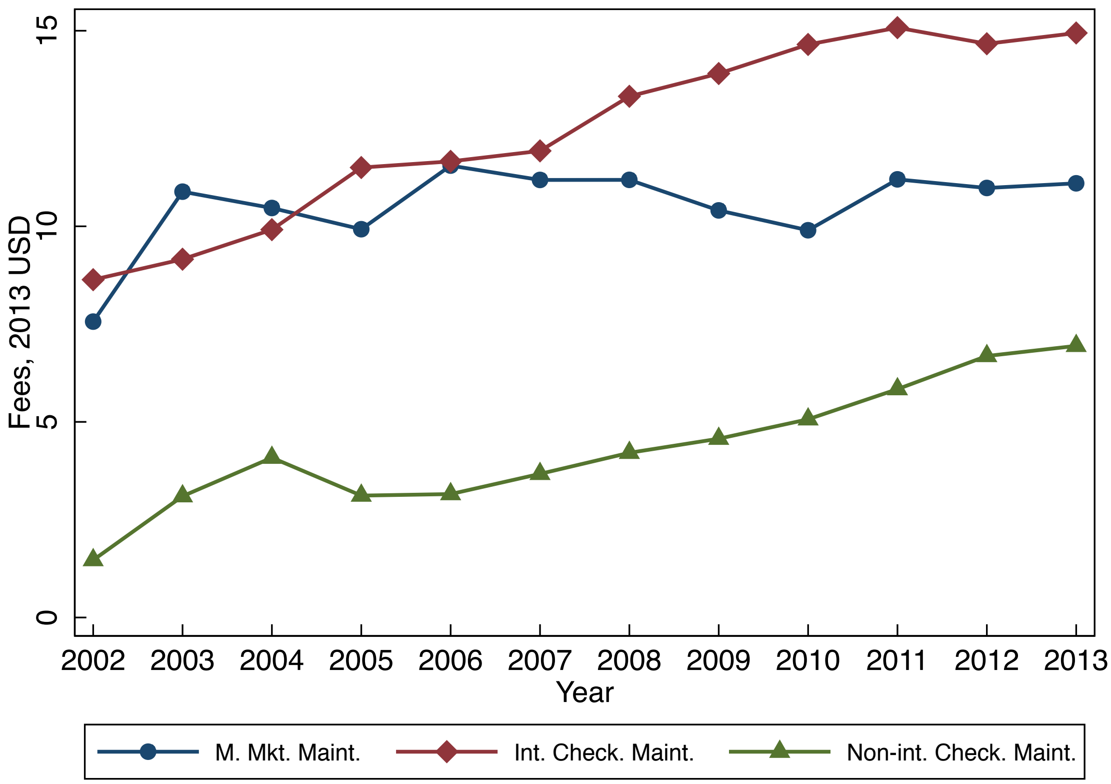
in 2002 to over \$6,800 in 2013. The other thresholds in the figure increase over time as well, with the money market account maintenance threshold rising from under \$400 to nearly \$4,400 and non-interest-bearing checking maintenance threshold rising from just over \$400 in 2002 to nearly \$1,200 in 2013. Based on these two figures, it is clear that the amount of maintenance fees and the thresholds below which they are charged have both been rising steadily over time.

Figure C.1C shows the average interest rates offered for 12-month CDs, money market accounts, and interest-bearing checking accounts as well as 10-year Treasury Constant Maturity rate in 2002 through 2013. The figure implies that, there is considerable variation over time in the spread between the Treasury rate and deposit rates on CDs and money market amounts. This spread is a measure of the margin banks charge their customers for the privilege of depositing their money with the institution. Compared to CD rates, there is less variation in the spread for interest-bearing checking account rates.

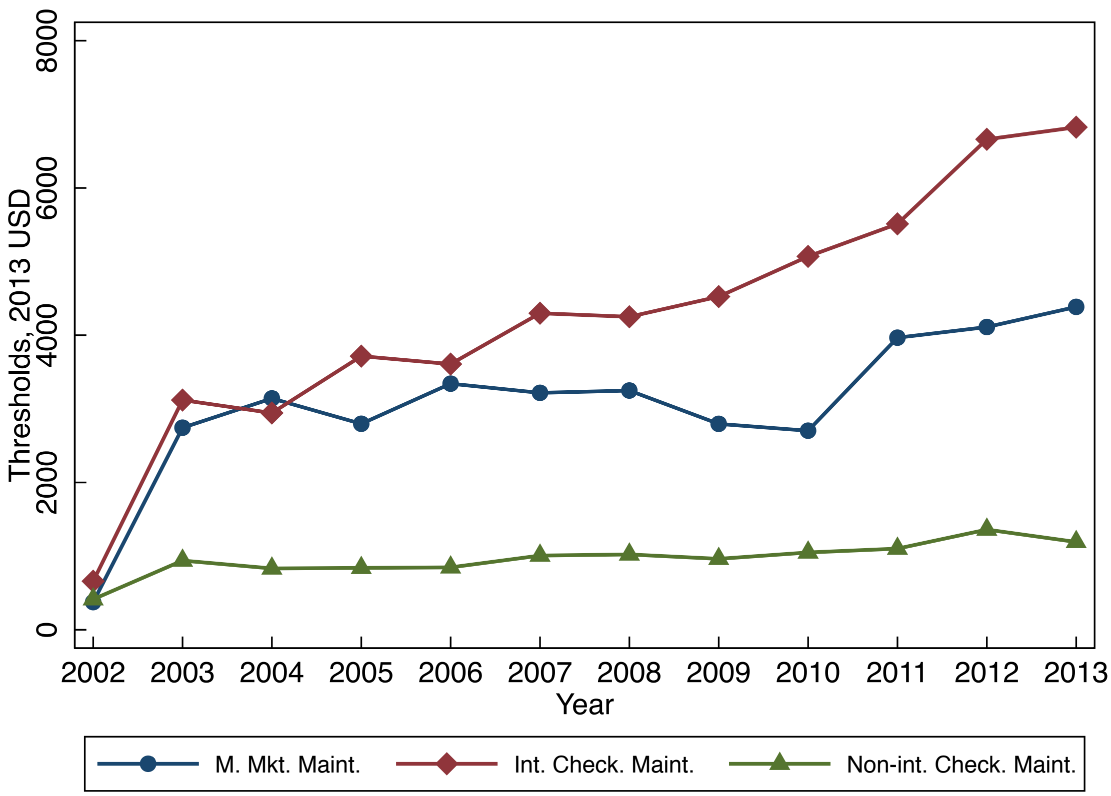
Appendix Figures

Figure C.1. Average Prices, 2002-2013

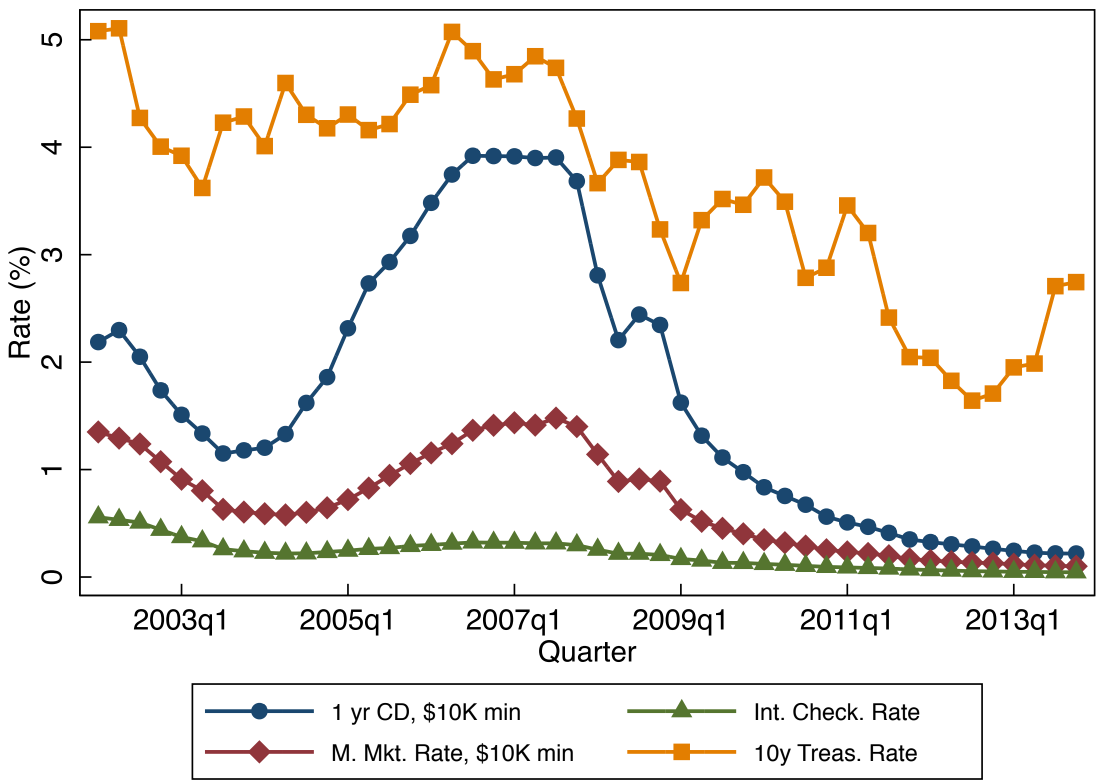
(A) Average Fees, 2002-2013. This figure shows the annual average fees charged by bank branches, from 2002 through 2013, in 2013 USD (adjusted for inflation using CPI). The fees in the figure are for money market account maintenance, interest-bearing checking account maintenance, and non-interest-bearing checking account maintenance. We average the last values for each of the above fees reported by each branch to RateWatch in a given year.



(B) Average Fee Thresholds, 2002-2013. This figure shows the annual average threshold below which maintenance fees are charged by bank branches, from 2002 through 2013, in 2013 USD (adjusted for inflation using CPI). The thresholds in this figure are for maintenance fees charged for money market accounts, interest-bearing checking accounts, and non-interest-bearing checking accounts. We average the last values for each of the above fee thresholds reported by each branch to RateWatch in a given year.



(C) Average Interest Rates, 2002-2013. This figure shows the quarterly average of the interest rate for deposit products offered by banks and the 10-year Treasury Constant Maturity rate from 2002 through 2013. The bank interest rates in this figure are for 12-month CDs with $10,000 minimum deposit, money market accounts, and interest-bearing checking accounts.



APPENDIX D

Ultimate ownership and bank competition: Panel IV first-stage regression tables

Table D.1. First stage of panel IV regressions instrumenting GHHI with index fund ownership

(A) First stage of panel IV regressions of time deposit spreads instrumenting GHHI with index fund ownership. This table shows the first stage of regressions of the effect of the GHHI instrumented with index fund ownership on time deposit spreads with 12-, 24-, and 36-month maturities. Spreads are calculated as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003M1 to 2013M6. Standard errors are clustered at the county level. Variable definitions are provided in the appendix. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Dependent Variable: GHHI		
	12-Month CD Spread	24-Month CD Spread	36-Month CD Spread
	(1)	(2)	(3)
Index Fund Ownership	0.0472***	0.0474***	0.0478***
	(0.00235)	(0.00236)	(0.00236)
Log Income	-0.0433	-0.0432	-0.0449
	(0.0306)	(0.0306)	(0.0315)
Log Population	-0.0606	-0.0601	-0.0580
	(0.0429)	(0.0434)	(0.0434)
Log(1+Market Cap)	0.000673***	0.000664***	0.000672***
	(8.44e-05)	(8.59e-05)	(8.32e-05)
Year FE	✓	✓	✓
Branch FE	✓	✓	✓
Observations	947,052	927,727	902,540
R-squared	0.911	0.911	0.912

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

(B) First stage of panel IV regressions of money market account maintenance fees, thresholds, and spreads on index fund ownership and panel IV regressions instrumenting GHHI with index fund ownership. This table shows the first stage of regressions of the effect of the GHHI instrumented with index fund ownership on money market account maintenance fees, maintenance fee thresholds, and interest rate spreads. Spreads are calculated as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003M1 to 2013M6. Standard errors are clustered at the county level. Variable definitions are provided in the appendix. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Dependent Variable: GHHI		
	Maintenance Fee	Maintenance Fee Threshold	Spread
	(1)	(2)	(3)
Index Fund Ownership	0.0531***	0.0539***	0.0473***
	(0.00277)	(0.00280)	(0.00239)
Log Income	-0.0798***	-0.0766***	-0.0438
	(0.0298)	(0.0279)	(0.0300)
Log Population	-0.0365	-0.0252	-0.0626
	(0.0592)	(0.0610)	(0.0431)
Log(1+Market Cap)	0.000668***	0.000674***	0.000658***
	(8.78e-05)	(8.82e-05)	(8.50e-05)
Year FE	✓	✓	✓
Branch FE	✓	✓	✓
Observations	533,815	488,666	911,361
R-squared	0.923	0.924	0.911

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

(C) First stage of panel IV regressions of interest checking account maintenance fees, thresholds, and spreads instrumenting GHHI with index fund ownership. This table shows the first stage of regressions of the effect of the GHHI instrumented with index fund ownership on interest checking account maintenance fees, maintenance fee thresholds, and interest rate spreads. Spreads are calculated as the difference between the 10-year Treasury Constant Maturity rate and the deposit rate, expressed as a percent of the Treasury rate. The sample includes all bank branches in RateWatch from the period 2003M1 to 2013M6. Standard errors are clustered at the county level. Variable definitions are provided in the appendix. While throughout the paper the HHI and GHHI are expressed on a scale of 0 to 10,000, we use a scale of 0 to 1 for the regressions.

	Dependent Variable: GHHI		
	Maintenance Fee	Maintenance Fee Threshold	Spread
	(1)	(2)	(3)
Index Fund Ownership	0.0535***	0.0531***	0.0473***
	(0.00283)	(0.00282)	(0.00236)
Log Income	-0.0761**	-0.0824***	-0.0425
	(0.0313)	(0.0310)	(0.0309)
Log Population	-0.0431	-0.0360	-0.0595
	(0.0623)	(0.0637)	(0.0437)
Log(1+Market Cap)	0.000708***	0.000698***	0.000668***
	(9.18e-05)	(8.37e-05)	(8.50e-05)
Year FE	✓	✓	✓
Branch FE	✓	✓	✓
Observations	532,634	490,230	913,328
R-squared	0.924	0.926	0.911

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

APPENDIX E

Ultimate ownership and bank competition: Construction of the GHHI index

For each retail banking market, we quantify the increase in market concentration from common ownership and cross-ownership using the Generalized HHI (GHHI) delta. Because some of the bank shareholders are also banks themselves, our first step in calculating the GHHI delta is solving for the ultimate financial interest and ultimate control shares of the different shareholders.[1]

We solve for ultimate financial interest as follows. Call π_j the operating profit of firm j. The overall profit of firm j, including the profit from the shares it holds in other firms is

$$\Pi_j = \pi_j + \sum_{k \neq j} d^*_{jk} \Pi_k,$$

where d^*_{jk} is the financial interest of firm j in firm k. In matrix form, the vector of overall profits is

$$\Pi = \pi + D^* \Pi,$$

[1]Our derivation of ultimate ownership and control is inspired by the work of Leontief (1941), Leontief (1966), Ellerman (1991), Gilo et al. (2006), Brito et al. (2013) and Brito et al. (2015). The main difference between our derivation and that of Brito et al. (2015) is that our methodology for calculating ultimate control shares makes the ultimate control shares add up to one for every firm, while the ultimate control shares implied by their methodology do not necessarily add up to one, and can in some cases be negative.

where D^* is the matrix of cross-financial interests of the industry firms. Solving for overall profits yields the equation

$$\Pi = (I - D^*)^{-1}\pi.$$

We define the objective function of external shareholder i as

$$U_i = \sum_k d_{ij}\Pi_j,$$

where d_{ij} is the direct financial interests of external shareholder in the (overall) profits industry firm j. Calling D the matrix of direct financial interests of external shareholders in the industry firms, the vector U of objective functions of external shareholders is

$$U = D\Pi = D(I - D^*)^{-1}\pi.$$

Thus, we define the ultimate financial interest matrix B as

$$B = D(I - D^*)^{-1}.$$

The element b_{ij} captures the ultimate financial interest of external shareholder i in the *operating* profits of firm j.

We solve for ultimate control in a similar way. The objective function of manager j can be written as

$$\omega_j = \sum_{k \neq j} c^*_{kj}\omega_k + \sum_i c_{ij}U_i$$

where c^*_{kj} is the control share of firm k in firm j, c_{ij} is the control share of external shareholder i in firm j, and U_i is the objective function of external shareholder i. We can write this in matrix form as

$$\Omega = C'^{*}\Omega + C'U,$$

where Ω is the vector of firm objective functions, C^{*} is the matrix of cross-control shares by other industry firms, C is the matrix of control shares of external shareholders.

Solving the system for Ω yields

$$\Omega = (I - C'^{*})^{-1}C'U.$$

Thus, the ultimate control shares Γ are given by

$$\Gamma = C(I - C^{*})^{-1}.$$

The element γ_{ij} captures the ultimate control share of external shareholder i in the objective function of firm j.

Note that the ultimate control shares for each firm add up to one. To see this, start from the initial control shares, which add up to one by definition:

$$C'\mathbb{1}_N + C'^{*}\mathbb{1}_K = \mathbb{1}_K$$

where $\mathbb{1}_N$ is a column vector of ones with number of rows equal to the number of external shareholders N, and $\mathbb{1}_K$ is a column vector of ones with number of rows equal to the number of industry firms. One can rewrite this as

$$C'\mathbb{1}_N = (I - C'^{*})\mathbb{1}_K$$

and then pre-multiply on both sides by $(I - C'^*)^{-1}$ to obtain

$$(I - C'^*)^{-1} C' \mathbb{1}_N = \mathbb{1}_K$$

$$\Gamma' \mathbb{1}_N = \mathbb{1}_K.$$

That is, the sum of each row of the ultimate control shares matrix Γ equals one. A similar derivation shows that ultimate financial interest shares for each firm add up to one as well.

Once we have solved for the ultimate ownership, we can apply the O'Brien and Salop (2000) formula directly to obtain the GHHI:

$$\text{GHHI} = s'Ws$$

where

$$W = diag(\Gamma' B)^{-1} \Gamma' B$$

is the matrix of weights that firms put in the profits of competitors' profits in their objective function, relative to their own profits. This formula can also be written in non-matrix form as

$$\text{GHHI} = \sum_j \sum_k \frac{\sum_i \gamma_{ij} \beta_{ik}}{\sum_i \gamma_{ij} \beta_{ij}} s_j s_k,$$

where γ_{ij} is the ultimate control share by shareholder i in firm j, and β_{ij} is the ultimate financial interest by shareholder i in firm j.

The GHHI delta, which measures the increase in market concentration due to common and cross-ownership, is the difference between the GHHI and the standard HHI:

$$\text{GHHI delta} = \text{GHHI - HHI}.$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Viral V Acharya, Thomas Cooley, and Matthew Richardson. *Manufacturing tail risk: A perspective on the financial crisis of 2007-2009*. Now Publishers Inc, 2010.

William Ackman. What happens when index funds control corporate america? *Pershing Square Capital Management Annual Shareholder Letter*, 2016.

Kenneth R. Ahern and Amy K. Dittmar. The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics*, 127(1):137–197, 2012.

Chunrong Ai and Edward C. Norton. Interaction terms in logit and probit models. *Economics Letters*, 80(1):123 – 129, 2003. ISSN 0165-1765. doi: http://dx.doi.org/10.1016/S0165-1765(03)00032-6. URL `http://www.sciencedirect.com/science/article/pii/S0165176503000326`.

Aigbe Akhigbe and James E McNulty. The profit efficiency of small us commercial banks. *Journal of Banking & Finance*, 27(2):307–325, 2003.

Alberto F. Alesina, Francesca Lotti, and Paolo Emilio Mistrulli. Do women pay more for credit? evidence from Italy. *Journal of the European Economic Association*, 11:45–66, 2013. ISSN 1542-4774. doi: 10.1111/j.1542-4774.2012.01100.x. URL `http://dx.doi.org/10.1111/j.1542-4774.2012.01100.x`.

Jason Allen, C Robert Clark, and Jean-François Houde. Market structure and the diffusion of e-commerce: Evidence from the retail banking industry. *Available at SSRN 1276336*, 2009.

Jason Allen, Robert Clark, Jean-Fran Houde, et al. The effect of mergers in search markets: Evidence from the canadian mortgage industry. *American Economic Review*, 104(10):3365–96, 2014.

Deniz Anginer, Asli Demirgüç-Kunt, and Min Zhu. How does bank competition affect systemic stability? *Journal of Financial Intermediation, Forthcoming*, 2013.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press, 2008.

Ian Appel, Todd Gormley, and Donald Keim. Passive investors, not passive owners. *Journal of Financial Economics*, forthcoming.

Kenneth Arrow. The theory of discrimination. *Discrimination in labor markets*, 3 (10):3–33, 1973.

Hadiye Aslan and Praveen Kumar. The product market effects of hedge fund activism. *Journal of Financial Economics*, 2015.

Jose Azar. *A new look at oligopoly: Implicit collusion through portfolio diversification.* PhD thesis, Princeton University, 2012.

José Azar, Martin C Schmalz, and Isabel Tecu. Anti-competitive effects of common ownership. *Ross School of Business Paper*, (1235), 2015.

John Jianqiu Bai, Daniel Carvalho, and Gordon Phillips. The impact of bank credit on labor reallocation and aggregate industry productivity. 2015.

Jonathan Baker. Overlapping financial investor ownership, market power, and antitrust enforcement: My qualified agreement with professor elhauge. *Harvard Law Review*, 2016.

Brad M. Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, pages 261–292, 2001.

Nicholas C Barberis. Psychology and the financial crisis of 2007–2008. *Financial Innovation: Too Much Or Too Little?*, page 15, 2013.

Lucian A Bebchuk, Alma Cohen, and Holger Spamann. Wages of failure: Executive compensation at bear stearns and lehman 2000-2008, the. *Yale J. on Reg.*, 27:257, 2010.

Marco Becht, Patrick Bolton, and Ailsa Röell. Corporate law and governance. *Handbook of Law and Economics*, 2:829–943, 2007.

Thorsten Beck. Bank competition and financial stability: friends or foes? *World Bank Policy Research Working Paper Series, Vol*, 2008.

Thorsten Beck, Asli Demirguc-Kunt, and Vojislav Maksimovic. Bank competition and access to finance: International evidence. *Journal of Money, Credit, and Banking*, 36(3):627–648, 2004.

Thorsten Beck, Asli Demirgüç-Kunt, and Ross Levine. Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5):1581–1603, 2006.

Thorsten Beck, Ross Levine, and Alexey Levkov. Big bad banks? the winners and losers from bank deregulation in the united states. *The Journal of Finance*, 65(5): 1637–1667, 2010.

Thorsten Beck, Olivier De Jonghe, and Glenn Schepens. Bank competition and stability: cross-country heterogeneity. *Journal of financial Intermediation*, 22(2): 218–244, 2013.

Gary S. Becker. *The Economics of Discrimination*. University of Chicago Press, second edition, 1971.

David N Beede, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms. Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11), 2011.

Andrea Bellucci, Alexander Borisov, and Alberto Zazzaro. Does gender matter in bank–firm relationships? evidence from small business lending. *Journal of Banking & Finance*, 34(12):2968–2984, 2010.

Itzhak Ben-David and Sumit Agarwal. Do loan officers' incentives lead to lax lending standards? Technical report, 2012.

Roland Bénabou. Groupthink: Collective delusions in organizations and markets. *The Review of Economic Studies*, page rds030, 2012.

Ola Bengtsson and David H. Hsu. How do venture capital partners match with startup founders? Working paper, SSRN, March 2010. URL `http://ssrn.com/abstract=1568131`.

Tobias Berg, Anthony Saunders, and Sascha Steffen. The total cost of corporate borrowing in the loan market: Don't ignore the fees. *The Journal of Finance*, 2015a.

Tobias Berg, Anthony Saunders, Sascha Steffen, and Daniel Streitz. Mind the gap: The difference between us and european loan rates. *Available at SSRN 2361066*, 2015b.

Allen N Berger. The profit-structure relationship in banking–tests of market-power and efficient-structure hypotheses. *Journal of Money, Credit and Banking*, pages 404–431, 1995.

Allen N Berger and Timothy H Hannan. The price-concentration relationship in banking. *The Review of Economics and Statistics*, pages 291–299, 1989.

Allen N Berger, Richard Joseph Rosen, Gregory F Udell, et al. The effect of market size structure on competition: The case of small business lending. Technical report, Federal Reserve Bank of Chicago, 2001.

Allen N Berger, Leora F. Klapper, and Rima Turk-Ariss. Bank competition and financial stability. *Journal of Financial Services Research*, 35:99–118, 2009.

Antje Berndt and Anurag Gupta. Moral hazard and adverse selection in the originate-to-distribute model of bank credit. *Journal of Monetary Economics*, 56(5):725–743, 2009.

Marianne Bertrand and Sendhil Mullainathan. Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy*, 111(5):1043–1075, 2003.

Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and J. *The American Economic Review*, 94(4):pp. 991–1013, 2004. ISSN 00028282. URL http://www.jstor.org/stable/3592802.

Sanjai Bhagat and Brian J Bolton. Bank executive compensation and capital requirements reform. *Available at SSRN 1781318*, 2013.

Sandra E Black and Philip E Strahan. Entrepreneurship and bank credit availability. *The Journal of Finance*, 57(6):2807–2833, 2002.

Patrick Bolton, Jose Scheinkman, and Wei Xiong. Executive compensation and short-termist behaviour in speculative markets. *The Review of Economic Studies*, 73(3): 577–610, 2006.

Patrick Bolton, Hamid Mehran, and Joel Shapiro. Executive compensation and risk taking. *Review of Finance*, page rfu049, 2015.

Audra L Boone and Joshua T White. The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, forthcoming.

Duarte Brito, Ricardo Ribeiro, and Helder Vasconcelos. Measuring unilateral effects in partial acquisitions. 2013.

Duarte Brito, António Osório, Ricardo Ribeiro, Helder Vasconcelos, et al. Unilateral effects screens for partial horizontal acquisitions: The generalized hhi and guppi. Technical report, Faculdade de Economia e Gestão, Universidade Católica Portuguesa (Porto), 2015.

Markus K Brunnermeier and Christian Julliard. Money illusion and housing frenzies. *Review of Financial Studies*, 21(1):135–180, 2008.

Markus K Brunnermeier and Stefan Nagel. Hedge funds and the technology bubble. *The Journal of Finance*, 59(5):2013–2040, 2004.

Candida G. Brush, Patricia G. Greene, and Lakshmi Balachandra andAmy E. Davis. Women entrepreneurs 2014: Bridging the gender gap in venture capital. Technical report, Arthur M. Blank Center for Entrepreneurship, Babson College, 2014.

Craig Burnside, Martin Eichenbaum, and Sergio Rebelo. Understanding booms and busts in housing markets. Technical report, National Bureau of Economic Research, 2011.

Paul S Calem and Gerald A Carlino. The concentration/conduct relationship in bank deposit markets. *The Review of Economics and Statistics*, pages 268–276, 1991.

John Y Campbell. Introduction to" asset prices and monetary policy". In *Asset Prices and Monetary Policy*, pages 1–7. University of Chicago Press, 2008.

Rodrigo Canales and Ramana Nanda. A darker side to decentralized banks: Market power and credit rationing in sme lending. *Journal of Financial Economics*, 105 (2):353 – 366, 2012.

Willard T Carleton, James M Nelson, and Michael S Weisbach. The influence of institutions on corporate governance through private negotiations: Evidence from tiaa-cref. *The Journal of Finance*, 53(4):1335–1362, 1998.

I. Casebourne, C. Davies, M. Fernandes, and N. Norman. Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic. 2012.

Claire Célérier and Adrien Matray. Unbanked households: Evidence of supply-side factors. Technical Report FIN-2014-1039, HEC Paris Research Paper, February 2014.

Nicola Cetorelli. Real effects of bank competition. *Journal of Money, Credit, and Banking*, 36(3):543–558, 2004.

Nicola Cetorelli and Philip E Strahan. Finance as a barrier to entry: Bank competition and industry structure in local us markets. *The Journal of Finance*, 61(1): 437–461, 2006.

Aaron K Chatterji and Robert C Seamans. Entrepreneurial finance, credit cards, and race. *Journal of Financial Economics*, 106(1):182–195, 2012.

Xia Chen, Jarrad Harford, and Kai Li. Monitoring: Which institutions matter? *Journal of Financial Economics*, 86(2):279 – 305, 2007.

Ing-Haw Cheng, Harrison Hong, and Jose A Scheinkman. Yesterday's heroes: Compensation and risk at financial firms. *The Journal of Finance*, 70(2):839–879, 2015.

Alex Chinco and Chris Mayer. Distant speculators and asset bubbles in the housing market. *Columbia Business School mimeo, April*, 2012.

Joao F Cocco. Portfolio choice in the presence of housing. *Review of Financial studies*, 18(2):535–567, 2005.

Iain M. Cockburn and Megan J. MacGarvie. Patents, thickets and the financing of early-stage firms: Evidence from the software industry. *Journal of Economics and Management Strategy*, 18(3):729–773, 2009. ISSN 1530-9134. doi: 10.1111/j.1530-9134.2009.00228.x. URL `http://dx.doi.org/10.1111/j.1530-9134.2009.00228.x`.

Lauren Cohen. Loyalty-based portfolio choice. *Review of Financial Studies*, 22(3): 1213–1245, 2009.

Shawn Cole, Martin Kanz, and Leora Klapper. Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers. *The Journal of Finance*, 70(2):537–575, 2015.

Robert N Collender and Sherrill Shaffer. Local bank office ownership, deposit control, market structure, and economic growth. *Journal of Banking and Finance*, 27(1): 27–57, 2003.

Financial Crisis Inquiry Commission et al. The financial crisis inquiry report. *US Government Printing Office*, 2011.

Sandrine Corvoisier and Reint Gropp. Bank concentration and retail interest rates. *Journal of Banking & Finance*, 26(11):2155–2189, 2002.

Council of Economic Advisers. Expanding opportunities for women in business. Issue brief, August 2015.

Joshua Coval, Jakub Jurek, and Erik Stafford. The economics of structured finance. *The Journal of Economic Perspectives*, pages 3–26, 2009.

Rachel Croson and Uri Gneezy. Gender differences in preferences. *Journal of Economic Literature*, 47(2):pp. 448–474, 2009. ISSN 00220515. URL http://www.jstor.org/stable/27739928.

CrunchBase. Crunchbase FAQ, 2014. URL https://info.crunchbase.com/about/faqs/.

Anthony W Cyrnak and Timothy H Hannan. Is the cluster still valid in defining banking markets? evidence from a new data source. *Antitrust Bull.*, 44:313, 1999.

Tri Vi Dang, Gary Gorton, and Bengt Holmström. Ignorance, debt and financial crises. *Yale University and Massachusetts Institute of Technology, working paper*, 2012.

David Dayen. The most important 2016 issue you don't know about. *The New Republic*, https://newrepublic.com/article/131412/important-2016-issue-dont-know, 2016.

Yuliya Demyanyk and Otto Van Hemert. Understanding the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1848–1880, 2011.

Elroy Dimson, Oğuzhan Karakaş, and Xi Li. Active ownership. *Review of Financial Studies*, forthcoming.

Itamar Drechsler, Alexi Savov, and Philipp Schnabl. The deposits channel of monetary policy. *Available at SSRN*, 2014.

Christopher Drew. Airlines under justice dept. investigation over possible collusion. *New York Times*, 1 July, 2015.

Gene Davolio. The market for borrowing stock. *Journal of financial economics*, 66 (2):271–306, 2002.

Alex Edmans and Qi Liu. Inside debt*. *Review of Finance*, 15(1):75–102, 2011.

Franklin R Edwards. Concentration in banking and its effect on business loan rates. *The Review of Economics and Statistics*, pages 294–300, 1964.

Mark Egan, Ali Hortaçsu, and Gregor Matvos. Deposit competition and financial fragility: Evidence from the us banking sector. Technical report, Working Paper, 2014.

Einer Richard Elhauge. Horizontal shareholding. *Harvard Law Review*, 109(5), 2016.

D Ellerman. Cross-ownership of corporations: A new application of input-output theory. *Metroeconomica*, 42(1):33–46, 1991.

Rüdiger Fahlenbrach and René M Stulz. Bank ceo incentives and the credit crisis. *Journal of Financial Economics*, 99(1):11–26, 2011.

Giovanni Favara and Mariassunta Giannetti. Forced asset sales and the concentration of outstanding debt: Evidence from the mortgage market. 2015.

Theodore G Flechsig. The effect of concentration on bank loan rates. *The Journal of Finance*, 20(2):298–311, 1965.

Stephen Foley and Ben McLannahan. Top us financial groups hold secret summits on long-termism. *http://www.ft.com/intl/cms/s/0/3e6a1b90-c8fb-11e5-be0b-b7ece4e953a0.html*, February 2016.

Christopher L Foote, Kristopher S Gerardi, and Paul S Willen. Why did so many people make so many ex post bad decisions? the causes of the foreclosure crisis. Technical report, National Bureau of Economic Research, 2012.

Mark J Garmaise and Tobias J Moskowitz. Bank mergers and crime: The real and social effects of credit market competition. *The Journal of Finance*, 61(2):495–538, 2006.

Nicola Gennaioli, Andrei Shleifer, and Robert Vishny. Neglected risks, financial innovation, and financial fragility. *Journal of Financial Economics*, 104(3):452–468, 2012.

Nicola Gennaioli, Andrei Shleifer, and Robert W Vishny. A model of shadow banking. *The Journal of Finance*, 68(4):1331–1363, 2013.

Kristopher Gerardi, Andreas Lehnert, Shane M Sherlund, and Paul Willen. Making sense of the subprime crisis. *Brookings Papers on Economic Activity*, 2008(2): 69–159, 2008.

Andra C Ghent and Marianna Kudlyak. Recourse and residential mortgage default: evidence from us states. *Review of Financial Studies*, page hhr055, 2011.

R Alton Gilbert and Adam M Zaretsky. Banking antitrust: Are the assumptions still valid? *REVIEW-FEDERAL RESERVE BANK OF SAINT LOUIS*, 85(6):29–52, 2003.

Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438:900–901, December 2005. doi: 10.1038/438900a.

David Gilo. The anticompetitive effect of passive investment. *Michigan Law Review*, pages 1–47, 2000.

David Gilo, Yossi Moshe, and Yossi Spiegel. Partial cross ownership and tacit collusion. *The RAND Journal of Economics*, 37(1):81–99, 2006.

David Gladstone and Laura Gladstone. *Venture Capital Handbook: An Entrepreneur's Guide to Raising Venture Capital*. FT Press, 2002.

Edward L Glaeser. A nation of gamblers: Real estate speculation and american history. *The American Economic Review*, 103(3), 2013.

Claudia Goldin and Cecilia Rouse. Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4):715–741, 2000.

Paul Gompers and Josh Lerner. *The money of invention: How venture capital creates new wealth*. Harvard Business School Press, Boston, MA, 2001.

Paul Gompers, Anna Kovner, Josh Lerner, and David Scharfstein. Performance persistence in entrepreneurship. *Journal of Financial Economics*, 96(1):18 – 32, 2010. ISSN 0304-405X. doi: http://dx.doi.org/10.1016/j.jfineco.2009.11.001. URL `http://www.sciencedirect.com/science/article/pii/S0304405X09002311`.

Paul Gompers, Vladimir Mukharlyamov, and Yuhai Xuan. The cost of friendship. Working Paper 18141, NBER, June 2012. URL `http://www.nber.org/papers/w18141`.

Paul Gompers, Vladimir Mukharlyamov, Emily Weisburst, and Yuhai Xuan. Gender effects in venture capital. Working paper 2445497, SSRN, 2014.

Paul Alan Gompers and Joshua Lerner. *The venture capital cycle*. MIT press, 2004.

Roger H Gordon. Do publicly traded corporations act in the public interest? Technical report, National Bureau of Economic Research, 1990.

Michael Gorman and William A. Sahlman. What do venture capitalists do? *Journal of Business Venturing*, 4(4):231 – 248, 1989. ISSN 0883-9026. doi: http://dx.doi.org/10.1016/0883-9026(89)90014-1. URL http://www.sciencedirect.com/science/article/pii/0883902689900141.

Todd A. Gormley, Nandini Gupta, and Anand Jha. Quiet life no more? corporate bankruptcy and bank competition. Technical report, SSRN, December 2015.

Robin Greenwood and David Scharfstein. The growth of finance. *The Journal of Economic Perspectives*, 27(2):3–28, 2013. ISSN 08953309. URL http://www.jstor.org/stable/23391688.

Hendrik Hakenes and Isabel Schnabel. Capital regulation, bank competition, and financial stability. *Economics Letters*, 113(3):256–258, 2011.

Timothy H Hannan. Bank commercial loan markets and the role of market structure: Evidence from surveys of commercial lending. *Journal of Banking & Finance*, 15 (1):133–149, 1991.

Timothy H Hannan. Market share inequality, the number of competitors, and the hhi: An examination of bank pricing. *Review of Industrial Organization*, 12(1): 23–35, 1997.

Timothy H Hannan and Allen N Berger. The rigidity of prices: Evidence from the banking industry. *The American Economic Review*, pages 938–945, 1991.

Timothy H Hannan and J Nellie Liang. The influence of thrift competition on bank business loan rates. *Journal of Financial Services Research*, 9(2):107–122, 1995.

Timothy H Hannan and Robin A Prager. The competitive implications of multimarket bank branching. *Journal of Banking & Finance*, 28(8):1889–1914, 2004.

Maryam Haque. Nvca yearbook 2016. http://www.nvca.org/?ddownload=2963, March 2016.

Robert S. Harris, Tim Jenkinson, and Steven N. Kaplan. Private equity performance: What do we know? *The Journal of Finance*, 69(5):1851–1882, 2014. ISSN 1540-6261. doi: 10.1111/jofi.12154. URL http://dx.doi.org/10.1111/jofi.12154.

Erik Heitfield and Robin A Prager. The geographic scope of retail deposit markets. *Journal of Financial Services Research*, 25(1):37–55, 2004.

Charles Himmelberg, Christopher Mayer, and Todd Sinai. Assessing high house prices: Bubbles, fundamentals, and misperceptions. Technical report, National Bureau of Economic Research, 2005.

Yael V. Hochberg, Alexander Ljungqvist, and Yang Lu. Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, 62 (1):251–301, 2007.

Yael V. Hochberg, Alexander Ljungqvist, and Yang Lu. Networking as a barrier to entry and the competitive supply of venture capital. *The Journal of Finance*, 65 (3):829–859, 2010.

Chang-Tai Hsieh, Erik Hurst, Charles I. Jones, and Peter J. Klenow. The allocation of talent and u.s. economic growth. Working Paper 18693, NBER, January 2013. URL `http://www.nber.org/papers/w18693`.

David H. Hsu. What do entrepreneurs pay for venture capital affiliation? *The Journal of Finance*, 59(4):1805–1844, 2004.

Jiekun Huang and Darren J. Kisgen. Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, 108(3):822–839, 2013.

Charles Kahn, George Pennacchi, and Ben Sopranzetti. Bank deposit rate clustering: Theory and empirical evidence. *Journal of Finance*, pages 2185–2214, 1999.

Charles Kahn, George Pennacchi, and Ben Sopranzetti. Bank consolidation and the dynamics of consumer loan interest rates*. *The Journal of Business*, 78(1):99–134, 2005.

Eugene Kandel, Konstantin Kosenko, Randall Morck, and Yishay Yafeh. The great pyramids of america: A revised history of us business groups, corporate ownership and regulation, 1930-1950. *NBER Working Paper*, 2013.

Steven N. Kaplan and Antoinette Schoar. Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance*, 60(4):1791–1823, 2005.

Steven N. Kaplan and Per Strömberg. Financial contracting theory meets the real world: An empirical analysis of venture capital contracts. *The Review of Economic Studies*, 70(2):281–315, 2003. doi: 10.1111/1467-937X.00245. URL `http://restud.oxfordjournals.org/content/70/2/281.abstract`.

Steven N. Kaplan, Berk A. Sensoy, and Per Strömberg. Should investors bet on the jockey or the horse? evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance*, 64(1):75–115, 2009. ISSN 1540-6261.

Matt Kaufman. CrunchBase, People+, and the EFF, 2013. URL `https://info.crunchbase.com/2013/11/crunchbase-people-and-the-eff/`.

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

Benjamin J Keys, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. Financial regulation and securitization: Evidence from subprime loans. *Journal of Monetary Economics*, 56(5):700–720, 2009.

Benjamin J Keys, Tanmoy Mukherjee, Amit Seru, Vikrant Vig, and Robert W Kolb. Did securitization lead to lax screening? evidence from subprime loans. *Lessons from the Financial Crisis: Causes, Consequences, and Our Economic Future*, pages 217–224, 2010.

Benjamin J Keys, Amit Seru, and Vikrant Vig. Lender screening and the role of securitization: evidence from prime and subprime mortgage markets. *Review of Financial Studies*, page hhs059, 2012.

Christopher R Knittel and Victor Stango. Price ceilings as focal points for tacit collusion: evidence from credit cards. *American Economic Review*, pages 1703–1729, 2003.

Donghoon Lee, Christopher J Mayer, and Joseph Tracy. A new look at second liens. Technical report, National Bureau of Economic Research, 2012.

Wassily Leontief. Structure of american economy, 1919-1929. *Harvard University Press*, 1941.

Wassily Leontief. Input-output economics. *Oxford University Press*, 1966.

Joshua Lerner. Venture capitalists and the decision to go public. *Journal of Financial Economics*, 35(3):293 – 316, 1994. ISSN 0304-405X. doi: http://dx.doi.org/10.1016/0304-405X(94)90035-3. URL `http://www.sciencedirect.com/science/article/pii/0304405X94900353`.

Michael Lewis. *The big short: Inside the doomsday machine*. WW Norton & Company, 2011.

Jules Lichtenstein. Issue brief: Demographic characteristics of business owners. Technical Report 2, US SBA Office of Advocacy, January 2014.

Jaegeum Lim and Jonathan Meer. The impact of teacher-student gender matches: Random assignment evidence from south korea. Working Paper 21407, NBER, July 2015. URL `http://www.nber.org/papers/w21407`.

Inessa Love and María Soledad Martínez Pería. How bank competition affects firms' access to finance. *The World Bank Economic Review*, page lhu003, 2014.

Dan Marom, Alicia Robb, and Orly Sade. Gender dynamics in crowdfunding (Kickstarter): Evidence on entrepreneurs, investors, deals and taste-based discrimination. Working paper, 2015.

David Martinez-Miera and Rafael Repullo. Does competition reduce the risk of bank failure? *Review of Financial Studies*, 23(10):3638–3664, 2010.

Christopher Mayer and Robert J Shiller. [bubble, bubble, where's the housing bubble?]. comments and discussion. *Brookings Papers on Economic Activity*, pages 51–67, 2006.

Joseph McCahery, Laura Starks, and Zacharias Sautner. Behind the scenes: The corporate governance preferences of institutional investors. *Journal of Finance*, forthcoming.

Brian T. Melzer and Donald P. Morgan. Competition and adverse selection in the small-dollar loan market: Overdraft versus payday credit. *NY Fed Staff Report no. 391*, 2014.

Atif Mian and Amir Sufi. The consequences of mortgage credit expansion: Evidence from the 2007 mortgage default crisis. Technical report, National Bureau of Economic Research, 2008.

Mark Mitchell, Todd Pulvino, and Erik Stafford. Limited arbitrage in equity markets. *Journal of Finance*, 57(2), 2002.

Robert R Moore, Thomas F Siems, and Richard S Barr. *Concentration, technology, and market power in banking: Is distance dead?* Federal Reserve Bank of Dallas, 1998.

Serge Moresi, Steven C Salop, and Yianis Sarafidis. A model of ordered bargaining with applications. *Available at SSRN 1287224*, 2008.

William Mullins. The governance impact of index funds: Evidence from regression discontinuity. *Work. Pap., Sloan Sch. Manag., Mass. Inst. Technol*, 2014.

Namepedia. About Namepedia, 2015. URL `http://namepedia.org/en/about/`.

Julie A Nelson. Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. *Journal of Economic Surveys*, 29(3):566–585, July 2015.

David Neumark and Steven A Sharpe. Market structure and the nature of price rigidity: evidence from the market for consumer deposits. *The Quarterly Journal of Economics*, pages 657–680, 1992.

Carol Ann Northcott et al. *Competition in banking: A review of the literature*. Bank of Canada Working Paper, 2004.

Daniel P O'Brien and Steven C Salop. Competitive effects of partial ownership: Financial interest and corporate control. *Antitrust Law Journal*, pages 559–614, 2000.

Francois Ortalo-Magne and Sven Rady. Housing market dynamics: On the contribution of income shocks and credit constraints. *The Review of Economic Studies*, 73 (2):459–485, 2006.

Valentina Paredes. A teacher like me or a student like me? role model versus teacher bias effect. *Economics of Education Review*, 39:38 – 49, 2014. ISSN 0272-7757. doi: http://dx.doi.org/10.1016/j.econedurev.2013.12.001. URL `http://www.sciencedirect.com/science/article/pii/S0272775713001684`.

Mitchell A Petersen and Raghuram G Rajan. The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics*, pages 407–443, 1995.

Steven J Pilloff and Stephen A Rhoades. Structure and profitability in banking markets. *Review of Industrial Organization*, 20(1):81–98, 2002.

Tomasz Piskorski, Amit Seru, and James Witkin. Asset quality misrepresentation by financial intermediaries: evidence from the rmbs market. *The Journal of Finance*, 2015.

Melanie Powell and David Ansic. Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18 (6):605–628, 1997.

Robin A Prager and Timothy H Hannan. Do substantial horizontal mergers generate significant price effects? evidence from the banking industry. *Journal of Industrial Economics*, pages 433–452, 1998.

LW Punt and MCJ Van Rooij. The profit-structure relationship, efficiency and mergers in the european banking industry: an empirical assessment. *Research Memorandum WO&E*, 604:1–36, 1999.

Manju Puri and Rebecca Zarutskie. On the life cycle dynamics of venture-capital- and non-venture-capital-financed firms. *The Journal of Finance*, 67(6):2247–2293, 2012. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2012.01786.x. URL `http://dx.doi.org/10.1111/j.1540-6261.2012.01786.x`.

Amiyatosh Purnanandam. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1881–1915, 2011.

Raghuram G Rajan. Has finance made the world riskier? *European Financial Management*, 12(4):499–533, 2006.

Raghuram G Rajan. *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton University Press, 2011.

Uday Rajan, Amit Seru, and Vikrant Vig. The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, 115(2):237–260, 2015.

Stephen A Rhoades. Market share inequality, the hhi, and other measures of the firm-composition of a market. *Review of Industrial Organization*, 10(6):657–674, 1995.

Alicia Robb, Susan Coleman, and Dane Stangler. Sources of economic hope: Women's entrepreneurship. Working paper, SSRN, November 2014.

Julio Rotemberg. Financial transaction costs and industrial performance. 1984.

Martin Ruckes. Bank competition and credit standards. *Review of Financial Studies*, 17(4):1073–1102, 2004.

David Scharfstein and Adi Sunderam. Market power in mortgage lending and the transmission of monetary policy. Technical report, working paper. Spread of mortgage rate to FHMLC survey rate, 2014.

Martin Schmalz. How passive funds prevent competition. *ericposner.com/martin-schmalz-how-passive-funds-prevent-competition/*, 2015.

Victor Sewell. Civilian labor force by sex, 1970-2012, October 2013. URL `http://www.dol.gov/wb/stats/Civilian_labor_force_sex_70_12_txt.htm`.

Steven A Sharpe. The effect of consumer switching costs on prices: a theory and its application to the bank deposit market. *Review of Industrial Organization*, 12(1): 79–94, 1997.

Robert Shiller. Commentary: Bubble, bubble, wheres the housing bubble. *Brookings Papers on Economic Activity*, 2006:1–59, 2006.

Robert J Shiller. Understanding recent trends in house prices and home ownership. Technical report, National Bureau of Economic Research, 2007.

Robert J Shiller. Derivatives markets for home prices. Technical report, National Bureau of Economic Research, 2008.

Michael Simkovic. Competition and crisis in mortgage securitization. *Ind. LJ*, 88: 213, 2013.

Todd Sinai and Nicholas S Souleles. Owner-occupied housing as a hedge against rent risk*. *The Quarterly journal of economics*, 120(2):763–789, 2005.

Margaret Hwang Smith and Gary Smith. Bubble, bubble, where's the housing bubble? *Brookings Papers on Economic Activity*, 2006(1):1–67, 2006.

Cindy K Soo. Quantifying animal spirits: news media and sentiment in the housing market. *The Stephen M. Ross School of Business at the University of Michigan Research Paper Series, Forthcoming*, 2013.

Geoffrey Tate and Liu Yang. Female leadership and gender equity: Evidence from plant closure. *Journal of Financial Economics*, 2014.

John Tierney. Daring to discuss women in science. *The New York Times*, June 7 2010. URL `http://www.nytimes.com/2010/06/08/science/08tier.html`.

Anne S. Tsui, Charles A. O'Reilly, et al. Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal*, 32(2):402–423, 1989.

J Wei, AA Nelson, and E Vytlacil. Liars loan? effects of origination channel and information falsification on mortgage delinqueny. *Columbia Business School Unpublished Manuscript*, 2011.

Rui Yao and Harold H Zhang. Optimal consumption and portfolio choices with risky housing and borrowing constraints. *Review of Financial studies*, 18(1):197–239, 2005.