The London School of Economics and Political Science

Essays on Household Finance, Venture Capital, and Labor

Zhongchen Hu

A thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of *Doctor of Philosophy*, March 2021

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 37,496 words.

Statement of conjoint work

I confirm that Chapter 3 was jointly co-authored with Ashwini Agrawal and Isaac Hacamo. I contributed 33% of this work.

Acknowledgement

I am forever grateful to my advisors Dirk Jenter and Ashwini Agrawal, for their generous support, trust, and inspiration. My countless interactions with them have significantly improved my academic work, and their influence on my thinking pervades every chapter of this thesis. I could not possibly have attempted this journey without their guidance. Thank you, Dirk and Ashwini! You are truly my role models for both research and life.

I thank Daniel Paravisini and Cameron Peng for their invaluable advice, as well as Ulf Axelson, Vicente Cuñat, Christian Julliard, Dong Lou, Ian Martin, Moqi Xu, Hongda Zhong, and many others for the numerous insightful interactions that have made my time at the LSE so immensely instructive. A special mention goes to Juanita González-Uribe for her continuous and invaluable support, which has greatly helped me in shaping my research agenda. I would also like to thank my co-authors, Isaac Hacamo and Peter Koudijs, for having taught me much more than I have ever imagined.

The six years at the LSE would not have been the same without the great friends I met during my PhD. In particular, I would like to thank my cohort members who have made my time at the LSE very special and memorable: Andreea Englezu, Bruce Iwadate, Francesco Nicolai, Marco Pelosi, Simona Risteska, Karamfil Todorov, and Yue Yuan. I am also grateful to other fellow PhD students: Lorenzo Bretscher, Juan Chen, Fabrizio Core, Brandon Han, Jiantao Huang, Jesus Gorrin, James Guo, Lukas Kremens, Olga Obizhaeva, Dimitris Papadimitriou, Alberto Pellicioli, Bernardo Ricca, Gosia Ryduchowska, Ran Shi, Su Wang, and Xiang Yin, for their friendship and support during all these years.

My deepest gratitude is to my family. I am indebted to my parents, Jinping Hu and Xiumei Zhong, for their unconditional life-long support that has enabled me to pursue this path. Most importantly, I thank my beloved wife, Qianru Du. For the past 12 years, she has always been there for me since the very beginning of my undergraduate study when I started my PhD dream. Without her everlasting love, endless encouragement, and abundant patience, I would never have gotten to where I am today. Finally, I dedicate this work to my newborn daughter, Muxi Hu. She is my greatest inspiration.

Abstract

In the first chapter, I study households' insurance decisions against flood risk. Flooding is the most costly natural disaster in the US, yet policymakers are puzzled by the low take-up for flood insurance. I argue that households are affected by the low salience of flood risk. Leveraging novel transaction-level data, I use two empirical strategies to support my hypothesis. First, I exploit a staggered campaign that publicizes already freely-available flood risk information. Insurance purchases increase by 30.6% in response. Second, I exploit salient flood events shared through social media. Households purchase significantly more insurance after their geographically distant peers experience floods. My results suggest that behavioral frictions have a major impact on households' insurance decisions.

Chapter two studies the role of venture capitalists (VCs) in the labor market for entrepreneurs. There is an ongoing debate on whether VCs bet on ideas or founders. Prior studies find that successful startups often have kept businesses stable but replaced founders; however, practitioners see founders as more critical. This paper aims to rationalize the two views. I analyze new hand-collected data and find that VCs redeploy entrepreneurs across portfolio companies, highlighting VCs' emphasis on human capital. I propose that VCs utilize private information about founders, and I show: (1) former VC partners continue influencing founders' mobility; (2) the redeployment positively predicts VC performance; (3) the redeployment is stronger where information is more asymmetric.

In the third chapter (co-authored with Ashwini Agrawal and Isaac Hacamo), we find that rank-and-file labor flows can be used to predict abnormal stock returns. Rank-and-file employees are becoming increasingly critical for many firms, yet we know little about how their employment dynamics matter for stock prices. We analyze new data from the individual CV's of public company employees and find that rank-and-file labor flows can be used to predict abnormal stock returns. Accounting data and survey evidence indicate that workers' labor market decisions reflect information about future corporate earnings. Investors, however, do not appear to fully incorporate this information into their earnings expectations. The findings support the hypothesis that rank-and-file employees' entry and exit decisions convey valuable insight into their employers' future stock performance.

Contents

List of Figures

List of Tables

Chapter 1

Salience and Households' Flood Insurance Decisions

Zhongchen Hu[1]

1.1 Introduction

Flooding is the most costly natural disaster in the US and is drawing increasing attention in public policy debates.[2] As flood risk materializes with low frequency but has catastrophic consequences for households' asset values and welfare, insurance is crucial for households to hedge this tail risk. However, the take-up rate of flood insurance is surprisingly low in the US, even though the government has spent $36.5 billion on subsidy to encourage flood insurance purchases.[3]

Yet why would flood insurance be underutilized by households? I test the

[1]I am thankful for the advice of Ashwini Agrawal, Fabrizio Core, Andreas Fagereng, Juanita González-Uribe, Isaac Hacamo, Bob Hartwig, Dirk Jenter, Peter Koudijs, Dong Lou, Diogo Mendes, Greg Niehaus, Daniel Paravisini, Cameron Peng, Paolo Sodini, Jan Starmans, Per Stromberg, and Su Wang. I am also grateful to seminar participants at the London School of Economics, University of South Carolina, Chinese University of Hong Kong, Chinese University of Hong Kong (Shenzhen), University of Toronto, Stockholm School of Economics, and BI Norwegian Business School, for helpful comments.

[2]According to the National Oceanic and Atmospheric Administration, in the 2010s, flood-related events caused losses of $658 billion, substantially outpacing other natural disasters.

[3]The Insurance Information Institute 2016 Survey suggests that only 12% of US households had flood insurance. Even in flood-prone areas, the take-up rate is as low as 30% (Kousky et al., 2018). Consistently, only 17% of the flooded homes by Hurricane Harvey were insured. Similar concerns have been documented for other insurance products and in other countries (Cole et al., 2013; Karlan et al., 2014; Banerjee et al., 2019; Finkelstein et al., 2019).

hypothesis that households do not buy flood insurance partly because the risk is not fully salient to them. I define salience as the phenomenon when one's attention is disproportionally directed to one portion of the environment and the information it contains (Taylor and Thompson, 1982). Models of salience and heuristics (Tversky and Kahneman, 1973, 1974; Bordalo et al., 2012, 2013; Kőszegi and Szeidl, 2013) suggest that the low salience of flood risk (in normal times) might lead to underestimation and low insurance take-up.

Identifying a causal effect of salience on insurance demand requires overcoming several challenges. First, the salience of flood risk is difficult to measure, as we do not observe households' attention. Second, the determinants of insurance demand—such as information about flood risk, actual risk, and attention—often move simultaneously, making it difficult to isolate the key driver. For example, in the aftermath of a flood, salience likely increases, but so may the actual flood risk, due to damage to infrastructure or to changes in geological conditions. Third, as the equilibrium price and quantity of insurance transactions are jointly determined by supply and demand, it is difficult to distinguish shifts in the supply curve from shifts in demand.

I overcome these challenges by using data from the National Flood Insurance Program (NFIP) and two quasi-experiments. The NFIP was created by the US Congress in 1968 to provide affordable household flood insurance. I obtain over 50 million transaction-level observations from January 2009 to August 2019, including information on policy start and end dates, premiums, coverage, house characteristics, and locations.

The NFIP is ideal for measuring shifts in insurance demand for two reasons. First, the supply is perfectly elastic. The insurance rates are fixed conditionally on given risk profiles, which primarily depend on government-designated risk zones. They do not otherwise vary by state, locality, or market conditions. Second, unlike most other property and casualty risks, flood risk has been shunned by private insurers (Horn and Webel, 2019), leaving few outside options for households.[4] Thus, standard models predict that the marginal household will go from

[4]The private market for flood insurance barely exists because the government heavily subsidizes the NFIP (see footnote 7). Cutting subsidies and privatization are the subject of contemporary policy debates, but irrelevant to this paper's research question.

not buying to buying flood insurance from the NFIP when its hedging demand becomes stronger. Aggregating at the county-month level, the change in the number of policies in-force thus captures shifts in the demand curve.

To identify the effect of salience, my first quasi-experiment exploits a staggered campaign that aims to increase public awareness of flood risk. The campaign has two components. First, it transforms an existing flood risk map from black-and-white to colored (Figure 1.1 presents an example). Second, when the new map is published in a county, the local government publicizes it in various ways (such as open houses and newspaper advertisements), which increases the salience of flood risk (see Section 1.2.2 for details).

Under the null hypothesis of perfect attention to flood risk, this reform should not significantly affect households' insurance purchases, for two reasons. First, the hard information about flood risk is identical in the two maps in Figure 1.1. Second, this information was already freely-available online. Hence, for a rational agent, to buy or not to buy flood insurance, the decision should have already been made.

Using a difference-in-differences strategy that exploits variation in the campaign timing across counties, I show that the number of flood insurance policies in-force increases by 30.6% after the localized salience shock. I find no evidence of a differential pre-trend in the treated counties. Following the campaign, the increase in insurance demand materializes quickly over several months and appears to be permanent, as I find that the incremental policies are renewed in subsequent years. This result is consistent with the hypothesis that households are inattentive to flood risk due to its low salience. While I cannot distinguish whether the risk information was known by households but ignored or whether it was simply unexploited, both are examples of limited attention to flood risk, as the information publicized by the campaign was always freely-available.[5]

I provide a set of additional empirical findings consistent with the salience hypothesis. First, I show that average house prices (measured by Zillow's home value index) in the treated counties drop by 2% after the campaign, suggesting

[5] Googling "flood insurance" or "flood risk" would instantly bring up information about the NFIP and its flood risk maps. In this setting, ignorance of the importance of hedging flood risk is effectively equivalent to ignorance of the freely-available flood risk information.

that when flood risk becomes more salient, it is more heavily priced into real estate values. This finding is consistent with households underweighting product attributes (such as a property's exposure to flooding) until they are made salient (Bordalo et al., 2013).

Second, I show that the effect is *weaker* where the salience of flood risk is already high. Specifically, flood occurrences in the past predict a smaller post-campaign increase in flood insurance purchases.[6] This correlation is monotonic—the longer the county has not had a flood, the larger the effect. In a similar vein, I also show that, given any value of n from 1 to 6, counties with no flood in the past n years before the campaign experience a larger treatment effect than do counties that had a flood in that period.

Third, I further consider several proxies for households' pre-campaign attention to flood risk. I find that the effect is *weaker* in counties with greater incomes, higher education, or more believers in global warming. Moreover, the effect is inverse U-shaped against the underlying risk level: the strongest impact of the campaign is observed in moderate-risk areas, likely because flood risk is already salient in high-risk areas and because insurance is not needed in low-risk areas.

I consider several alternative explanations for my findings. First, the relationship between campaign and flood insurance demand might not be causal if both are driven by a third factor. One alternative explanation is that the campaigns are introduced across geographies that experience increasing underlying flood risk, which simultaneously drives up insurance purchases. In other words, the treatment might endogenously capture an existing trend in underlying flood risk and insurance demand. However, the parallel pre-trend, shown in Figure 1.5, refutes this explanation.

A related non-causal explanation is that the campaign follows a local flooding event, which simultaneously induces insurance purchases. This explanation differs from the previous one if flooding is unpredictable, in which case the pre-trend test does not help. Therefore, I provide additional evidence to cast doubt on

[6] There is a large literature showing that past experiences affect subsequent decision making. See, for example, Malmendier et al. (2011); Malmendier and Nagel (2011); Dittmar and Duchin (2016); Malmendier and Nagel (2016); Bernile et al. (2017); Schoar and Zuo (2017); Kuchler and Zafar (2019).

this alternative explanation. First, past floods cannot predict campaign timing. Second, there is no change in flood frequency around treatment, while the alternative hypothesis predicts bunching before it. Third, anecdotally, local officials explicitly declared that the campaign is unrelated to recent floods.[7]

Next, I investigate whether there is a purely rational explanation for why the campaign leads to increasing demand for flood insurance. While my data does not allow me to fully rule out all rational hypotheses discussed below, I present evidence suggesting that my findings are unlikely to be driven by purely rational mechanisms.

The first purely rational hypothesis I consider is that having the new maps on the Internet reduces households' transaction cost of finding or processing flood risk information. However, this is unlikely, as the existing black-and-white maps were always online, and people could also easily acquire information from local NFIP agents or toll-free hotlines. I provide further evidence to support this argument in Section 1.4.6.2. For example, I find that the campaign also affects the intensive margin of existing policyholders, who are already informed by their current policies (the most crucial piece of information being their risk zones). Yet, they increase their insurance coverage, which is unlikely due to a reduction in information cost.

Second, many households probably lack insurance knowledge (which might be costly to learn), even if they are perfectly aware of flood risk. Thus, a potential rational channel is that the campaign educates households about how insurance works. While financial illiteracy is a plausible factor of low insurance take-up (Cai et al., 2020), it is unlikely the driving mechanism of my quasi-experiment. For example, following the NFIP's campaign, there are no spillover effects in other insurance products, such as earthquake or health insurance. I present additional evidence in Section 1.4.6.3.

The third purely rational hypothesis I consider is that the informational content of the existing and new maps is fundamentally different. Anecdotally, some areas of some maps might have been updated, while being colorized, to reflect

[7] For example, on the official website of the City of Alexandria, Virginia, it says, "This effort [developing new maps] is unrelated to recent flooding the city has experienced from flash flooding in July 2019 and, more recently, on July 23 [2020]." See Section 1.4.6.1 for details.

higher or lower flood risk. However, I present evidence in Section 1.4.6.4 suggesting that the scope of updating is small, consistent with the examples in Figure 1.1 and the Web Appendix. For example, I show that the campaign does not cause existing policyholders to cancel or to stop renewing their flood insurance policies. If the map modification had been extensive, we should observe cancellations induced by risk downgrades.

To further support the salience hypothesis and avoid the information-based alternative explanations, my second strategy exploits variation in social connectedness on Facebook and leverages non-local flooding events. This strategy complements the first one by making it unequivocal that the local flood risk does not change and that there is no new information about the local risk. Specifically, for a given flooding event (e.g., in Boston), I examine flood insurance purchases in geographically distant states. Within the same far-away state (e.g., California), I compare changes in flood insurance purchases in counties that are more versus less socially connected to Boston, before and after the flood in Boston. The idea is that when an individual sees Facebook friends sharing flood experiences, the risk becomes more salient.

Pooling all major floods (that triggered federal assistance) between 2010 and 2019 in an event study design, I find that the number of flood insurance policies in-force increases by 1%–5% in counties that are more socially connected to the flooded area, compared to the less connected counterfactuals in the same distant state. I find no pre-trend, and I document two additional findings consistent with the salience hypothesis. First, the effect is monotonic in the strength of social connectedness. Second, the most damaging floods cause the most pronounced effect across social networks. My strategy is similar to that of Bailey et al. (2018a), who show that friends' house-price experiences affect one's own housing investment decisions. As regional housing markets are possibly correlated while geographically distant floods are much less likely to be, my setting makes it arguably easier to rule out rational learning as an alternative explanation.

This paper does not claim that the increase in flood insurance purchases—or that the NFIP in general—is socially optimal. It is likely welfare-improving for households, as the NFIP offers heavily subsidized insurance rates, and it is the government's declared goal to encourage more people to acquire flood insurance.[8] A back-of-the-envelope calculation suggests that the expected net benefit of buying flood insurance is $1,240 per year per flood-prone household.[9] However, a complete analysis on social welfare needs to take government expenditures into account, which is beyond the scope of this paper.[10]

This paper contributes to a rapidly growing literature on behavioral household finance and is among the first to study households' insurance decisions against rare disaster risks. There is considerable evidence that households make suboptimal decisions because of limited attention. For recent empirical findings from various contexts, see Barber and Odean (2008); Chetty et al. (2009); Choi et al. (2009); Finkelstein (2009); Brown et al. (2010); Malmendier and Lee (2011); Lacetera et al. (2012); Hastings and Shapiro (2013); Stango and Zinman (2014); Busse et al. (2015); Andersen et al. (2020). The cost of inattention, in these examples, is typically small but incurred frequently and cumulatively. In contrast, evaluating low-probability high-consequence risks is likely a distinct task for households, and acquiring insurance is an increasingly critical household decision, yet understudied by the literature. My results suggest that households are especially vulnerable to behavioral biases when assessing low-probability risks. This finding is important to policymakers, as neglecting rare disasters can be costly for households.

A closely related paper is Gallagher (2014), which shows that flood insurance purchases increase after a local flooding event. My work differs in several dimensions. First, Gallagher (2014) documents a long-term but diminishing effect of personal experiences, while I provide evidence of a persistent effect in a different setting.[11] Second, his NFIP data is yearly, aggregated, and covers an earlier pe-

[8]The NFIP suggests that eliminating the subsidy would cause aggregate premiums to increase by 50-75% (Hayes and Neal, 2011). Consistently, as premiums are set below actuarial levels, they cannot fully cover claims: the cumulative difference is −$5.85 billion over the past 20 years. The NFIP's operating expenses further worsen its financial condition—it owed $20.5 billion to the Treasury as of December 2019 (excluding a $16 billion debt canceled by Congress in 2017).

[9]The estimation is based on the following assumptions: an NFIP-defined 1% inundation probability p.a., an average premium of $993, an average coverage of $227,112, and an average deductible of $3,831.

[10]Wagner (2019) provides a framework for studying the welfare effects of the NFIP. Her results suggest that enforcing a flood insurance mandate will increase social welfare.

[11]In Gallagher (2014), the effect is rather stable around 8% in event years 1 to 8, but it

riod (1980-2007). Third, his results are open to different interpretations: actual risk may change in the flooded area (e.g., due to foundation erosion), households may learn about flood risk and exposures from local and nearby floods, and households may overweight recent experiences.

This paper also adds to the nascent literature on the effects of climate risk on household behaviors and economic outcomes. A number of studies examine whether climate risk, in particular sea-level rise, is capitalized into real estate values (Giglio et al., 2015; Keenan et al., 2018; Bernstein et al., 2019; Baldauf et al., 2020; Murfin and Spiegel, 2020) and mortgages (Issler et al., 2019; Ouazad and Kahn, 2019). Other papers study how personal experiences of climate change or natural disasters affect beliefs about climate risk (Li et al., 2011; Zaval et al., 2014; Dessaint and Matray, 2017; Chang et al., 2018; Anderson and Robinson, 2019; Choi et al., 2020). These studies typically demonstrate a short-term impact of a shock (e.g., a day of unusual weather). The substantial and persistent effects of salience via campaigns and social networks, documented in this paper, have unique implications for effective climate policy, given that to offer subsidized insurance rates is costly yet has limited success.

Finally, this paper contributes to the emerging field of social finance (Hirshleifer, 2020) for studying how social interactions shape financial decision-making. See Kuchler and Stroebel (2020) for a review of recent empirical work at the intersection of social finance and household finance. My second quasi-experiment suggests that salient events—in particular, natural disasters—are transmitted through social networks. In my first setting, social interactions may also help to amplify the campaign's effect on salience.

The remainder of the paper is organized as follows. Section 1.2 describes the institutional background and details the data. Section 1.3 describes the two empirical strategies. Section 1.4 presents results from the staggered campaign. Section 1.5 presents results using social connectedness and geographically distant floods. Section 1.6 concludes.

---

becomes statistically insignificant in years 9 and 10. In my setting, Appendix Figure A1.4 shows a persistent effect.

1.2 Institutional Background and Data

1.2.1 The National Flood Insurance Program

The US Congress founded the National Flood Insurance Program (NFIP) in 1968, and as of 2019, it covers all 50 states and 3,053 (out of 3,143) counties. The program creates flood hazard maps for participating communities (subdivisions of counties, such as townships, villages, and cities), and only residents in participating communities are eligible to buy NFIP policies. The insurance premiums primarily depend on risk zones (set centrally by the government) and also vary by the choices of coverage and house characteristics. They do not otherwise vary by state, locality, or market conditions.

The NFIP data is maintained by the Federal Emergency Management Agency (FEMA).[12] I obtain more than 50 million transaction-level observations between January 2009 and August 2019, including policy effective dates, policy termination dates, premiums, coverage, deductibles, first policy dates, cancellation dates, and house characteristics. The policies are annually renewed, and renewals appear as separate transactions in the dataset. Broad location information (such as census tract, county FIPS code, and community code) is available, but specific properties cannot be identified due to privacy protection (geographic coordinates are truncated to one decimal point).

Based on the policy effective dates and termination dates, I can calculate the number of policies in-force (a stock measure) and the number of policies purchased (a flow measure) in a given month for a given county. By definition, the change in the number of policies in-force between two consecutive months $t-1$ and t is equal to the number of policies purchased minus the number of policies expired in t. Additionally, knowing the policyholder's first policy date allows me to determine whether the anonymized transaction is a first-time purchase or a renewal. Since I do not observe data for 2008, I can not perfectly impute the number of policies in-force in 2009. Therefore, my analysis starts from January 2010. The data granularity and the 10-year panel allow me to zoom in on households' flood insurance demand in the very short-run as well as to keep track of

[12] Source: https://www.fema.gov/openfema-data-page/fima-nfip-redacted-policies-v1

its long-term dynamics.

Table 1.1 presents descriptive statistics of the NFIP. Panel A shows that in an average month, there are 5.29 million policies in-force nationally. From these policyholders, the program receives \$3.32 billion in premium for \$1.26 trillion in coverage.[13] The nationwide average annual premium per policy is \$628, and the average coverage per policy is \$238 thousand. Panel B shows that the cross-sectional heterogeneity is stark. While the average county purchases 1,766 policies, the median is only 120. Appendix Figure A1.1 shows a geographical heat map of the number of policies in-force. As one would expect, coastal counties have the highest densities. The variation in the average insurance premium, reported in Panel B, is due to differences in flood risk across counties, rather than to price differentiation.

1.2.2 Flood Risk Maps and Campaigns

The Risk Mapping, Assessment, and Planning (Risk MAP) program was launched in 2009 by FEMA, aiming to increase public awareness of flood risk. This campaign transforms existing black-and-white flood risk maps into colored ones and publicizes them. Before the reform, people could obtain paper maps from local NFIP agents or offices, find scanned copies online, or call toll-free NFIP hotlines to acquire relevant information.

1.2.2.1 Map Transformation

Figure 1.1 shows an example—the Town of Colfax (community code: 220077), Grant Parish (county code: 22043), Louisiana—to illustrate what users can see on the flood risk maps (available at msc.fema.gov/portal). Figure 1.1.a shows a scanned copy of the legacy paper map. It was published on November 16, 1995 and was in use until the new colored map (Figure 1.1.b) became available on June 16, 2016.

These two maps convey identical information about the flood risk in the Town

[13]Compared to other property and casualty insurance in terms of aggregate premiums (2017 data): earthquake (\$2.9B), aircraft (\$1.5B), mortgage guaranty (\$5.0B), burglary (\$0.3B), and fire (\$11.6B).

of Colfax (although the jurisdiction boundary is slightly different). The key information is the risk zone designation. The blue areas in Figure 1.1.b are the Special Flood Hazard Areas (SFHA) which are expected to have a 1% (or higher) flooding probability per year (referred to as 100-year floodplains), and the brown areas are of median risk, with a 0.2% (to 0.99%) flooding probability per year (referred to as 500-year floodplains).

1.2.2.2 Local Campaigns

Around the publication of the new colorized maps in a county, FEMA instructs the local government to run a campaign (known as the "Flood Risk Open House") to increase public attention. FEMA provides customized marketing packages and templates (known as the "Toolkit") to advertise the Open House by placing advertisements in local newspapers and on radio, distributing flyers, and posting announcements on community websites and social media. The full toolkit can be found on FEMA's website.[14] Figure 1.2 presents examples of actual advertisements and announcements.

1.2.2.3 Visual Salience

Visually, the new maps are more appealing and might increase the salience of flood risk, even if the hard information about flood risk does not change on the maps. Figure 1.1.b appears visually more salient than Figure 1.1.a because flood risk is more intuitively highlighted by the water-like color and because the visibility of actual houses and streets in satellite views might create resonance and draw attention.

To support this argument, I leverage a machine-learning-based methodology called the Saliency Attentive Model (SAM), developed by Cornia et al. (2018) to predict human eye fixations on an image. I put the existing and new maps side by side as one input image. Appendix Figure A1.2 presents the output returned by the SAM; the overlay heat map represents the predicted eye fixations. The

[14] See http://townofvanburen.com/wp-content/uploads/2016/08/Onondaga-Open-House-Community-Packet-FINAL.pdf, for an example of a tailored toolkit that FEMA sent to Onondaga County, NY.

results show that the colored map draws more attention than the black-and-white map.

My data does not allow me to disentangle the above two salience effects, as the publication of new maps and the marketing events take place simultaneously in a given county. I provide evidence in Section 1.4 suggesting that the marketing campaigns are likely to be the driving mechanism.

1.2.3 Staggered Rollout

Revisiting Figure 1.1.b, the dates shown on each small area indicate the publication dates of the new flood hazard maps. For instance, the Town of Colfax (so as other communities from the same county) published its new colored map on June 16, 2016, whereas the neighboring county to the west did so on July 6, 2015. For the dot-shaded area to the south-west (part of a different county), its colored map is not yet available, and the latest version is still the scanned black-and-white map from September 5, 1984 (which is easily obtainable from the same online portal).

I obtain all communities' map publication dates from FEMA's Community Status Information (CSI) database. The data suggests that the rollout happens at the county level, and communities within the same county typically publish their new maps simultaneously, consistent with the open house examples discussed in the previous section. Since the other essential data for my analysis, such as flood occurrences and housing prices, is not as granular as communities, I examine the staggered campaign at the county level. When there are disparities among communities, I define the treatment time of a county as the calendar month in which more than 50 percent of its communities simultaneously publish their new maps.[15] I verify that the results are robust to various alternative definitions of county-level rollout (see Section 1.4.1 for details).

Figure 1.3 maps the staggered rollout by time and county. The darker the shade, the more recent the treatment dates; and the unshaded areas are the

[15]The 50 percent threshold is not a cumulative measure. Instead, it means at least half of the communities publish the new maps simultaneously in one specific month. This criterion by construction captures the unique treatment event at the county level.

untreated. There is no discernible clustering, and there is obvious variation within regions and states. (Figure 1.1.b is a consistent micro example showing three adjacent counties with different campaign timing.) I also verify that the timing is uncorrelated with either the level of flood risk (see Appendix Figure A1.3) or past flood occurrences (see Section 1.4.6.2).

One limitation of the CSI database is that it only records the latest map publication dates, while FEMA aims to review its maps every five years. Thus, a concern is that some counties may have published new maps twice, but I only observe the latter. This might lead to underestimation, as the first treatment is ignored and the county is coded as control until the second treatment.[16] However, Appendix Section A presents evidence suggesting that FEMA largely falls short of its goal. Hence, the date I observe is likely to be the county's only map publication in the 2010s.

1.2.4 Presidential Disaster Declaration

I identify flooding incidents using the Presidential Disaster Declaration database. The declaration process was established in 1988 (by the Stafford Act) for local and state governments to request federal natural disaster assistance. This database provides information on disaster ID numbers, declaration dates, incident begin and end dates, declared states and counties, and incident types. I categorize certain incident types—Severe Storm, Hurricane, Flood, Coastal Storm, and Typhoon—as flood-related events. I identify 419 flood-related declarations over my sample period; one declaration typically includes multiple affected counties.

1.2.5 Social Connectedness by Facebook

Bailey et al. (2018b) aggregate anonymized information from the universe of friendship links between all Facebook users as of April 2016 to produce a county-by-county social connectedness measure. I initially obtained the data through a non-disclosure data-sharing agreement; the data was later made open source

[16] Presumably, the first treatment is stronger as it combines a marketing campaign and an enhancement to visual salience (from black-and-white to colored), whereas the second treatment has only the campaign.

by Facebook. Bailey et al. (2018b) calculate the Social Connectedness Index (SCI) for a pair of counties as the number of Facebook friendship links between individuals located in those two counties. They further create a measure called the *relative probability of friendship* by dividing the SCI for county i and j by the product of the number of Facebook users in the two counties. If this measure is twice as large, it means that a given Facebook user in county i is about twice as likely to be connected with a given Facebook user in county j. I denote the *relative probability of friendship* by $p_{i,j}$ and use it to measure county-by-county social connectedness.

1.3 Empirical Strategies

1.3.1 Staggered Campaign for Flood Risk Maps

My first strategy exploits the staggered campaign for the publication of new flood risk maps, which increases the salience of flood risk. Leveraging variation in campaign timings across counties, the identifying assumption of my difference-in-differences strategy is that, without the campaign, flood insurance purchases would have moved in parallel in the treated and untreated counties.

I first estimate the following canonical staggered difference-in-differences regression, which models treatment events as immediate and permanent shifts in the outcome:

$$Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \epsilon_{it} \tag{1.1}$$

In this two-way fixed effects model, unit and time are specified as county and year-month. Y_{it} measures the number of policies in-force in county i at time t. I normalize the number in January 2010 to 100 for each county, so that the results are not dominated by extremely large counties. Let t_i^* denote the campaign time for county i. $Campaign_{it}$ is an indicator variable $\mathbb{1}(t > t_i^*)$ that turns on if county i has published its new colored flood risk map at time t; this term is set to zero for untreated counties for any t. α_i and λ_t are unit and time fixed effects, respectively. The coefficient β measures the change in the outcome following treatment. Standard errors are clustered at the county level to allow for arbitrary dependence of ϵ_{it} across t within i.

The identifying assumption requires the treatment timing to be uncorrelated with the outcome. If this assumption is not satisfied, the treated counties might already diverge from the controls before the treatment date. Also, the change in flood risk salience, due to the campaign, might not affect households' insurance decisions immediately; instead, the impact might develop gradually. To accommodate these possibilities, I also estimate the following nonparametric model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-\underline{L}}^{k=\overline{L}} \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it} \tag{1.2}$$

The primary advantage of regression (1.2) is that it allows me to flexibly and visually assess the pattern of outcomes relative to the event date. $\{\beta_k\}$ for $k < 0$ correspond to the pre-trends, and $\{\beta_k\}$ for $k > 0$ measure the dynamic effects of the campaign. These effects are measured relative to β_{-1}, which is omitted.

1.3.2 Social Connectedness and Geographically Distant Floods

My second strategy exploits exogenous variation in flood risk salience caused by cross-sectional variation in social connectedness with a geographically distant flooded area. The idea is that when an individual sees Facebook friends sharing flood experiences, the possibility of going through a flood becomes more salient.

For a given flooding event f, I identify a set of flooded counties $\{j\}_f$ and a set of geographically distant counties $\{i\}_f$ (that are at least 750 miles away from the flooded area). Then, within a far-away state, I define the treatment (control) group as counties that are more (less) socially connected with the flooded area. As the flooded area $\{j\}_f$ may consist of several counties, I calculate county i's social connectedness with the flooded area as a weighted average of the county-by-county relative probability of friendship $p_{i,j}$ (discussed in Section 1.2.5):

$$p_{i,f} = \sum_{\{j\}_f} w_j * p_{i,j} \tag{1.3}$$

where w_j represents population-weighting or equal-weighting. Within a *state*, county i is coded as treated (control) with respect to event f, if $p_{i,f}$ is above (below) its state median.

Figure 1.4 depicts one specific flooding event—Hurricane Florence in September 2018. It shows a heat map of social connectedness ($p_{i,f}$, blue-shaded) with the flooded area (red-shaded). The uncolored counties located less than 750 miles away from the flooded area are not included in this particular event study with respect to Hurricane Florence.

I stack individual event studies (with respect to every flooding event f) and cluster the standard errors at the county level. Each event study features the following difference-in-differences regression:

$$Y_{it} = \beta_0 + \beta_1 * Connected_i \times Post_t + \beta_2 * Connected_i + \beta_3 * Post_t + \epsilon_{it} \quad (1.4)$$

$Connected_i$ is the treatment dummy; $Post_t$ is the post-event dummy (that turns on if t is after the flooding time). The outcome variable Y_{it} measures the insurance demand in county i at time t, which is defined as in regression (1.1). The interaction $Connected_i \times Post_t$ is the key explanatory variable of interest.

By construction, I define treated and control counties conditionally on being in the same state, so that they are likely to have similar climatological and economic conditions. Even if counties in the same state have distinct climates and flood risk, the difference-in-differences construct of regression (1.4) teases out the fixed differences.

More formally, in order to interpret the estimate of β_1 as the causal effect of friends' flood experiences, I need to assume that insurance purchases in treated and control counties would have evolved in parallel without treatment (i.e., a distant flooding event). I test for parallel pre-trends by replacing $Post_t$ with a sequence of event time dummies $\{\mathbb{1}(t = t^* + k)\}$. The coefficients $\{\beta_1^k\}$ on the interaction terms $\{Connected_i \times \mathbb{1}(t = t^* + k)\}$ allow me to examine the patterns in insurance purchases in the months before and after the flood experienced by geographically distant friends. I present evidence of parallel pre-trends in Section 1.5.1. I discuss potential concerns about this event study approach in Section

1.5.4 and consider an alternative empirical methodology.

1.4 Empirical Findings: Risk-map Campaign

In the following two sections, I test the hypothesis that households will demand more flood insurance if the salience of flood risk increases. Section 1.4 presents the results of my first quasi-experiment, which exploits the staggered campaign across US counties to increase public awareness of flood risk. I also provide additional evidence supporting the hypothesized channel of salience and evaluate several alternative explanations.

1.4.1 Main Results: Salience and Insurance Purchases

Table 1.2 presents the staggered difference-in-differences estimates of regression (1.1). I examine the change in the number of policies in-force (with January 2010 normalized to 100). The standard errors are clustered at the county level to adjust for autocorrelation.[17] Column 1 shows that the average number of insurance policies goes up dramatically by 21.39, following the campaign for flood risk maps. Note that this is the mean treatment effect over the entire post period, and it corresponds to an increase of 19.91% relative to the mean of 107.42 at event time zero.

The result is robust to a variety of alternative specifications and sample restrictions. For example, in column 2, I add county-level covariates, including the average premium per policy and the average coverage per policy. In column 3, I exclude the never-treated counties from my analysis; thus, to control for any underlying trends, the staggered difference-in-differences estimation uses only counties that have not yet had or have already had the campaign. This specification addresses the concern that some never-treated counties (e.g., located in the desert) are poor counterfactuals. In both columns, the coefficient estimate has a similar magnitude and statistical significance.

[17] In fact, the clustered standard errors are almost 10 times larger than the non-clustered ones, suggesting strong positive autocorrelation. This is not surprising, as the policies are renewed annually.

I assess the robustness of my treatment construction in several dimensions. First of all, column 4 of Table 1.2 considers a continuous treatment approach and obtains a similar result. As detailed in Section 1.2.3, in my baseline discrete specification, a county is defined as treated if more than half of its communities publish and publicize the new maps in the same month. In column 4, I refine the treatment indicator $Campaign_{it}$ to be the cumulative fraction of treated communities. In the extreme case, in which all communities from the same county always act simultaneously, the continuous and binary approaches are identical.

Appendix Table A1.1 addresses the concern that population is not evenly distributed across communities. I refine the treatment definition to be population-weighted; that is, a county becomes treated in a given month if more than half of its population gets exposures to the campaign in that month. Since population data is not available at the community level, I use the number of policyholders as a proxy. Columns 1 through 3 show that my results hold. Additionally, in columns 4 and 5, I use different thresholds to construct the binary treatment; the results are consistent.

Appendix Table A1.2 shows that the results are also insensitive to sample selection. In my baseline specification, I estimate regression (1.1) using the full sample and do not impose any restrictions on leads and lags. In this table, I restrict the event window by various specifications of different leads and lags. The estimates demonstrate a similar magnitude and statistical significance across all specifications.

Figure 1.5 shows the dynamic effects of the campaign on insurance purchases. It plots the coefficient estimates of the event time dummies from regression (1.2). Crucially, there is no evidence of a differential pre-trend in the treated counties. This addresses the concern that the treatment assignment endogenously captures trends in the outcome variable; for instance, the government might expedite the reform in counties that are experiencing an increase in underlying flood risk. The parallel pre-trend refutes such an explanation and validates the identifying assumption (see more discussions in Section 1.4.6.1).

Figure 1.5 also shows that, following the campaign, insurance purchases rise sharply. The impact is visually apparent and persistent. In the first six months, households in the treated counties collectively buy, on average, 26.07 more flood insurance policies (corresponding to a 24.27% increase over the mean of 107.42 at event time zero). This gap continues widening to 32.84 (or 30.57%) at the end of the first year post treatment. Thereafter, the number of policies in-force remains stable and does not revert. Recall that the policies are renewed annually. Thus, comparing event months 1-6 with event months 13-18, we could posit that the additional purchases made immediately after the campaign are renewed in the next year. To illustrate that the impact is not transitory, Appendix Figure A1.4 extends the same plot to a horizon of 8 years and shows that the pattern persists in the long run.

1.4.2 First-time Purchases and Renewals

In this section, I investigate the surge in insurance demand in more detail. I examine the number of policies purchased (a flow measure), complementing the previous analysis of the number of policies in-force (a stock measure). The main benefit of the flow measure is a clear decomposition of all purchases, at any given point in time, into two parts—first-time new buyers and existing policyholders (i.e., renewal decisions).

Figure 1.6.a depicts the dynamic effects on the number of policies purchased. Similar to Figure 1.5, there is no evidence of a differential pre-trend in the treated counties. There is a sharp spike immediately following the campaign—the transaction volume increases by 80 percent in event month 2. Interestingly, the dynamic effects display a recursive pattern. For instance, the second spike of insurance purchases occurs in event month 14, suggesting that the additional purchases are renewed one year later.

In the next step, I decompose the number of insurance purchases into policies bought by first-time buyers (Figure 1.6.b) and policies renewed by existing policyholders (Figure 1.6.c). Note that these classifications of buyer types are time-specific.

Figure 1.6.b shows that the number of first-time buyers triples immediately after the campaign. Noticeably, in the subsequent years, there is no new wave of

new buyers. This result suggests that the increase in insurance demand seems to be driven by the campaign in a short period of time, rather than by a more gradual information transmission process (see Section 1.4.6.2 for a concrete discussion). This result also suggests that the primary component of the salience shock is the marketing efforts (discussed in Section 1.2.2.2) instead of the enhancement on visual salience (discussed in Section 1.2.2.3).

Will the first-time new buyers induced by the flood-risk-map campaign (for example in event month 2 in Figure 1.6.b) renew their policies? By construction, one year later, they are coded as existing policyholders. Hence, their renewal decisions are reflected in event month 14 in Figure 1.6.c, which depicts a spike of transactions.

The other interesting observation from Figure 1.6.c is that the campaign has little impact on the extensive margin of existing policyholders, as we do not see households canceling or failing to renew their policies around event time zero. This finding has important implications for differentiating among alternative explanations (see details in Section 1.4.6.4).

1.4.3 Salience and House Prices

The salience theory suggests that consumers place greater weight on product attributes that are salient (Bordalo et al., 2012, 2013; Kőszegi and Szeidl, 2013). In this section, I study a product that is closely related to flood risk—houses.

Conceptually, when flood risk is not salient, people pay limited attention to a property's flood risk exposure (a product attribute). As a result, this low-salience risk is not fully priced into real estate values. My hypothesis implies that when the campaign for flood risk increases its salience, we should expect flood risk to be more heavily priced—that is, to drag down house prices.[18]

Empirically, I examine the effect of the flood-risk-map campaign on house prices in the same framework as regression (1.1). I obtain the monthly county-

[18] Even for insured houses, this implication is valid, provided that there are frictions between insurance and housing markets, for two reasons. First, the NFIP's coverage is capped at $250,000. Second, the NFIP does not cover temporary living expenses, let alone the opportunity cost of time and inconvenience.

level home value index from Zillow (with January 2010 normalized to 100).[19] Zillow provides separate indices for different types of homes, such as all homes, single-family homes, top-tier homes, bottom-tier homes, and homes with 1, 2, 3, 4, or 5+ bedrooms. For robustness, I also use the yearly county-level house price index from the Federal Housing Finance Agency (FHFA), in which case, I collapse my NFIP data on an annual basis as well.

Table 1.3 reports the coefficient estimates using the Zillow data. The baseline result, presented in column 1 of Panel A, suggests that the average treatment effect on house price index (all homes) is −1.86, corresponding to a 1.80% decrease relative to the mean value of 103.50 at event time zero. Column 2 includes covariates, such as the county-level median household income and unemployment rate. While they show significant explanatory power (unreported for brevity), the coefficient estimate of $Campaign_{it}$ is fairly unaffected. In column 3, I examine single-family homes only. The result is qualitatively and quantitatively similar.

Columns 4 and 5 demonstrate the heterogeneous effects across the distribution of house prices. Zillow measures its house price index for the top-tier (bottom-tier) homes by using properties with values within the 65th to 95th (5th to 35th) percentile range for a given county. The results suggest that the campaign has a stronger impact on the cheaper properties, as the price drop for top-tier houses is 0.4 percentage point smaller than that for bottom-tier houses.

Panel B of Table 1.3 examines the house price indices of homes with various numbers of bedrooms. In all cases, there is a significant decrease in house price after the localized campaign for flood risk maps in a given county. Interestingly, the price drop is monotonically smaller in magnitude for houses with an increasing number of bedrooms. This result is consistent with the comparison in Panel A between the top-tier and bottom-tier homes.

Figure 1.7 shows the dynamic effects (using the baseline index of all homes). Firstly and crucially, there is little evidence of a differential pre-trend in the treated counties. Following the campaign, the house price index is apparently trending downwards. This result suggests that flood risk, once it attracts more public attention and has higher salience, is more extensively incorporated into

[19] Source: Zillow, Inc., https://www.zillow.com/research/data/

asset prices. It is interesting to notice that unlike the immediate surge in flood insurance purchases (see Figure 1.5), the downward trend in house prices continues for a few years and then stabilizes around −2%. A plausible explanation is that housing transactions are much less liquid than insurance purchases.

For robustness, Appendix Figure A1.5 presents the results of using the alternative annual house price index obtained from the FHFA. The downward post-campaign trend has a similar pattern, and the estimate is larger in magnitude.

1.4.4 Salience and Past Flood Occurrence

As flooding events would certainly draw the attention of those flooded and increase the salience of flood risk, my hypothesis predicts a smaller impact of the flood-risk-map campaign if the county has recently experienced a flood. Moreover, the correlation should be somewhat monotonic—the longer the county has not experienced a flood, the stronger effect the campaign should have.

I repeat my main analysis in subsamples of counties that have not had any flood in the n years prior to the campaign. Panel A of Table 1.4 reports the results up to $n = 6$ (82% of the counties had at least one flood in the previous six years). I find an almost monotonic pattern going from columns 1 through 6. For instance, in counties without a flood in the previous year, the campaign causes the insurance demand to increase by about 20%, but in counties without a flood for at least six years, the impact of the campaign is twice as large.

As the frequency of flooding correlates with the inherent risk, it is worth showing that I do not merely replicate the variations in the underlying risk. To measure the inherent flood risk, I leverage the Special Flood Hazard Area (SFHA) defined by FEMA, which is expected to have a one-percent or higher probability of being inundated in any given year. In Appendix Table A1.3, I condition on counties with low or high flood risk (below- or above-median proportions of SFHA), and I find the same monotonicity seen in Panel A of Table 1.4.

Panel B of Table 1.4 presents an alternative perspective to consider the correlation between salience, flood occurrence, and the effect of the campaign. In Panel B, I examine counties that had at least one flood in the past n years before the campaign. Given a value of n (i.e., fixing the column), my hypothesis predicts a larger effect in Panel A than in Panel B. The results, across all the columns, are consistent with this prediction.

1.4.5 Salience and Heterogeneity across Counties

In this section, I demonstrate the heterogeneity of my finding by examining several factors that pertain to households' attention to flood risk. My hypothesis predicts that the post-campaign increase in insurance purchases should be more substantial in counties where flood risk is more likely to be neglected.

1.4.5.1 Heterogeneity by Income and Education

I first consider income and education as a proxy for households' awareness of the importance of mitigating flood risk. To assess if people react differently to the campaign, I interact the post-treatment indicator $Campaign_{it}$ with $Income_i$ and $Education_i$ in regression (1.1). $Income_i$ measures county i's median household income (in \$1,000), and $Education_i$ measures the share (in percent) of population with college degrees in county i. The data is obtained from the US Census Bureau.

Columns 1 and 2 of Table 1.5 report the coefficient estimates of the interactions. The negative coefficients suggest that counties with more income and education are less responsive to the campaign for the new colorized flood risk maps. Quantitatively, \$1,000 more in income or 1 percentage point more of college attainment is associated with 0.65 percentage point less growth in insurance purchases following the campaign.

My strategy controls for existing differences in levels of insurance purchases (e.g., the wealthier are more likley to acquire insurance in the first place), and I estimate the heterogeneous effect of the campaign. My result suggests that inattention to flood risk is potentially more concerning in areas with less education and income.

1.4.5.2 Heterogeneity by Beliefs about Global Warming

Households' beliefs about global warming is another useful proxy for their ex-ante attention to flood risk. I obtain data from the Yale Climate Opinion survey (Howe et al., 2015), which has been used in several recent studies of climate risk (Bernstein et al., 2019; Baldauf et al., 2020). My main measure, $ClimateOpinion_i$, is the percentage of people, in county i, who answered "Yes" to the question of whether they think global warming will harm them personally.

Column 3 of Table 1.5 reports the coefficient estimate of $ClimateOpinion_i$ interacted with $Campaign_{it}$. The result suggests that a 1-percentage-point increase in the proportion of people who worry about the consequence of global warming leads to a 1.56-percentage-point decrease in the post-campaign growth in flood insurance purchases. Column 4 shows that this effect is not subsumed by the heterogeneity in income.

1.4.5.3 Heterogeneity by Inherent Risk Level

Because people living in areas with higher inherent flood risk are likely to pay more attention to it, the campaign-induced salience shock should be less pronounced in flood-prone areas. Along this line, I consider several analyses.

First, I define a binary variable $\mathbb{1}(Coastal)_i$, equal to 1 if county i is located in a coastal state, as a proxy for high flood risk. Second, for every county, I calculate the proportion of policies originated in the Special Flood Hazard Area (SFHA) over the total number of policies and define $\mathbb{1}(HighRisk)_i$ as a binary variable indicating if county i is above the median. As before, I examine their interactions with $Campaign_{it}$ to capture any heterogeneous treatment effect. The negative coefficients, reported in columns 5 and 6 of Table 1.5, suggest that the effect is stronger in areas with lower flood risk. Appendix Figure A1.6 illustrates the point graphically and verifies that the parallel pre-trend assumption is satisfied in all subsamples.

Additionally, I exploit the heterogeneous effect within counties. My transaction data allows me to see if a policy was originated by a household living in the SFHA or not. I compare the post-campaign growth in insurance purchases in a treated county's SFHA regions and non-SFHA regions. Columns 1 and 2 of Appendix Table A1.4 suggest that the growth is less than 10 percent in the SFHA regions, but more than 50 percent in the non-SFHA regions. Column 3 effectively runs a triple-differences regression, showing that the difference is statistically significant. Consistently, column 4 shows that the proportion of SFHA policies decreases by 4.7 percentage points in the treated counties. These results suggest that the reform induces disproportionately more households in moderate- or low-risk areas to acquire flood insurance.

1.4.5.4 U-shaped Effect and Inherent Risk Level

Following up on the previous section, I explore the variation in flood risk in a more granular scope. The campaign's effect is likely to be non-linear in flood risk: on the one hand, in high-risk areas, the salience is already high; on the other hand, in low-risk areas, flood insurance is less relevant. Therefore, we expect to see the strongest response to the campaign in moderate-risk areas—that is, an inverse U-shape.

In column 7 of Table 1.5, I consider a continuous measure $RiskLevel_i$, which is the proportion of SFHA policies in county i (in percent), interacting its quadratic form with the post-treatment dummy $Campaign_{it}$. The result, in column 7, is consistent with the prediction: the estimated coefficients imply that the largest effect occurs when $RiskLevel_i$ equals 30 percent.

To illustrate this point graphically, I divide all counties into quintiles based on $RiskLevel_i$ and estimate the treatment effect in each quintile. Figure 1.8.a plots the coefficient estimate of the effect of the flood-risk-map campaign on the number of flood insurance policies in-force. Again, we observe an inverse U-shape. The moderate-risk counties (quintiles 2 and 3) experience the most substantial increase in insurance purchases following the campaign, whereas there is little response in the bottom and top quintiles.

To complement the above evidence supporting the salience hypothesis, I document the same pattern of heterogeneous effects in the post-campaign housing markets. Figure 1.8.b plots the coefficient estimate of the treatment effect on house prices in quintiles of counties based on the flood risk measure $RiskLevel_i$.

Consistently, we see a U-shape: the moderate-risk counties display the largest decrease in house prices, whereas there is little response in the bottom and top quintiles.

1.4.6 Alternative Explanations

In this section, I evaluate a number of alternative explanations for my main findings and present empirical evidence to characterize their relevance.

1.4.6.1 Non-causal Interpretations

I first address concerns pertaining to the validity of the assumption of quasi-random treatment. The observed positive relationship between campaign and flood insurance demand might not be causal if both are driven by a third factor. One alternative explanation is that the new colorized maps are introduced across geographies that experience increasing underlying flood risk, which simultaneously drives up insurance purchases. In other words, the campaign process might endogenously capture an existing trend in underlying flood risk and insurance demand. However, the parallel pre-trend, shown in Figure 1.5, refutes this explanation. It suggests that the staggered campaign does not capture any existing trends in demand.

However, even in the absence of a pre-trend, the reform could still be endogenous. One specific alternative mechanism is that the campaign follows a local flooding event, and households buy more flood insurance after recently experiencing a flood, due to either rational learning or behavioral bias. Many papers show that individuals overweight recent experiences (Greenwood and Nagel, 2009; Malmendier and Nagel, 2011; Dessaint and Matray, 2017), including floods (Gallagher, 2014). This explanation differs from the previous one if flooding is unpredictable, in which case the pre-trend test might not help. (For instance, two identical counties could have identical dynamics of flood insurance policies in-force, until a random one of them is hit by a flood.)

Therefore, I provide additional evidence to cast doubt on this alternative explanation. First, Appendix Table A1.6 shows that past floods cannot predict the campaign timing. Second, Appendix Figure A1.7 shows that there is no difference in flooding frequency around the campaign, although if this alternative hypothesis were true, we would expect to see bunching before it. Third, anecdotally, local officials explicitly declared that the effort to publicize the new maps is unrelated to recent local flooding events. For example, in an announcement posted on the official website of the City of Alexandria, Virginia, it says, "This effort [developing new maps] is unrelated to recent flooding the city has experienced from flash flooding in July 2019 and, more recently, on July 23 [2020]."[20]

1.4.6.2 Rational Explanation: Information Cost

After addressing the concerns of non-causal interpretations, I continue to explore if there is a purely rational alternative explanation for why the flood-risk-map campaign leads to increasing demand for flood insurance.

The first purely rational alternative explanation I consider is that the reform reduces the transaction cost of finding and/or processing flood risk information. In other words, the observed marginal households did not buy flood insurance before the campaign because they found it too costly to acquire information about flood risk, not because of a psychological cost related to inattention.

While my data does not allow me to completely rule out this alternative explanation, I present evidence that it is unlikely to be the primary channel. First, the information cost, although difficult to quantify, was likely minimal. Households could always obtain the existing maps from FEMA's website, and it was not a hassle to visit their local NFIP offices or call its toll-free hotlines. Moreover, as detailed in Section 1.2.2, the flooding probabilities are fairly easy to extract from the maps using the self-explanatory legends. Therefore, it seems unlikely that the transaction cost is as high as $1,240 (the back-of-the-envelope benefit of buying the government-subsidized flood insurance).

Second, under this alternative hypothesis, existing policyholders should not behave differentially before and after the campaign, as they are already informed about their risk designations and insurance prices, stated in the current policies. However, I find that the campaign also affects the intensive margin of the already

[20] See https://www.alexandriava.gov/floodmap for full information.

insured households: they increase coverage per policy by 2.6 percent.

Third, as discussed previously in Section 1.4.2, the campaign induces a large number of first-time new buyers immediately after the event but not anymore in the future. If the driving mechanism of my result is through information provision and through reduction of information cost, we would expect to observe a more gradual process for more and more households to acquire the information.

Fourth, it is difficult for this alternative explanation to reconcile the heterogeneity of my results. For example, Section 1.4.5.4 shows that the treatment effect is stronger in moderate-risk areas than in high-risk areas. It is not clear why the reduction in information cost would be heterogeneous across locations.

Fifth, I consider an additional quasi-experiment featuring a reduction in transaction cost. Appendix Section B details the empirical setting, but in a nutshell, in July 2014, the government upgraded its online portal to provide improved search functionalities and greater convenience to users. I find that this direct reduction in search cost does not stimulate insurance demand (see Appendix Table A1.5).

1.4.6.3 Rational Explanation: Financial Literacy

While the transaction cost of acquiring flood risk information is minimal (as discussed above), the cost of financial education could be high. Thus, even if households are perfectly aware of flood risk, they might not understand how insurance works. An alternative explanation for my results is that the campaign, open house in particular, educates households about insurance knowledge. Cai et al. (2020) shows that financial literacy is effective in promoting insurance take-up.

However, based on anecdotal evidence and government documentation, financial education is not part of the campaign—instead, to increase public awareness of flood risk is the key goal. Moreover, existing policyholders presumably already understand how insurance works, yet as mentioned in the previous section, they increase their insurance coverage following the campaign.

I also show that there is no spillover effect across other insurance products. Firstly, I examine the effect of the flood-risk-map campaign on households' purchases of health insurance, leveraging county-year-level data obtained from the US Census Bureau's Small Area Health Insurance Estimates (SAHIE) program from 2010 to 2018. Using health insurance coverage (in percent) as the dependent variable in regression (1.1), I find that the treatment effect is economically and statistically insignificant (0.042 with a t-statistic of 0.61). Secondly, I investigate earthquake insurance purchases. Due to data limitations, I am only able to obtain county-level panel data for one state—from the Missouri Department of Insurance. I find that following the NFIP's campaign, the number of earthquake insurance policies in-force increases insignificantly by 0.68 percent (t-statistic=0.37).

1.4.6.4 Rational Explanation: Change in Informational Content

The third rational alternative explanation I consider is that the informational content of the existing black-and-white maps and the new colorized maps is fundamentally different and that the observed increase in insurance purchases merely reflects Bayesian updating. This is a valid concern, as anecdotally some maps may have been updated during the transformation process.

However, I present evidence that the scope of this alternative channel is small. First, the Web Appendix presents many examples comparing the two maps. Visually, the flood risk information in the two formats appears to be almost identical. I also find consistent anecdotal evidence; for example, a press release for the new maps for Sussex County, New Jersey, quotes Mary Colvin, Acting Mitigation Director for FEMA, Region II, as saying: "The new, preliminary map does not present any major changes in the flood plain."[21]

Second, the government suggests that the update component can go either way. Specifically, for the purpose of my argument, some areas should be downgraded to median or low-risk, as "overstated hazards can result in potentially unnecessary construction costs and incorrect insurance rating decisions." If this alternative explanation is prevailing, we should observe some households cancel or fail to renew their policies (even if the update is possibly systematically

[21] See the county's website: https://www.sussex.nj.us/cn/news/index.cfm?TID=7&NID=17552.

upwards—that is, if disproportionately more areas are designated with a higher risk). However, my result, presented in Section 1.4.2, suggests that the campaign does not have any significant impact on renewals or cancellations.

Third, I show, in Section 1.4.5.3, that the post-campaign effect is stronger in areas with lower inherent flood risk. To reconcile this result, the alternative mechanism needs to assume that the systematic increases in flood risk (reflected in the new maps) are disproportionately more significant in lower-risk areas. To the best of my knowledge, this assumption is not anecdotally supported.

Fourth, using a complementary sample (only yearly) of the program from 1980 to 2000, I find that map updates (from paper to paper) do not stimulate insurance purchases.[22] It suggests that FEMA's map modifications are likely to be small in scope.

1.4.6.5 Changes in Insurance Price

Is it possible that the campaign comes along with reductions in insurance price? This alternative hypothesis is not possible, as the federal government sets the NFIP rates and prohibits price discrimination across localities. Therefore, county officials and agents have no authorization to amend insurance prices while publicizing the new maps.[23]

Nevertheless, to empirically evaluate this alternative explanation, I check whether there are any price differences before and after the campaign. I consider three county-level measures of insurance price: the average premium per policy, the average premium per policy per $1,000 coverage, and the average premium per policy scaled by the proportion of high-risk policies (to account for changes in risk composition). I also examine the county-level median household income which pertain to the real cost of purchasing flood insurance. Appendix Figure A1.8 shows that there is no evidence to support this alternative explanation.

[22] The transaction-level information from the NFIP is available only for after 2009. For the earlier periods, the program provides the year-end statistics of the number of policies in-force for all counties.

[23] There were two nationwide reforms regarding insurance rates: the Biggert-Waters Flood Insurance Reform Act of 2012 and the Homeowner Flood Insurance Affordability Act of 2014. These affect all counties simultaneously and are controlled for by my difference-in-differences strategy.

1.5 Empirical Findings: Second Strategy

In this section, I present the findings of my second quasi-experiment, developed in Section 1.3.2. It complements the results of my first strategy and supports the hypothesis that households' flood insurance decisions are sensitive to the salience of flood risk.

1.5.1 Social Connectedness and Geographically Distant Floods

I exploit variations in social connectedness on Facebook and leverage non-local floods. The main idea is that a geographically distant flood should not change or generate any information about the actual local flood risk. However, non-local flood news potentially conveys differential degrees of salience to the local households, depending on the strength of social connectedness.

Using the event study framework of regression (1.4), I compare the changes in flood insurance policies in-force across counties in the same state with high versus low levels of social connectedness to a geographically distant flooded area. The connected (treated) counties are defined by having a connectedness measure above the state median.

Table 1.6 reports the coefficient estimates of regression (1.4). To reject the null hypothesis, we expect a positive difference-in-differences estimate. In column 1 of Panel A, the estimate of 1.11 is a 0.94-percent increase over the average number of policies in-force (117.66) at event time zero. It suggests that, when individuals see geographically distant Facebook friends sharing flood experiences, increasing the salience of flood risk, more households decide to acquire flood insurance.

This result is robust to a variety of alternative specifications and sample restrictions. For example, column 2 uses an equal-weighting scheme in equation (1.3) to measure a county's social connectedness with the flooding counties. Columns 3 and 4 construct the analysis sample using counties that are at least 500 miles or 1,000 miles away from a given flood. The effect remains positive and statistically significant.

In Panel B of Table 1.6, I further check the robustness of my results. Because the Facebook data is a snapshot as of April 2016, one concern is that the county-by-county social connectedness might be time-varying. In particular, the measure might not be a good proxy for the status back in the early 2010s. Panel B of Table 1.6 uses a sample period between 2014 and 2017. My findings are not sensitive to the choices of sample periods.

Figure 1.9 presents the dynamic effects of the flood experiences of geographically distant friends. Most crucially, there is no evidence of a differential pre-trend in the connected counties before the flood, which supports the identifying assumption that the less-connected counties are an appropriate counterfactual for the more-connected counties in the same state.

1.5.2 Salience and Heterogeneity

1.5.2.1 Monotonicity in Social Connectedness

The salience hypothesis implies that the effect should be monotonic in the strength of social connectedness. The most-connected counties should show the largest increases in insurance demand and the least-connected counties should show the least change.

Recall that my baseline analysis (in Table 1.6) defines the treatment or control group as respectively above or below the state-median value of the social connectedness measure as per equation (3). In Panel A of Table 1.7, I use a sharper approach and compare the top versus bottom quartiles. The estimates are larger in magnitude across all specifications. For example, the estimate of 2.20 in column 1 is a 1.88-percent increase over the mean of 117.09 at event time zero. This suggests that the salience of flood news, transmitted through social networks, is monotonic in the strength of social connectedness.

1.5.2.2 Significant Floods

My hypothesis also implies that the most damaging floods should cause the most pronounced salience effect across social networks. Panel B of Table 1.7 tests this prediction. I restrict my event study to 18 floods that were characterized as significant by FEMA.[24] Across all specifications, the estimate is more than twice as large in magnitude as the baseline in Table 1.6. For example, the estimate of 3.08 in column 1 is a 2.64-percent increase over the mean of 116.73 at event time zero; column 4 generates the largest estimate (5.10%) in this empirical design. This result suggests that a natural disaster's salience, transmitting across social networks, is order-preserving.

1.5.3 Alternative Explanation: Migration

Households may move after a flooding event, plausibly to places where families and friends live. Thus, an alternative explanation for my finding is that the more-connected counties receive more incoming households migrating from the flooded area than the less-connected counties. To assess its empirical relevance, I use the county-to-county migration data produced by the Internal Revenue Service (IRS), which is based on year-to-year address changes reported on individual income tax returns filed with the IRS.

The data suggests that long-distance migration is uncommon and hence unlikely the driving factor of my result. The average number of migrating households, from a flooded area to a 750-mile-away county, within one year of the flood, is only 3.6 (the median is 0). The magnitude is small compared to the average number of households (35,637) in a county. Moreover, there is little difference between the more- and less-connected counties. Using the number of migrating households as the dependent variable in regression (1.4), the difference-in-differences estimate is 0.99 (t-statistic=0.94). The estimate is economically and statistically insignificant, suggesting that the increase in insurance demand in geographically distant counties is unlikely due to migration.

1.5.4 Limitation and Alternative Methodology

In this section, I address a concern of my event-study design to link social connectedness with flood insurance purchases. The advantage of my second strategy,

[24] See https://www.fema.gov/significant-flood-events for FEMA's list of significant flood events. A significant event is defined as a flooding event with 1,500 or more paid losses.

as discussed above, is that each flooding event characterizes a standard difference-in-differences analysis, which allows for a straightforward verification of parallel pre-trends. However, the disadvantage is that a county could be involved (as either treated or control) in more than one events, which entails duplicating observations if the event windows overlap. In other words, instead of one observation per county per time, the event study approach has one observation per county per event per time.

I first address this problem of non-independent observations by clustering the standard errors at the county level (as in Tables 1.6 and 1.7). The other approach commonly adopted by empirical researchers is to only use large events, in the hope that they are sufficiently far apart. The analysis presented in Panel B of Table 1.7, with only the largest 18 floods included, is undertaken in this spirit.

In the following, I consider an alternative empirical approach used by Bailey et al. (2018a). Applying their terminology to my setting, I construct an index, $FriendFlood^N_{i,t_1,t_2}$, to measure the average flood experience of county i's social network N between t_1 and t_2. The largest social network N is the universe of all other counties; a restricted network can include only geographically distant ones. Let $\theta^N_{i,j}$ be the share of county i's friends in network N who live in county j, and let $Flood_{j,t_1,t_2}$ be the number of floods in county j between t_1 and t_2. The key explanatory variable is constructed as:

$$FriendFlood^N_{i,t_1,t_2} = \sum_j \theta^N_{i,j} * Flood_{j,t_1,t_2} \tag{1.5}$$

Bailey et al. (2018a) instrument for the house price experiences of all friends with the house price experiences of geographically distant friends to identify the causal impact of friends on an individual's housing investment decisions. As my primary interest is on the distant floods, I focus on the reduced-form to capture the average effect of geographically distant friends. Specifically, I estimate the following regression, with my baseline specification taking t_1 to be 12 months before t_2:

$$log(Policies)_{i,t} = \beta * FriendFlood^{Distant}_{i,t-12,t} + FE_{state \times time} + \epsilon_t \tag{1.6}$$

I control for the state$\times$time fixed effects, which allow me to isolate the effects of friends' flood experiences on the insurance decisions of counties located in the same state at the same time.

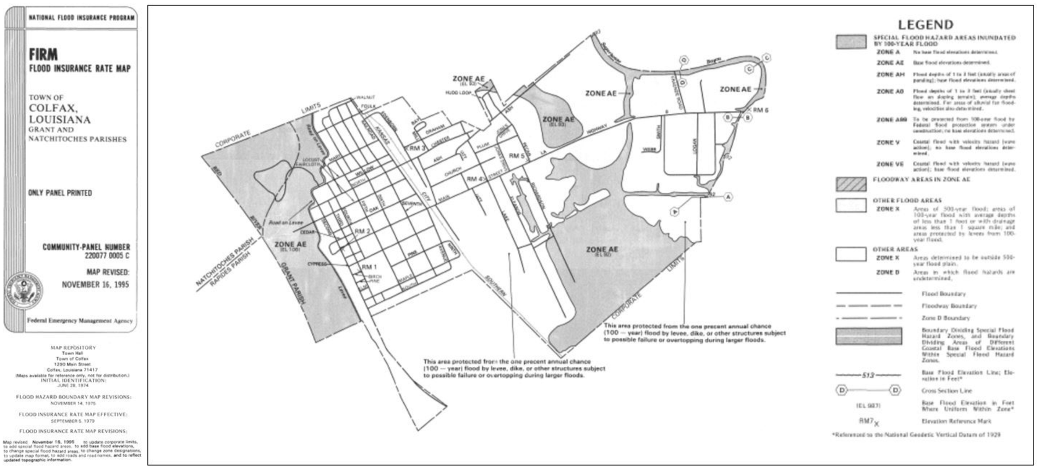
Panel A of Table 1.8 presents results from regressions (1.6). The estimate in column 1 suggests that every flood experienced by friends living in geographically distant counties (at least 750 miles away) increases the local insurance demand by 1.3%. Columns 2 through 4 show that my result is robust to a variety of specifications with different measurement windows of floods. Columns 5 through 7 show that my result is also insensitive to a variety of definitions of "geographically distant".

Panel B of Table 1.8 repeats the analysis by focusing on the experiences of significant floods only (as defined in Section 1.5.5.2). Across all columns, the estimates in Panel B are more than twice the baseline in Panel A. For example, the estimate in column 1 means that when geographically distant friends experience a significant flood, the local county's demand for flood insurance increases by 4.2%. Consistent with my event study methodology, these findings suggest that a distant flood's salience effect, transmitting across social networks, is order-preserving.

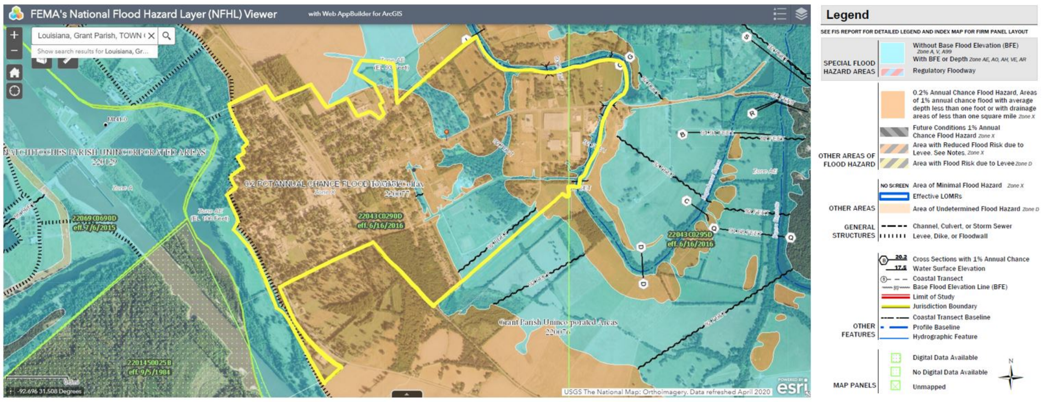
1.6 Conclusion

This paper examines how households make insurance decisions against flood risk. I use two empirical strategies to identify a causal effect of the salience of flood risk on households' willingness to acquire flood insurance. My results suggest that households pay limited attention to flood risk, due to its low salience. But because the expected cost of neglecting flood risk is large, US policymakers are seeking ways to stimulate insurance take-up. My findings suggest that one effective way is to increase the salience of flood risk by running campaigns to enhance public awareness, presenting risk information in more salient formats to households, and covering non-local flood news on local media. This insight could be widely generalized to other types of tail risk (especially natural disaster risk) and to other countries.

1.7 Figures



(a) Black-and-white Map Published on November 16, 1995

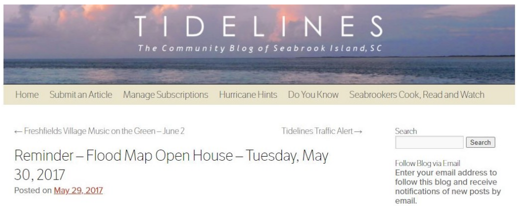


(b) Colorized Map Published on June 16, 2016

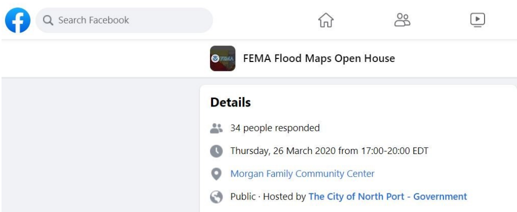
Figure 1.1. An Example of a National Flood Insurance Program Flood Hazard Map

This figure shows the flood hazard maps developed by the National Flood Insurance Program for the Town of Colfax, Grant Parish, Louisiana. Figure (a) is a scanned copy of the legacy black-and-white paper map, which was published on November 16, 1995. For readability, only the most relevant information is presented here, and the full copy can be found at https://msc.fema.gov/portal. Figure (b) shows the corresponding colorized map published on June 16, 2016. The two maps present identical information about the flood risk in the Town of Colfax (except that the jurisdiction boundary is slightly different). *Source*: The Federal Emergency Management Agency (FEMA).

(a) Open House Invitation on Local News

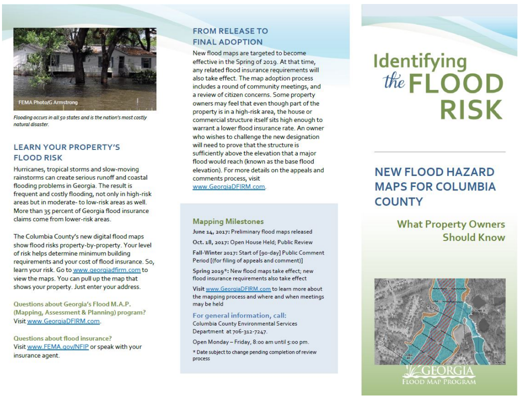


(b) Advertisement on Community Blog



(c) Open House Invitation via Facebook

(d) Announcement of New Maps Publication



(e) Brochure

(f) Local Newspaper

Figure 1.2. Local Advertisements of Flood Risk Open Houses and Map Publication

This figure presents examples of county governments advertising Flood Risk Open Houses and the publication of new flood maps.

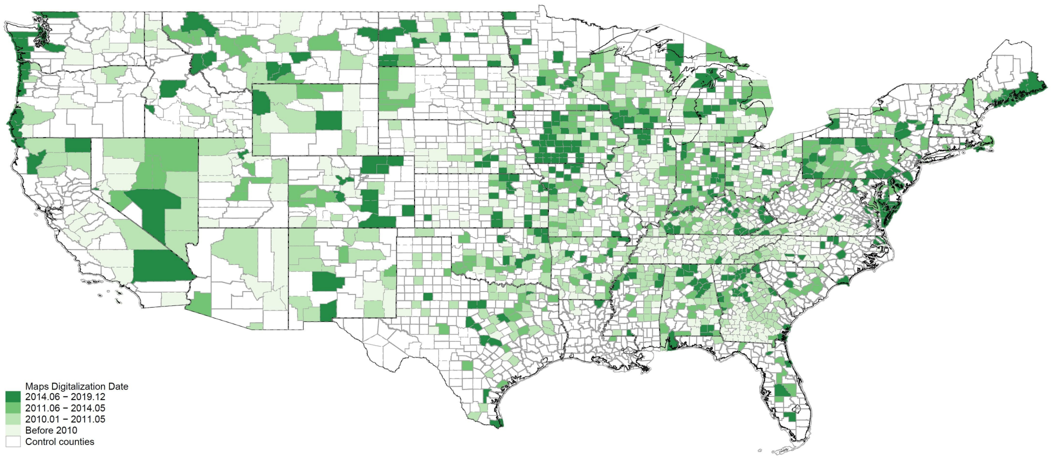


Figure 1.3. Empirical Strategy 1: Staggered Campaign for Flood Risk Maps

The figure shows the flood-risk-map campaign by county and time. The darker shade represents the more recent publication date of the new maps. The unshaded counties represent the untreated group.

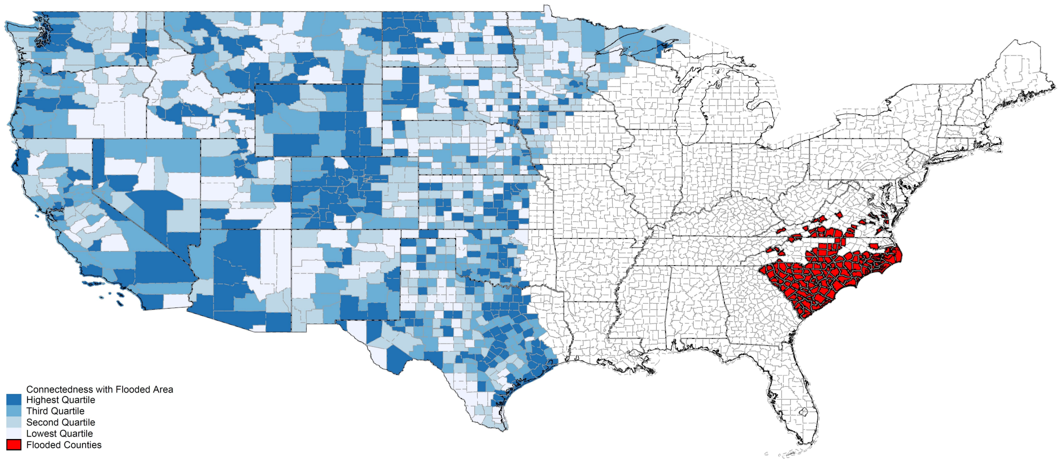


Figure 1.4. Empirical Strategy 2: Social Connectedness and Geographically Distant Floods

This figure shows one specific example to illustrate the empirical design of the second quasi-experiment. Hurricane Florence hit South Carolina, North Carolina, and Virginia in September 2018. The flooding counties are red-shaded on the map. The blue shades depict the heat map of social connectedness with the flooding area. Only counties located at least 750 miles away from the flooding area are considered.

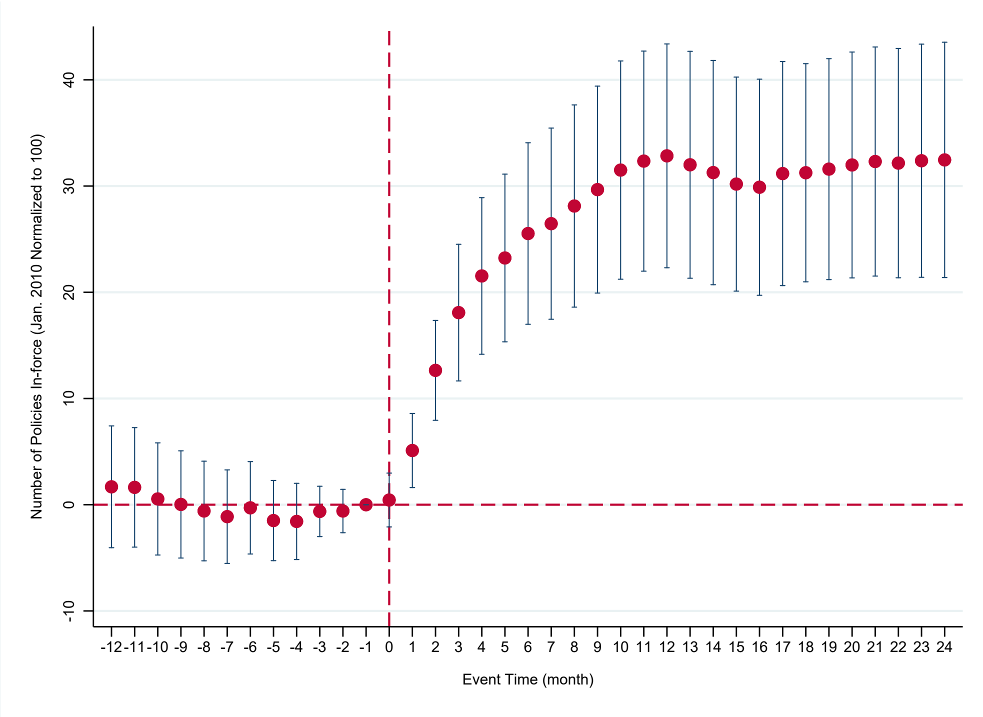
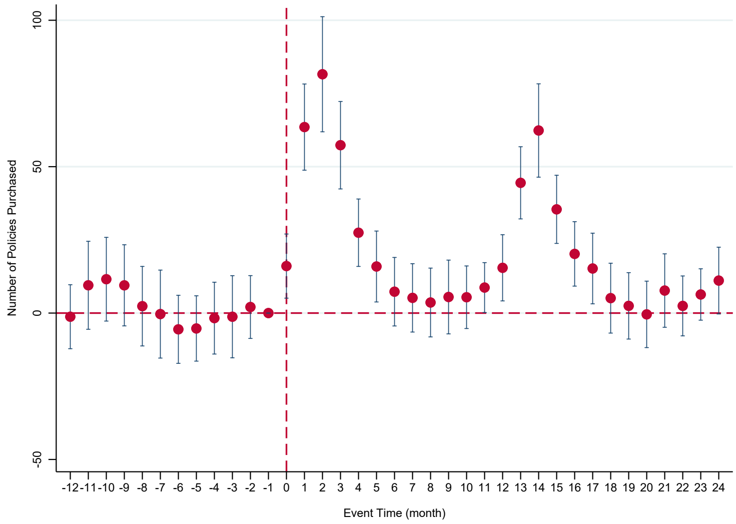
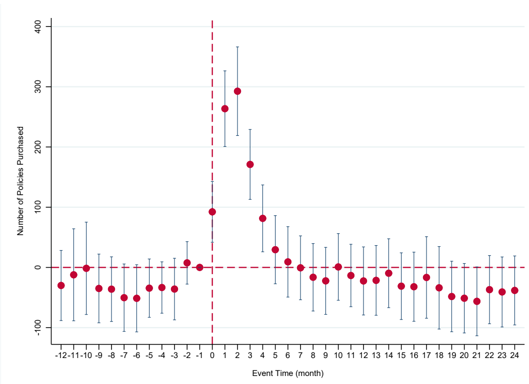


Figure 1.5. The Impact of Flood-risk-map Campaign on Insurance Policies In-force

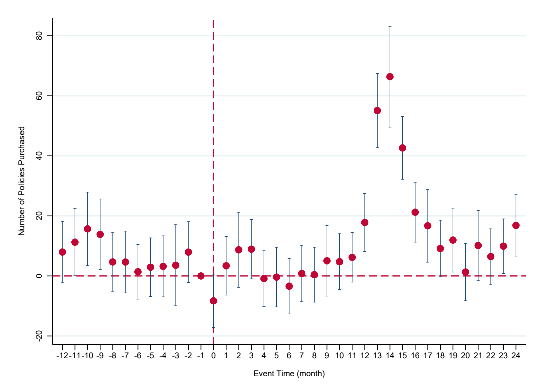
This figure shows the dynamic effects of flood-risk-map campaign on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. t_i^* is the publication time of the new maps in county i. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) All Transactions



(b) New Purchases



(c) Renewals

Figure 1.6. The Impact of Flood-risk-map Campaign on Insurance Purchases (Flow Measure)

This figure plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. In figure (a), Y_{it} measures the number of flood insurance policies purchased in county i in month t. In figure (b), Y_{it} measures the number of flood insurance policies purchased by first-time new buyers in county i in month t. In figure (c), Y_{it} measures the number of flood insurance policies renewed by existing policyholders in county i in month t. In all cases, Y_{it} is normalized with the value of January 2010 being 100. All the other variables are defined as per Figure 3. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

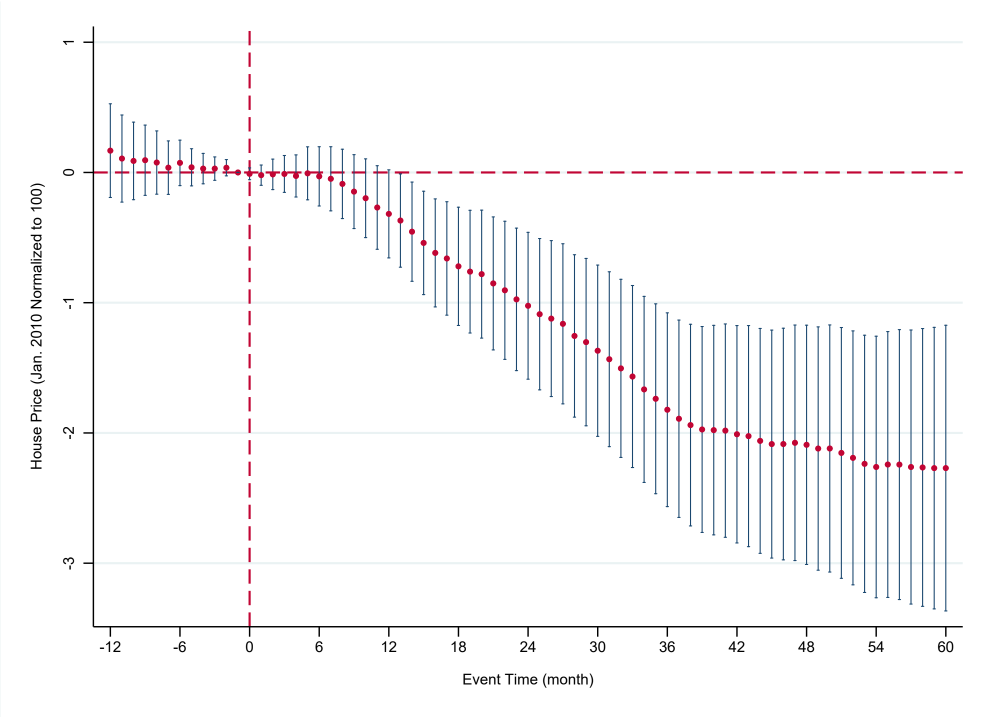
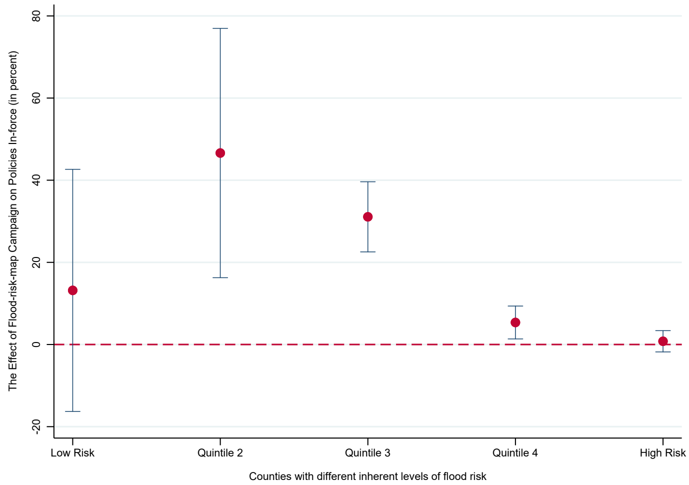
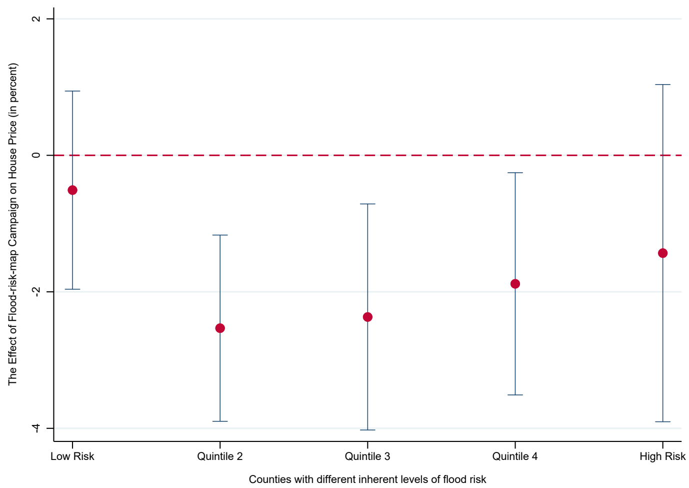


Figure 1.7. The Impact of Flood-risk-map Campaign on House Price

This figure shows the dynamic effects of the flood-risk-map campaign on house prices. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $HousePrice_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable $HousePrice_{it}$ is the house price index (with January 2010 normalized to 100) in county i in month t. t_i^* is the calendar month when county i publicizes its new maps. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) Heterogeneous effect on insurance purchases by risk-quintiles



(b) Heterogeneous effect on house prices by risk-quintiles

Figure 1.8. The Impact of Flood-risk-map Campaign across Subsamples

This figure shows the coefficient estimate of $Campaign_{it}$ in different subsamples from the regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \epsilon_{it}$. All counties are divided into quintiles based on a measure of inherent flood risk $RiskLevel_i$, which is the proportion of Special Flood Hazard Areas (SFHA). In figure (a), the dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. In figure (b), Y_{it} measures the average house prices (with January 2010 normalized to 100) in county i in month t. The main explanatory variable $Campaign_{it}$ is a binary variable indicating if county i has publicized the new maps at time t. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 90% confidence intervals.

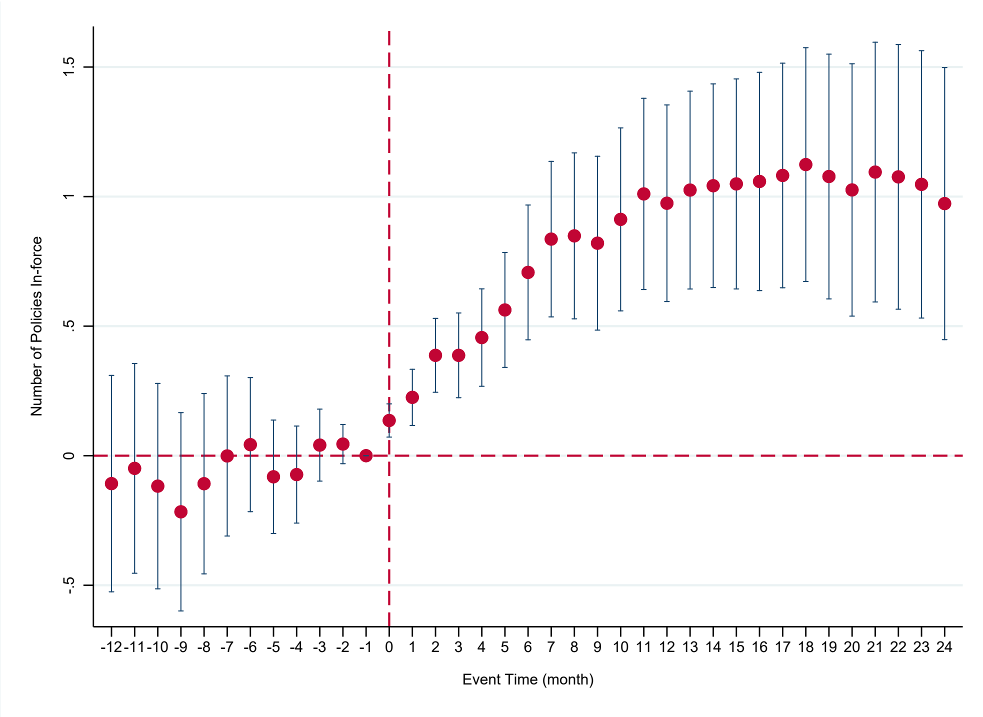


Figure 1.9. The Impact of Friends' Flood Experiences on Insurance Purchases

This figure shows the dynamic effects of geographically distant friends' flooding experiences on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ from the event study design: $Y_{it} = \beta_0 + \sum_k \beta_1^k * Connected_i \times \mathbb{1}(t = t^* + k) + \beta_2 * Connected_i + \sum_k \beta_3^k * \mathbb{1}(t = t^* + k) + \epsilon_{it}$. For notational brevity, the event index f is omitted from the equation. $\{\beta_1^k\}$ are measured relative to $\beta_1^{k=-1}$ which is omitted. For a given flood event f and the associated flooding counties $\{j\}_f$, $Connected_i$ is a binary variable indicating if county i is socially connected with the flooding area, which is defined as having a value of the connectedness measure above the state-median. t^* is the occurrence month of the geographically distant flood. The analysis sample consists of only counties that are at least 750 miles away from the flooding area. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

1.8 Tables

Table 1.1: Descriptive Statistics of the NFIP

This table presents descriptive statistics that characterize the National Flood Insurance Program (NFIP) data in my sample from January 2010 to August 2019. Panel A summarizes the time-series variations for the entire program. *Policies in-force* is the number of effective insurance policies in a given month. *Premium* is the total dollar amount of premiums collected from the policies in-force in a given month. *Coverage* is the total dollar amount of coverages for the policies in-force in a given month. *Flood-prone policies* is calculated as the number of policies in the Special Flood Hazard Area (SFHA) over the total number of policies. The NFIP creates risk maps and designates flood zones, and the SFHA is defined as the area that has a 1-percent or higher probability to be inundated in any given year. Panel B summarizes the cross-sectional variations of the data at the county level.

	Panel A: Nationwide Time-Series Varation				
	mean	s.d.	25^{th} pctl.	50^{th} pctl.	75^{th} pctl.
Policies in-force (m)	5.29	0.22	5.08	5.32	5.51
Premium ($b)	3.32	0.13	3.26	3.30	3.41
Coverage ($t)	1.26	0.03	1.24	1.26	1.28
Premium per policy ($)	628	35.8	599	647	654
Coverage per policy ($k)	238	12.1	229	239	249
Flood-prone policies (%)	52.2	3.8	49.6	52.9	55.7
	Panel B: County-level Cross-Sectional Variation				
	mean	s.d.	25^{th} pctl.	50^{th} pctl.	75^{th} pctl.
Policies in-force	1,766	12,563	31	120	437
Premium ($k)	1,108	6,251	21	88	325
Coverage ($m)	421	3,023	4.8	20	83
Premium per policy ($)	754	358	554	697	876
Coverage per policy ($k)	185	68	137	185	232
Flood-prone policies (%)	53.3	23.8	37.8	55.2	70.8

Table 1.2: Insurance Demand and Flood-risk-map Campaign

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \epsilon_{it}$. The panel covers 3,053 counties from January 2010 to August 2019. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. The main explanatory variable $Campaign_{it}$ is a binary variable indicating if county i has publicized its new flood risk maps at time t; this term is set to zero for the control counties without campaigns. $Campaign_{it}$ aggregates the community-level campaign process to the county level: treatment is defined as more than 50 percent of the communities in county i runs the campaign in the same month. Alternatively, $Campaign_{it}$ is defined as a continuous variable that equal to the cumulative fraction of treated communities in county i in month t. α_i and λ_t are the county and year-month fixed effects. X_{it} are the covariates: $Premium_{it}$ is the average premium per policy (in $), and $Coverage_{it}$ is the average coverage per policy (in $k). Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
Campaign	21.39***	21.41***	20.57***	24.97***
	(5.66)	(5.66)	(5.13)	(5.78)
Premium		-0.003	-0.005	
		(0.01)	(0.01)	
Coverage		-0.069	0.005	
		(0.05)	(0.10)	
Observations	347,852	347,852	176,251	347,852
R-squared	0.69	0.69	0.75	0.69
Include never-treated	Y	Y	N	Y
Treatment construction	Discrete	Discrete	Discrete	Continuous

Table 1.3: House Prices and Flood-risk-map Campaign

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \epsilon_{it}$. The dependent variable Y_{it} measures the house price index of county i in month t. The house price data is obtained from Zillow. Zillow provides separate county-level house price indices for different types of houses, such as All Homes, Single-Family Homes, Top-tier Homes (within the 65th to 95th percentile range for a given county), Bottom-tier Homes (within the 5th to 35th percentile range for a given county), and Homes with 1, 2, 3, 4 or 5+ bedrooms. The main explanatory variable $Campaign_{it}$ is an indicator variable which turns on if county i has publicized its new flood risk maps at time t; this term is set to zero for the control counties without events. α_i and λ_t are the county and time fixed effects. X_{it} are the covariates. *Income* is the median household income (in $1,000); *Unemployment* is the unemployment rate (in percent). Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)
	Panel A: Zillow Home Value Index with Various Price Ranges				
Campaign	-1.86***	-1.69***	-1.66***	-1.56***	-1.99***
	(0.46)	(0.42)	(0.43)	(0.38)	(0.45)
Observations	312,901	289,881	289,851	289,994	289,623
R-squared	0.80	0.79	0.79	0.79	0.79
House Type	All Homes	All Homes	Single-Family	Top Tier	Bottom Tier
Covariates	N	Y	Y	Y	Y
	Panel B: Zillow Home Value Index with Various Sizes				
Campaign	-2.25***	-2.23***	-1.64***	-1.45***	-1.28***
	(0.52)	(0.45)	(0.42)	(0.41)	(0.42)
Observations	256,995	287,065	289,833	288,203	281,498
R-squared	0.72	0.75	0.79	0.80	0.79
House Type	1 bedroom	2 bedrooms	3 bedrooms	4 bedrooms	5+ bedrooms
Covariates	Y	Y	Y	Y	Y

Table 1.4: Salience and Past Flood Occurrence

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \epsilon_{it}$. All the variables are defined as per Table 2. Panel A runs the regression in subsamples of counties that have not had any flood in the previous n years prior to the flood-risk-map campaign. Panel B uses subsamples of counties that have had at least one flood in the previous n years before the campaign. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
	Panel A: no flood in the past n years					
Campaign	22.37***	25.74***	22.94**	28.08**	35.14**	47.79**
	(6.86)	(8.79)	(9.44)	(12.08)	(15.62)	(21.17)
Observations	125,560	93,121	66,571	47,745	30,718	19,358
R-squared	0.72	0.71	0.72	0.71	0.71	0.69
	Panel B: with a flood in the past n years					
Campaign	14.45***	13.84***	19.13***	17.24***	16.11***	16.09***
	(4.85)	(4.60)	(6.30)	(5.42)	(4.93)	(4.60)
Observations	48,104	80,543	107,093	125,919	142,946	154,306
R-squared	0.83	0.79	0.77	0.77	0.77	0.77

Table 1.5: Heterogeneous Effects of the Flood-risk-map Campaign

This table shows results from the regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \gamma * Campaign_{it} \times Z_i + \epsilon_{it}$. Z_{it} captures the heterogeneity of ex-ante awareness of flood risk across counties. $Income_i$ is the median household income (in $1,000) in county i. $Education_i$ is the percentage of people with college degrees (in percent). $ClimateOpinion_i$ is the percentage of people (in percent) who answered "Yes" to the question of whether they think global warming will harm them personally, which is obtained from the Yale climate opinion survey (Howe et al., 2015). $\mathbb{1}(Coastal)_i$ is a binary variable indicating if county i is from a coastal state or not. $\mathbb{1}(HighRisk)_i$ is a binary variable indicating if county i's proportion of the Special Flood Hazard Area (SFHA) is above the nationwide median. The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year. $RiskLevel_i$ is a continuous proxy for county i's flood risk level, which is the proportion of SFHA (in percent). $RiskLevel_i^2$ is the square of $RiskLevel_i$. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Campaign$	50.55***	35.42***	81.11***	112.61***	45.96***	135.49***	21.46
	(13.97)	(9.41)	(24.85)	(34.08)	(12.38)	(38.94)	(21.44)
$Campaign \times$							
$Income$	-0.65**			-0.59**		-0.78***	
	(0.27)			(0.27)		(0.29)	
$Education$		-0.65**					
		(0.31)					
$ClimateOpinion$			-1.56**	-1.72**		-1.54**	
			(0.63)	(0.67)		(0.71)	
$\mathbb{1}(Coastal)$					-21.61**	-19.73**	
					(9.09)	(9.18)	
$\mathbb{1}(HighRisk)$					-26.96***	-29.76***	
					(10.21)	(10.53)	
$RiskLevel$							1.02
							(0.83)
$RiskLevel^2$							-0.017**
							(0.0079)

Table 1.6: Flood Insurance Purchases and Social Connectedness

This table shows results from the event study: $Y_{it}^{f} = \beta_0 + \beta_1 * Connected_i^f \times Post_t^f + \beta_2 * Connected_i^f + \beta_3 * Post_t^f + \epsilon_{it}^f$. For a given flood event f and the associated flooding counties $\{j\}_f$, county i's social connectedness to $\{j\}_f$ is measured by the relative probability of Facebook friendship $p_{i,f} = \sum_{\{j\}_f} w_j * p_{i,j}$, where $p_{i,j}$ is the county-by-county probabilities obtained from Bailey et al. (2018b). w_j represents population-weighting or equal-weighting scheme. $Connected_i$ is a binary variable indicating if county i is socially connected with the flooding area, which is defined as having a value of $p_{i,f}$ above the state-median. The analysis sample consists of only counties that are geographically distant to the flooding area. Three different choices of distance threshold are considered: being 500, 750 and 1,000 miles. $Post_t$ is a binary variable indicating post-flood periods. The dependent variable Y_{it} measures the insurance demand in county i in month t, which is defined as per Table 2. Panel A uses the full sample period from January 2010 to August 2019. Panel B uses a restricted sample period from January 2014 to December 2017. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
	Panel A: Full Sample Period 2010-2019			
$Connected * Post$	1.11***	0.90***	0.86***	1.54***
	(0.20)	(0.20)	(0.19)	(0.24)
Observations	23,369,392	23,369,392	32,114,276	16,335,272
	Panel B: Restricted Sample Period 2014-2017			
$Connected * Post$	1.28***	0.86***	0.77***	1.72***
	(0.28)	(0.26)	(0.23)	(0.35)
Observations	8,760,517	8,760,517	12,028,168	6,019,329
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750	750	500	1000

Table 1.7: Heterogeneity of Social Connectedness and Flood Salience

This table shows results from the event study: $Y_{it}^f = \beta_0 + \beta_1 * Connected_i^f \times Post_t^f + \beta_2 * Connected_i^f + \beta_3 * Post_t^f + \epsilon_{it}^f$. Panel A and B consider two deviations from the baseline specifications presented in Table 6; otherwise, the variables are defined as per Table 6. In Panel A, $Connected_i$ is defined as a binary variable that equals one if county i has a connectedness measure in the top quartile of the state and equals zero if county i has a connectedness measure in the bottom quartile of the state. Panel B only includes the significant flood events defined by the FEMA. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
	Panel A: Top vs. Bottom Quartiles of Connectedness			
$Connected * Post$	2.20***	1.74***	1.91***	2.98***
	(0.30)	(0.30)	(0.30)	(0.37)
Observations	11,965,523	11,965,523	12,576,973	8,101,031
	Panel B: Subsample of Significant Floods			
$Connected * Post$	3.08**	4.57***	4.12**	5.95***
	(1.43)	(1.44)	(1.64)	(2.02)
Observations	688,352	688,352	1,107,478	381,157
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750	750	500	1000

Table 1.8: Alternative Methodology of Estimating the Causal Effect of Social Connectedness

This table shows results from regression: $log(Policies)_{i,t} = \beta * FriendFlood^{Distant}_{i,t-k,t} + FE_{state\times time} + \epsilon_t$. Following the methodology proposed by Bailey et al. (2018a), $FriendFlood^{N}_{i,t-k,t}$ measures the average flood experience of a county i's social network N between $t-k$ and t. $FriendFlood^{N}_{i,t-k,t}$ is calculated as the weighted average as $\sum_j \theta^N_{i,j} * Flood_{j,t-k,t}$, where $\theta^N_{i,j}$ is share of county i's friends in network N who live in county j, and $Flood_{j,t-k,t}$ is the number of floods in county j between $t-k$ and t. A geographically distant network $N = Distant$ is a set of counties that are certain miles away from county i. Columns 1 through 4 show results of using 750 miles as the threshold; columns 5 through 7 use 250, 500, and 1,000 miles, respectively. The measurement window (i.e. k) of floods takes values of 3, 6, 12 or 24 months. $FE_{state\times time}$ are the state×time fixed effects. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Geographically Distant Friends' Flood Experiences						
$FriendFlood^{Distant}_{i,t-k,t}$	0.013***	0.012***	0.011***	0.013***	0.009***	0.011***	0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	321,511	349,443	340,120	284,729	321,511	321,511	321,511
R-squared	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	Panel B: Geographically Distant Friends' Significant Flood Experiences Only						
$FriendFlood^{Distant}_{i,t-k,t}$	0.042***	0.041***	0.041***	0.033***	0.043***	0.067***	0.042***
	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.006)	(0.004)
Observations	321,511	349,443	340,120	284,729	321,511	321,511	321,511
R-squared	0.42	0.42	0.42	0.42	0.42	0.42	0.42
Distance Threshold	750 miles	750 miles	750 miles	750 miles	250 miles	500 miles	1,000 miles
Flood Window	12 months	3 months	6 months	24 months	12 months	12 months	12 months

1.9 Appendix

A. Multiple Publications of Risk Maps

As discussed in Section 1.2.3, the FEMA's Community Status Information only provides the publication date of the latest map. As the FEMA aims to review their maps every five years, in principle, a county may have two publication dates during my 10-year sample period, and in which case, the first publication should capture the treatment of interest. In this section, I present a set of evidence to show that the FEMA fails this goal, and in reality, new publications take place much longer than every five years.

First, in an official audit report titled "FEMA Needs to Improve Management of Its Flood Mapping Programs" published in September 2017, evidence suggests that more than half of the database falls behind schedules.

Second, according to the FEMA's Community Status Information (as of June 2020), almost 75% of the communities have an effective date more than five years old, i.e., the latest update was before June 2015. Moreover, 37% or 13% of the maps are more than 10 or 20 years old. These statistics indicate that the FEMA has been struggling to keep pace with its goal.

Third, I have downloaded the Community Status Information at two points in time—December 2019 and June 2020. By comparing the effective dates in the two downloads, I can identify a sample of communities that have published new maps in 2020, i.e., the communities with two different dates in the two downloads. For these communities, I can impute the time spell between the two publications. I find 412 such cases in total, and on average, it takes 11.5 years to publish a new map.

B. The New Map Service Center and Information Cost

In July 2014, the FEMA launched a new online portal, known as the Map Service Center (MSC), to replace the legacy one. The new portal enhances address search, integrates products, improves user interface, and provides a variety of other upgrades and new features. More details can be found in the FEMA's newsletter.[25]

Appendix Figure A1.9 presents screenshots of the new and old websites, obtained from the Wayback Machine. Consistent with the timeline discussed above, the old portal's last appearance was on July 22, 2014, and the new MSC is available since July 28, 2014. As shown, the announcement on the website said, "Welcome to the New FEMA Flood Map Service Center! A series of major changes, including a complete site redesign, have taken effect on the MSC. All flood hazard products are now available free of charge, and the former products catalog has been replaced with an integrated Search All Products feature that allows you to find and download all products for a geographic area."

In this setting, I construct the treatment group as counties that had published the new maps before the new MSC. Thus, households in the treated counties have experienced using both portals. In comparison, the control group consists of counties without new maps yet. Thus, the upgrade is irrelevant. I estimate a standard difference-in-difference model:

$$Y_{it} = \beta_0 + \beta_1 * Treated_i \times Post_t + \beta_2 * Treated_i + \beta_3 * Post_t + \epsilon_{it} \tag{1.7}$$

$Treated_i$ is the treatment dummy indicating whether county i has published new risk maps. $Post_t$ is a binary variable indicating if t is posterior to July 2014. β_1 is the difference-in-differences estimate of interest.

[25] https://www.fema.gov/media-library-data/1405342400259-4b9d70489f7e9f6ffd90ba001182f112/The+New+FEMA+Flood+Map+Service+Center.pdf

Appendix Figures and Tables

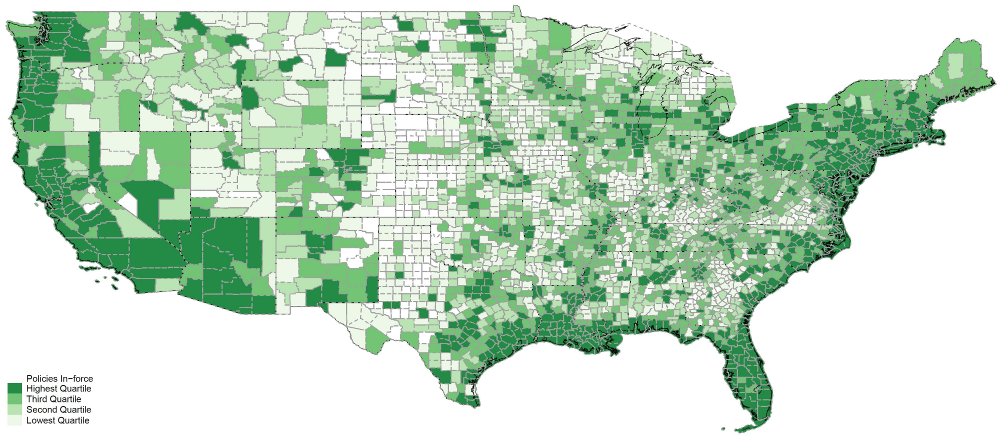


Figure A1.1. The Number of Policies In-force (Averages of 2010-2019)

This figure shows a heat map of the geographical distribution of the number of flood insurance policies in-force at the county level. Darker shades represent higher densities.

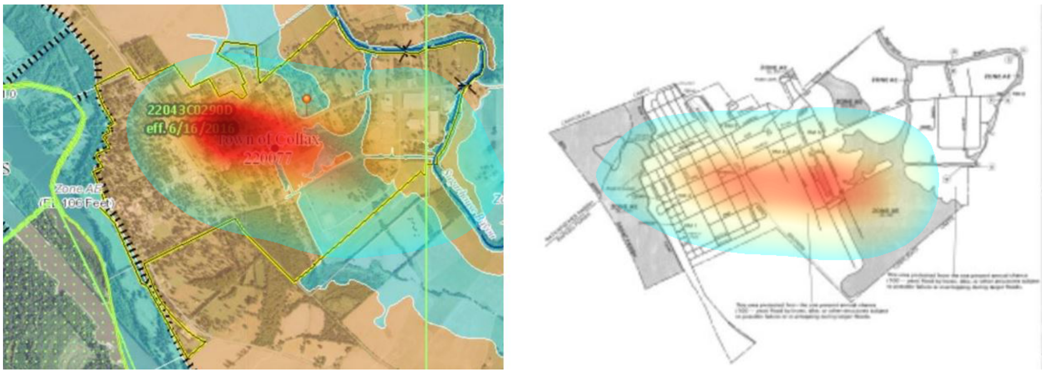


Figure A1.2. Predicted Human Eye Fixations

This figure shows the new colorized flood hazard map and the corresponding black-and-white existing map side by side as one image, and it shows an overlay heat map of predicted human eye fixations on the image. The prediction is generated by a machine-learning-based methodology called the Saliency Attentive Model (SAM) developed by Cornia et al. (2018). The darker the heat map, the more attention is allocated to that area of the image. *Source*: the Federal Emergency Management Agency.

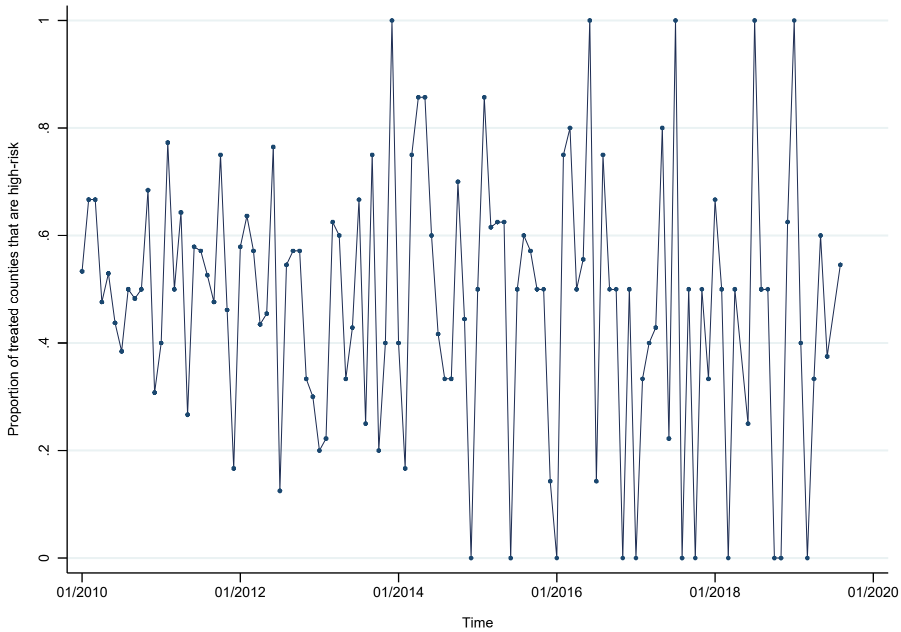


Figure A1.3. Campaign Timing and Flood Risk Levels

This figure plots the time-series of the share of high-risk counties among newly treated counties. For each month, the y-variable is calculated as the number of high-risk counties that publicize new flood risk maps divided by the total number of counties that publicize new flood risk maps in that month. A county is defined as high-risk if its proportion of the Special Flood Hazard Area (SFHA) is above the nationwide median. The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year.

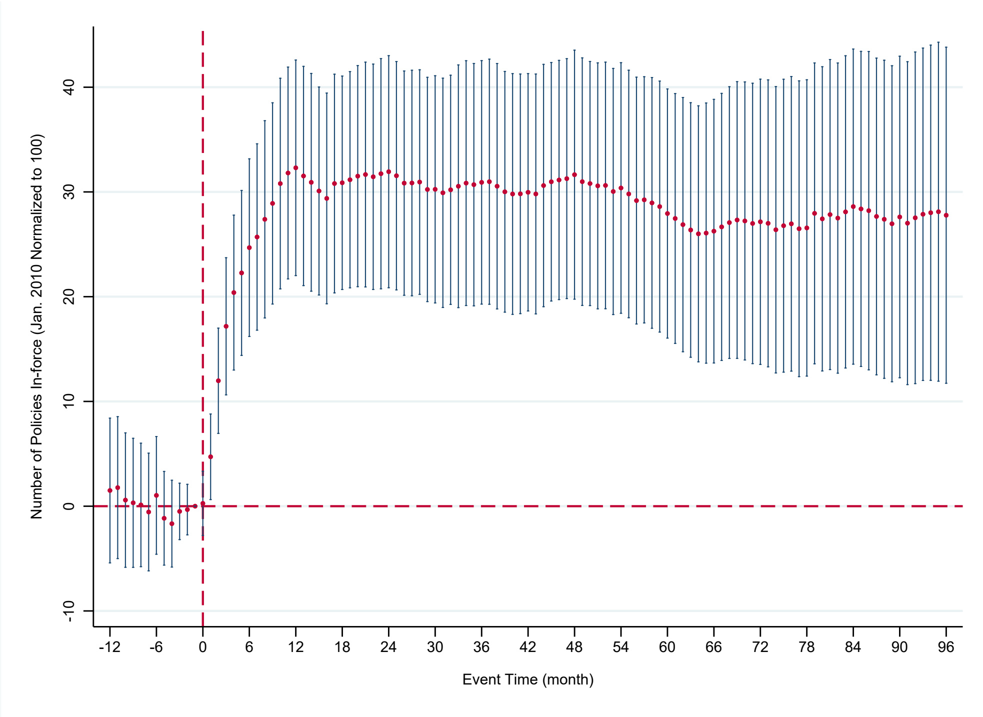


Figure A1.4. The Long-run Impact of Flood-risk-map Campaign

This figure shows the dynamic effects of the flood-risk-map campaign on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. t_i^* is the publication time of the new risk maps in county i. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

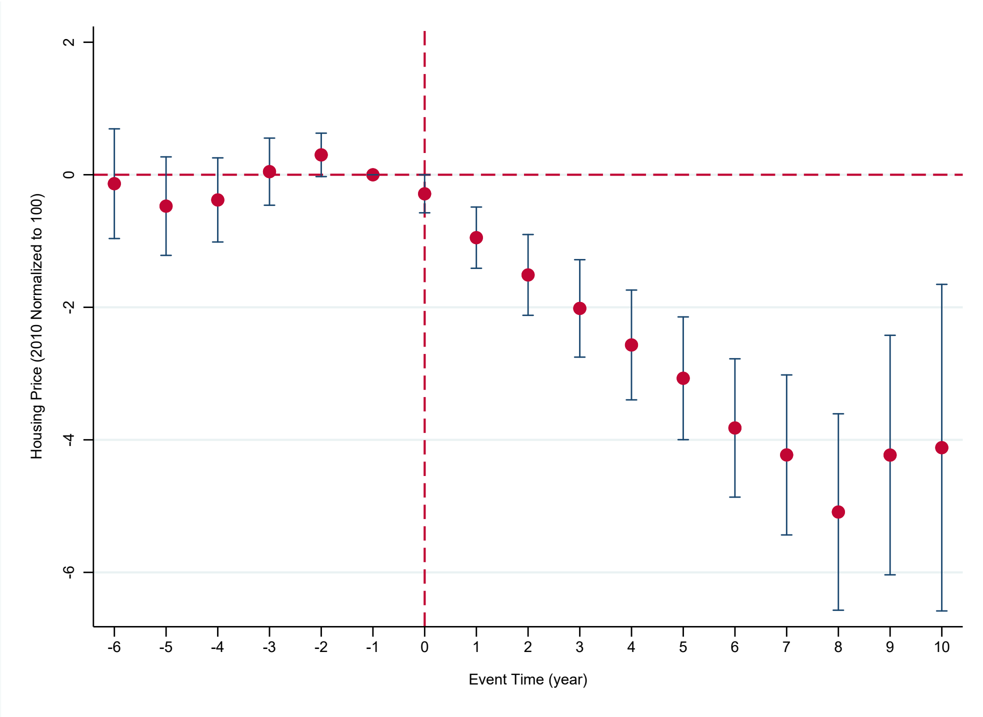
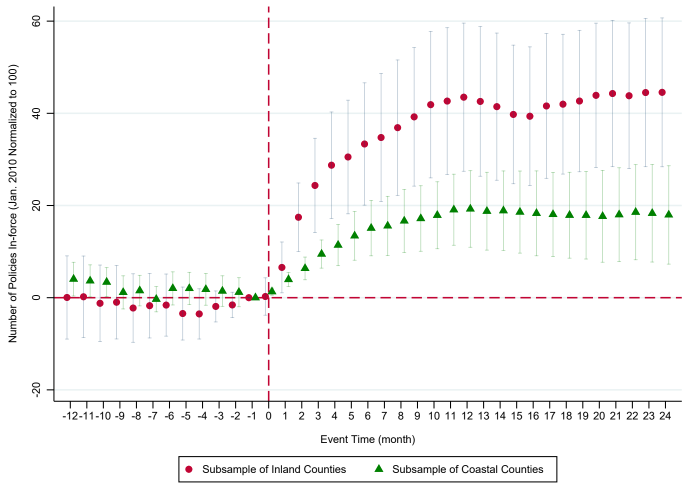
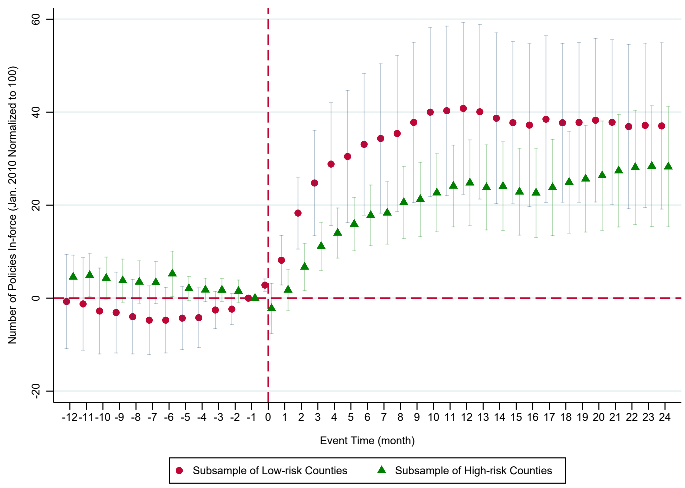


Figure A1.5. The Impact of Flood-risk-map Campaign on House Prices

This figure shows the dynamic effects of the flood-risk-map campaign on county-level house price index. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $HousePrice_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable $HousePrice_{it}$ is the house price index (with January 2010 normalized to 100) in county i in year t, which is obtained from the Federal Housing Finance Agency. t_i^* is the year when county i publicizes its new flood risk maps. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) Subsamples of Inland vs. Coastal Counties



(b) Subsamples of High vs. Low Share of SFHA

Figure A1.6. Heterogeneity by Flood Risk Across Counties

This figure shows the dynamic effects of the flood-risk-map campaign on insurance purchases, in subsamples. All the variables and the regression are defined as per Figure 3. Figure (a) splits the sample by whether the county is from a coastal state or not. Figure (b) splits the sample by whether the county has an above- or below-median value of the proportion of the Special Flood Hazard Area (SFHA). The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

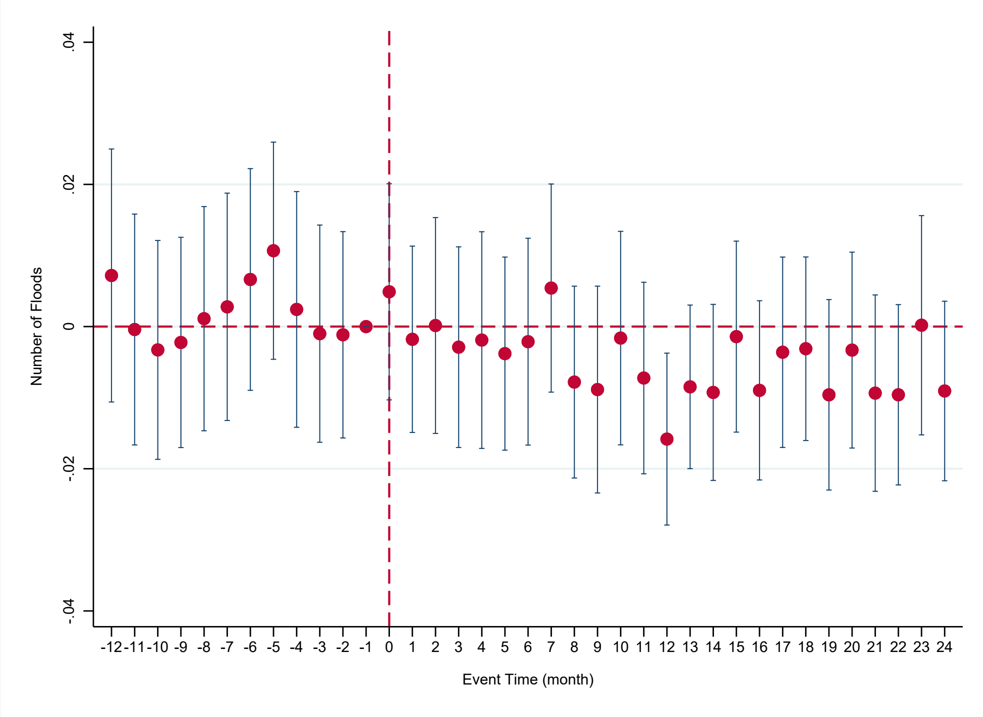
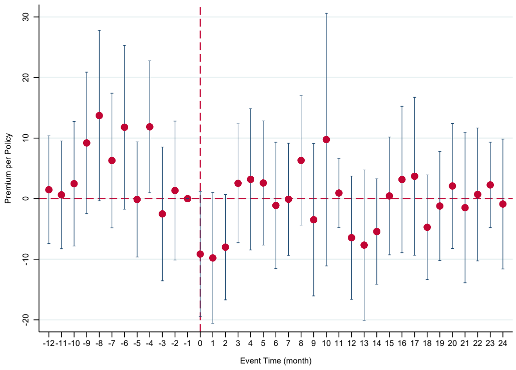
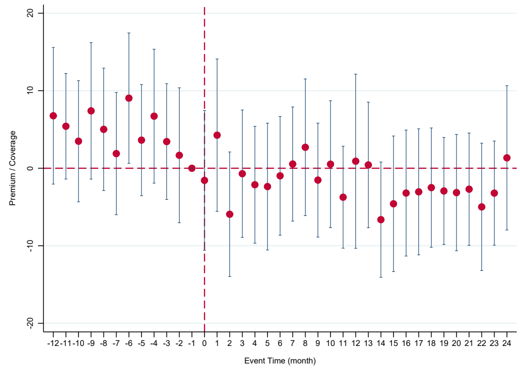


Figure A1.7. The Impact of Flood-risk-map Campaign on Flood Occurrence

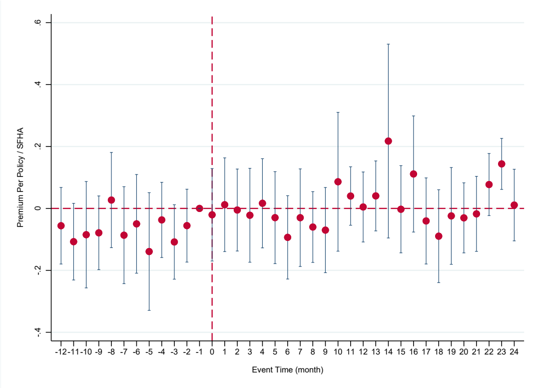
This figure shows the dynamic effects of the flood-risk-map campaign on flood occurrence. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. The dependent variable Y_{it} measures the number of floods in county i in month t. All the other variables are defined as per Figure 3. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



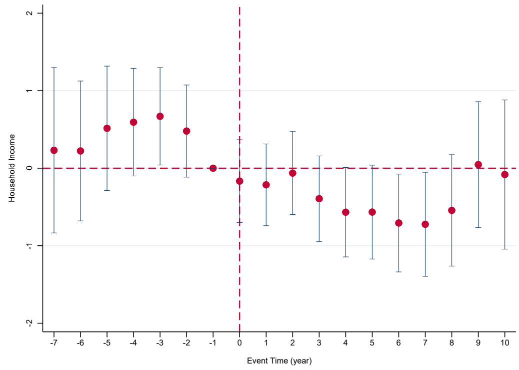
(a) Premium



(b) Premium/Coverage



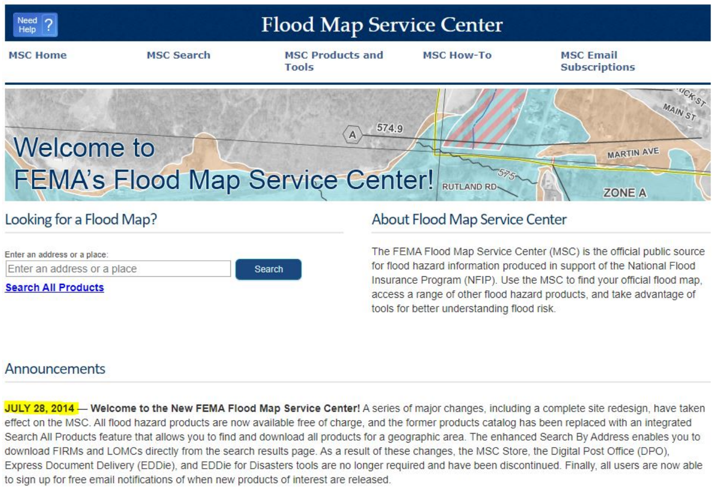
(c) Premium/SFHA



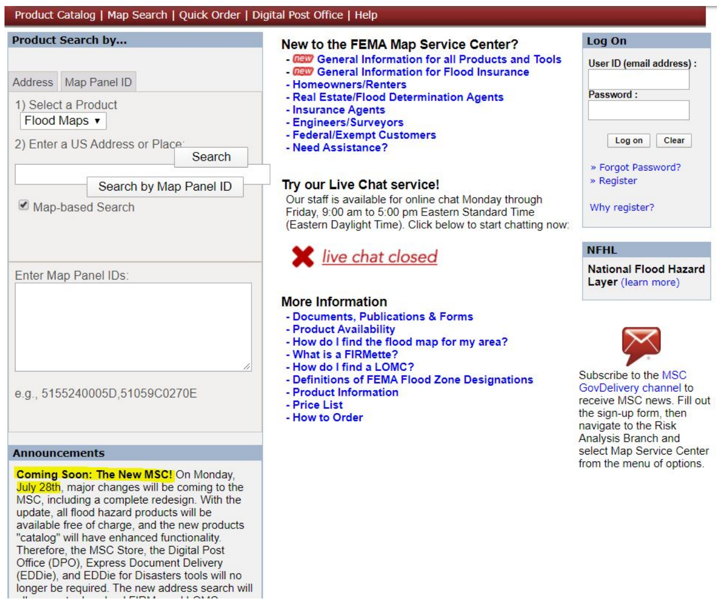
(d) Household Income

Figure A1.8. Flood-risk-map Campaign and Price of Flood Insurance

This figure shows the dynamic effects of the flood-risk-map campaign on the cost of buying a flood insurance policy. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. In figure (a), Y_{it} measures the average premium per policy in county i in month t. In figure (b), Y_{it} measures the average premium per policy per \$1000 coverage in county i in month t. In figure (c), Y_{it} measures the average premium per policy divided by the fraction of SFHA in county i in month t. In figure (d), Y_{it} measures the median household income in county i in year t. All the other variables are defined as per Figure 3. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) New Portal of the Map Service Center



(b) Legacy Portal of the Map Service Center

Figure A1.9. The New FEMA Map Service Center (MSC) Launched in July 2014

This figure shows the FEMA's new and old portal of its online GIS database. The screenshots were taken on different dates in July 2014 by the Wayback Machine. The new Map Service Center (MSC) was officially launched on 28 July, 2014. The last appearance of the old website (in the library of the Wayback Machine) was on 22 July, 2014. *Source*: the Federal Emergency Management Agency.

Table A1.1: Insurance Demand and Flood-risk-map Campaign

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \epsilon_{it}$. The main explanatory variable $Campaign_{it}$ is a binary variable indicating if county i has publicized the new flood risk maps at time t; this term is set to zero for the control counties without campaigns. $Campaign_{it}$ aggregates the community-level campaign process to the county level: treatment is defined as that more than 50, 75 or 90 percent of the population in county i gets exposures to the campaign in the same month. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)
Campaign	20.89***	20.87***	18.63***	25.25***	26.28***
	(6.78)	(6.66)	(6.23)	(6.61)	(7.03)
Premium		-0.003	-0.023*		
		(0.007)	(0.012)		
Coverage		-0.070	-0.024		
		(0.050)	(0.097)		
Observations	347,852	347,852	241,271	347,852	347,852
R-squared	0.69	0.69	0.70	0.69	0.69
Include never-treated	Y	Y	N	Y	Y
Treatment definition	>50%	>50%	>50%	>75%	>90%

Table A1.2: Various Specifications of Event Windows

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \epsilon_{it}$. All variables are defined as per Table 2. In this table, each specification considers a different sample choice, where only observations within a certain leads and lags around the campaign date are included. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Campaign	20.37***	20.50***	20.55***	20.02***	21.52***	21.63***	21.59***	20.93***
	(5.63)	(5.53)	(5.42)	(5.24)	(5.73)	(5.63)	(5.51)	(5.30)
Observations	289,251	274,681	259,459	244,035	284,868	270,298	255,076	239,652
R-squared	0.66	0.65	0.64	0.63	0.66	0.65	0.64	0.63
Leads	60	60	60	60	48	48	48	48
Lags	60	48	36	24	60	48	36	24
Campaign	23.19***	23.24***	23.04***	22.09***	24.32***	24.41***	24.10***	22.93***
	(5.99)	(5.87)	(5.73)	(5.47)	(5.90)	(5.82)	(5.71)	(5.49)
Observations	279,583	265,013	249,791	234,367	273,062	258,492	243,270	227,846
R-squared	0.66	0.65	0.64	0.63	0.67	0.66	0.65	0.64
Leads	36	36	36	36	24	24	24	24
Lags	60	48	36	24	60	48	36	24

Table A1.3: Salience and Past Floods

This table repeats the analysis in Table 4 with an extra restriction on sample selections. The subsamples of counties are constructed such that: (1) they have not had any flood in the previous n years prior to the flood-risk-map campaign; (2) they have low (in Panel A) or high (in Panel B) inherent flood risk, which is defined as having a below- or above-median value of *RiskLevel* (defined in Table 5). This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \epsilon_{it}$. All the variables are defined as per Table 4. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
	Panel A: Low-risk counties & No flood in the past n years					
Campaign	31.91***	40.31**	35.24**	40.41**	49.53*	66.49*
	(12.21)	(15.80)	(16.35)	(20.23)	(25.96)	(34.14)
Observations	66,359	49,761	37,265	26,884	17,745	11,598
R-squared	0.70	0.69	0.70	0.70	0.69	0.69
	Panel B: High-risk counties & No flood in the past n years					
Campaign	5.055*	5.714*	6.324*	6.361*	11.570**	17.794**
	(2.795)	(3.377)	(3.642)	(3.516)	(4.513)	(7.763)
Observations	59,212	43,586	29,648	20,855	12,857	7,760
R-squared	0.784	0.834	0.849	0.847	0.813	0.379

Table A1.4: Insurance Purchases in SFHA and Non-SFHA

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \epsilon_{it}$. The dependent variable Y_{it} differs in columns 1 through 4. The first specification examines the number of policies in-force held by SFHA households, with the value of January 2010 normalized to 100. The second specification examines the number of policies in-force held by non-SFHA households, with the value of January 2010 normalized to 100. The third specification takes the difference between the two. The fourth specification examines the fraction of policies held by SFHA households relative to the total policies in-force. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
Campaign	7.96**	53.02***	-43.64**	-0.047***
	(3.34)	(20.11)	(20.46)	(0.005)
Constant	106.70***	112.85***	-5.17	0.55***
	(1.27)	(7.66)	(7.83)	(0.002)
Dependent Variable	SFHA Policies	Non-SFHA Policies	SFHA – Non-SFHA	SFHA/Total
Observations	329,114	339,042	320,484	347,672
R-squared	0.75	0.65	0.65	0.86

Table A1.5: Upgrade of Map Service Center and Insurance Purchase

This table shows results from the difference-in-differences regression: $Y_{it} = \beta_0 + \beta_1 * Treated_i \times Post_t + \beta_2 * Treated_i + \beta_3 * Post_t + \epsilon_{it}$. $Treated_i$ is the treatment dummy indicating whether county i has publicized new flood risk maps or not. $Post_t$ is a binary variable indicating if t is posterior to July 2014, which is the launch date of the FEMA's new online portal (called the Map Service Center). The other variables are defined as per Table 2. β_1 is the difference-in-differences estimate of interest. I run the regression in a sample from July 2011 to July 2017, i.e. three years before and after the launch of the new portal. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
$Treated \times Post$	4.52	4.85	6.07	6.14	5.98	6.14
	(3.15)	(3.16)	(4.04)	(4.07)	(4.61)	(4.68)
Observations	216,401	216,401	216,401	216,401	216,401	216,401
R-squared	0.006	0.009	0.011	0.014	0.012	0.015
Covariates	N	Y	N	Y	N	Y
Treatment Definition	>50%	>50%	>75%	>75%	>90%	>90%

Table A1.6: Past Floods Predicting Campaign Timing

This table shows results from the regression: $Campaign_{it} = \sum_k \beta_k * Flood_{i,t-k} + \epsilon_{it}$. $Campaign_{it}$ is a binary variable indicating county i publicizes its new flood risk maps in month t. $Flood_{i,t-k}$ is a binary variable indicating county i has a flood in time $t-k$. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)
$Flood_t$	0.001		0.001	0.001	0.001
	(0.001)		(0.001)	(0.001)	(0.001)
$Flood_{t-1}$		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
$Flood_{t-2}$				-0.000	-0.000
				(0.001)	(0.001)
$Flood_{t-3}$				0.000	0.000
				(0.001)	(0.001)
$Flood_{t-4}$					0.000
					(0.001)
$Flood_{t-5}$					0.001
					(0.001)
$Flood_{t-6}$					0.000
					(0.001)
$Constant$	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	359,740	356,611	356,611	350,361	341,014
R-squared	0.00	0.00	0.00	0.00	0.00

Chapter 2

Redeploying the Jockeys: Do VCs Create Internal Labor Markets For Entrepreneurs?

Zhongchen Hu[1]

2.1 Introduction

There is an ongoing debate among practitioners and academics, whether venture capitalists (VCs) bet on a startup's business idea ("the horse") or founders ("the jockey").[2] Kaplan et al. (2009) suggest VCs bet on the horse, as successful startups often have kept businesses stable but replaced founders. However, practitioners report or reveal greater weights on the jockey across all stages (Gompers et al., 2020; Bernstein et al., 2017). Competing theories of the firm underlie the debate: VCs should focus on the founders if human assets are the critical resources (Wernerfelt, 1984; Rajan and Zingales, 2001); on the contrary, VCs should pay substantial attention to the ideas if nonhuman assets define the firm (Grossman

[1]I am thankful for the advice of Ashwini Agrawal, Ulf Axelson, Fabrizio Core, Vicente Cuñat, Juanita González-Uribe, Isaac Hacamo, Dirk Jenter, Dong Lou, Radoslawa Nikolowa, Daniel Paravisini, Marco Pelosi, Morten Sørensen, Moqi Xu, and Hongda Zhong. I am also grateful to seminar participants at the London School of Economics and the EFA Doctoral Tutorial for helpful comments.

[2]For brevity, I use VCs as an abbreviation interchangeably for either venture capitalists or venture capital firms (partnerships of venture capitalists). When necessary, I distinguish the two by VC partner and VC firm.

and Hart, 1986; Hart and Moore, 1990).

This paper hypothesizes that VCs redeploy founders (jockeys) to different firms (horses), rationalizing the seeming contradiction between what VCs do and say. Rajan's (2012) model suggests that a founder's human capital is critical to differentiate her firm in the early stage, but in order to raise substantial funds, the firm's human capital needs to be replaceable so that outside financiers can obtain control rights. The founder turnover in IPO firms documented by Kaplan et al. (2009) is consistent with this model and highlights the importance of non-human assets. However, the replaced founders are not necessarily undervalued and ditched; instead, VCs can "recycle" them to young startups. Creating an internal labor market (ILM) reflects VCs' emphasis on human assets—we should not observe an ILM if human assets do not matter or VCs do not care.

More generally, the ILM provides a novel view of how VCs create value, which is to match talent to firms. Intuitively, the ILM is not limited to founders from successful ventures only. Unsuccessful startups constitute a significant fraction of the market, yet we know little about how VCs react to failed founders. Leveraging the redeployability of human assets via the ILM, VCs can mitigate their risk due to underdiversified investments on startups with inherent high failure rates. Moreover, with the ILM, the deal selection is not necessarily a case-by-case independent decision. Because of the real option value, VCs can strategically invest in good founders with mediocre ideas and vice versa.

Testing whether VCs actively redeploy entrepreneurs across firms is challenging. First, it is difficult to identify the causal effect of VCs on labor mobility. The matching of both VCs and entrepreneurs to firms is endogenous. It is at least in part driven by unobservable factors, such as an affinity for specific technologies, industries, or other firm characteristics. Thus, labor mobility across firms in a VC's portfolio might be driven by these shared affinities rather than through connections to the common VC. Second, answering this question requires collecting data on the employment histories of VC-backed founders, which has not previously been done in a systematic manner. Additionally, the need to track every founder's job-to-job movements, especially across small entrepreneurial firms, further raises the empirical bar.

I overcome the data challenge by hand-collecting a new data set of 17,724 VC-backed founders with their full resumes on LinkedIn, which covers the universe of startups invested by the largest 200 VCs. I identify 7,353 cases of founder-startup separations, and I track the founders' post-separation employment choices. To the best of my knowledge, this paper is the first to provide a systematic description of founders' labor mobility across entrepreneurial firms in a large and comprehensive sample. About 60 percent of the departing founders find a new job within my sample of VC-backed firms, while outside options include creating new startups, retiring, or joining firms outside my sample.[3]

I find strong evidence of an internal labor market within VCs' portfolios. The likelihood of a departing founder staying in a given financing VC's portfolio is 5.17%, which is more than twice the 1.52% probability of joining a *matched* non-financing VC's portfolio, which is matched to have the same geographical and industrial focuses. The difference reduces but remains sizable at 2.35 percentage points after including VC-month, pairwise location, and pairwise industry fixed effects.[4] These fixed effects alleviate the concern that the unobserved time-variant attributes of VCs (such as portfolio size or demand for talent) or the shared affinities between VCs and founders (such as industry specialization or geographical proximity) are driving the result.

Alternatively, I use future-financing VCs as the control group, and I obtain similar results. As VC financing is commonly staged via a series of fundraising rounds, *after* a founder's departure, the startup may continue thriving and receiving new funding from new VCs (termed *future-financing* in this paper). Financing and future-financing VCs both invest in the same startup, indicating similar investment choices, but the departing founder has no direct connection to the future-financing VC. Compared with the matched sample, I find this approach is less affected by the inclusion of the aforementioned fixed effects, suggesting that

[3] I do not count serial entrepreneurs as redeployment as I only investigate the existing VC-backed firms as prospective employers. For example, Elon Musk left Paypal in 2000, founded SpaceX in 2002, and joined Tesla in 2004. SpaceX and Tesla received financing from Draper Fisher Jurvetson (DFJ) afterward. In my analysis, Elon's post-Paypal career is not considered as finding a new job in DFJ-backed firms.

[4] The pairwise location fixed effects are a set of dummies for 25×25 area-pairs (a combination of the startup's location and the VC's geographical focus). Similarly, the pairwise industry fixed effects are constructed from 46 business category groups. See Section 2.3.1 for details.

future-financing VCs are similar to financing VCs with respect to unobserved affinities.

To further address the concern that unobserved affinities confound founders' post-separation mobility and VCs' ex-ante selection of portfolio firms, I consider a second-difference that exploits within-VC variations. The idea is that VCs should have little incentive to redeploy valuable human capital into exited firms, as the investment has already paid off. While both types of VCs have such incentive discontinuity, the one with the non-financing VC should be irrelevant for the focal departing founder of interest. Analytically, I track her probability of finding a new job in the *past* and *contemporaneous* portfolios of the financing and the matched non-financing VCs.

I find that the difference between the contemporaneous portfolios is 5.17% versus 1.52%, whereas it is 1.85% versus 0.76% between the past portfolios. Collectively, I obtain a difference-in-differences estimate of 2.56 percentage points, mimicking the previous result of the single-difference approach with fixed effects. This extra propensity should attribute to the VC's direct effects. The identifying assumption is that the affinities-related confounder, which drives the departing founder to join her financing VC's companies than the non-financing VC's, is constant across the past and contemporaneous portfolios.

I propose a mechanism to explain the link between VCs and entrepreneurs' mobility. The ILM that I document is consistent with a model of asymmetric information, where VCs possess private information about entrepreneurs' quality and traits, motivated by the theory of asymmetric employer learning from labor economics (Waldman, 1984; Greenwald, 1986; Bernhardt, 1995). This informational advantage can be established ex-post, through VCs' active investing behaviors, such as monitoring, as well as ex-ante through pre-deal due diligence. As a result, VCs would retain valuable human capital that they have recognized in their portfolios, but they have less precise information (or none at all) about the entrepreneurs in other's portfolios.

In the following, I provide a set of additional findings to flesh out the mechanism of VCs using private information to reallocate talent. First, I show that VC partners continue redeploying entrepreneurs, even after they have joined new

VC firms. Who learns about founders, and who carries the VC-entrepreneur relationship? The information and connections are likely possessed by VC partners instead of being VC firms' organizational capital, consistent with Ewens and Rhodes-Kropf (2015). I collect all VC partners' resumes, track their job changes, and find that founders' mobility follows. Specifically, a departing founder is 63.2% more likely to join her former VC partner's new portfolio, compared to a matched control. Moreover, this effect does not exist if the moving partner and the departing founder do not share an overlapping period in the old VC firm.

The second supporting evidence is that a more active ILM is associated with a higher VC fund return, controlling for existing predictors, such as fund size and past performance. This result is consistent with my hypothesis of asymmetric learning: by reallocating valuable human capital to enhance the other portfolio firms' productivities, the VC should ultimately boost its overall return. The observed differences in ILMs could be because of VCs' differentiated capability of identifying distinctive traits and talent.

Third, I document several pieces of suggestive evidence of heterogeneity in accordance with my hypothesis. For example, the ILM is stronger for younger founders (whose ability is less known) and stronger for CEO- than CTO-founders (as CEOs' managerial skills are harder to verify than CTOs' technical skills). Besides, along the dimension of VCs, the redeployment is more likely to be made by lead VCs (who primarily interact with the founders in syndicated deals) than co-investing VCs.

I also consider two alternative mechanisms consistent with VCs' redeployment of entrepreneurs. While my analysis does not allow me to distinguish between all possible explanations, I present arguments that the asymmetric learning hypothesis discussed above is likely the primary driving factor.

The first alternative posits that the only reason VCs link departing founders (seeking new opportunities) to portfolio companies (hiring talent) is that the VCs have both sides' phone numbers. I refer to it as the Rolodex hypothesis. The key distinction is what friction of the entrepreneurial labor market VCs overcome. In the asymmetric learning hypothesis, the friction is information imperfection about founders' ability; in the Rolodex hypothesis, the friction is information imperfection about job creation and job-seekers' vacancies (timing and location). However, the key issue of the Rolodex mechanism is that it does not speak to significant value creation and VC performance.

The second alternative mechanism is that the ILM may serve as career insurance to entrepreneurs. Having a stronger ILM allows the VCs to extract rents, like reputation (Hsu, 2004). The insurance premium is effectively reflected as a lower ex-ante valuation (i.e., larger ownership shares for a given amount of financing), and the VCs fulfill job referrals ex-post. However, if the redeployment is purely about claiming insurance from VC firms, we should not observe founders following their VC partners' job changes. Also, this story does not easily explain the heterogeneous effects. For example, it is unclear why the CEO is insured more often than the CTO, considering the whole funding team accepts the same valuation.

Overall, this paper provides a novel perspective to understand in what sense VCs value human capital as they say, and more generally, to understand how VCs create value.[5] I stress that my results do not indicate that nonhuman assets are not crucial. Instead, I view the ILM documented in this paper as a complementary rationale adding to the literature. While the correlation between IPO and business stability (as opposed to founder turnover) highlights the importance of nonhuman assets (Kaplan et al., 2009), my finding of VCs redeploying (rather than ditching) the replaced founders suggests that VCs value human assets too. Moreover, the ILM is also active for founders from failed ventures (underexplored in this literature), suggesting that VCs identify and value talent. This argument is consistent with the evidence that the ILM is positively associated with VC fund return.

Relatedly, this paper contributes to the literature on active investing by VCs, specifically founder replacement. Hellmann and Puri (2002) show that VC-backed companies are more likely to replace the initial founder-CEOs with professional managers. Ewens and Marx (2018) identify a positive causal impact of founder

[5]These questions are important because VC-backed startups propel economic growth in terms of job creation (Samila and Sorenson, 2011; Haltiwanger et al., 2013; Adelino et al., 2017; Puri and Zarutskie, 2012) and innovation (Kortum and Lerner, 2000; Acs and Audretsch, 1988; Ueda and Hirukawa, 2008; Acharya and Xu, 2017), but they are particularly risky to finance due to high failure rates and severe information asymmetry (Hall and Lerner, 2010).

replacement on firm performance. While these papers emphasize the benefits of professionalization on startups, my results suggest that VCs' ability to redeploy founders is likely to affect both VCs' and founders' incentives and thus reduce frictions in facilitating professionalization.

This paper also sheds light on the role of VC's common ownership. Prior studies find that the interactions between portfolio firms in a common VC are disproportionally more likely. González-Uribe (2020) focuses on exchanges of innovation resources; Lindsey (2008) analyzes strategic alliances; Gompers and Xuan (2009) study mergers and acquisitions. My work examines the labor mobility of VC-backed founders, and my emphasis is on the intermediary role of VCs to overcome information asymmetry and reallocate human capital to efficient use. Besides, while these prior studies do not evaluate the implication of value creation, I show that a stronger ILM predicts a higher return.

Finally, this paper broadly relates to the literature on entrepreneurship. Entrepreneurs bear substantial risk yet earn no risk premium (Moskowitz and Vissing-Jørgensen, 2002; Hamilton, 2000), which begs the question of why people become entrepreneurs. Manso (2016) argues that if one accounts for the option value of returning to salaried jobs, then the risk-adjusted return is higher than thought. The ILM documented in this paper raises another factor that could mitigate entrepreneurial risk.

The remainder of the paper is organized as follows. Section 2.2 outlines the conceptual framework for my analysis and develops the hypotheses. Section 2.3 details the data. Section 2.4 describes the empirical strategies. Section 2.5 presents the empirical findings. Section 2.6 concludes.

2.2 Conceptual Framework and Hypothesis

I consider a theoretical framework that bridges labor economics theories with the venture capital literature to explain the link between VCs and the mobility of founders in their portfolios. The framework develops a number of testable hypotheses and forms the basis for the subsequent empirical analysis.

In canonical theories of labor mobility, recruiting firms can have an informational disadvantage relative to the preceding employers about workers' productivity, which is termed asymmetric employer learning in the literature (Waldman, 1984; Greenwald, 1986; Bernhardt, 1995). When workers enter the labor market, their ability is unknown, and only the incumbent employer learns about it by direct observations.

In my setting, an entrepreneur's ability, which critically determines her startup's probability to succeed, is likely private information. How do VCs learn about an entrepreneur's ability? Firstly, VC's investment decisions may have already reflected selection on unobservable quality through pre-deal investigations. More importantly, VCs engage in ex-post value-adding activities that require frequent interactions, such as board meetings, through which VCs would gradually establish informational advantage relative to outsiders. As a result, VCs identify high-ability founders in their portfolios, but they do not have as precise information (or none at all) about the founders in other's portfolios.

Therefore, the model's central prediction is that when a founder separates from her initial startup, the financing VC is more likely to retain her inside its portfolio than a non-financing VC. This mechanism also predicts several follow-up implications. First, after a VC partner has changed VC firm, she is likely to keep materializing her private information and redeploying the entrepreneurs from the first VC firm into her new portfolio. Second, the mechanism implies the ILM is value-adding to VC performance. Third, the redeployment should be active where information asymmetry is severe.

2.3 Data

To conduct this study, I hand-collect a large sample of VC-backed startups and founders. The primary objective is to track entrepreneurs' job-to-job movements across portfolio firms. In this section, I describe details of the data construction, present descriptive statistics, and discuss the advantages and limitations of this unique data set.

2.3.1 Data on VCs Financing and Portfolio Firms

The main data set begins with a comprehensive investment record of the largest 200 VCs from 1981 to 2018 provided by Crunchbase, a commercial database of VCs and startups. It contains deal-level information on the transaction date, deal size, investment round, investors, and startup-level details.

The sample consists of 11,090 VC-backed firms and 20,600 founders. Panel A of Table 2.1 presents summary statistics on the portfolio firms. The average number of founders per firm is 2.01, which is consistent with prior work.[6] The average firm is founded in 2008 and raises $68.97 million via 1.75 fundraising rounds participated by 1.67 VCs. Figure 2.1 depicts the distribution of firms by geography and industry (measured by the "business category group" from Crunchbase).[7] This information is essential for constructing fixed effects to control for the proximity between VCs' and startups' locations and industries.

There are several advantages to Crunchbase. First, it collects founders' names and their personal social media profiles, which are typically unavailable in alternative databases. Second, Crunchbase also collects granular firm-level data, such as the headquarter and business category. Third, it keeps track of IPOs and acquisitions (i.e., VC exits). Lastly, Crunchbase has comprehensive coverage on VC deals, which I cross-validate with Preqin, a reliable database that gets recent academic traction (Harris et al., 2014a; Kaplan and Lerner, 2016).

2.3.2 Data on Founders' Employment Histories

The core data is the employment histories of startup founders. With these, I can identify when an entrepreneur founds her startup, when she departs (if at all), which companies she joins thereafter (if at all), and whether these companies are backed by any VCs in my sample. Following the sample of VC-backed founders collected in Section 2.3.1, I manually search and download their resumes from LinkedIn. Over 400 million users use LinkedIn for job hunting and profes-

---

[6] Ewens and Marx (2018) report an average team size of 2.15; Kaplan et al. (2009) report 1.9; Beckman (2006) reports 2.2.

[7] On Crunchbase, every company is tagged with multiple (out of 470) business categories. These detailed tags are then mapped to 46 broader category groups. The complete list can be found online: https://support.crunchbase.com/hc/en-us/articles/360009616373.

sional networking, and users upload self-reported resumes, which are visible to the general public, such as headhunters.

One major limitation is that not everyone creates a LinkedIn profile. For my sample, I identify 86% (17,724 out of 20,600) of the founders on LinkedIn. Regarding the missing resumes, the concern is: if an entrepreneur's decision on whether to create a LinkedIn profile is systematically associated with a higher (lower) tendency to join a new employer from her financing VC's portfolio, then I would overestimate (underestimate) the redeployment effect of VCs.

I address this sampling issue from two aspects. First, despite missing these founders' LinkedIn resumes, I do have their startups' firm-level information from Crunchbase. The missing sample appears statistically similar to the full sample, based on the observable firm characteristics (see Appendix Table A2.1). It suggests the lack of LinkedIn profiles may be reasonably idiosyncratic. Second, in the worst hypothetical scenario where I assume none of the missing founders are retained by their financing VCs, the redeployment effect is still economically and statistically significant.[8]

2.3.3 Identifying Founder-Startup Separations

For every founder-startup pair, I identify the startup on the founder's resume. This exercise cross-validates that a correct resume is collected and allows me to extract the time spell during which the founder works in the focal startup. Hence, I can finalize a sample of departing founders with their corresponding separation times.

The final sample consists of 7,353 separations, and Panel B of Table 2.1 presents the descriptive statistics. Upon separation, the average time spent in the startup is 5.37 years, and the average total work experience is 15.73 years. 36% of the departing founders are CEOs, and 27% are CTOs. Knowing these characteristics, I can explore the heterogeneity of VCs' ILMs across different types of

[8]Specifically, as if I had the resumes of the missing founders, I simulate the same separation process as the observed sample. I assume none of these hypothetical departing founders joins their financing VCs' portfolio companies. Even with such extreme negative bias imposed on the missing data, the key finding is preserved. The redeployment effect is 29.3% smaller in magnitude, yet it is still positive and statistically significant.

founders.

Due to data limitations, it is not possible to know the precise reason for separation. On the one hand, when startups fail, the founders inevitably move on. On the other hand, founders may quit while the startups are growing, either because of personal preference or strategic replacement. If the startup never has an exit or never raises any funding again after the founder's departure, I assume the separation is due to failures, which accounts for 38% of the sample. To be clear, having new financing certainly does not guarantee eventual success—but I only use it to proxy for whether the startup has a viable business going forward, at the time of the specific founder's departure.

Appendix Figure A2.1 presents anonymized resumes to illustrate the two different separation reasons. Figure (a) shows a case of failure, as the founder wrote, "Raised $2.4M from A16z and NEA...The company was not successful." Subsequently, as the resume indicates, in August 2016, the founder joined company S., which was also backed by A16z. Figure (b) shows an example where the founder departed when the initial startup was growing. Company E. was founded in 2010 and had Series A and B by NEA in 2011 and 2013. The CTO-founder departed in February 2015 and then joined another NEA-backed company L. At the time, company E. was thriving—it had a Series C in 2016 and was eventually acquired by WeWork in 2019.

2.3.4 Tracking Departing Founders' Mobility

Panel C of Table 2.1 presents the descriptive statistics of where the departing founders find new employment opportunities after separations. Many of them (45.82%) stay in the ecosystem of startups that are contemporaneously held by VCs. If I extend the set of new employers to the historically VC-backed firms, the fraction increases to 57.01%.

Most interestingly, the probability of finding a new job in a financing VC's contemporaneous portfolio is 5.17%, whereas it is only 0.74% in a non-financing VC's. The endogenous matching between startups and VCs, due to shared affinities, potentially explains this difference. Addressing this concern forms the core of the following sections. As a preliminary examination in this table, I check the matched non-financing VCs (by industry and location) and the future-financing VCs. The probability increases to 1.52% and 2.49%, respectively, but it is still distinctly lower than the financing VCs.

Let me conclude my discussion of the labor mobility data with a note on how I accurately track job-to-job movements. I exploit a nice digital feature of LinkedIn. When creating a LinkedIn resume, a user self-links each work experience to the employer's LinkedIn page. As a result, from a completed resume, one click on a company's logo (or name) will redirect the browser to the corresponding company's LinkedIn page. It works because a URL, which is unique by definition, is embedded. Therefore, besides all the information visible by eyes—company name, time spell, job title, and job description—I also collect the embedded URL to be my company identifier. It allows me to identify job changes across my sample firms.

2.3.5 Data on VC Partners

Following the same procedures detailed above, I search and collect the LinkedIn resumes of 2,324 venture capitalists who have ever worked as partners in any of the VC firms in my sample. I identify 130 venture capitalists who have moved from one VC firm to another, both times in partner positions. This information is essential for further investigations on the VC-entrepreneur relationships, as discussed in the introduction (also see Section 2.4.3).

2.3.6 Data on VC Performance

Lastly, I obtain VC fund-level cash flows from Preqin, which allow me to compute various common measures of VC performance, such as MIC (multiple of invested capital), IRR (internal rate of return), and PME (public market equivalent). For my sample, I identify 157 VC firms (out of 200) and 402 funds on Preqin. Incompleteness is because Preqin may miss some VCs that do not have public pension fund investors. Fortunately, this potential bias of Preqin seems minimal according to Harris et al. (2014a).[9]

2.4 Empirical Strategy

2.4.1 First Difference: Financing vs. Non-financing VC

This section (and the next) aims to isolate VCs' impact on a departing founder's future employment decision. If we simply compare the propensity of finding a new job in the financing VC's portfolio companies with the propensity in the non-financing VC's, the estimate would suffer from selection bias. The central concern is that companies select into VC portfolios because of some unobservable factors, such as affinities for specific technologies or industries. The shared affinities would drive the departing founder to join another firm in the common portfolio.

In this section, I illustrate my baseline single-difference setting, where I choose the benchmark non-financing VCs carefully, and I consider an extensive set of fixed effects. Specifically, for a given founder-startup separation, I define the benchmark in two ways. The first approach is matching by location and industry. That means the benchmark VC must have the same geographical and industrial focuses as the financing VC.[10]

The alternative benchmark choice is the future-financing VC. As VC fundraising is commonly staged, a departing founder potentially faces three types of VCs: (1) a financing VC; (2) a future-financing VC, which is a specific type of non-financing VC that has not invested in her startup as of her separation time but will invest in a future round (observed ex-post); (3) a never-financing VC, which never invests in the startup. The financing and future-financing VCs share a similar investment affinity, but the departing founder has no direct connection to the future-financing VC.

The empirical model (referred to as the first-difference) takes the following

[9] US and UK pension funds are required to provide detailed investment information under the Freedom of Information Acts. Preqin primarily obtains data on this basis, but it also sources data voluntarily from general partners and limited partners. Harris et al. (2014a) document that Preqin yields qualitatively and quantitatively similar performance results to other trustworthy databases.

[10] I define the geographical (industrial) focus of a VC firm by the most common location (business category) of the startups in its portfolio.

form:

$$Y_{ij} = \beta * Financing_{ij} + \lambda_{jt} + \lambda^{loc\times loc} + \lambda^{ind\times ind} + \epsilon_{ij} \tag{2.1}$$

where i indexes a specific separation between a founder and her startup.[11] The dependent variable Y_{ij} is a binary variable that equals one if the departing founder i ends up working in any company that VC j contemporaneously holds in its portfolio. In most analyses, I define the dependent variable over the entire post-separation spell, i.e., it switches on if the departing founder i finds a new job in VC j's portfolio anytime after she departs. The independent variable $Financing_{ij}$ is a binary variable indicating if VC j is founder i's financing VC or not.

Model (2.1) is essentially a cross-sectional regression, yet the separation time, indexed by t, plays an important role. Specifically, I only investigate the *contemporaneous* portfolio of VC j at time t, i.e., the collection of startups that VC j invested before t and has not yet exited. λ's are the fixed effects. λ_t accounts for the general trend of searching for new jobs in VC-backed companies, and λ_{jt} (subsuming λ_t if included) accounts for the time-variant attributes of the VC, such as portfolio size. $\lambda^{loc\times loc}$ represents the pairwise location fixed effects, which are a set of dummies for the 25×25 area-pairs (a combination of the startup's location and the VC's geographical focus). Similarly, the industry fixed effects $\lambda^{ind\times ind}$ are also pairwise, constructed from the 46 business categories (see Section 2.3.1 for detailed descriptions of location and industry).

2.4.2 Difference-in-Differences

In this section, I consider a second difference, to address further the concern that unobserved affinities confound entrepreneurs' post-separation mobility and VCs' ex-ante selection of portfolio firms. My strategy exploits the within-VC variations in a VC's incentive, regarding where to redeploy valuable human capital. The idea is that after exiting from a portfolio company, the VC should care little about the future trajectory of this company since the payoff has already been realized.

I take the same first-difference between the hiring probability of the financing VC's *past* portfolio and the matched non-financing VC's *past* portfolio. In

[11] For brevity, index i refers to either the founder or the startup, depending on the context.

contrast to the contemporaneous portfolio used in the previous section, a VC's past portfolio is the collection of startups that the VC has already exited before the founder's departure of interest. Differencing the two first-differences could remove the unobservable selection based on shared affinities. Thus, the remaining difference-in-differences should capture the financing VC's direct impact. The regression takes the following form:

$$\begin{aligned} Y_{ijk} = &\beta_1 * Financing_{ij} \times Holding_{jk} + \beta_2 * Financing_{ij} \\ &+ \beta_3 * Holding_{jk} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ijk} \end{aligned} \tag{2.2}$$

where k represents the past or contemporaneous group of portfolio firms. The dependent variable Y_{ijk} is a binary variable that equals one if departing founder i ends up working in group k of VC j. The independent variable of interest is the interaction between $Financing_{ij}$ and $Holding_{jk}$, where $Financing_{ij}$ is a binary variable indicating if VC j has financed startup i or not, and $Holding_{jk}$ is a binary variable indicating if VC j still holds shares of group k or has exited. The coefficient on the interaction is a difference-in-differences estimate. The fixed effects are defined as per regression (2.1).

For a given separation, the treatment represents a discontinuity in the financing VC's incentive to redeploy the departing founder to its contemporaneous portfolio firms. The strategy relies on the assumption that the affinities-related confounder, which drives the departing founder to join her financing VC's portfolio companies than to join the non-financing VC's, is constant across the two VCs' past and contemporaneous portfolios. I present evidence to support this identifying assumption in Section 2.5.2.

Let me end with a discussion on an empirical limitation of the strategy: the precise timing of VC's exits is not publicly available. Following the standard approach, I use portfolio firms' IPO and M&A events as a proxy.[12] Note that secondary sales of private equities are possible, but the data is unavailable. Since this market is tiny and illiquid, this concern is not too worrying, and the bias, if

[12]The IPO lockup period (usually 180 days) is common (Brav and Gompers, 2003). I collect this information from the IPO prospectuses using the SEC's Edgar database. 177 companies in my sample are listed outside the U.S., for which I assume a 180-day lockup. Results are robust to excluding them.

any, would lead to an underestimation.[13]

2.4.3 Who Learns About Founders?

This section aims to investigate further on the VC-entrepreneur relationship. Consistent with Ewens and Rhodes-Kropf (2015), any private information about startup founders is likely to be possessed by VC partners, instead of as part of the organizational capital. Thus, we should expect VC partners to continue utilizing their private information to redeploy the entrepreneurs they know well, even after they have changed VC firms.

I define a VC partner's job change as a venture capitalist transiting from one VC firm (the "old VC") to another (the "new VC"), both times in partner positions. For every VC partner's move, I identify a set of entrepreneurs such that: they are also in the "old VC" *before* this partner leaves, and they depart from their initial startups *after* this partner has left. Thus, by construction, these departing founders and the moving VC partner share an overlapping period in the "old VC". I analyze these founders' post-separation employment choices.

In this setting, the treated is the "new VC" that hires the specific moving VC partner. I define the control VC to hire the same number of new partners in the given year as the treated. Also, I require the control VC to have the same industry, location, and portfolio size as the treated.[14] I examine whether the departing founder is more likely to find new employment in the "new VC". The effect is estimated in the flowing form:

$$Y_{ij} = \beta * FormerPartner_{ij} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ij} \tag{2.3}$$

where i indexes a departing founder and j indexes a VC firm. $FormerPartner_{ij}$ represents the treatment, and it equals one if departing founder i's former VC partner has moved to VC j. The fixed effects are defined as per regression (2.1).

[13]The bias is likely to be one direction. That is, the VC has sold its shares in the private secondary market before the firm's IPO or M&A. In this case, the firm would be erroneously considered as a contemporaneous portfolio company in my difference-in-differences analysis. This measurement error would go against the actual contemporaneous group.

[14]In a given year, I calculate the number of portfolio firms for every VC and then classify VCs into deciles. The matching on portfolio size is by exact decile.

2.4.4 ILM and VC Performance

To examine whether stronger ILMs are associated with better VC performance, I estimate a simple model as below:

$$Return_j = \beta * ILM_j + \gamma * X_j + \lambda_v + \epsilon_j \tag{2.4}$$

where j represents a VC fund started in vintage year v. λ_v controls for the year fixed effects. Standard errors are clustered by VC firms. This regression examines whether the intensity of ILM can explain VC performance in addition to other well-known predictors X_j in the literature, such as fund size (quadratic) and past performance (Phalippou and Gottschalg, 2009; Harris et al., 2014b; Robinson and Sensoy, 2011).

The dependent variable is the VC fund return measured by MIC (multiple of invested capital), IRR (internal rate of return), or PME (public market equivalent) proposed by Kaplan and Schoar (2005). The key component of regression (2.4) is to measure how active a VC is in redeploying its startup founders. I propose a relative measure:

$$ILM_j = \frac{Join_{j,j}}{Separation_j} - \frac{Join_{j,-j}}{Separation_{-j}} \tag{2.5}$$

$Separation_j$ counts departing founders from VC j's portfolio, and $Join_{j,j}$ computes the number of departing founders who are from j and join j for new employment. Similarly, $Separation_{-j}$ counts departing founders from portfolios other than j, and $Join_{j,-j}$ tracks founders who join j from the outside.

A naive measure is to calculate the fraction of departing founders who are retained inside, i.e., the first component of formula (5). However, such an absolute measure does not necessarily mean the VC spends greater efforts in making redeployment. Instead, it could be that its portfolio firms are more successful and thus attract more labor flows irrespectively.

2.5 Empirical Results

2.5.1 First Difference: Financing vs. Non-financing VC

My single-difference analysis, as detailed in Section 2.4.1, examines the comparison between a departing founder's probability of finding a new job in her financing VC's portfolio and a comparable non-financing VC's portfolio.

Table 2.2 presents the estimates of regression (2.1). Values are expressed in percentages. The binary outcome variable is measured over the entire post-separation spell: it switches on if the departing founder i joins any portfolio company of VC j after the separation time t.

Panel A of Table 2.2 uses the baseline matched sample, which requires an exact match of VC's geographical and industrial focuses. Column 1 shows that, without any fixed effect, the difference in the hiring probability between the financing VC's portfolio and the matched non-financing VC's portfolio is 3.65 percentage points (5.17% vs. 1.52%), which is consistent with the summary statistics in Table 2.1. Columns 2 through 5 show that including VC-month, pairwise location, and pairwise industry fixed effects, reduces the coefficient. The fixed effects alleviate the concern that some unobservables, such as common specialization or time-variant portfolio size, are driving the result. In the full specification (i.e., column 5), a departing founder is 2.35 percentage points more likely to find new employment in a financing VC's portfolio, representing an increase of 155% relative to the baseline probability of 1.52 percent.

Panel B of Table 2.2 uses future-financing VCs as an alternative benchmark, which in part addresses the selection issue. Compared to Panel A, the estimate is smaller in the absence of fixed effects, and it is less affected by the fixed effects. This result is consistent with the idea that future-financing VCs have accounted for some of the selection. In the full specification (i.e., column 5), a departing founder is 2.40 percentage points more likely to find a new job in a financing VC's portfolio than in a future-financing VC's. The difference is likely due to the lack of direct connection between the departing founder and the future-financing VC.

The similarity between the estimates (with full fixed effects) in Panel A and B suggests that the fixed effects may adequately capture the unobserved affinities. Panel C of Table 2.2 provides further supporting evidence in this regard. The idea is to use future-financing VCs as the placebo financing VCs and compare them to the "totally unrelated" never-financing VCs. If my empirical model fails to account for some unobservables, we should expect a positive placebo effect, because the omitted confounders would drive the departing founder to join the future-financing VC's portfolio firms.

Columns 1 and 2 of Panel C show that the portfolio firms backed by the future-financing VCs are indeed more attractive than the never-financing VCs. Most importantly, in columns 3 through 5, once I include more fixed effects, the estimate is no longer statistically significant. Overall, the evidence suggests that the fixed effects appropriately address the selection issue.

The above results are robust to various observation windows for post-separation employment. Table 2.3 presents estimates of regression (2.1) over different post-separation windows, from one year to the entire spell (as in Table 2.2). For instance, a 5-year window means the outcome dummy turns on if departing founder i finds a new job in any portfolio company of VC j *within 5 years* after she departs from her initial startup. For brevity, Table 2.3 only shows estimates with full fixed effects.

The estimate is fairly unaffected by varying the observation windows. Specifically, the estimate using the one-year window is only slightly smaller than the estimate using the full post-separation spell. It suggests that VC's redeployment of entrepreneurs primarily occurs during the first year after the founder-startup separations.

2.5.2 Difference-in-Differences

The analysis in the previous section suggests the fixed effects competently account for the unobserved affinities. This section aims to provide further evidence to address the endogeneity problem. The difference-in-differences strategy, detailed in Section 2.4.2, exploits the changes in VCs' incentives when exiting from portfolio firms: redeploying human capital to current investment makes sense, but less so for firms already exited.

2.5.2.1 Main Results

I take the same first-difference between the hiring probability of the financing VC's *past* portfolio and the non-financing VC's *past* portfolio, replicating the baseline comparison between the contemporaneous portfolios.

Figure 2.2 depicts the raw results. The right panel shows the first-difference in the hiring probability between the *contemporaneous* portfolios (5.17% vs. 1.52% as in Section 2.5.1). The left panel shows the first-difference between the *past* portfolios (1.85% vs. 0.76%). Collectively, Figure 2.2 demonstrates a difference-in-differences estimate of 2.56 percentage points. It has a similar magnitude to the previous result of the single-difference approach with fixed effects.

Table 2.4 presents the coefficient estimates of regression (2.2). Column 1 does not include any fixed effects, and the result corroborates the finding of Figure 2.2. In columns 2 through 5, the coefficient of interest remains stable across specifications, and it is worth noting that the coefficient on $financing_{ij}$ becomes insignificant. This is consistent with the discussion in Section 2.5.1 that the fixed effects should account for the unobserved affinities that drive labor flows to the financing VCs' portfolios.

One concern of my strategy is that VCs' specializations or affinities might change over time, in which case, firms exited a long time ago might not be a suitable control for the current portfolio firms. In robustness checks, I define the past portfolios as firms exited less than 10 or 5 years ago. Appendix Table A2.2 reports the results. The estimate remains qualitatively and quantitatively similar.

The other concern is that past firms are subject to attrition and survival biases. However, as long as the attrition rate is not systematically different across the financing and the matched non-financing VCs, it would not affect the difference-in-differences estimate. Besides, the results in Appendix Table A2.2 also address this concern: using only recent exits (i.e., less than 5 years) should keep the survival bias minimal. In the same vein, the next section presents evidence that the result holds with even narrower windows around exits.

2.5.2.2 Validity of the Strategy

The crucial identifying assumption of my strategy is: the affinities-induced probabilistic difference in hiring a departing founder, between firms backed by financing VCs and by non-financing VCs, is constant across different exit timings. If true, any extra propensity towards the financing VC's contemporaneous portfolio should be due to the VC's direct redeployment efforts.

Figure 2.3 helps to shed light on the plausibility of this assumption. I partition Figure 2.2 into smaller groups by exit timings (relative to the founder's separation time t). By construction, I make each group contain an equal number of firms (one-tenth of the total), so that the probabilities are comparable across groups. For example, the first group to the right of the threshold (i.e., the separation time t) represents firms that are still in the VCs' portfolios at time t and will be exited within the next 15 months (observed ex-post). The bar in Figure 2.3 shows the probability of the departing founder joining a specific group. Dark grey represents the financing VC, and light grey represents the matched non-financing VC. The hatched bar represents the difference. I emphasize that the plot does not trace out a time series. Instead, for a given founder's separation, each bar represents different prospective employers according to their exit timings relative to the separation time.

Figure 2.3 shows that, for the exited firms, the difference between the financing and non-financing VCs (i.e., the hatched bar) is fairly constant, which supports the identifying assumption. On the other side, for firms that are still in portfolios, there is a significant jump, mainly driven by the financing VC. There is no obvious discontinuity for the non-financing VC across the threshold, suggesting that the incentive discontinuity of the non-financing VC is irrelevant for the departing founder of interest.

As Figure 2.3 only presents the raw comparisons, the hatched bar is significantly positive even for the exited firms. The fixed effects can fully explain this. Figure 2.4 provides evidence from the regression, where I replace $Holding_{jk}$ in regression (2.2) with a set of dummies for each group and interact them with

$Financing_{ij}$. Figure 2.4 presents the coefficient estimates of the interaction terms, with the 95% confidence intervals. As shown, for the exited groups, the coefficients are no longer different from zero, and there is a significant jump in the contemporaneous groups.

2.5.2.3 Placebo Tests

Aiming to support the identification strategy further, I conduct two placebo tests. In the first experiment, I shift the actual separation time t backward by 2, 4, or 6 years. The control and treatment groups are constructed the same as before, except benchmarking to the shifted threshold. Taking the backward 4-year shift as an example: portfolio firms exited by the VCs more than 48 months ago constitute the *past* group and firms exited during $(t - 48, t)$ constitute the "*contemporaneous*" group in this placebo. Panel A, B, and C of Table 2.5 present the placebo estimates. In all cases, the effect is insignificant. This result suggests that the financing VC does not differentiate the exited firms, since the VC equally care less about them.

In the second placebo, I replace the actual financing VC with a non-financing VC with the same geographical and industrial focuses. Although it is still true that the placebo VC's incentive is discontinuous between the pre-exit and post-exit portfolio firms, such discontinuity should be irrelevant for the departing founder. The insignificant estimate reported in Panel D of Table 2.5 confirms this prediction.

To briefly wrap up the findings by far, the results in Section 2.5.1 and 2.5.2 are evidence that VCs create ILMs to redeploy entrepreneurs across their portfolio companies. The rest of the paper is delegated to explore the mechanisms of the ILMs. In particular, in the following sections, I present additional empirical findings to characterize the relevance of the model of asymmetric learning, detailed in Section 2.2.

2.5.3 VC Partners' Job Changes and Founders' Mobility

The hypothesis of asymmetric learning implies that even after a VC partner has changed VC firm, she will keep materializing her private information and redeploying the entrepreneurs from the first VC firm into her new portfolio.

In this section, I exploit VC partners' job changes to trace their impact on entrepreneurs' mobility. As detailed in Section 2.4.3, I identify 130 cases of partner-to-partner job transitions of venture capitalists, and accordingly, I construct a sample of 1,339 departing founders, who have experienced that a partner of the financing VC (termed "old VC") moved to another VC (termed "new VC").

Regression (2.3) computes the difference between the probability of a departing founder joining the "new VC" (treated) and a matched VC (control). The control VC is constructed to invest in the same industry and location, to hire the same number of new partners in the given year, and to have the same number of portfolio firms in the given year, as the treated.

Panel A of Table 2.6 reports the estimates of regression (2.3). In column 1, I find that a departing founder is 0.98 percentage point more likely to find a new job in the "new VC" that her former VC partner has joined. The constant (1.56%) in this regression is the average probability of joining the control VC's portfolio companies. This value echoes the unconditional mean likelihood (1.52%) of a departing founder joining a non-financing VC's portfolio (as in Table 2.1).

The identifying assumption of this analysis is: a VC partner's choice of the new VC firm is unrelated to a departing founder's future employment choice, conditioning on the matching criteria and the fixed effects. In other words, in the absence of the VC partner's job transition, the departing founder would have the same likelihood of joining the treated or the control VC's portfolio companies.

Recall that my previous analysis in Section 2.5.1 suggests that the fixed effects plausibly account for the unobservable affinities. In columns 2 through 5 of Table 2.6, I include the fixed effects. I find that the coefficient estimate of $FormerPartner$ is not much affected. In fact, in the full specification, it increases slightly to 1.18 percentage points. This result favors the identifying assumption and suggests that the endogeneity problem is less a concern in this quasi-experiment.

To shed further light on the identifying assumption, I run the following placebo test. For a given job transition of a VC partner, I focus on entrepreneurs who obtain funding from the "old VC" *after* this partner has left. I identify 2,318 such cases. By construction, these placebo entrepreneurs did not collaborate with the moving VC partner in the "old VC". Panel B of Table 2.6 presents the placebo results. Consistent with the identifying assumption, I do not find any significant redeployment effect.

2.5.4 Internal Labor Market and VC Performance

The mechanism of asymmetric learning implies that the redeployment of valuable human capital within a VC's portfolio should add value to the VC's performance. If some VCs are better at reallocating talent, we should expect the ILM intensity to be heterogeneous. By regression (2.4) detailed in Section 2.4.4, I examine whether variations in the ILM intensity can explain VC performance.

Table 2.7 uses three different metrics to measure VC fund returns (MIC, IRR, and PME).[15] The key explanatory variable, the intensity of a VC's ILM, is measured by formula (5), as discussed in Section 2.4.4. Panel A of Table 2.7 uses the absolute measure of ILM, and Panel B uses the relative measure. The relative measure accounts for how attractive the VC's portfolio firms are to outside labor flows. Thus, it addresses the concern that the portfolio attracts entrepreneurs irrespectively (from both inside and outside) because its companies are thriving. Compared with Panel A, the estimates in Panel B have smaller magnitudes, suggesting that the relative measure does capture a certain degree of endogeneity. Nevertheless, the positive and statistically significant estimates indicate that the ILMs are positively associated with VC performance.

I note that this section does not intend to establish a causal impact of the ILM on VC performance. There could exist a confounding factor, such as VC's ability (orthogonal to the covariates: size and past performance), that simultaneously drives the variations in ILM intensity and VC performance. My result only

[15]The public market benchmark in PME is VW S&P500. For robustness, I have checked the results hold when I use Nasdaq and NYSE indices, both VW and EW, as the benchmark.

suggests that the ILM helps to explain the variations in VC performance that is not spanned by the existing predictors known in the literature.

2.5.5 Asymmetric Information and Heterogeneity

The hypothesis of asymmetric learning implies that the redeployment effect is heterogeneous across founders with different information asymmetry, and across VCs with different learning capability. I present five pieces of suggestive evidence to support this prediction.

2.5.5.1 Heterogeneity by Founders' Work Experience

The redeployment effect should be stronger for younger people with less track record. The reason is that the information about younger entrepreneurs' ability is likely to be more asymmetric, and hence the informational advantage of financing VCs is more critical.

Panel A of Table 2.8 investigates the redeployment effect for departing founders partitioned by their years of work experience. I examine subsamples of founders with above and below the median years of experience, and I present the difference-in-differences estimates, respectively. The central takeaway is that the VC's redeployment effect is about 40% larger for the less-experienced than the more-experienced (3.22% vs. 2.23%).

2.5.5.2 Heterogeneity by Founders' Roles

The two most common roles in startups—CEO and CTO—are likely to have different sensitivity to information asymmetry. CEOs typically need relatively more soft skills (such as leadership), whereas CTOs are responsible for the more easily observable aspects of a startup's performance (such as product launch). When searching for new jobs, a departing CEO is likely to face greater information asymmetry because the soft skills are harder to verify. Therefore, my hypothesis predicts that the VC's ILM is more relevant for departing CEO-founders.

Panel B of Table 2.8 divides founders into CEOs and CTOs based on their roles in the initial startups. Founders who take both roles or neither (such as

CFOs) are excluded. I find VC's redeployment intensity is two times stronger for CEOs than CTOs (3.64% vs. 1.31%). This result suggests that CEOs' ability, compared to CTOs, is harder to verify, and hence the connections to VCs matter more.

2.5.5.3 Heterogeneity by Separation Types

The financing VC is likely to know better whether a startup's failure is because of bad luck or bad quality. However, I do not rule out that the financing VC also knows better if a startup's success is because of good luck or good quality. I assume failures convey asymmetrically less negative signals to outsiders than successes for positive signals, as startup failures are relatively more common than successes.

Since the motives of departure are unobservable in data, I classify them in the following way. After a specific founder-startup separation, if the startup never has an exit or never raises any funding again, I assume the reason for separation is due to business failures. Otherwise, I assume the startup is thriving, but the founder quits for other reasons (see Section 2.3.3 for details).

Panel C of Table 2.8 shows the regression results in subsamples. I find the financing VCs redeploy both types of departing founders. Consistent with my hypothesis, the intensity is higher for the ones from failed startups than thriving startups (3.47% vs. 2.33%).

2.5.5.4 Heterogeneity by Investors' Types

VCs tend to syndicate investments with a lead investor taking charge of the project, whereas the co-investors often merely provide capital. In this case, the frequent interactions with the founding team make it easier for the lead VC to acquire private information. Therefore, my hypothesis predicts a stronger redeployment effect in the lead VC's ILM.

Panel D of Table 2.8 examines subsamples of lead VCs and co-investing VCs. The results show that both the lead and co-investing VCs are more likely to redeploy the entrepreneurs than the matched non-financing VCs. Consistent with my hypothesis, the effect is stronger for the lead VCs than the co-investing VCs (3.15% vs. 2.15%).

2.5.5.5 Heterogeneity by Geography

Nearly half of the VC-backed startups are headquartered in California (primarily in Silicon Valley), as shown in Figure 2.1. In the more concentrated California market, information asymmetry is likely to be less severe. Thus, in California, we expect to observe a weaker impact of VCs on entrepreneurs' mobility.

Panel E of Table 2.8 partitions the sample into California and non-California, where the California subsample includes departing founders and VCs who are both in California. The results suggest that the VC's ILM is significant in both subsamples, and it is stronger outside California, where information is more asymmetric (3.11% vs. 2.15%).

2.5.6 Alternative Explanations

In the final part of this paper, I assess two alternative explanations for VCs' redeployment of startup founders. While my analysis does not allow me to distinguish between all possible explanations, I present arguments that the asymmetric learning hypothesis, as discussed in the previous sections, is likely the main driving force.

2.5.6.1 Rolodex

One may argue that a VC links departing founders (who are looking for new jobs) to her portfolio companies (which are hiring new employees or partners) simply because her Rolodex (i.e., contacts) contains the phone numbers of both sides. The key labor market friction underlying this hypothesis is information imperfection about job creation and job-seekers' vacancies. In contrast, the friction underlying the asymmetric learning hypothesis is information imperfection about the founders' ability and traits.

While this theory explains the observed phenomenon of the VC-created ILMs, it is inconsistent with some of my findings. First, the key issue of this story is that it does not imply significant value creation, as evidenced in Section 2.5.4. Second, the underlying friction of information imperfection about job creation and job-seekers' vacancies leads to search costs in the canonical labor economics models (Diamond, 1982; Mortensen, 1982; Pissarides, 1985). To reconcile the heterogeneity of my findings, the model requires heterogeneous search costs—conceptually, more costly the information imperfection, stronger the ILM. However, it is unclear and perhaps counter-intuitive why the search costs appear to be higher for CEO, young, and failed founders.

2.5.6.2 Career Insurance

The other alternative hypothesis is that VCs do not acquire any private information; instead, they create ILMs to provide career insurance to entrepreneurs: if the initial startup fails, VCs facilitate the departing founder's job searches. Therefore, a stronger ILM provides greater insurance to entrepreneurs, which allows the VC to extract rents. In this context, the insurance premium is effectively reflected as entrepreneurs accepting a lower ex-ante valuation (i.e., larger ownership shares for a given amount of financing). In this model, the entrepreneurs' ability can be homogeneous.

This explanation is consistent with some of my findings. First, VCs introduce the departing founders to new jobs as per their ex-ante insurance agreement. Second, the positive correlation between ILMs and VC performance can be explained by VCs' bargaining power to charge higher premiums, i.e., lower valuations or prices.

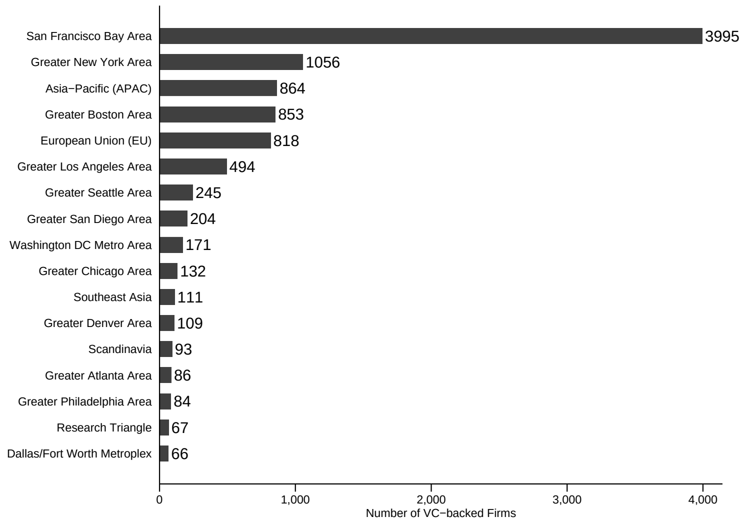
However, this alternative explanation is inadequate in the following aspects. First, as the whole founding team accepts the same valuation, it is not clear why the VCs would insure the CEOs more often ex-post than the CTOs. Second, the heterogeneity between lead VCs and co-investing VCs is also against the insurance story. Since investors coordinate on one valuation, they should provide the same insurance. Third, if it is only about claiming insurance from the VC firms, we should not observe the founders following their VC partners' job changes.

Overall, I view the key mechanism of the ILM as VCs knowing private information about the founders' ability, but I acknowledge that all three mechanisms, in reality, are probably simultaneously going on.

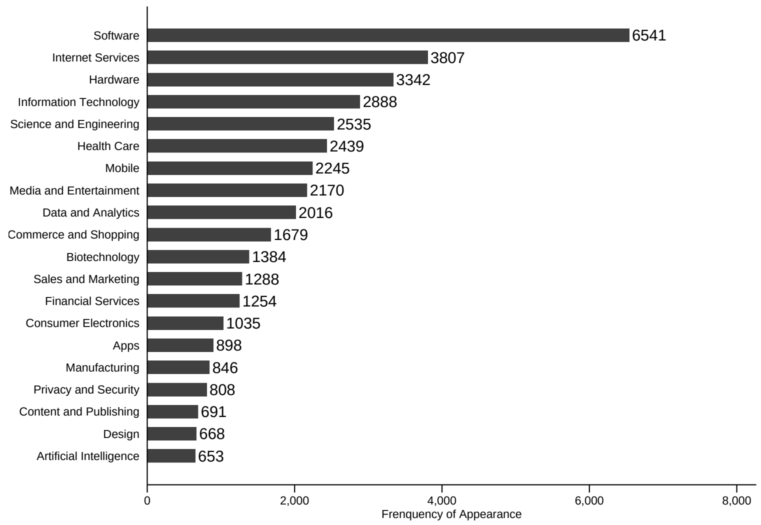
2.6 Conclusion

This paper adds to the vast literature of venture capital that seeks to understand whether VCs emphasize human capital and how VCs add value. The unique contribution of this paper is evidence that VCs redeploy entrepreneurs across their portfolio companies. My findings highlight VCs' emphasis on human assets. Given the evolving nature of a firm's defining resources and VCs' long investment horizon involving firms in different stages, this paper provides a new perspective to think about how VCs create value, which is to match talent to firms.

2.7 Figures



(A) Headquarter Location



(B) Business Category

Figure 2.1. Geographical and Industrial Distribution of VC-backed Firms

This figure shows the geographical and industrial distribution of the VC-backed startups in my sample. Panel (A) depicts the headquarter location of the startups, and Panel (B) depicts the business category (classified by Crunchbase) of the startups.

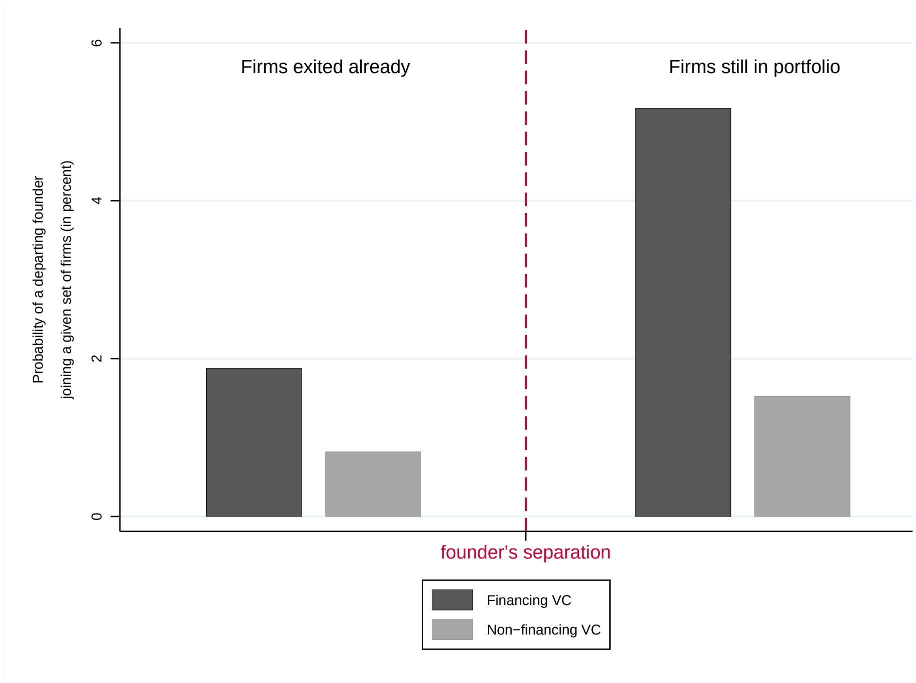


Figure 2.2. Difference-in-Differences

This figure shows the probability of a departing founder joining a specific set of VC-backed firms. The left two bars represent firms that have already been exited at the time of the founder's departure of interest (i.e., the past portfolios), while the two bars on the right represent firms that are still held by the VCs at the time (i.e., the contemporaneous portfolios). The dark (light) grey represent firms invested by the financing (matched non-financing) VC.

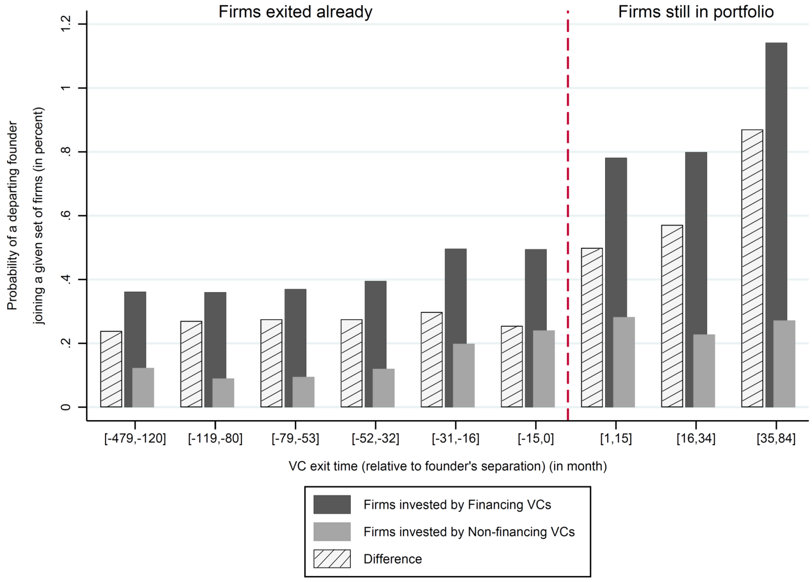


Figure 2.3. Examination of Diff-in-diff Assumption

This figure shows the probability of a departing founder joining a specific set of VC-backed firms. I partition the firms into small groups as per their exit timings relative to the separation time of the departing founder of interest. By construction, each group contains equal number of firms (one-tenth of the total), so that the probabilities are comparable across groups. Groups to the left of the threshold (i.e., the focal separation time) represent firms that have already been exited by the VCs, while groups to the right represent firms that are still held in portfolios as of the separation time. The dark (light) grey represent firms invested by the financing (matched non-financing) VC. The hatched bar represents the difference between the financing and the matched non-financing VC.

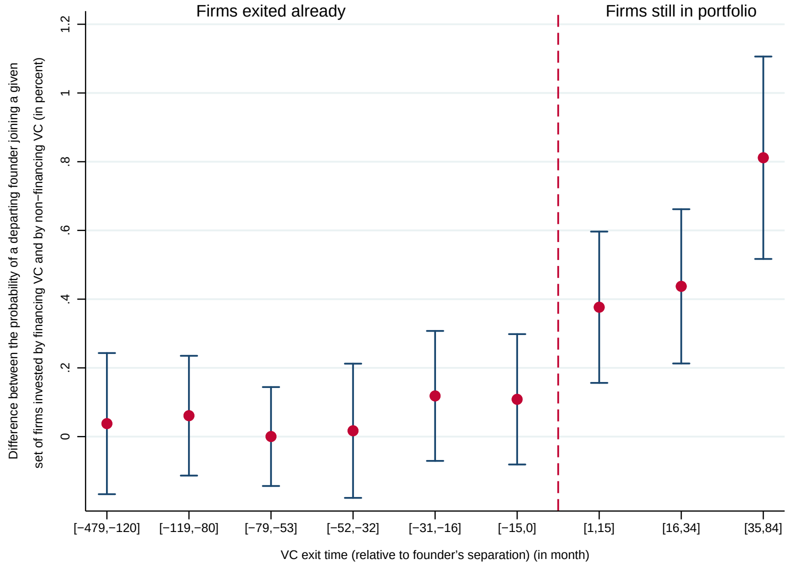


Figure 2.4. Examination of Diff-in-diff Assumption (Regression-based)

This figure plots the coefficient estimates of the interaction terms in the regression: $Y_{ijg} = \sum_g \beta_1^g * Financing_{ij} \times D_{jg} + \beta_2 * Financing_{ij} + \sum_g \beta_3^g * D_{jg} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ijg}$. The bins around the coefficient estimates represent the 95% confidence intervals. g represents the small groups partitioned as per their exit timings relative to the separation time of the departing founder of interest. D_{jg} indicates the specific group. The dependent variable Y_{ijg} is a binary variable that equals 1 if the departing founder i ends up working in group g of VC j. $Financing_{ij}$ is a binary variable indicating if VC j is the financing VC or the matched non-financing VC of startup i. The leftest group (i.e., firms that have been exited longest time ago) is omitted as the benchmark.

2.8 Tables

Table 2.1: Summary Statistics

This table presents an overview of the data set of VC-backed founders and startups. Panel A shows the firm characteristics of the 11,090 companies backed by the largest 200 VCs. Panel B describes the founders from the 7,353 cases of founder-startup separations. *Tenure* is the number of years during which the departing founder works in the initial startup. *Experience* is the total work experience, which is measured from the first job after graduation from universities to the separation time. $\mathbb{1}_{\text{CEO}}$ ($\mathbb{1}_{\text{CTO}}$) is a dummy variable indicating whether the departing founder works as a CEO (CTO) in the initial startup. $\mathbb{1}_{\text{Startup Failed}}$ indicates whether the founder leaves because the startup fails (approximated by whether the startup raises new VC financing or has an exit after the separation). Panel C reports the post-separation employment of the departing founders.

Panel A: Portfolio Firms

	Num. of Founders	Founded Year	Num. of Rounds	Num. of VCs	Total Funding (m$)	$\mathbb{1}_{\text{IPO}}$	$\mathbb{1}_{\text{M\&A}}$
mean	2.01	2008	1.75	1.67	68.97	0.072	0.33
s.d.	1.08	5.87	1.09	1.05	257.07	0.26	0.45
25^{th} pctl	1	2004	1	1	8	0	0
50^{th} pctl	2	2009	1	1	22	0	0
75^{th} pctl	3	2012	2	2	60.6	0	1

Panel B: Departing Founders & Initial Startups

	Tenure (yrs)	Experience (yrs)	$\mathbb{1}_{\text{CEO}}$	$\mathbb{1}_{\text{CTO}}$	$\mathbb{1}_{\text{Startup Failed}}$
mean	5.37	15.73	0.36	0.27	0.38
s.d.	3.6	7.72	0.48	0.44	0.49
25^{th} pctl	2.92	10.08	0	0	0
50^{th} pctl	4.5	15	0	0	0
75^{th} pctl	7	20.25	1	1	1

Panel C: Post-separation Mobility

	Joining any VC-backed firms (past & current)	Joining any current VC-backed firms	Joining a financing VC's portfolio	Joining a non-financing VC's portfolio	Joining a matched non-financing VC's portfolio	Joining a future-financing VC's portfolio
probability	57.01%	45.82%	5.17%	0.74%	1.52%	2.49%

Table 2.2: First Difference: Financing vs. Non-financing VC

This table presents results of the first-difference analyses between the probability of a departing founder joining a financing VC's portfolio and a benchmark non-financing VC's. The estimates are from the following empirical form: $Y_{ij} = \beta_1 * Financing_{ij} + \lambda_{jt} + \lambda^{loc\times loc} + \lambda^{ind\times ind} + \epsilon_{ij}$. i indexes a specific separation between a founder and her initial startup. t represents the separation time (month-level). Y_{ij} is a binary variable that equals to 1 if the departing founder i ends up working in any company in VC j's portfolio anytime post-separation. The location (industry) fixed effects are pairwise: one dummy variable for every 25×25 location-pairs (46×46 industry-pairs) between the startup and the VC. Panel A uses the matched non-financing VCs (by industry and location) as the benchmark. Panel B uses the future-financing VCs as the benchmark, where future-financing VCs are defined as VCs that only invest in the startup after the founder's departure of interest. Panel C examines a placebo test in which I switch on the $Financing_{ij}$ dummy for the future-financing VCs and compare them with the never-financing VCs. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Matched Non-Financing VC as Benchmark				
Financing	3.651***	3.708***	2.363***	2.404***	2.352***
	(0.245)	(0.248)	(0.297)	(0.327)	(0.330)
R^2	0.010	0.012	0.113	0.124	0.126
	Panel B: Future-Financing VC as Benchmark				
Financing	2.674***	2.233***	2.270***	2.286***	2.402***
	(0.485)	(0.399)	(0.542)	(0.585)	(0.583)
R^2	0.002	0.005	0.134	0.145	0.148
	Panel C: Placebo: Future-Financing vs. Never-Financing				
Financing	1.756***	1.798***	0.384	0.380	0.334
	(0.435)	(0.436)	(0.319)	(0.346)	(0.346)
R^2	0.000	0.000	0.016	0.018	0.018
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

Table 2.3: Measuring Employment Across Various Post-Separation Periods

This table presents results of the first-difference regression: $Y_{ij} = \beta_1 * Financing_{ij} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ij}$. Y_{ij} is a binary variable that equals to 1 if the departing founder i ends up working in any company in VC j's portfolio within τ year(s) after her separation. τ is specified with different values as in columns 1 through 7. All the other variables are defined as per Table 2. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Post-separation	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Period	1 year	2 years	3 years	4 years	5 years	10 years	entire spell
	Panel A: Matched Non-Financing VC						
Financing	2.021***	2.303***	2.279***	2.268***	2.276***	2.346***	2.352***
	(0.296)	(0.305)	(0.314)	(0.319)	(0.324)	(0.329)	(0.330)
R^2	0.130	0.129	0.128	0.127	0.127	0.126	0.126
	Panel B: Future-Financing VC						
Financing	2.380***	2.606***	2.491***	2.647***	2.728***	2.390***	2.402***
	(0.437)	(0.470)	(0.517)	(0.524)	(0.525)	(0.583)	(0.583)
R^2	0.152	0.149	0.149	0.148	0.148	0.148	0.148
VC×Month	Y	Y	Y	Y	Y	Y	Y
Location×Location	Y	Y	Y	Y	Y	Y	Y
Industry×Industry	Y	Y	Y	Y	Y	Y	Y

Table 2.4: Difference-in-Differences: Past vs. Contemporaneous Portfolios

This table presents results of the regression: $Y_{ijk} = \beta_1 * Financing_{ij} \times Holding_{jk} + \beta_2 * Financing_{ij} + \beta_3 * Holding_{jk} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ijk}$. k represents the past or contemporaneous portfolios. The dependent variable Y_{ijk} is a binary variable that equals 1 if the departing founder i ends up working in group k of VC j. $Holding_{jk}$ is a binary variable indicating if VC j still holds shares of group k or has exited. The other variables are defined as per Table 2. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
$Financing \times Holding$	2.564***	2.564***	2.564***	2.754***	2.755***
	(0.291)	(0.291)	(0.291)	(0.312)	(0.313)
$Financing$	1.085***	1.134***	0.169	0.070	0.012
	(0.156)	(0.157)	(0.188)	(0.208)	(0.210)
$Holding$	0.753***	0.753***	0.753***	0.715***	0.715***
	(0.150)	(0.151)	(0.152)	(0.163)	(0.163)
$Constant$	0.764***	0.738***	1.240***	1.006***	0.603***
	(0.087)	(0.087)	(0.111)	(0.184)	(0.217)
R^2	0.013	0.014	0.078	0.085	0.086
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

Table 2.5: Difference-in-Differences: Placebo Tests

This table presents results of the regression: $Y_{ijk} = \beta_1 * Financing_{ij} \times Holding_{jk} + \beta_2 * Financing_{ij} + \beta_3 * Holding_{jk} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ijk}$. In Panel A, B, and C, the actual separation time t is shifted backward by 2, 4 or 6 years, respectively. The past and contemporaneous portfolios are constructed in the same procedures as per Table 4, except benchmarking to shifted separation time. All the variables are define as per Table 2. In Panel D, the true financing VC is replaced with a non-financing VC that invests in the same location and industry. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
	Panel A: 2 Years Shift				
$Financing \times Holding$	-0.264*	-0.264*	-0.264*	-0.232	-0.232
	(0.138)	(0.138)	(0.139)	(0.150)	(0.150)
	Panel B: 4 Years Shift				
$Financing \times Holding$	0.174	0.174	0.174	0.091	0.091
	(0.138)	(0.138)	(0.139)	(0.150)	(0.150)
	Panel C: 6 Years Shift				
$Financing \times Holding$	0.223	0.223	0.223	0.146	0.146
	(0.138)	(0.138)	(0.139)	(0.150)	(0.150)
	Panel D: Placebo Financing VCs				
$Financing \times Holding$	-0.045	-0.045	-0.045	-0.006	-0.006
	(0.207)	(0.207)	(0.206)	(0.214)	(0.214)
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

Table 2.6: VC Partners' Job Transitions and Entrepreneurs' Mobility

This table presents results of the regression: $Y_{ij} = \beta_1 * FormerPartner_{ij} + \lambda_{jt} + \lambda^{loc\times loc} + \lambda^{ind\times ind} + \epsilon_{ij}$. The independent variable $FormerPartner_{ij}$ equals 1 if there is a former VC partner of departing founder i having moved to VC j before founder i's separation; $FormerPartner_{ij}$ equals 0 if VC j is the control VC, which is matched to invest in the same location and industry, to hire the same number of VC partner in the given year, and to have the same portfolio size, as the treated. Panel A uses a sample of departing founders in the old VC such that they share a common time period with the moving VC partner. Panel B examines a placebo test using a sample of departing founders such that they join the old VC's portfolio after the VC partner has left. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Baseline				
$FormerPartner$	0.976*	0.975*	0.972*	1.021*	1.177**
	(0.540)	(0.540)	(0.540)	(0.582)	(0.584)
$Constant$	1.554***	-	-	-	-
	(0.337)	-	-	-	-
R^2	0.001	0.006	0.075	0.093	0.100
	Panel B: Placebo				
$FormerPartner$	-0.403	-0.404	-0.405	-0.331	-0.268
	(0.346)	(0.346)	(0.346)	(0.389)	(0.399)
$Constant$	1.647***	-	-	-	-
	(0.262)	-	-	-	-
R^2	0.000	0.003	0.050	0.062	0.066
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

Table 2.7: ILM and VC Performance

This table examines the correlation between ILMs and VC performance. VC's fund return is measured by MIC, IRR or PME (relative to VW S&P 500). The regression takes the following form: $Return_j = \beta * ILM_j + \gamma * X_j + \lambda_v + \epsilon j$. j represents a VC fund started in vintage year v. λ_v controls for the year fixed effects. The covariates X_j include past fund return and a quadratic form of fund size. Panel A defines the ILM by an absolute measure: $ILM_j = \frac{Join_{j,j}}{Separation_j}$, whereas Panel B uses a relative measure: $ILM_j = \frac{Join_{j,j}}{Separation_j} - \frac{Join_{j,-j}}{Separation_{-j}}$. $Separation_j$ counts departing founders from VC j's portfolio and $Join_{j,j}$ computes the number of departing founders who are from j and join j for new employment. Similarly, $Separation_{-j}$ counts departing founders from portfolios other than j and $Join_{j,-j}$ tracks founders who join j from the outside. The standard errors are clustered by VC firms and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

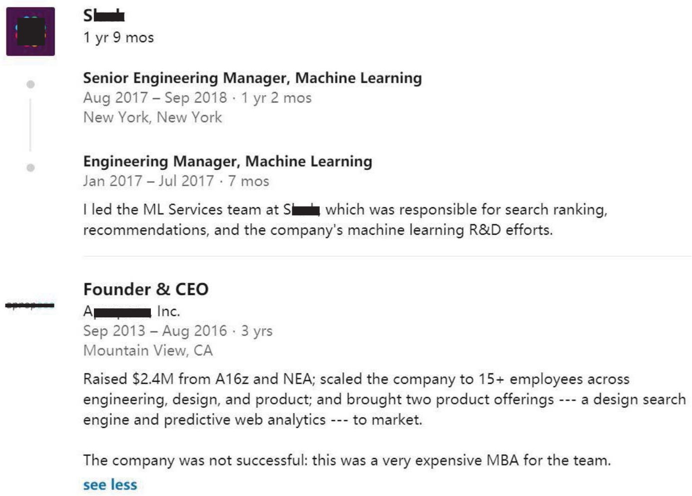
	Panel A: Absolute Measure of ILM					
	MIC		IRR		PME	
ILM	3.133***	3.266***	0.579***	0.581**	2.088***	2.083***
	(0.875)	(0.918)	(0.214)	(0.221)	(0.624)	(0.650)
R^2	0.429	0.422	0.485	0.486	0.351	0.349
	Panel B: Relative Measure of ILM					
	MIC		IRR		PME	
ILM	2.567**	2.425**	0.414*	0.405	1.517**	1.491*
	(1.065)	(1.099)	(0.214)	(0.250)	(0.759)	(0.777)
R^2	0.417	0.410	0.479	0.476	0.339	0.336
Quadratic size	N	Y	N	Y	N	Y
Past performance	N	Y	N	Y	N	Y
Year FE	Y	Y	Y	Y	Y	Y

Table 2.8: Heterogeneity of the Internal Labor Market

This table presents results of the baseline difference-in-differences estimates as per Table 4 in different subsamples. Panel A examines departing founders with below or above the median years of work experience. Panel B examines departing founders who work as CEOs or CTOs in the initial startups. Panel C examines departing founders from failed startups or from growing startups. Panel D examines lead VCs or co-investing VCs' redeployment of founders relative to the benchmark matched non-financing VCs. Panel E examines a subsample in which both the VC and the departing founder are located in California and a subsample in which either the VC or departing founder is outside California. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Founder's Work Experience				
Less Experienced	3.057***	3.057***	3.057***	3.216***	3.218***
	(0.411)	(0.411)	(0.414)	(0.442)	(0.442)
More Experienced	2.100***	2.100***	2.100***	2.294***	2.295***
	(0.412)	(0.411)	(0.409)	(0.443)	(0.443)
	Panel B: Founder's Role				
CEO	3.470***	3.470***	3.470***	3.636***	3.637***
	(0.449)	(0.448)	(0.453)	(0.484)	(0.485)
CTO	1.211**	1.211**	1.211**	1.311**	1.312**
	(0.512)	(0.511)	(0.519)	(0.558)	(0.559)
	Panel C: Separation Type				
Failed Startup	3.071***	3.071***	3.071***	3.470***	3.471***
	(0.479)	(0.479)	(0.482)	(0.530)	(0.530)
Growing Startup	2.263***	2.263***	2.263***	2.329***	2.330***
	(0.366)	(0.366)	(0.363)	(0.386)	(0.386)
	Panel D: VC's Type				
Lead VC	2.888***	2.888***	2.888***	3.144***	3.145***
	(0.275)	(0.275)	(0.273)	(0.295)	(0.295)
Co-investing VC	1.959***	1.959***	1.959***	2.143***	2.145***
	(0.310)	(0.310)	(0.309)	(0.332)	(0.332)
	Panel E: Geography				
Non-California	2.650***	2.650***	2.650***	3.107***	3.107***
	(0.395)	(0.395)	(0.394)	(0.459)	(0.458)
California	2.500***	2.500***	2.500***	2.500***	2.502***
	(0.427)	(0.427)	(0.428)	(0.428)	(0.429)
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

2.9 Appendix



S[redacted]
1 yr 9 mos

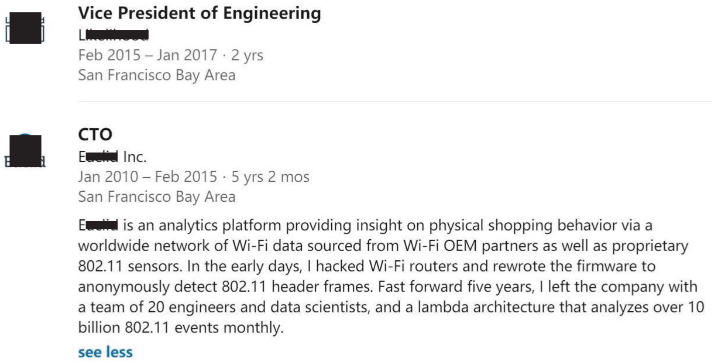
Senior Engineering Manager, Machine Learning
Aug 2017 – Sep 2018 · 1 yr 2 mos
New York, New York

Engineering Manager, Machine Learning
Jan 2017 – Jul 2017 · 7 mos
I led the ML Services team at S[redacted] which was responsible for search ranking, recommendations, and the company's machine learning R&D efforts.

Founder & CEO
A[redacted] Inc.
Sep 2013 – Aug 2016 · 3 yrs
Mountain View, CA
Raised $2.4M from A16z and NEA; scaled the company to 15+ employees across engineering, design, and product; and brought two product offerings --- a design search engine and predictive web analytics --- to market.

The company was not successful: this was a very expensive MBA for the team.
see less

(a) Failing Separation



Vice President of Engineering
L[redacted]
Feb 2015 – Jan 2017 · 2 yrs
San Francisco Bay Area

CTO
E[redacted] Inc.
Jan 2010 – Feb 2015 · 5 yrs 2 mos
San Francisco Bay Area
E[redacted] is an analytics platform providing insight on physical shopping behavior via a worldwide network of Wi-Fi data sourced from Wi-Fi OEM partners as well as proprietary 802.11 sensors. In the early days, I hacked Wi-Fi routers and rewrote the firmware to anonymously detect 802.11 header frames. Fast forward five years, I left the company with a team of 20 engineers and data scientists, and a lambda architecture that analyzes over 10 billion 802.11 events monthly.
see less

(b) Growing Separation

Figure A2.1. Examples of Different Separation Types

Figure (a) shows an example of separation that is due to startup failure. Company names are anonymized. According to Crunchbase, the initial startup A. was founded in 2013, obtained financing from Andreessen Horowitz (also known as A16z) in September 11, 2014 and was closed in 2016. The resume shows that the founder left A. in August, 2016, and subsequently joined company S., which was also in Andreessen Horowitz's portfolio at the time. S. was financed by Andreessen Horowitz in April 1, 2010. Figure (b) shows an example where the founder departed when the initial startup was viable. According to Crunchbase, the initial startup E. was founded in 2010 and had two VC fund-raising rounds in November 3, 2011 and February 21, 2013 from investors including New Enterprise Associates (also known as NEA). The CTO-founder departed in February, 2015 and joined company L., which was also backed by NEA (Series A in 2012), as a Vice President of Engineering. At the time, company E. was still growing and it had a Series C in January 14, 2016 and it was acquired by WeWork in February 7, 2019.

Table A2.1: Summary Statistics (Missing LinkedIn Resumes)

This table presents firm-level summary statistics for startups that have at least one founder who is not identified on LinkedIn. There are 2,876 missing resumes (out of 20,600 founders).

	Num. of Founders	Founded Year	Num. of Rounds	Num. of VCs	Total Funding (m$)	$\mathbb{1}_{\text{IPO}}$	$\mathbb{1}_{\text{M\&A}}$
mean	2.24	2007	1.80	1.75	103.52	0.13	0.28
s.d.	1.29	7.82	1.13	1.13	457.99	0.33	0.45
25^{th} pctl	1	2003	1	1	9.8	0	0
50^{th} pctl	2	2008	1	1	29.5	0	0
75^{th} pctl	3	2012	2	2	93	0	1

Table A2.2: Difference-in-Differences: Using Different Past Portfolios

This table presents results of the regression: $Y_{ijk} = \beta_1 * Financing_{ij} \times Holding_{jk} + \beta_2 * Financing_{ij} + \beta_3 * Holding_{jk} + \lambda_{jt} + \lambda^{loc \times loc} + \lambda^{ind \times ind} + \epsilon_{ijk}$. i represents a separation between a founder and a startup. t represents the separation time. k represents the past or contemporaneous portfolios of VC j. Past portfolios are defined using exited firms within the past 10 years (Panel A) or 5 years (Panel B). The dependent variable Y_{ijk} is a binary variable that equals 1 if the departing founder i ends up working in group k of VC j. $Financing_{ij}$ is a binary variable indicates if VC j is the financing VC or the matched non-financing VC of startup i. $Holding_{jk}$ is a binary variable indicating if VC j still holds shares of group k or has exited. The other variables are defined as per Table 2. The robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Past Portfolio of 10 Years				
$Financing \times Holding$	2.670***	2.670***	2.670***	2.871***	2.873***
	(0.287)	(0.287)	(0.287)	(0.308)	(0.308)
R^2	0.013	0.015	0.078	0.086	0.087
	Panel B: Past Portfolio of 5 Years				
$Financing \times Holding$	2.656***	2.656***	2.656***	2.852***	2.853***
	(0.249)	(0.249)	(0.248)	(0.267)	(0.267)
R^2	0.015	0.016	0.080	0.088	0.089
Month	N	Y	N	N	N
VC×Month	N	N	Y	Y	Y
Location×Location	N	N	N	Y	Y
Industry×Industry	N	N	N	N	Y

Chapter 3

Information Dispersion Across Employees and Stock Returns

Ashwini Agrawal, Isaac Hacamo, and Zhongchen Hu[1]

3.1 Introduction

Rank-and-file employees are becoming an increasingly critical factor of production for many companies (Zingales, 2000). These changes suggest that firms' workforce dynamics have important consequences for firm performance. We know little, however, about how employee entry and exit matter for firms' stock prices. Investors may ignore these dynamics if they believe that the information contained in rank-and-file labor flows is sufficiently spanned by other sources of data that are used to value securities.

To date, there has only been limited study of the implications of labor flows for asset prices. The main difficulty in addressing this issue stems from the empirical challenge of collecting granular data on employment dynamics at the firm-level. Standard datasets that are typically used to analyze workers and firms, such as Compustat and matched employer-employee administrative data, often lack

[1] We are grateful to seminar and conference participants at the London School of Economics, Georgia Tech, and the 29th Mitsui Finance Symposium for Asset Pricing at the University of Michigan, for helpful comments and feedback. We are especially thankful for the advice of Pedro Barroso, Pierre Colin Dufresne, Vicente Cunat, Alex Edmans, Daniel Ferreira, Samuel Hartzmark, Juhani Linnainmaa, Dong Lou, Ian Martin, Zhenyu Wang, and Wenyu Wang.

precise information on the timing of employee entry and exit. These limitations make it difficult to assess how the employment dynamics of rank-and-file workers matter for stock returns.

We overcome this challenge by collecting new data from LinkedIn, one of the world's largest online professional networks. We analyze the CV's of individual users of the platform, and identify the start and end dates of job spells to construct a sample of monthly labor flows at Russell 1000 firms. Using this data, we assess whether rank-and-file employees' entry and exit decisions reflect information that can be used to predict stock returns.

More concretely, we propose and test the hypothesis that rank-and-file labor flows reflect information observed by workers that is not incorporated into prices by investors. The intuition behind our hypothesis can be understood as a bridge between theories of worker job search and theories of investor behavior. We hypothesize that workers observe informative signals about the firm's future prospects, and use these signals to update their wage expectations at the firm. In response to negative (positive) signals, workers become more likely to exit (join) the firm. Net labor flows reflect the aggregation of this information across workers. If investors do not infer this information from labor flows and incorporate it into stock prices immediately, either because it takes time for information to percolate through the market (Hong and Stein, 1999) or because investors are subject to behavioral biases (Barberis et al., 1998; Daniel et al., 1998), then we will observe a link between labor flows and future stock returns.

We document a number of empirical findings that support our hypothesis. First, we use calendar-time portfolio analysis to show that labor flows can predict future abnormal stock returns. We evaluate a trading strategy in which we short (long) firms that experience high (low) net labor outflows, where net labor outflows are calculated as the difference between gross labor outflows and inflows over a given month, divided by total employment at the start of the month. The strategy yields a statistically significant abnormal return of 0.42% per month (or more intuitively, 4.98% per year). The results are robust to a variety of alternative specifications, such as strategies that use different sorting window lengths, sample period start dates, return weighting schemes, and different benchmark factor models. We also show that labor flows can explain stock returns in Fama-Macbeth regressions that control for a variety of factors associated with well-known stock return anomalies.

Second, we present evidence that equity analysts and investors do not appear to fully incorporate information from labor flows into their corporate earnings expectations. Equity analysts, for example, consistently overestimate (underestimate) the earnings of firms that experience high (low) net labor outflows. The negative correlation that we observe between net labor outflows and analyst errors is robust to numerous explanatory controls for earnings surprises and analyst biases documented by prior studies such as Hughes et al. (2008) and So (2013). Additionally, event study evidence reveals that stock prices decrease (increase) significantly in response to negative (positive) earnings surprises. These findings suggest that investors behave similarly to equity analysts, and fail to adequately formulate earnings expectations to reflect the information contained in labor flows.

Third, we show that our results are stronger for firms that are financially opaque to investors. We construct a number of well-established proxies for financial transparency, and find that the link between labor flows and stock returns is especially pronounced for firms that are less transparent to investors. For example, for newly listed firms that are likely to be harder for investors to evaluate given their limited operating histories, we find that our trading strategy yields abnormal returns of approximately 1.1% per month in the immediate three years that follow an initial public offering (IPO). These and other findings suggest that the failure to account for labor flows is especially costly in instances where it is already difficult to value the firm's assets correctly.

Fourth, we show that the link between gross labor outflows and abnormal stock returns is more pronounced than the link between gross labor inflows and abnormal stock returns. Although our hypothesis pertains to both workers who enter and exit firms, the mechanisms by which these two types of workers obtain information about the firm's future prospects are likely to differ. For example, some studies argue that employees within a company observe information about the firm's future prospects while on the job (Brown and Matsa (2016); Baghai et al. (2018); Bassamboo et al. (2015); Cowgill and Zitzewitz (2015)). Other studies claim that prospective workers gather information about a firm through its current employees (for example, Holzer (1988); Harry (1987); Ioannides and Datcher Loury (2004); Bayer et al. (2008); Cingano and Rosolia (2012); Hacamo and Kleiner (2017)). Our results support both information mechanisms that underlie our hypothesis.

Fifth, we conduct a large-sample survey of the actual LinkedIn users who comprise our sample, and show evidence that workers themselves confirm making entry and exit decisions in accordance with our hypothesis. For example, workers report that their employers' future prospects factored heavily into their past entry and exit decisions, and were generally more important to their decisions than idiosyncratic factors such as family considerations. We also show that prospective employees frequently report gathering information about a firm through its existing employees before deciding whether to join the company. These findings indicate that the workers in our analysis corroborate behavior that reflects our hypothesis.

We also present evidence of one example of the types of information that rank-and-file workers may observe about the firm's future prospects (under our hypothesis, workers may observe a variety of signals that pertain to firm performance). We argue that workers who are central to the operations of a firm observe information about the firm's future production costs. To support this claim, we first show that increases in net labor outflows are predictive of reductions in corporate earnings, primarily through increases in operating expenses and SG&A; we do not observe any significant correlations between labor flows and revenues. We then show that the net labor flows of high-skilled workers such as engineers, scientists, and middle managers—workers who are central to the firm's operations and able to directly observe the firm's production process—are highly predictive of abnormal stock returns, whereas other types of labor flows in our sample are less informative about future performance.

We consider several alternative explanations for our findings, and present theoretical and empirical arguments to characterize their relevance. For example, we assess whether the abnormal stock returns that we document are transitory phenomena that are subject to reversal over longer time horizons; if so, then labor flows may not reflect fundamental information that is materially important for stock prices. In contrast to this hypothesis, however, we show that our main results are not subject to reversal over longer-time horizons. Instead, estimates suggest that investors slowly incorporate information contained in labor flows into stock prices over time.

A second alternative explanation for the findings is that labor flows may simply reflect the hiring and firing decisions of well-informed top executives who possess inside information about the firm's future prospects (Myers and Majluf, 1984), rather than information used by rank-and-file employees to make entry and exit decisions on their own. We show, however, that top executives' insider trading patterns do not correlate with labor flows in a manner that is consistent with this explanation. Moreover, if executives possess insider information that leads to hiring and firing decisions in ways that affect stock prices (as our results demonstrate), then executives would likely be required to disclose this information to comply with fair disclosure rules such as SEC Rule 10(b)-5. The fact that analysts and investors fail to accurately forecast earnings in line with labor flows, however, suggests that top executives do not communicate (and therefore likely: possess) such information.

A third alternative explanation for the findings is that the abnormal stock returns that we observe may simply reflect employment adjustment costs caused by worker flows. To refute this hypothesis, we present three arguments. First, we construct several proxies for employment adjustment costs across firms, and show that our results are actually stronger for firms that have low adjustment costs rather than high adjustment costs. Second, we present survey evidence that individual workers in our sample do not believe that hiring costs—a first-order component of adjustment costs—are significant enough to impact stock returns. Third, we argue that if labor flows simply reflect adjustment costs that are significant enough to impact stock returns, then presumably managers should be cognizant of these costs and communicate such issues to investors. Our findings on insider trading and analyst earnings forecasts, however, are inconsistent with such behavior.

A fourth alternative explanation for our findings is that the abnormal stock returns that we document may simply reflect missing risk factors in the benchmark models for the equity cost of capital. In contrast to this explanation, however, the accounting data and survey evidence indicate that labor flows contain information about earnings levels rather than discount rates alone. Moreover, if labor flows only reflected information about discount rates, then we should not observe our predicted links between labor flows and earnings surprises, nor should we observe our predicted market reactions to earnings announcements.

Finally, we argue that our results do not stem from labor flows constituting private information that is unavailable to investors. We demonstrate that labor flows are publicly observable and fairly straightforward to analyze in real-time. For example, labor flows for individual companies can be constructed at low cost using LinkedIn's own search engine, as well as many other sources of data such as social media platforms and news databases. Consistent with this assessment, we show that our results hold even when we limit our analysis to post-2005 and post-2010 sample years—time periods that follow the public launch of LinkedIn and its rise in popularity among workers.

Our paper adds to a nascent, but rapidly growing literature that seeks to understand how the firm's labor force dynamics matter for asset prices and corporate behavior. Some studies, for example, argue that labor mobility and hiring adjustment costs impact the firm's equity cost of capital (Donangelo, 2014; Belo et al., 2017, 2014). Fedyk and Hodson (2019) present a descriptive analysis of a firm's technical and social skillsets and its equity returns. Other related papers argue that employee satisfaction surveys can be used to predict stock prices (Edmans, 2011; Green et al., 2017; Sheng, 2019).

The unique contribution of this study is evidence supporting the hypothesis that rank-and-file labor flows reflect information observed by workers that is not incorporated into stock prices by investors. The findings are important because they illustrate the costs of asset valuations that ignore the labor market activities of rank-and-file workers. The scope of this problem is potentially large given the increasing reliance of many firms on human capital as a factor of production.

The remainder of this paper is as follows. Section 3.2 outlines the conceptual framework for our analysis. Section 3.3 details the construction and sampling properties of our data. Section 3.4 presents the empirical findings. Section 3.5 concludes.

3.2 Conceptual Framework

3.2.1 Hypothesis

We propose the hypothesis that rank-and-file labor flows reflect information observed by workers that is not immediately incorporated into prices by investors. The intuition behind our hypothesis can be understood as a bridge between canonical theories of job search in labor economics with standard theories of investor behavior in finance. Additionally, the microfoundations of our hypothesis are supported by a number of empirical studies in the labor and finance literatures.

In the canonical model of job search (Mortensen, 1986; Cahuc et al., 2014), workers face various labor market frictions and incur search costs when looking for a job. While conducting their search, individuals face an exogenously specified distribution of wage offers and may receive income during unemployment spells. Workers formulate expectations over the discounted stream of wages they expect to earn from an employer over time. In equilibrium, a worker obeys the following rule: accept any wage offer that exceeds her reservation wage, where the reservation wage is an endogenously determined threshold that reflects the exogenous parameters discussed above.

One of the standard comparative statics of the canonical search model is that a worker is more likely to take up (turn down) an alternative job, in response to an exogenous reduction (increase) in the income that she expects to earn from a given employer. In our setting, workers may observe various signals about a firm's future performance that influence their wage expectations at the firm. These signals may include information about the firm's production costs, information about the CEO's productivity, or any other factors that impact the total amount of firm surplus available to workers, where firm surplus is defined as the difference between revenues and non-labor production costs.[2]

Our hypothesis applies to both workers currently employed by the firm, as well as prospective workers who may join the firm. The specific channels through which each group of workers obtains information about the firm may differ, however. For example, a number of papers argue that rank-and-file employees of firms are able to observe information about a firm's future prospects through their on-the-job activities (Brown and Matsa, 2016; Agrawal and Tambe, 2016; Baghai et al., 2018; Bassamboo et al., 2015; Cowgill and Zitzewitz, 2015). Other studies show that outside workers also gather information about a prospective firm's prospects, through peer networks of workers who are already employed by the firm (Holzer, 1988; Harry, 1987; Ioannides and Datcher Loury, 2004; Bayer et al., 2008; Cingano and Rosolia, 2012; Hacamo and Kleiner, 2017).

The net sum of entry and exit decisions reached by workers at a given firm constitutes a firm's net labor flow. Labor flows, therefore, reflect an aggregation of the information observed by workers that pertains to the firm's future prospects. This information can be used to predict stock returns, if investors fail to extract this information from labor flows and immediately incorporate it into stock prices. There are several models of investor behavior that could give rise to such an outcome. For example, investors may not fully incorporate information contained in labor flows into asset prices because they are too conservative or too overconfident in their preexisting views about firms' future earnings (Barberis et al., 1998; Daniel et al., 1998). It may also be the case that the information reflected by labor flows simply takes time to permeate through financial markets (Hong and Stein, 1999).

3.2.2 Empirical Predictions

Our hypothesis generates a number of testable empirical predictions. First, our hypothesis predicts that high (low) net labor outflows reflect negative (positive) signals about future stock returns. High net labor outflows stem from workers observing negative net signals about a firm's future prospects. If investors fail to

[2]Workers may even use these signals to make inference about future stock prices, which could further impact their wage expectations.

incorporate this information into their expectations, stock prices for high (low) net labor outflow firms will be higher (lower) than their fundamental values. Our hypothesis, therefore, predicts that higher net labor outflows will be associated with lower future abnormal stock returns, ceteris paribus.

Second, our hypothesis implies that net labor outflows will be predictive of corporate earnings surprises. As per the reasoning described above, investors will overestimate the earnings of firms that experience high net labor outflows, since investors do not infer that net labor outflows reflect negative signals about firm surplus. Similarly, our hypothesis implies that investors will underestimate the earnings of firms that experience high net labor inflows.

Third, our hypothesis has implications for how abnormal returns should vary across different types of firms. For example, we should expect to see a stronger link between labor flows and abnormal stock returns for firms that are less financially transparent to investors. If investors do not incorporate labor flows into their corporate valuations, then their ability to formulate accurate expectations for financially opaque firms—where investors already face difficulty in predicting corporate performance—will be especially poor.

Fourth, our hypothesis applies to both workers who are currently employed by a given firm, as well as prospective workers who may join a firm. The mechanisms through which information is collected by the two groups of workers are slightly different. Nevertheless, we should observe an empirical link between gross outflows and stock returns, as well as an empirical link between gross inflows and stock returns. Moreover, we should also observe that both current employees as well as outside workers report awareness of the firm's future prospects when they make entry and exit decisions.

3.2.3 Example of Information Content in Labor Flows: Production Costs

Under our hypothesis, labor flows may encompass a variety of signals that rank-and-file workers observe about the firm's future performance. The main contribution of our paper is to establish that investors do not immediately extract and incorporate this information into stock prices. The various types of information that workers observe cannot be directly measured or quantified, as workers' information sets are inherently intangible. However, we use our data to consider one potential source of information that may be reflected in labor flows which is unique to the literature.

Specifically, we conjecture that key workers in the firm's operations are able to observe information about the firm's non-labor production costs. Software engineers, for example, may observe unexpected production setbacks that cause the firm to increase operating expenses such as IT purchases or marketing costs. If workers observe signals about the firm's future production costs, then net labor outflows should correlate positively with future operating expenses, and correlate negatively with corporate earnings. Furthermore, the link between labor flows and stock returns should be especially pronounced for workers who perform tasks that are critical to the core operations of the firm, as these workers are more likely to directly observe non-labor production costs. We use our data to evaluate these predictions, and present additional findings to shed light on the types of information that is reflected in rank-and-file labor flows.

3.3 Data

To test our hypothesis, we sample data from LinkedIn, one of the largest online business networking platforms in the U.S. Our sampling procedures are designed to meet three competing constraints: computational feasibility, population representativeness, and economic relevance. In this section, we describe our data collection methods, we present descriptive statistics of our dataset, and we discuss various sampling issues that pertain to our empirical analysis. Further details are provided in the Appendix where indicated.

3.3.1 Dataset Construction

3.3.1.1 Worker-Firm Panel Dataset

We collect data on individual workers registered on LinkedIn, where users upload self-reported information from their CV's to the website. The typical information available for a user includes data on an individual's educational background and employment history (i.e. current and past employment spells). The educational background includes information on schools attended, start and end dates, degrees obtained, and educational specialties such as college major. Each employment spell record includes the job title, full name of the employer, start and end dates, detailed job description, and geographic location. Each employment spell is also linked to the employer's firm-level profile on LinkedIn; this profile contains information such as the location of the firm's headquarters, its industry, size, and number of employees on LinkedIn.

From the universe of individual worker profiles available on LinkedIn, we use a randomized sampling strategy to collect data for over 1 million employees who have worked for publicly traded companies in the U.S. Our main sample consists of Russell 1000 constituents as of 2018; we choose this sample for three reasons. First, the Russell 1000 covers more than 90% of all traded equities in the U.S., which ensures that our results generalize across a wide range of firms. Second, we identify firms in the Russell 1000 as of a recent time period, to maximize the number of employee records that we are able to observe on LinkedIn. Third, we use the Russell 1000 index to provide a sample definition that is unlikely to vary with labor flow data, thereby minimizing potential sample selection bias. To show that our results are not sensitive to the choice of firms in our sample, we also collect and analyze the labor flows of over 1 million employees of firms that are part of the Russell 1000 as of 2006, as well as labor flows for firms that undertake an initial public offering (IPO) between 1985 and 2016.[3]

We access data on LinkedIn using publicly-available search tools provided by

[3] Due to computational constraints, we are unable to collect data on all Russell 1000 constituents across all sample years, however, these alternative samples provide useful evidence of the robustness of our results to the choice of sample firms. These data also help us address additional issues in Section 3.4.

online search engines such as Google and Yahoo. We use these tools to search for the profiles of workers on LinkedIn who report any instances of working for a sample firm in their employment histories. Specifically, the inputs that are entered into the search engines are text strings that contain company names followed by a randomly generated alphanumeric character. This procedure returns a sample of individual user profile results that we collect for our analysis.

Using this sample, we create an employee-employer matched panel dataset that covers employment histories for individual workers at Russell 1000 firms between 1985 and 2016. We then use this data to aggregate individual employment spells across firms, and construct sample measures of firm employment levels and employment entry and exit every month. The net labor outflow of workers in a given firm-month is defined as the ratio of the difference between the total number of employees who exit a firm minus the total number of employees who join the firm (during the month), divided by the total stock of employment at the firm at the beginning of the month. We construct these measures for all workers in the sample, and also create these measures across various worker classifications, based on job descriptions and educational characteristics from LinkedIn profiles.

We merge our firm-labor sample to several other datasets. For example, we use CRSP and Compustat to obtain stock prices and employer-quarter measures of accounting variables such as balance sheet and income statement figures. We collect insider trading data for Russell 1000 executives from Thomson Reuters' Insider Filing database. We also gather data on equity analyst earnings forecasts from the Institutional Brokers Estimate System (I/B/E/S). Further details on our data construction are presented in the Appendix.

3.3.1.2 Survey Dataset

We conduct a large-sample survey of the individuals who appear in our worker-firm panel data, to provide additional findings that we use to test our hypothesis. We ask users questions that describe their decision-making process when they chose whether to exit or join firms in our sample. We randomly select 2,500 users from our worker-firm panel dataset, and contact each user individually using LinkedIn's e-messaging service. Each message contains a link to a survey hosted on Surveymonkey.com.

Each survey contains three questions, and the content of the questions differs for workers who comprise the inflow versus outflow samples (we survey 1,250 workers from each of these samples). The questions (detailed in the Appendix), pertain to the main hypothesis that we test, as well as alternative hypotheses that we consider in the empirical analysis. For example, we ask workers in our outflow sample about the importance of their employer's future prospects when deciding to leave their jobs.

The responses to the questions correspond to a numerical scale of 1 (Not important at all) to 5 (Very important). We are thus able to quantify the average score and the standard deviation of scores that correspond to each question. We obtained approximately 400 responses to our survey, for an overall response rate of 16%, over a period of six months.

3.3.2 Sample Descriptive Statistics

Our worker-firm panel dataset consists of 1,500,457 job records held by 1,028,356 employees across Russell 1000 firms. Table 3.1 (Panel A) presents summary statistics that describe the workers in our sample. The most frequently observed sample occupations are middle-managers, engineers, and office administrators, followed by consultants, scientists, and finance staff. In terms of education, approximately 37.7% (12.45%) of our sample reports earning a Bachelor's degree (high school degree) as their highest level of educational attainment. The average length of labor market experience for a worker in our sample is 5.63 years.

To place these statistics into the proper context and understand the population of workers that is represented by our sample, it is helpful to compare our sample with the LinkedIn population and the U.S. labor force. The population of workers on LinkedIn represents a substantial fraction of the U.S. workforce. Although exact figures are not available, there are more than 160 million past and current U.S. users on LinkedIn as of October 2019; the current U.S. labor force is approximately 164 million workers.[4]

[4]These figures are taken from: www.statista.com/statistics/272783/linkedins-membership-worldwide-by-country/; www.dlt.ri.gov/lmi/laus/us/usadj.htm.

LinkedIn contains a large absolute number of workers across a variety of occupations and industries, as illustrated in Figure 3.1. The differences in the total numbers of workers on LinkedIn compared to the U.S. labor force stem from a variety of factors. Most important is the fact that online professional networking is relatively less important for workers in certain segments of the labor force. For example, Figure 3.1 illustrates that LinkedIn represents a high fraction of workers in the U.S. labor force who are employed in the finance, information technology, and business services sectors. In contrast, LinkedIn contains smaller numbers of workers in the U.S. labor force who are employed in the manufacturing, trade, and transportation sectors. Though the absolute numbers of LinkedIn workers is still high in these sectors, online professional networking in these industries is likely to be of less importance.

LinkedIn is also more likely to represent younger, more educated workers than the U.S. labor force as a whole. In the U.S. labor force, for example, the fraction of workers whose highest level of educational attainment is a college (high school) degree is approximately 24% (26%) as per BLS statistics. Taken together, these data illustrate that the LinkedIn population represents a very large fraction of the U.S. labor force, with over- and under- sampling of various occupations and industries.

Our dataset represents a sample of workers that have been employed by Russell 1000 firms sometime between 1985 and 2016. Figure 3.2 depicts the industry distribution of workers in our sample, compared to the total population of Russell 1000 workers on LinkedIn and the total population of Russell 1000 workers on Compustat. Although each of these data sources is subject to measurement error and therefore inadequate for providing precise employment numbers for Russell 1000 firms, data comparisons across these sources provide a general sense of the data that we analyze.

Figure 3.2 illustrates that our sample contains a large cross-section of workers across many industries. The figure also shows that the distribution of workers across most industries is similar to that of the LinkedIn population and Compustat. There are differences across groups in the sampling rates of specific industries; for example, our data over-samples workers in information technology and financial services, and under-samples workers in trade, transportation, and utilities. As discussed below, we assess the importance of these sampling differences in our empirical analysis.

Table 3.1 (Panel B) describes the characteristics of companies in our sample. The average firm size is $25 billion in assets, while the average market capitalization is $12.3 billion in equity. Table 3.1 (Panel C) describes our measures of firm-level labor flows in the sample. The average number of labor outflows (inflows) over in a given month is 4.1 (5.4) outgoing (incoming) workers across firms in the sample. Standardized net labor outflows (which we refer to simply as net labor outflows for brevity) have an average value of -0.007, and a standard deviation of 0.049, across all firm-months in the sample. Intuitively, this figure implies that the average monthly change in employment observed for sample firms is roughly an increase of 0.7% over the sample period. The 5th percentile of net outflows is -0.056, while the 95% percentile of net outflows is 0.028. These figures illustrate wide heterogeneity in monthly labor flows observed across Russell 1000 firms.

3.3.3 Sampling Issues

The main strength of our data is granular information on employee entry and exit at public companies. Other commonly used datasets, such as Compustat and administrative employer-employee matched surveys, often lack precise information on the timing of labor flows. Our data enable us to test whether rank-and-file labor flows contain information that is useful for predicting abnormal stock returns at a monthly frequency. An important concern for our empirical analysis is assessing the extent to which measurement error, sample selection biases, and population representativeness impact the interpretation of our findings. We discuss each of these issues below and cite further analysis in Section 3.4.

3.3.3.1 Measurement Error

There are two potential sources of measurement error in our data. First, users may provide incorrect information about their employment histories on LinkedIn, for example, by changing the start and end dates of various positions. Second, our sample measures of labor flows may be imprecise, because we do not observe all workers in the LinkedIn population, and we are unable to observe workers in the labor force who are not on LinkedIn.

We perform several analyses that suggest that measurement error is unlikely to be a major concern for our results. First, we present evidence that users report accurate information to LinkedIn. Figure 3.3 illustrates that the lengths of employment spells implied by our dataset closely match the lengths of employment spells for workers in the U.S. labor force (based on Current Population Survey data).

Second, the veracity of LinkedIn data is supported by the fact that employers often run background checks on workers to verify employment histories and educational acheivements, and users can be identified for posting false information on LinkedIn because this data is publicly available. Third, as we discuss in our empirical analysis below, we characterize the degree of measurement error in our labor flow measures using employment data from other data sources such as Compustat, and we show that are our results are stronger for samples where measurement error is likely to be relatively low.

3.3.3.2 Sample Selection

Another important concern is the over- and under-sampling of particular workers at Russell 1000 firms. As discussed earlier, Figure 3.2 shows that our sample over-represents younger, more highly educated workers across specific occupations and industries, as compared to the general population of workers employed by Russell 1000 firms.

To address this concern, in our empirical analysis we conduct bootstrap procedures to re-create samples that more closely mirror the population of workers on LinkedIn. We also create subsamples of data that are restricted to specific types of workers, to control for worker characteristics which may be subject to uneven sampling. Finally, we examine subsamples of data where sampling rates of the population are relatively high versus relatively low. Using these different samples, we perform our main tests, and show that our findings are robust to these different sampling schemes, which suggests that our main results are unlikely to be driven by significant sample selection bias.

3.3.3.3 Population Representativeness

Our data represent a large, economically meaningful segment of the labor force employed by Russell 1000 firms. We do not observe workers who do not use LinkedIn, nor do we study all firms outside of the Russell 1000. These caveats imply that we are unable to assess whether the labor flows of all workers contain information that is useful for predicting stock returns, nor are we able to claim that our hypothesis holds for publicly traded firms outside the Russell 1000. Nevertheless, we believe the dataset that we analyze is important, because it captures the labor dynamics for a large segment of the workforce that is employed by firms that represent approximately 90% of all U.S. traded equities. As such, the data enable us to test whether there are *any* labor flows that contain information that is useful for predicting stock returns for Russell 1000 firms.

3.4 Empirical Findings

In this section, we present three sets of empirical findings. First, we present evidence that supports the main predictions of our hypothesis. Second, we provide suggestive evidence of one example of information content that is reflected in labor flows. Third, we evaluate a number of alternative explanations for our findings.

3.4.1 Labor Flows and Stock Returns

3.4.1.1 Calendar-Time Portfolio Analysis

The first prediction of our hypothesis is that labor flows contain information that can explain abnormal stock returns. To test this prediction, we evaluate a trading strategy that is based on firms' labor flows. Specifically, we measure the returns of a portfolio constructed each month that shorts firms that experience high net labor outflows, and longs firms that experience low net labor outflows, over the previous month.

The long-short portfolio returns are measured against factor models that are commonly used to estimate a firm's equity cost of capital. In our main specifications, we present results using the Fama-French five factor model (Fama and French, 2015).

$$R_{p,t} = \alpha + \beta \cdot MP_t + s \cdot SMB_t + h \cdot HML_t + p \cdot RMW_t + i \cdot CMA_t + \epsilon_t$$

where t denotes the calendar month, $R_{p,t}$ is the monthly return of our portfolio, and the monthly explanatory factors such as MP_t and SMB_t are defined as per Fama and French (2015). In our main specifications, we sort firms into quartiles based on their realized net labor outflows over a given month; the dependent variable captures the differences in returns between firms in the top and bottom quartiles. The main coefficient of interest is the intercept (i.e. the "alpha"). Intuitively, the intercept is a measure of the average monthly abnormal return generated by the portfolio.

The coefficient estimates and raw returns for our main specifications are presented in Table 3.2. In column 1, the alpha of 0.415% per month is statistically significant, and stems from raw returns of 1.813% per month for the long portfolio and 1.522% per month for the short portfolio. Intuitively, the results imply that a trading strategy based on net labor outflows generates abnormal returns of approximately 4.98% per year. Column 2 indicates that the alpha is similar in magnitude when measuring portfolio returns using a value-weighting scheme.

The results are robust to a variety of alternative specifications and sample restrictions. For example, in columns 3 and 4 of Table 3.2, we construct our portfolios by sorting firms into terciles rather than quartiles of realized net labor outflows. The significant alpha estimates in both columns indicate that our results are not sensitive to more coarse distributions of labor flows across firms. In columns 5 through 8, we restrict our sample to years when the LinkedIn platform is publicly available, and we also remove NBER-defined recessionary periods in the final two columns. The results across these columns indicate that our main findings are not driven by periods of time when labor flows are potentially harder to observe by investors using LinkedIn, nor are they driven by recessionary periods when stock returns are especially low.

Additionally, we report that the returns generated by the trading strategy stem from returns that are monotonic in labor flows across portfolios; our results are not simply driven by the returns of portfolios with extreme labor flows. As shown in Table 3.2, the alphas for the top (bottom) quartiles of net labor outflows are generally negative (positive) and statistically significant, while the alphas for the middle quartiles are statistically indistinguishable from zero.

We further illustrate our results' robustness by presenting regression estimates across a variety of alternative specifications in the Appendix. For example, the findings presented in Appendix Table A3.1 show that our trading strategy generates consistent abnormal returns irrespective of whether we alter the length of the sorting windows used to calculate labor flows (from one month to six months), the percentile ranking schemes used to allocate firms across the long and short portfolios (terciles to quintiles), or the sample start years (1985, 1995, 2005, and 2010).

We also show that our results do not appear to be adversely impacted by measurement error in our labor flow measures. Appendix Table A3.2 illustrates that the link between labor flows and stock returns is, in fact, stronger when we examine subsamples of data where measurement error in labor flows is likely to be low. In particular, when we restrict our sample of analysis to firms in which we observe relatively high fractions of the total worker population in our sample—as measured by either Compustat or LinkedIn—we observe greater return predictability of labor flows. Therefore, because our results grow stronger as measurement error decreases, it is likely that our full sample results understate the ability of labor flows to predict abnormal stock returns.

In Appendix Table A3.3, we show that our results are robust to sample selection concerns that pertain to the workers in our data. The sampling strategy that we develop to collect our dataset is designed to generate a random sample of workers across firms. However, because our sample is ultimately comprised of return results from internet search engines, it is possible that the sample of workers that we collect is non-random across dimensions such as unobservable worker quality. To explore the relevance of this concern, we perform bootstrap analyses to generate samples of data where the distribution of workers who attend highly ranked universities (a proxy for unobservable worker quality) in our samples match those of the population. We show that even for these samples, our results hold. These results suggest that our estimates do not appear to be significantly biased by sample selection issues at the worker-level.

Appendix Table A3.4 shows that our results hold using alternative benchmark models such as 6-factor specifications that control for momentum, liquidity, and investment behavior (Panels A and B). Panel C illustrates that our findings are also robust to alternative factor measurement methods (Hou et al., 2015). Appendix Table A3.5 shows that our results do not appear specific to the choice of firms that we analyze. For example, our results remain similar if we analyze firms that comprise the Russell 1000 as of 2006; these results suggest that our main sample results are not subject to survivorship bias or any other unique features of firms that are in the Russell 1000 as of 2018. We also show that our results persist even if we exclude firms that IPO between 1985 and 2016 from our analysis. Because our results hold for liquid stocks with low transaction costs, it is unlikely that our full sample findings are significantly impacted by trading costs.

Firm-level Fama-Macbeth regressions of stock returns provide further support for our hypothesis. The specifications presented in Appendix Table A3.6 control for factors associated with well-established anomalies, such as firm size, investment, book-to-market ratio, and operating profitability. We also control for financial distress risk (Campbell et al., 2008), using five different measures of distress studied in the literature. First, we include a traditional measure of financial distress–the Altman Z-score (Altman, 1968). Second, we use the O-score as an alternative to the Z-score for predicting financial distress (Ohlson, 1980). Third, following Bharath and Shumway (2008), we use measures of "naïve'' Merton distances to default. Fourth, we control for whether firms file for Chapter 7 (bankruptcy) or Chapter 11 (liquidation). Lastly, we include an indicator for whether a firm's credit rating falls below investment grade (S&P BBB).

The regression estimates reported in Appendix Table A3.6 generally show that net labor outflows can explain stock returns even after controlling for various firm characteristics such as financial distress indicators. The estimated coefficient on net labor outflows is stable across specifications and statistically significant. These findings indicate that our results are unlikely to simply reflect anomalies associated with factors previously studied in the literature. Instead, the estimates provide further support for our hypothesis.

Overall, these findings illustrate that rank-and-file labor flows can be used to predict abnormal stock returns. Granular measures of employee entry and exit have significant explanatory power that is robust to a variety of specification choices. The findings support the hypothesis that rank-and-file labor flows contain information that investors do not fully incorporate into stock prices.

3.4.1.2 Labor Flows and Earnings Expectations

Our hypothesis posits that labor flows explain abnormal stock returns partly because investors do not fully incorporate information from labor flows into corporate earnings expectations. Therefore, a second empirical prediction of our hypothesis is that labor flows can predict investors' earnings forecast errors. We test this prediction in two ways.

First, we examine the earnings expectations of equity analysts. Equity analysts are a useful proxy for well-informed investors, as they are incentivized to formulate accurate forecasts of corporate earnings. Our hypothesis suggests that labor outflows can predict analyst forecast errors: in particular, net labor outflows should correlate negatively with future earnings surprises.

To measure earnings surprises, we compute the difference between analysts' earnings per share (EPS) forecasts with the realized earnings per share announced by firms in our sample. Specifically, for a given firm i in month t, we calculate the mean $\mu_{i,t}$ and standard deviation $\sigma_{i,t}$ of analysts' earnings per share forecasts for the firm's next upcoming quarterly earnings announcement. The standardized unexpected earnings (SUE) for firm i in month t is defined as:

$$SUE_{i,t} = \frac{actual_{i,t}^{ex\text{-}post} - \mu_{i,t}}{\sigma_{i,t}}$$

Intuitively, the SUE is the difference between the actual EPS realized by the firm minus the mean forecasted EPS across all equity analysts in a given month, normalized by the standard deviation of the EPS forecasts observed for that month.

We use this measure to estimate firm-month panel regressions of earnings surprises on realized net labor flows, controlling for a variety of known predictors of earnings forecast errors, following So (2013).[5] The key measure of interest is the estimated coefficient for net labor outflows. We test whether the estimated coefficient is negative and statistically significant, as our hypothesis predicts.

The results are presented in Table 3.3. In column 1, the coefficient for net labor outflows is negative and statistically significant, consistent with our hypothesis: increases in net labor outflows lead to more negative earnings surprises. In columns 2 through 4, we add controls such as firm and month fixed effects and other known predictors of unexpected earnings, and the coefficient for net labor outflows remains similar in magnitude and statistical significance. The results also remain the same when we restrict the sample years to 2005 and afterwards, suggesting that our main results are not driven by the potential difficulty of observing labor flow data from LinkedIn.

The findings are consistent with our hypothesis. The data indicate that differences between analyst earnings forecasts and realized corporate earnings can be partly explained by net labor outflows. Higher net labor outflows reflect worsening earnings prospects, yet equity analysts do not appear to factor this information into their forecasts prior to earnings announcements.

To buttress this evidence, in our second analysis, we estimate market reactions to earnings announcements in our sample. According to our hypothesis, if investors fail to incorporate information from labor flows into earnings expectations, then we should see a negative (positive) stock price reaction to negative (positive) earnings surprises. Such evidence would illustrate that investors behave similarly to equity analysts when forecasting corporate earnings and factoring their expectations into stock prices.

Figure 3.4 depicts event study analysis of earnings announcements for firms in our sample. We define the event window to be 10 days around earnings an-

[5]See also Hughes et al. (2008) for work that documents predictable components of earnings forecast errors and analyst biases.

nouncements, and we estimate factor loadings over daily returns for up to 100 days, starting 50 days before the start of the event window. We graphically depict average cumulative abnormal returns (CAR) and corresponding 95% confidence intervals generated each day of the event window.

The results in Figure 3.4 indicate that negative (positive) earnings surprises generate negative (positive) and significant cumulative abnormal returns in the immediate days surrounding earnings announcements. We find that the size and significance of these CAR's remains the same when we vary specification parameters such as the lengths of the event and estimation windows. These results suggest that the market fails to anticipate realized earnings in a manner that mirrors equity analysts.

To further reinforce this conclusion, we use the estimates from Table 3.3 to decompose equity analysts' earnings forecast errors into a component predicted by net labor outflows and a component explained by other factors (i.e. the residual). Appendix Table A3.7 shows that market reactions to earnings announcements are partly explained by labor flow-driven components of earnings surprises. These results further corroborate the view that investors behave similarly to equity analysts and fail to incorporate information from labor flows into their earnings expectations.

Collectively, these results indicate that investors do not appear to incorporate information from labor flows into earnings expectations. Consistent with our hypothesis, we find that higher net labor outflows lead to more negative earnings forecast errors and lower abnormal stock returns. The patterns are robust to a variety of different empirical specifications, and highlight the value of earnings-related information that can be extracted from rank-and-file workers' entry and exit decisions.

3.4.1.3 Heterogeneity of Findings Across Firms

We demonstrate the heterogeneity of our results across different types of firms in our sample. As discussed in Section 3.2, our hypothesis predicts that the link between net labor outflows and abnormal stock returns should be stronger for firms that are less financially transparent to investors. We test this prediction by examining our main results across firms that differ across measures of financial transparency that have been established by the prior literature.

For example, numerous accounting studies, such as Brown and Martinsson (2018), proxy for financial transparency by using well-cited measures of earnings reporting quality from Leuz et al. (2003). Like others, we assume that firms that engage in greater earnings management are less likely to be transparent to investors, since these firms exhibit greater discretion in reporting their accounting data to the public. The first measure of earnings management that we examine is the ratio of the firm's absolute value of accruals scaled by the absolute value of the firm's cash flow from operations. This ratio provides a time-varying measure of a firm's financial transparency, and is used in widely cited accounting studies such as Cohen et al. (2008).

We split our sample into firms with high vs. low levels of firm transparency, and we repeat our portfolio analysis and Fama-Macbeth regressions for firms in each of these subsamples. In Panel A of Table 3.4, the alphas for firms with low transparency are almost uniformally positive, statistically significant, and of larger economic magnitude than the respective alphas for firms of high transparency. Panel B of Table 3.4 uses Fama-MacBeth regressions to provide a formal test of the differences in returns explained by labor flows. Panel B shows that the average direct effect of net labor flows on stock returns is negative, and that the interaction term between financial transparency and labor flows is also negative and statistically significant. These findings illustrate that not only do net labor flows have a negative average effect on stock returns, but that the negative effect is especially pronounced for firms with low levels of financial transparency.

In the Appendix, we present additional results using alternative measures of financial transparency. For example, we estimate earnings management across sample firms by computing the ratio of the firm's standard deviation of operating earnings divided by the standard deviation of the firm's cash flow from operations. Low values of this measure indicate that managers exercise greater discretion to smooth reported earnings. The third measure of earnings management that we examine is the correlation between changes in accounting accruals and changes in operating cash flows for a given firm. The fourth measure is the ratio of small profits to small losses, using after-tax earnings scaled by total assets. Increases in both the third and fourth measures imply greater earnings management and hence lower transparency. The fifth measure that we employ is based on the firm's age. We analyze firms that IPO during the sample period, and perform our portfolio analysis over different periods of time after an IPO. We posit that recent IPO firms are likely to be less transparent to investors than older companies with longer operating histories, as investors gather more information about the firm over time.

The results in Appendix Table A3.8 show that the alphas generated by samples of firms with low levels of transparency are generally greater in magnitude than the alphas that correspond to highly transparent firms. Appendix Table A3.9 shows that the alphas are particularly large immediately following the IPO date of newly listed firms—on the order of 1.1% per month—and that the abnormal returns slowly converge over time to the alphas of established firms. These results are consistent with the hypothesis that labor flows are likely to be especially informative for firms with low levels of financial transparency.

3.4.1.4 Net Labor Flows and Gross Labor Flows

As explained in Section 3.2, our hypothesis is relevant for both the firm's existing employees as well as the firm's prospective workers, though the specific information transmission mechanisms are slightly different between the two types of workers. Existing employees may obtain information about firm prospects through their job activities, while outside workers may gather this information through their peer networks within firms. Net labor flows combine the information contained in gross labor outflows and gross labor inflows, thus, our main results are depicted using net labor flows because they are more informative than gross flows alone.

Nevertheless, to understand the relative empirical importance of gross labor outflows versus gross labor inflows, we perform our portfolio analysis using each of these gross labor flows in isolation. For example, we repeat our portfolio construction and return analysis as per Section 3.4.1.1, but sort firms into quartiles based on their gross labor outflows rather than their net labor outflows. We also repeat these procedures using gross labor inflows. Our hypothesis predicts that we should observe positive abnormal returns when we long (short) firms with low (high) gross outflows and long (short) firms with high (low) gross inflows.

The results of these two sets of analyses are presented in Table 3.5. Panel A illustrates that our trading strategy generally leads to positive abnormal returns when we sort firms based on gross labor outflows. The results are statistically significant and similar in magnitude across our equal-weighted portfolio return specifications (the results are positive, but less significant in our value-weighted schemes, possibly because gross flows contain less information than net flows for large-cap companies).

Panel B illustrates that gross inflows appear to have some explanatory power for abnormal returns, though the empirical link is relatively weaker than the documented effects of gross outflows. The abnormal returns for the gross inflow strategies are only statistically significant in three out of eight specifications. Moreover, the abnormal returns for six out of eight gross inflow specifications in Panel B are smaller in magnitude than the returns for the corresponding gross outflow specifications in Panel A.

The findings indicate that gross labor flows can explain abnormal stock returns, consistent with our hypothesis. Moreover, the data show that gross labor outflows are more informative than gross labor inflows. This evidence suggests that the information observed by the firm's existing employees is likely to be more precise than then information gathered by prospective workers outside the firm.

3.4.1.5 Survey Evidence

In addition to the statistical findings presented above, we also present survey evidence that supports the hypothesis that rank-and-file labor flows contain information that investors do not incorporate into stock prices. As described in Section 3.3, we ask a random sample of the actual individuals in our dataset several questions that pertain to their past labor market decisions. The answers to these questions corroborate the mechanisms that underlie our hypothesis.

For employees whose exit decisions comprise the gross labor outflows in our sample, we asked on a scale of 1 (Not important at all) to 5 (Very important): "How important were the future prospects of your employer when deciding whether to leave and find a new job?" Of the 169 responses received, Figure 3.5 shows that the average score for this question was 4.39 (with a standard deviation of 0.74). To provide a benchmark against which this score can be compared, we also asked these workers: "How important were personal circumstances when choosing whether to leave your employer"? The average score for this question was 3.62 (with a standard deviation of 1.31). The difference in average scores between the two questions is statistically significant at the 5% level.

For workers whose entry decisions comprise the gross labor inflows in our sample, we asked on a scale of 1 (Not important at all) to 5 (Very important): "Did you gather information from existing (or former) employees before deciding whether to join a prospective employer"? Of the 230 responses received, the average score for this question was 3.92 (with a standard deviation of 1.19). For comparability, we also asked them: "How important was publicly available information in deciding whether to join a prospective employer"? The average score for this question was 3.76 (with a standard deviation of 1.10). There are no statistically significant differences in the average scores between the two questions.

The survey answers demonstrate several points that are consistent with the underlying premises of our hypothesis. First, the evidence illustrates that existing employees use information about their employer's future prospects when making exit decisions. This information is relatively important, as it is highly valued relative to idiosyncratic, personal factors that also drive employees' exit decisions.

Second, the evidence also shows that prospective workers make entry decisions based on information about an employer's future prospects. The survey answers indicate that many workers obtain information about a firm's future prospects through their network of contacts that are already employed by the firm, consistent with existing literature. Moreover, this information appears to be as important to workers as the information collected from publicly available sources about a firm's future prospects–information that investors presumably use when valuing stocks. The survey evidence thus supports the notion that labor flows reflect information that rank-and-file employees have about a firm's future prospects.

3.4.2 Example of Information Content in Labor Flows: Production Costs

The findings presented thus far support the main contribution our paper: we show that rank-and-file labor flows contain information that investors do not incorporate into stock prices. Under our hypothesis, labor flows may encompass a variety of signals that workers observe about the firm's future performance. Although these signals cannot be directly measured, as workers' information sets are inherently unobservable, we provide suggestive evidence of a unique information channel that is reflected in labor flows: we argue that workers who are central to the core operations of the firm possess valuable information about the firm's future production costs.

3.4.2.1 Labor Flows and Production Costs

We demonstrate that labor flows can predict future production costs in our sample by examining firm quarterly operating earnings and its components, such as SG&A (sales, general and administration expenses), operating costs, and revenues. Table 3.6 (Panel A) reports the coefficients on net labor outflows in the following regression specification:

$$y_{i,t+1} = a + b * NetLaborOutflow_{i,t} + FEs + \epsilon_{i,t+1}$$

where t denotes the fiscal quarter. The dependent variable $y_{i,t+1}$ is measured as either SG&A, operating costs, revenues, or earnings after depreciation and amortization (EBIT) in the next quarter. All dependent variables are normalized by the book value of the firm's assets. The key independent variable $NetLaborOutflow_{i,t}$ is the net labor outflow of the current quarter. We also include firm and year-quarter fixed effects.

Table 3.6 illustrates that higher net labor outflows predict higher future operating expenses in our sample. The results are driven by the positive correlation between net labor outflows and SG&A. There is little correlation between labor flows and revenues. The sum of these effects is the observed negative correlation between labor flows and future earnings depicted in the final columns of Panel A. As higher net labor outflows lead to lower firm wage bills without offsetting effects on revenues, the results in Table 6 imply that the link between net labor outflows and earnings is driven by increased non-labor production costs, part of which are reflected in SG&A and operating expenses.

3.4.2.2 Labor Flows of Employees Central to the Firm's Production

We also show that the labor flows of specific types of workers are particularly informative about future stock returns. In particular, we examine employees who are directly involved in the firm's day-to-day operations; these workers are likely to observe shocks to the firm's production capabilities. For example, engineers often witness production setbacks that require the firm to incur additional future expenditures. To the extent that these increased expenditures leave less revenue surplus to be distributed among employees—consistent with the results depicted in Panel A of Table 3.6—there is an increased likelihood that workers will exit the firm, ceteris paribus.

To examine this issue empirically, we exploit data from workers' job titles, educational backgrounds, and career paths, to examine the links between specific types of labor flows and stock returns in our sample. For example, we identify the following major occupational categories in our data: engineers, scientists, middle managers, finance staff, office administrators, and consultants. We evaluate our trading strategy using the net labor outflows of workers that belong to each of these occupational categories. We perform similar analyses using the labor flows of workers distinguished by their educational attainment levels and years of work experience, respectively.

The results of these analyses are presented in Panels B through D of Table 3.6. In Panel B, we observe positive abnormal stock returns for our trading strategy when we sort firms based on the net labor flows of engineers, scientists, and middle managers. In contrast, we observe statistically insignificant abnormal stock returns for our trading strategy when we sort on the flows of finance personnel, office administrators, and consultants. Within each class of occupations, the results are generally similar across columns and therefore robust to a variety of alternative specifications. Panels C and D show that we observe positive abnormal stock returns when we sort firms based on the flows of workers with high levels of work experience and workers with relatively higher levels of educational attainment; these workers likely possess human capital that is critical to the operations of the firm.

Taken together, the results in Table 3.6 support the view that rank-and-file labor flows partly reflect information that pertains to the firm's productive capabilities. The correlations between labor flows and various accounting figures illustrate that higher net labor outflows are predictive of increased production costs. Furthermore, the links between labor flows and stock returns are especially pronounced among workers who are central to the operations of the firm and most able to directly observe these costs during the production process.

3.4.3 Alternative Explanations

We consider a number of alternative explanations for our main findings. In this section, we detail each of these alternative explanations, and then present theoretical and empirical arguments to characterize their relevance.

3.4.3.1 Return Reversal or Return Persistence?

We assess whether the abnormal stock returns that we document are subject to reversal over longer time horizons or whether they persist over time. If the returns reverse over longer horizons, then labor flows may not contain fundamental information that is materially important for stock prices. Instead, labor flows may simply correlate with transitory phenomena that temporarily influence prices in the short-run.

To evaluate this possibility, we repeat our sorting procedure, and estimate the long-short portfolio's returns over the subsequent months that follow the initial one-month period that generates the main results presented in Table 3.2. We test whether the trading strategy yields negative abnormal stock returns during these subsequent months. Table 3.7 presents the results of this analysis for each specification depicted in Table 3.2.

Each row in Table 3.7 corresponds to the specific month over which returns are calculated; the alphas in the first row reflect the main results presented in Table 3.2. As illustrated in the second and third rows, the alpha generally remains positive when we examine the long-short portfolio's returns during the second and third months following the initial one-month return period. The remaining rows, however, show that any subsequent abnormal returns are no longer statistically different from zero. These patterns are similar across all specifications (columns) of Table 3.7.

The data show little evidence of negative abnormal stock returns in any future periods. Instead, the data reveal gradually decreasing, positive abnormal returns for up to three months, followed by statistically insignificant abnormal returns. This evidence indicates that our main results are not subject to reversal over longer-time horizons. In fact, the findings suggest that investors slowly incorporate information contained in labor flows into stock prices over time.

3.4.3.2 Top Executive Inside Information

Another alternative explanation for the findings is that labor flows simply reflect the hiring and firing decisions of well-informed top executives who possess inside information about the firm's future prospects (Myers and Majluf, 1984). This explanation differs slightly from our hypothesis, in that labor flows reflect information possessed by top executives rather than information possessed by rank-and-file employees per se.

We present two arguments that suggest that the observed links between labor flows and stock returns do not simply reflect inside information possessed by top executives. First, we show that top executives' insider trades do not correlate with labor flows. If top-level executives make operating decisions such as hiring and firing that reflect inside information, then presumably top-level executives should also trade their holdings to capitalize on this information. Thus, we should expect to see a correlation between labor flows and insider trades.

To test this prediction, we aggregate executive insiders' monthly net sales (open market sales minus open market purchases), normalized by the total number of outstanding shares, and examine whether insider sales correlate with net labor flows. As Panel A of Table 3.8 illustrates, we find little evidence of any statistical link between labor flows and insider trades. Panel B shows that these results persist even when the analysis is limited to insider trades that are considered "opportunistic" (Cohen et al., 2012). Furthermore, in Appendix Table A3.10, we find that insider trades are poor predictors of whether firms are in the high vs. low net labor outflow quartiles studied in our portfolio analysis.

Second, we note that top executives are generally required by fair disclosure rules such as SEC Rule 10(b)-5 to disclose material information that is relevant to investors. If executives receive information that causes them to make hiring and firing decisions in ways that impact stock prices (as our results already demonstrate), then executives would be compelled to disclose this information to investors. The fact that equity analysts systematically fail to forecast earnings accurately in line with labor flows suggests that top executives do not communicate (and therefore likely: possess) such information.

3.4.3.3 Do Labor Flows Measure Adjustment Costs?

Another alternative hypothesis is that the abnormal stock returns that we document may simply reflect employment adjustment costs caused by worker flows. For example, employee departures can cause firms to incur costs of worker replacement such as hiring and training expenses; these expenditures may lead to lower future stock returns. To address this "adjustment cost" hypothesis, we present three arguments.

First, we construct several proxies for employment adjustment costs across firms. Using these measures, we show that our results are slightly stronger for firms that actually have low adjustment costs rather than high adjustment costs. We proxy for employment adjustment costs by measuring the labor market tightness faced by firms in the sample. Firms in tighter labor markets will likely incur greater worker replacement costs, since it is relatively harder to replace workers in a labor market that is tight rather than slack.

To construct this proxy, we compute each state's share of the total unemployed labor force in the U.S.; we assume that firms in states with above-median shares of the unemployed labor force will have different (and most likely: lower) adjustment costs than firms in states with below-median shares of the unemployed labor force. Our assumption is based on the idea that firms in states with relatively high shares of the unemployed labor force will likely face lower labor search costs, and thus lower adjustment costs, when replacing workers, since these firms will more easily be able to find available workers with similar skills within local geographic proximity.

As Table 3.9 indicates, when we restrict our sample to firms in either group of states, we find results for each group that mirror our full sample findings. However, when comparing the estimates between the two groups, we see that the results are slightly stronger for firms in states with higher unemployed labor force shares (i.e. lower adjustment costs). This evidence suggests that employment adjustment costs are unlikely to account for our findings, since reasonable proxies for employment adjustment costs have little explanatory power for the labor flow-stock return patterns that we document.

Second, in our LinkedIn survey of individuals who comprise our dataset, we ask questions that pertain to the adjustment cost hypothesis, and present evidence that individual workers in our sample do not believe that hiring adjustment costs are a major determinant of stock returns. When we ask workers on a scale of 1 (Not important at all) to 5 (Very important), whether they perceived hiring costs to be significant enough to impact stock prices for their specific occupations, they reported an average score of 2.06. This low average score is identical for both subsamples of inflows and outflows, and suggests that workers do not consider hiring costs—a first-order component of adjustment costs—to be significant enough to impact stock returns.

Third, we argue that other reasonable implications of the adjustment cost hypothesis do not appear to be supported by the data. For example, if the adjustment cost hypothesis truly explains the observed connection between labor flows and firm performance, then presumably managers should be cognizant of these adjustment costs. Moreover, managers would be legally obligated by fair disclosure rules to communicate these costs to equity analysts and shareholders if these costs are materially important to investors.

Our findings on insider trading, analyst expectations, and market reactions to earnings announcements, however, fail to support these implications. The lack of any significant insider trading patterns suggests that managers do not act upon adjustment costs triggered by departing workers. The evidence we find on equity analyst expectations and market surprises to earnings announcements also suggests that neither analysts nor investors appear to incorporate adjustment costs into their earnings forecasts.

Collectively, these findings suggest that the adjustment cost hypothesis is unlikely to explain the evidence that we document. The statistical findings we show for firms facing labor market tightness are inconsistent with empirical implications of the adjustment cost hypothesis, the survey evidence shows that workers do not view hiring adjustment costs as significant, and many of our other findings are inconsistent with reasonable implications of the adjustment cost hypothesis.

3.4.3.4 Discount Rates or Cash Flows?

One alternative hypothesis for our findings is that the abnormal stock returns that we document may simply reflect missing risk factors in the benchmark models for the equity cost of capital. For example, Belo et al. (2017, 2014) and Donangelo (2014) argue that labor adjustment costs and labor mobility impact the firm's cost of capital in ways that are not captured by commonly-used factor models. These "missing" factors might fully explain the abnormal returns that we associate with labor flows.

We believe that this hypothesis is unlikely to be the sole explanation for our findings. Both the accounting data and the survey evidence indicate that labor flows contain information about earnings levels rather than discount rates alone. Moreover, if labor flows only reflected information about discount rates, then we should not observe any link between labor flows and analyst earnings forecast errors, nor should we observe our predicted market reactions to earnings announcements. The findings, however, show that labor flows partly reflect information about the *level* of firms' earnings, rather than the *riskiness* of firms' earnings alone. Thus, the link that we document between labor flows and stock returns is unlikely to simply represent cost of capital differences across firms.

3.4.3.5 Are Labor Flows Publicly Observable?

Another explanation for our findings is that labor flows may not be publicly observable to investors in real-time; instead, it may take time for investors to observe and trade upon information otherwise contained in labor flows. Our documented links between labor flows and abnormal returns may thus essentially reflect private information that investors are unable to utilize.

While it is inherently difficult to measure the real-time availability of historical data, we argue that this alternative hypothesis is unlikely to hold true. First, workers generally update their CV's on LinkedIn and social media site quite rapidly upon a change of jobs (Ryan, 2016; Shuey, 2017; Lombard, 2016). We corroborate this claim by conducting a separate survey of almost 30 LinkedIn users in our dataset, in which we ask users how frequently they update their LinkedIn profiles after a job change. Over 76% of users report updating their LinkedIn profile within one month of starting a new job. In fact, over 50% of all respondents voluntarily add that they update their online CVs immediately after switching positions.

Second, labor flows for individual companies can be constructed in real-time at low cost using LinkedIn's own search engine, as the platform allows users to search for the total number of current or past employees at a firm using relatively few steps. There are also many other publicly available sources of data, aside from LinkedIn, which can be used to construct labor flows across firms. For example, patent data that is publicly available from the U.S. Patent Office contains information on key scientists and their company affiliations. Other sources of information, such as social media platforms, news databases, etc., can be used to collect information on key workers who enter and exit specific firms. Given the returns to collecting this data and the resources available to investors such as hedge funds and mutual funds, we believe that investors can collect this information and incorporate it into their trading strategies at low cost.

To further support our argument, we show in Table 3.2 that our results hold even when we limit our sample to labor flows that take place after 2005 (when LinkedIn is available for public searches, as verified by the internet archive Wayback Machine (web.archive.org)). We also show in Appendix Table A3.1 that our results hold if we analyze our sample using data from 2010 onwards, when significantly more users joined the platform and richer labor data becomes available. These findings indicate that our results hold in time periods when labor flow data is relatively easier to collect by investors; our findings are not driven by sample periods where it may have been harder for investors to observe worker flows.

3.5 Conclusion

This paper adds to a nascent, but rapidly growing literature that seeks to understand how the firm's labor force characteristics matter for asset prices and corporate behavior. The unique contribution of this paper is evidence that the firm's rank-and-file labor dynamics reflect information that can be used to explain stock returns. The findings in our paper suggest that workers observe information that investors fail to extract from employees' labor market decisions.

A natural next step for research is to shed light on other aspects of corporate behavior that are impacted by the firm's labor dynamics. Casual observation suggests that corporate investment and financing decisions are intimately related to the entry and exit decisions of rank-and-file employees. Formal study of these issues is lacking, however, and there are a number of poorly understood concerns that arise when evaluating the relationship between labor flows and firm behavior. For example, the hiring rates of specific workers likely impact the timing and choice of investment projects, while exit rates of key personnel likely impact security issuance decisions. The findings in our paper suggest that these issues are fruitful areas for further inquiry.

3.6 Figures

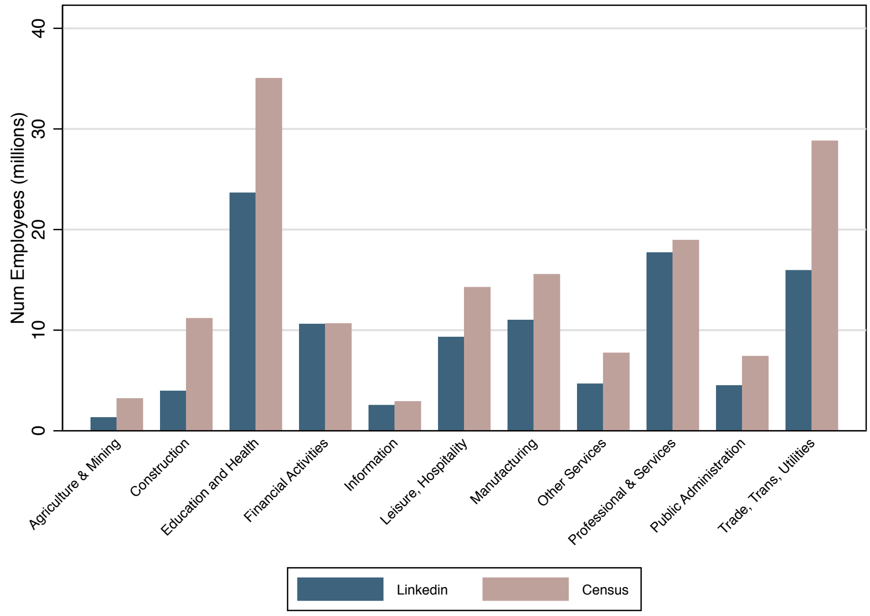


Figure 3.1. Industry Distribution of Workers in LinkedIn and the U.S. Labor Force

This figure depicts the distribution of employment across industries for workers in the LinkedIn population and workers in the U.S. labor force as of 2018. The horizontal axis corresponds to industries defined by two-digit NAICS codes, and the vertical axis corresponds to employment figures reported in millions. Labor force employment estimates are based on Census data maintained by the U.S. Bureau of Labor Statistics (BLS).

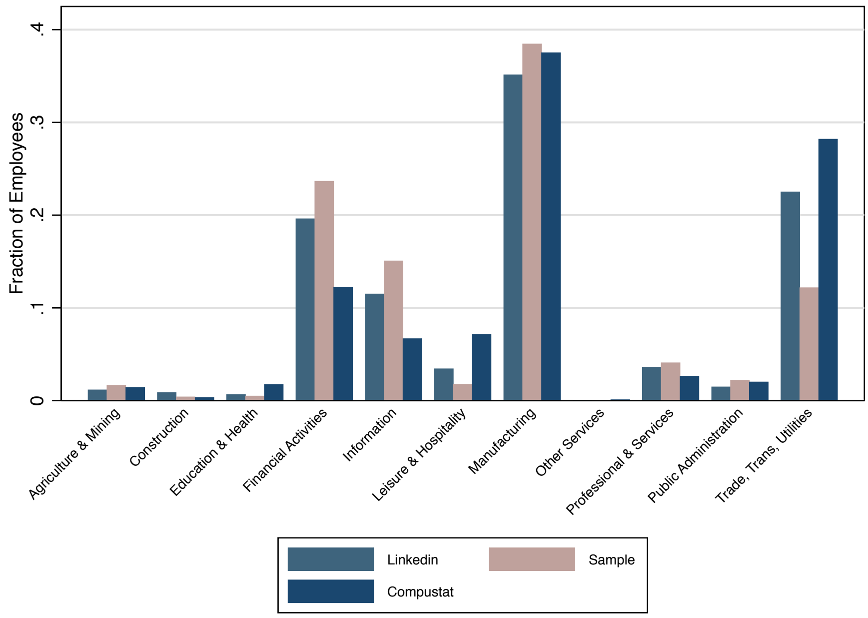


Figure 3.2. Industry Distribution of Workers at Russell 1000 Firms

This figure depicts the distribution of employment across industries for workers employed by *Russell 1000* firms, using three sources of data: the LinkedIn population, our LinkedIn sample, and Compustat. For each firm in the Russell 1000, we estimate the total size of the firm's workforce, and then assign all employees at the firm to the two-digit primary NAICS code of the firm as measured in Compustat. The horizontal axis corresponds to two-digit NAICS industries, and the vertical axis corresponds to the fraction of employees across industries within each data source.

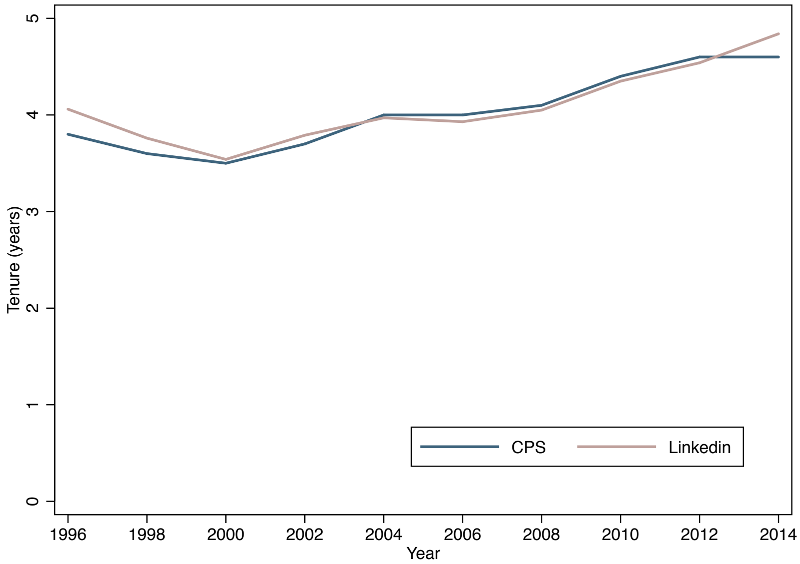
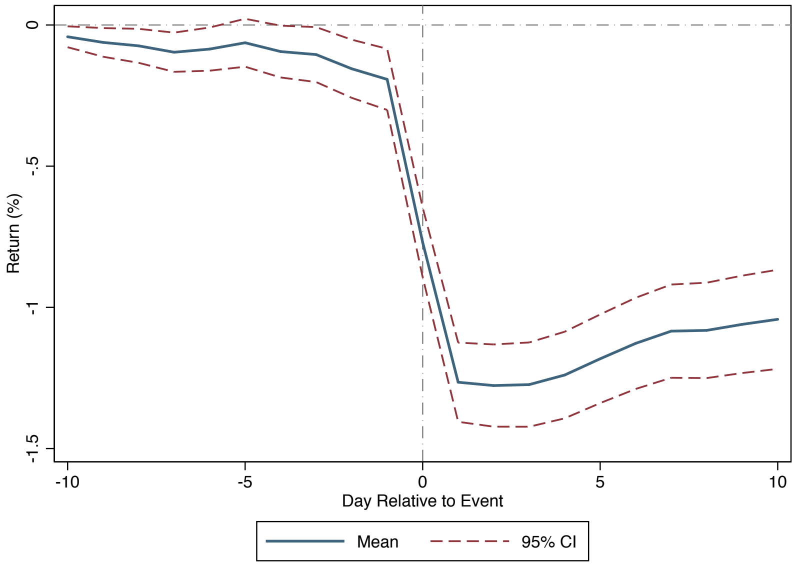
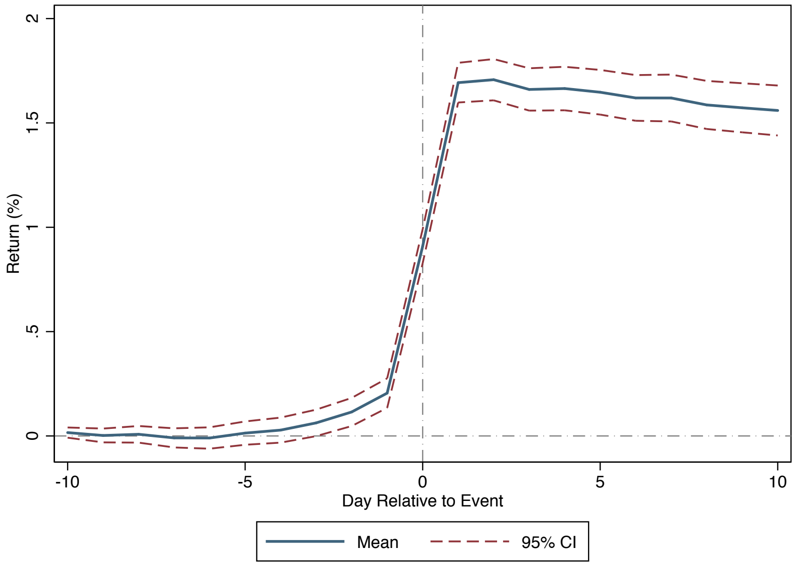


Figure 3.3. Job Tenures of Workers in LinkedIn and the U.S. Labor Force

This figure reports the lengths of job tenures for workers in our LinkedIn sample and workers in the U.S. labor force. Job tenures for workers in LinkedIn are measured using the start and end dates of employment spells listed on worker CV's. Job tenures for workers in the labor force are measured using the U.S. Current Population Survey (CPS) Job Tenure Supplement for respondents aged 15 years and older. The horizontal axis corresponds to the year of observed employment spell, and the vertical axis corresponds to the length of job tenure reported in years.



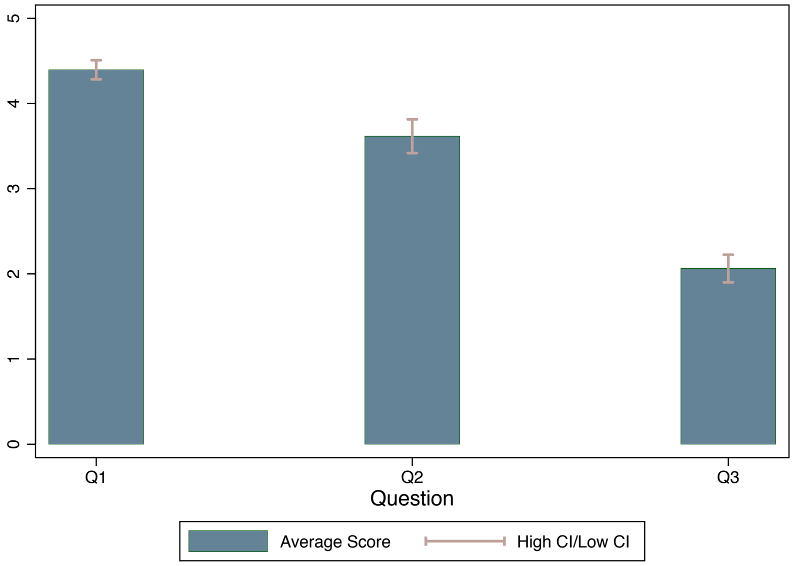
(A) Negative earnings surprises



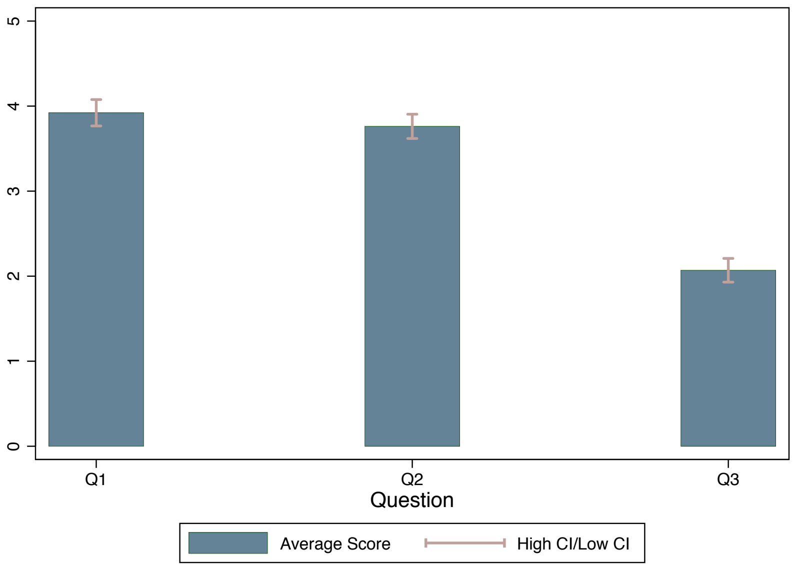
(B) Positive earnings surprises

Figure 3.4. Stock Price Reactions to Corporate Earnings Announcements

This figure presents event study analysis of stock price reactions to earnings announcements of sample firms. Panel A (B) depicts the mean cumulative abnormal stock returns and 95% confidence intervals around negative (positive) earnings surprises, measured over 10-day event windows around earnings announcement dates. Earnings announcements in the sample are characterized as negative (positive) surprises if the average earnings-per-share (EPS) forecast of equity analysts in the quarter preceding the earnings announcement is lower (greater) than the realized EPS that is announced by the firm. Benchmark factor loadings are estimated using daily returns for 100 days, starting 50 days prior to the start of the event window. The horizontal axis corresponds to the day relative to the earnings announcement date, and the vertical axis corresponds to the average cumulative abnormal stock return measured in percentage terms.



(A) Outflow survey



(B) Inflow survey

Figure 3.5. LinkedIn Survey Responses

This figure depicts the responses to survey questions administered to individual workers in our sample. Panel A (B) presents the average scores (and 95% confidence intervals) for each question asked of workers in our outflow (inflow) survey. Individual questions for each survey are listed in the appendix. Scores are obtained for 230 (169) workers in the outflow (inflow) survey samples. The horizontal axis corresponds to the specific question asked in each survey, and the vertical axis presents the average response score.

3.7 Tables

Table 3.1: Descriptive Statistics for Sample Workers and Firms

This table presents descriptive statistics that characterize the workers and firms in our sample. The sample contains 1,500,457 individual employment records for 1,028,356 employees at Russell 1000 firms between 1985 and 2016. Panel A summarizes data at the level of employment record. *Occupations* are inferred from individuals' job titles as described in the Data section; panel A shows six of the most common occupations in our sample. *Experience* is the cumulative years worked for an individual prior to the start of an employment spell. *Education* refers to the highest level of educational attainment reached by a given worker. Panel B presents summary statistics of firms in our data set, averaged across all firm-years in the sample. *Total assets* is the book value of assets. *Market value of equity* is the number of shares outstanding times the closing share price as of the most recent date for which data is available. *B/M of equity* is the ratio of book value of equity to the market value of equity. *Return on assets* is defined as the ratio of net income to the book value of assets. *Leverage* is defined as the ratio of total short-term and long-term debt obligations to total book value of assets. Panel C summarizes labor flows at the firm-level over a 1-month period. The *Outflow* (*Inflow*) is computed as the total number of employees whose job spells at a given company ends (begins) in a given month. The *Net outflow* is computed as the difference between the *Outflow* and *Inflow*. The *Standardized net outflow* is *Net outflow* divided by the total number of employees that work at the firm as of the beginning of the month.

	A. Employee characteristics						
Occupation	Engineers	Scientists	Mid-managers	Admin.	Finance	Consultants	Others
Obs.	215,111	93,620	326,228	145,196	20,853	92,085	607,364
Frac.	14.34%	6.24%	21.74%	9.67%	1.39%	6.14%	40.48%
Education	PhD	MBA	Master's	Bachelor's	High school	Unreported	
Obs.	58,210	218,314	193,708	565,256	186,859	278,110	
Frac.	3.88%	14.55%	12.91%	37.67%	12.45%	18.54%	
Experience (years)	Mean	SD	5th pctl	25th pctl	50th pctl	75th pctl	95th pctl
	5.63	5.90	0.25	1.50	3.67	7.83	17.83
	B. Firm characteristics						
	Mean	SD	5th pctl	25th pctl	50th pctl	75th pctl	95th pctl
Total assets ($b)	25.19	117.52	0.19	1.37	4.51	14.26	82.58
Equity market value ($b)	12.26	32.42	0.15	1.17	3.48	9.43	50.45
B/M of equity	0.59	0.77	0.092	0.26	0.45	0.77	1.38
Return on assets (%)	1.14	4.13	-2.14	0.36	1.17	2.30	4.74
Total employees (,000s)	28.17	77.40	0.26	2.46	8.20	25.01	118.50
Leverage	0.59	0.23	0.19	0.45	0.59	0.74	0.93
	C. Monthly labor flows						
	Mean	SD	5th pctl	25th pctl	50th pctl	75th pctl	95th pctl
Outflow	4.10	15.27	0	0	0	3	18
Inflow	5.43	18.13	0	0	1	4	23
Net outflow	-1.33	9.89	-9	-2	0	1	4
Standardized net outflow	-0.0071	0.049	-0.056	-0.011	0	0.0014	0.028

Table 3.2: Results from Calendar-time Portfolio Return Analysis

This table presents coefficient estimates from our calendar-time portfolio return analysis. Each month, firms are sorted into quartiles (or terciles) based on the net labor outflows realized over the previous one month. The long (short) portfolio consists of firms with the lowest (highest) realized net labor outflows. The long-short portfolios are rebalanced monthly and returns are computed using both value- (VW) and equal-weighted (EW) specifications. Abnormal returns are assessed using the five-factor model (Fama and French, 2015): $r_{p,t} = \alpha + \beta * MP_t + s * SMB_t + h * HML_t + r * RMW_t + c * CMA_t + \epsilon_t$, where MP is the market premium calculated as the value weighted market return on all NYSE-Amex-Nasdaq stocks minus the 1-month Treasury-bill rate, SMB (small minus big) is the average return of small firms minus the average return of big firms, HML (high minus low) is the average return of value (high book-to-market) firms minus the average return of growth (low book-to-market) firms, RMW (robust minus weak) is the average return of robust-profitability firms minus the average return of weak-profitability firms, and CMA (conservative minus aggressive) is the average return of firms with low investment minus the average returns of firms with high investment. Crisis periods are defined by NBER recession dates. Sample firms correspond to the Russell 1000 index as of June 2018. Monthly returns and alphas are reported in percentages for the long-short portfolio and each individual quartile (tercile), and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
α (%)	0.415***	0.337**	0.316***	0.250*	0.389***	0.306**	0.364***	0.285**
	(3.546)	(2.114)	(2.868)	(1.818)	(3.685)	(2.080)	(3.501)	(2.003)
MP	-0.096***	-0.086*	-0.082***	-0.090**	-0.095**	0.042	-0.041	0.060
	(-3.057)	(-1.847)	(-2.750)	(-2.200)	(-2.515)	(1.003)	(-1.146)	(1.279)
SMB	0.130***	0.143**	0.077	0.074	0.085	0.126*	0.096*	0.084
	(2.603)	(2.070)	(1.592)	(1.285)	(1.208)	(1.752)	(1.699)	(1.086)
HML	-0.229***	-0.176**	-0.205***	-0.087	-0.312***	-0.287***	-0.276***	-0.212**
	(-3.505)	(-2.051)	(-3.299)	(-1.175)	(-4.499)	(-3.946)	(-4.167)	(-2.273)
RMW	0.030	0.018	0.051	0.020	-0.137	0.002	-0.135*	-0.128
	(0.316)	(0.145)	(0.569)	(0.208)	(-1.358)	(0.015)	(-1.763)	(-1.013)
CMA	-0.093	-0.304**	-0.102	-0.403***	-0.112	-0.191	-0.148	-0.248**
	(-0.738)	(-2.099)	(-0.860)	(-3.003)	(-0.933)	(-1.270)	(-1.510)	(-2.057)
R^2	.162	.146	.132	.135	.351	.214	.322	.191
Raw long return (%)	1.813	1.235	1.781	1.242	1.461	0.973	1.823	1.395
Raw short return (%)	1.522	1.067	1.569	1.115	1.188	0.662	1.576	1.126
Low NetOutflows (α%)	0.196***	0.188*	0.116*	0.124*	0.218**	0.163	0.214**	0.236**
Quartile 2 (α%)	-0.007	0.002	–	–	0.016	0.031	0.014	-0.046
Middle tercile (α%)	–	–	-0.010	-0.007	–	–	–	–
Quartile 3 (α%)	-0.044	0.046	–	–	-0.073	-0.117	-0.103	-0.087
High NetOutflows (α%)	-0.220**	-0.149*	-0.200*	-0.125*	-0.170*	-0.143*	-0.150*	-0.049
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table 3.3: Net Labor Outflows and Earnings Surprises

This table reports coefficient estimates for an ordinary least squares model of corporate earnings surprises regressed on net labor outflows and various controls: $SUE_{i,t+j} = \beta_0 + \beta_1 NetOutflows_{i,t} + \beta_2 E^+_{i,t} + \beta_3 NEGE_{i,t} + \beta_4 ACC^-_{i,t} + \beta_5 ACC^+_{i,t} + \beta_6 AG_{i,t} + \beta_7 DD_{i,t} + \beta_8 DIV_{i,t} + \beta_9 PRICE_{i,t} + \beta_{10} BTM_{i,t} + \epsilon_{i,t}$. Standardized unexpected earnings ($SUE_{i,t+j}$) is defined as $(EPS^{actual}_{i,t+j} - \mu^{forecast}_{i,t})/\sigma^{forecast}_{i,t}$, where $EPS^{actual}_{i,t+j}$ is the next EPS of firm i announced in month $t+j$, $\mu^{forecast}_{i,t}$ is the mean of financial analysts' forecast reported in month t, and $\sigma^{forecast}_{i,t}$ is the standard deviation of the forecasts made in month t. $NetOutflows_{i,t}$ is the net labor outflows of firm i from month $t-1$ to t. Following So (2013), we include the following controls: earnings per share when earnings are positive and zero otherwise (E^+), a binary variable indicating negative earnings ($NEGE$), negative and positive accruals per share (ACC^-, ACC^+), the percentage change in total assets (AG), a binary variable indicating zero dividends (DD), dividends per share (DIV), share price ($PRICE$), and book-to-market value (BTM). Firm and year-month fixed effects are also included. Robust standard errors are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NetOutflows_{i,t}$	-0.706*	-0.889**	-0.892**	-1.233***	-0.906*	-0.741	-1.187**	-1.109**
	(0.426)	(0.421)	(0.384)	(0.375)	(0.493)	(0.484)	(0.461)	(0.448)
$E^+_{i,t}$			-0.029	0.048**			-0.069***	0.038*
			(0.018)	(0.019)			(0.019)	(0.020)
$NEGE_{i,t}$			-0.359***	-0.082*			-0.512***	0.130**
			(0.039)	(0.046)			(0.048)	(0.058)
$ACC^-_{i,t}$			-0.045***	-0.014			-0.023*	0.021
			(0.012)	(0.012)			(0.013)	(0.014)
$ACC^+_{i,t}$			0.002	-0.029**			-0.002	-0.057***
			(0.013)	(0.013)			(0.016)	(0.016)
$AG_{i,t}$			0.336***	0.270***			0.298***	0.094
			(0.066)	(0.065)			(0.084)	(0.083)
$DD_{i,t}$			0.506***	-0.160***			0.396***	-0.070
			(0.044)	(0.062)			(0.066)	(0.100)
$DIV_{i,t}$			-0.001	-0.004***			-0.003*	-0.003**
			(0.001)	(0.002)			(0.001)	(0.002)
$PRICE_{i,t}$			0.003***	-0.002***			0.002***	-0.003***
			(0.000)	(0.000)			(0.000)	(0.000)
$BTM_{i,t}$			0.007	0.006			0.014	0.008
			(0.010)	(0.010)			(0.011)	(0.011)
$Constant$	0.786***	0.785***	0.311***	1.087***	1.052***	1.053***	0.784***	1.315***
	(0.015)	(0.014)	(0.045)	(0.062)	(0.018)	(0.017)	(0.066)	(0.099)
R^2	.000	.078	.005	.113	.000	.095	.003	.124
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Time FE	N	Y	N	Y	N	Y	N	Y
Firm FE	N	Y	N	Y	N	Y	N	Y

Table 3.4: Portfolio Return Analysis across Firms with Varying Financial Transparency

This table presents our portfolio return analysis and Fama-MacBeth regression estimates for firms with varying degrees of financial transparency. Following Leuz et al. (2003), we proxy for financial transparency by measuring the extent to which firms engage in earnings management. Earnings management is measured by the absolute value of a firm's accruals scaled by the absolute value of cash flow from operations. Panel A reports coefficient estimates from portfolio return analysis of firms characterized by high versus low measures of financial transparency. All other variables are defined in Table 2. Abnormal returns are assessed using the five-factor model. For brevity, we do not report coefficient estimates for all factors. Alphas are expressed as monthly percentages, and t-statistics are in parentheses. Panel B reports the results of Fama and MacBeth (1973) cross-sectional regressions of monthly returns on lagged individual firm characteristics. Market betas are estimated from time-series regressions of individual excess returns on market premiums. In the cross-section, firm size ME, book-to-market B/M, operating profitability OP and investment INV are measured following Fama and French (2015). NLO is the net labor outflows of an individual firm in the prior month. $LowT$ is our measure of financial transparency that is decreasing in earnings management (i.e., higher values imply lower transparency), and t-statistics are reported in parentheses and are based on Newey and West (1987) standard errors corrected for autocorrelation using 12 lags. * p<.1; ** p<.05; *** p<.01.

	A. Portfolio return analysis							
Low transparency α (%)	0.732***	0.553**	0.575***	0.406*	0.554***	0.268	0.487***	0.506**
	(3.765)	(2.226)	(3.235)	(1.776)	(3.415)	(1.212)	(2.807)	(2.446)
High transparency α (%)	0.310**	0.529***	0.221*	0.371**	0.397***	0.378*	0.443***	0.393*
	(2.326)	(2.835)	(1.674)	(2.220)	(2.914)	(1.926)	(3.497)	(1.935)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude
	B. Fama − MacBeth cross − sectional regressions							
	β	$ln(ME)$	$ln(B/M)$	$ln(OP)$	$ln(INV)$	NLO	$NLO * LowT$	
Fama-MacBeth coefficient	0.411	-0.271***	-0.096	0.141**	-0.008	-0.016*	-0.017**	
	(1.513)	(-7.016)	(-1.028)	(1.999)	(-0.420)	(-1.719)	(-1.773)	

Table 3.5: Portfolio Return Analysis Based on Gross Outflows and Gross Inflows

This table reports coefficient estimates from our portfolio return analysis, where firms are sorted into long and short portfolios based on either gross labor outflows (panel A) or (negative) gross labor inflows (panel B) each month. Abnormal returns are assessed using the five-factor model (Fama and French, 2015): $r_{p,t} = \alpha + \beta * MP_t + s * SMB_t + h * HML_t + r * RMW_t + c * CMA_t + \epsilon_t$. The long (short) portfolio consists of firms with the lowest (highest) realized gross labor flows. Alphas are expressed as monthly percentages. All other variables are defined as per Table 2. For brevity, we do not report coefficient estimates for all factors. The t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Outflows							
α (%)	0.208**	0.012	0.244***	-0.013	0.205**	0.175	0.199**	0.137
	(2.483)	(0.107)	(3.164)	(-0.119)	(2.043)	(1.122)	(2.002)	(0.928)
	B. Negative inflows							
α (%)	0.159*	0.383***	0.098	0.252**	0.092	-0.028	0.104	-0.011
	(1.902)	(3.005)	(1.276)	(2.411)	(0.911)	(-0.188)	(1.037)	(-0.080)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table 3.6: Information Content in Labor Flows: Production Costs

Panel A of this table reports OLS regression coefficients of corporate fundamentals regressed on net labor outflows: $y_{i,t+1} = a + b * NetLaborOutflow_{i,t} + \lambda_i + \lambda_t + \epsilon_{i,t+1}$. The independent variable $NetLaborOutflow_{i,t}$ is the firm's net labor outflow of the current quarter, normalized by the total number of employees in the firm at the beginning of the quarter. The dependent variable $y_{i,t+1}$ is obtained from Compustat and is either: SG&A (sales, general, and administration) expenses, operating expenses, revenues, or operating income after depreciation and amortization, in the following quarter. All dependent variables are normalized by book total assets at the beginning of the current quarter. Quarter-year and firm fixed effects are also included as independent variables. Panels B through D present the alphas from portfolio return analysis using labor flows of different subsets of workers. Each row corresponds to the worker characteristic that is used to compute labor flows. Panel B presents the results for long-short portfolios sorted by the net labor outflows of workers within specific occupations. Panel C presents the results for long-short portfolios sorted by workers with above- versus below-sample median levels of labor market experience. Panel D presents the results for long-short portfolios sorted by workers with different levels of educational attainment. Abnormal returns are assessed using the five-factor model (Fama and French, 2015), and alphas are in monthly percentage terms. All other variables are defined in Table 2. The t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Accounting fundamentals							
	SG&A expense		Operating expense		Revenues		Operating income	
Net labor outflow	0.019***	0.016**	0.016**	0.013**	0.004	-0.006	-0.014***	-0.016***
	(3.167)	(2.286)	(2.011)	(1.857)	(0.410)	(-0.758)	(-4.603)	(-5.333)
Starting year	1985	2005	1985	2005	1985	2005	1985	2005
	B. Employee occupation							
Engineers	0.414**	0.476**	0.414**	0.475**	0.339**	0.188	0.351***	0.090
	(2.575)	(2.509)	(2.572)	(2.504)	(2.459)	(0.952)	(2.623)	(0.484)
Scientists	0.268*	0.164	0.268*	0.164	0.333**	0.395*	0.393***	0.448**
	(1.793)	(0.918)	(1.793)	(0.918)	(2.579)	(1.945)	(2.812)	(2.187)
Managers	0.306***	0.434***	0.198*	0.321**	0.325***	0.389**	0.335***	0.502***
	(2.632)	(2.958)	(1.780)	(2.455)	(3.144)	(2.417)	(3.266)	(3.658)
Administration	0.113	0.057	0.105	0.056	0.161	0.045	0.220	-0.020
	(0.861)	(0.362)	(0.802)	(0.353)	(1.239)	(0.260)	(1.609)	(-0.115)
Finance	0.012	0.228	0.012	0.228	0.052	-0.131	0.198	-0.282
	(0.056)	(0.863)	(0.056)	(0.863)	(0.245)	(-0.397)	(1.188)	(-1.452)
Consultant	0.102	0.123	0.102	0.123	0.243*	0.228	0.293**	0.158
	(0.588)	(0.599)	(0.588)	(0.599)	(1.867)	(1.267)	(2.496)	(0.835)
	C. Employee work experience							
High	0.241**	0.285**	0.228**	0.245**	0.332***	0.329*	0.432***	0.353**
	(2.398)	(2.049)	(2.343)	(1.991)	(2.859)	(1.658)	(3.938)	(2.164)
Low	0.144	0.234**	0.076	0.216*	0.318	0.330*	0.304	0.242
	(1.480)	(1.971)	(0.885)	(1.942)	(1.650)	(1.903)	(1.533)	(1.458)
	D. Employee education							
PhD/MBA/Master's	0.359***	0.252**	0.372***	0.332**	0.358**	0.336*	0.344**	0.297
	(3.905)	(2.084)	(3.960)	(2.481)	(2.565)	(1.741)	(2.504)	(1.534)
Bachelor's/High school	0.093	0.191*	0.060	0.210**	0.117	0.215	0.136*	0.146
	(1.297)	(1.677)	(0.928)	(2.048)	(1.591)	(1.644)	(1.840)	(1.202)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table 3.7: Return Persistence in Portfolio Return Analysis

This table presents results from portfolio return analysis using monthly returns computed following N-month gaps between the trading period and the sorting period, where N ranges from 0 to 6 (N=0 corresponds to the main results presented in Table 2). Each month, firms are sorted into quartiles (or terciles) based on the net labor outflows realized over the previous $N+1$ month. The long (short) portfolio consists of firms with the lowest (highest) realized net labor outflows. Abnormal returns are assessed using the five-factor model (Fama and French, 2015). Each cell contains the coefficient estimate for alpha in monthly percentage terms. All other variables are defined in Table 2. The *t*-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-month gap	0.415***	0.337**	0.316***	0.250*	0.389***	0.306**	0.364***	0.285**
	(3.546)	(2.114)	(2.868)	(1.818)	(3.685)	(2.080)	(3.501)	(2.003)
1-month gap	0.542***	0.346	0.439***	0.185	0.267	0.077	0.332**	0.087
	(3.242)	(1.638)	(2.929)	(1.056)	(1.590)	(0.325)	(2.277)	(0.376)
2-month gap	0.216*	0.215	0.105	0.176	0.187	-0.116	0.273**	-0.051
	(1.669)	(1.196)	(0.871)	(1.001)	(1.523)	(-0.542)	(2.335)	(-0.256)
3-month gap	0.103	0.258	-0.072	-0.005	0.064	0.041	0.163	0.212
	(0.743)	(1.138)	(-0.545)	(-0.025)	(0.453)	(0.212)	(1.231)	(1.099)
4-month gap	0.218	0.401**	0.098	0.141	0.071	0.314	0.135	0.238
	(1.506)	(2.254)	(0.793)	(0.968)	(0.488)	(1.463)	(0.985)	(1.088)
5-month gap	0.225	0.271	0.140	0.167	-0.117	0.043	-0.050	0.223
	(1.436)	(1.332)	(0.993)	(1.072)	(-0.862)	(0.231)	(-0.380)	(1.179)
6-month gap	0.309**	0.290	0.124	0.175	0.199	0.152	0.151	0.209
	(1.984)	(1.607)	(1.024)	(1.194)	(1.429)	(0.744)	(1.157)	(1.019)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table 3.8: Insider Trading and Labor Flows

This table presents ordinary least squares regression estimates of insider trades as a function of net labor flows: $NetLaborOutflow_{i,t+L} = a + b * InsiderTrade_{i,t} + \lambda_i + \lambda_t + \epsilon_{i,t+L}$, where L = 0, 1, 2, 3, 4, 5, or 6 month(s). The main independent variable $InsiderTrade_{i,t}$ is the number of net shares sold by the insiders of firm i in month t, normalized by the number of outstanding shares at the beginning of the month. Insiders are categorized as either "routine traders" or "opportunistic traders" following Cohen et al. (2012). Panel A uses the sample of all insiders, and panel B only uses "opportunistic traders" who are more likely to possess and exploit insider information. The dependent variable is the net labor outflows computed after L month(s) following the observed insider trades in a given month. All specifications include year-month fixed effects and firm fixed effects. Robust standard errors are in parentheses. * p<.1; ** p<.05; *** p<.01.

	A. All insiders						
$InsiderTrade$	0.001	-0.001	-0.006	-0.002	-0.006	-0.005	0.008
	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.006)
R^2	.060	.058	.054	.060	.060	.057	.054
L	0 month	1 month	2 months	3 months	4 months	5 months	6 months
	B. Opportunistic insiders						
$InsiderTrade$	0.020	0.065	0.008	0.035**	0.004	0.000	-0.036
	(0.023)	(0.044)	(0.021)	(0.015)	(0.021)	(0.020)	(0.028)
R^2	.092	.096	.106	.111	.089	.092	.100
L	0 month	1 month	2 months	3 months	4 months	5 months	6 months

Table 3.9: Portfolio Return Analysis for Firms Facing Varying Labor Adjustment Costs

This table reports results of portfolio return analysis for firms that face high versus low labor adjustment costs. We proxy for labor adjustments costs using the share of total labor force unemployment that belongs to a state in which the firm's headquarters are located; above-median (below-median) unemployment shares correspond to low (high) labor adjustment costs faced by firms. Panel A (Panel B) reports the estimates for firms located in states with high (low) unemployment shares. Abnormal returns are assessed using the five-factor model (Fama and French, 2015). Alphas are expressed as monthly percentages. All other variables defined as per Table 2. The t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. High unemployment share (low labor adjustment costs)							
α (%)	0.541***	0.627***	0.376**	0.556***	0.359**	0.338*	0.336**	0.277
	(3.207)	(2.947)	(2.422)	(3.146)	(2.210)	(1.720)	(1.993)	(1.419)
	B. Low unemployment share (high labor adjustment costs)							
α (%)	0.275**	0.131	0.213*	0.094	0.421***	0.333*	0.399***	0.389**
	(2.027)	(0.722)	(1.681)	(0.560)	(3.118)	(1.727)	(3.049)	(2.124)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

3.8 Appendix

Table A3.1: Portfolio Return Analysis under Alternative Specifications

This table presents results from our portfolio return analysis under a variety of alternative specifications to the main sample results presented in Table 2. Each month, firms are sorted into quartiles/terciles/quintiles based on the net labor outflows realized over the previous 1, 2, 3, and 6 month(s). The long (short) portfolio consists of firms with the lowest (highest) realized net labor outflows. The long-short portfolios are rebalanced monthly and returns are computed using both value- (VW) and equal-weighted (EW) specifications. The sample runs from January 1985 (or January 2005) to December 2016, including or excluding the NBER recession periods. Abnormal returns are assessed using the five-factor model (Fama and French, 2015): $r_{p,t} = \alpha + \beta * MP_t + s * SMB_t + h * HML_t + r * RMW_t + c * CMA_t + \epsilon_t$, where MP is the market premium calculated as the value weighted market returns of all NYSE-Amex-Nasdaq stocks minus the 1-month Treasury-bill rate, SMB (small minus big) is the average return of small firms minus the average return of big firms, HML (high minus low) is the average return of value (high book-to-market) firms minus the average return of growth (low book-to-market) firms, RMW (robust minus weak) is the average return of robust-profitability firms minus the average return of weak-profitability firms, and CMA (conservative minus aggressive) is the average return of firms with low investment minus the average returns of firms with high investment. Returns and alphas are in monthly percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

α (%)	0.431***	0.400**	0.475***	0.389**	0.289**	0.440***	0.327**	0.355*
	(3.035)	(2.129)	(3.226)	(2.032)	(2.345)	(2.650)	(2.614)	(1.990)
Sorting window	1 month	1 month	1 month	1 month	1 month	1 month	2 months	2 months
Starting year	1995	1995	1995	1995	2010	2010	2010	2010
Portfolio cutoff	Quartile	Quartile	Quartile	Quartile	Quintile	Quintile	Quintile	Quintile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Exclude	Exclude	Include	Include	Include	Include
α (%)	0.389***	0.392**	0.406***	0.428***	0.380***	0.401**	0.348**	0.509**
	(3.062)	(2.425)	(3.134)	(2.675)	(3.169)	(2.462)	(2.366)	(2.450)
Sorting window	2 months	2 months	2 months	2 months	3 months	3 months	3 months	3 months
Starting year	1985	1985	1985	1985	1985	1985	1995	1995
Portfolio cutoff	Quartile	Quartile	Quartile	Quartile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Exclude	Exclude	Include	Include	Include	Include
α (%)	0.405***	0.277*	0.505***	0.332**	0.574***	0.347*	0.442***	0.273*
	(4.226)	(1.704)	(5.394)	(2.332)	(4.782)	(1.958)	(4.322)	(1.714)
Sorting window	3 months	3 months	6 months	6 months	6 months	6 months	6 months	6 months
Starting year	2005	2005	1985	1985	1985	1985	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Exclude	Exclude	Include	Include	Include	Include	Exclude	Exclude

Table A3.2: Analysis of Measurement Error in Portfolio Return Results

This table presents analysis of measurement error in our portfolio return results. We perform our portfolio analysis across firms for which our sample represents high versus low fractions of total employment, where total employment is measured using LinkedIn and Compustat. In panel A, we calculate the ratio of the total employees on LinkedIn for a given firm, divided by the total employees who work at the firm as reported by Compustat in 2018. In Panel B, we calculate the ratio of the firm's current and past employees in our sample, divided by the number of the firm's current and past employees in the LinkedIn population. Across both panels, we split our sample into firms with above versus below median employment coverage ratios, and we assume that firms with high (low) employment coverage ratios have less (more) measurement error in labor flow estimates. We repeat our portfolio return analysis (as per Table 2) for firms in each of these subsamples. Abnormal returns are assessed using the five-factor model (Fama and French, 2015). All other variables are defined in Table 2. Alphas are expressed as monthly percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. LinkedIn population versus Compustat							
Above median	0.309*	0.406**	0.439**	0.521**	0.598***	0.348*	0.552***	0.314
	(1.685)	(2.043)	(2.135)	(2.271)	(3.655)	(1.679)	(3.500)	(1.386)
Below median	0.286**	0.195	0.315***	0.246	0.132	0.153	0.131	0.129
	(2.528)	(1.146)	(2.717)	(1.384)	(0.816)	(0.892)	(0.836)	(0.664)
	B. LinkedIn sample versus LinkedIn population							
Above median	0.472***	0.482**	0.344**	0.316*	0.361**	0.360	0.345**	0.203
	(3.093)	(2.529)	(2.442)	(1.855)	(2.547)	(1.620)	(2.451)	(0.997)
Below median	0.367***	0.131	0.333**	0.228	0.347***	0.324*	0.332***	0.420**
	(2.711)	(0.743)	(2.517)	(1.356)	(3.636)	(1.939)	(3.362)	(2.424)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table A3.3: Analysis of Worker Sample Selection Bias in Portfolio Analysis Results

This table presents two sets of bootstrapped analyses of our trading strategy. The first analysis, depicted in Panel A, constructs a new sample of employee-weighted labor flows for each firm-month. Weights are calculated as the ratio of the number of firm employees in the LinkedIn population who belong to a specific category, divided by the number of firm employees in the LinkedIn population who do not belong to the same category; these weights are applied to individual workers in the sample who belong to categories that are undersampled by our dataset relative to the LinkedIn population. The second analysis, depicted in panel B, resembles a Monte Carlo simulation method. From our original dataset of employment records by firm, we create a new sample by drawing (without replacement) employment records for workers who belong to categories that are undersampled by our dataset relative to the LinkedIn population, until the new sample distribution of workers across categories matches that of the LinkedIn population. The worker category that we examine in both analyses is whether a worker attended a college listed in the top-20 universities ranked by *U.S. News* as of 2018. For each set of analyses, we draw 10,000 random samples of firm-month observations, and estimate the abnormal returns of our trading strategy for each of these random samples. We calculate the average and standard deviation of the alpha estimates across all random samples, and report these values below. We estimate abnormal returns using the five-factor model (Fama and French, 2015). All variables are defined in Table 2. Monthly alphas are expressed as percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Resampling method 1							
α (%)	0.391***	0.316**	0.308***	0.242*	0.366***	0.306**	0.360***	0.300**
	(3.301)	(1.979)	(2.906)	(1.784)	(3.491)	(2.151)	(3.548)	(2.201)
	B. Resampling method 2							
α (%)	0.347***	0.296***	0.284***	0.231***	0.312***	0.240***	0.316***	0.230***
	(6.732)	(4.139)	(6.405)	(3.506)	(6.792)	(3.358)	(8.281)	(3.752)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis period	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table A3.4: Portfolio Return Analysis Using Alternative Factor Models

This table presents results of portfolio analysis using alternative factor models. We repeat our analysis as per Table 2, but use different factors to estimate abnormal returns generated by our trading strategy. Each month, firms are sorted into quartiles (or terciles) based on the net labor outflows realized over the previous one month. The long (short) portfolio consists of firms with the lowest (highest) realized net labor outflows. The long-short portfolios are rebalanced monthly and returns are computed using both value- and equal-weighted specifications. Panel A adds the liquidity factor Pástor and Stambaugh (2003) to our benchmark Fama-French five-factor model. Panel B adds the momentum factor (Carhart, 1997). Panel C uses the q-factor model proposed by Hou et al. (2015). Monthly returns and alphas are in percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. FF 5-factor + liquidity							
α (%)	0.410***	0.328**	0.311***	0.242*	0.389***	0.305**	0.381***	0.290**
	(3.508)	(2.115)	(2.831)	(1.822)	(3.681)	(2.068)	(3.565)	(2.001)
MP	-0.084**	-0.067	-0.071**	-0.073*	-0.090**	0.036	-0.037	0.061
	(-2.511)	(-1.557)	(-2.216)	(-1.950)	(-2.288)	(0.848)	(-1.055)	(1.310)
SMB	0.132***	0.145**	0.078	0.075	0.082	0.129*	0.100*	0.085
	(2.611)	(2.124)	(1.601)	(1.326)	(1.198)	(1.771)	(1.759)	(1.104)
HML	-0.226***	-0.171**	-0.202***	-0.083	-0.313***	-0.285***	-0.267***	-0.209**
	(-3.468)	(-2.006)	(-3.240)	(-1.124)	(-4.615)	(-3.942)	(-4.016)	(-2.166)
RMW	0.033	0.023	0.054	0.024	-0.135	-0.001	-0.131*	-0.126
	(0.345)	(0.183)	(0.596)	(0.250)	(-1.312)	(-0.007)	(-1.715)	(-1.013)
CMA	-0.094	-0.307**	-0.104	-0.405***	-0.109	-0.195	-0.152	-0.250**
	(-0.748)	(-2.102)	(-0.868)	(-3.003)	(-0.920)	(-1.297)	(-1.550)	(-2.054)
LIQ	-0.029	-0.048	-0.027	-0.040	-0.012	0.014	-0.024	-0.009
	(-1.591)	(-1.491)	(-1.464)	(-1.242)	(-0.517)	(0.642)	(-1.041)	(-0.270)
R^2	.168	.156	.138	.143	.353	.216	.328	.192
	B. FF 5-factor + momentum							
α (%)	0.282***	0.232	0.187**	0.173	0.363***	0.291**	0.333***	0.260*
	(2.816)	(1.559)	(2.001)	(1.295)	(3.445)	(1.997)	(3.170)	(1.789)
MP	-0.059**	-0.057	-0.047**	-0.069	-0.057*	0.063	-0.034	0.065
	(-2.320)	(-1.195)	(-1.974)	(-1.616)	(-1.928)	(1.626)	(-0.994)	(1.445)
SMB	0.106***	0.124**	0.054	0.060	0.075	0.120*	0.082	0.073
	(2.935)	(1.974)	(1.537)	(1.121)	(1.395)	(1.712)	(1.444)	(0.963)
HML	-0.064	-0.046	-0.046	0.007	-0.201***	-0.222***	-0.237***	-0.182**
	(-1.103)	(-0.619)	(-0.848)	(0.097)	(-3.207)	(-3.281)	(-3.472)	(-2.040)
RMW	-0.042	-0.039	-0.019	-0.021	-0.163**	-0.013	-0.155**	-0.143
	(-0.617)	(-0.344)	(-0.299)	(-0.233)	(-2.053)	(-0.105)	(-2.015)	(-1.126)
CMA	-0.224**	-0.408***	-0.230**	-0.478***	-0.164	-0.221	-0.142	-0.243**
	(-2.321)	(-3.332)	(-2.500)	(-3.752)	(-1.584)	(-1.518)	(-1.506)	(-2.018)
UMD	0.247***	0.195***	0.240***	0.142***	0.155***	0.090***	0.081**	0.062
	(7.875)	(3.837)	(8.186)	(3.658)	(4.236)	(2.813)	(2.139)	(1.225)
R^2	.401	.239	.381	.192	.499	.252	.349	.200

(Continued)

(Continued)

	C. q-factor							
α (%)	0.272**	0.222	0.173	0.147	0.265**	0.263*	0.272**	0.194
	(2.220)	(1.414)	(1.530)	(1.066)	(2.085)	(1.802)	(2.328)	(1.307)
MP	-0.066**	-0.056	-0.052*	-0.056	-0.054	0.048	-0.030	0.073
	(-2.282)	(-1.211)	(-1.931)	(-1.339)	(-1.520)	(1.137)	(-0.833)	(1.610)
ME	0.207***	0.198**	0.148**	0.103*	0.105*	0.109	0.113	0.113
	(3.134)	(2.206)	(2.483)	(1.688)	(1.759)	(1.576)	(1.623)	(1.591)
IA	-0.315***	-0.492***	-0.289***	-0.459***	-0.391***	-0.428***	-0.310***	-0.364***
	(-3.942)	(-4.949)	(-3.860)	(-5.037)	(-4.020)	(-4.100)	(-3.164)	(-3.455)
ROE	0.244***	0.208**	0.257***	0.165**	0.256***	0.174**	0.124*	0.146*
	(3.258)	(2.480)	(3.567)	(2.176)	(2.666)	(2.427)	(1.757)	(1.689)
R^2	.199	.172	.183	.138	.304	.158	.160	.141
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table A3.5: Portfolio Return Analysis across Alternative Firm Samples

This table reports the results of our portfolio return analysis using alternative samples of firms. Columns 1 to 4 correspond to a sample of firms that are members of the Russell 1000 as of 2006. Columns 5 to 8 correspond to firms in the Russell 1000 as of 2018 (our main sample), but exclude firms that undertake an initial public offering (IPO) between 1985 and 2016. Abnormal returns are assessed using the five-factor model (Fama and French, 2015). All variables are defined in Table 2. Monthly alphas are expressed as percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
α (%)	0.370***	0.332**	0.325***	0.232*	0.375**	0.367***	0.252*	0.292**
	(2.869)	(2.004)	(2.887)	(1.669)	(2.443)	(3.054)	(1.789)	(2.582)
MP	-0.064*	-0.070	-0.062*	-0.051	-0.111**	-0.087**	-0.079*	-0.073**
	(-1.675)	(-1.562)	(-1.853)	(-1.320)	(-2.492)	(-2.561)	(-1.893)	(-2.245)
SMB	0.070	0.129*	0.066	0.113**	0.147**	0.110**	0.087	0.059
	(1.173)	(1.854)	(1.264)	(2.002)	(2.243)	(2.028)	(1.431)	(1.131)
HML	-0.222***	-0.244**	-0.183***	-0.174**	-0.173*	-0.211***	-0.048	-0.190***
	(-2.676)	(-2.446)	(-2.599)	(-2.302)	(-1.939)	(-3.049)	(-0.630)	(-2.900)
RMW	-0.056	-0.142	-0.007	-0.037	0.060	0.046	0.060	0.050
	(-0.574)	(-1.232)	(-0.083)	(-0.398)	(0.512)	(0.436)	(0.584)	(0.495)
CMA	-0.063	-0.155	-0.092	-0.160	-0.317**	-0.055	-0.432***	-0.052
	(-0.496)	(-1.040)	(-0.848)	(-1.276)	(-2.172)	(-0.403)	(-3.147)	(-0.417)
R^2	.113	.151	.105	.108	.137	.117	.120	.094
Starting year	1985	1985	1985	1985	1985	1985	1985	1985
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Tercile	Tercile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW

Table A3.6: Fama-MacBeth Regression Results

This table reports coefficient estimates of Fama and MacBeth (1973) cross-sectional regressions of monthly returns on individual lagged firm characteristics. Market beta is estimated from time-series regressions of individual excess returns on market premiums. In the cross-section, firm size ME, book-to-market B/M, operating profitability OP, and investment INV are measured following Fama and French (2015). $NetOutflows$ is the net labor outflows of an individual firm in the prior month. Financial distress is measured in five ways. The *Z-score* is derived from Altman (1968). The *O-score* is derived from Ohlson (1980). $CreditRating$ is a dummy variable that equals one if the firm is rated as investment grade (BBB and above) by the Standard & Poor's Issuer Credit Rating (ICR), and zero otherwise. $Chapter7/11$ is a dummy variable that equals one if the firm filed for bankruptcy under Chapter 7 or Chapter 11 in a given year, and zero otherwise. $MertonDD$ is the distance to default (DD) as per Bharath and Shumway (2008). We report t-statistics in parentheses, which are based on Newey and West (1987) standard errors corrected for autocorrelation using 12 lags. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β	0.420	0.405	0.333	0.141	0.398	0.300	0.136	0.382	0.287
	(1.553)	(1.503)	(1.188)	(0.541)	(1.452)	(1.104)	(0.520)	(1.403)	(1.060)
$ln(ME)$	-0.275***	-0.268***	-0.291***	-0.179***	-0.271***	-0.278***	-0.176***	-0.265***	-0.271***
	(-7.220)	(-7.013)	(-7.229)	(-4.206)	(-7.157)	(-7.053)	(-4.162)	(-6.950)	(-6.748)
$ln(B/M)$	-0.100	-0.006	-0.047	-0.063	-0.102	-0.159	-0.023	-0.012	-0.073
	(-1.074)	(-0.070)	(-0.448)	(-0.657)	(-1.088)	(-1.532)	(-0.239)	(-0.137)	(-0.737)
$ln(OP)$	0.138*	0.164**	0.155	0.143**	0.139*	0.096	0.147**	0.164**	0.116
	(1.963)	(2.415)	(1.629)	(2.043)	(1.955)	(1.082)	(2.259)	(2.391)	(1.353)
$ln(INV)$	-0.012	-0.014	-0.005	-0.024	-0.012	-0.004	-0.024	-0.014	-0.007
	(-0.627)	(-0.735)	(-0.237)	(-1.097)	(-0.588)	(-0.160)	(-1.121)	(-0.694)	(-0.311)
$NetOutflows$	-0.042**	-0.043**	-0.035**	-0.039**	-0.045**	-0.034**	-0.041***	-0.045**	-0.036**
	(-2.115)	(-2.125)	(-2.037)	(-2.807)	(-2.214)	(-2.180)	(-2.991)	(-2.257)	(-2.258)
Z-score		0.024***					0.027	0.023***	0.022***
		(3.665)					(1.080)	(3.518)	(2.632)
O-score			-0.053**						
			(-2.147)						
$CreditRating$				-0.229**			-0.249**		
				(-2.271)			(-2.332)		
$Chapter7/11$					0.329			0.325	
					(0.603)			(0.598)	
$MertonDD$						-0.018			-0.018
						(-0.784)			(-0.796)

Table A3.7: Earnings Surprises and Market Responses

This table reports coefficient estimates of cumulative abnormal returns ($CAR_{i,t}$) around earnings announcements regressed on equity analyst earnings surprises and net labor outflows. Abnormal returns are computed using the Fama-French five-factor model (Fama and French, 2015) over an event window of 20 days surrounding the earnings announcement date of firm i in month t. The estimation period for factor loadings is set to be -250 days up to -30 days from the announcement date. $NetOutflows_{i,t-1}$ is the net labor outflows of firm i in the prior month $t-1$. Standardized unexpected earnings ($SUE_{i,t}$) is defined as $(EPS^{actual}_{i,t} - \mu^{forecast}_{i,t-1})/\sigma^{forecast}_{i,t-1}$, where $EPS^{actual}_{i,t}$ is the EPS of firm i announced in month t, $\mu^{forecast}_{i,t-1}$ is the mean of financial analysts' forecast reported in the prior month $t-1$, and $\sigma^{forecast}_{i,t-1}$ is the standard deviation of the forecasts made in month $t-1$. The $SUE_{i,t}$ is expressed as the sum of two components. $\widehat{SUE}^{Labor}_{i,t}$ is the component of standardized unexpected earnings that is explained by the prior month's net labor outflows as per the regression specification in Table 3. $\widehat{SUE}^{Residual}_{i,t}$ is the component of standardized unexpected earnings that is not explained by labor flows (i.e., the residual). $CAR_{i,t}$ is regressed on $SUE_{i,t}$, $NetOutflows_{i,t-1}$, $\widehat{SUE}^{Labor}_{i,t}$, and $X = \widehat{SUE}^{Residual}_{i,t}$, with and without year-month fixed effects. Robust standard errors are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable= $CAR_{i,t}$					
$SUE_{i,t}$	0.257***	0.264***				
	(0.037)	(0.038)				
$NetOutflows_{i,t-1}$			-2.050*	-2.671**		
			(1.195)	(1.203)		
$\widehat{SUE}^{Labor}_{i,t}$					1.868*	2.334**
					(1.058)	(1.061)
$\widehat{SUE}^{Residual}_{i,t}$					0.257***	0.264***
					(0.037)	(0.038)
Observations	43,646	43,646	43,646	43,646	43,646	43,646
R^2	.015	.037	.000	.022	.015	.037
Year-month FE	N	Y	N	Y	N	Y

Table A3.8: Portfolio Return Analysis across Firms with Varying Financial Transparency

This table reports coefficient estimates from our portfolio return analysis for firms characterized by high versus low measures of financial transparency, as a supplement to our results in Table 4. Following Leuz et al. (2003), we proxy for financial transparency using three additional measures of earnings reporting quality, and assume that firms that engage in greater earnings management are less likely to be transparent to investors. In panel A, transparency is measured as the ratio of the firm's standard deviation of operating earnings divided by the standard deviation of the firm's cash flow from operations. In panel B, transparency is proxied by the correlation between changes in accounting accruals and changes in operating cash flows for the firm. In panel C, we compute transparency using the ratio of small profits to small losses, where small losses (profits) are defined as after-tax earnings scaled by total assets being in the range −0.01 to 0 (0 to 0.01). Across all measures, higher values correspond to higher firm transparency. All other variables are defined in Table 2. Abnormal returns are assessed using the five-factor model. For brevity, we do not report coefficient estimates for all factors. Alphas are expressed as monthly percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Alternative transparency measure 1							
Low	0.538***	0.496**	0.418**	0.266	0.504***	0.332*	0.484***	0.422**
	(2.912)	(2.285)	(2.343)	(1.330)	(2.945)	(1.941)	(2.818)	(2.177)
High	0.454***	0.568***	0.340**	0.418**	0.452***	0.404*	0.445***	0.500**
	(3.118)	(2.645)	(2.497)	(2.245)	(3.202)	(1.862)	(2.880)	(2.161)
	B. Alternative transparency measure 2							
Low	0.424***	0.504***	0.324**	0.283*	0.474***	0.397*	0.437***	0.506**
	(2.779)	(2.641)	(2.336)	(1.664)	(3.634)	(1.886)	(3.225)	(2.537)
High	0.491***	0.169	0.382**	0.196	0.415***	0.132	0.435***	0.216
	(2.831)	(0.737)	(2.334)	(0.913)	(2.681)	(0.641)	(2.766)	(1.060)
	C. Alternative transparency measure 3							
Low	0.495***	0.430**	0.399***	0.326**	0.504***	0.291*	0.477***	0.443***
	(3.776)	(2.467)	(3.208)	(2.061)	(4.129)	(1.776)	(3.718)	(2.751)
High	0.431	0.564*	0.318	0.436	0.216	0.482	0.317	0.434
	(1.539)	(1.723)	(1.243)	(1.453)	(0.795)	(1.306)	(1.187)	(1.106)
Starting year	1985	1985	1985	1985	2005	2005	2005	2005
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Quartile	Quartile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW
Crisis periods	Include	Include	Include	Include	Include	Include	Exclude	Exclude

Table A3.9: Portfolio Return Analysis of IPO Firms

This table presents portfolio return analysis for a sample of 3,612 firms that undertake initial public offerings (IPO) between 1995 and 2016. We perform our analysis over different periods of time after an IPO. Panel A (B) analyzes firms in the immediate three (five) years that follow an IPO. Panel C contains the full sample of firms across sample years. All variables are defined as per Table 2. Abnormal returns are assessed using the five-factor model (Fama and French, 2015). Monthly alphas are in percentages, and t-statistics are in parentheses. * p<.1; ** p<.05; *** p<.01.

	IPO less than 3 years				*IPO less than 5 years*				*All IPO firms*			
α (%)	1.183***	1.116**	1.104***	1.028**	1.178***	1.087***	1.150***	0.954**	0.993***	0.864**	0.935***	0.832**
	(2.948)	(2.240)	(2.840)	(2.158)	(3.343)	(2.636)	(3.395)	(2.449)	(3.159)	(2.203)	(3.063)	(2.214)
MP	0.048	0.164	0.068	0.132	-0.008	0.137	0.003	0.152	-0.020	0.064	-0.010	0.086
	(0.517)	(1.264)	(0.758)	(1.041)	(-0.098)	(1.215)	(0.044)	(1.419)	(-0.258)	(0.614)	(-0.133)	(0.851)
SMB	0.094	0.154	0.076	0.090	-0.023	-0.004	-0.036	-0.052	-0.055	0.037	-0.061	0.030
	(0.583)	(0.748)	(0.485)	(0.463)	(-0.159)	(-0.023)	(-0.253)	(-0.311)	(-0.400)	(0.221)	(-0.447)	(0.184)
HML	-0.382**	-0.677***	-0.334**	-0.506**	-0.445***	-0.750***	-0.433***	-0.551***	-0.361***	-0.611***	-0.354***	-0.557***
	(-2.299)	(-3.027)	(-2.137)	(-2.258)	(-3.087)	(-3.971)	(-3.183)	(-2.663)	(-2.773)	(-3.435)	(-2.836)	(-3.165)
RMW	0.371*	0.280	0.385*	0.192	0.413**	0.191	0.426**	0.193	0.227	0.122	0.262	0.151
	(1.670)	(0.913)	(1.775)	(0.669)	(1.984)	(0.826)	(2.122)	(0.921)	(1.201)	(0.516)	(1.468)	(0.673)
CMA	0.346	0.230	0.336	0.077	0.302	0.246	0.301	0.037	0.315	0.008	0.304	-0.048
	(1.085)	(0.621)	(1.092)	(0.230)	(1.018)	(0.754)	(1.061)	(0.128)	(1.157)	(0.023)	(1.172)	(-0.151)
R^2	.030	.060	.029	.044	.063	.093	.066	.072	.045	.099	.048	.094
Starting year	1985	1985	1985	1985	1985	1985	1985	1985	1985	1985	1985	1985
Portfolio cutoff	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Tercile	Tercile	Quartile	Quartile	Tercile	Tercile
EW/VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW

Table A3.10: Insider Trading and Portfolio Allocation by Labor Flows

This table presents OLS regression estimates of our portfolio sorting outcomes on insider trades: $D^{High}_{i,t+L}$ ($D^{Low}_{i,t+L}$) $= a + b * InsiderTrade_{i,t} + \lambda_i + \lambda_t + \epsilon_{i,t+L}$, where L = 0, 1, 2, 3, 4, 5, or 6 month(s). The dependent variable $D^{High}_{i,t+L}$ ($D^{Low}_{i,t+L}$) is a binary variable that equal one if firm i is sorted into the top (bottom) quartile of net labor outflows in month $t + L$ (computed after L month(s) following the observed insider trades in a given month t). The main independent variable $InsiderTrades_{i,t}$ is the number of net shares sold by the insiders of firm i in month t, normalized by the number of outstanding shares at the beginning of the month. All specifications include year-month fixed effects and firm fixed effects. Robust standard errors are in parentheses. * p<.1; ** p<.05; *** p<.01.

	A. High net outflow quartile						
$InsiderTrade$	-0.035	0.014	-0.036	-0.035	-0.021	-0.018	-0.001
	(0.058)	(0.040)	(0.034)	(0.032)	(0.036)	(0.042)	(0.036)
R^2	.081	.085	.081	.081	.082	.083	.082
L	0 month	1 month	2 months	3 months	4 months	5 months	6 months
	B. Low net outflow quartile						
$InsiderTrade$	0.076	-0.042	0.073	-0.010	-0.019	0.023	-0.133
	(0.106)	(0.039)	(0.048)	(0.049)	(0.050)	(0.049)	(0.098)
R^2	.079	.077	.077	.079	.077	.077	.077
L	0 month	1 month	2 months	3 months	4 months	5 months	6 months

Bibliography

V. Acharya and Z. Xu. Financial dependence and innovation: The case of public versus private firms. *Journal of Financial Economics*, 124(2):223–243, 2017.

Z. J. Acs and D. B. Audretsch. Innovation in large and small firms: an empirical analysis. *The American economic review*, pages 678–690, 1988.

M. Adelino, S. Ma, and D. Robinson. Firm age, investment opportunities, and job creation. *The Journal of Finance*, 72(3):999–1038, 2017.

A. Agrawal and P. Tambe. Private equity and workers' career paths: the role of technological change. *The Review of Financial Studies*, 29(9):2455–2489, 2016.

E. I. Altman. Financial rations, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):589–609, 1968.

S. Andersen, J. Y. Campbell, K. M. Nielsen, and T. Ramadorai. Sources of inaction in household finance: Evidence from the danish mortgage market. *American Economic Review*, 2020.

A. Anderson and D. T. Robinson. Knowledge, fear and beliefs: Understanding household demand for green investments. *Swedish House of Finance Research Paper*, 2019.

R. Baghai, R. Silva, V. Thell, and V. Vig. Talent in distressed firms: Investigating the labor costs of financial distress. *Available at SSRN 2854858*, 2018.

M. Bailey, R. Cao, T. Kuchler, and J. Stroebel. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276, 2018a.

M. Bailey, R. Cao, T. Kuchler, J. Stroebel, and A. Wong. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018b.

M. Baldauf, L. Garlappi, and C. Yannelis. Does climate change affect real estate prices? only if you believe in it. *The Review of Financial Studies*, 33(3):1256–1295, 2020.

A. Banerjee, A. Finkelstein, R. Hanna, B. A. Olken, A. Ornaghi, and S. Sumarto. The challenges of universal health insurance in developing countries: Evidence from a large-scale randomized experiment in indonesia. Technical report, National Bureau of Economic Research, 2019.

B. M. Barber and T. Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2): 785–818, 2008.

N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.

A. Bassamboo, R. Cui, and A. Moreno. Wisdom of crowds in operations: Forecasting using prediction markets. *Available at SSRN 2679663*, 2015.

P. Bayer, S. L. Ross, and G. Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196, 2008.

C. M. Beckman. The influence of founding team company affiliations on firm behavior. *Academy of Management Journal*, 49(4):741–758, 2006.

F. Belo, X. Lin, and S. Bazdresch. Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, 122(1):129–177, 2014.

F. Belo, J. Li, X. Lin, and X. Zhao. Labor-force heterogeneity and asset prices: The importance of skilled labor. *The Review of Financial Studies*, 30(10):3669–3709, 2017.

D. Bernhardt. Strategic promotion and compensation. *The Review of Economic Studies*, 62(2): 315–339, 1995.

G. Bernile, V. Bhagwat, and P. R. Rau. What doesn't kill you will only make you more risk-loving: Early-life disasters and ceo behavior. *The Journal of Finance*, 72(1):167–206, 2017.

A. Bernstein, M. T. Gustafson, and R. Lewis. Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics*, 134(2):253–272, 2019.

S. Bernstein, A. Korteweg, and K. Laws. Attracting early-stage investors: Evidence from a randomized field experiment. *The Journal of Finance*, 72(2):509–538, 2017.

S. T. Bharath and T. Shumway. Forecasting default with the merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369, 2008.

P. Bordalo, N. Gennaioli, and A. Shleifer. Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285, 2012.

P. Bordalo, N. Gennaioli, and A. Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013.

A. Brav and P. A. Gompers. The role of lockups in initial public offerings. *The Review of Financial Studies*, 16(1):1–29, 2003.

J. Brown and D. A. Matsa. Boarding a sinking ship? An investigation of job applications to distressed firms. *The Journal of Finance*, 71(2):507–550, 2016.

J. Brown, T. Hossain, and J. Morgan. Shrouded attributes and information suppression: Evidence from the field. *The Quarterly Journal of Economics*, 125(2):859–876, 2010.

J. R. Brown and G. Martinsson. Does transparency stifle or facilitate innovation? *Management Science*, 65(4):1600–1623, 2018.

M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso. The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, 130(1):371–414, 2015.

P. Cahuc, S. Carcillo, and A. Zylberberg. *Labor Economics*. MIT press, 2014.

J. Cai, A. de Janvry, and E. Sadoulet. Subsidy policies and insurance demand. *American Economic Review*, 110(8):2422–53, 2020.

J. Y. Campbell, J. Hilscher, and J. Szilagyi. In search of distress risk. *The Journal of Finance*, 63(6):2899–2939, 2008.

M. M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1): 57–82, 1997.

T. Y. Chang, W. Huang, and Y. Wang. Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies*, 85(3):1609–1634, 2018.

R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–77, 2009.

D. Choi, Z. Gao, and W. Jiang. Attention to global warming. *The Review of Financial Studies*, 33(3):1112–1145, 2020.

J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick. Reinforcement learning and savings behavior. *The Journal of finance*, 64(6):2515–2534, 2009.

F. Cingano and A. Rosolia. People I know: Job search and social networks. *Journal of Labor Economics*, 30(2):291–332, 2012.

D. A. Cohen, A. Dey, and T. Z. Lys. Real and accrual-based earnings management in the pre-and post-sarbanes-oxley periods. *The accounting review*, 83(3):757–787, 2008.

L. Cohen, C. Malloy, and L. Pomorski. Decoding inside information. *The Journal of Finance*, 67(3):1009–1043, 2012.

S. Cole, X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. Barriers to household risk management: Evidence from india. *American Economic Journal: Applied Economics*, 5 (1):104–35, 2013.

M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE Transactions on Image Processing*, 27(10):5142–5154, 2018.

B. Cowgill and E. Zitzewitz. Corporate prediction markets: Evidence from Google, Ford, and firm X. *The Review of Economic Studies*, 82(4):1309–1341, 2015.

K. Daniel, D. Hirshleifer, and A. Subrahmanyam. Investor psychology and security market under-and overreactions. *The Journal of Finance*, 53(6):1839–1885, 1998.

O. Dessaint and A. Matray. Do managers overreact to salient risks? evidence from hurricane strikes. *Journal of Financial Economics*, 126(1):97–121, 2017.

P. A. Diamond. Wage determination and efficiency in search equilibrium. *The Review of Economic Studies*, 49(2):217–227, 1982.

A. Dittmar and R. Duchin. Looking in the rearview mirror: The effect of managers' professional experience on corporate financial policy. *The Review of Financial Studies*, 29(3):565–602, 2016.

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

A. Edmans. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640, 2011.

M. Ewens and M. Marx. Founder replacement and startup performance. *The Review of Financial Studies*, 31(4):1532–1565, 2018.

M. Ewens and M. Rhodes-Kropf. Is a vc partnership greater than the sum of its partners? *The Journal of Finance*, 70(3):1081–1113, 2015.

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

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

A. Fedyk and J. Hodson. Trading on talent: Human capital and firm performance. *Available at SSRN 3017559*, 2019.

A. Finkelstein. E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124 (3):969–1010, 2009.

A. Finkelstein, N. Hendren, and E. F. Luttmer. The value of medicaid: Interpreting results from the oregon health insurance experiment. *Journal of Political Economy*, 127(6):2836–2874, 2019.

J. Gallagher. Learning about an infrequent event: evidence from flood insurance take-up in the united states. *American Economic Journal: Applied Economics*, pages 206–233, 2014.

S. Giglio, M. Maggiori, J. Stroebel, and A. Weber. Climate change and long-run discount rates: Evidence from real estate. Technical report, National Bureau of Economic Research, 2015.

P. A. Gompers and Y. Xuan. Bridge building in venture capital-backed acquisitions. In *AFA 2009 San Francisco Meetings Paper*, 2009.

P. A. Gompers, W. Gornall, S. N. Kaplan, and I. A. Strebulaev. How do venture capitalists make decisions? *Journal of Financial Economics*, 135(1):169–190, 2020.

J. González-Uribe. Exchanges of innovation resources inside venture capital portfolios. *Journal of Financial Economics*, 135(1):144–168, 2020.

T. C. Green, R. Huang, Q. Wen, and D. Zhou. Wisdom of the employee crowd: Employer reviews and stock returns. Technical report, Working paper, 2017.

B. C. Greenwald. Adverse selection in the labour market. *The Review of Economic Studies*, 53 (3):325–347, 1986.

R. Greenwood and S. Nagel. Inexperienced investors and bubbles. *Journal of Financial Economics*, 93(2):239–258, 2009.

S. J. Grossman and O. D. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719, 1986.

I. Hacamo and K. Kleiner. Competing for talent: Firms, managers and social networks. *Kelley School of Business Research Paper*, (17-34), 2017.

B. H. Hall and J. Lerner. The financing of r&d and innovation. In *Handbook of the Economics of Innovation*, volume 1, pages 609–639. Elsevier, 2010.

J. Haltiwanger, R. S. Jarmin, and J. Miranda. Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361, 2013.

B. H. Hamilton. Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political economy*, 108(3):604–631, 2000.

R. S. Harris, T. Jenkinson, and S. N. Kaplan. Private equity performance: What do we know? *The Journal of Finance*, 69(5):1851–1882, 2014a.

R. S. Harris, T. Jenkinson, S. N. Kaplan, and R. Stucke. Has persistence persisted in private equity? evidence from buyout and venture capital funds. 2014b.

H. Harry. Informal job search and black youth unemployment. *American Economic Review*, 77 (3):446–52, 1987.

O. Hart and J. Moore. Property rights and the nature of the firm. *Journal of political economy*, 98(6):1119–1158, 1990.

J. S. Hastings and J. M. Shapiro. Fungibility and consumer choice: Evidence from commodity price shocks. *The quarterly journal of economics*, 128(4):1449–1498, 2013.

T. L. Hayes and D. A. Neal. Actuarial rate review: In support of the recommended october 1, 2011, rate and rule changes. *Washington, DC: Federal Emergency Management Agency*, 2011.

T. Hellmann and M. Puri. Venture capital and the professionalization of start-up firms: Empirical evidence. *The journal of finance*, 57(1):169–197, 2002.

D. Hirshleifer. Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, 75(4):1779–1831, 2020.

H. J. Holzer. Search method use by unemployed youth. *Journal of Labor Economics*, 6(1):1–20, 1988.

H. Hong and J. C. Stein. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184, 1999.

D. P. Horn and B. Webel. Private flood insurance and the national flood insurance program. *Congressional Research Service*, 2019.

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

P. D. Howe, M. Mildenberger, J. R. Marlon, and A. Leiserowitz. Geographic variation in opinions on climate change at state and local scales in the usa. *Nature Climate Change*, 5 (6):596–603, 2015.

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

J. Hughes, J. Liu, and W. Su. On the relation between predictable market returns and predictable analyst forecast errors. *Review of Accounting Studies*, 13(2-3):266–291, 2008.

Y. M. Ioannides and L. Datcher Loury. Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093, 2004.

P. Issler, R. Stanton, C. Vergara-Alert, and N. Wallace. Mortgage markets with climate-change risk: Evidence from wildfires in california. *Available at SSRN 3511843*, 2019.

S. N. Kaplan and J. Lerner. Venture capital data: Opportunities and challenges. Technical report, National Bureau of Economic Research, 2016.

S. N. Kaplan and A. Schoar. Private equity performance: Returns, persistence, and capital flows. *The journal of finance*, 60(4):1791–1823, 2005.

S. N. Kaplan, B. A. Sensoy, and P. 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.

D. Karlan, R. Osei, I. Osei-Akoto, and C. Udry. Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652, 2014.

J. M. Keenan, T. Hill, and A. Gumber. Climate gentrification: from theory to empiricism in miami-dade county, florida. *Environmental Research Letters*, 13(5):054001, 2018.

S. Kortum and J. Lerner. Assessing the contribution of venture capital to innovation. *RAND journal of Economics*, pages 674–692, 2000.

B. Kőszegi and A. Szeidl. A model of focusing in economic choice. *The Quarterly journal of economics*, 128(1):53–104, 2013.

C. Kousky, H. Kunreuther, B. Lingle, and L. Shabman. The emerging private residential flood insurance market in the united states. *Wharton Risk Management and Decision Processes Center*, 2018.

T. Kuchler and J. Stroebel. Social finance. Technical report, National Bureau of Economic Research, 2020.

T. Kuchler and B. Zafar. Personal experiences and expectations about aggregate outcomes. *The Journal of Finance*, 74(5):2491–2542, 2019.

N. Lacetera, D. G. Pope, and J. R. Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–36, 2012.

C. Leuz, D. Nanda, and P. D. Wysocki. Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, 69(3):505–527, 2003.

Y. Li, E. J. Johnson, and L. Zaval. Local warming: Daily temperature change influences belief in global warming. *Psychological science*, 22(4):454–459, 2011.

L. Lindsey. Blurring firm boundaries: The role of venture capital in strategic alliances. *The Journal of Finance*, 63(3):1137–1168, 2008.

J. Lombard. Don't update that linkedin profile just yet—here's when and why you might want to hold off. *Media Bistro*, December, 09 2016.

U. Malmendier and Y. H. Lee. The bidder's curse. *American Economic Review*, 101(2):749–87, 2011.

U. Malmendier and S. Nagel. Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416, 2011.

U. Malmendier and S. Nagel. Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87, 2016.

U. Malmendier, G. Tate, and J. Yan. Overconfidence and early-life experiences: the effect of managerial traits on corporate financial policies. *The Journal of finance*, 66(5):1687–1733, 2011.

G. Manso. Experimentation and the returns to entrepreneurship. *The Review of Financial Studies*, 29(9):2319–2340, 2016.

D. T. Mortensen. The matching process as a noncooperative bargaining game. In *The economics of information and uncertainty*, pages 233–258. University of Chicago Press, 1982.

D. T. Mortensen. Job search and labor market analysis. *Handbook of Labor Economics*, 2: 849–919, 1986.

T. J. Moskowitz and A. Vissing-Jørgensen. The returns to entrepreneurial investment: A private equity premium puzzle? *American Economic Review*, 92(4):745–778, 2002.

J. Murfin and M. Spiegel. Is the risk of sea level rise capitalized in residential real estate? *The Review of Financial Studies*, 33(3):1217–1255, 2020.

S. C. Myers and N. S. Majluf. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2):187–221, 1984.

W. K. Newey and K. D. West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787, 1987.

J. A. Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pages 109–131, 1980.

A. Ouazad and M. E. Kahn. Mortgage finance in the face of rising climate risk. Technical report, National Bureau of Economic Research, 2019.

L. Pástor and R. F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685, 2003.

L. Phalippou and O. Gottschalg. The performance of private equity funds. *The Review of Financial Studies*, 22(4):1747–1776, 2009.

C. A. Pissarides. Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *The American Economic Review*, 75(4):676–690, 1985.

M. Puri and R. Zarutskie. On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms. *The Journal of Finance*, 67(6):2247–2293, 2012.

R. G. Rajan. Presidential address: The corporation in finance. *The Journal of Finance*, 67(4): 1173–1217, 2012.

R. G. Rajan and L. Zingales. The firm as a dedicated hierarchy: A theory of the origins and growth of firms. *The Quarterly Journal of Economics*, 116(3):805–851, 2001.

D. Robinson and B. Sensoy. Manager compensation, ownership, and the cash flow performance of private equity funds. *Unpublished working paper*, 2011.

L. Ryan. How long should i wait before adding my new job to my linkedin profile? *Forbes*, November, 20 2016.

S. Samila and O. Sorenson. Venture capital, entrepreneurship, and economic growth. *The Review of Economics and Statistics*, 93(1):338–349, 2011.

A. Schoar and L. Zuo. Shaped by booms and busts: How the economy impacts ceo careers and management styles. *The Review of Financial Studies*, 30(5):1425–1456, 2017.

J. Sheng. Asset pricing in the information age: Employee expectations and stock returns. *Available at SSRN 3321275*, 2019.

J. Shuey. How soon should you update your linkedin profile. *Personal Branding Blog*, February, 24 2017.

E. C. So. A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics*, 108(3):615–640, 2013.

V. Stango and J. Zinman. Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *The Review of Financial Studies*, 27(4):990–1030, 2014.

S. E. Taylor and S. C. Thompson. Stalking the elusive" vividness" effect. *Psychological review*, 89(2):155, 1982.

The Federal Emergency Management Agency (FEMA). OpenFEMA Dataset: FIMA NFIP Redacted Policies - v1. URL http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm.

A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

M. Ueda and M. Hirukawa. Venture capital and industrial'innovation'. *Available at SSRN 1242693*, 2008.

K. Wagner. Adaptation and adverse selection in markets for natural disaster insurance. *Available at SSRN 3467329*, 2019.

M. Waldman. Job assignments, signalling, and efficiency. *The RAND Journal of Economics*, 15(2):255–267, 1984.

B. Wernerfelt. A resource-based view of the firm. *Strategic management journal*, 5(2):171–180, 1984.

L. Zaval, E. A. Keenan, E. J. Johnson, and E. U. Weber. How warm days increase belief in global warming. *Nature Climate Change*, 4(2):143–147, 2014.

Zillow, Inc. Zillow Home Value Index (ZHVI). URL https://www.zillow.com/research/data/.

L. Zingales. In search of new foundations. *The Journal of Finance*, 55(4):1623–1653, 2000.