Where do Friends Matter for Founders? The Role of Social Connections in U.S. Entrepreneurship*

Robert B. Fluegge† and Mike Bailey‡

March 2025

Abstract

How important are social connections in the decision to start a new business? We assemble a new dataset using de-identified social network data from 670,000 U.S. firm owners and 78 billion friendships on Facebook linked to data on the firms founded by users. The data demonstrate that firm owners interact disproportionately with other firm owners and that social connections to existing firm owners increase an individual's probability of starting a successful new firm. Estimates from a quasi-experimental design among new college students imply that going from the 10th to the 90th percentile of connections to firm owners increases a student's probability of founding a firm after graduation by 111% of the base rate. Students are particularly likely to start a firm in the same narrowly-defined industry as the firm owners they are connected to, and effects are stronger when the student and the firm owner are the same gender. Using a quasi-experimental design identified using the timing of moves across commuting zones, we then provide evidence that the strength of network spillovers varies over geographic areas and is positively correlated with local rates of entrepreneurship. To assess the economic significance of the spillovers and their geographic variation, we construct and estimate a model of firm entry using moments from the quasi-experimental designs. Simulated policy counterfactuals suggest that targeting a hypothetical business-creating intervention to high-impact areas increases firm entry by 110% relative to targeting low-entrepreneurship places, with the difference driven primarily by large spillovers in areas with high existing entry rates.

*We thank Raj Chetty, Ed Glaeser, Josh Lerner, Kunal Sangani, Jesse Shapiro, Andrei Shleifer, Scott Stern, Drew Johnston, and James Stratton, as well as participants at numerous seminars for helpful feedback. Outstanding research assistance was provided by Valentina Juricek and Jonathan Kang.

†Hoover Institution, email: rfluegge@stanford.edu

‡Meta, email: mcbailey@meta.com

I Introduction

Interactions among entrepreneurs, and the associated knowledge spillovers that may occur, have long been viewed as central to economic growth and agglomeration (e.g., Marshall 1890; Jacobs 1969). As a result, policymakers and private organizations implement a variety of interventions aimed at reducing barriers to entry for new businesses.[1] However, more recent attempts to uncover direct individual-level evidence of spillovers find that entrepreneurs have only a modest effect on successful firm formation among direct peers in the workplace or classroom (e.g., Lerner and Malmendier 2013). These results cast doubt on whether any intervention aimed at supporting new entrepreneurs can be an effective tool for spurring new economic activity beyond the individual targeted by the intervention.

In this paper, we show that knowledge spillovers among entrepreneurs are economically meaningful despite the seemingly small magnitudes estimated in typical quasi-experimental settings. To do so, we construct a new de-identified dataset from 133 million individuals between the ages of 25 and 59 and their connections, termed 'friends' on Facebook, linked to establishment-level data covering nearly all new businesses formed in the U.S. Firms are identified via their Facebook page and linked to data on their size and market characteristics. 670,000 firm owners are then linked to firms via self-reported ownership on their Facebook profile.

Using these data, we make three points. First, entrepreneurs produce substantial spillovers because of the *scale* of their networks. That is, connections to entrepreneurs are sufficiently numerous that even modest effects at the level of a single connection imply important forces at the level of the entire economy. Second, the extent to which individuals have access to valuable entrepreneurs through their networks varies substantially across individual characteristics and location. This variation suggests that some individuals who might otherwise start successful firms lack the social resources to do so. Third, we suggest that this variation has implications for policy by constructing an economic model of entry spillovers. Using the estimated model, we argue that

[1] Policies intended to promote entrepreneurship are motivated and described in (Josh Lerner 2009). Such policies are of particular recent interest given evidence of a long-run decline in U.S. business dynamism (Decker et al. 2014; Kwon et al. 2024). New firms also play a disproportionate role in the labor market (Haltiwanger et al. 2013), contribute substantially to productive innovation (e.g., Acs and Audretsch 1988; Akcigit, Caicedo, et al. 2018), and are intimately related to market concentration (Bresnahan and Reiss 1991, De Loecker et al. 2020)

the effect of a given intervention on firm entry is likely to depend strongly on the places and subgroups targeted, with most of the variation driven by spillovers.

In the first part of the paper, we document that the average individual has many connections to entrepreneurs and that individual-level connections to entrepreneurs are strongly associated with firm creation. The average individual has 6.76 friendships with business owners. Each friendship with a business owner is associated with a 0.1 percentage point increase in the probability that an individual owns a firm, a strong relationship relative to an overall entrepreneurship rate of roughly 1% in the data, and almost nobody with zero social connections to business owners starts a firm.[2] The number of connections to business owners is a stronger predictor of firm ownership than demographic characteristics, socioeconomic position, local rates of entrepreneurship, or characteristics of the local financial market. Entrepreneurs cluster particularly strongly within narrowly-defined industries: an entrepreneur with 2 friends in one of 1059 6-digit NAICS industries is 75 times more likely to own a business in that industry compared to the average entrepreneur.

However, the interpretation of these descriptive patterns is complicated by potential selection in friendship formation, making it difficult to draw policy conclusions without causal evidence. In the second part of the paper, we develop two quasi-experimental empirical designs in order to estimate the causal effect of social connections on individual-level founding behavior.

The first strategy, which we refer to as the "College design," estimates the effect of connections to existing owners on post-graduate founding behavior for individuals entering college. The strategy compares students who were born in the same year and attended the same college but were exposed to a different number of business owners among *friends of friends* (i.e., 2nd-degree connections) during the highly social first weeks after their arrival on campus. Unlike most connections to business owners, we show that 2nd-degree connections to business owners formed among college first-year students during the first weeks of college are unrelated to the number of business owners that the student already knew, suggesting that students with a higher propensity for entrepreneurship after college do not disproportionately select well-connected friends during this period.

[2]While many startups in the U.S. economy do not produce substantial economic activity (Robert W. Fairlie et al. 2023), we show in Section II that the firms identified by our linking procedure tend to be relatively successful. For firms which have linked revenue data, we also show that networks are an important determinant of in the formation of firms with greater than $1 million in annual revenue.

Under the College design, we estimate that each 2nd-degree connection between a student and a business owner increases the probability that the student starts a firm after graduation by 0.007 percentage points. The effect magnitude implies that going from the 10th percentile of 2nd-degree connections to business owners among first year college students to the 90th percentile, equivalent to a difference of 39 connections, increases the probability of founding a firm by 0.27 percentage points compared to a base entrepreneurship rate of roughly 0.25% among recent college graduates. This scaling demonstrates a key theme of the quasi-experimental results: modest spillover effects on entrepreneurship at the connection-level, such as those found in previous quasi-experimental studies, translate to economically meaningful aggregate social spillovers due to the size of individuals' extended networks. Furthermore, the firms created as a result of these college connections are not negatively selected; there is no evidence that the marginal and average firms have different probabilities of reaching $1 million in annual revenue.

Using the same quasi-experimental variation, we then show that the effect of connections is highly industry-specific. One extra connection to a firm owner in one of 1059 narrowly-defined industries classified at the 6-digit NAICS level increases the probability that an individual founds in that industry nearly 10 times the industry-specific base rate, but does not increase the probability of founding in closely related industries. Under the assumption that students in the first weeks of college do not select friends with connections in the precise industry where the student is likely to found a firm years later after graduation, the industry-specific results imply that social connections have a causal effect on the type of entrepreneurship that individuals pursue. Formally, the precise industry designations provide a set of overidentifying restrictions such that any selection effect which violates our identifying assumptions would need to covary with the quasi-experimental treament at the 6-digit industry level, narrowing the set of potential threats to identification.

The importance of social connections to firm owners suggests that potential entrepreneurs who lack such connections may face substantial barriers to successful firm entry. We provide evidence of meaningful individual-level variation in both connections to entrepreneurs and treatment effects by gender and socioeconomic status. Estimates from the College design imply that females are more likely to start a business if connected to a female firm owner but not a male firm owner.

However, the average woman has 13.4% fewer connections to firm owners of the same-gender compared to the average male. Similarly, using estimates of socioeconomic status developed in previous work (Chetty et al. 2022a), we show that the number of connections to entrepreneurs is strongly positively related to income. For example, an individual in the top 10% of the income distribution has 89% more connections to entrepreneurs than those below the 90th percentile of the income distribution. These differences manifest early in life: meaningful gaps in connections to entrepreneurs exist at age 18 for individuals who grow up in high-income v.s. low-income households.

Given the importance of spillovers in models of agglomeration, and the resulting popularity of policies intended to encourage local startups, differences in connections to entrepreneurs across geographic markets are of particular interest. We show that individuals living in commuting zones in where the average resident has an above-median rate of entrepreneurship in their network are connected to 47% more entrepreneurs than individuals in below-median commuting zones, a substantial difference relative to a gap of 57% in the rate of entrepreneurship. However, the College design does not have the statistical power to precisely estimate geographic heterogeneity in spillover effects, which our model will demonstrate can play an important role in policy design.

In order to identify areas where spillovers matter most, we implement a second quasi-experimental design which we call the "Movers design". The Movers design exploits the timing of a friend's migration across commuting zones to isolate quasi-experimental variation in the exposure of friends to entrepreneurial places.[3] To implement this approach, we use a series of stacked difference-in-differences (Sun and Abraham 2021) which estimate the effect of a friend's migration on founding behavior among non-movers separately for every pair of commuting zones, representing the non-mover's location and the moving friend's destination. While estimates at the level of commuting zone pairs are noisy, projecting pair-level estimates onto the mover's destination allows us to identify the characteristics of places where social connections have a high-impact on entrepreneurship.

The destination-specific effect is highest when the mover's destination has high rates of entrepreneurship, measured using the annual number of jobs per capita created by new businesses, and is largest when the mover is likely to form new connections to business owners based on pre-

[3]See Bell et al. 2019 and Moretti 2021 for analogous quasi-experimental variation in studies of patenting behavior.

move friending patterns. Movers who are least likely to make friends with entrepreneurs based on pre-move characteristics have no effect on their friends, and migration to a commuting zone which does not produce many successful new businesses has little effect on the friends of the mover. Intuitively, the patterns imply that a mover only has a large effect on their friends at home when the mover has both the opportunity to make friends with successful entrepreneurs and has a history of forming such friendships. These results hold even when considering only the creation of new firms with revenues in excess of $1 million annually as the outcome. We interpret the results as suggesting that the Movers design operates via the mover's friending behavior in their destination, which changes the composition of the non-mover's 2nd-degree connections.

In the third part of the paper, we construct an economic model in which potential entrepreneurs may learn productive skills from their social connections in order to assess the magnitude of the spillovers and provide guidance on the policy implications of our findings. We then illustrate the economic significance of the spillovers by using the model to analyze the total number of new firms generated by a hypothetical intervention aimed at reducing barriers to entry for entrepreneurs. In the appendix, we provide an extension of the model demonstrating that the entry effect contains information about the welfare impacts of an intervention.[4]

Under the model, the overall effect of the hypothetical intervention on firm entry in equilibrium can be additively decomposed into (1) a direct effect on entry for the individual targeted by the intervention, and (2) a spillover effect that scales the target individual's effect on each of their connections by the target's centrality in the network. The relevant notion of centrality in the model, known as Katz-Bonacic centrality (Katz 1953), captures the scale of the target individual's spillover effect on both direct connections and indirect connections throughout the networked economy. Notably, the additive separability between direct and spillover effects in the model's solution suggests a potential trade-off between the two when targeting the hypothetical intervention.

Our estimation strategy combines estimates of mean centrality in the cross section with scaled moments from the quasi-experimental designs to separately estimate the direct and spillover effects.

[4]We show that the entry effect is proportional to the welfare effect when spillovers only reduce fixed costs. This welfare result holds even when the equilibrium number of firms is different from the social optimum.

Estimating each component separately allows us to empirically evaluate the potential trade-off between direct effects and spillovers but requires directly estimating Katz-Bonacic centrality, which is difficult to calculate analytically in large networks. To perform the estimation, we introduce a technical innovation in the form of a recently-developed sampling estimator from the network science literature (Lin et al. 2021). Given the relevance of Katz-Bonacic centrality to many linear peer effects models, we additionally report economy-wide average centrality estimates for a range of parameter values which we hope will be useful in future research.

In our first counterfactual exercise, we separately analyze the direct and "total effect" on firm entry, including spillovers, from a hypothetical intervention targeted to a random individual in each commuting zone. We start by comparing the relationship between each effect and existing entrepreneurship rates at the commuting zone level. The direct effect is negatively correlated with rates of entrepreneurship; the hypothetical intervention produces 27% fewer firms among target individuals who live in commuting zones with above-median entrepreneurship rates compared to places below the median. However, when combining our estimates of the direct effect and social spillovers to estimate the total effect, we find that relationship reverses: the total effect is *higher* when the policy is targeted to places with high rates of entrepreneurship. The shift when accounting for spillovers occurs because residents of high-entrepreneurship cities tend to produce larger spillovers than residents of cities with little entrepreneurship, and demonstrates that the trade-off between direct and total effects suggested by the model has empirical relevance.

We then simulate the effect of the hypothetical policy administered in commuting zones with an above-median total effect. Our estimated model implies that targeting an average person in high-effect commuting zones increases firm entry by 110% relative to the same intervention in a commuting zones with low rates of entrepreneurship. The improvement largely comes from targeting places where spillovers are highest: targeting a commuting zone with above median-direct effects only increases firm entry by 33% relative to places with below median entrepreneurship rates.

However, place-specific estimates alone do not provide guidance on how network characteristics beyond location might be used to target policies effectively. To illustrate the potential gains from individual-level policy targeting, we extend the analysis to consider how individual-level

characteristics of prospective entrepreneurs and their connections affect the direct and spillover effects of policies. First, we demonstrate that network composition can be used to identify the characteristics of high-impact targets, who are more likely to start a business when given a valuable connection. Our analysis shows that individuals with many existing connections to entrepreneurs are particularly likely to be on the margin of starting a new firm, including individuals who have many entrepreneur friends despite living in low-entrepreneurship places. Under the model, the magnitudes imply that a policy targeted to a highly connected person in a low-entrepreneurship commuting zone is expected to have an 24% higher direct effect than a policy targeted at the average person in the same place.

We conclude the counterfactual analysis by considering policies intended to reduce barriers to entrepreneurship for under-represented subgroups by simulating the effects of a hypothetical policy aimed at promoting female entrepreneurship. To do so, we provide an extension of the model that combines place- and subgroup-specific moments from each quasi-experimental design. Under the extended model, a policy targeted at high-impact women in commuting zones with the highest total effect creates 301% more female-owned businesses than a policy targeted to high-impact women in commuting zones with low rates of entrepreneurship.

Contributions Relative to Existing Literature The key innovation of this paper is the use of large-scale social network data to provide quasi-experimental evidence on both the overall magnitude of social spillovers from individual-level entrepreneurship and on policy-relevant heterogeneity. In doing so, we expand on existing work which uses quasi-experimental evidence to estimate spillovers in entrepreneurship among a subset of peers, typically in classrooms and workplaces (Nanda and Sørensen 2010, Josh Lerner and Malmendier 2013). The social spillovers we estimate supplement existing indirect evidence that interactions play a role in entrepreneurship based on the observed clustering of entrepreneurs across, spatial, industrial, and demographic lines.[5] The heterogeneity in network exposure to entrepreneurs by demographic subgroup and childhood income in our data demonstrates that the heterogeneous treatment effects estimated in workplaces and classrooms

[5] e.g., Ellison et al. 2010; Glaeser, W. R. Kerr, et al. 2010; A. Chatterji et al. 2014; Guzman and Stern 2015 ; R. Andrews et al. 2022; W. R. Kerr and Mandorff 2023; Mohammadi and Shafi 2024

reflect important patterns in broader social interactions (Wallskog 2023, Mertz et al. 2024).[6] The results also expand our understanding of the preferences, characteristics, experiences, and opportunities distinct to entrepreneurs (Chetty et al. 2025).[7] Beyond personal characteristics, a vibrant literature has catalogued the importance of external inputs to successful entrepreneurship, with particular focus on capital availability and capital market mechanisms.[8]

The evidence that entrepreneurship produces broad social spillovers, even among firms unlikely to produce technical innovations, helps to clarify the nature of nonrival knowledge central to many models of economic growth.[9] The evidence is particularly relevant to models of creative destruction in which entry and competition can play a key role (Howitt and Aghion 1998; Akcigit and Van Reenen 2023) and to theories that formally distinguish the role of entrepreneurs from technical innovators (e.g., Acs, Braunerhjelm, et al. 2009; M. J. Andrews et al. 2022). We interpret our results as evidence that non-technical knowledge of business opportunities or practices among entrepreneurs and managers is likely to produce economically meaningful knowledge spillovers.[10] Our evidence that entrepreneurship spillovers are intimately tied to spatial variation in economic activity bridges models of regional growth based on entrepreneurship (Duranton and Puga 2023, Fluegge 2024) with empirical evidence that interpersonal interactions vary substantially across place, particularly for under-represented subgroups (Athey et al. 2021). [11]. Moreover, our result

[6]Barriers to entrepreneurship for subgroups has long been a subject of study in economics and finance (e.g., Bradford and Osborne 1976; R. Fairlie et al. 2022; Howell and Nanda 2024). These strands of inquiry supplement an older literature in economic sociology, with origins in work studying the importance of weak ties and structural gaps (Granovetter 1973; Burt 1992), which suggests that founders' social networks are a key determinant of their location decisions (Sorenson and Audia 2000) and success (RS Burt and Raider 2002; Stuart and Sorenson 2005; Bhide 2000).

[7]Further evidence is provided in Kihlstrom and Laffont 1979; Hamilton 2000; Hurst and Pugsley 2011; Lazear 2005; Levine and Rubinstein 2017, Azoulay et al. 2020; Bhandari, Kass, et al. 2024; Chodavadia et al. 2024). The importance of social networks to entrepreneurs echos a long literature on the importance of knowledge, experience, and good business practices among entrepreneurs and managers for firm productivity (e.g., Jovanovic 1982, Jovanovic and Rob 1989; Bloom and Van Reenen 2007; Smith et al. 2019, Bhandari and McGrattan 2021, Denes et al. 2025). While the current paper focuses on effects along the extensive margin of firm creation, our measurement of network-based spillovers also suggests that existing microeconomic evidence on productivity spillovers between firms and firm owners in developed economies may run in part through social networks (e.g., Greenstone et al. 2010; Cai and Szeidl 2018; Guiso et al. 2021).

[8]Key papers in the financing of new ventures include Evans and Jovanovic 1989; P. A. Gompers and Josh Lerner 2002, P. Gompers et al. 2010; Golosov et al. 2024, Chodorow-Reich et al. 2024; S. P. Kerr et al. 2022; Davis and Haltiwanger 2024). Other work has examined the role of government policy (Howell 2017) and workplace circumstances (Babina, Fedyk, et al. 2024).

[9]Leading examples include Romer 1990, Aghion and Howitt 1992, Grossman and Helpman 1992, Lucas 1988, Klette and Kortum 2004

[10]We view this evidence as an important finding in light of recent macroeconomic theories that knowledge spillovers from entry may be small or negative (Klenow and Li 2024). Recent macroeconomic research has also suggested that falling productivity spillovers and the changing role of entrepreneurs in market structures may explain a host of economic phenomena in the U.S. (Akcigit and Van Reenen 2023; Deb 2024). In this sense, our study parallels a long literature finding evidence for spillovers in technical innovation and patenting (e.g., Jaffe et al. 1993; Audretsch and Feldman 1996; Ellison et al. 2010; Akcigit, Caicedo, et al. 2018; Iaria et al. 2018; Bell et al. 2019; Akcigit, Hanley, et al. 2021; Moretti 2021; Babina and Howell 2024; Atkin et al. 2022; Kalyani et al. 2025)

[11]Our results provide concrete individual-level data to support a long history of research across the social sciences suggesting that local interactions and entrepreneurship are closely related (Chinitz 1961; Jacobs 1969; Saxenian 1996; Stuart and Sorenson 2003; Delgado et al. 2010; Glaeser, S. P. Kerr, et al. 2015)

that spillover effects operate strongly within narrow industries informs influential models of the relationship between the allocation of entrepreneurial talent and economic growth (Baumol 1990; Murphy et al. 1991). While we do not provide evidence on other (potentially negative) spillovers from entry, our benchmark welfare results demonstrate that the welfare effects of a marginal intervention can be positive even in the presence of business-stealing externalities.[12][13]

Our counterfactual policy simulations are intended to inform the evaluation and implementation of entrepreneurship-focused economic development programs. Interventions evaluated by this literature include business education or training in the U.S. (e.g., Kosanovich et al. 2001; Robert W Fairlie et al. 2015) and abroad (Field, Jayachandran, and Pande 2010; Oosterbeek et al. 2010; Huber et al. 2014;), contract allocation (Lerner 1999; A. K. Chatterji et al. 2014), grant programs (Babina, He, et al. 2023), and the establishment of local business development centers (Chrisman 2017).[14] Beyond interventions specific to entrepreneurship, our study relates to the growing literature on policy targeting in networked economies.[15] Our technical innovation in this regard is the introduction of recently developed tools from network science, which allow for the estimation of a class of theoretically-motivated centrality measures in the very large networks relevant to our setting.

Finally, the results contribute to a fast-growing literature on the importance of social networks for economic phenomena (Bailey, Rachel Cao, et al. 2018). Social networks, including those represented on the Facebook platform, have been shown to play a role in household financial decisions (Bailey, Ruiqing Cao, et al. 2018), economic opportunity (Raj Chetty et al. 2022a Raj Chetty et al. 2022b), cultural assimilation (Bailey, D. Johnston, Koenen, et al. 2024), migration decision (Koenen and D. Johnston 2024), and may play a role distinct role in firm entry as consumers learn

[12] In this sense, our work provides evidence on a subset of the externalities relevant to a developing literature on optimal taxation of high-earning entrepreneurs (Lockwood et al. 2017, Jakobsen et al. 2024, Kleven 2025)

[13] Other types of positive spillovers have been hypothesized to operate outside of social networks. For example, exposure to new firms may help important non-entrepreneurs like investors and employees learn more about the needs of entrepreneurial ventures (Josh Lerner 2009). Furthermore, despite survey evidence described in SectionII that the Facebook platform does capture a variety of 'real world' interactions, spillovers may occur through interpersonal interactions that are not represented on the Facebook platform.

[14] Entrepreneurship-focused interventions are frequently targeted directly at individual entrepreneurs within well-defined geographic markets: of the 16 public small business assistance programs studied in a review (Gu et al. 2008), 10 list the entrepreneur's personal characteristics in their eligibility criteria and 8 are explicitly locally targeted or administered. Place-based development policies in the U.S., including tax incentives, cost between $22 billion and $101 billion annually (Bartik 2020).

[15] See Elliott and Golub 2019; Galeotti et al. 2020; Boucher et al. 2024 for recent theoretical work on targeting in networks. The usefulness of network data in targeting interventions have long been recognized in the economic development literature, including in experiments aimed at promoting the formation of micro-enterprises (Miguel and Kremer 2004; Banerjee et al. 2013; Field, Jayachandran, Pande, and Rigol 2016).

about products (Bailey, D. Johnston, Kuchler, et al. 2022, Campbell et al. 2024).[16] A substantial body of work has further demonstrated that social networks, and peer effects more generally, likely play an important role in educational settings (e.g., Calvó-Armengol et al. 2009; Sacerdote 2011). We interpret these studies, particularly those on the economics of education, as bolstering the view that social connections are important for the transmission of human capital.

Layout The paper proceeds as follows. Section II describes the data and presents key descriptive relationships. Section III develops the quasi-experimental designs. Section IV specifies the model. Section V estimates the model and simulates policy counterfactuals. Section VI concludes.

II Data and Descriptive Evidence

II.A Data Collection and Matching

II.A.1 Sample Construction

Our primary analysis sample, drawn on February 14, 2025, includes de-identified information on individuals residing in the U.S. aged 30-49 who were active on the Facebook platform at least once in the previous 30 days, had at least 30 U.S.-based Facebook friends, and for whom we measure a home ZIP code. The primary sample consists of 37,661,123 users, equal to 81% of the population in that age range in the 2015-2019 American Community Survey (ACS). The age range was chosen because previous survey evidence indicates that individuals between 30 and 49 use Facebook at the highest rates (Pew Research Center 2024). This range also covers the ages that many entrepreneurs found a firm for the first time (Azoulay et al. 2020).

Social ties are measured using 'friendships', the on-platform term for links between users. Both individuals must agree to form an on-platform friendship, and most Facebook friends have interacted in person (Jones et al. 2013). We therefore interpret an on-platform friendship as representing an off-platform friendship or acquaintance.

To create a dataset which combines information on founders with firm-level data, we link records on these users and their friends with data on currently operating firms. Linked firms appear to be

[16] See Mobius and Rosenblat 2014 for a related overview on social learning

relatively successful; 16% of linked firms have greater than $1 million in annual revenue, and more than 93% of firms have survived for more than 4 years. Appendix Figure E5 presents a histogram of the number of owners by the year that they first started their firm.

Below, we describe the basic linking process and resulting sample characteristics. A schematic summarizing the linkage between each source is presented in Appendix Figure E1. Detailed information on the measurement, linkages, and sample inclusion criteria are provided in Appendix Section A.1.

II.A.2 Establishment-Level Data: Customer Graph

The first source, termed the Customer Graph, consists of 8 distinct datasets compiled from 3rd-party data providers which measure founding date and industry at the establishment-level for U.S. firms. Establishments are matched across these 8 data sources to create a dataset with one record per establishment.

The yellow series in Appendix Figure E2 plots the number of new establishments in the Customer Graph by year between 2011 and 2022. There is a large spike between 2015 and 2016, driven entirely by a jump in the number of observations from the largest data provider in that period. We benchmark these data against an estimate of the number of firms registered in each year until 2016, presented in the green series. Estimates were compiled using data from 46 state-level registration datasets, adjusted to account for firms that were registered in Delaware but operate elsewhere (R. Andrews et al. 2022). Since the registration data only include registered firms and omit proprietorships, we present a simple adjustment in the pink series which scales the number of registered businesses by the ratio of all businesses to registered businesses in a series of surveys of businesses with websites administered by GoDaddy (*Venture Forward* 2023). The results of the benchmarking indicate that the number of new establishments in the Customer Graph track the number of registrations in each year until 2016, at which time the private establishments data spikes to the estimated number of all firms that enter including proprietorships. Appendix Figure E3 presents a scatterplot of number of new Customer Graph firms per capita vs the number of new registered firms per capita at the county level, which have a population-weighted correlation

of 0.69 in 2016.

II.A.3 Linking Establishment Data to Facebook Pages

Establishment-level records from the Customer Graph are matched to Facebook Pages using the name, location, and contact information of the establishment. Pages that are likely to be branches of a larger organization based on their posting behavior are omitted. After matching, there are 3.5 million matched establishments that can be identified in both the Customer Graph data and as Facebook pages. Appendix Figure E4 presents a sample of matched pages. Each page in the figure is actively posting on the platform and has hundreds of 'followers' who receive notifications about the establishment's on-platform activity. The sample of pages also provides concrete examples of the types of firms likely to be matched under this procedure: firms tend to be small- or medium-sized, include both consumer and business-facing industries, and sometimes fall into traditional 'high-human capital' sectors like business consulting. Appendix Figure E6 presents match rates by sector between the Pages and Customer Graph data.

The relatively low rate at which establishments in the Customer Graph data can be matched to Facebook pages appears to be due to the fact that many establishments in the Customer Graph are not economically meaningful entities. The red series in Appendix Figure E2 indicates the number of new employer firms in each year until 2020 from the Business Dynamics Statistics. The number of new employer firms in each year is substantially smaller than the number of registrations. Mirroring this pattern, random samples from the Customer Graph data indicate that a large proportion of firms in those data are registered to a residential address and have no online presence. Figure 1a compares the number of matched pages in the Facebook data with the number of new employer firms between 2015 and 2020 in the Business Dynamics Statistics for the 9 largest 2-digit NAICS sectors. Figure 1b compares the number of new matched pages per capita vs the number of new employer firms per capita in 2016 at the county level. The two have a population-weighted correlation of 0.72. These comparisons suggests that the 3.5 million pages that can be matched to the Customer Graph data are disproportionately economically meaningful entities, and are fairly representative of the set of new employer business across both industries and local markets.

II.A.4 Identifying Firm Owners

To match establishments to founders, we use the self-reported 'employer' and 'job title' fields on users' public Facebook profiles. We identify owners of a business as users who list a page as their employer and who lists a job title that includes the terms 'founder', 'owner', or 'CEO'.[17] We omit pages that have more than 4 individuals who have ever claimed to own the business associated with the page, and provide robustness checks which verify our baseline estimates hold when only considering firms that have exactly one owner.

This matching procedure identifies about 695,000 firm founders, including 299,000 who founded at least one firm since 2015, out of the analysis sample of 73 million U.S. users who list an employer and job title. This dataset allows us to identify the start date and industry (from Customer Graph) as well as the social network of the founder. Match rates between verified matched pages and owners by 2-digit NAICS sector are presented in Appendix Figure E6. Table 1 presents sample statistics for the analysis sample. Age and gender distributions of the firm owners we identify are compared to data from the Annual Business Survey in Figure E7, which shows that our sample is younger and more female than the owners of employer businesses covered by the survey.

II.B Descriptive Facts about Entrepreneur Networks

II.B.1 Entrepreneurs are Friends with Entrepreneurs

Figure 2 presents the distribution of the number of connections, termed 'friends' on the platform, who own a business separately for people who own vs. do not own a business themselves. The distributions exhibit evidence of substantial 'homophily', the tendency for similar individuals to interact, among entrepreneurs. Nearly 25% of non-owners know zero owners, while less than 3% of owners have no friends who are owners. Figure 3a shows that an additional 10 friends who own a business is associated with an increase in the probability that an individual owns a business from roughly 0 to 0.015, a large difference given that roughly 1% of the sample owns a business. This

[17] We include 'owner' and 'CEO' in the fuzzy string match because random samples of firm pages indicate that many users who are clearly the founder list themselves as one of those two titles. For example, page-profile pairs which share the same name often have a job title listed as 'owner' or 'CEO'. Appendix Figure E10 demonstrates that the key descriptive relationships hold when restricting the fuzzy string match to include only users with "Founder" in their job title.

pattern is primarily driven by differences in the composition of friends rather than the total number of friends: Appendix Figure E9 demonstrates that the fraction of friends who own a business is highly predictive of one's own probability of starting a firm.

Figure 3b shows that the number of friends who own a business is highly predictive of stricter definitions of business success, not just founding behavior, as the probability that an individual owns a firm with greater than $1 million in self-reported annual revenue is also strongly increasing in the owner's entrepreneur friends. Analogous plots for firms with greater than $10 in annual revenue and greater than 50 employees are presented in Appendix Figure E10.

II.B.2 The Predictive Power of Friendships vs Other Covariates

How does the predictive power of relationships with business owners compare to other predictors? Figure 4 demonstrates that, while other demographic and market-specific factors contain some information, the number of friends in an individual's network is the most predictive individual-level characteristic. Appendix Figure E11 demonstrates that the relationship also holds across time: recent founding behavior is predicted by friendships that started prior to founding. However, a substantial portion of owner friends can be explained by behavior around the time an individual founds their firm. Appendix Figure E12 shows that individuals make friends with owners at the highest rate in the year when the individual founds their firm, and Appendix Figure E13 shows that the relationship holds for friendship with more experienced owners that founded prior to 2010.

II.B.3 Industry-Specific Friendships

Friendships with other owners are particularly likely within narrowly-defined industries. In the roughly 75% of the sample for which we measure industry at the 6-digit NAICS level, which distinguishes 1059 distinct industries, Figure 5 demonstrates that having just two friends in an industry increases the relative probability that a given owner works in that industry by 75 times. The same is not true for other industries in the same industry group (4-digit NAICS) or sector (2-digit NAICS). Appendix Figure E14 presents an illustrative example of this phenomenon: the number of friends who own a bike shop is highly predictive of whether that individual owns a

bike shop themselves. The number of friends who own a sports wholesale business, and to a lesser extent the number who own a bike repair business, are also predictive of bike shop ownership, while the number of friends who own an electronics shop is not predictive. Appendix Figure E15 presents the distribution of friends in the same industry, industry group, and sector among owners.

II.B.4 Connections to Owners by Income and Gender

To what extent do friendships with owners, and access to the social resources such connections may entail, differ by socioeconomic status and gender?

Figure 6 demonstrates that the average number of friendships with firm owners increases dramatically with income. The pattern is particularly stark at the top of the income distribution: the average individual above the 90th percentile of the income distribution has 89% more friendships with firm owners than those below the 90th percentile.

Gender differences in connections to entrepreneurs require a more nuanced discussion. Quasi-experimental evidence in Section V.C.2 will indicate that connections to entrepreneurs have the largest causal effect when the individual and the connected entrepreneur are the same gender. For example, aspiring female entrepreneurs benefit most from connections to female business owners. This pattern of gender concordance which in our data mirrors evidence from existing quasi-experimental studies in workplaces and classrooms (Wallskog 2023, Mertz et al. 2024).

Accordingly, Table 2 provides statistics on the distribution of connections to owners separately by the gender of the individual and the business owner. The statistics indicate that men have 14.42% more friendships with business owners of the same gender than women have with the same gender. In this sense, potential male entrepreneurs appear to derive a greater total benefit from their social networks despite men and women having similar total connections to business owners.

II.B.5 Geographic Variation in Owner Networking

The extent to which residents of an area makes friends with entrepreneurs varies across geographic regions. Figure 7 plots a county-level map of the average fraction of a person's friends that own a business. Roughly 1% of friends in a county in the bottom 33 percent of (population weighted)

owner networking are entrepreneurs, compared to 2% in a county above the 66th percentile.[18]

While differences in owner networking across places are related to the rate of local business ownership, they also reflects differences in rates of friendship formation among the general population. Appendix Figure E16, which shows a county-level scatterplot of the average fraction of a person's friends who own a business vs. the in-sample entrepreneurship rate, demonstrates substantial unexplained variation in friendships with entrepreneurs. Appendix Table D3 provides data on the fraction of entrepreneurs among residents and the fraction of entrepreneurs among friends of residents in our data for the 30 largest commuting zones.

III Quasi-Experimental Designs and Evidence

Friendships between business owners may reflect causal effects of friendships or nonrandom network formation. This section develops two quasi-experimental designs to uncover the causal effect of connections to entrepreneurs on business formation. The first is based on friendships formed in college, and the second exploits the migration of friends to high-entrepreneurship commuting zones.

III.A College Friends

We first provide direct evidence that friendships between entrepreneurs have a causal effect on business formation by exploiting plausibly exogenous friending behavior in the first semester of college. We start by showing that friendships which provide indirect connections to business owners in the highly social first weeks of college are unrelated to the number of business owners that were already in an individual's network, which we interpret as evidence that unobserved personal characteristics do not drive the results.

Figure 8 shows the mean number of new friends made by week in the year that an individual turns 18 separately for college attendees and non-attendees. The shaded region highlights the period from mid-August to Thanksgiving, which corresponds to the period during which a school on the semester system typically holds first-year orientation activities and classes. The group of college attendees makes substantially more friends in this period than either non-attendees during

[18] The implied average number of friends who are entrepreneurs differs from the entrepreneurship rate in the data because entrepreneurs have more total friends than non-entrepreneurs.

the same period or the same group of attendees earlier in the year. This provides an identifiable group of friends who are plausibly different from the standard friending composition of college attendees.

III.A.1 Design

For our quasi-experimental treatment we use characteristics of the 2nd-degree connections (i.e., friends of friends) formed during the first 12 weeks of the academic year, since direct friendships are almost entirely formed among other students starting college and who are unlikely to own businesses themselves. The individual-level treatment for person i will be defined as the number of new second degree connections j who satisfy the following conditions:

1. i and j were not friends and had no mutual friends prior to the year that individual i turned 18.

2. During weeks 33-44 (mid-August to Thanksgiving) of the year i turned 18, i becomes friends with one of j's friends, whom j had known since before the year i turned 18.

3. j had started a business prior to the year that i turned 18.

These conditions ensure that only i's friending behavior are driving the results: j's friending behavior and founding activity are held constant.

Define T_i to be the number of individuals j who satisfy these conditions wrt person i. The distribution of T_i for collegegoers in the sample is presented in Appendix Figure E17.

The identifying assumption is that T_i is not systematically related to person i's underlying propensity to start a business, which is plausible if person i and j have little information about each other at the time they form the friendship. This might be the case if, for example, people make friends with randomly assigned roommates (e.g., Sacerdote 2001), neighbors, or members of the same orientation groups. Section III.A.3 presents evidence that new 2nd-degree connections during this period are unrelated to student characteristics.

III.A.2 Baseline Estimates

Based on the linearity in levels exhibited in Figure 3a, we begin with a linear probability model using the following estimating equation:

$$Y_i = \tau_{sc} + \beta T_i + \gamma X_i + \epsilon_i \tag{1}$$

for person i who attended school s and was born in year c, and τ a school x birth-cohort fixed effect. People that did not attend college are omitted from the estimation. We set Y_i to be 1 if person i started a firm 4 or more years after turning 18, and zero otherwise. The vector X_i includes lagged values of T_i for each 4-week period in the same year that i turns 18. For example, the first element of X_i is equal to the number of new 2nd-degree connections to owners made during the first 4 weeks of the year i turned 18.

OLS estimates of β in are presented in Table 3. The magnitude implies that one extra 2nd-degree connection to a business owner increases the probability that an individual starts a business after graduation by 7.9 x $10^{-}5$ percentage points. This connection-level effect is relatively modest, and implies that many connections would be required to have an economically meaningful impact on founding behavior. However, scaling the effect by the standard deviation of the number of 2nd-degree connections to business owners among all first-year college students, roughly 2.0 x $10^{-}5$ 2nd-degree connections, implies an effect of 3.95 percentage points per standard deviation, a meaningful effect relative to a base entrepreneurship rate of 3.19 in this sample. The results are not driven by co-founding behavior, as Column (4) exhibit similar patterns when the outcome is defined to be starting a firm without co-founders. Appendix Table D5 demonstrates that the findings are robust to various control schemes and outcome definitions.

Success of Marginal Firms Table D5 repeats the baseline regression models when the outcome is redefined as starting a firm with greater than $1 million in annual revenue. We find similar effects as a fraction of the base rate. This similarity in relative effect sizes suggests that marginal entrepreneurs induced to start a firm have similar success as the average entrepreneur in the sample

of collegegoers.

To provide direct evidence that the marginal and average firm have similar success, we estimate the probability of reaching $1 million in annual revenue conditional on having started a marginal firm. Estimation is performed via 2SLS, where the "reduced form" estimates the effect of 2nd-degree connections on an indicator for starting a firm with greater than $1 million in annual revenue and the "first stage" estimates the effect on starting any firm. The results, presented in Column (8) of Table D5, suggest that the conditional probability of becoming a high-revenue firm among marginal firms is similar to the conditional probability among all firms in the sample.

Effects of Indirect Connections Several pieces of evidence suggest that the results cannot be explained by solely by shifting entrepreneurship among direct connections.

First, our instruments do not induce existing direct connections to start businesses at rates high enough to explain the results. Conditioning on the direct connection not having started a business does not appreciably change the effect magnitudes. Note that finding an effect even when the existing connection does not start a business does not imply that direct connections do not matter: indeed, we find that the effects in the movers design are substantially stronger when the mover does start a business. Rather, the effect occurs even when the mover does not start a business because base rates of entrepreneurship are low.

Second, one may be concerned that the baseline estimates reflect the fact that the second-degree connections to entrepreneurs used in the quasi-experimental design may later become direct connections. However, such conversions from 2nd-degree connection to direct friendship are rare: fewer than 1% of 2nd-degree connections to owners formed at the start of college convert to first degree connections by the end of the sample.

Finally, the design allows us to estimate the effect of the quasi-experimental treatment on friends which the focal individual made prior to starting college. These existing friends of collegegoers are *three* links removed from the entrepreneurs used to define the variation. Table 3 demonstrates that despite this meaningful distance (in network space), quasi-experimental exposure to entrepreneurs increases the probability that an individual will found a firm. This effect is not driven by prior rates of entrepreneurship: Section III.A.3 will demonstrate that the average number of firms

started by existing friends before the focal individual went to college has no association to the quasi-experimental treatment.

III.A.3 Falsification Tests

The identification assumption underlying the baseline causal estimates depend on future founders not differentially forming 2nd-degree connections with existing business owners at the start of college. While this assumption cannot be tested directly, we provide two complementary pieces of evidence that the conditions for identification are satisfied.

First, the extent to which a students makes 2nd-degree connections to business owners during college is unrelated to the student's friendship formation patterns before turning 18. The rightmost estimate in Figure 9 plots an estimate of β for a falsification test corresponding to Equation 1, where the outcome is replaced by the number of friendships with owners before the year that i turns 18. The falsification test gives a precisely estimated zero "effect" of the treatment on pre-college friending patterns, consistent with the identification assumption.

Second, we show that the zero estimate is not due to poor measurement of the number of existing connections (the falsification outcome). If pre-18 friending patterns were mostly noise, it should not be predictable using friending patterns outside the treament window either. However, each of the estimates on the left side of Figure 9, corresponding to estimates of γ for the falsification test, strongly reject a zero effect. The outcome and all covariates are standardized so each estimate can be interpreted as the partial correlation between the number of owners i had in their network before the year they attend college and the number of new owners they add to their network in a given period. For example, the estimate corresponding to weeks 1-4 tells us that an individual i who had 1SD more connections to owners on Dec 31 of the year before i turned 18 made 0.08SD more connections to owners during January of the year i turned 18.

The pattern in this figure is stark: the number of existing connections to owners predicts friending behavior in every period of the year in which person i turns 18 until i starts college (i.e., every element of γ is positive in the falsification test). However, pre-18 network characteristics are unrelated to connections made in the first 12 weeks of the semester (i.e., we cannot reject $\beta = 0$

in the falsification test). Thus, we conclude that for any selection to contaminate our estimated treatment effects, the selection must be unrelated to friending behavior in all periods leading up to the start of college.

III.A.4 Within-Industry Identification

We next investigate whether the clustering of owner networks within narrowly defined industries that we documented in Section II.B reflects a causal effect.

There is a formal sense in which within-industry estimates can be interpreted as causal evidence of exposure effects distinct from the baseline College design. The identifying assumption in the industry-specific estimates is that students who have a higher-propensity to found a business in one of more than 1000 industries at the 6-digit NAICS level do not preferentially form links with others that have connections in that precise industry during the first weeks of college. We believe it is unlikely that students strategically form 2nd-degree connections in the exact industry where they are likely to found a firm years later, while forming no connections in closely related industries. Thus, the industry-specificity of the effect suggests that our estimates reflect a causal effect rather than nonrandom network formation. Formal identification in class-specific models relies on overidentification conditions to distinguish between models of selection, derived in (Bell et al. 2019).

To do so, we would like to estimate an equation of the following form:

$$Y_{i,k} = \alpha_k + \tau_{s,c} + \beta T_{i,k} + \epsilon_{i,k} \tag{2}$$

where $Y_{i,k} = 1$ if person i started a firm in industry k (for the 1059 industries k defined by 6-digit NAICS codes) and $T_{i,k}$ the number of new 2nd degree connections in industry k.

Estimating Equation 2 directly among all students is not computationally feasible.[19] However, a numerically identical estimate of β can be produced by regressing the fraction of individuals who start a business on the number of connections who own a business in that industry, across cells

[19]For 1 million students and 1000 industries, estimating Equation 2 directly requires a student-by-industry-level dataset of 1 billion rows, which exceeds our computational memory constraints.

defined by school × cohort × industry × number of connections (Bell et al. 2019). In order to increase the power of the within-industry design, we use as a treatment all connections made in weeks 33-44 of the year a college-goer turns 18 rather than only connections made within the same college who are the same age.

Table 3 Column (5) present estimates of β from the collapsed approach. Overidentification tests in Columns (6)-(8) replace the explanatory variable T_{ik} with the number of connections in a different industry but same industry group, same sector, or different sector, similar to the approach in Bell et al. 2019. The effects load entirely on the same narrowly-defined industry: an extra 2nd-degree connection in industry k has no effect on firm creation in related industries. We interpret this pattern as evidence against models of selection in which students interested in entrepreneurship select connections in an industry-agnostic manner.

III.B Friend Migration

The College quasi-experimental design demonstrates that connections to entrepreneurs matter in the decision to start a firm. However, the design does not have the statistical power to estimate place-specific effects which could inform geographic targeting of entrepreneurship policies. To address this gap, we implement a second quasi-experimental empirical design, termed the "Movers" design, which exploits the timing of friend migration across commuting zones.

III.B.1 Design and Estimation: The Timing of Moves

To make the identifying variation behind the Movers design clear, we start by providing a concrete example that illustrates the type of comparisons we wish to make in the data. Two individuals, Person A and Person B, both live in Boston. Person A is friends with Person X, who moves to Silicon Valley in 2016. Person B is friends with Person Y, who moves to Silicon Valley in 2019. How does Person A's founding behavior differ from Person B's behavior between 2016 and 2019, in the years after X moved but before Y moved?

We generalize and operationalize the basic idea above by constructing stacked difference-in-differences estimates for each pair of commuting zones in the United States. The first city in the

pair is the residence of the non-mover, which corresponds to Boston for Persons A and B in the example above. The second city is the destination of the mover, which corresponds to Silicon Valley for Persons X and Y in the example. Constructing estimates at the level of commuting zone pairs ensures that quasi-experimental comparisons are only made between two individuals who live in the same place and have a friend move to the same destination. Only friendships formed prior to the start of the panel (that is, before 2015) are used in the design so that endogenous link formation during the movers sample does not contaminate estimates.

Formally, let $\mathcal{C}$ denote the set of commuting zones and $c(i,t) \in \mathcal{C}$ denote the commuting zone where person i resides at time t. The statistical model is expressed in Equation 3:

$$Y_{it} = \alpha_{ndi} + \lambda_{ndt} + \sum_{k \neq -1} \beta_{ndk} \mathbb{1}(t - D_{id} = k) + \epsilon_{ndit} \tag{3}$$
$$\forall (n,d) \in \mathcal{C} \times \mathcal{C} \text{ s.t. } n \neq d,\ c(i,t) = n$$

where Y_{it} is one if user i has founded a firm by year t and zero otherwise, D_{id} is defined as the first year in which person i has a friend move to destination d among friendships formed before the first year of the panel. The omitted category is the year prior to the first friend moving to destination d. Each coefficient β_{ndk} estimates the dynamic treatment effect in founding behavior for an individual living in place n, k years after they first have a friend who moves to place d. The identifying assumption underpinning this empirical design is that the entrepreneurship rate among friends of movers would have evolved in parallel for movers who moved at different times.

We estimate the set of β_{ndk} for each CZ-pair (n,d) using a stacked difference-in-differences estimator, which addresses potential contamination from variation in treatment timing under heterogeneous treatment effects (Sun and Abraham 2021). We consider moves between 2016 and 2020, and use firm formation between 2015 and 2022 as the outcome. Appendix Figure E18 provides evidence that moves can be accurately measured in our data by showing that user interactions with pages corresponding to local businesses switch from origin to destination at the time of move.

III.B.2 Effects by Mover Destination

Given the set of β_{ndk}, we first show that a nonmover is more likely to start a firm when their friends move to highly entrepreneurial places.

Figure 10 plots averages of β_{ndk} over nonmover commuting zones n, where the destinations d are binned into three groups: high-, medium-, and low-entrepreneurship, defined by CZ-level population-weighted terciles of the average number of new jobs per capita created by firms less than 1 year old between 2010 and 2020. Binning by the entrepreneurship rate in the destination and comparing the results across bins allows us to evaluate potential explanations for any effects we might find.[20]

The estimates indicate that friends of a mover to a city with many successful entrepreneurs are 0.02pp more likely to have started a firm 3 years after the move.

Conversely, friends of a mover who goes to a city that creates few successful employer businesses do not found more businesses after the move. This pattern narrows the types of shocks that threaten identification. For the results to be explained by unobserved shocks rather than network exposure, the shock must (1) be time-varying within the set of individuals who live in the same place and have a friend move to the same destination, and (2) only effect the non-mover's entrepreneurship among individuals who had a friend move to a high-entrepreneurship destination (but not a low-entrepreneurship destination).

There is little evidence of pretrends in any of the bins prior to the friend moving, implying that individuals exposed to the quasi-experimental treatment earlier do not start firms at different rates prior to the move. Appendix Figure E19 presents a set of estimates of the effect for high-entrepreneurship destinations separately by treatment cohort for the 2020 control cohort.

The absence of pretrends also suggests that the effects are not driven by differential reporting of firms in the data. Note that our linkage to firm-level data allows us to measure the year in which a firm was actually founded rather than the year that the firm was reported on the platorm or

[20] Formally, we form a destination-specific average for CZ d as

$$\frac{1}{\sum_n N_{nd}} \sum_{n \in \mathcal{C}} \beta_{ndk} N_{nd} \tag{4}$$

for $N_{nd} \equiv \sum_i \mathbb{1}(D_{id} < \infty \ \& \ c(i,t) = n)$.

claimed by the founder. Because of this feature of the linked data, any effect of friend migration on whether a firm is reported on Facebook would show up in this design as the firm being formed prior the the move. The zero effect prior to the migration event therefore provide evidence that reporting and claiming of firms among entrepreneurs does not differ between the treatment and control group.

We also provide estimates of the interaction between the annual exposure effects and the entrepreneurship rate in the mover's destination. To do so, we estimate the equation:

$$\hat{\beta}_{ndk} = \alpha_{nd} + \tau k + \beta_{interaction}\ EshipRate_d \times k + \epsilon_{ndk} \tag{5}$$

For $k \geq -1$, such that the trend is taken over all post-move periods in the difference-in-differences estimator. The coefficient $\beta_{interaction}$ describes how the time trend in β_{ndk} varies with the rate of entrepreneurship in the destination, where entrepreneurship is measured by the average number of jobs added per capita in the commuting zone by firms less than 1 year old. Column (3) in Table 5 present estimates of $\beta_{interaction}$.

III.B.3 Within-Sector Effects by Destination Sectoral Composition

The patterns hold even when considering sector-specific effects. Let $s \in \mathcal{S}$ denote a sector measured at the 2-digit NAICS level. We estimate the following sector-specific version of Equation 3:

$$Y_{its} = \alpha_{ndis} + \lambda_{ndt} + \sum_{k \neq -1} \beta_{ndks} \mathbb{1}(t - D_{id} = k) + \epsilon_{ndits} \tag{6}$$
$$\forall (n, d, s) \in \mathcal{C} \times \mathcal{C} \times \mathcal{S} \text{ s.t. } n \neq d,\ c(i,t) = n$$

Appendix Figure E21 recreates Figure 10 where the destination market d is defined at the commuting zone-by-sector level, and finds qualitatively similar patterns. We formally estimate place-by-sector interactions in a within-sector version of Equation 5:

$$\hat{\beta}_{ndks} = \alpha_{nds} + \tau k + \beta_{interaction}\ EshipRate_{ds} \times k + \epsilon_{ndks} \tag{7}$$

such that $EshipRate_{ds}$ denotes the entrepreneurship rate in commuting zone d and sector s. The sectoral composition of each commuting zone is calculated from the county-level data in the business dynamics statistics.

Column (4) of Table 5 provides estimates of Equation 5 at the destination-by-sector level. Column (5) replaces the explanatory variable with the entrepreneurship rate in Sectors other than s, and Column (6) includes both the same-sector and other-sector terms. Effects load primarily on the same-sector entrepreneurship rate rather than the cross-sector entrepreneurship rate, similar to the industry-specific results in the College Design. We interpret this place-by-sector specificity as indicating that individuals exposed to places that are highly entrepreneurial within a specific sector are particularly likely to produce spillover effects within that sector, and are less likely to produce spillover effects outside that sector.

III.B.4 Mover Friending Behavior and Mover Destination

We next provide evidence consistent with the hypothesis that the effect in the Movers design operates through the mover's connections to entrepreneurial individuals in the destination. Directly comparing movers who make different numbers of friends with entrepreneurs in the destination may reflect reverse causality. Instead we identify movers who are likely to make friends with entrepreneurs in the destination based on pre-move friending patterns. Appendix Figure E20 demonstrates that friendships with entrepreneurs prior to 2015 is a strong predictor of the number of entrepreneur friends a person makes post-2015 relative to other individual-level covariates in our data.

Accordingly, Figure 12 presents a version of the baseline Movers difference-in-difference estimates which cut by the pre-2015 propensity for the mover to befriend business owners. The results indicate that friends of a mover to a high-entrepreneurship CZ only benefit when the mover is likely to befriend business owners based on their pre-2015 friending patterns. Friends of movers

who make few entrepreneur friends see no boost to their probability of starting a firm, and friends of movers who typically make many entrepreneur friends but move to a low-entrepreneurship CZ benefit from the move much less than friends of those going to an entrepreneurial place.

IV Model and Estimation Approach

What can the data tell us how spillover effects interact with policies aimed at promoting entrepreneurship? To answer this question, we start by describing a linear peer effects model which incorporates human capital spillovers through social networks. We describe human capital as a determinant of entry which reflects the gains from information that help an entrepreneur better understand how to start and run their business. For example, fixed costs of entry may depend on the entrepreneur's understanding of how to identify opportunities, secure financing, get proper permits, or engage in appropriate marketing to make their product viable in the market.

We consider a series of counterfactuals aimed at understanding how spillovers impact the treatment effects of a hypothetical policy intervention which can be targeted across space or subgroups.[21] The results of the model estimation and policy simulations using are described in Section V.

Appendix Section Appendix B: provides a substantial extension of the basic model in order to derive two welfare results that motivate the focus on entry. The first result describes conditions under which policies that shift down cost functions and produce spillovers will increase aggregate welfare. The second shows that the effect of a hypothetical intervention on firm creation provide a sufficient statistic proportional to the welfare effect when it operates by reducing fixed costs.

IV.A Entry, Spillovers, and Individual-level Treatment Effects

IV.A.1 Setup

The economy consists of individuals $i \in \mathcal{I}$ distributed across geographic markets $c \in \mathcal{C}$. Each individual i supplies one unit of labor inelastically, which can be used to either earn an exogenous

[21] One prominent example of a large-scale policy aimed at connecting individuals with knowledge of a field is the Service Corps of Retired Executives (SCORE). SCORE operates by training volunteers with executive experience to deliver mentorship and run workshops on management or business-relevant technical skills for new and aspiring small business owners. Like many similar programs in the U.S., these services are organized through more than 200 local chapters.

wage or run a profit-maximizing firm. Let Y_i be an indicator variable which describes whether an individual starts a firm.

At time t, individual i has entrepreneurial human capital $E_{i,t}$, which reduces the sunk cost of entering the market and is not observed by the econometrician. Following the linearity in levels exhibited in Figure 3a, we model firm entry as:

$$Y_{i,t} = \beta_{c(i,t)} E_{i,t} + \epsilon_{i,t} \tag{8}$$

Where $Y_{i,t}$ is one if i starts a business at time t and zero otherwise. Individual i's human capital contributes to entrepreneurship to a greater or lesser extent depending on the barriers to entry in i's city of residence at time t, $c(i,t)$, which is parameterized by $\beta_{c(i,t)}$. Cities with high barriers to entry correspond to a low value of β_c, such that a given change in an individual's level of entrepreneurial human capital has a smaller entry effect.

An individual's human capital E_i is a linear function of the individual's unobserved type α_i, a policy instrument D_i, and the sum of the human capital of their friends.

$$E_{i,t} = \alpha_i + \delta D_{i,t} + \gamma \sum_{j \in \mathcal{I}} g_{i,j,t} E_{j,t} \tag{9}$$

The sum of the human capital of friends is expressed using the indicator $g_{i,j,t}$ which equals one if i and j are friends at time t and zero otherwise. To reflect the undirected nature of our network data, we assume that connections are symmetric such that $g_{i,j,t} = g_{j,i,t}$. Indicators $g_{i,j,t}$ are collected into an adjacency matrix G_t with dimensions $|\mathcal{I}| \times |\mathcal{I}|$. We consider the outcome of interest to be whether an individual has ever started a business, friendship links to be permanent once formed, and treatment to indicate whether an individual has ever been treated, such that $g_{i,j,t'} \geq g_{i,j,t}$ and $D_{i,t'} \geq D_{i,t} \;\; \forall \; t' > t$.

IV.A.2 Treatment Effects in the Human Capital Equilibrium

This model implies a unique equilibrium in $\vec{E}_t$, which we denote $\vec{E}_t^*$, where Equation 9 is satisfied for all individuals at all points in time. [22] The equilibrium for person i can be expressed as:

$$E_{i,t}^* = \alpha_i + \delta D_{i,t} + \sum_{n=1}^{\infty} \sum_{j \in \mathcal{I}} \gamma^n [G_t^n]_{i,j} (\alpha_j + \delta D_{j,t}) \tag{10}$$

where $[G_t^n]_{i,j}$ denotes the (i,j) element of the n-th power of the adjacency matrix G_t.

We can derive the effect of an increase in the treatment $D_{i,t}$ by one unit for person i at time t at the equilibrium. This effect, which we call $\tau_{i,t}$, can be written as:

$$\tau_{i,t} \equiv \delta \beta_{c(i,t)} + \delta \sum_{n=1}^{\infty} \sum_{j} \gamma^n \left[G_t^n\right]_{i,j} \beta_{c(j,t)}$$

To keep the notation simple, it will be convenient to explicitly label some of the quantities in the model. Let $K_{i,t} \equiv \sum_{j,n} \gamma^n \left[G_t^n\right]_{i,j}$ such that $K_{i,t}$ is the Katz-Bonacic centrality of person i at time t. We can then define $A_{i,t} \equiv \frac{1}{K_{i,t}} \sum_{j,n} \gamma^n \left[G_t^n\right]_{i,j} \alpha_j$ and $B_{i,t} \equiv \frac{1}{K_{i,t}} \sum_{j,n} \gamma^n \left[G_t^n\right]_{i,j} \beta_{c(j,t)}$, such that $A_{i,t}$ and $B_{i,t}$ can be interpreted as a weighted average of α_j and $\beta_{c(j,t)}$, respectively, over individuals j with weights equal to the length-discounted number of paths between i and j at time t. Under these definitions, the effect $\tau_{i,t}$ can be rewritten as

$$\tau_{i,t} = \delta \beta_{c(i,t)} + \delta K_{i,t} B_{i,t} \tag{11}$$

Equation 11 has a straightforward interpretation and provides some immediate insights into the distribution of effects from a hypothetical intervention in our networked economy. The first term, $\delta \beta_c(i,t)$, is equal to the effect of the intervention on the targeted individual and depends on both the strength of the intervention δ and the barriers to entry in i's market, $\beta_c(i,t)$. This term

[22] The equilibrium is unique when γ is less than $\frac{1}{||G||}$ for $||G||$ the largest eigenvalue of the adjacency matrix. Appendix Section C.0.1 provides evidence that this technical condition is satisfied.

reflects the portion of the effect that operates in a world where network effects are equal to zero. We refer to this as the "Direct Effect" of the intervention.

The second term, $\delta K_{i,t} B_{i,t}$, is equal to the spillover effect resulting from the intervention. It is a combination of the composition of barriers to entry in individual i's network, represented by $B_{i,t}$, and the scale of i's extended network, represented by $K_{i,t}$. The multiplicative structure of the spillover effects implies that spillovers are largest when the individual targeted by the intervention both has many connections and the average connection faces low barriers to entry.

The additive decomposition of the total effect $\tau_{i,t}$ into a direct effect and spillover effect suggests that there will be a trade-off between the effects to the extent that the two are anti-correlated. Section V.B demonstrates using the estimated model that this trade-off appears to be empirically relevant in our setting.

IV.B Modeling the Movers Design

We next show how the individual-level treatment effect from the baseline model will be aggregated up to the level of a geographic market and estimated using the variation from the Movers design. Let $\overline{\mathcal{C}} \subseteq \mathcal{C}$ denote a subset of cities.

Consider the average annual effect of a friend moving to destination d on a non-mover in commuting zone n, which we will denote Δ_{nd}. Appendix Section Appendix C: demonstrates that Δ_{nd} is a combination of the network characteristics in d, the barriers to entry in n, and the human capital endowments among the mover's new friends post-move. However, since we've modeled the human capital endowments and treatment effects of the intervention as linearly separable in Equation 11, our estimation approach requires that $\Delta_{n,d}$ be purged of variation from human capital endowments among new connections in the destination.

Let $A_{n,d}$ denote the average human capital endowments of new friends of the movers. Then we can define two quantities:

1. Let $SP_c \equiv \mathbb{E}\left[\frac{\Delta_{n,d}}{A_{n,d}} | d = c\right]$ be the average rate at which non-migrant friends of movers to city c found firms, corrected for the human capital endowments of new friends formed in the destination. We will refer to this quantity as a city's "Spillover Production".

2. Let $NP_c \equiv \mathbb{E}\left[\frac{\Delta_{n,d}}{A_{n,d}} | n = c, d \in \overline{\mathcal{C}}\right]$ be the average rate at which non-migrants residing in city c found firms when a friend moves to a city in the set $\overline{\mathcal{C}}$, corrected for the human capital endowments of new friends formed in the destination.[23] We will refer to this quantity as a city's "Networking Productivity".

Direct Effects Appendix Section Appendix C: demonstrates that the direct effect of an intervention on individuals in a given city $c(i,t)$ who take up treatment, equal to $\delta\beta_{c(i,t)}$, can be expressed as:

$$\delta\beta_{c(i,t)} = \frac{1}{C} \cdot NP_c \tag{12}$$

for a constant C, meaning that the direct effect of an intervention in place c is proportional to the networking productivity in that place. This result allows for a description of the patterns of direct effects across place based on estimates from the quasi-experimental design.

Total Effects We will consider counterfactual policies in which the target has a Katz-Bonacic centrality equal to the mean centrality of the population, $\mathbb{E}[K_{it}]$. In this case, the city-level total effect is proportional to a linear combination of city-specific Networking Productivity and Spillover Production:

$$\mathbb{E}[\tau_{i,t} | c(i,t) = c] = \frac{1}{C}\left(NP_c + SP_c \cdot \mathbb{E}[K_{i,t}]\right) \tag{13}$$

where NP_c and SP_c correspond to quantities from the movers design, and C is the same constant of proportionality as in the direct effect. In combination, Equation 12 and Equation 13 imply that the term $SP_c \cdot \mathbb{E}[K_{i,t}]$ is proportional to the spillover effect from the hypothetical intervention targeted to person i with centrality equal to the mean.

[23] The choice of $\mathcal{C}$ affects estimates proportionally. The constant of proportionality will drop out in the correlational analysis. In estimation, we set $\overline{\mathcal{C}}$ to be cities in the top population-weighted tercile of startup employment per capita.

The choice to consider counterfactual policies targeted to individuals with the same Katz-Bonacic centrality has two advantages. First, fixing centrality at the mean ensures that the results are not driven by measurement error in the extent to which individuals report their true social network on the Facebook platform. This is important given the existing evidence that patterns of Facebook usage differ by region in the U.S. Such variation is not an issue in the estimation moments derived from the Movers design, which conditions on commuting zone. The second advantage is a practical one: fixing centrality at the mean allows us to provide an estimate of centrality with lower sampling error, since our estimator of Katz-Bonacic centrality relies on a sampling approach that will reduce precision when estimated among subgroups.

The cost of fixing centrality at the mean appears to be small for the purpose of understanding the importance of spillover effects relative to direct effects. Our approach, which neglects interaction effects between network size and composition, will tend to underestimate spillovers when K_{it} and B_{it} are positively correlated. This is likely to be true in practice given that the number of friends and the fraction of friends who are entrepreneurs are correlated 0.3 in the cross section. Formal justification for our approach is given in the appendix, which provides conditions under which K_{it} and B_{it} are conditionally mean-independent at the city level.

IV.C Place-Specific Estimation

Our first series of counterfactuals require estimating the components of Equation 13 at the commuting zone level. To do so, we estimate each term other than the proportionality constant C separately, which allows us to compare the direct effect and the total effect. This term-by-term approach requires the use of new tools from the network science literature in the estimation of Katz-Bonacic centrality for large networks, which we discuss in Section IV.C.2.

IV.C.1 Networking Productivity and Spillover Production

Commuting zone Networking Productivity (NP_c) and Spillover Production (SP_c) are estimated using variation from the Movers design, combined with city-specific estimates of the human capital endowments.

To estimate $\Delta_{n,d}$, we take the linear trend in β_{ndk} over post-treatment periods. Estimation is performed by regressing post-treatment β_{ndk} from Equation 4 on relative time k. The estimating equation is

$$\hat{\beta}_{ndk} = \alpha_{nd} + \Delta_{nd} k + \epsilon_{ndk} \tag{14}$$

for $k \geq -1$, weighting by the sum of treatment and control observations used to construct each $\hat{\beta}_{ndk}$. Regressions are estimated separately for each commuting zone pair (n, d)

Commuting zone-specific human capital endowments $A_{n,d}$ are estimated using a set of fixed effect regressions in a panel within each commuting zone.

$$Y_{i,t} = \eta_i + \nu_t + \beta_{local,c} LocalFriends_{i,t} + \beta_{nonlocal,c} NonlocalFriends_{i,t} + \epsilon_{i,t} \; \forall \; c(i,t) = c \tag{15}$$

where $LocalFriends_{i,t}$ is the number of friends in the same commuting zone as i at time t, and $NonlocalFriends_{i,t}$ is the number of friends in a different commuting zone at time t. The ratio $\widehat{\beta}_{local,c}/\widehat{\beta}_{nonlocal,c}$ is used as our estimate of $A_c/\overline{A}$ for a constant $\overline{A}$. This fixed effects approach is derived in Appendix Section C.0.2.

Networking Productivity and Spillover Production are then estimated as means over destinations and non-mover commuting zones, respectively:

$$\widehat{NP}_c = \frac{1}{\sum_{d \in \overline{\mathcal{C}}} N_{n,c}} \sum_{d \in \overline{\mathcal{C}}} \widehat{\Delta}_{c,d} \frac{\widehat{\beta}_{nonlocal,c}}{\widehat{\beta}_{local,c}} \cdot N_{c,d} \tag{16}$$

$$\widehat{SP}_c = \frac{1}{\sum_{n \in \mathcal{C}} N_{n,c}} \sum_{n \in \mathcal{C}} \widehat{\Delta}_{n,c} \frac{\widehat{\beta}_{nonlocal,c}}{\widehat{\beta}_{local,c}} \cdot N_{n,c} \tag{17}$$

IV.C.2 Katz-Bonacic Centrality

We estimate Katz-Bonacic centrality using a new sampling-based approach from the network science literature (Lin et al. 2021). The sampling estimator is intended to provide a method for

estimation of individual-level Katz-Bonacic centrality in networks with many nodes, such as the Facebook social graph. The estimation approach and benchmarking is discussed in greater detail in Section C.0.1.

The estimator uses the fact that individual-level Katz-Bonacic centrality can be expressed as the sum of paths to all nodes in the network, discounted by an attenuation factor that scales scales geometrically with the length of a path. By sampling nodes and counting both direct and indirect paths, an unbiased estimator can be constructed which appropriately weights sampled indirect connections using the sampling probability. We implement this estimator over the entirety of the sample of individuals and friends used in the Movers design.

Appendix Figure E23 benchmarks our implementation of the sampling algorithm against true Katz-Bonacic centrality in a publicly-available sample of the Facebook network containing 4031 nodes (Jure Leskovec n.d.). Appendix Section C.0.1 discusses how we set the estimator's parameters in the validation sample to match the parameters in the analysis sample. Our estimates and true centrality are correlated 0.99 at the individual level in the validation sample.

Estimating γ In order to provide an estimate of Katz-Bonacic centrality, the estimator requires knowing the attenuation factor γ, which we defined in Equation 9, as an input. γ determines the relative importance of direct and indirect connections in equilibrium.

We use an estimator for γ that exploits variation from the Movers design in order to estimate the relative effect of a mover on a non-mover and on their friends. The idea is that, under the model, moving to a high entrepreneurship destination affects entry among both non-mover friends and non-mover friends of friends. If the non-mover friends and friends of friends face the same barriers to entry, then the ratio of the effects identifies the attenuation parameter.

To operationalize this approach, denote the Networking Productivity in place c among 2nd-degree connection of the mover (i.e., among friends of the non-mover) who live in c: $NPfof_c$. This is exactly analogous to the definition of NP_c, except that the outcome is the fraction of friends of the non-mover who start a firm among friends living in c. We estimate γ as the ratio of average $NPfof_c$ and NP_c

$$\hat{\gamma} = \sum_{c \in \mathcal{C}} \widehat{NPfof}_c \cdot N_c / \sum_{c \in \mathcal{C}} \widehat{NP}_c \cdot N_c \tag{18}$$

We formally derive the approach in the Appendix.

Because Katz-Bonacic centrality features prominently in many linear models of peer effects, centrality estimates may be of interest for other outcomes beyond entrepreneurship. However, the attenuation factor γ may differ when considering other types of human capital which may matter for different outcomes. With this in mind, Appendix Table D7 describes quantiles of Katz-Bonacic centrality for both our estimate $\hat{\gamma}$ and for other values of γ which may be relevant in other contexts. We hope that these additional calculations can inform peer effect models in future studies.

V Targeting Policy Counterfactuals

V.A Direct Effects

Figure 11b shows the signal correlation between the direct effects estimated by Equation 18 and commuting zone characteristics. The modest negative correlations between the direct effect and measures of entrepreneurship indicate that non-movers in a low-entrepreneurship commuting zones benefit somewhat more from a friend moving to high-entrepreneurship commuting zone compared to non-movers in a high-entrepreneurship commuting zone. However, the relationships between networking productivity and entrepreneurship rates are weaker than the relationship between the direct effect and other CZ-level covariates. The direct effect is particularly high in places with low incomes and low college attendance rates. Table D8 lists the top 10 commuting zones by direct effect, where estimates are shrunk by the signal-to-noise ratio to minimize mean squared error (Chetty and Hendren 2018).

V.B Total Effects and Place-Based Policy

Figure 11 plots the signal correlation between the estimated total effect and commuting zone characteristics. Notably, the predicted total effect from a hypothetical policy is positively correlated

with local entrepreneurship rates. This is somewhat surprising when compared with the modest negative correlation between the direct effect and entrepreneurship rates found in Figure 11b, as the change in the sign of the relationship suggests that spillover effects are strong enough to dominate the predicted direct effect of a policy.

From a targeting perspective, the results imply that interventions of the type we consider are less effective in low-entrepreneurship places despite having a somewhat larger direct effect.

V.B.1 Policy Counterfactuals

To illustrate the implications of our estimates for targeting, we return to the question we posed at the beginning of the paper: how should an intervention be targeted to maximize business creation?

In order to perform the relevant calculations, one needs to know the proportionality between the total effect from our estimated model and the effect of the hypothetical policy on human capital. For illustrative purposes, we set the scaling factor such that the treatment effect of the hypothetical intervention is the same magnitude as having a friend move to a high-entrepreneurship commuting zone.

Table 6 shows that assigning the hypothetical policy to 100 average people in a commuting zone above the 50th percentile of our total effect distribution creates 0.274 new businesses in total. The model predicts that the assignment creates 0.006 new businesses among the 100 participants, equal to the direct effect scaled by 100, and creates 0.268 new businesses among their first and second-degree social networks.

We can compare this predicted (scaled) effect to assigning the scaled policy to 100 average people in a commuting zone with a direct effect above the 50th percentile. Doing so is expected to create 0.004 new businesses among the participants, a gap of 46% relative to high-impact commuting zones under the model. However, the model predicts that an intervention targeted to low-entrepreneurship commuting zones creates only 0.094 businesses via spillovers, only 35% of the spillovers created in the high-impact commuting zones.

V.C Individual-Level Effects and Targeting

How can policymakers effectively target interventions beyond the program's location? In this section, we demonstrate how the direct and spillover effects of policies may depend on network characteristics. First, we show using the Movers design that networking productivity increases substantially with the number of entrepreneurs in an individual's network, suggesting that the direct effect of a policy also differ by an individual's network composition. We then demonstrate how spillover effects vary with connection-level characteristics by showing that firm formation effects in the College design are substantially stronger when the student and owner are of the same gender. We conclude by estimating an extension of the model that allows for targeting beyond place-based policy using both the College and Movers designs to identify effective targets for programs that aim to bolster female entrepreneurship.

V.C.1 Direct Effects: Existing Exposure to Entrepreneurs

The simplest way in which network information assists policy targeting within place is by identifying individuals with large direct effects. We illustrate this point using an individual's pre-treatment connections to entrepreneurs, which has a strong positive relationship to networking productivity in the Movers design. Figure E24 bins the difference-in-differences plot by both the entrepreneurship rates in the nonmover's commuting zone and by the number of friendships with entrepreneurs that the non-mover had prior to 2015. We construct estimates of networking productivity separately for nonmovers who have zero entrepreneur friends before 2015, a low number of entrepreneur friends (defined as having fewer than the 1% that would be expected under random friendship formation), and a high number of entrepreneur friends (more entrepreneurs than expected under random friendship formation).

Networking productivity is highest among nonmovers who have many entrepreneur friends but live in a low-entrepreneurship place. Most of the difference in effect between this "high-impact" group and the average nonmover is explained by having having many entrepreneur friends prior to the move. A nonmover who had many entrepreneur friends and lives in a high-entrepreneurship CZ still reaps substantial benefits, while a nonmover with no pre-2015 entrepreneur friends sees no

benefit regardless of the entrepreneurship rate in their city. We interpret this pattern as reflecting the fact that individuals who had either no opportunity or no inclination to befriend entrepreneurs prior to 2015 are likely not marginal; that is, they are unlikely to engage in entrepreneurship without a more substantial intervention.

V.C.2 Spillover Effects: Owner and Student Gender

We next consider how connection-level characteristics influence the spillover effects from entrepreneurs. To do so, we cut the College estimates by the gender of both the student and the business owner.

Effects are strongest when the gender of the student matches the gender of the business owner. Table 4 provides estimates from a regression which includes the number of male and female owner connections separately, both interacted with the gender of the student. The estimates imply that males owners have large effects on male students and female owners have relatively large effects on females, while cross-gender estimates are indistinguishable from zero.

V.C.3 Estimating the Extended Model for High-Impact Women

How can individual-level characteristics inform entrepreneurship policy targeting? In this section, we present estimates from an extended model that considers effects which may vary by individual-level characteristics.

The idea behind the extended model is that policymakers may want to increase entrepreneurship among a particular subgroup by targeting high-impact individuals. For example, organizations like Lean In seek to foster valuable connections for women interested in leadership or management. The key difference in the extended model relative to the baseline model is that estimates of subgroup-specific effects, which we can calculate in the College design, are required to appropriately scale the social multiplier. Other than the subgroup-specific treatment effects, the remainder of the social multiplier which can be calculated from commuting zone spillover production estimated separately by subgroup in the Mover's design. Intuitively, the reason that the Movers design alone is not sufficient is that estimates from the Movers design, which varies exposure to entrepreneurial

individuals using friends who move to high-entrepreneurship places, do not distinguish between the effect of connections to different types of people within a place. However, we can infer the distribution of type-specific effects using heterogeneity across types in the College design, and apply that distribution to appropriate estimates from the Movers design to calculate the social multiplier. The technical details of the extended model are developed in Appendix Section C.1.

We illustrate the value of our individual-level causal estimates by using them to predict the effects of entrepreneurship interventions aimed at maximizing female entrepreneurship under the extended model. Based on the evidence that individuals with many entrepreneur friends are particularly likely to start businesses when treated, we consider a hypothetical intervention which targets "high-impact" women whose networks over-represent business owners, and consider the spillover effects on business creation among all women.

Table 7 indicates that targeting an intervention for the average high-impact women in the top 50% of the (population weighted) CZ distribution produces 24% more female-owned businesses among the target population and 344% more female-owned businesses among the network of the target population than targeting the average woman in a commuting zone with below-median rates of female entrepreneurship.

VI Discussion

This paper has presented new quasi-experimental evidence on the importance of social connections for the creation of new businesses. While most previous research on social networks and entrepreneurship has focused on the direct effects of random or quasi-random exposure to potentially helpful peers on business creation or growth, our results demonstrate that the diffusion of business creation spillovers through existing social networks results in substantial social multipliers.

Existing policies which might harness social spillovers could range from networking opportunities and mentorship for aspiring entrepreneurs to subsidies for local businesses. Our analysis does not provide guidance on which types of policies might be most effective, but it does provide guidance on how any effective policy can be targeted across places and groups of people. In particular, our model predicts that policies targeted at individuals with many entrepreneur connections will

have a particularly large direct effects on the target population, while policies that create new entrepreneurs in existing high-entrepreneurship regions produce substantial spillovers.

The pattern of effects also suggests potential new approaches to targeting entrepreneurship programs. For example, the large effects of valuable connections on individuals who are already friends with entrepreneurs suggest that referrals by existing entrepreneurs to business creation programs like startup accelerators may help administrators of those programs identify high-impact candidates. Furthermore, spillovers from an intervention's target can matter through particularly narrow channels, for instance by helping aspiring entrepreneurs distinguish between highly specific industries when choosing a career path. Such specificity suggests the possibility that similarly narrow or bespoke interventions, such as providing aspiring entrepreneurs the opportunity to learn industry-specific technical or management skills, may be particularly effective when targeted towards high-impact individuals in the network.

More broadly, our results speak to the benefits of targeting interventions towards influential members of a network rather than targeting solely individuals with the highest direct effects from the intervention. Policies which take advantage of positive spillovers in arenas where people learn from or inspire each other, like entrepreneurship, innovation, or human capital accumulation, have the potential to create increasing returns and ultimately bolster economic growth. Mapping the structure of social networks and charting the paths of social spillovers in such markets is therefore a particularly promising avenue for future research and policy.

References

Acs, Zoltan J and David B Audretsch (1988). "Innovation in large and small firms: an empirical analysis". In: *The American economic review*, pp. 678–690.

Acs, Zoltan J, Pontus Braunerhjelm, David B Audretsch, and Bo Carlsson (2009). "The knowledge spillover theory of entrepreneurship". In: *Small business economics* 32, pp. 15–30.

Aghion, Philippe and Peter Howitt (1992). "A Model of Growth Through Creative Destruction". In: *Econometrica* 60.2, pp. 323–351. (Visited on 01/15/2025).

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018). *Dancing with the stars: Innovation through interactions*. Tech. rep. National Bureau of Economic Research.

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde (2021). "Back to basics: Basic research spillovers, innovation policy, and growth". In: *The Review of Economic Studies* 88.1, pp. 1–43.

Akcigit, Ufuk and John Van Reenen (2023). *The Economics of Creative Destruction*. Cambridge: Harvard University Press.[644].

Andrews, Michael J, Aaron Chatterji, Josh Lerner, and Scott Stern (2022). *The role of innovation and entrepreneurship in economic growth*. University of Chicago Press.

Andrews, RJ, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern (2022). "The Startup Cartography Project: Measuring and mapping entrepreneurial ecosystems". In: *Research Policy* 51.2, p. 104437. ISSN: 0048-7333. DOI: https://doi.org/10.1016/j.respol.2021.104437.

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt (2021). "Estimating experienced racial segregation in US cities using large-scale GPS data". In: *Proceedings of the National Academy of Sciences* 118.46, e2026160118.

Atkin, David, M Keith Chen, and Anton Popov (2022). *The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley*. Tech. rep. National Bureau of Economic Research.

Audretsch, David B and Maryann P Feldman (1996). "R&D spillovers and the geography of innovation and production". In: *The American economic review* 86.3, pp. 630–640.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda (2020). "Age and high-growth entrepreneurship". In: *American Economic Review: Insights* 2.1, pp. 65–82.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024). "Artificial intelligence, firm growth, and product innovation". In: *Journal of Financial Economics* 151, p. 103745.

Babina, Tania, Alex Xi He, Sabrina T Howell, Elisabeth Ruth Perlman, and Joseph Staudt (2023). "Cutting the innovation engine: how federal funding shocks affect university patenting, entrepreneurship, and publications". In: *The Quarterly Journal of Economics* 138.2, pp. 895–954.

Babina, Tania and Sabrina T Howell (2024). "Entrepreneurial spillovers from corporate R&D". In: *Journal of Labor Economics* 42.2, pp. 469–509.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2018). "Social connectedness: Measurement, determinants, and effects". In: *Journal of Economic Perspectives* 32.3, pp. 259–280.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel (2018). "The economic effects of social networks: Evidence from the housing market". In: *Journal of Political Economy* 126.6, pp. 2224–2276.

Bailey, Michael, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2024). "Social networks shape beliefs and behavior: Evidence from social distancing during the covid-19 pandemic". In: *Journal of Political Economy Microeconomics* 2.3, pp. 463–494.

Bailey, Michael, Drew Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (July 2022). "Peer Effects in Product Adoption". In: *American Economic Journal: Applied Economics* 14.3, pp. 488–526. DOI: 10.1257/app.20200367. URL: https://www.aeaweb.org/articles?id=10.1257/app.20200367.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Ayush Kumar, and Johannes Stroebel (2025). *Cross-Gender Social Ties Around the World*. Tech. rep. National Bureau of Economic Research.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson (2013). "The diffusion of microfinance". In: *Science* 341.6144, p. 1236498.

Bartik, Timothy J (2020). "Using place-based jobs policies to help distressed communities". In: *Journal of Economic Perspectives* 34.3, pp. 99–127.

Baumol, William J. (1990). "Entrepreneurship: Productive, Unproductive, and Destructive". In: *Journal of Political Economy* 98.5, Part 1, pp. 893–921. DOI: 10.1086/261712.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019). "Who Becomes an Inventor in America? The Importance of Exposure to Innovation". In: *The Quarterly Journal of Economics* 134.2, pp. 647–713.

Bhandari, Anmol, Tobey Kass, Thomas J May, Ellen McGrattan, and Evan Schulz (2024). *On the nature of entrepreneurship*. Tech. rep. National Bureau of Economic Research.

Bhandari, Anmol and Ellen R McGrattan (2021). "Sweat equity in US private business". In: *The Quarterly Journal of Economics* 136.2, pp. 727–781.

Bhide, Amar (2000). *The origin and evolution of new businesses*. Oxford University Press.

Bloom, Nicholas and John Van Reenen (2007). "Measuring and explaining management practices across firms and countries". In: *The quarterly journal of Economics* 122.4, pp. 1351–1408.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou (2024). "Toward a general theory of peer effects". In: *Econometrica* 92.2, pp. 543–565.

Bradford, William D and Alfred E Osborne (1976). "The entrepreneurship decision and black economic development". In: *The American Economic Review* 66.2, pp. 316–319.

Bresnahan, Timothy F and Peter C Reiss (1991). "Entry and competition in concentrated markets". In: *Journal of Political Economy* 99.5, pp. 977–1009.

Burt, R (1992). *Structural Holes*. Cambridge, MA: Harvard University Press.

Burt, RS and HJ Raider (2002). "Creating careers: women's paths to entrepreneurship". In: *Unpublished manuscript, University of Chicago*.

Cai, Jing and Adam Szeidl (2018). "Interfirm Relationships and Business Performance". In: *The Quarterly Journal of Economics* 133.3, pp. 1229–1282.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009). "Peer effects and social networks in education". In: *The review of economic studies* 76.4, pp. 1239–1267.

Campbell, Arthur, Philip Ushchev, and Yves Zenou (2024). "The network origins of entry". In: *Journal of Political Economy* 132.11, pp. 3867–3916.

Chatterji, Aaron, Edward Glaeser, and William Kerr (2014). "Clusters of entrepreneurship and innovation". In: *Innovation policy and the economy* 14.1, pp. 129–166.

Chatterji, Aaron K, Kenneth Y Chay, and Robert W Fairlie (2014). "The impact of city contracting set-asides on black self-employment and employment". In: *Journal of Labor Economics* 32.3, pp. 507–561.

Chetty, R, JN Friedman, N Hendren, MR Jones, and SR Porter (2019). *The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility*. Tech. rep. National Bureau of Economic Research Working Paper w25147.

Chetty, R and N Hendren (2018). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-level Estimates". In: *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey,

Pablo Barbera, Monica Bhole, and Nils Wernerfelt (2022a). "Social Capital I: Measurement and Associations with Economic Mobility". In: *Nature* 608.7921, 108–121.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barbera, Monica Bhole, and Nils Wernerfelt (2022b). "Social Capital II: Determinants of Economic Connectedness". In: *Nature* 608.7921, 122–134.

Chinitz, Benjamin (1961). "Contrasts in Agglomeration: New York and Pittsburgh". In: *The American Economic Review Papers and Proceedings* 51.2, pp. 279–289.

Chodavadia, Saheel A, Sari Pekkala Kerr, William R Kerr, and Louis J Maiden (May 2024). *Immigrant Entrepreneurship: New Estimates and a Research Agenda*. Working Paper 32400. National Bureau of Economic Research. DOI: 10.3386/w32400.

Chodorow-Reich, Gabriel, Plamen T Nenov, Vitor Santos, and Alp Simsek (2024). *Stock Market Wealth and Entrepreneurship*. Tech. rep. National Bureau of Economic Research.

Chrisman, James J (2017). *Economic impact of Small Business Development Center counseling activities in the United States: 2015–2016*. Tech. rep. Small Business Development Center.

Chung, Fan, Linyuan Lu, and Van Vu (2003). "Spectra of random graphs with given expected degrees". In: *Proceedings of the National Academy of Sciences* 100.11, pp. 6313–6318.

Davis, Steven J and John Haltiwanger (2024). "Dynamism diminished: The role of housing markets and credit conditions". In: *American Economic Journal: Macroeconomics* 16.2, pp. 29–61.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger (2020). "The Rise of Market Power and the Macroeconomic Implications". In: *The Quarterly Journal of Economics* 135.2, pp. 561–644.

Deb, Shubhdeep (2024). *How Market Structure Shapes Entrepreneurship and Inequality*. Mimeo.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda (2014). "The role of entrepreneurship in US job creation and economic dynamism". In: *Journal of Economic Perspectives* 28.3, pp. 3–24.

Delgado, Mercedes, Michael E Porter, and Scott Stern (2010). "Clusters and entrepreneurship". In: *Journal of economic geography* 10.4, pp. 495–518.

Denes, Matthew R, Spyridon Lagaras, and Margarita Tsoutsoura (Jan. 2025). *Entrepreneurship and the Gig Economy: Evidence from U.S. Tax Returns*. Working Paper 33347. National Bureau of Economic Research. DOI: 10.3386/w33347.

Duranton, Gilles and Diego Puga (2023). "Urban growth and its aggregate implications". In: *Econometrica* 91.6, pp. 2219–2259.

Elliott, Matthew and Benjamin Golub (2019). "A Network Approach to Public Goods". In: *Journal of Political Economy* 127.2, pp. 730–776.

Ellison, Glenn, Edward L Glaeser, and William R Kerr (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns". In: *American Economic Review* 100.3, pp. 1195–1213.

Evans, David S and Boyan Jovanovic (1989). "An estimated model of entrepreneurial choice under liquidity constraints". In: *Journal of political economy* 97.4, pp. 808–827.

Fairlie, Robert, Alicia Robb, and David T Robinson (2022). "Black and white: Access to capital among minority-owned start-ups". In: *Management Science* 68.4, pp. 2377–2400.

Fairlie, Robert W, Dean Karlan, and Jonathan Zinman (2015). "Behind the GATE experiment: Evidence on effects of and rationales for subsidized entrepreneurship training". In: *American Economic Journal: Economic Policy* 7.2, pp. 125–161.

Fairlie, Robert W., Zachary Kroff, Javier Miranda, and Nikolas Zolas (2023). *The Promise and Peril of Entrepreneurship: Job Creation and Survival among U.S. Startups*. MIT Press.

Federal Deposit Insurance Corporation (2019). *Summary of Deposits (SOD) Data*. Accessed: 2025-01-31. URL: https://banks.data.fdic.gov/bankfind-suite/SOD/customDownload.

Field, Erica, Seema Jayachandran, and Rohini Pande (2010). "Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India". In: *American Economic Review* 100.2, pp. 125–129.

Field, Erica, Seema Jayachandran, Rohini Pande, and Natalia Rigol (2016). "Friendship at work: Can peer effects catalyze female entrepreneurship?" In: *American Economic Journal: Economic Policy* 8.2, pp. 125–153.

Fluegge, Robert B (2024). *Death, Destruction, and Growth in Cities: Entrepreneurial Capital and Economic Geography After the 1918 Influenza*. Tech. rep. DOI: https://dx.doi.org/10.2139/ssrn.4173228.

Galeotti, Andrea, Benjamin Golub, and Sanjeev Goyal (2020). "Targeting interventions in networks". In: *Econometrica* 88.6, pp. 2445–2471.

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr (2015). "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines". In: *The Review of Economics and Statistics* 97.2, pp. 498–520.

Glaeser, Edward L, William R Kerr, and Giacomo AM Ponzetto (2010). "Clusters of Entrepreneurship". In: *Journal of Urban Economics* 67.1, pp. 150–168.

Golosov, Mikhail, Michael Graber, Magne Mogstad, and David Novgorodsky (2024). "How Americans respond to idiosyncratic and exogenous changes in household wealth and unearned income". In: *The Quarterly Journal of Economics* 139.2, pp. 1321–1395.

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein (2010). "Performance persistence in entrepreneurship". In: *Journal of financial economics* 96.1, pp. 18–32.

Gompers, Paul A and Josh Lerner (2002). *The money of invention: how venture capital creates new wealth*.

Granovetter, Mark S (1973). "The strength of weak ties". In: *American journal of sociology* 78.6, pp. 1360–1380.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings". In: *Journal of Political Economy* 118.3, pp. 536–598.

Grossman, Gene M and Elhanan Helpman (1992). *Innovation and growth: Technological competition in the global economy*.

Gu, Qian, Lynn A Karoly, and Julie M Zissimopoulos (2008). *Small business assistance programs in the United States: an analysis of what they are, how well they perform, and how we can learn more about them*. Tech. rep.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi (2021). "Learning entrepreneurship from other entrepreneurs?" In: *Journal of labor economics* 39.1, pp. 135–191.

Guzman, Jorge and Scott Stern (2015). "Where is silicon valley?" In: *Science* 347.6222, pp. 606–609.

Gyourko, Joseph, Albert Saiz, and Anita Summers (2008). "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index". In: *Urban studies* 45.3, pp. 693–729.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda (2013). "Who creates jobs? Small versus large versus young". In: *Review of Economics and Statistics* 95.2, pp. 347–361.

Hamilton, Barton H (2000). "Does entrepreneurship pay? An empirical analysis of the returns to self-employment". In: *Journal of Political economy* 108.3, pp. 604–631.

Howell, Sabrina T (2017). "Financing innovation: Evidence from R&D grants". In: *American economic review* 107.4, pp. 1136–1164.

Howell, Sabrina T and Ramana Nanda (2024). "Networking frictions in venture capital, and the gender gap in entrepreneurship". In: *Journal of Financial and Quantitative Analysis* 59.6, pp. 2733–2761.

Howitt, Peter and Philippe Aghion (1998). "Capital accumulation and innovation as complementary factors in long-run growth". In: *Journal of Economic Growth* 3, pp. 111–130.

Huber, Laura Rosendahl, Randolph Sloof, and Mirjam Van Praag (2014). "The effect of early entrepreneurship education: Evidence from a field experiment". In: *European Economic Review* 72, pp. 76–97.

Hurst, Erik and Benjamin Wild Pugsley (2011). *What do small businesses do?* Tech. rep. National Bureau of Economic Research.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2018). "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science". In: *The Quarterly Journal of Economics* 133.2, pp. 927–991.

Jacobs, Jane (1969). *The Economy of Cities*. New York: Random House.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". In: *the Quarterly journal of Economics* 108.3, pp. 577–598.

Jakobsen, Katrine, Henrik Kleven, Jonas Kolsrud, Camille Landais, and Mathilde Muñoz (2024). *Taxing Top Wealth: Migration Responses and their Aggregate Economic Implications*. Tech. rep. National Bureau of Economic Research.

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H.e Fowler (2013). "Inferring Tie Strength from Online Directed Behavior". In: *PLOS ONE* 8.1, pp. 1–6.

Jovanovic, Boyan (1982). "Selection and the Evolution of Industry". In: *Econometrica: Journal of the econometric society*, pp. 649–670.

Jovanovic, Boyan and Rafael Rob (1989). "The growth and diffusion of knowledge". In: *The Review of Economic Studies* 56.4, pp. 569–582.

Jure Leskovec (n.d.). *Social circles: Facebook*. Accessed: 2025-02-23. URL: https://snap.stanford.edu/data/ego-Facebook.html.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun (Jan. 2025). "The Diffusion of New Technologies*". In: *The Quarterly Journal of Economics*, qjaf002. DOI: 10.1093/qje/qjaf002.

Katz, Leo (1953). "A new status index derived from sociometric analysis". In: *Psychometrika* 18.1, pp. 39–43.

Kerr, Sari Pekkala, William R Kerr, and Ramana Nanda (2022). "House prices, home equity and entrepreneurship: Evidence from US census micro data". In: *Journal of Monetary Economics* 130, pp. 103–119.

Kerr, William R and Martin Mandorff (2023). "Social networks, ethnicity, and entrepreneurship". In: *Journal of Human Resources* 58.1, pp. 183–220.

Kihlstrom, Richard E and Jean-Jacques Laffont (1979). "A general equilibrium entrepreneurial theory of firm formation based on risk aversion". In: *Journal of political economy* 87.4, pp. 719–748.

Klenow, Peter J and Huiyu Li (Sept. 2024). *Entry Costs Rise with Growth*. Working Paper 32974. National Bureau of Economic Research. DOI: 10.3386/w32974.

Klette, Tor Jakob and Samuel Kortum (2004). "Innovating firms and aggregate innovation". In: *Journal of political economy* 112.5, pp. 986–1018.

Kleven, Henrik (Jan. 2025). *Externalities and the Taxation of Top Earners*. Working Paper 33345. National Bureau of Economic Research. DOI: 10.3386/w33345.

Koenen, Martin and Drew Johnston (2024). *Social Ties and Residential Choice: Micro Evidence and Equilibrium Implications*. Mimeo.

Kosanovich, William T, Heather Fleck, Berwood Yost, Wendy Armon, and Sandra Siliezar (2001). "Final Report: Comprehensive Assessment of Self-Employment Assistance Programs". In: *Arlington, VA: DTI Associates*.

Kwon, Spencer Y, Yueran Ma, and Kaspar Zimmermann (2024). "100 Years of Rising Corporate Concentration". In: *American Economic Review* 114.7, pp. 2111–2140.

Lazear, Edward P. (Oct. 2005). "Entrepreneurship". In: *Journal of Labor Economics* 23.4, pp. 649–680.

Lerner, J (1999). "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program." In: *Journal of Business* 72.3, pp. 285–318.

Lerner, Josh (2009). *Boulevard of broken dreams: why public efforts to boost entrepreneurship and venture capital have failed–and what to do about it*. Princeton University Press.

Lerner, Josh and Ulrike Malmendier (2013). "With a little help from my (random) friends: Success and failure in post-business school entrepreneurship". In: *The Review of Financial Studies* 26.10, pp. 2411–2452.

Levine, Ross and Yona Rubinstein (2017). "Smart and illicit: who becomes an entrepreneur and do they earn more?" In: *The Quarterly Journal of Economics* 132.2, pp. 963–1018.

Lin, Mingkai, Wenzhong Li, Lynda J Song, Cam-Tu Nguyen, Xiaoliang Wang, and Sanglu Lu (2021). "Sake: Estimating katz centrality based on sampling for large-scale social networks". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15.4, pp. 1–21.

Lockwood, Benjamin B, Charles G Nathanson, and E Glen Weyl (2017). "Taxation and the Allocation of Talent". In: *Journal of Political Economy* 125.5, pp. 1635–1682.

Lucas, Robert E (1988). "On the Mechanics of Economic Development". In: *Journal of Monetary Economics* 22, pp. 3–42.

Mankiw, N Gregory and Michael D Whinston (1986). "Free entry and social inefficiency". In: *The RAND Journal of Economics*, pp. 48–58.

Marshall, Alfred (1890). *Principles of Economics*. First edition. London: Macmillan and Co.

Mertz, Mikkel, Maddalena Ronchi, and Viola Salvestrini (2024). "Female representation and talent allocation in entrepreneurship: The role of early exposure to entrepreneurs". In: *Available at SSRN 4920176*.

Miguel, Edward and Michael Kremer (2004). "Worms: identifying impacts on education and health in the presence of treatment externalities". In: *Econometrica* 72.1, pp. 159–217.

Mobius, Markus and Tanya Rosenblat (2014). "Social learning in economics". In: *Annu. Rev. Econ.* 6.1, pp. 827–847.

Mohammadi, Ali and Kourosh Shafi (2024). "Learning Entrepreneurship From My Neighbor". In: *Available at SSRN 4741824*.

Moretti, Enrico (2021). "The effect of high-tech clusters on the productivity of top inventors". In: *American Economic Review* 111.10, pp. 3328–3375.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny (1991). "The Allocation of Talent: Implications for Growth". In: *The Quarterly Journal of Economics* 106.2, pp. 503–530.

Nanda, Ramana and Jesper B Sørensen (2010). "Workplace peers and entrepreneurship". In: *Management science* 56.7, pp. 1116–1126.

Oosterbeek, Hessel, Mirjam Van Praag, and Auke Ijsselstein (2010). "The impact of entrepreneurship education on entrepreneurship skills and motivation". In: *European economic review* 54.3, pp. 442–454.

Pew Research Center (2024). *Social Media Fact Sheet*. Accessed: 2025-02-23. URL: `https://www.pewresearch.org/internet/fact-sheet/social-media/`.

Romer, Paul M (1990). "Endogenous technological change". In: *Journal of political Economy* 98.5, Part 2, S71–S102.

Sacerdote, Bruce (2001). "Peer effects with random assignment: Results for Dartmouth roommates". In: *The Quarterly journal of economics* 116.2, pp. 681–704.

— (2011). "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In: *Handbook of the Economics of Education*. Vol. 3. Elsevier, pp. 249–277.

Saxenian, AnnaLee (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick (2019). "Capitalists in the Twenty-first Century". In: *The Quarterly Journal of Economics* 134.4, pp. 1675–1745.

Sorenson, Olav and Pino G Audia (2000). "The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940–1989". In: *american Journal of Sociology* 106.2, pp. 424–462.

Stuart, Toby E and Olav Sorenson (2003). "Liquidity events and the geographic distribution of entrepreneurial activity". In: *Administrative science quarterly* 48.2, pp. 175–201.

— (2005). "Social networks and entrepreneurship". In: *Handbook of entrepreneurship research: Interdisciplinary perspectives*, pp. 233–252.

Sun, Liyang and Sarah Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

U.S. Census Bureau (2000). *2000 Census Summary File 3*. Accessed: [Insert Date of Access]. URL: https://www.census.gov/programs-surveys/decennial-census/data/datasets.2000.html.

— (2010a). *2010 Census Planning Database*.

— (2010b). *2010 Census Summary File 1*. URL: https://www.census.gov/programs-surveys/decennial-census/data/datasets.2010.html.

— (2010c). *American Community Survey (ACS): Selected Social Characteristics*. URL: https://www.census.gov/programs-surveys/acs/.

— (2015a). *American Community Survey (ACS) 5-Year Estimates Public Use Microdata Sample (PUMS), 2011-2015*. URL: https://www.census.gov/programs-surveys/acs/data/pums.html.

— (2015b). *County Business Patterns*. Accessed: 2025-02-02. URL: https://www.census.gov/data/datasets/2015/econ/cbp/2015-cbp.html.

— (2010-2020). *Business Dynamics Statistics (BDS)*. URL: https://www.census.gov/data/tables/time-series/econ/bds/bds-tables.html.

— (2010–2019). *County Population Estimates: 2010s*. URL: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par_textimage_70769902.

U.S. Small Business Administration (2019). *7(a) Loan Data*. Accessed: 2025-01-31. URL: https://data.sba.gov/dataset/7-a-504-foia.

Venture Forward (2023). Accessed: 2023-10-31.

Wallskog, Melanie (2023). *Entrepreneurial Spillovers Across Coworkers*. Tech. rep. Working paper, Duke University.

Table 1: Sample Statistics

	Sample Statistic	All Individuals (1)	Has Work Experience (2)	All Owners (3)	Founders Since 2015 (4)	Student Sample (5)	Post-Grad Owners (6)
Network Characteristics							
Friends	Mean	700.0	794.9	998.5	1051.1	1370.8	1630.5
	Median	523	631	843	904	1300	1572
	Std. Dev	584.7	604.0	659.0	671.5	541.0	574.4
Entrepreneur Friends	Mean	3.6	4.5	15.2	14.4	6.1	18.7
	Median	2	2	11	10	4	13
	Std. Dev	5.6	6.4	15.6	14.8	6.2	19.3
Firm Characteristics							
Years Since Founding	Mean			11.3	6.9		4.8
	Median			10	7		5
	Std. Dev			6.5	2.1		2.8
Share >$1m Annual Revenue				0.16	-		-
Share with 6-digit NAICS				0.68	0.72		0.63
Share with 2-digit NAICS				0.94	0.91		0.96
Demographics							
Share Attended College		0.33	0.41	0.48	0.47	1	1
Share Female		0.56		0.47	0.48	0.64	0.54
Age	Mean	41.3		45.9	42.5	29.6	30.8
	Median	40		46	41	30	31
	Std. Dev	11.7		9.8	9.5	2.1	1.9
Zip Code Median Income	Mean	72.8	73.6	80.2	77.7	89.3	84.5
($1000s)	Median	66.1	67.1	74.1	71.7	83.2	79.3
	Std. Dev	30.8	30.5	32.0	30.8	37.9	35.1
Sample Size		107,715,276	63,339,757	527,070	295,588	788,946	1,986

Notes: This table presents summary statistics for the primary samples used in the analysis. Column (1) includes all individuals in the primary analysis sample, which restricts to users aged 18-64 with between 50 and 3000 friends. Columns (2) and (3) restrict to all business owners and business owners with a business created since 2015, respectively. Column (4) restricts to students used in the College design, which includes college graduates born before 1998 who made at least 5 friends at the same school during the fall of the year the user turned 18. Column (5) further restricts to the owners in the college sample who started businesses in the year they turned 22 or later. Sample restrictions and definitions for the full sample are described in Section II. Sample restrictions for the College design are described in Section III.A.

Table 2: Average Number and Average Fraction of Owner Friends by Own Gender and Friend Gender

		Owner Friends		
Own Gender	Friend Gender	Average Number	Average Fraction	Number of Observations
Pooled	Pooled	3.41	0.00509	10,428,120
Pooled	Male	1.78	0.00607	10,428,120
Pooled	Female	1.64	0.00434	10,428,120
Male	Pooled	3.15	0.00509	4,715,032
Male	Male	1.97	0.00609	4,715,032
Male	Female	1.18	0.00400	4,715,032
Female	Pooled	3.64	0.00509	5,713,089
Female	Male	1.61	0.00604	5,713,089
Female	Female	2.03	0.00461	5,713,089

Notes: This table presents the average number of owner friends and average fraction of owner friends by an individual's own gender and their friends' gender. The table is structured to show differences in owner friend counts across different gender pairings. The final column provides the total number of observations.

Table 3: College Design: Effects of Owner Connections Overall and Within Industry

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner	(4) Sole Owner	(5) Fraction Owners	(6) Fraction Owners	(7) Fraction Owners	(8) Fraction Owners
Owner Friends of Friends (FoFs)	7.09×10^{-5} (2.54×10^{-5})	4.93×10^{-5} (2.53×10^{-5})	7.03×10^{-5} (3.01×10^{-5})	4.64×10^{-5} (2.13×10^{-5})				
Owner FoFs in Same 6-Digit Industry					1.35×10^{-5} (1.29×10^{-6})			
Owner FoFs in Same 4-Digit Industry but Different 6-Digit Industry						-3.53×10^{-7} (6.14×10^{-7})		
Owner FoFs in Same 2-Digit Industry but Different 4-Digit Industry							2.67×10^{-7} (1.39×10^{-7})	
Owner FoFs in Different 2-Digit Industry								3.33×10^{-8} (1.51×10^{-8})
Control for Pre-College Owner FoFs		X	X	X	X	X	X	X
Control for # Friendships in First 12 Weeks of College			X					
School by Birth Cohort Fixed Effects	X	X	X	X	X	X	X	X
6-Digit Industry Fixed Effects					X	X	X	X
Unit of Observation	Individual	Individual	Individual	Individual	Industry by School by Birth Cohort by # Owner FoFs	Industry by School by Birth Cohort by # Owner FoFs	Industry by School by Birth Cohort by # Owner FoFs	Industry by School by Birth Cohort by # Owner FoFs
Number of Cells	598,892	598,892	598,892	598,892	17,848,368	18,367,512	26,199,301	90,578,242
Mean of Dep. Var	0.0027	0.0027	0.0027	0.0019	1.29×10^{-6}	1.29×10^{-6}	1.29×10^{-6}	1.29×10^{-6}
SD of Dep. Var	0.052	0.052	0.052	0.043	2.44×10^{-4}	2.24×10^{-4}	2.83×10^{-4}	4.51×10^{-4}
Mean of Ind. Var	3.30	3.30	3.30	3.30	0.001615	0.00571	0.111	2.76
SD of Ind. Var	3.00	3.00	3.00	3.00	0.0452	0.0864	0.415	3.60
p90 - p10 difference in Owner FoFs in year student turns 18	39	39	39	39				

Notes: This table presents univariate estimates from the College design for college students who turn 18 between 2006 and 2016. The outcome is firm ownership starting four or more calendar years after the year that the student turns 18. The explanatory variable is the number of 2nd-degree links (friends of friends) to firm owners formed during weeks 33-44 of the year the student turns 18, corresponding to the first 12 weeks of college under the semester system. Columns (1) and (2) estimate the effect of exposure to any owner on students, and are estimated at the student level as described in Section III.A.2. Column (3) adds a control for the number of friendships formed by the student in the treatment period during the first 12 weeks of college. Column (4) redefines the outcome to be whether a student starts a firm with no co-founders. Columns (5)-(8) estimates regressions at the 6-digit NAICS Industry by school by birth-year by number of owner friends-of-friends level as described in Section III.A.4. Bins with greater than 10 owner friends of friends are dropped. Point estimates from the binned approach are numerically identical to the same regressions run at the individual level. Standard errors clustered at the birth cohort by college level are reported in parentheses.

Table 4: College Design: Gender-Specific Exposure Effects

A. Gender Specific Exposure Effects					
Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner	(4) Is Owner	(5) Is Owner
Is Male					0.001
					(1.93×10^{-4})
Male Owner Friends of Friends		-1.23×10^{-5}		2.21×10^{-4}	-2.81×10^{-5}
		(3.90×10^{-5})		(6.71×10^{-5})	(4.14×10^{-5})
Female Owner Friends of Friends	9.58×10^{-5}		2.19×10^{-4}		1.34×10^{-4}
	(4.10×10^{-5})		(6.87×10^{-5})		(4.34×10^{-5})
Male Student x Male Owner Friends of Friends					1.85×10^{-4}
					(8.03×10^{-5})
Male Student x Female Owner Friends of Friends					-8.97×10^{-6}
					(8.43×10^{-5})
Control for Pre-College Owner FoFs	X	X	X	X	X
School x Birth Cohort Fixed Effects	X	X	X	X	X
Sample	Females	Females	Males	Males	Pooled
Observations	491,641	491,641	281,378	281,378	773,019
Mean of Dependent Variable	0.0024	0.0024	0.0035	0.0035	0.0028
SD of Dependent Variable	0.049	0.049	0.059	0.059	0.053
Mean of Independent Variable	1.55	1.56	1.40	1.63	
SD of Independent Variable	1.86	1.81	1.75	1.90	

Notes: This table estimates gender-specific exposure effects in the college design. Column (1) estimates the coefficient from a univariate regression of an indicator for founding a firm after college on exposure to male owners. Column (2) repeats Column (1) using exposure to female owners as the independent variable. Column (3) estimates a multivariate regression where firm founding is regressed on exposure to male and female owners simultaneously, where both are interacted with an indicator for whether the student is male.

Table 5: Movers Design: Annual Exposure Effects By Friend Destination CZ Entrepreneurship

Dependent Variable	(1) Fraction Owner	(2) Fraction Owner	(3) Fraction Owner	(4) Fraction Owner	(5) Fraction Owner	(6) Fraction Owner
Years After First Friend Moves	-6.62×10^{-5} (1.82×10^{-5})	-5.25×10^{-5} (1.61×10^{-5})	-5.40×10^{-5} (1.57×10^{-5})	5.19×10^{-7} (3.26×10^{-7})	-1.77×10^{-6} (7.28×10^{-7}	-2.54×10^{-6} (7.34×10^{-7})
Destination Entrepreneurship Rate		0.0046 (8.38×10^{-4})				
Destination Entrepreneurship by Years After First Move	0.007 (0.0011)	0.0062 (0.0010)	0.0062 (9.85×10^{-4})			
Destination Entrepreneurship In Same Sector by Years After First Move				0.0014 (3.64×10^{-4})		0.0013 (3.63×10^{-4})
Destination Entrepreneurship In Other Sectors by Years After First Move					2.31×10^{-4} (4.59×10^{-5})	2.06×10^{-4} 4.63×10^{-5})
Fixed Effects	none	none	Destination CZ by Nonmover CZ	Destination CZ by Nonmover CZ by Sector	Destination CZ by Nonmover CZ by Sector	Destination CZ by Nonmover CZ by Sector
Unit of Observation	Destination CZ by Nonmover CZ by Relative Year by Sector	Destination CZ by Nonmover CZ by Relative Year by Sector	Destination CZ by Nonmover CZ by Relative Year by Sector	Destination CZ by Nonmover CZ by Relative Year by Sector	Destination CZ by Nonmover CZ by Relative Year by Sector	Destination CZ by Nonmover CZ by Relative Year by Sector
Number of Cells	1,368,248	1,368,248	1,368,248	32,128,297	32,128,297	32,128,297
Mean of Dependent Variable	5.016×10^{-5}	5.016×10^{-5}	5.016×10^{-5}	1.824×10^{-6}	1.824×10^{-6}	1.824×10^{-6}
SD of Dependent Variable	0.0048	0.0048	0.0048	9.586×10^{-4}	9.586×10^{-4}	9.586×10^{-4}
Mean of Jobs Variable	0.016	0.016	0.016			
SD of Jobs Variable	0.0048	0.0048	0.0048			
Mean of Same Sector Jobs				8.69×10^{-4}		8.69×10^{-4}
SD of Same Sector Job				8.43×10^{-4}		8.43×10^{-4}
Mean of Other Sector Jobs					0.015	0.015
SD of Other Sector Jobs					0.0047	0.0047

Notes: This table estimates interactions between the destination entrepreneurship rate and the post-treatment trend in the Movers Design. Column (3) is the baseline specification expressed in Equation 5. Columns (1) and (2) build up to the baseline specification by estimating Equation 5 without destination CZ by nonmover CZ fixed effects.

Table 6: Commuting Zone Targeting under Estimated Baseline Model

	Commuting Zones in top 50% of Total Impact	Commuting Zones below 50th Percentile of Entrepreneurship Rate	Ratio of (1)/(2)
	(1)	(2)	(3)
Firms Created among Target Population	0.006	0.004	1.46
	[0.005,0.007]	[0.002,0.005]	
Firms Created among Individuals who were Not Directly Targeted	0.268	0.094	2.86
	[0.188,0.359]	[-0.009,0.196]	
Total Firms Created	0.274	0.098	2.80
	[0.193,0.365]	[-0.005,0.200]	

Notes: This table provides estimates of the effect of a hypothetical policy targeted at 100 individuals under the baseline model using causal estimates from the Movers design under the assumptions described in Section V. The policy is assumed to have the same commuting zone-specific direct effect as having a friend move to a high-entrepreneurship commuting zone. Column (1) estimates the direct effect, spillover effect, and total effect from a policy targeted at commuting zones with a total effect above the median under the estimated model. Column (2) estimates the effects from a policy targeted at a commuting zone with below-median entrepreneurship rates. We report credible sets at the 95% level calculated from a Bayesian bootstrap with 1000 iterations.

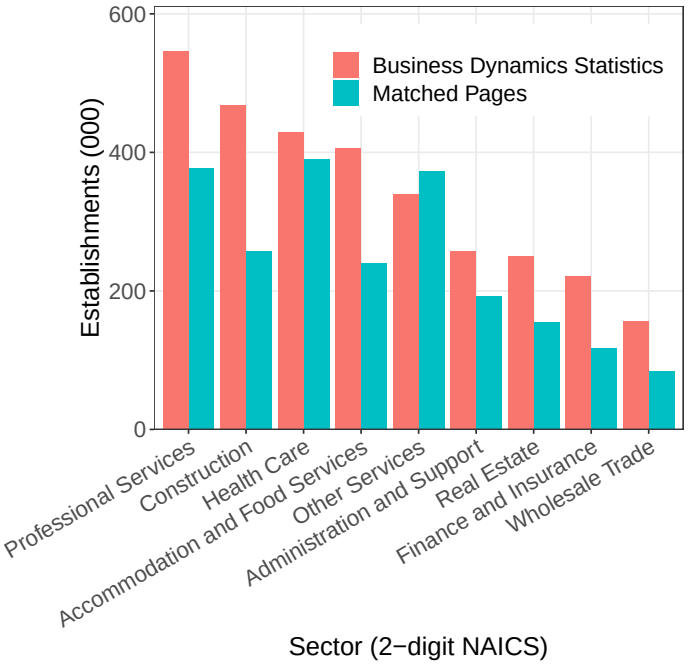
Table 7: Individual-Level Targeting under Estimated Individual-Level Model

	Women in top 50% of Total Impact	Commuting Zones below 50th Percentile of Entrepreneurship Rate	Ratio of (1)/(2)
	(1)	(2)	(3)
Firms Created among Targeted High-Impact Women	0.015	0.012	1.24
Firms Created among Women who were Not Directly Targeted	0.342	0.077	4.44
Total Female-Owned Firms Created	0.357	0.089	4.01

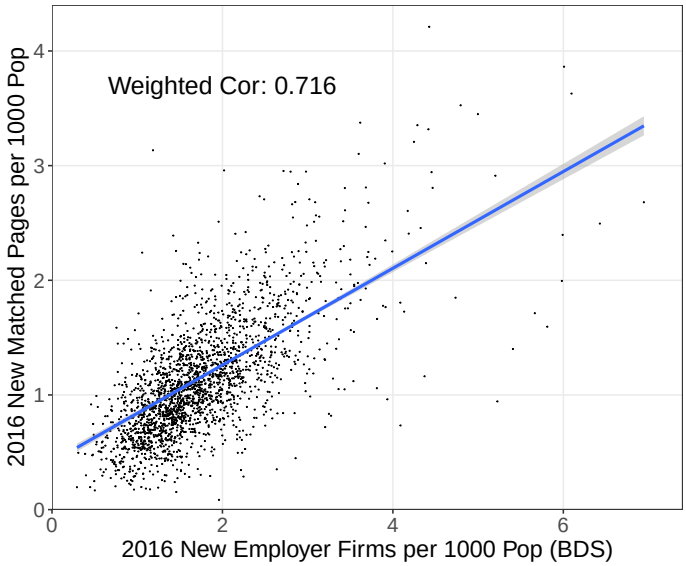
Notes: This table provides estimates of the effect of a hypothetical policy targeted at 100 high-impact women under the extended model, estimated by combining causal estimates from the Movers and College designs under the assumptions described in Section C.1. The policy is assumed to have the same commuting zone-specific direct effect as having a friend move to a high-entrepreneurship commuting zone. Column (1) estimates the direct effect, spillover effect, and total effect from a policy targeted at women whose network is $> 1\%$ owners in commuting zones with a total effect above the median for the target population under the estimated model. Column (2) estimates the effects from a policy targeted at a commuting zone with below-median entrepreneurship rates.

Figure 1: New Matched Pages vs New Employer Establishments in Business Dynamics Statistics

(a) New Establishments between 2015-2020 by 2-Digit NAICS Sector

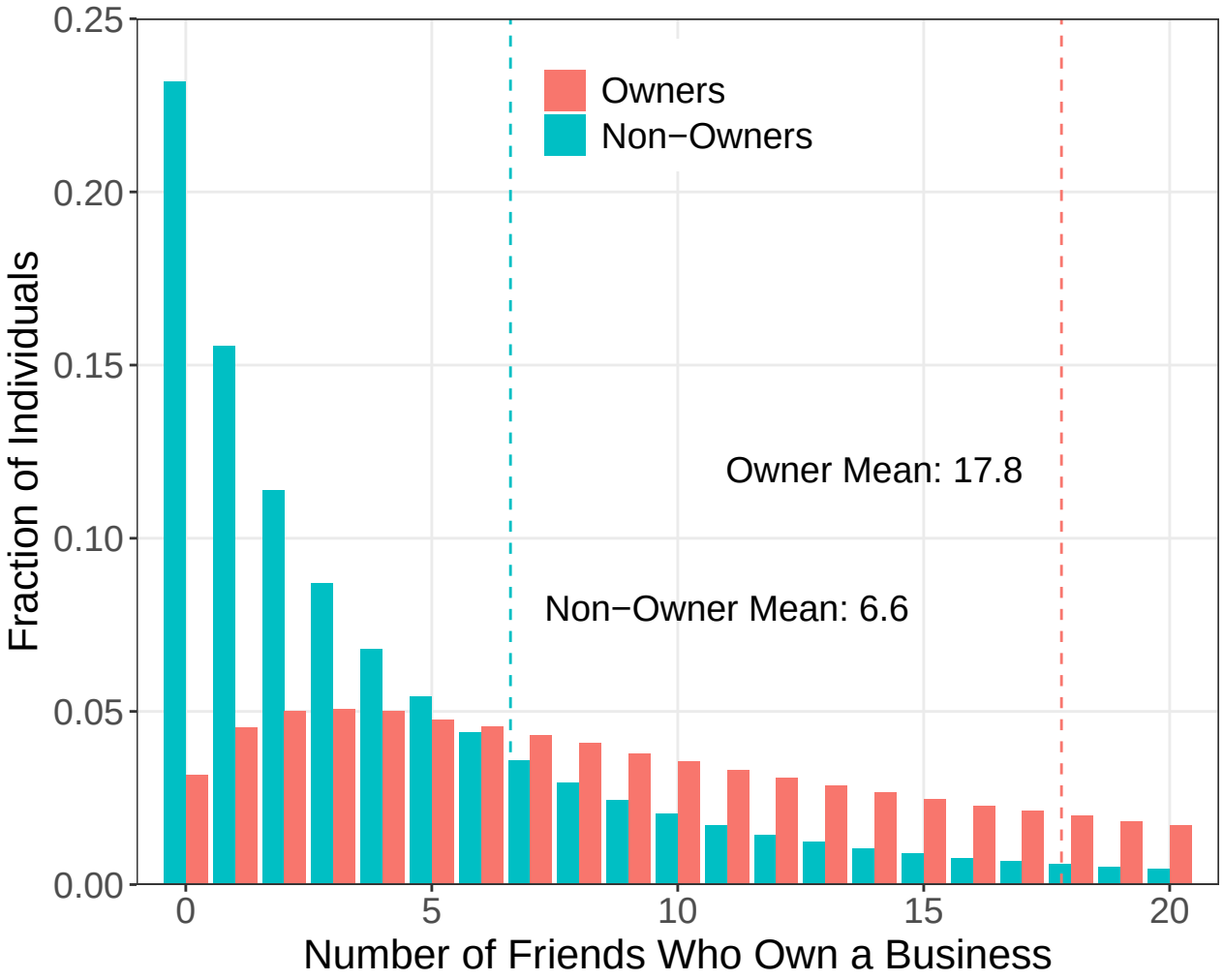


(b) New Establishments per Capita in 2016: County-Level



Notes: This figure compares the number of matched Facebook pages to the number of new employer establishments in the Business Dynamics Statistics. Panel A depicts the number of new establishments between 2015 and 2020 in the 9 largest 2-digit NAICS sectors. Panel B presents a county-level scatterplot of the number of new matched pages and new employer establishments per capita in 2016. The two have a population-weighted correlation of 0.716. Facebook pages are matched to 3rd-party data on U.S. establishments using name, location, and publicly available contact information. The matching process is described in Section II.

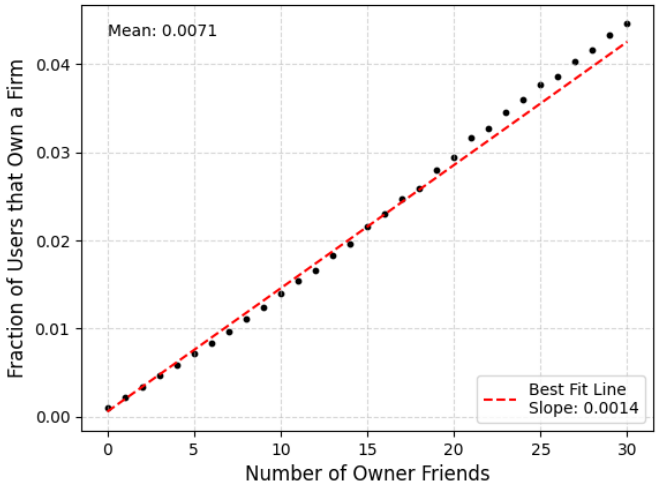
Figure 2: Number of Friends who Own a Business by Owner Status



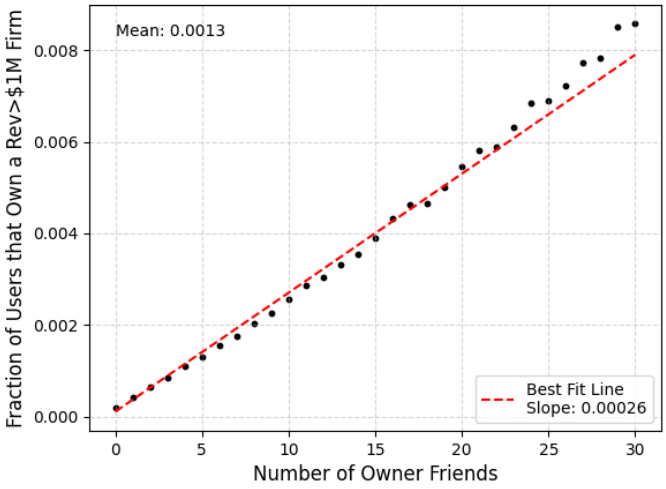
Notes: This figure presents the distribution across individuals in the number friends who own a business, separately for individuals identified as business owners and non-owners. Business ownership among individuals and their friends is identified using self-reported employer and job title on public Facebook profiles. The average owner has roughly 18 friendships with other business owners, while the average non-owner has less than 6 friendships with business owners. We restrict to individuals between the ages of 18-64 with more than 100 friends on Facebook. Sample statistics are described in Table 1.

Figure 3: Owner Status by Number of Owner Friends

(a) Fraction Owner by Number of Owner Friends

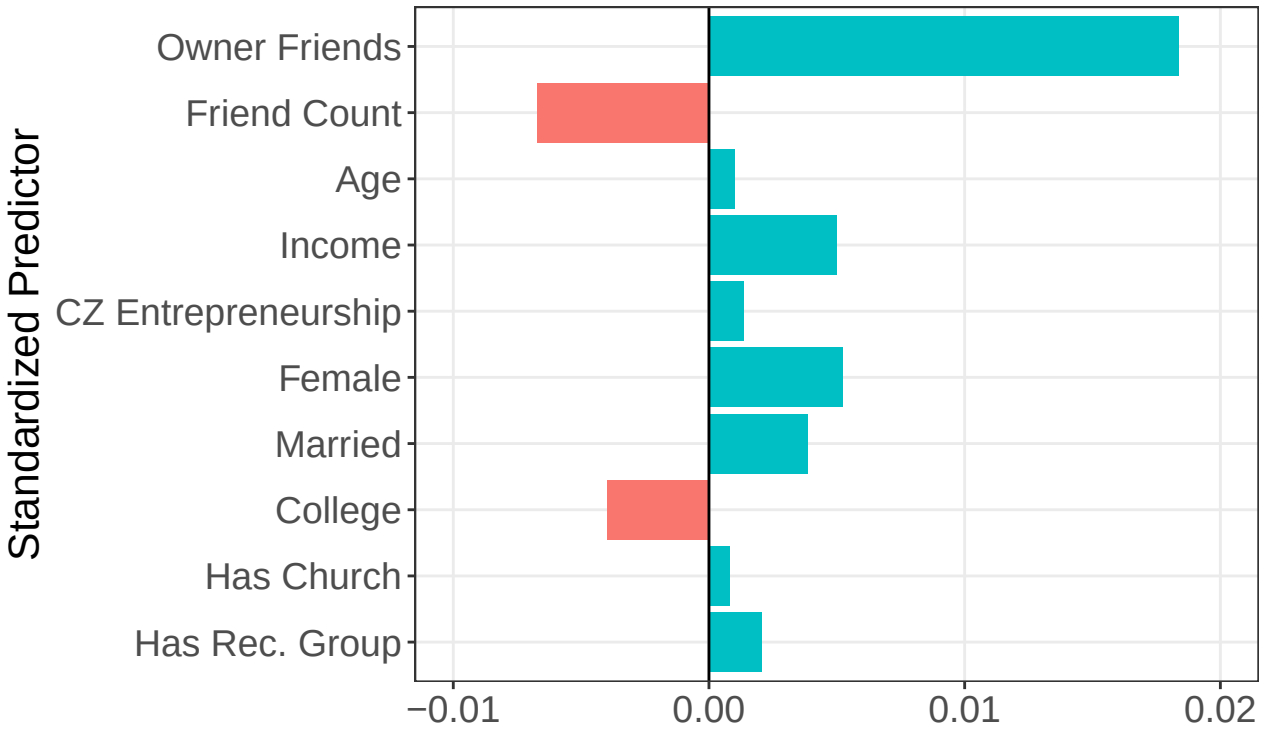


(b) Fraction Owner with Annual Rev. $1 Million by Number of Owner Friends



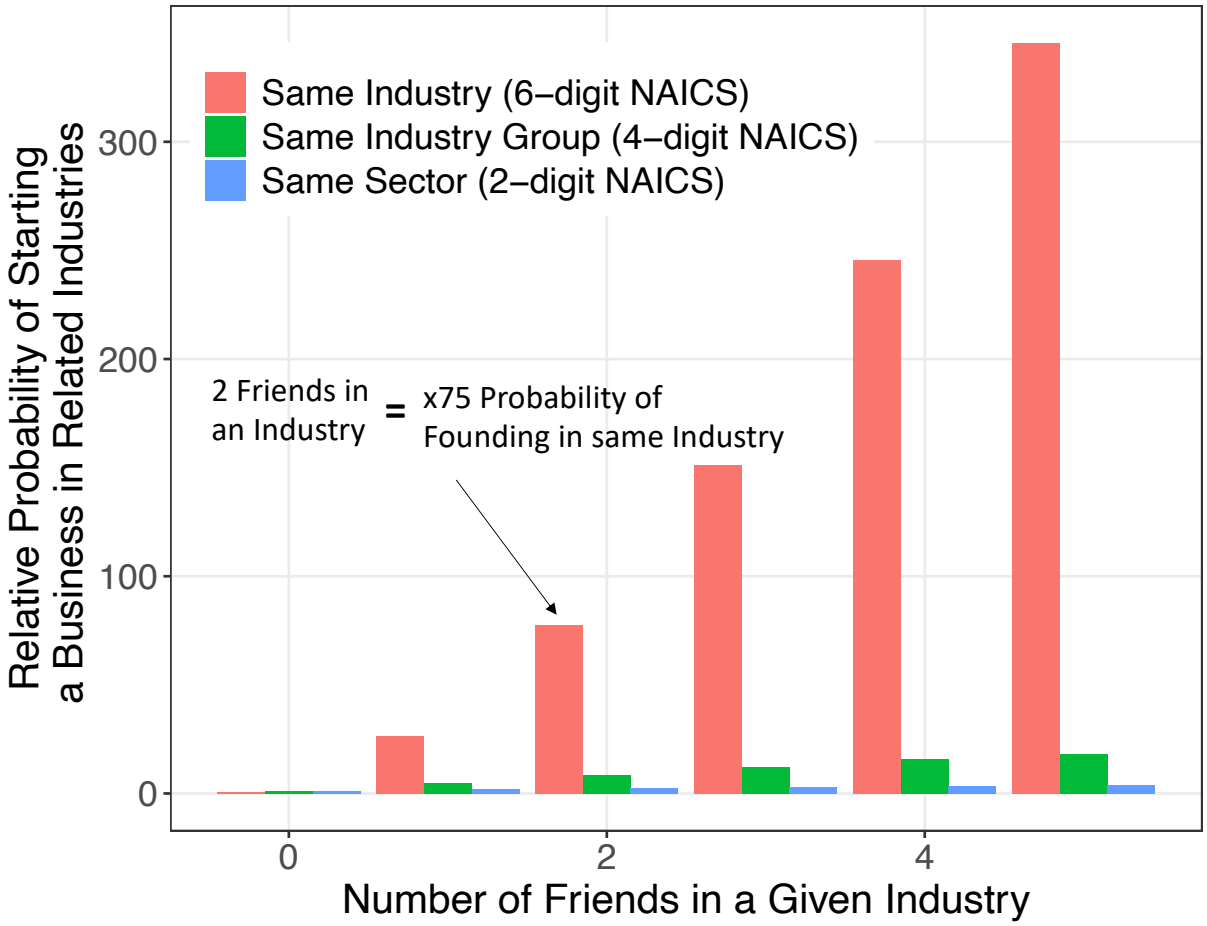
Notes: This figure shows that the number of an individual's friends who own a business predicts whether the individual owns a business among individuals with more than 100 friends. Panel A shows that rates of total business formation increase with the number of friends who own a business. Panel B shows that rates of business formation for businesses with greater than $1 million in self-reported annual revenue increase with the number of friends who own a business. Business ownership among individuals and their friends is identified using self-reported employer and job title on public Facebook profiles. The number of employees is measured using the number of individuals on Facebook who have ever listed the page associated with the owner's business as the individual's employer. Analogous plots in Appendix Figure E9 recreate the analysis using the fraction of friends as the explanatory variable, indicating that the relationship is not driven by the number of friends.

Figure 4: Predictors of Firm Ownership (Multivariate Regression)



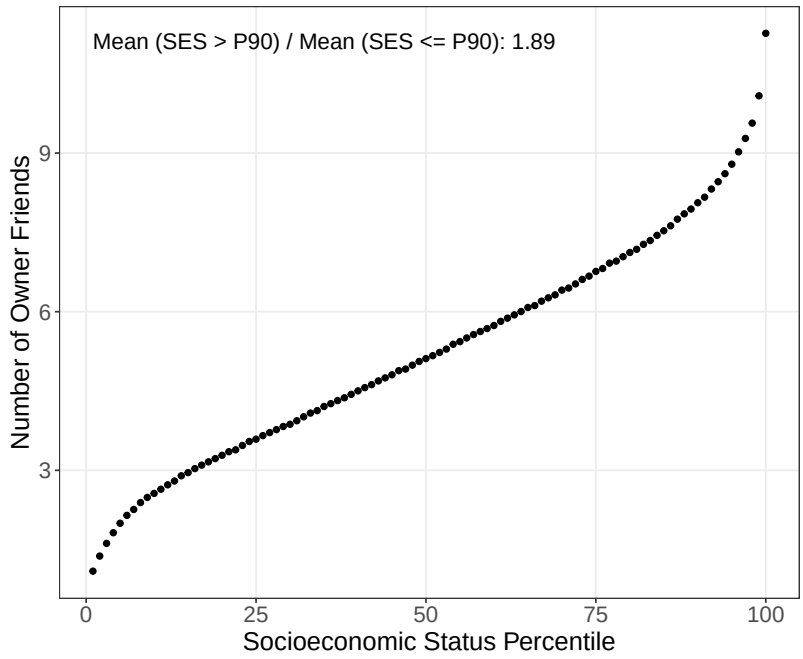
Notes: This figure compares the predictive power of individual-level covariates on an individual's firm ownership. Bars indicate the coefficient from a regression of an indicator for firm ownership on the covariate. Continuous covariates, including the number of owner friends, the number of total friends, age, income, and the entrepreneurship rat in an individual's commuting zone, are standardized across individuals. Dichotomous predictors gender, college attendance, and relationship status are self-reported. Dichotomous predictors 'Has Church' and 'Has Rec. Group' are equal to one if the individual is a member of at least one Facebook group that corresponds to either a church or a recreational group, or likes a page corresponding to each.

Figure 5: Industry Choice by Friend Industry



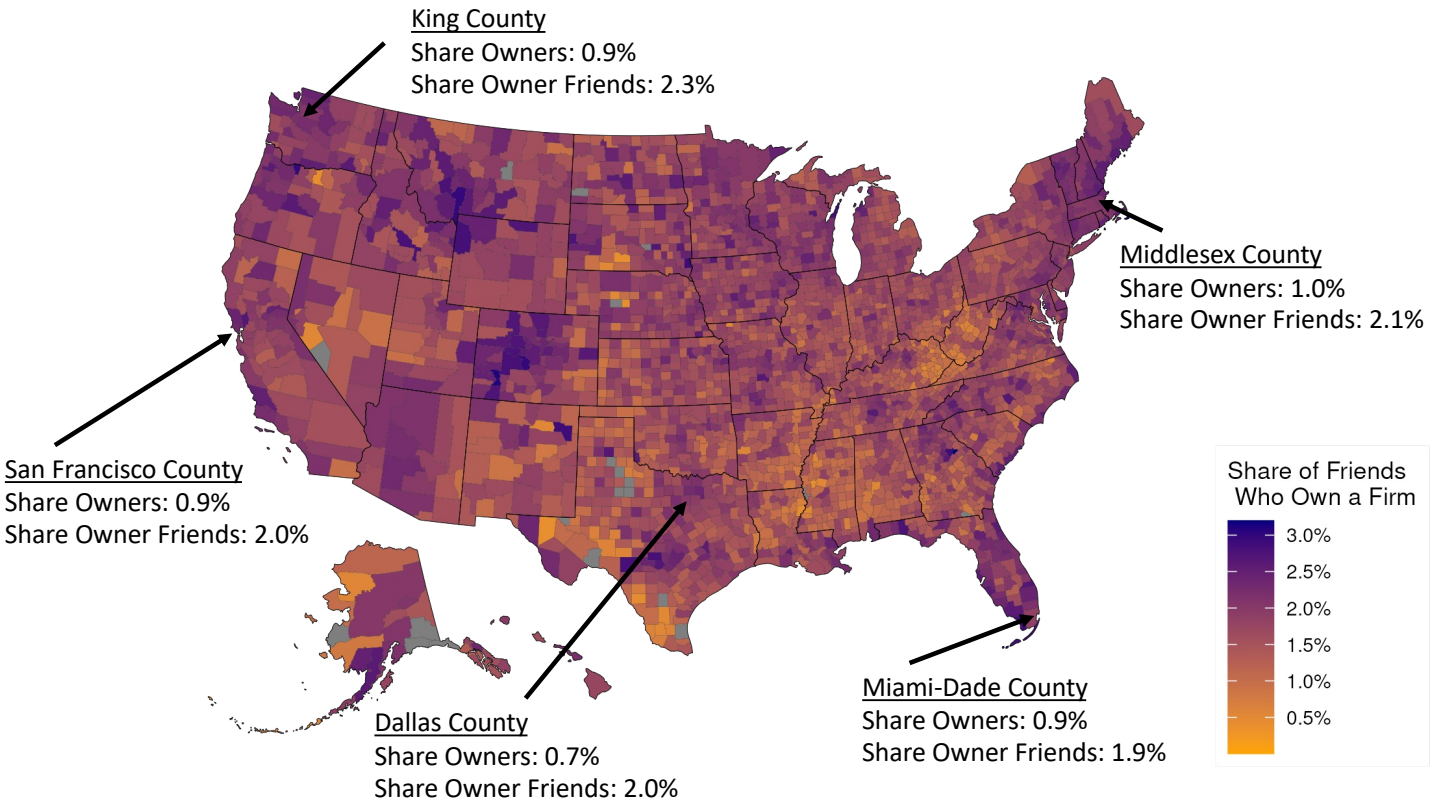
Notes: This figure plots the relative probability of starting a business in a given industry class compared to the baseline probability of starting in that class, separately by the number of friends in the same and related industry classes and estimated among all owners. The relative probability for the number of friends f and class k is calculated by taking the ratio of the probability that an individual who owns a business in k has f friends who own a business in k, over the probability that an individual not in k has f friends who own a business in k. The industry corresponding to an owner's firm is measured at the 6-digit NAICS level in linked Customer Graph data. Details on the linkage procedure are given in Section II. Sample statistics are presented in Table 1.

Figure 6: Average Number of Owner Friends by Socioeconomic Status



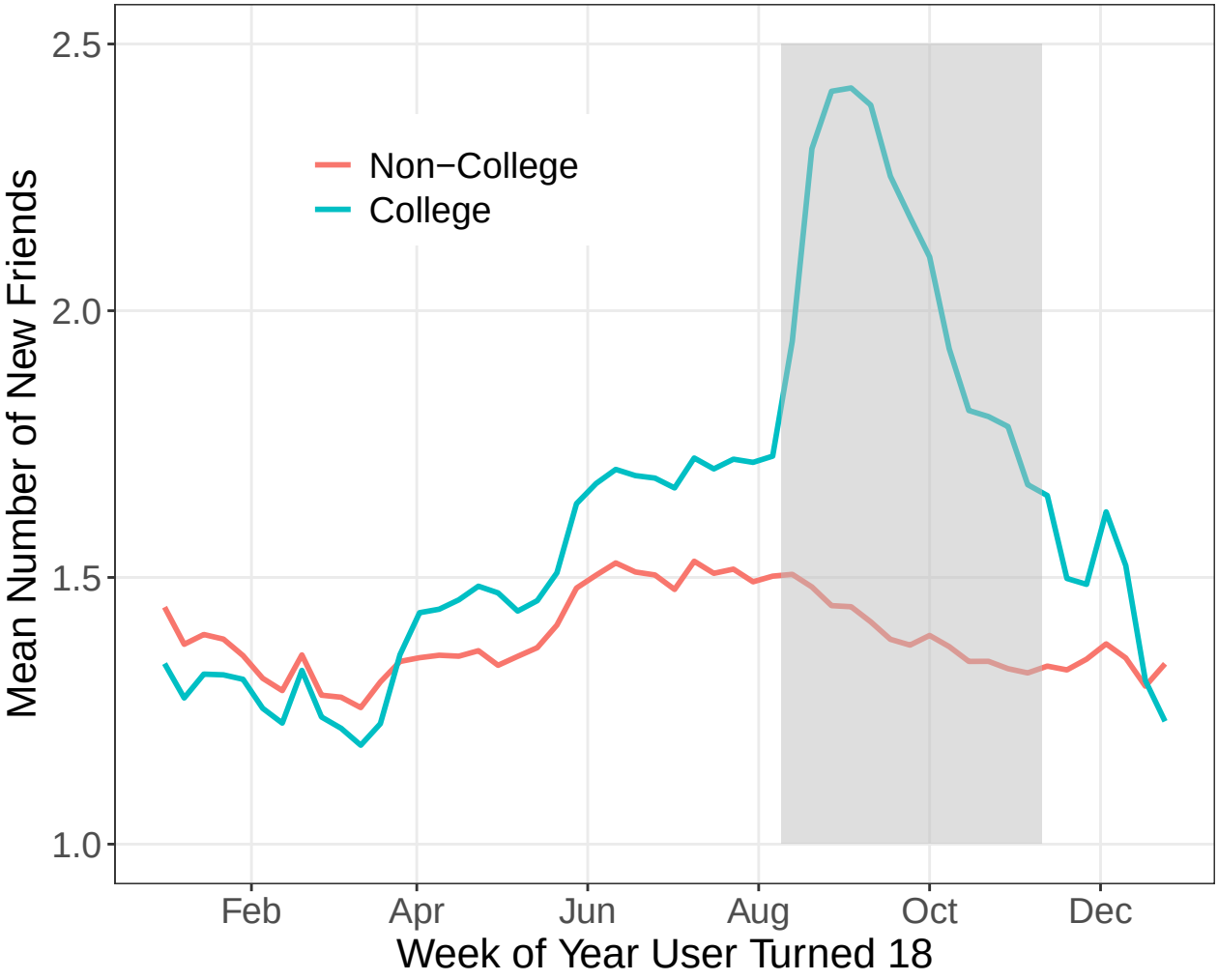
Notes: This figure plots the average number of friends who have started a firm by the income percentile in adulthood. Income measurement is described in Appendix Section A.1.

Figure 7: Average Share of Friends Who Own a Business



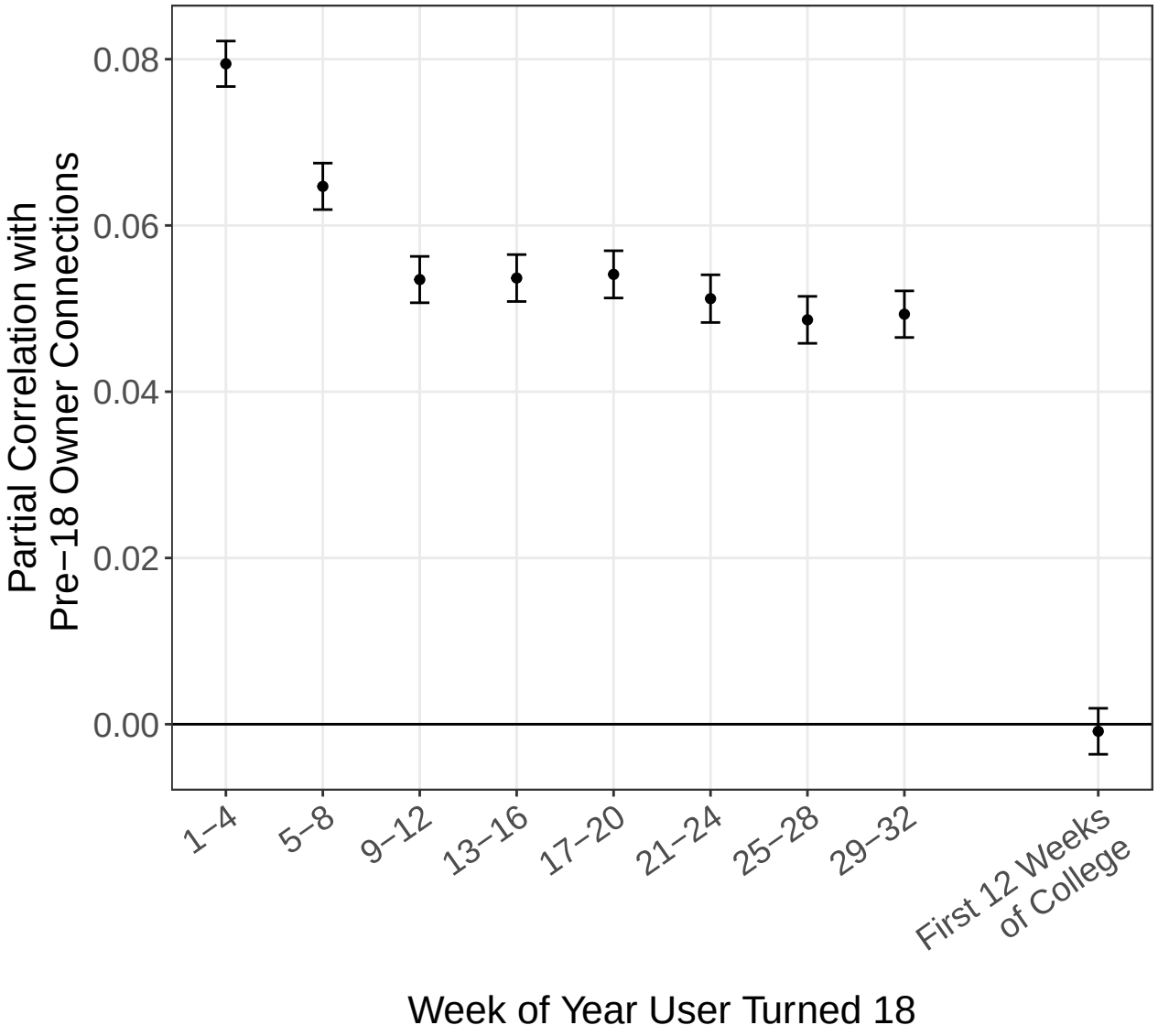
Notes: This figure plots the share of friends that own a business for the average person in each county in the U.S., restricting to counties with more than 30 users. An individual's county is measured using the modal zip code corresponding to their ZIP address in the first 3 weeks of December 2022.

Figure 8: College Friending Rate Spike



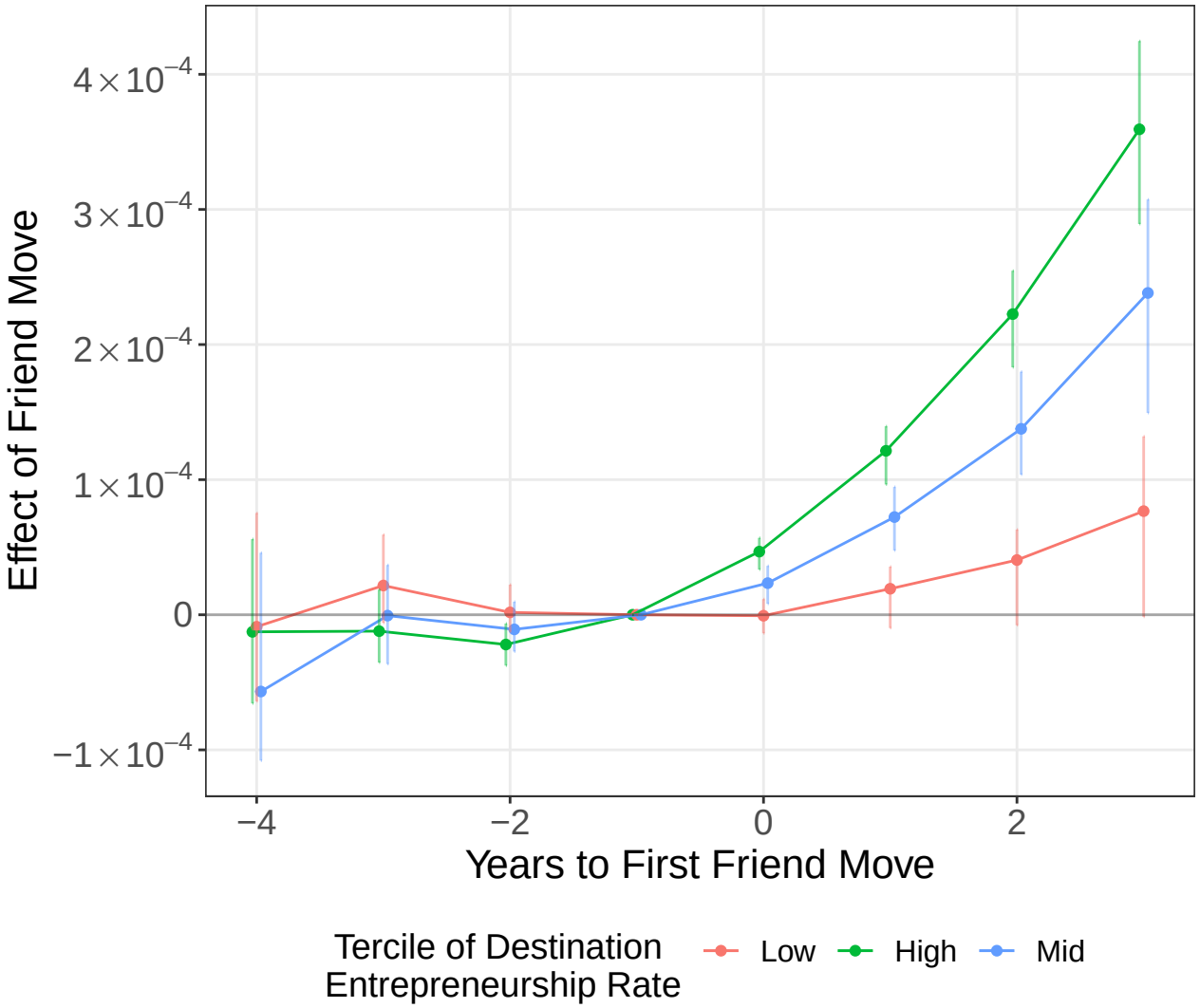
Notes: This figure presents the average number of new friends made in the week an individual turned 18, separately for users that went to college and did not go to college. College attendance is measured using self-reported college on an individual's Facebook profile. The shaded area corresponds to weeks 33-44, mid-August to mid-November, which is when a student entering a college on the semester system would begin orientation activities and classes. Sample statistics representing the sample of college attendees is presented in Table 1.

Figure 9: College Network Formation and Existing Network Composition



Notes: This figure plots coefficients from an individual-level multivariate regression in the sample of college attendees. The outcome in the regression is the standardized number of friendships the individual had with business owners as of January 1 in the year the individual turned 18, restricted to business owners who had founded a business prior to the year the individual turned 18. The independent variables are the standardized number of new 2nd-degree friendships with owners for each period in the year the individual turned 18. New 2nd-degree friendships are calculated by counting the number of owner friends among new friends formed by the individual, where the owner started their firm and formed a friendship with the mutual connection prior to the year that the individual turned 18. The 'First 12 Weeks of College' corresponds to weeks 33-44, which is roughly mid-August to mid-November, and represents only 2nd-degree connections where the mutual friend formed via a mutual friend who was born in the same year and attended the same college as the individual. Further details on the calculation of the independent variables is given in Section III.A.

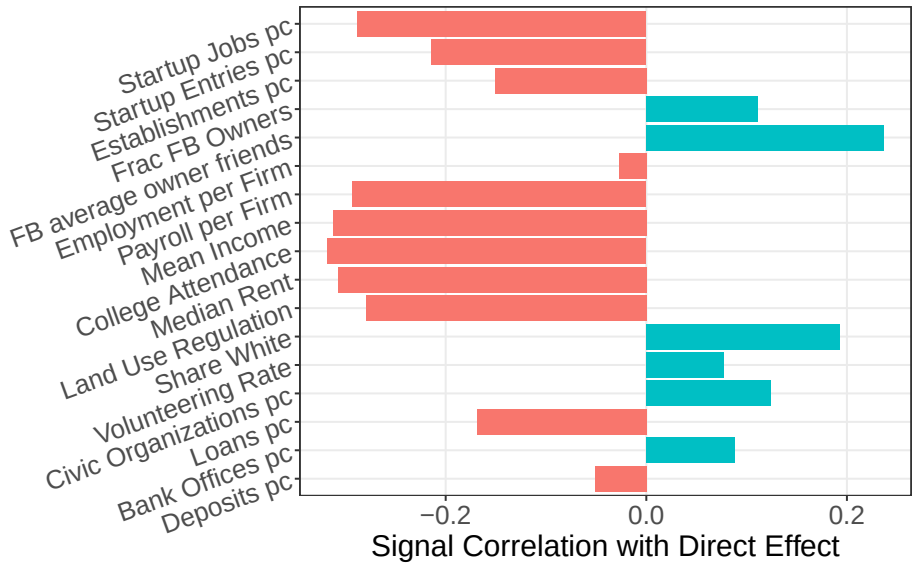
Figure 10: Effect of Friend Migration by Destination Entrepreneurship Rate



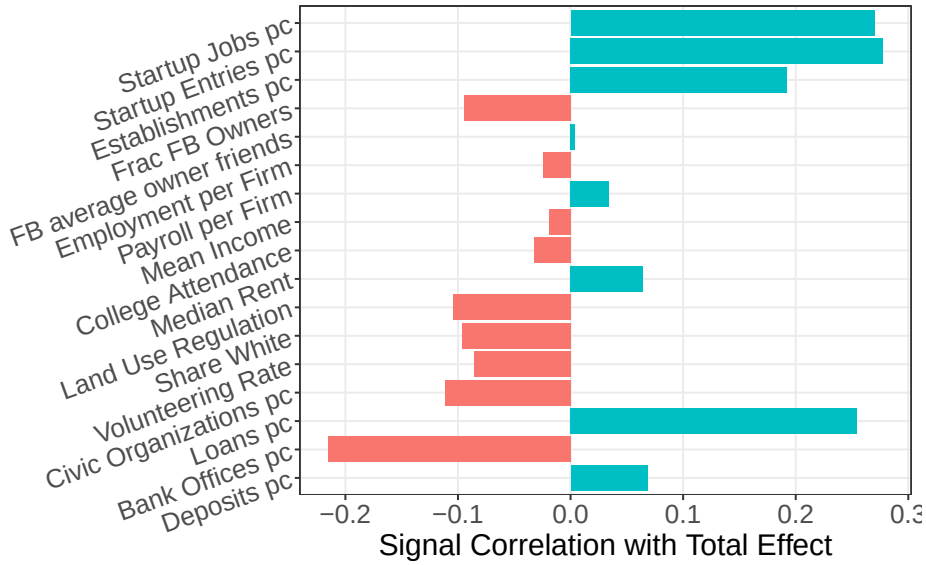
Notes: This figure plots difference in differences estimates from the Movers quasi-experimental design, binned by terciles of the entrepreneurship rate in the mover's destination commuting zone and averaged across all non-mover commuting zones. The outcome is the rate of firm founding among non-movers who are friends with a mover. Moves between 2016 and 2019, inclusive, are used as treatment. Control groups include movers between 2017 and 2020. Firm ownership is measured up to the year prior to the move in the control group. Local entrepreneurship rates used to differentiate between destinations are defined as the average number of new jobs per capita from firms less than 1 year old, measured between 2010 and 2020 in the Business Dynamics Statistics. Pointwise 95% credible intervals are calculated via Bayesian bootstrap with 1000 iterations. Details of the estimation strategy are presented in Section III.B.

Figure 11: Predictors of Combined Networking Productivity and Spillover Production

(a) Direct Effect

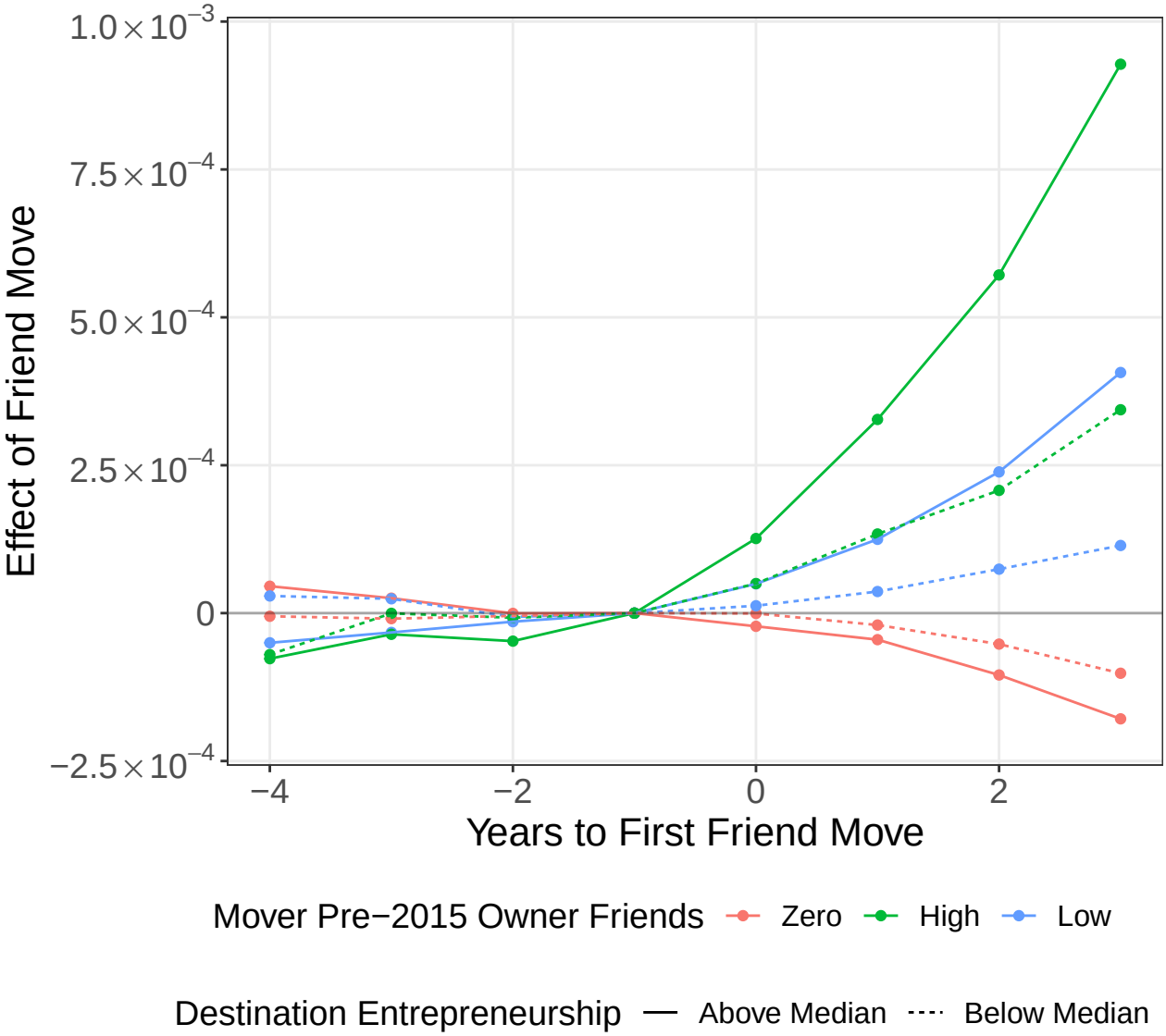


(b) Total Effect



Notes: This figure presents signal correlations between the estimated "total effect" of an intervention in a given commuting zone and commuting zone-level covariates. The total effect is estimated by calculating the direct effect and social multiplier separately from cz-level estimates of networking productivity and spillover production and then combining. Signal correlations are calculated by adjusting the raw correlation by the square root of the signal to noise ratio, which is the portion of the total variance of the cz-level total effect estimates which is not explained by noise. Noise variance is estimated as the average variance of estimates across Bayesian bootstrap samples. Information on the model and estimation approach are presented in Sections IV.B, and results are presented in Section V.B.

Figure 12: Effect of Friend Migration by Destination Entrepreneurship Rate and Mover's Pre-Move Network



Notes: This figure plots difference in differences estimates from the Movers quasi-experimental design, binned by the mover's friendships with entrepreneurs prior to 2015 and by the entrepreneurship rate in the mover's destination commuting zone and averaged across all non-mover commuting zones. The 3 bins of network exposure to business owners are defined by individuals who have no friendship with entrepreneurs, individuals with fewer owner friends than would be expected from random sampling, and individuals with more owner friends than would be expected from random sampling. The outcome is the rate of firm founding among non-movers who are friends with a mover. Moves between 2016 and 2019, inclusive, are used as treatment. Control groups include movers between 2017 and 2020. Firm ownership is measured up to the year prior to the move in the control group. Local entrepreneurship rates used to differentiate between destinations are defined as the average number of new jobs per capita from firms less than 1 year old, measured between 2010 and 2020 in the Business Dynamics Statistics. Details of the estimation strategy are presented in Section III.B.

Appendix A: Supplementary Information on Dataset Construction and Measurement

A.1 Data Construction and Matching

The analysis in this paper requires linking users (and their networks) on the Facebook platform to the firms founded by those users. The linking is performed across three datasets: a set of firm-level data compiled from a variety of sources termed the "Customer Graph", data from Facebook Pages which are used as on-platform representations of establishments, and public individual user profiles.

The linkage is performed by first matching firms in the Customer Graph to Pages using the contact information of the firm, then linking Pages to users via the users' self-reported employment and job title. A schematic summarizing the linkage is presented in Appendix Figure E1.

Below, the construction of each of the datasets and the steps for matching across records are discussed in detail.

A.1.1 Construction of the Customer Graph

Data on firms are collected into a national establishment-level dataset termed the "Customer Graph". The Customer Graph consists of data from the following sources: Dunn and Bradstreet (DnB), ORB companies, Clearbit, Facebook Customer Relationship Management (CRM), Salesforce, Crunchbase, Dow Jones, and LeanData Sales Intelligence (LSI).

The processing and matching between datasets is performed internally by Meta using a custom match algorithm. Records from each source are first deduplicated by matching the standardized data against itself and combining any self-matches into a single row. Each source is then iteratively matched against each of the other sources using the same match algorithm used in the deduplication step. Disagreements are reconciled via Meta's own field-by-field evaluations of data accuracy in each dataset, such that when a given field disagrees in two matched records, the field from the dataset with the highest accuracy (as assessed by Meta) is used. The algorithm chooses randomly between the two records when field-specific accuracy is tied.

The custom matching algorithm applied at each step combines deterministic and probabilistic matching approaches. Records are first matched deterministically based on website, such that two records with the same website are always treated as a match. For records which are not matched deterministically, the name, address, and website are vectorized using a text embedding procedure. Candidate matches are then identified by binning at the state level and applying a proprietary k-nearest-neighbors algorithm on the embedding space within states. Finally, candidate matches are accepted or rejected by applying a Bayesian matching model to the embeddings, with priors and a decision rule calibrated to a set of manual matches performed internally by Meta.

A.1.2 Matching Customer Graph and Establishment Pages

In order to match the Customer Graph data to the establishment pages, we first identify pages that are likely to correspond to establishments by collecting the set of pages that U.S. users self-report as their 'employer' on their public Facebook profile. We then filter these pages to those who self-report their country as 'United States' and self-identify as a 'Local Business' page, which eliminates public and non-profit organizations like churches. We cull pages from the data which do not have a user who administers the page but which has posted in the past 28 days in order to eliminate branches of large organizations and multi-level marketing organizations. We consider only pages which have fewer than 5 self-reported owners, where ownership is measured using the approach described in Section A.1.3.

Pages are then matched to Customer Graph records in two steps. First, the matching algorithm from the Customer Graph merge (described in Section A.1.1) is applied to generate an initial set of matches. We then supplement these matches by manually cleaning the website and phone number reported by the business page, and perform a deterministic merge on the cleaned website and phone number across the entire U.S. For firms matched to multiple pages, firms are de-duplicated and field-level disagreements are reconciled by prioritizing the data from the Customer Graph matching algorithm. If that is unavailable, then fields are reconciled using matches on cleaned website data. Finally, if website data is unavailable, then reconciliation is done using the matches on cleaned phone number data.

A.1.3 Matching Users to Establishment Pages

Firm owners are identified using self-reported 'employer' and 'job title' fields on users' public Facebook profiles. We take as a set of possible firms those pages that have been matched to the Customer Graph as described in Section A.1.2. For the purposes of the baseline analysis, we denote as an owner any user matched to a firm in our data whose position title is fuzzy string matched to 'founder', 'owner', or 'ceo'. In various robustness checks, we restrict the matches to only individuals whose position title is fuzzy string matched to 'founder'. If the position title is not similar to any of these words, then we do not label them as an owner or founder.

The founding date of a firm is identified as the date that the firm was founded from the sources in Customer Graph. The date that firm's page was created on the platform and the date when a user self-identified as an owner are not used in the analysis in order to minimize the chances that any effects we find are driven by on-platform reporting behavior rather than the establishment of a firm.

We then collapse the data to the user-level. When one user is identified as an owner of multiple firms, we use only data from the firm with the earliest founding date as measured in Customer Graph. The data should therefore be interpreted as identifying the first date that an individual founded a firm, and all firm-level characteristics correspond to the first firm.

A.2 Measurement of User and Firm Characteristics

A.2.1 Location Measurement for Users and Firms

Annual home location is measured at the user level using daily zip code data collected when users are using the Facebook platform and assembled in Koenen and D. Johnston 2024. First, using this daily zip code data, we aggregate to the user-year level by taking the modal zip code observed for each U.S. user between December 1 and December 20 of each year starting in 2015. Zip codes are then aggregated up to the commuting zone by year-level to produce a panel of locations for each user. A user is identified as having moved in the year t if we observe a different commuting zone between December 1st and 20th of year t compared to the same period in year $t-1$.

Page-level locations are measured at the zip code level and aggregated up to the county or commuting zone level depending on the analysis. We use zip codes that are either self-reported by the page or inferred by Meta, rather than locations measured in the Customer Graph, in order to locate firms based on their physical location rather than the area they were registered.

To validate our measurement of user-level migration events, we use the fact that users may 'like' or 'follow' geographically-linked pages. On the platform, 'likes' and 'follows' allow users to publicly or privately elect to prioritize viewing content produced by the target page. For each user, we construct a user by commuting zone by year-level panel of total likes or follows of local business pages who we identifying as residing in a given commuting zone. Appendix Figure E18 displays the average number of likes and follows in the origin vs. destination commuting zone for a user that we identify as having moved in the reference year 0. Appendix Figure E18 displays the averages disaggregated by year of move.

A.2.2 Socioeconomic Status

Measurement of individual socioeconomic status follows the approach and data in (Bailey, D. M. Johnston, et al. 2025). To estimate SES, a large survey was conducted of users in which they report income buckets. A predictive model was then trained on a variety of individual-level features including geographic location, on-platform behavior, device characteristics, and self-reported demographic information. We use the output of this predictive model, applied to the whole sample for which enough features are measured, as our estimate of socioeconomic status.

Appendix B: Extension: Welfare Model

B.1 Setup

B.1.1 Preferences

There is a unit mass of individuals indexed by i. Individuals have quasi-linear utility

$$U = x_0 + u(x_1) \tag{19}$$

where the competitively-produced x_0 is numeraire and sets the wage. We denote the aggregate inverse demand function for the good x_1 as $P(Q)$ and will ignore income effects throughout.

B.1.2 Network Formation

Individuals are arranged in an unweighted undirected network G without self-links. Define $g_{i,j} = 1$ if i and j agree to be friends, and $g_{i,j} = 0$ if not. Each dyad that forms produces a surplus $s_{i,j}$ for each individual which depends on the types of the individuals, denoted α.

$$s_{i,j} = [\phi(\alpha_i, \alpha_j) + \mu_{i,j}]\, g_{i,j} \tag{20}$$

where $\phi(a, b) = \phi(b, a)$ and the error $\mu_{i,j}$ is an iid random variable which is symmetric such that $\mu_{i,j} = \mu_{j,i}$.

B.1.3 The Firm Formation Decision

Each individual i supplies one unit of labor inelastically, which can be used to either earn the exogenous wage equal to w or run a profit-maximizing firm that sells q_i units of the homogenous good x_1 at cost $c(q_i)$, from which the individual receives all operating profits. Let Y_i be an indicator variable which describes whether an individual starts a firm and N be the number of firms, such that $N \equiv \int_0^1 Y_i \, di$. An individual's production q_i should they enter the market may in principle be a function of the decisions of every potential entrepreneur such that we can write q_i as a function $q_i(\vec{Y}_{-i})$ for $\vec{Y}_{-i}$ the vector of entry decisions of all agents other than i. Let $Q(\vec{Y}) \equiv \int_0^1 q_i(\vec{Y}_{-i})di$ be the total quantity of good x_1 produced by the economy.

Individual i has entrepreneurial human capital E_i, which reduces the fixed and sunk cost of production. The cost of firm entry is equal to $\psi(E_i)\kappa$ for $\kappa > 0$ and $\psi(\cdot)$ a decreasing convex function bounded below at zero. Human capital E_i depends on the human capital of other agents in the network according to Equation 9. We consider a static version of the model where the equilibrium is achieved at each time t, and therefore drop the time subscripts.

B.1.4 Timing

Decisions proceed in two steps. In the first step, individuals myopically form a network by maximizing the surplus from their links. In the second step, individuals then maximize utility given their network. Individuals do not account for the effect of their network formation decisions on their consumption decisions.

Step 1 : i maximizes $\sum_j s_{i,j}$

Step 2 : i maximizes utility from consumption

B.1.5 Competitive Externalities

To characterize the relationship between the free-entry equilibrium number of firms N^e and the socially optimal number of firms N^* (given the nature of post-entry competition) we require 4 assumptions, adapted from Mankiw and Winston (1986) with a modest adaptation that allows for heterogeneous human capital among entrepreneurs:

Assumption 1. $q_i(\vec{Y}_{-i}) = q_i(N) \ \ \forall i$

Assumption 1 says entry decisions may only produce spillovers through the number of total firms, which is continuous. It follows from assumption 1 that $Q(\vec{Y}) = Q(N)$. We can then state the following assumptions:

Assumption 2. $Q(N) > Q(N') \ \forall \ \ N > N'$ and $\lim_{N\to\infty} Q(N) = M < \infty$

Assumption 3. $q_i(N) < q_i(N') \ \forall \ i, \ N > N'$

Assumption 4. $P(Q) - c'(q_i) \geq 0 \ \forall i$

Assumption 2 says that aggregate output rises with the number of firms and is bounded, which guarantees that the free-entry equilibrium number of firms is well-defined. Assumption 3 says that an operating firm's output falls with the number of firms in the economy. Assumption 4 requires that equilibrium price is not below marginal cost.

B.2 Welfare Results

We start by showing that there is generically excess firm entry in this economy.

Proposition 1. *The number of firms in a free-entry equilibrium, N^e is weakly greater than the socially optimal number of firms N^*. The result holds strictly when Assumption 3 is a strict inequality.*

Proof. The proof is presented in Appendix Section B.4 □

The result is similar to established results with minor changes that allow for heterogeneous human capital (Mankiw and Whinston 1986). The intuition is that when firm entry reduces the quantity that incumbent firms can bring to the market, typically termed a "business-stealing externality", firms fail to internalize the negative effects of their entry on aggregate production.

We next show that an intervention which reduces entry costs will increase social welfare despite the presence of business-stealing externalities:

Proposition 2. *The derivative of total welfare with respect to the policy D is proportional to $\frac{d\mathbb{E}[Y_i|D_i]}{dD_i}$ when $\psi(\cdot)$ is linear.*

Proof. The proof is presented in Appendix Section B.5 □

The intuition is that when firms are profit maximizing and firm entry only produces spillovers via the total number of firms, then the only first-order effect of D from a welfare perspective is to reduce the fixed costs facing inframarginal firms (either directly or through network spillovers). Under linearity of the cost function, the combined direct and indirect effects of D_i are proportional to the direct effect of the intervention on each individual, which can be estimated in a discrete choice estimation procedure.

B.3 Derivations from the Welfare Model not tied to a Specific Proposition

B.3.1 Network Formation

Because individuals maximize the sum of surplus from their network and because surplus is symmetric for both members of the dyad, $G_{i,j}$ is set such that the link forms iff the associated surplus is positive:

$$G_{i,j} = 1\{s_{i,j} > 0\}$$

The probability that a link forms between two individuals is therefore

$$P(G_{i,j} = 1|\alpha_i, \alpha_j) = F_\mu(\phi(\alpha_i, \alpha_j)) \tag{21}$$

for $F_\mu(\cdot)$ the CDF of μ_{ij}.

B.3.2 Human Capital Diffusion

Recall that at each point in time human capital E_i is modeled as follows in Equation 9:

$$E_i = \alpha_i + \delta D_i + \gamma \sum_j g_{ij} E_j$$

where α_i is an individual's endowment of human capital, $D_i \in \{0, 1\}$ is a potential intervention, and g_{ij} is a binary variable indicating where there is a link between i and j.

We can represent this relationship in matrix form by stacking each individual to form a vector:

$$\vec{E} = \vec{\alpha} + \delta \vec{D} + \gamma G \vec{E}$$

for $E = [E_1, E_2, ...]'$ and G the adjacency matrix.

Solving this equation gives

$$\vec{E} = (I - \gamma G)^{-1} \left(\vec{\alpha} + \delta \vec{D} \right)$$

where $(I - \gamma G)$ is invertible when $\gamma < \frac{1}{||G||}$, the reciprocal of the largest eigenvalue of the adjacency matrix.

Leaving aside for a moment that we don't observe E_i directly, this equation presents a computational hurdle in our environment because solving it requires inverting a matrix that is on the order of 100 million by 100 million.

However, we can exploit a matrix identity to both develop intuition for the equation and to motivate a suitable estimator. Note that:

$$(I - \gamma G)^{-1} = \sum_{n=0}^{\infty} \gamma^n G^n = I + \gamma G + \gamma^2 G^2 + ...$$

where the i, j element of matrix G^n can be interpreted as the number of *paths* of length n between nodes i and j. At an individual level, this gives the equation:

$$E_i = \alpha_i + \gamma T_i + \gamma \sum_j g_{ij}(\alpha_j + \delta D_j) + \gamma^2 \sum_j \left[G^2\right]_{ij} (\alpha_j + \delta D_j) + ...$$

where $\left[G^n\right]_{ij}$ is the i, j element of G^n (interpreted as the number of paths of length n between i and j).

Note that the equilibrium is defined without reference to barriers to entry $\beta_{c(i,t)}$. Barriers to entry simply map the unobserved human capital equilibrium, defined by the vector $\vec{E^*}$ to observable entry decisions. We consider perturbations of D_i around this equilibrium, holding constant other parameters, and interpret the results as partial equilibrium effects.

One consequence of this approach is that the linear functional form does not matter to the economics of the problem; it is simply a convenient reduced-form statistical model, motivated by the linearity in levels observed in the descriptive exercises. The equilibrium solution and perturbation exercise can be recast in a generalized linear model where $\mathbb{E}[Y_i|E_i] = B(E_i)$ for some given continuous link function $B(\cdot)$.

B.4 Proof of Proposition 1

We prove Prop 1 link under more general conditions that allow the marginal cost of production to vary with human capital. In a free entry equilibrium, entrepreneurs with high human capital E_i will enter until the marginal entrant has human capital E^e such that

$$\pi(E^e) = w \tag{22}$$

Because utility is quasilinear, social welfare can be expressed as the sum of consumer and producer surplus:

$$W(N) = \int_0^Q P(s)ds - \int_{E_m}^{\infty} \psi_c(E)c(q) - \psi_k(E)\kappa - w \; dF_E(E) \tag{23}$$

Where E_m is the marginal entrepreneur's human capital given N firms. That is, $E_m = F_E^{-1}(1-N)$ for $F_E^{-1}(\cdot)$ the quantile function of E. The quantity is maximized at N^* such that

$$W'(N^*) = 0 \tag{24}$$

The derivative of the social welfare function wrt the number of firms is

$$\begin{aligned} W'(N) = P(Q)\frac{dQ}{dN} - \int_{E_m}^{\infty} \psi_c(E)c'(q)\frac{dq}{dN}dF_E(E) \\ + \left[\psi_c(E_m)c(q(E_m)) + \psi_k(E_m)\kappa - w\right]\frac{dE_m}{dN} \end{aligned}$$

Where $P(Q)\frac{dQ}{dN} = P(Q)\left[\int_{E_m}^{\infty}\frac{dq}{dN}dF_E(E) - q(E_m)\frac{dE_m}{dN}\right]$. we can group terms to express the derivative of social welfare as

$$W'(N) = -\frac{dE_m}{dN}\left(\pi(E_m) - w)\right) + \int_{E_m}^{\infty}\left[P(Q) - \psi_c(E)c'(q)\right]\frac{dq}{dN}dF_E(E) \tag{25}$$

The second term is strictly negative if Assumption 3 holds strictly and is zero if Assumption 3 is an equality. Because $-\frac{dE_m}{dN}$ is strictly positive, this implies that at the social optimum the marginal firm's profits are weakly greater than the entrepreneur's outside option $\pi(E^*) \geq w$.

However, since we know that the free entry equilibrium condition is satisfied when marginal profits equal the wage, it will be sufficient to show that the marginal firm's profits are decreasing in the number of firms.

Taking the derivative of the marginal firm's profits and grouping terms gives:

$$\begin{aligned} \frac{d\pi(E_m)}{dN} =& qP'(Q)\frac{dQ}{dN} + \left[P(Q) - \psi_c(E_m)c'(q)\right]\frac{dq}{dN} \\ & - \left[\psi_c'(E_m)c(q) + \psi_k'(E_m)\kappa\right]\frac{dE_m}{dN} \end{aligned} \tag{26}$$

Under Assumptions 1-3 each term is weakly negative. It must therefore be the case that $E^* \geq E^e$, which implies $N^e \geq N^*$.

B.5 Proof of Prop 2

Consider consumer and producer surplus separately. With quasi-linear utility, consumer surplus is equal to

$$W_{cons} = \int_0^Q P(s) - p^e ds \tag{27}$$

for $p^e = P(Q)$ the equilibrium price. The derivative wrt D_i is

$$\frac{dW_{cons}}{dD_i} = \left[P(Q) - p^e\right]\frac{dQ}{dD_i} - Q\frac{dp^e}{dD_i} = -Q\frac{dp^e}{dD_i} \tag{28}$$

Producer surplus is

$$W_{prod} = \int_{E^e}^{\infty} p^e q - c(q) - \psi(E)\kappa - w \; dF_E(E) \tag{29}$$

Because $p^e q(E^e) - c(q(E^e)) - \psi(E^e)\kappa = w$ by the free entry condition, the derivative wrt D_i is simply the integral of the derivative of the indirect utility of the inframarginal entrepreneurs, which is equal to the partial derivative (holding constant quantity produced) by the envelope theorem:

$$\frac{dW_{prod}}{dD_i} = \int_{E^e}^{\infty} \frac{dV(E)}{dD_i} dF_E(E) = \int_{E^e}^{\infty} \frac{\partial V(E)}{\partial D_i} dF_E(E)$$

Because utility is quasi-linear, this quantity is equal to the partial derivative of profits wrt D_i, which is:

$$\int_{E^e}^{\infty} \frac{\partial V(E)}{\partial D_i} dF_E(E) = \int_{E^e}^{\infty} \frac{\partial \pi(E)}{\partial D_i} dF_E(E) \tag{30}$$

We can break up the partial derivative of profits into the derivative of revenue minus the derivative of costs and re-index the derivative of costs to be across people, which is a function of own treatment, the treatment to first degree connections, and so on:

$$\int_{E^e}^{\infty} \frac{\partial \pi(E)}{\partial D_i} dF_E(E) = \int_{E^e}^{\infty} q \frac{dp^e}{dD_i} - \int_0^1 Y_i \left(\delta \sum_{nj} \gamma^n \left[G^n \right]_{ij} \psi'(E_j) \kappa \right) di$$

Adding the derivatives of consumer and producer surplus gives

$$\frac{dW_{cons}}{dD_i} + \frac{dW_{prod}}{dD_i} = -\delta\kappa \int_0^1 Y_i \sum_{nj} \gamma^n \left[G^n \right]_{ij} \psi'(E_j) \, di$$

because $\int_{E^e}^{\infty} q \frac{dp^e}{dD_i} \, dF_E(E) = Q \frac{dp^e}{dD_i}$.

Since $\psi'(\cdot)$ is negative and γ is positive, the integral term must be negative, so the whole expression has the same sign as δ.

In the simple case where $\psi(E) = a + bE$ for $b < 0$, the change in total surplus wrt D_i becomes

$$-b\delta\kappa \int_0^1 Y_i \sum_{nj} \gamma^n \left[G^n \right]_{ij} di$$

To relate this quantity to the data, note that the probability that an individual owns a firm is equal to

$$P(Y_i = 1 | D_i) = P(p^e q^e - c(q^e) - a - b\kappa C_i - \gamma b \kappa D_i > w)$$

for $D_i \perp C_i \equiv E_i - \gamma D_i$ (ignoring self-links), where $\{p^e q^e, c(q^e), w\}$ are all constant in partial equilibrium. A discrete-choice estimating framework using a randomly-assigned D_i as the independent variable and individual-level firm formation as the outcome can then directly identify $-\beta b \kappa$, which is proportional to the welfare object of interest.

Appendix C: Modeling and Estimating the Movers Design

The Movers design estimates a set of pairwise DiDs around the first friend moving to destination d among non-movers living in city n. In this section, we derive the relationship between the Movers design and the treatment effect of interest τ.

Assume that the total human capital endowments in the network of person i in city $c(i,t)$ grows according to the equation:

$$\frac{d}{dt} \left(K_{i,t} A_{i,t} \right) = a_{c(i,t)} U_i \cdot \frac{d}{dt} K_{i,t} \tag{31}$$

where

$$\frac{d}{dt}K_{i,t} = R_{i,t}\phi_{c(i,t)} \tag{32}$$

for $R_{i,t}$ the observed rate at which i makes friends at time t, a_c the unobserved city-specific multiplier that scales an unobserved individual-specific preference for friends' human capital U_i, and ϕ_c the city-specific multiplier that scales the observed friending rate.

We model the observed friending rate $R_{i,t}$ as itself a function of characteristics of the individual's location, parameterized by ψ_c, and an unobserved individual-level term V_i.

$$R_{i,t} = \psi_{c(i,t)}V_i \tag{33}$$

For a person who has never moved and was born at time $t_{i,0}$, Katz-Bonacic centrality is equal to:

$$K_{i,t} \equiv \phi_{c(i,t)}\psi_{c(i,t)}\int_{t_{i,0}}^{t} V_i = \phi_{c(i,t)}\psi_{c(i,t)}V_i(t - t_{i,0}) \tag{34}$$

We can now rewrite the treatment effect of interest $\tau_{i,t}$ from Expression 11 the effect as

$$\tau_{i,t} = \delta\beta_{c(i,t)} + \delta\phi_{c(i,t)}\psi_{c(i,t)}V_i(t - t_{i,0})B_{i,t} \tag{35}$$

and the average effect for permanent residents of city $c(i,t) = c$ at age $t - t_{i,0}$ who form friendships at rate $R_{i,t} = x$ as

$$\begin{aligned}&\mathbb{E}\left[\tau_{i,t}|c(i,t) = c, R_{i,t} = x, t - t_{i,0}\right] \\ &= \delta\beta_c + \delta\phi_c(t - t_{i,0})x\mathbb{E}\left[B_{i,t}|c(i,t) = c, R_{i,t} = x, t - t_{i,0}\right]\end{aligned} \tag{36}$$

For a person j who moves to CZ d and is friends with i who lives in CZ n, the post-move trend in entrepreneurship for person i estimates the following quantity:

$$\gamma a_d\phi_d\psi_d\beta_n\mathbb{E}\left[U_jV_j|c(i,t) = n, c(j,t) = d\right]$$

Note that γ multiplies the entire expression because all effects on person j's network are attenuated by γ in person i's network when i and j are friends.

Under the model, conditioning on the destination $c(j,t) = d$ and the observed rate of post-move friend formation at $R_{j,t} = x$ holds constant the individual-level component of friending V_j such that:

$$\begin{aligned}&\gamma a_d\phi_d\psi_d\beta_n\mathbb{E}\left[U_jV_j|c(i,t) = n, c(j,t) = d, R_{j,t} = x\right] \\ &= \gamma a_d\phi_d\beta_n x\mathbb{E}\left[U_j|c(i,t) = n, c(j,t) = d, R_{j,t} = x\right]\end{aligned} \tag{37}$$

Given a quantity proportional to $a_d\mathbb{E}\left[U_j|c(i,t) = n, c(j,t) = d, R_{j,t} = x\right]$ with constant of proportionality $\overline{\alpha}$, which we provide a method to estimating in the next section, we can divide out the average level of human capital from the trend to arrive at an nonmover-location-by-destination quantity reflecting the network growth rate in the destination (evaluated at a level x), $\phi_d x$, and the barrier to entry in the nonmover's location, β_n.

$$\text{CZ-pair Effect} = \overline{\alpha}\gamma\phi_d\beta_n x \tag{38}$$

We define the "Spillover Production" (SP) as the expectation of Expression 38 over nonmover location, which is:

$$SP_d(x) = \overline{\alpha}\gamma\phi_d x\mathbb{E}_i\left[\beta_{c(i,t)}|c(j,t) = d, R_{j,t} = x\right] \tag{39}$$

where the term $\mathbb{E}_i\left[\beta_{c(i,t)}|c(j,t)=d,R_{j,t}=x\right]$ reflects the average barriers to entry faced by *friends* of individuals living in city d (which will in general be different from β_d).

We assume that the average barriers to entry faced by friends of movers to a city d reflects the path-weighted average of barriers to entry among the extended network of residents of d, meaning that

$$\mathbb{E}\left[B_{i,t}|c(i,t)=d,R_{i,t}=x\right]=\mathbb{E}\left[\beta_{c(i,t)}|c(j,t)=d,R_{j,t}=x\right] \tag{40}$$

Together, Equations 39 and 40 imply

$$SP_d(x)=\overline{\alpha}\gamma\phi_d x\mathbb{E}_i\left[B_{i,t}|c(i,t)=c,R_{i,t}=x\right] \tag{41}$$

We analogously define the "Networking Productivity" (NP) as the expectation of the CZ-pair effect in Expression 38 over a fixed set of destinations $\overline{\mathcal{C}}$, which is:

$$NP_n(x)=\overline{\alpha}\gamma\beta_n x\mathbb{E}\left[\phi_c|c\in\overline{\mathcal{C}}\right] \tag{42}$$

where the term $x\mathbb{E}\left[\phi_c|c\in\overline{\mathcal{C}}\right]$ reflects the average growth rate of the Katz centrality of residents of cities $\overline{\mathcal{C}}$ who make friends at rate x. In estimation, we will define $\overline{\mathcal{C}}$ to be cities in the top population-weighted tercile of startup job growth, reflecting the substantial effect that connections to high-entrepreneurship cities exhibit in the difference-in-differences design.

This gives us the pieces we need to show that city-specific estimates of Networking Productivity can be combined with estimates of Spillover Production, conditional on age and the friending rate, in order to estimate a quantity proportional to the treatment effect of interest from Expression 36:

$$\begin{aligned}\mathbb{E}\left[\tau_{i,t}|c(i,t)=c,R_{i,t}=x,t-t_{i,0}\right]&=\delta\beta_c+\delta\phi_c(t-t_{i,0})x\mathbb{E}\left[B_{i,t}|c(i,t)=c,R_{i,t}=x,t-t_{i,0}\right]\\&\propto\\NP_c(x)+\omega SP_c(x,t-t_{i,0})&=\overline{\alpha}\gamma\beta_c x\mathbb{E}_c\left[\phi_c|c\in\overline{\mathcal{C}}\right]+\omega\overline{\alpha}\gamma\phi_c x\mathbb{E}_i\left[B_{i,t}|c(i,t)=c,R_{i,t}=x,t-t_{i,0}\right]\end{aligned}$$

for $\omega=(t-t_{i,0})x\mathbb{E}_c\left[\phi_c|c\in\overline{\mathcal{C}}\right]$, which is equal to the average Katz centrality of permanent residents of $\overline{\mathcal{C}}$ at age $t-t_{i,0}$ who make friends at rate x.

C.0.1 Katz Centrality Estimator and Validation

Our estimate for Katz-Bonacic Centrality follows the individual-level estimator described in Lin et al. 2021, which we implement. The estimator operates by sampling a focal individual's extended network (that is, connections at network distance ≥ 2), calculating centrality using the series representation for the sampled extended network, and then adjusting the calculated centrality to account for the rate at which the network was sampled.

The estimator requires three parameters:

1. Depth d, equal to the maximum distance from the focal individual at which the estimator will sample.
2. The sampling rate r, which is the rate at which individuals at distance ≥ 2 from the focal individual are sampled.
3. The attenuation factor γ.

In the analysis sample, we set $d=2$, $r=0.001$, and γ is estimated as described in Section IV.C.2. For the validation exercise depicted in Figure E23, we use a publicly available dataset of Facebook connections which contains 4031 nodes (Jure Leskovec n.d.). We set $d=2$ and $r=0.001$ as in the analysis sample.

For γ, we use the fact that the attenuation factor is bounded above by $\frac{1}{||G||}$, where $||G||$ is equal to the largest eigenvalue of the adjacency matrix. In very large networks calculating the largest eigenvalue is difficult. However, in a wide class of random graphs with a fat-tailed power law degree distribution, the largest eigenvalue converges to the square root of the maximum expected degree as the number of nodes grows (Chung et al. 2003). In our analysis sample, the maximum degree is 3000, so the maximum value of γ in the analysis sample is bounded above by roughly $\gamma\leq\frac{1}{\sqrt{3000}}\approx 0.018$. Our estimate of γ, described in Section IV.C.2, is equal to 0.012. In the validation exercise, which has fewer nodes and a higher upper bound on the attenuation factor, we set the attenuation factor to be equal to $\frac{0.012}{0.018}\times\frac{1}{||G||}$ for $||G||$ the true maximum eigenvalue in the validation sample. This ensures that the ratio of the attenuation factor to its upper bound is the same in the validation sample and the analysis sample.

C.0.2 Estimating the Human Capital Scaling Factor

Notice that in the derivation of our estimator, a correction for levels of human capital in the mover's destination is required to arrive at Expression 38. A correction is required in this model because the effect of the hypothetical treatment D is independent of an individual's human capital, a condition which follows directly from the assumptions that the treatment effect δ is constant across people and is additively separable from human capital.

Under the functional form in the model, the correction factor is the reciprocal of $a_d \mathbb{E}\left[U_j | c(i,t) = n, c(j,t) = d, R_{j,t} = x, t - t_{i,0}\right]$, where j is the mover and a_d is the CZ-level component of human capital endowments among a resident's network for residents of CZ d. Intuitively, this quantity divides out the component of the destination-specific affect that is attributed to the level of human capital among new connections that are formed in the mover's destination. We estimate each component of the expression separately.

To form an estimate of the common component a_d, first note that an individual's propensity to become an entrepreneur as a function of their network can be split by the location where each connection resides:

$$\begin{aligned} Y_{i,t} = \beta_{c(i,t)}\alpha_i + \beta_{c(i,t)} \sum_{j,n} \gamma^n [G_t^n]_{i,j} \alpha_j \times 1\{c(i,t) = c(j,t)\} \\ + \beta_{c(i,t)} \sum_{j,n} \gamma^n [G_t^n]_{i,j} \alpha_j \times 1\{c(i,t) \neq c(j,t)\} \end{aligned} \tag{43}$$

If we assume that friend formation can be split as $\frac{d}{dt}K_{i,t} = \frac{d}{dt}K_{i,t,local} + \frac{d}{dt}K_{i,t,nonlocal}$ such that

$$\frac{d}{dt} \sum_{j,n} \gamma^n [G_t^n]_{i,j} \alpha_j \times 1\{c(i,t) = c(j,t)\} = a_{c(i,t)} U_i \cdot \frac{d}{dt} K_{i,t,local}$$

and

$$\frac{d}{dt} \sum_{j,n} \gamma^n [G_t^n]_{i,j} \alpha_j \times 1\{c(i,t) = c(j,t)\} = \overline{a} U_i \cdot \frac{d}{dt} K_{i,t,nonlocal}$$

Then we can write an individual's entrepreneur status as:

$$Y_{i,t} = \beta_{c(i,t)}\alpha_i + (t - t_{i,0})\beta_{c(i,t)} a_{c(i,t)} U_i \cdot \frac{d}{dt} K_{i,t,local} + (t - t_{i,0})\beta_{c(i,t)} \overline{a} U_i \cdot \frac{d}{dt} K_{i,t,nonlocal} \tag{44}$$

Therefore, a multivariate regression of entrepreneur status on connections in the same CZ and connections in other CZs with individual fixed effects produces the coefficient $(t - t_{i,0})\beta_{c(i,t)} a_{c(i,t)} U_i$ on the same-CZ connections and $(t - t_{i,0})\beta_{c(i,t)} \overline{a} U_i$ on other connections. The ratio of these two coefficients estimates a quantity proportional to the CZ component of human capital in the network:

$$(t - t_{i,0})\beta_{c(i,t)} a_{c(i,t)} U_i / (t - t_{i,0})\beta_{c(i,t)} \overline{a} U_i = a_{c(i,t)} / \overline{a} \tag{45}$$

While estimating individual-level changes in indirect connections is computationally prohibitive at scale, we approximate these quantities by substituting the change in the number of local vs. other friends for $\frac{d}{dt}K_{i,t,local}$ and $\frac{d}{dt}K_{i,t,nonlocal}$ respectively and estimating an OLS regression of the form

$$Y_{i,t} = \nu_t + \beta_{local} F_{local} + \beta_{nonlocal} F_{nonlocal} + \epsilon_{i,t} \tag{46}$$

for $F_{local} \equiv \sum_j G_{i,j} \times 1\{c(i,t) = c(j,t)\}$ and $F_{nonlocal} \equiv \sum_j G_{i,j} \times 1\{c(i,t) \neq c(j,t)\}$, and using the ratio $\widehat{\beta}_{local} / \widehat{\beta}_{nonlocal}$ as our estimate of $\alpha_c / \overline{\alpha}$.

C.1 Targeting Subgroups within Places

Consider an intervention that directly targets individuals of a given type who live in each city. Denote the type targeted by the program as $\theta \in \Theta$. We will derive the direct effects of targeting a person of type θ and the spillover effects onto other individuals of type θ.

Let the barriers to entry vary by place and subgroup such that

$$Y_{it} = \beta_{\theta_i, c(i,t)} E_{it} + \epsilon_{it} \tag{47}$$

The effect of a one unit increase in D_i for person i of type θ on entrepreneurship for the target and connections of types $\overline{\Theta}$, which we call $\tau_{\theta,\overline{\Theta},c}$ will be:

$$\tau_{it\theta} = \delta \beta_{\theta_i, c(i,t)} + K_{it\theta} B_{it\theta} \tag{48}$$

Where $K_{it\theta} \equiv \sum_{j:\theta_j=\theta} \sum_n \gamma^n [G_t^n]_{ij}$ denotes the extended network of i who are of type θ and $B_{it\theta} \equiv \frac{1}{K_{it\theta}} \sum_{j:\theta_j=\theta} \sum_n \gamma^n [G_t^n]_{ij} \beta_{\theta_j, c(j,t)}$ is the analogous weighted average of barriers to entry. Note that $\sum_\theta K_{it\theta} = K_{it}$ such that overall centrality can be additively decomposed into type-specific centralities.

$$\frac{d}{dt}(K_{it\theta} A_{it\theta}) = a_{\theta, c(i,t)} U_{i\theta} \cdot \frac{d}{dt} K_{it\theta} \tag{49}$$

$$\frac{d}{dt} K_{it\theta} = \sum_{s \in \Theta} R_{its} \phi_{s\theta, c(i,t)} \tag{50}$$

$$R_{it\theta} = \psi_{\theta, c(i,t)} V_{i\theta} \tag{51}$$

for $A_{it\theta} \equiv \frac{1}{K_{it\theta}} \sum_{j:\theta_j=\theta} \sum_n \gamma^n [G_t^n]_{ij} \alpha_j$ the weighted average of endowments among connections of type θ, $a_{\theta,c}$ the city-specific multiplier on the human capital of type θ, $U_{i\theta}$ the preference for friends human capital by type θ, $R_{it\theta}$ the rate at which person i makes friends with people of type θ, $\phi_{s\theta,c}$ the city- and type-specific multiplier that scales the observed friending rate of type s to account for the size of the extended network of type θ, $V_i\theta$ the individual-specific preference for friending individuals of type θ, and $\psi_{\theta,c}$ the city- and type-specific multiplier the scales individual-level preferences for type θ.

We can express type specific centrality in terms of underlying parameters:

$$K_{it\theta} = (t - t_{i,0}) \sum_{s,\theta} \phi_{s\theta, c(i,t)} \psi_{s, c(i,t)} V_{is} \tag{52}$$

Now imagine implementing the Movers design, restricting to non-movers of type θ. The estimand given (n, d, θ) is equal to

$$\sum_{\eta \in \Theta} \gamma a_{\eta d} \beta_{n\theta} \sum_{s \in \Theta} \phi_{s\eta d} \psi_{sd} \mathbb{E}[U_{j\eta} V_{j\eta} | c(i,t) = n, c(j,t) = d] \tag{53}$$

Conditioning on $\vec{R}_{jt} = \vec{x}$ where elements of the vector are types s results in the expression

$$\sum_{\eta \in \Theta} \gamma a_{\eta d} \beta_{n\theta} \sum_{s \in \Theta} \phi_{s\eta d} x_s \mathbb{E}[U_{j\eta} | c(i,t) = n, c(j,t) = d, \vec{R}_{jt} = \vec{x}] \tag{54}$$

Given estimates of $a_{\eta d} \mathbb{E}[U_{j\eta} | c(i,t) = n, c(j,t) = d, \vec{R}_{jt} = \vec{x}]$ separately by type η, one could again divide out the level of human capital. In the empirical application, we will assume separability such that

$$\begin{aligned} &\sum_{\eta \in \Theta} \gamma a_{\eta d} \beta_{n\theta} \sum_{s \in \Theta} \phi_{s\eta d} x_s \mathbb{E}[U_{j\eta} | c(i,t) = n, c(j,t) = d, \vec{R}_{jt} = \vec{x}] \\ &= \left(\sum_{\eta \in \Theta} a_{\eta d} \mathbb{E}[U_{j\eta} | c(i,t) = n, c(j,t) = d, \vec{R}_{jt} = \vec{x}] \right) \left(\gamma \beta_{n\theta} \sum_{\eta, s \in \Theta} \phi_{s\eta d} x_s \right) \end{aligned} \tag{55}$$

and divide out the average level of human capital. We thus arrive at a Movers estimand equal to:

$$\overline{\alpha}\gamma\beta_{n\theta}\sum_{\eta,s\in\Theta}\phi_{s\eta d}x_s \tag{56}$$

We can then define analogs of Networking Productivity and Spillover Production:

$$SP_{\theta c}(\vec{x}) = \overline{\alpha}\gamma\sum_{\eta,s}\phi_{s\eta c}x_s\mathbb{E}[B_{i\theta}|c(i,t)=c,\vec{R}_{it\theta}=\vec{x}] \tag{57}$$

where we assume that $\mathbb{E}[B_{i\theta}|c(i,t)=c,\vec{R}_{it\theta}=\vec{x}] = \mathbb{E}[\beta_{\theta c}|c(j,t)=c,\vec{R}_{jt\theta}=\vec{x}]$. And

$$NP_{\theta c}(\vec{x}) = \overline{\alpha}\gamma\beta_{\theta c}\sum_{\eta s}x_s\mathbb{E}[\phi_{s\eta c}|s,\eta,c\in\overline{\mathcal{C}}] \tag{58}$$

We finish by relating these quantities to the type-specific treatment effect of interest:

$$\mathbb{E}[\tau_{it\theta}|\theta,c(i,t)=c,\vec{R}_{it}=\vec{x},t-t_{i,0}] = \delta\beta_{\theta c}+\delta(t-t_{i,0})\sum_{s\in\Theta}\phi_{s\theta c}x_s\mathbb{E}[B_{i\theta}|c(i,t)=c,\vec{R}_{it}=\vec{x},t-t_{i,0}] \tag{59}$$

$$\propto$$

$$NP_{\theta c}(\vec{x})+\omega_\theta SP_{\theta c}(\vec{x},t-t_{i,0}) = \overline{\alpha}\gamma\beta_{\theta c}\sum_{\eta s}x_s\mathbb{E}[\phi_{s\eta c}|s,\eta,c\in\overline{\mathcal{C}}]+\omega_\theta\overline{\alpha}\gamma\sum_{\eta,s}\phi_{s\eta c}x_s\mathbb{E}[B_{i\theta}|c(i,t)=c,\vec{R}_{it\theta}=\vec{x}]$$

for

$$\omega_\theta = (t-t_{i,0})\sum_{\eta s}x_s\mathbb{E}[\phi_{s\eta c}|s,\eta,c\in\overline{\mathcal{C}}]\cdot\frac{\sum_s\phi_{s\theta c}x_s}{\sum_{\eta,s}\phi_{s\eta c}x_s} \tag{60}$$

where $(t-t_{i,0})\sum_{\eta s}x_s\mathbb{E}[\phi_{s\eta c}|s,\eta,c\in\overline{\mathcal{C}}]$ is the overall Katz-Bonacic centrality in the destination set and the fractional term represents the ratio of type-specific centrality $K_{it\theta}$ to overall centrality.

Appendix D: Supplemental Tables

Table D1: Overview of Covariates

Covariate	**Geographic Level Data Outside Set**	**Years Covered**	**Description**
Deposits per Capita	Commuting Zone	2019	Average bank deposit amounts per capita by CZ. *Original Source:* Federal Deposit Insurance Corporation 2019
Offices per Capita	Commuting Zone	2019	Number of bank branches per capita in a given CZ. *Original Source:* Federal Deposit Insurance Corporation 2019
Loans per Capita	Commuting Zone	2019	Number of SBA 7(a) loans issued per capita by CZ. *Original Source:* U.S. Small Business Administration 2019
Job Creation Rate	Commuting Zone	2010 – 2020	Rate of job creation due to new business establishments in a CZ, calculated as the average annual number from 2010 to 2020, divided by the 2010 population. *Original Source:* U.S. Census Bureau 2010-2020
Entry Rate	Commuting Zone	2010 – 2020	The rate of new firm entries in a CZ, calculated as the average annual number from 2010 to 2020, divided by the 2010 population. *Original Source:* U.S. Census Bureau 2015b
Firms per Capita	Commuting Zone	2015	Number of business establishments per capita in a CZ, calculated as the total number of establishments divided by the 2010 total population. *Original Source:* U.S. Census Bureau 2015b
Log Population	Commuting Zone	2010	Natural logarithm of the total population. *Original Source:* U.S. Census Bureau 2010–2019
Employees per Firm	Commuting Zone	2015	Average number of employees per business establishment in a CZ. *Original Source:* U.S. Census Bureau 2015b
Payroll per Firm	Commuting Zone	2015	Average payroll per business establishment in a CZ. *Original Source:* U.S. Census Bureau 2015b

Covariate	Geographic Level Data Outside Set	Years Covered	Description
New Employer Establishments (Sector)	National	2011 – 2020	Number of new employer establishments by sector. *Original Source:* U.S. Census Bureau 2010-2020
New Employer Establishments	County	2011 – 2020	Number of new employer establishments. *Original Source:* U.S. Census Bureau 2010-2020
Mean Household Income	Commuting Zone	2000	Mean household income, obtained from the 2000 Census. *Original Source:* U.S. Census Bureau 2000[24]
College Graduate Share	Commuting Zone	2010	Share of the population aged 25+ with at least a bachelor's degree. *Original Source:* U.S. Census Bureau 2010c [25]
Median Rent	Commuting Zone	2015	Median rent for two-bedroom units in renter-occupied housing. *Original Source:* U.S. Census Bureau 2015a[26]
Wharton Residential Land Use Regulatory Index	Commuting Zone	2008	Index measuring the strictness of land use regulations in a CZ. Constructed using city-level data and crosswalked to CZs. *Original Source:* Gyourko et al. 2008[27]
White Population Share	Commuting Zone	2010	Proportion of the population identifying as White (non-Hispanic). *Original Source:* U.S. Census Bureau 2010b [28]
Mail Return Rate	Commuting Zone	2010	Share of expected Census mail returns that were completed. *Original Source:* U.S. Census Bureau 2010a [29]
Firm Owner Demographics	National	2017	Firm owner demographics by age group and sex. Includes the total number of firm owners in each age category and the percentage breakdown for male, female, and all owners. *Original Source:* U.S. Census Bureau 2010a
New Firm Registrations	National	2011 – 2016	Proportion of new firm registrations. *Original Source:* R. Andrews et al. 2022

Covariate	Geographic Level Data Outside Set	Years Covered	Description
New Firm Registrations	County	2016	Proportion of new firm registrations by county in 2016 across all U.S. states and the District of Columbia. *Original Source:* R. Andrews et al. 2022
Volunteering Rate	Commuting Zone	2022	The percentage of Facebook users who are members of a group which is predicted to be about 'volunteering' or 'activism' based on group title and other group characteristics. *Original Source:* Raj Chetty et al. 2022a
Civic Orgs per Capita	Commuting Zone	2022	The number of Facebook Pages predicted to be "Public Good" pages based on page title, category, and other page characteristics, per 1,000 users in the county. *Original Source:* Raj Chetty et al. 2022a

Notes: Indicated variables–i.e., Mean Household Income, College Graduate Share, Median Rent, Wharton Residential Land Use Regulatory Indiex, White Population, and Mail Return Rate–are based on U.S. Census data, originally processed and cleaned by Chetty, Friedman, et al. 2019.

Table D2: Average Number of Owner Friends by Own gender, friend gender, and decile of owner friends

		Owner Friends Decile										
Own Gender	Friend Gender	1	2	3	4	5	6	7	8	9	10	Obs per Decile
Pooled	Pooled	0.00	0.00	0.00	0.60	1.00	1.92	2.88	4.47	7.15	16.14	11,387,646
Pooled	Male	0.00	0.00	0.00	0.00	0.29	1.00	1.35	2.26	3.74	9.11	11,387,646
Pooled	Female	0.00	0.00	0.00	0.00	0.14	1.00	1.22	2.15	3.53	8.37	11,387,646
Male	Pooled	0.00	0.00	0.00	0.27	1.00	1.56	2.50	3.96	6.48	15.73	5,242,460
Male	Male	0.00	0.00	0.00	0.00	0.34	1.00	1.48	2.43	4.07	10.38	5,242,460
Male	Female	0.00	0.00	0.00	0.00	0.00	0.39	1.00	1.47	2.52	6.42	5,242,460
Female	Pooled	0.00	0.00	0.00	0.88	1.23	2.12	3.30	4.87	7.57	16.44	6,145,186
Female	Male	0.00	0.00	0.00	0.00	0.24	1.00	1.23	2.12	3.46	8.04	6,145,186
Female	Female	0.00	0.00	0.00	0.00	0.77	1.00	1.85	2.68	4.32	9.69	6,145,186

Notes: This table presents the average number of owner friends by an individual's own gender, their friends' gender, and decile of owner friends. The table is structured to show differences in owner friend counts across gender pairings and ownership deciles, where Decile 1 represents individuals with the fewest owner friends and Decile 10 represents those with the most. The final column provides the number of observations per decile.

Table D3: Average Fraction of Friends who are Owners and Average Fraction of Users who are Owners for the 30 Most Populous Commuting Zones

Commuting Zone	Avg. Fraction of Friends who are Owners	Avg. Fraction of Users who are Owners
Atlanta, GA	0.00494	0.00534
Baltimore, MD	0.00484	0.00455
Boston, MA	0.00599	0.00518
Bridgeport, CT	0.00629	0.00602
Chicago, IL	0.00450	0.00390
Cleveland, OH	0.00545	0.00495
Dallas, TX	0.00341	0.00275
Denver, CO	0.00667	0.00697
Detroit, MI	0.00521	0.00481
Fort Worth, TX	0.00367	0.00315
Houston, TX	0.00294	0.00243
Las Vegas, NV	0.00404	0.00399
Los Angeles, CA	0.00288	0.00219
Miami, FL	0.00324	0.00304
Minneapolis, MN	0.00662	0.00548
New York, NY	0.00338	0.00265
Newark, NJ	0.00438	0.00380
Orlando, FL	0.00409	0.00376
Philadelphia, PA	0.00514	0.00472
Phoenix, AZ	0.00551	0.00570
Pittsburgh, PA	0.00582	0.00550
Sacramento, CA	0.00367	0.00298
San Antonio, TX	0.00319	0.00264
San Diego, CA	0.00378	0.00280
San Francisco, CA	0.00336	0.00236
San Jose, CA	0.00282	0.00196
Seattle, WA	0.00617	0.00573
St. Louis, MO	0.00609	0.00560
Tampa, FL	0.00479	0.00433
Washington DC, DC	0.00445	0.00368

Table D4: Average Number of Owner Friends by SES Decile and Owner Friends Decile

	Owner Friends Decile										
SES Decile	1	2	3	4	5	6	7	8	9	10	Obs per Decile
1	0.00	0.00	0.00	0.00	0.00	0.00	0.40	1.00	1.60	4.71	917,617
2	0.00	0.00	0.00	0.00	0.00	0.60	1.00	1.60	2.71	7.26	976,405
3	0.00	0.00	0.00	0.00	0.43	1.00	1.36	2.25	3.72	9.25	995,069
4	0.00	0.00	0.00	0.16	1.00	1.12	2.00	3.03	4.82	11.26	1,028,660
5	0.00	0.00	0.00	0.82	1.00	1.87	2.60	3.83	6.00	13.27	1,064,406
6	0.00	0.00	0.39	1.00	1.57	2.29	3.35	4.76	7.22	15.27	1,097,200
7	0.00	0.00	0.86	1.20	2.00	2.94	4.16	5.73	8.42	17.28	1,126,067
8	0.00	0.25	1.00	1.76	2.54	3.55	4.84	6.71	9.65	19.33	1,150,997
9	0.00	0.58	1.26	2.10	3.12	4.30	5.63	7.71	10.88	21.63	1,172,037
10	0.00	0.89	1.78	2.73	3.79	5.20	6.73	8.94	12.55	25.01	1,190,925

Notes: This table reports the average number of owner friends by an individual's SES decile and owner friends decile. SES deciles rank individuals from lowest (1) to highest (10) socioeconomic status, while owner friends deciles classify individuals by the number of owner friends.

Table D5: Appendix Table: College Design Robustness

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner	(4) Is Founder	(5) Is Founder	(6) Is Owner Rev > $1M	(7) Is Owner Rev > $1M	(8) Is Owner Rev > $1M	(9) Is Owner
Owner Friends of Friends	6.75×10^{-5}	7.81×10^{-5}	1.13×10^{-4}	3.40×10^{-5}	2.63×10^{-5}	1.95×10^{-5}	2.12×10^{-5}		
	(2.53×10^{-5})	(2.53×10^{-5})	(5.23×10^{-5})	(1.52×10^{-5})	(1.69×10^{-5})	(9.97×10^{-6})	(2.11×10^{-5})		
Is Owner								0.25	
								(0.12)	
Owner Friends of Friends Among Friends who did Not Start a Firm									1.10×10^{-4}
									(3.24×10^{-5})
Number of Cells	598,892	598,892	598,892	598,892	598,892	598,892	598,892	598,892	551,927
Mean of Dep. Variable	0.0027	0.0027	0.0027	0.0010	.0010	.0004	.0004		0.0028
SD of Dep. Variable	0.0519	0.0519	0.0519	0.0301	0.0301	0.0195	0.0195		0.053
Mean of Ind. Variable	3.30	3.30	1.12	3.30	3.30	3.30	1.11		2.54
SD of Ind. Variable	3.00	3.00	1.34	3.00	3.00	3.00	1.34		2.55

Table D6: College Design: SES-Specific Exposure Effects

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner
Owner Friends of Friends Above 90th Percentile Income	1.85×10^{-4} (7.40×10^{-5})		1.31×10^{-4} (7.55×10^{-5})
Owner Friends of Friends Below 90th Percentile Income		1.50×10^{-4} (4.96×10^{-6})	1.29×10^{-4} (5.07×10^{-5})
Observations	767,329	767,329	767,329
Mean of Dependent Variable	0.0028	0.0028	0.0028
SD of Dependent Variable	0.053	0.053	0.053
Mean of Independent Variable	0.49	1.07	
SD of Independent Variable	0.86	1.38	

Notes: This table estimates income-specific exposure effects in the college design. Column (1) estimates the coefficient from a univariate regression of an indicator for founding a firm after college on exposure to high-income owners. Column (2) repeats Column (1) using exposure to low-income owners as the independent variable. Column (3) estimates a multivariate regression with both high- and low-income owners.

Table D7: Distribution of Katz-Bonacic Centrality Estimates
Means by Quartile

Attenuation Factor	Q1	Q2	Q3	Q4
0.001	0.0812	0.1996	0.3576	0.8378
	(0.0021)	(0.0022)	(0.0039)	(0.0248)
0.005	0.7554	2.0411	3.6858	10.3771
	(0.0213)	(0.0264)	(0.0412)	(0.4515)
0.010	2.4905	6.6566	12.2260	33.9827
	(0.0732)	(0.0829)	(0.1325)	(1.5569)
0.015	5.1067	13.6762	25.4764	72.5456
	(0.1501)	(0.1742)	(0.3058)	(3.1150)
0.012269	3.3428	9.2454	17.7998	50.2177
	(0.0951)	(0.1289)	(0.2156)	(2.3048)

Notes: This table shows the distribution of Katz centrality estimates across quartiles for different attenuation factors, with standard errors in parentheses. Estimates were obtained using a sampling-based SAKE method with inverse sampling probability correction and averaged over multiple iterations. The attenuation factor scales centrality values, with higher factors leading to larger estimates in higher quartiles.

Table D8: Top 10 CZs with the Highest Shrunk Total Effect Among the 50 Most Populated CZs

CZ Name	Total Effect	Direct Effect
Charlotte	1.40×10^{-3}	2.7×10^{-5}
Los Angeles	8.78×10^{-4}	8.0×10^{-6}
Detroit	8.36×10^{-4}	1.1×10^{-5}
New York	6.79×10^{-4}	-2.0×10^{-6}
Baltimore	6.48×10^{-4}	8.0×10^{-6}
Atlanta	6.36×10^{-4}	1.6×10^{-5}
Dallas	6.18×10^{-4}	-3.0×10^{-6}
Denver	6.04×10^{-4}	-1.6×10^{-5}
Las Vegas	5.74×10^{-4}	2.6×10^{-5}
Manchester	5.54×10^{-4}	3.6×10^{-5}

Notes: This table lists the top 10 commuting zones (CZs) with the highest shrunk total effect among the 50 most populated CZs. The total effect measures the overall impact of networking on entrepreneurship, while the direct effect isolates the portion not mediated by network exposure. Shrinkage was applied to reduce estimation noise: the direct and total effects were adjusted by shrinking towards the mean effect, where the shrinkage factor is the ratio of signal variance to total variance as in (Bell et al. 2019).

Appendix E: Supplemental Figures

Figure E1: Summary of Data Linkage Relationships

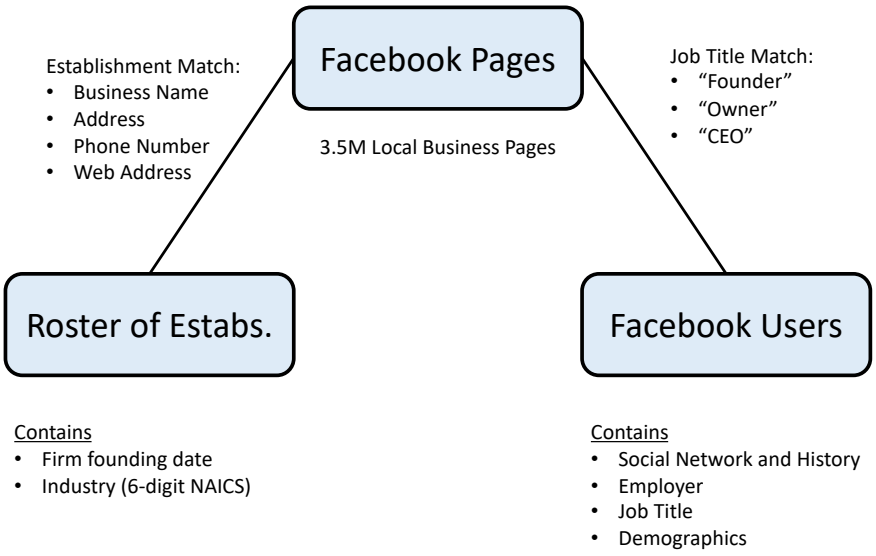
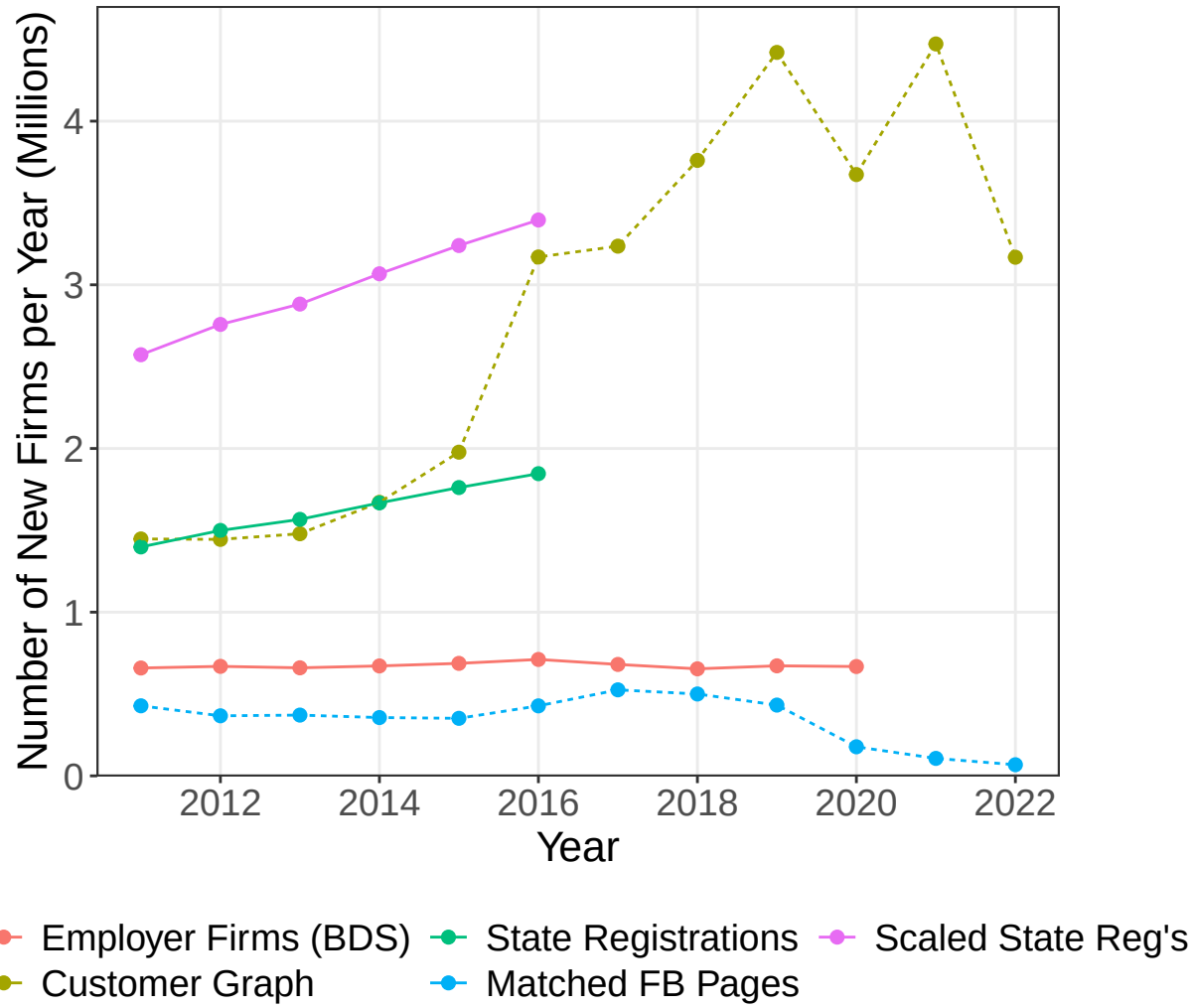
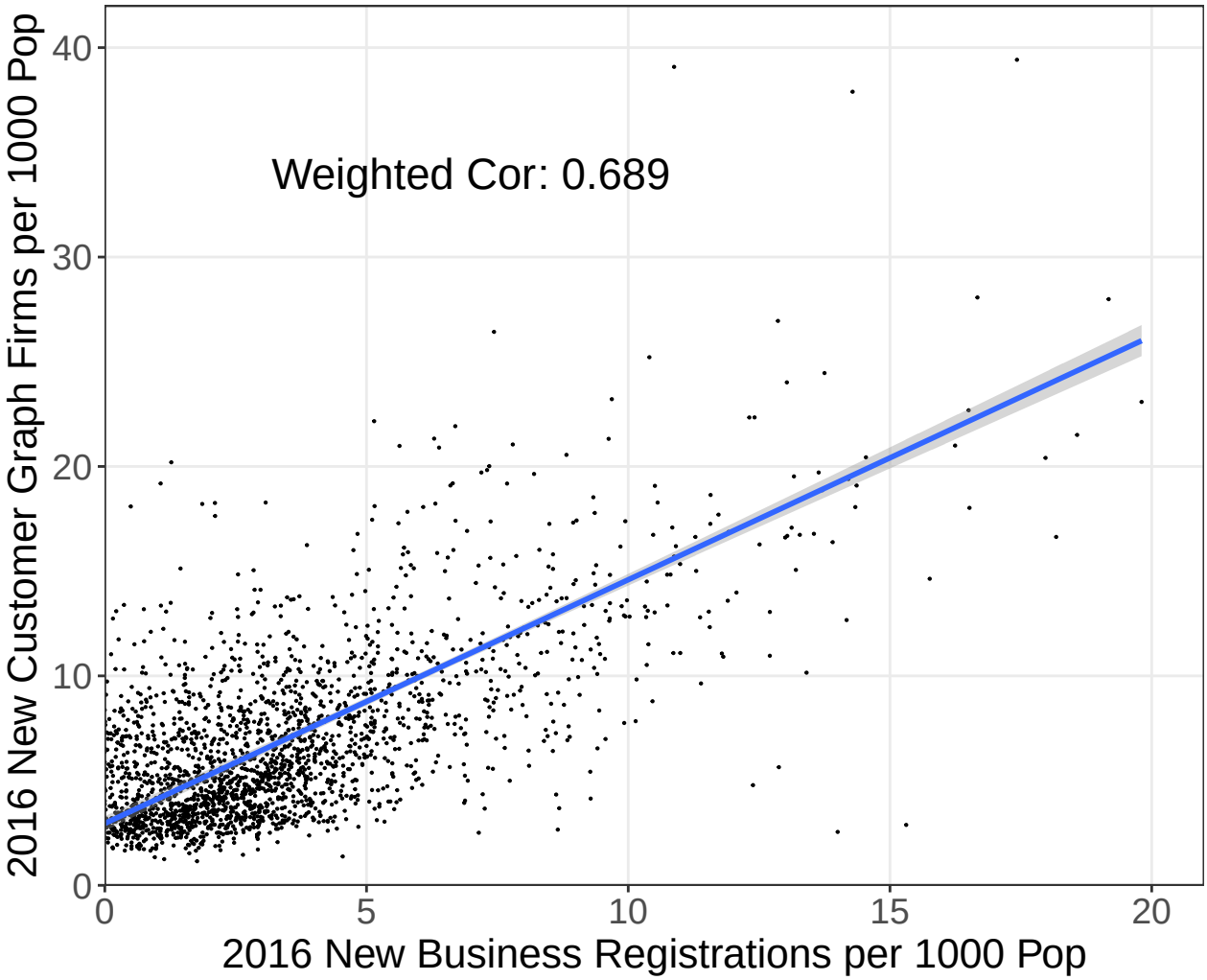


Figure E2: Benchmarking National Firm Counts



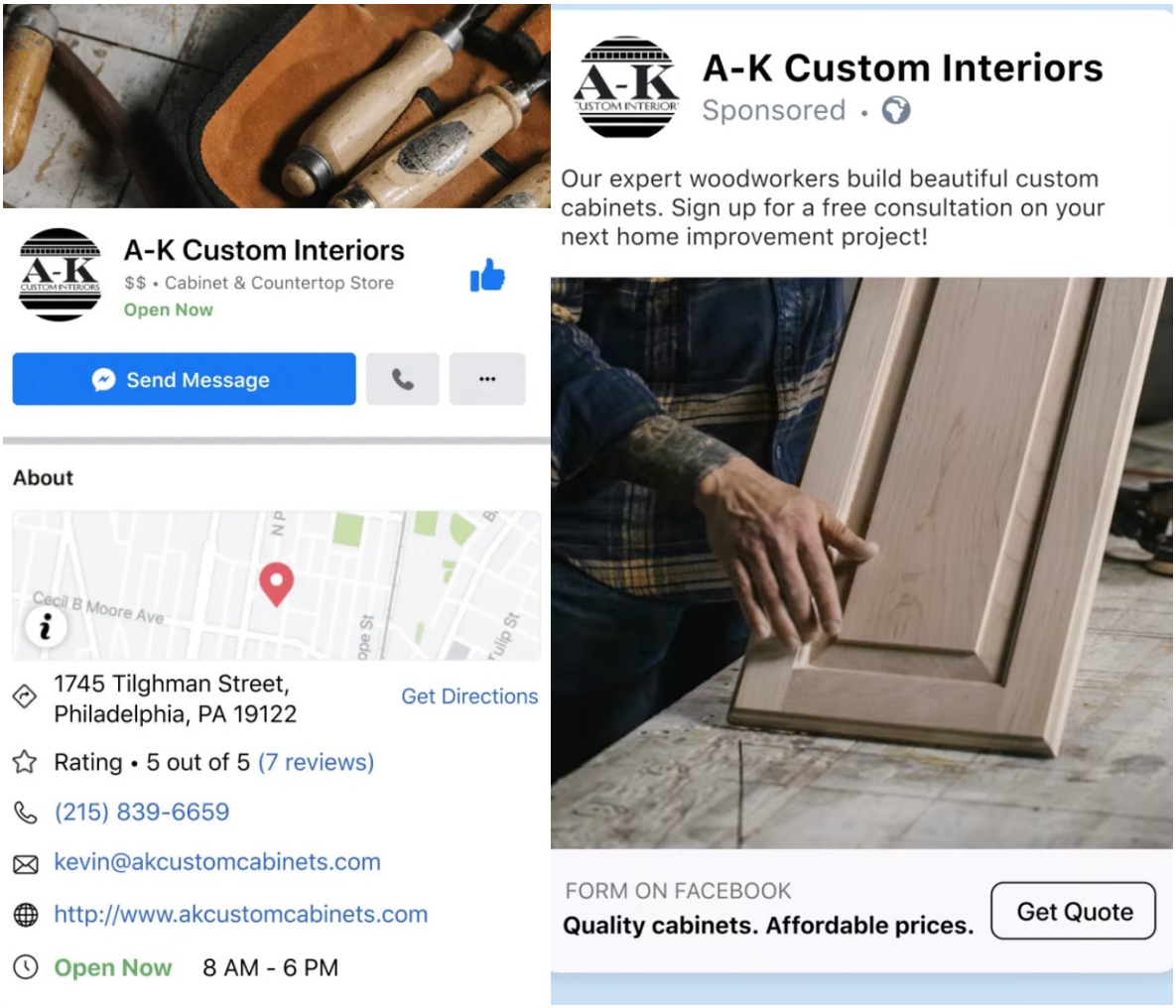
Notes: This figure plots the number of new firms per year from 2011 to 2022. It compares the number of new firms in the Customer Graph with a variety of publicly available datasets, including the Business Dynamics Statistics from the U.S. Census Bureau and state registration data. Further details and analysis are provided in Section II.

Figure E3: Benchmarking All New Firm Registrations in 2016: County Level



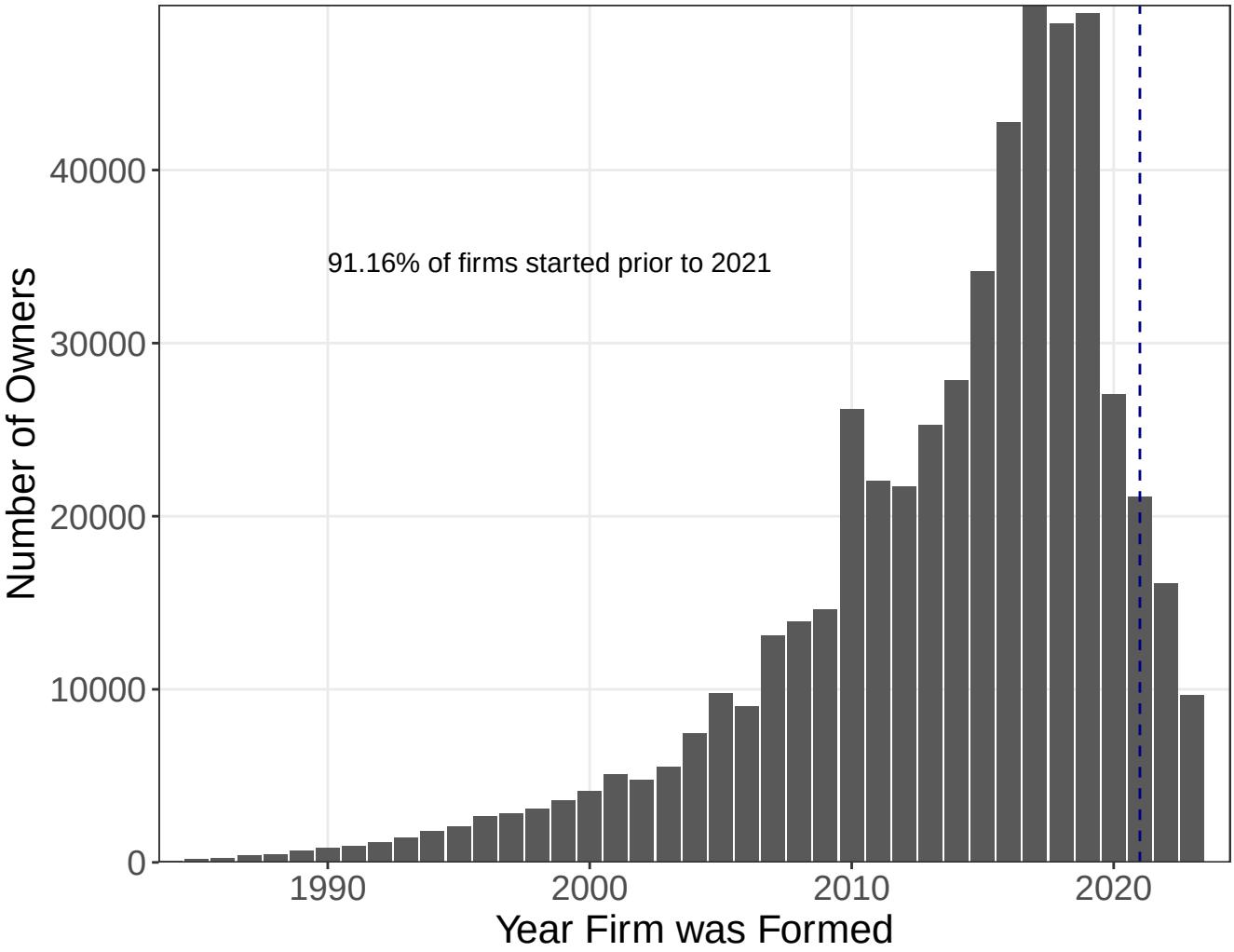
Notes: This figure repeats the analysis performed in Figure 1b, which compares the number of matched Facebook pages per capita to the number of new employer establishments per capita in 2016. However, this figure instead uses data from (R. Andrews et al. 2022) to compare the number of new Customer Graph firms per capita with the number of new business registrations per capita in 2016. All analysis is done on the county level. Details on the results are provided in Section II.

Figure E4: Sample Facebook Business Page



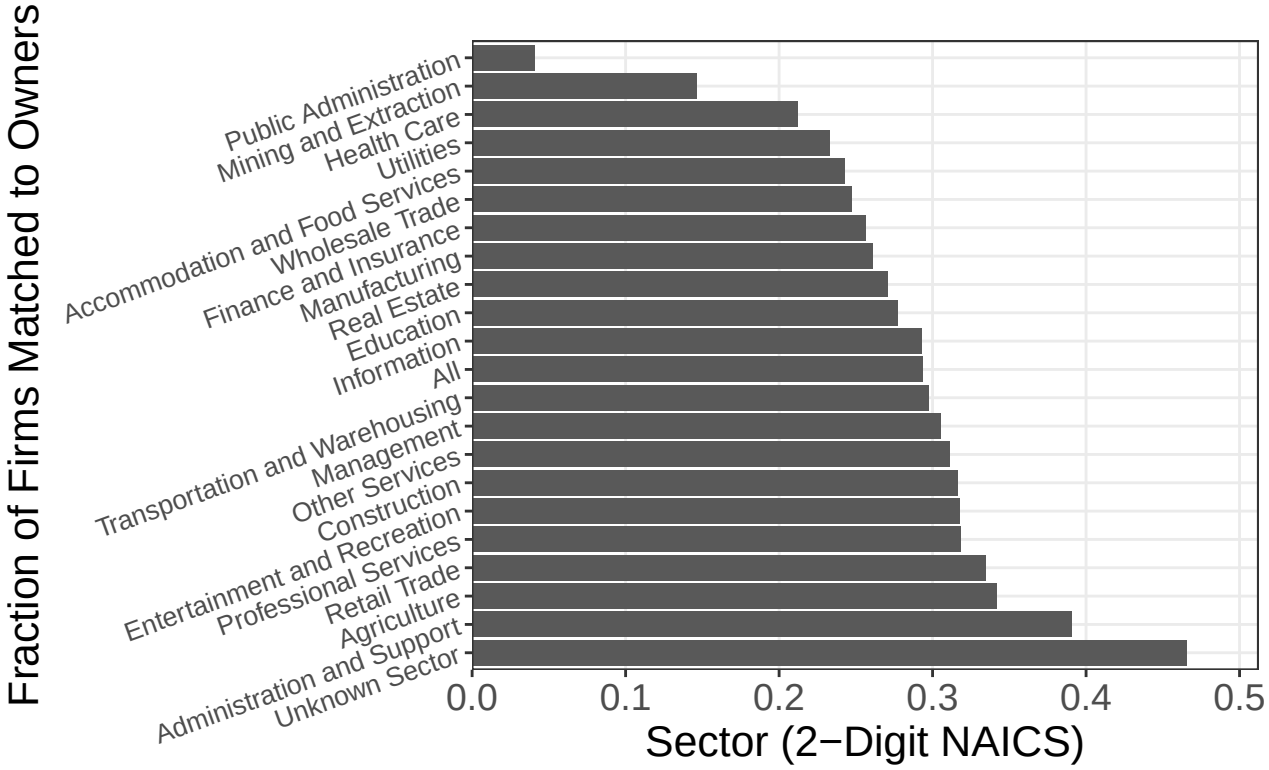
Notes: This figure presents a sample Facebook page for a local business. We use Facebook pages to connect business owners with Facebook profiles to firmographic data. Details of the linking process are presented in Section II.

Figure E5: Number of Firm Owners on Facebook by Firm Start Year



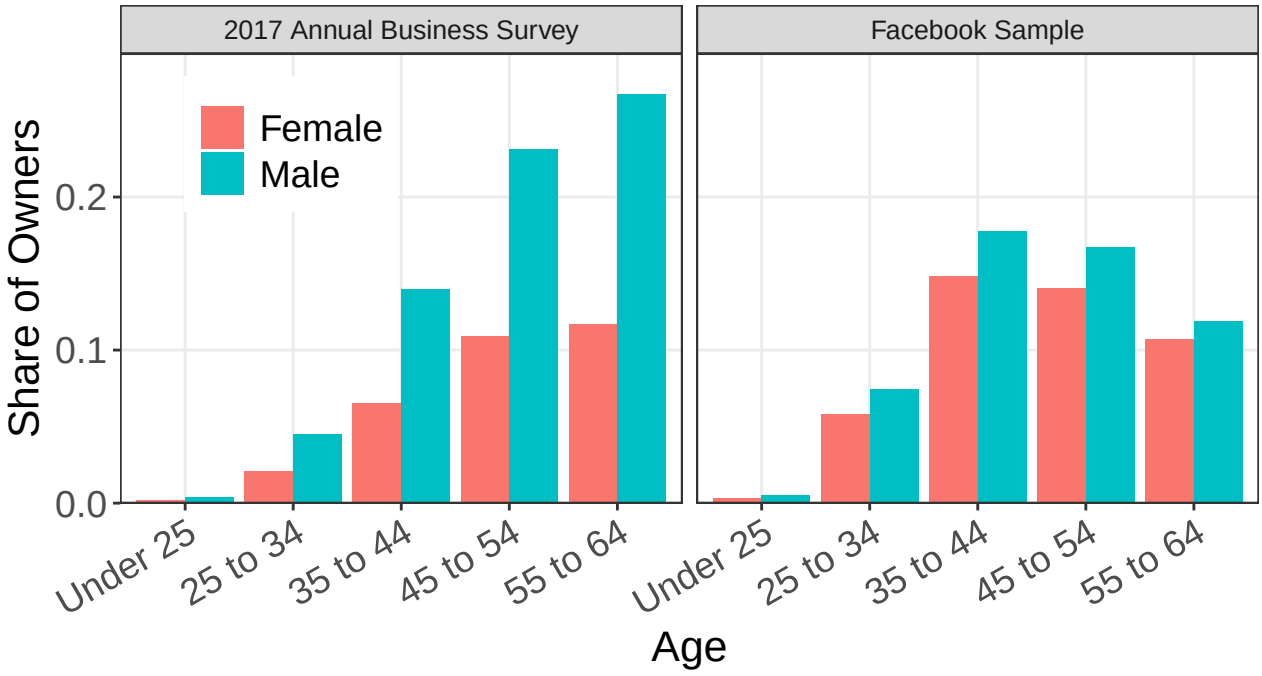
Notes: This figure displays the number of firm owners in the primary sample by the year that they started their first firm. Note that the set of firms for this figure are those that survive the matching process described in Section A.1 where data from the Customer Graph and Establishment Pages are matched. This matching procedure identifies about 695,000 firm founders, from which we create this figure. A detailed description of the dataset and matching process can be found in Section II.

Figure E6: Verified Pages Matched to Owners by Industry



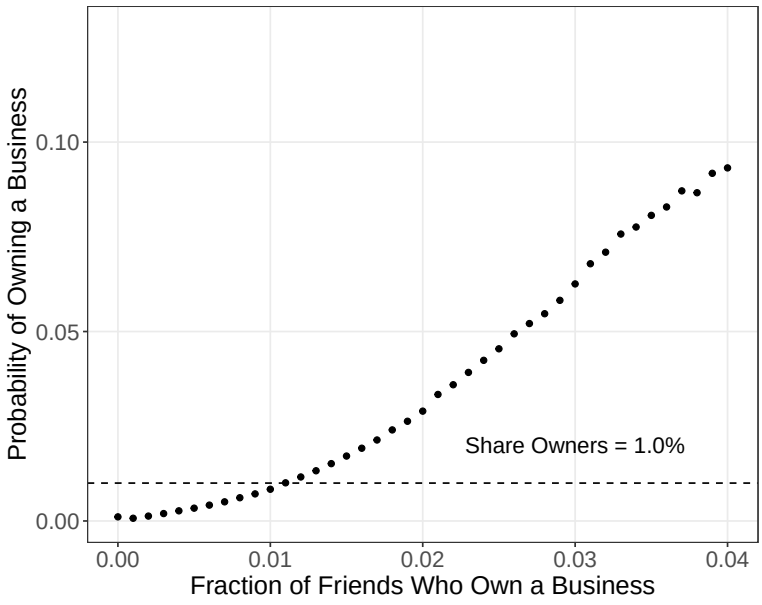
Notes: This figure displays the match rate by industry between users and verified matched firms at the 2-digit NAICS level. Note that the set of firms for this figure are those that survive the matching process described in Section A.1 where data from the Customer Graph and Establishment Pages are matched. This matching procedure identifies about 695,000 firm founders, from which we create this figure. A detailed description of the dataset and matching process can be found in Section II.

Figure E7: Owner Demographics: Annual Business Survey vs Facebook Sample



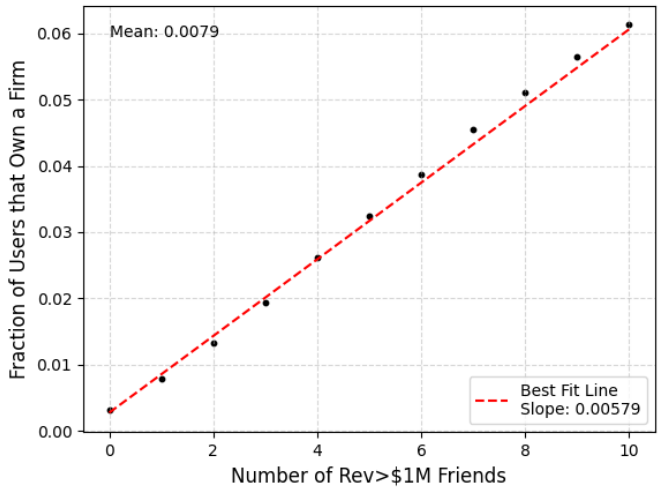
Notes: This figure compares the age and gender distributions of firm owners identified in our data and in the 2017 Annual Business Survey (ABS). Our data contain 830,000 individuals that can be matched to pages in 3rd-party data, which is approximately 21% as large as the number of owners of employer establishments estimated by the ABS. Owners identified in the Facebook data tend to be younger and more female than owners in the ABS. Owners are matched to pages using the individual's self-reported employer and job title on their public Facebook profile. Details of the matching procedure are described in Section II

Figure E8: Firm Ownership by Rates of Firm Ownership Among Friends



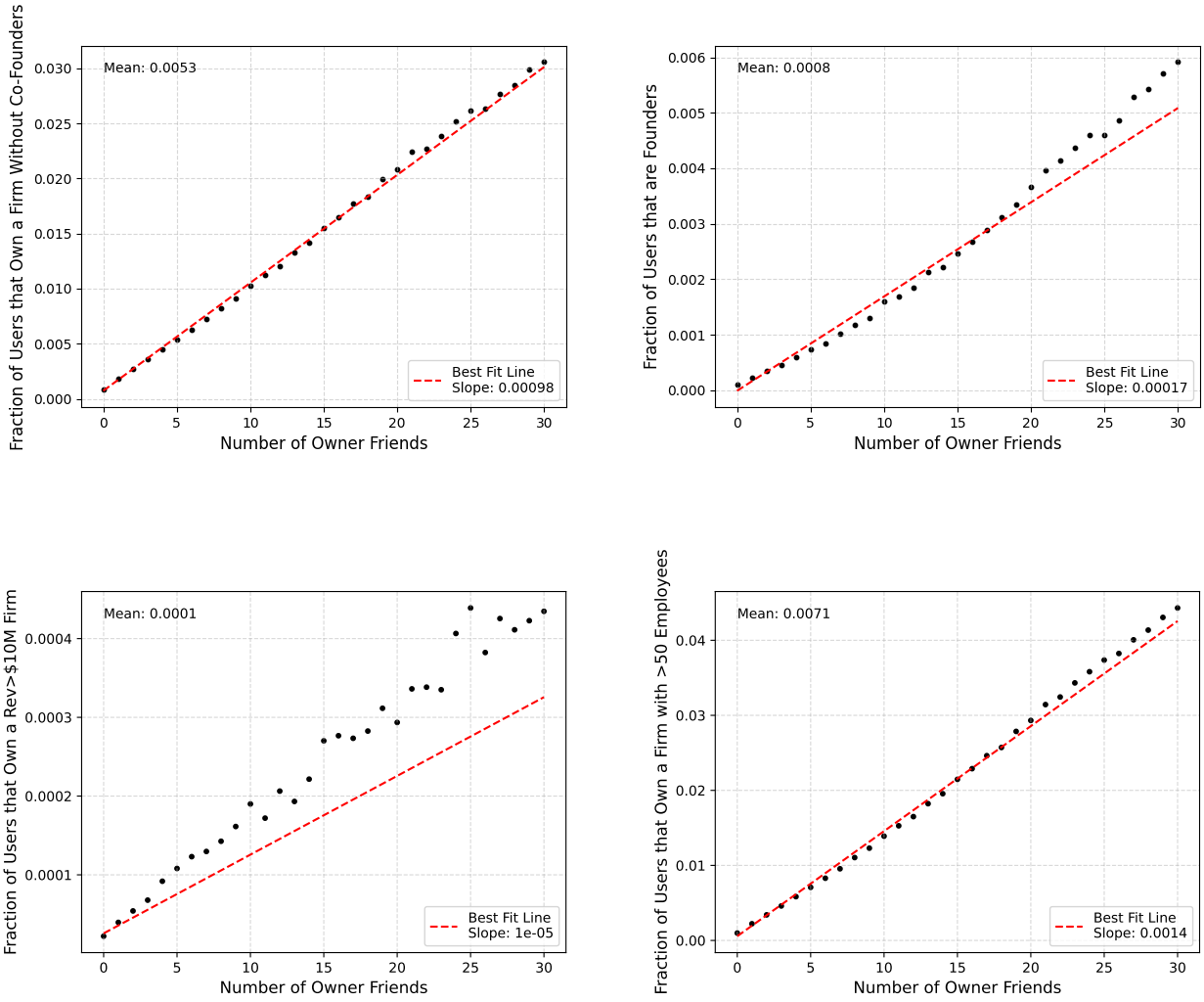
Notes: This figure shows that the fraction of an individual's friends who own a business predicts whether the individual owns a business among individuals with more than 100 friends.

Figure E9: Firm Ownership by Number of Firm Owners with $1 Million in Annual Revenue Among Friends



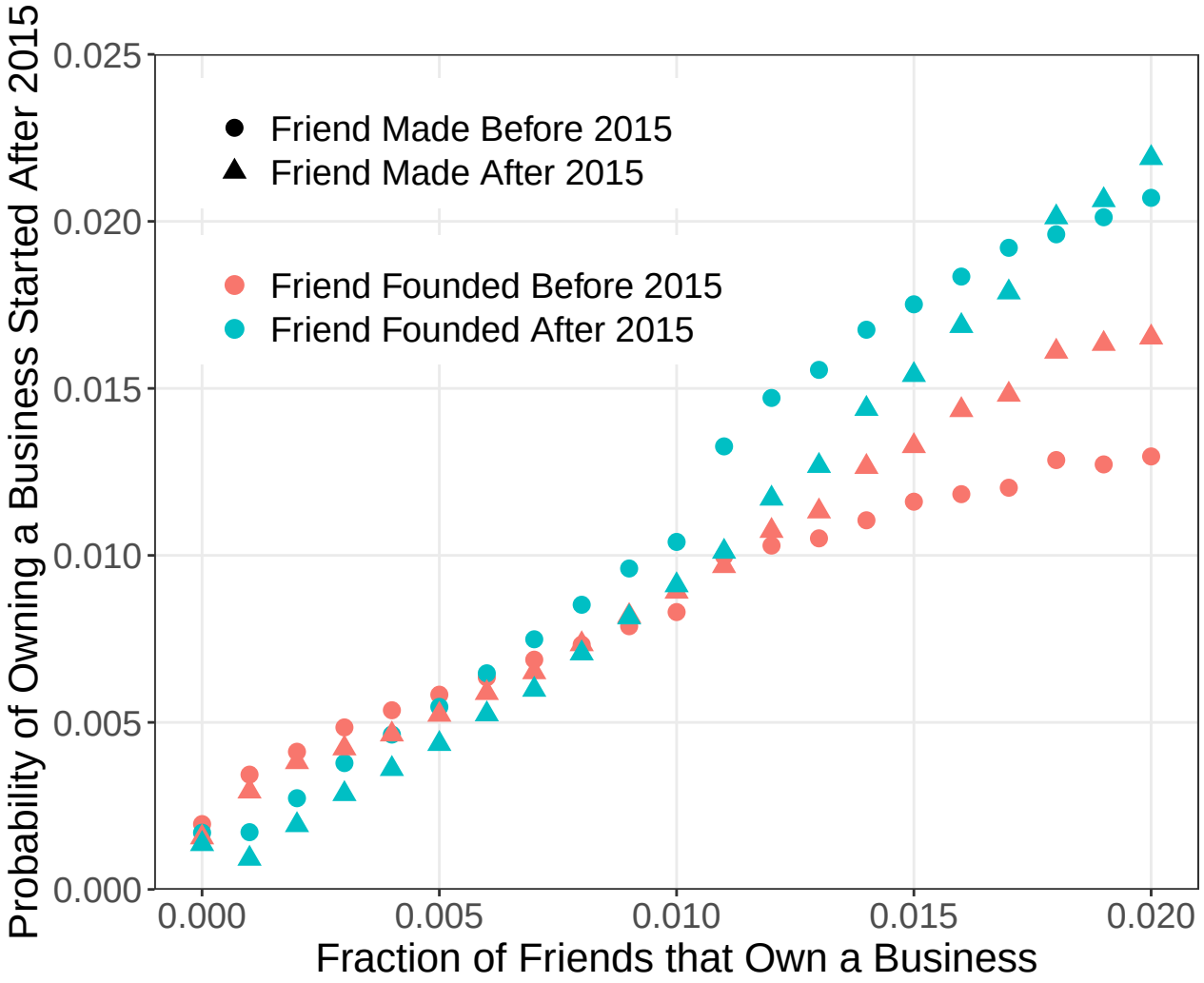
Notes: This figure shows that the number of an individual's friends who own a business with $1 Million in annual revenue predicts whether the individual owns a business.

Figure E10: Firm Ownership by Number of Firm Owners Among Friends: Alternative Owner Definitions and Outcomes



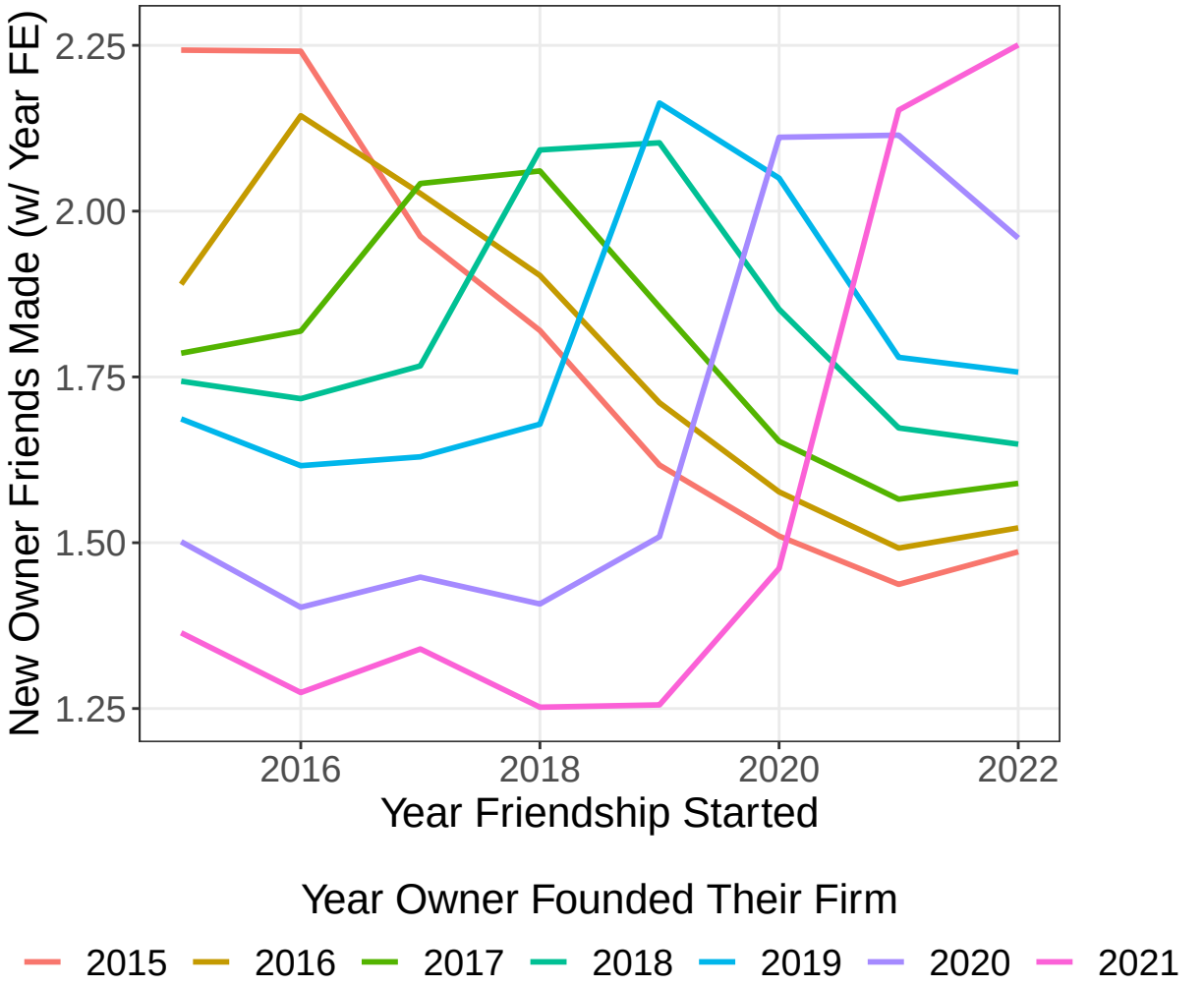
Notes: This figure presents scatterplots of the fraction of individuals who own a firm by the number of friends they have who own a firm for alternative definitions of the outcome. The upper left plot restricts to firm owners that have zero co-founders. The upper right plot restricts the definition of firm owners in the sample only match those who fuzzy string match the term "founder". The lower left considers as the outcome whether an individual starts a firm with $10 million in revenue, and the lower right considers as the outcome whether an individual starts a firm with 50 employees.

Figure E11: Founding Rates by Friend Ownership Rates and Friendship Length



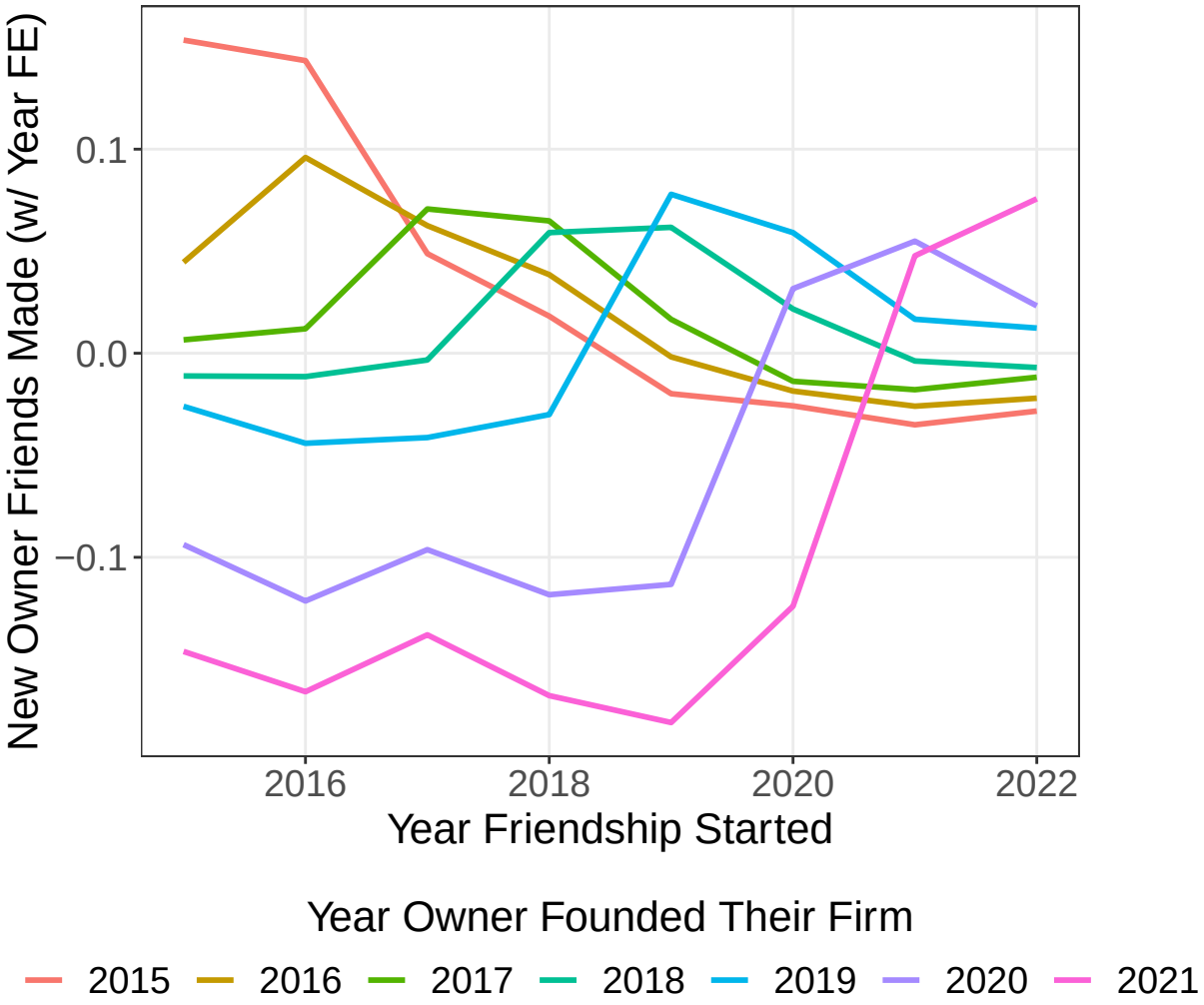
Notes: This figure repeats the analysis from Figure E9 but disaggregated by the year the friendship was formed and the year the friend founded their firm. This figure primarily demonstrates that the relationship between entrepreneurship and friendship holds across time. Further detail is provided in Section II.B.

Figure E12: Owners Meet Owners in the Year they Start a Business



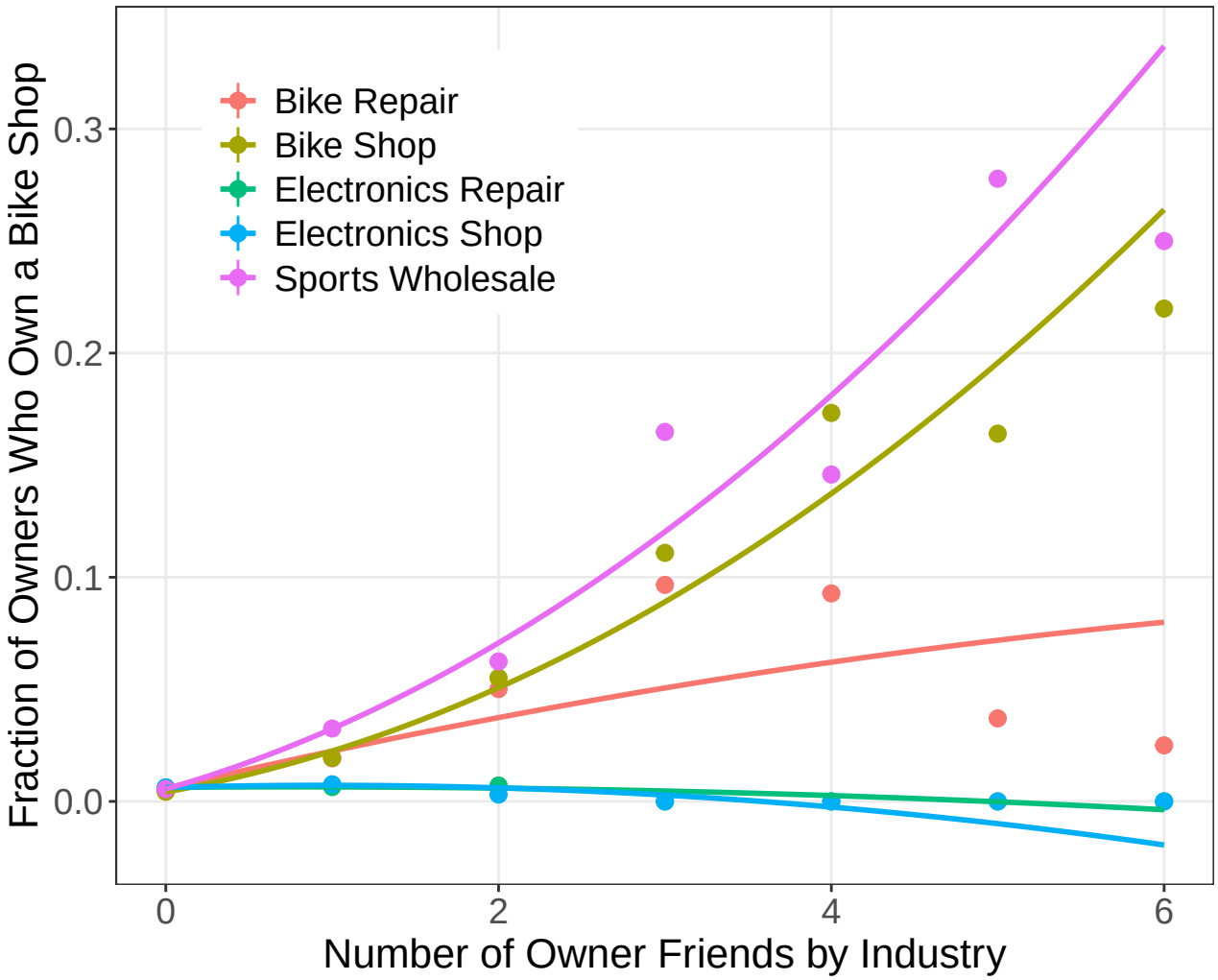
Notes: This figure displays the yearly rates at which owners befriend owners, separated by the year that the focal owner started their firm. This specification includes year fixed effects and adds back the overall mean across all periods. This figure is further analyzed in Section II.B.

Figure E13: Owners Meet pre-2010 Owners in the Year they Start a Business



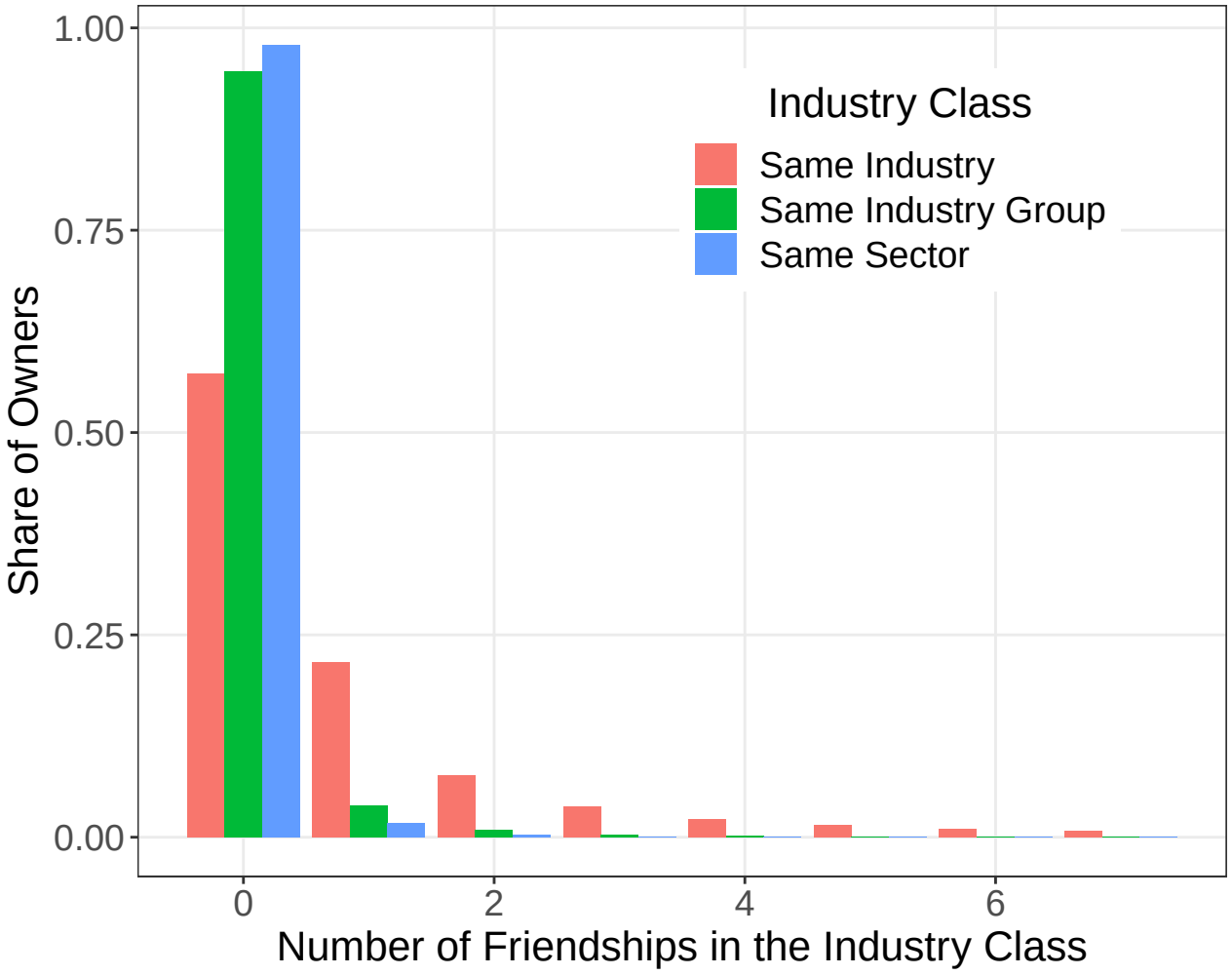
Notes: This figure performs the same analysis as Figure E12, but only measures friendships with owners who started their firm prior to 2010. This figure is described further in Section II.B.

Figure E14: Bike Shop Owner Friends



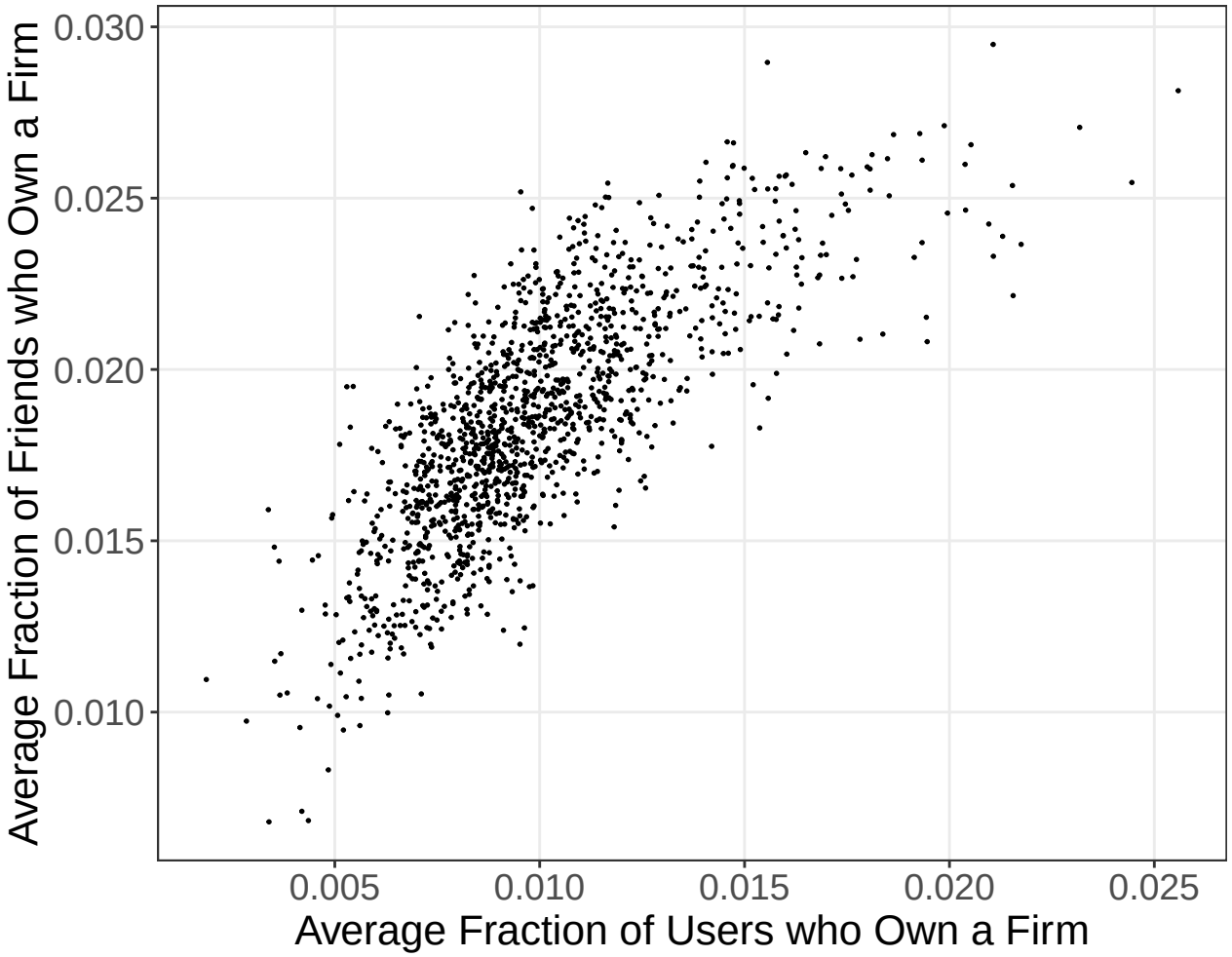
Notes: This figure provides an example of industry-specific friending. It plots the fraction of owners who own a bike shop relative to the number of friends that are within each specific industry. The 6-digit NAICS codes corresponding to each of the relevant industries is: Bike Repair - 811490; Bike Shop - 336991; Electronics Repair - 811210; Electronics Shop - 449210; Sports Wholesale - 423910. This figure is described further in Section II.B.

Figure E15: Distribution of Owner Friends by Industry Similarity



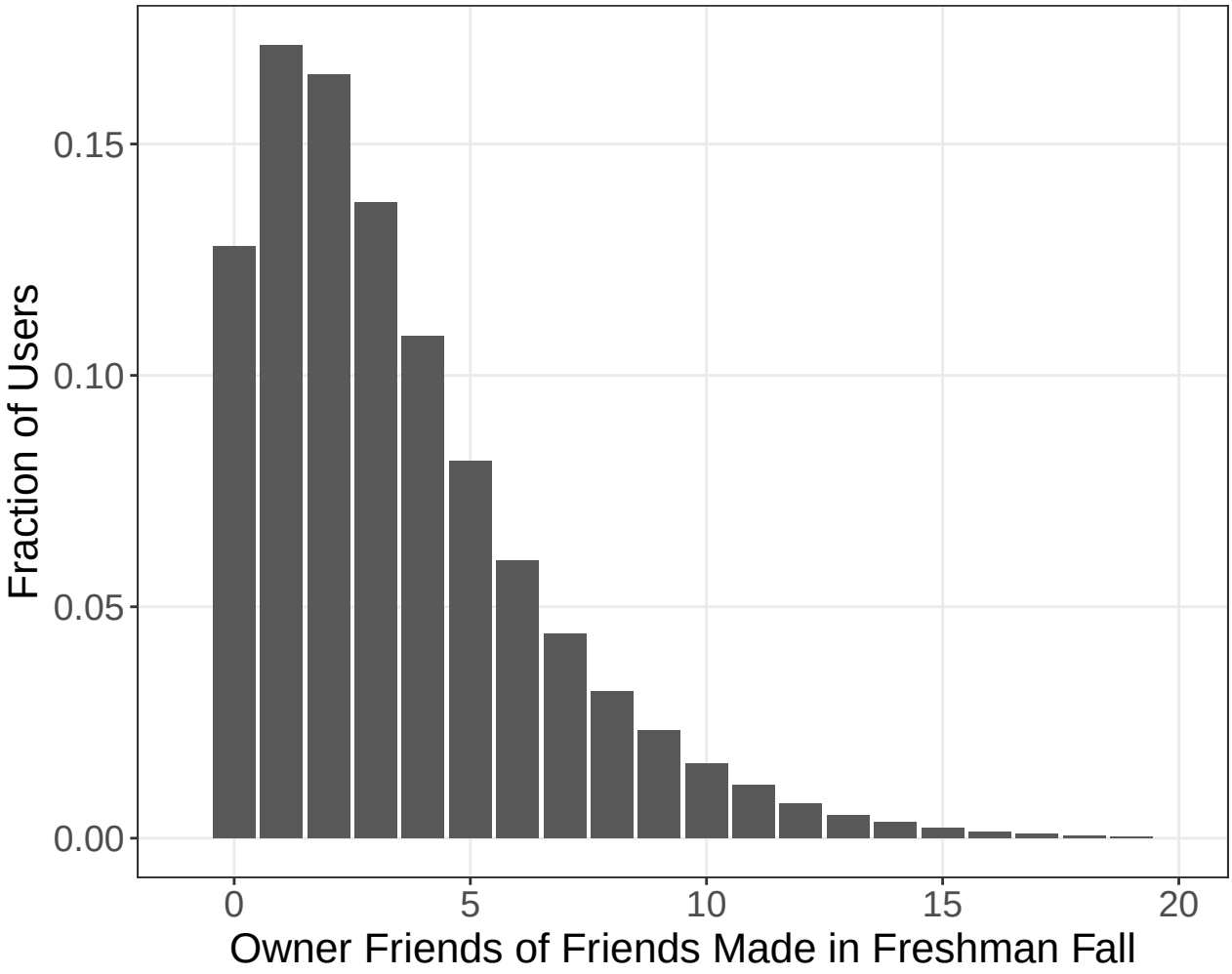
Notes: This figure shows the number of owner friends present at each industry level. Each color is a separate distribution showing the number of owner friends in each industry level separated by NAICS class. So, the red color graphs the distribution of friends in the same 6-digit NAICS industry, the green color graphs the distribution of friends in the same 4-digit NAICS industry group, and the blue color graphs the distribution of friends in the same 2-digit NAICS sector. Detailed analysis is given in Section II.B.

Figure E16: Average Fraction of Friends who Own a Firm vs Fraction of Individuals who Own a Firm by County



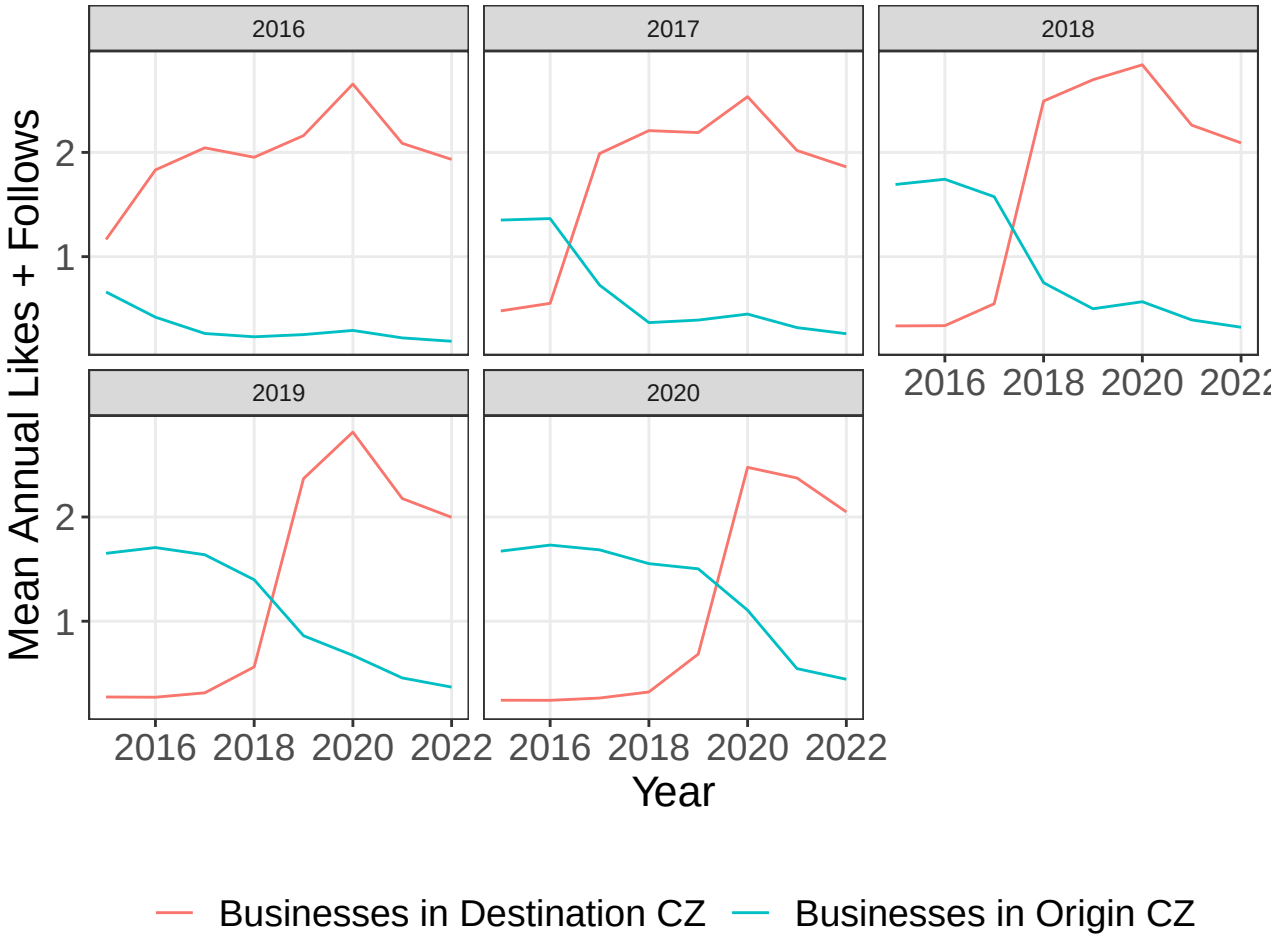
Notes: This figure is a county-level scatterplot comparing the fraction of friends who are owners with the fraction of users who own a firm. Further analysis is given in Section II.B.5.

Figure E17: Distribution of Owner Friends-of Friends in College



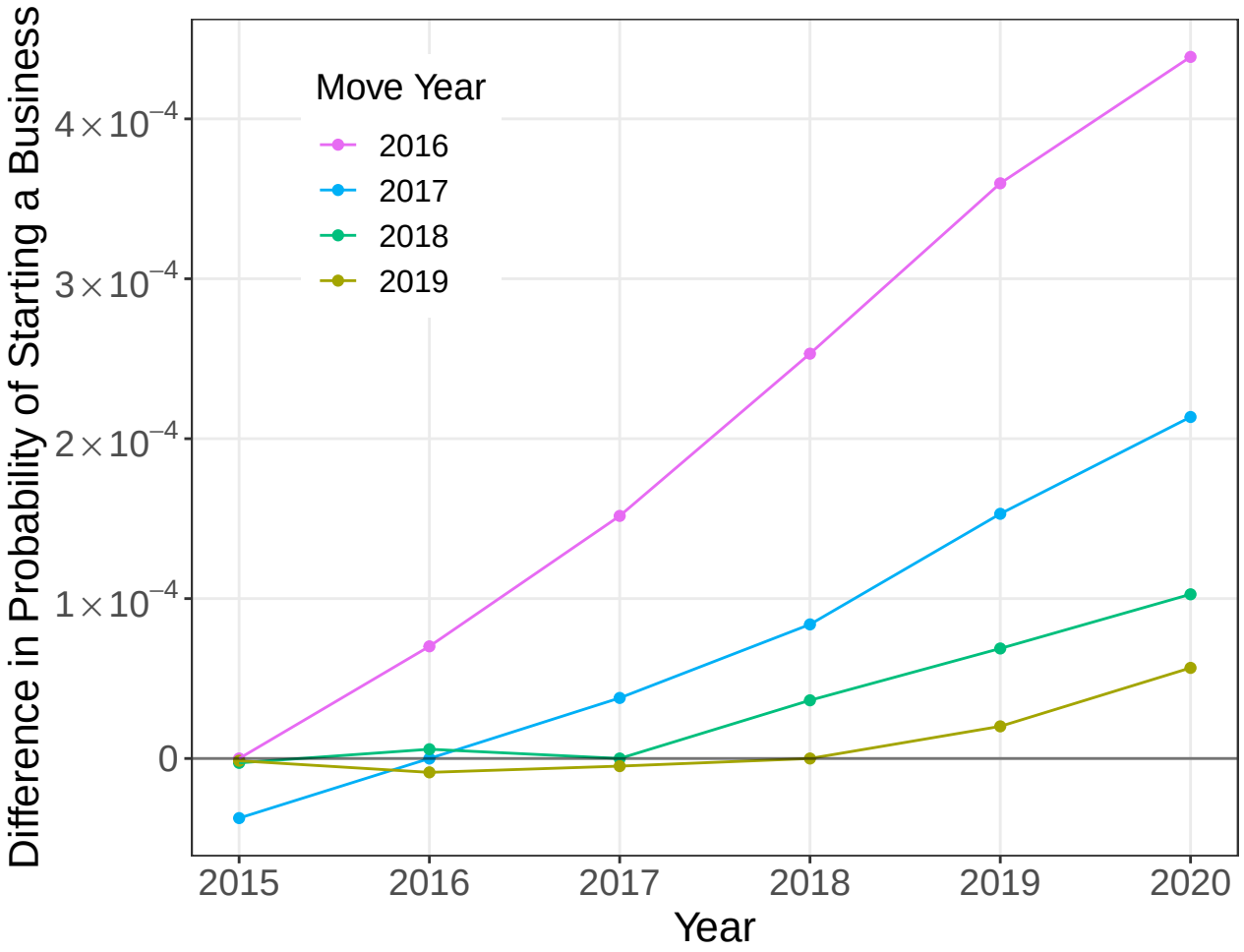
Notes: This figure plots the number of owner friends-of friends in the College Design. For this figure, in order to ensure that only an individual's friending behavior and activity are driving the results, we impose a number of specific characteristics on 2nd-degree connections formed during the first 12 weeks of the academic year. Specifically, for some person i and a new second-degree connection j, we must satisfy the following conditions: 1. i and j were not friends and had no mutual friends prior to i turning 18. 2. Between weeks 33-44 of the year i turned 18, i becomes friends with one of j's friends (whom j had known before the year i turned 18). 3. j started a business before the year i turned 18. These specifications ensure j's friending behavior and founding activity are held constant while analyzing the impact of i's friending behavior. Detail on the design of the model is given in Section III.A.

Figure E18: Interactions with Local Business Pages by Year of Move



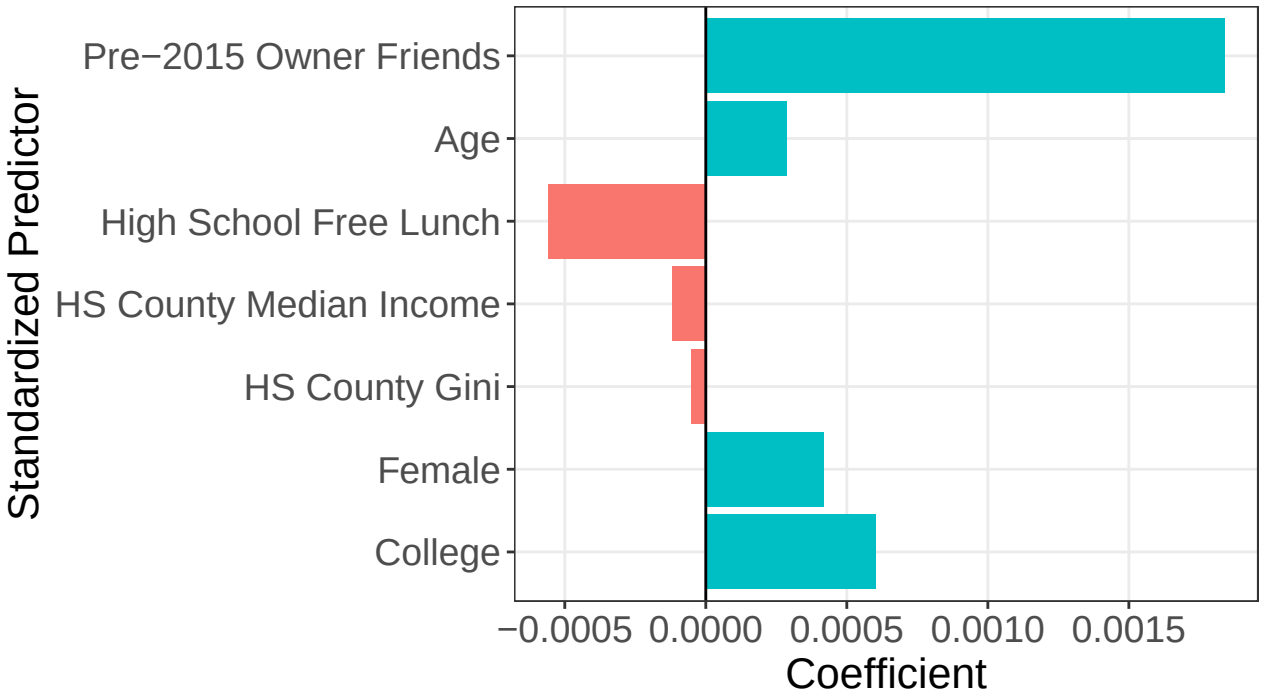
Notes: This figure plots the average mean interactions (likes plus follows) between movers and local business pages in the origin vs. the destination for movers in each year between 2016 and 2020. Further information is provided in Section III.B

Figure E19: Effect of a Friend Moving to a High-Entrepreneurship CZ by Move Year: Friends of 2020 Movers as Control



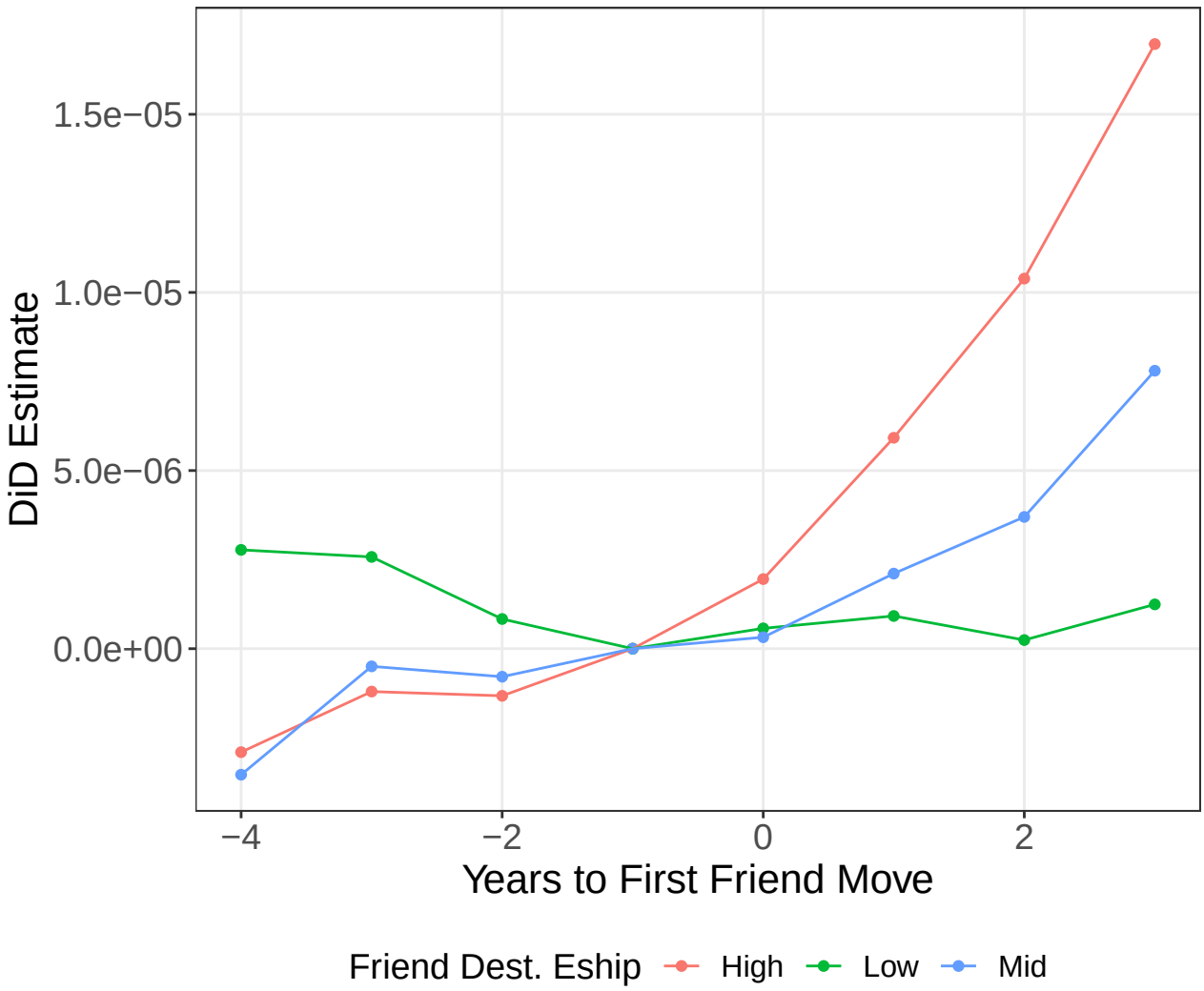
Notes: This figure plots the difference in probability of starting a business, using 2020 as the control year. A difference-in-differences design is used to analyze the impact of move year on probability. Information on the design is given in Section III.B

Figure E20: Predictors of Fraction post-2015 Friends who are Owners



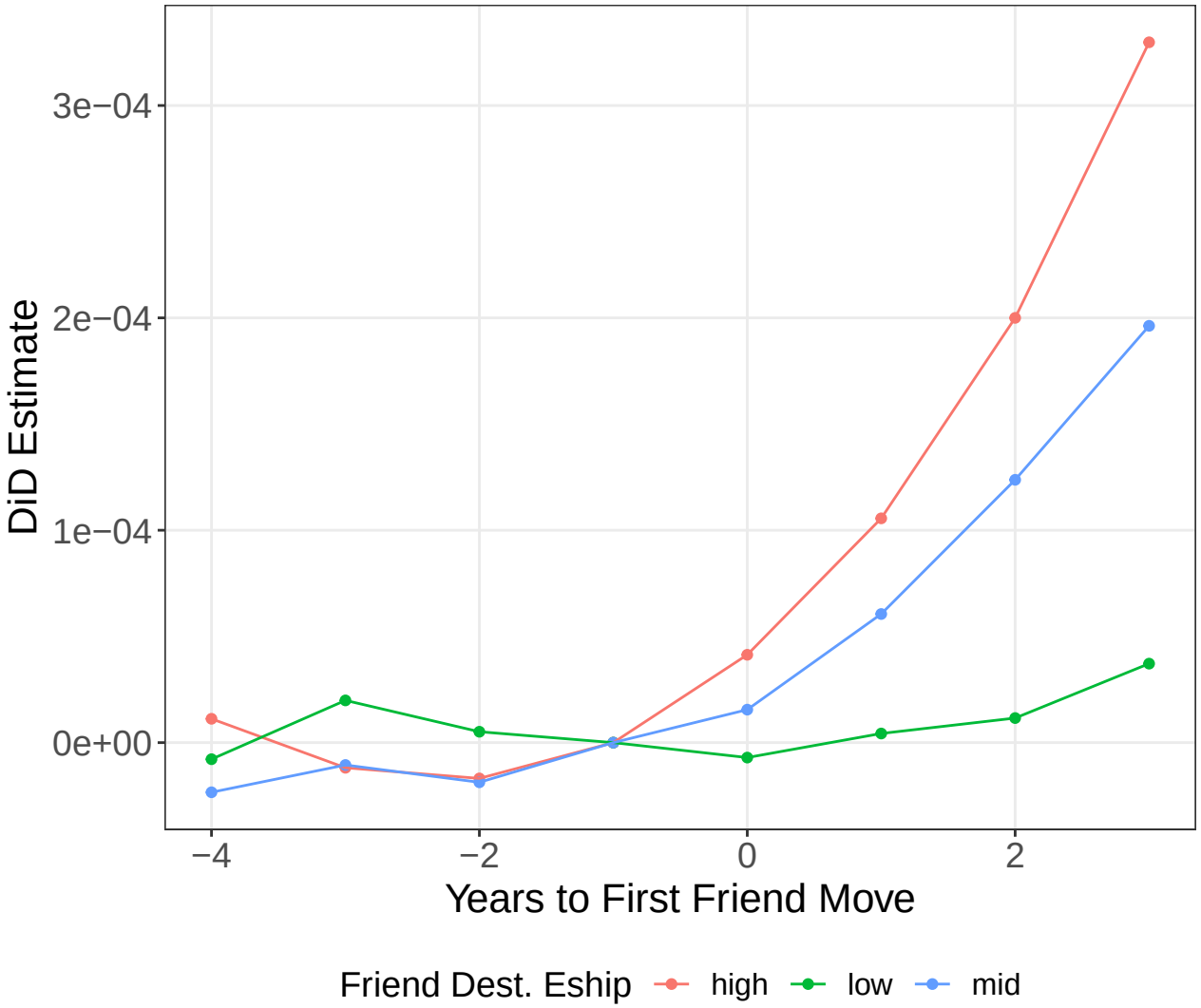
Notes: This figure plots the coefficients obtained from an individual-level regression of each of the listed predictors on the number of owner friends made after 2015. All explanatory variables other than female and college are standardized. Further analysis of the figure is given in Section III.B

Figure E21: Commuting-Zone by Sector Difference in Differences



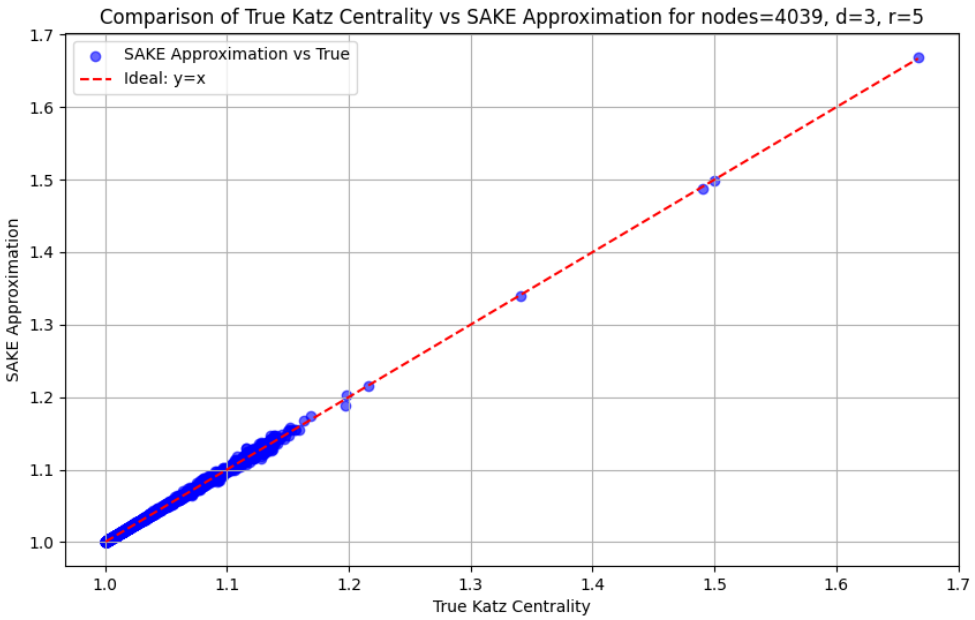
Notes: This figure repeats the analysis in Figure 10 but defines the destination market at the commuting zone-by-sector level. The same technical designs are used from Figure 10 to produce the results in this figure. Further analysis of the figure and comparisons to Figure 10 are given in Section III.B.

Figure E22: Effect of a Friend Moving to a High-Entrepreneurship CZ: Mover does not Start a Business



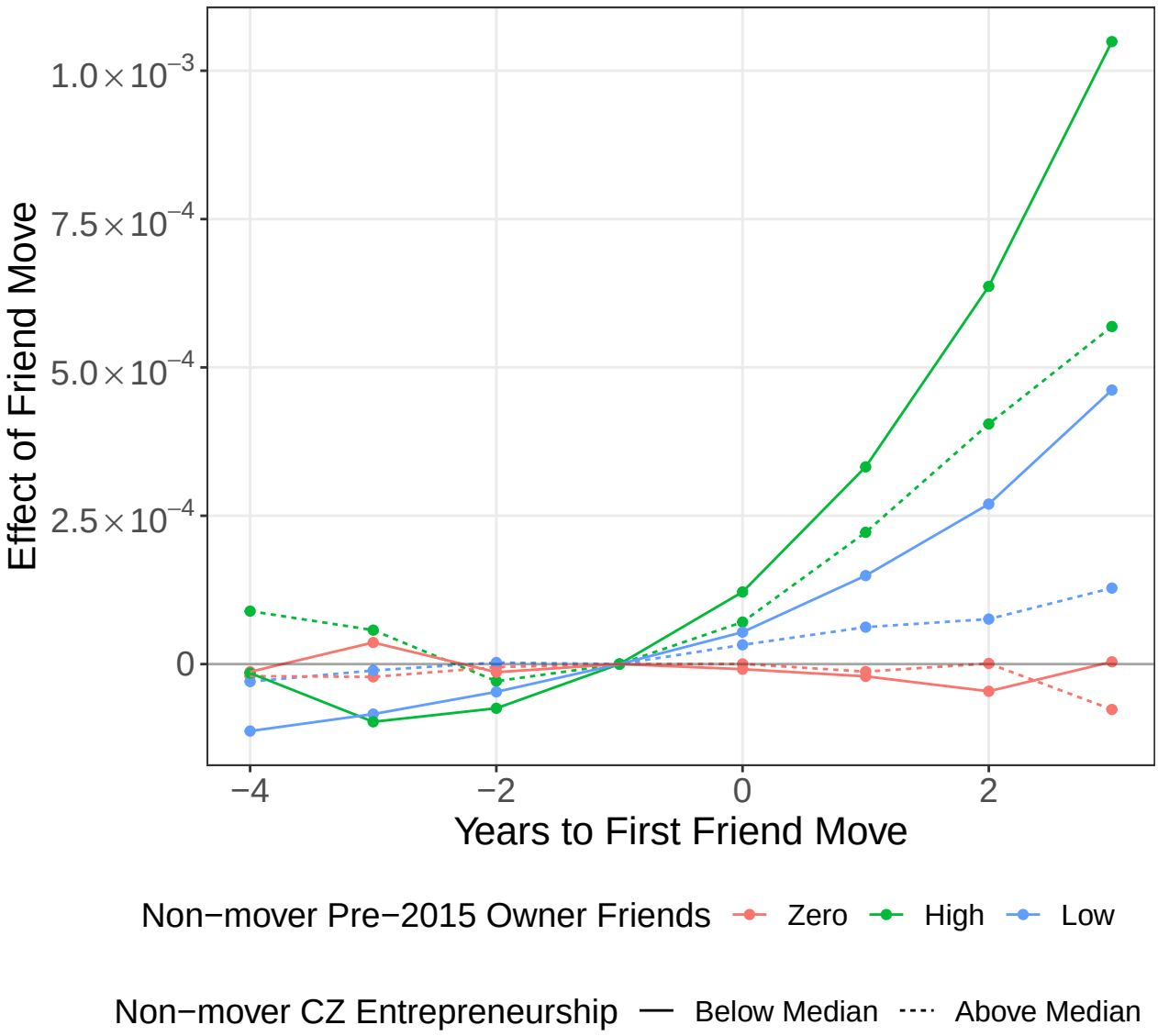
Notes: This figure repeats the analysis in Figure 10 but only considers movers who do not start a business. As with Figure E21, the same technical designs are used from Figure 10 to produce the results in this figure. Further analysis of the figure and comparisons to Figure 10 are given in Section III.B.

Figure E23: Peformance of Katz Centrality Estimator in Public Data



Notes: This figure gives an individual-level scatterplot that validates our Katz-Bonacic estimator. As further background, we estimate Katz-Bonacic centrality using a sampling-based approach described in (Lin et al. 2021). This estimator provides a method for estimation of individual-level Katz-Bonacic centrality in networks with many nodes, and this figure benchmarks our implementation of the sampling algorithm against true Katz-Bonacic centrality. Specifically, it does this in a sample of the Facebook network with 4039 nodes ((Jure Leskovec n.d.)). Details on the design of the model and estimator are given in B.1.

Figure E24: Movers Design: Effect of Friend Migration by Non-Mover's Commuting Zone Entrepreneurship Rate and Pre-Move Network



Notes: This figure presents difference in difference estimates of the effect of a mover to a commuting zone with above-median entrepreneurship rates, as measured by the average number of new jobs per capita at firms less than one year old between 2010 and 2020 in the Business Dynamics Statistics. Estimates are by the fraction of the non-mover's pre-2015 network that are business owners, and by above/below median rates of entrepreneurship in the non-mover's commuting zone. The 3 bins of network exposure to business owners are defined by individuals who have no friendship with entrepreneurs, individuals with fewer owner friends than would be expected from random sampling, and individuals with more owner friends than would be expected from random sampling. The outcome is the rate of firm founding among non-movers who are friends with a mover. Moves between 2016 and 2019, inclusive, are used as treatment. Control groups include movers between 2017 and 2020. Firm ownership is measured up to the year prior to the move in the control group. The approach is described further in Section V.C.1.