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

Robert B. Fluegge† and Mike Bailey‡

July 2025

Abstract

How important are social connections in the decision to start a new business? We assemble a new dataset using privacy protected social network data from 628,000 U.S. firm owners and 70 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 74% 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 from 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 Movers quasi-experimental design. Simulated policy counterfactuals suggest that targeting a hypothetical business-creating intervention to high-impact areas increases firm entry by 32% compared to targeting low-entrepreneurship places, with the difference driven primarily by large spillovers in areas with high existing entry rates.

*We thank Raj Chetty, Shubhdeep Deb, 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 1970). 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 interaction-based spillovers among entrepreneurs are economically meaningful despite the seemingly small magnitudes estimated in typical quasi-experimental settings. To do so, we construct a new privacy protected dataset from 128 million individuals born between 1966 and 2000 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. 628,000 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 evaluate the implications of this variation for policy by constructing an economic model of entry spillovers. Using the estimated model, we argue that the effect of a given intervention on firm entry is likely to depend strongly on the places and subgroups

[1] Policies intended to promote entrepreneurship are motivated and described in (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., Z. J. Acs and Audretsch 1988; 2018), and are intimately related to market concentration (Bresnahan and Reiss 1991, De Loecker et al. 2020)

targeted, with most of the variation across potential targets 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 4.7 friendships with business owners. Each friendship with a business owner is associated with a 0.13 percentage point increase in the probability that an individual owns a firm, a strong relationship relative to an overall entrepreneurship rate of 0.88% 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 either demographic characteristics or individual-level socioeconomic status. Entrepreneurs cluster particularly strongly within narrowly-defined industries: an entrepreneur with 2 friends in one of 1059 6-digit NAICS industries is 86 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.

Under the College design, we estimate that each 2nd-degree connection between a student

[2]While many startups in the U.S. economy do not produce substantial economic activity (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.

and a business owner increases the probability that the student starts a firm after graduation by 0.005 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 41 connections, increases the probability of founding a firm by 0.19 percentage points compared to a base entrepreneurship rate of roughly 0.26% among our sample of collegegoers. This scaling demonstrates a key theme of the quasi-experimental results: modest spillover effects on entrepreneurship at the connection-level, similar to those found in previous quasi-experimental studies, translate to economically meaningful aggregate social spillovers due to the size of individuals' extended networks. Moreover, there is no evidence that firms created as a result of these college connections are negatively selected, as the marginal firms have a higher probability of reaching $1 million in annual revenue compared to the average firm in the sample.

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 by nearly 7 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 treatment *only* 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 make connections with female entrepreneurs at a rate that is 35%

lower than the average male makes connections with male entrepreneurs. Similarly, using estimates of socioeconomic status developed in previous work (Chetty et al. 2022c), 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 147% more connections to entrepreneurs with more than $1 million in annual revenue compared to individuals below the 90th percentile of the income distribution. These partly reflect differences in resources early in life: meaningful gaps in connections to entrepreneurs exist in adulthood 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. Our data demonstrate that there is substantial variation in such connections across areas. For example, individuals living in commuting zones where the average resident has an above-median rate of entrepreneurship in their network are connected to 52% more entrepreneurs than individuals in below-median commuting zones. 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 geographic markets where spillovers matter most, we implement a second quasi-experimental design which we call the "Movers design". The Movers design exploits the timing of friends' migration decisions across local labor markets, measured using commuting zones (CZs), 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, where rates of entrepreneurship can be measured in a variety of ways. The effect

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

is primarily driven by movers who are likely to form new connections to business owners based on pre-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. This pattern implies that a mover only has a large effect on their non-mover connections when the mover has both the opportunity to make friends with successful entrepreneurs and has a history of forming such friendships. 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. The model allows us to assess the economic magnitude of the spillovers and provide guidance on the policy implications of our findings. We illustrate the 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, Bonacich 1987), captures the scale of the target individual's spillover effect on both immediate connections and indirect connections throughout the networked economy. Notably, the additive separability between direct and spillover effects in the model's solution allows for 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 Movers quasi-experimental design to separately estimate the direct and spillover

[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.

effects. Estimating each component separately allows us to empirically evaluate the potential trade-off between direct effects and spillovers under different approaches to targeting. However, this approach requires 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 statistics on individual-level centrality distribution 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 fewer firms among targeted 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 possible 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 low entrepreneurship, high direct effects, and high total effects. Our estimated model implies that targeting an average person in commuting zones with above-median total effects increases firm entry by 32% relative to the same intervention in a commuting zones with below-median rates of entrepreneurship. The improvement largely comes from targeting places where spillovers are highest: targeting places with high total effects increases entry among those not directly targeted by the intervention by 53% relative to a policy targeting places with below median entrepreneurship rates.

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. Our analysis shows that individuals with many existing connections to entrepreneurs are particularly likely to be on the margin of starting a new firm. To use this information, we extend the model to account for within-place heterogeneity and estimate the extended model, allowing effects to vary by network composition. Counterfactual simulations from the estimated extension show that targeting policies based on network composition mitigates, but does not completely eliminate, the tradeoff between direct effects and spillover effects.

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, 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 reflect important patterns in broader social interactions (Wallskog 2025, Mertz et al. 2024).[6] The results also expand our understanding of the preferences, characteristics, experiences, and opportunities that distinguish entrepreneurs (Chetty, Dossi, et al. 2025).[7] Beyond personal characteristics, a

[5]e.g., Ellison et al. 2010; Glaeser et al. 2010; Chatterji et al. 2014; Guzman and Stern 2015 ; Andrews et al. 2022; Kerr and Mandorff 2023; Mohammadi and Shafi 2024; Akcigit et al. 2025. Recent work finds that the entry of "third places" such as coffee shops may also facilitate spillovers in entrepreneurship (Choi et al. 2024).

[6]Barriers to entrepreneurship for subgroups has long been a subject of study in economics and finance (e.g., Bradford and Osborne 1976; Fairlie et al. 2022e; 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 (Burt and Raider 2002; Stuart and Sorenson 2005; Bhide 2000).

[7]Further evidence is provided in Kihlstrom and Laffont 1979; Hamilton 2000; Lazear 2005; Hurst and Pugsley 2011; **levine2017smart**, 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

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 et al. 2009; Andrews et al. 2022a). Given that the modal firm in our data (and in the U.S. economy) is unlikely to be associated with new intellectual property, we interpret our results as suggestive evidence that non-technical knowledge of business opportunities or practices among entrepreneurs and managers may 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 2025) with empirical evidence that interpersonal interactions vary substantially across place, particularly for under-represented subgroups (Athey et al. 2021).[11] Moreover, our result 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

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; Gompers and Lerner 2002, Gompers et al. 2010; Golosov et al. 2024, Chodorow-Reich et al. 2024; Kerr et al. 2022f; Davis and Haltiwanger 2024, Akcigit et al. 2025). Other work has examined the role of government policy (Howell 2017) and workplace circumstances (Babina et al. 2024a).

[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 2025). 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 et al. 2018; Iaria et al. 2018; Bell et al. 2019; Akcigit et al. 2021; Moretti 2021; T. 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 1970; Saxenian 1996; Stuart and Sorenson 2003; Delgado et al. 2010; Glaeser et al. 2015b)

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. and abroad, contract allocation, grant programs, and the establishment of local business development centers.[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. Social networks, including those represented on the Facebook platform, play a role in household financial decisions, economic opportunity, cultural assimilation, migration decision, and may play a role distinct role in firm entry as consumers learn about products.[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.

[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 (Lerner 2009). Furthermore, despite survey evidence described in Section II 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] Kosanovich et al. 2001; Fairlie et al. 2015a; E. Field et al. 2010; Oosterbeek et al. 2010; Huber et al. 2014; J. Lerner 1999; Chatterji et al. 2014; Babina et al. 2023; Chrisman 2017. 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 et al. 2016).

[16] Bailey et al. 2018a; Bailey et al. 2018b; Chetty et al. 2022c; Chetty et al. 2022d; Bailey et al. 2024b; Koenen and Johnston 2024; Bailey et al. 2022b, Campbell et al. 2024. See Mobius and Rosenblat 2014 for a related overview on social learning

II Data and Descriptive Evidence

This section briefly describes the data and presents descriptive results. Section II.A summarizes the key steps in the creation of the dataset. Section II.B details the key patterns in the social networks of entrepreneurs.

II.A Data Collection and Matching

II.A.1 Sample Summary

Our primary analysis sample, drawn on July 3, 2025, includes privacy protected information on all individuals residing in the U.S. born between 1966 and 2000 who were active on the Facebook platform at least once in the previous 30 days, had between 30 and 3000 U.S.-based Facebook friends, and for whom we measure a home ZIP code. The primary sample consists of 128 million individuals, equal to 86% of the population born between 1966-2000 in the 2015-2019 American Community Survey (ACS).[17] The cohorts in our analysis sample also cover the ages at which entrepreneurs are most likely to found their firms (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. In our descriptive exercises, we also provide evidence that patterns of interaction among entrepreneurs we identify are stronger when restricting attention to friends predicted to have the most in-person interaction based on in-sample survey evidence.

To create a dataset which combines information on entrepreneurs with firm-level data, we link individual-level data with data on currently operating firms from a variety of third-party sources using self-reported job titles. Firms that can be linked to founders are relatively successful; 18% of linked firms have greater than $1 million in annual revenue, and 92% of linked firms have survived for more than 4 years. For analyses in which individual-level entrepreneurship is the outcome, we restrict the analysis sample to individuals who ever report an employer and job title on their public

[17] Survey evidence indicates that individuals between 30 and 49 use Facebook at the highest rates (Pew Research Center 2024).

Facebook profile. This restricted sample contains 72 million individuals, or 65% of the currently employed population born between 1966-2000 in the ACS.

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

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

Our first data 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. 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 spike between 2015 and 2016, driven entirely by an increase in the number of observations from the largest data provider in that period. We benchmark these data against an estimate of the annual number of business registrations through 2016, presented in the green series. Benchmark estimates were compiled using data from 46 state-level registration datasets, adjusted for firms that were registered in one state (such as Delaware) but operate elsewhere (Andrews et al. 2022). To account for the omission of proprietorships in the registration data, we also present a simple adjustment in the pink series which scales the number of business registrations by the ratio of all firms to registered firms derived from surveys of firms 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 number of establishments in the Customer Graph spikes to the estimated number of all new businesses including proprietorships. Appendix Figure E3 presents a scatterplot of number of new Customer Graph establishments per capita vs the number of new business registrations per capita at the county level, which have a population-weighted correlation of 0.69 in 2016. Descriptions and sources of

county- and commuting zone-level entrepreneurship measures are presented in Appendix Table D1.

Establishment-level records from the Customer Graph are then matched to establishment-level Facebook Pages using the name, location, and contact information of the establishment. After matching, there are 3.5 million matched Pages that can be linked between to the Customer Graph. We refer to pages which can be linked to establishments in the Customer Graph data as "verified Pages".

We provide a variety of benchmarking analyses for the set of verified Pages. The blue series in Appendix Figure E2 presents the number of new verified Pages by year, which we compare to the number of new employer establishments in the Business Dynamics Statistics (BDS) in the red series. Appendix Figure E4 shows that the number of verified Pages lines up well with the number of new employer establishments between 2015 and 2020 in the BDS for the 9 largest 2-digit NAICS sectors. These comparisons suggests that the verified Pages are likely to be economically meaningful entities, and are fairly representative of the set of new employer establishments across both industries and geographic markets.[18] Appendix Figure E5 presents an example Page.

II.A.3 Identifying Entrepreneurs

To match establishments to entrepreneurs, we use the self-reported 'employer' and 'job title' fields on the public Facebook profiles of individual users. We identify entrepreneurs as users who list a Page as their employer and who list a job title that satisfies a fuzzy string match to the terms 'Founder', 'Owner', or 'CEO'.[19] We omit verified Pages that have more than 4 individuals who have ever claimed to own the business associated with the Page, and provide robustness checks to verify that our analysis holds when considering only Pages claimed by exactly one owner.

The matching procedure between verified Pages and individuals identifies 627,747 entrepreneurs, including 321,544 who entered the market since 2015, out of the sample of 72 million U.S. Facebook users who have ever listed at least one job with a job title. This dataset allows us to analyze the social network history of the entrepreneur influences an entrepreneur's starting year, industry, and

[18] Our data slightly over-represent the "Other Services" sector, which is made up primarily of personal services like gyms and salons.

[19] We include 'Owner' and 'CEO' in the fuzzy string match because random samples of verified Pages indicate that many users who are clearly the founder list themselves as one of those two titles. For example, Page-user pairs which share the same name often have a job title listed as 'Owner' or 'CEO'. Robustness checks for our descriptive and quasi-experimental analyses demonstrate that the key relationships hold when restricting the fuzzy string match to include only users with "Founder" in their job title.

success (from Customer Graph). Table 1 presents sample statistics for the analysis sample. Age and gender distributions of the entrepreneurs we identify are presented in Appendix Figure E6. Our sample is slightly younger and more female than the respondents in the Annual Business Survey, which targets owners of employer firms in the United States. Appendix Figure E7 demonstrates that nearly all entrepreneurs started a firm more than 4 years ago, with a substantial number having started prior to 2015. Table 2 presents the average number and share of owners in individuals' networks. Appendix Table D2 breaks down the number of owner friends by decile of the individual-level distribution.

II.B Descriptive Homophily among Entrepreneurs

The strongest and most consistent pattern in our newly linked data is the tendency for entrepreneurs to form links with other entrepreneurs. The disproportionate level of interaction, or 'homophily', among entrepreneurs is demonstrated in Figure 1. The plot presents the distribution of the number of connections who own a business separately for people who own vs do not own a business themselves. The average entrepreneur knows more than 21 business owners, while the average non-entrepreneur knows only 4.6 business owners. Differences at the bottom of the distribution are even more stark; nearly a third of non-owners know zero owners, while fewer than 3% of owners have no friends who are owners. Figure 2a shows that going from zero to 6 friends who own a business is associated with an increase in the probability that an individual is an entrepreneur from roughly 0% to 0.82%, similar to the base entrepreneurship rate of 0.88% in the sample. The observed homophily is primarily driven by differences in the composition of friends rather than the total number of friends: Appendix Figure E8 demonstrates that the fraction of friends who own a business is highly predictive of one's own probability of starting a firm.

The number of friends who own a business is also predictive of business success. Figure 2b shows that each friendship with a business owner is associated with an increase in the probability that an individual owns a firm with greater than $1 million in self-reported annual revenue by 0.024%, or 16% of the base rate. Analogous plots showing that the pattern holds for firms with greater than $10 in annual revenue and greater than 50 employees are presented in Appendix Figure E9.

Appendix Figure E10 demonstrates that the number of friends with annual revenue of greater than $1 million is somewhat stronger predictor of firm ownership than the number firm owners with any level of revenue.

The observed homophily appears to reflect in-person interactions rather than online-only connections. Using in-sample survey data on in-person interactions as training data, we form friendship-level predictions of off-platform tie strength using a variety of friendship-level predictors, including frequency of messages and tagging in photos. Appendix Figure E11 demonstrates that the descriptive patterns of homophily are stronger when restricting to the top ten closest friends by this metric. Appendix Section A.2.4 describes our predictions of closeness in detail.

II.B.1 The Predictive Power of Connections vs Other Covariates

How does the predictive power of relationships with business owners compare to other predictors? Figure 3 demonstrates that, while demographic and socioeconomic factors have some predictive power for individual-level entrepreneurial status, the number of owners in an individual's network is by far the most predictive characteristic. To show this, we first provide a set of coefficients from cross-sectional univariates regressions of an indicator for being an entrepreneur on the number of connections to owners and on the total number of friends. We compare these coefficients to coefficients from univariate regressions on individual characteristics including income, college attendance, age, gender, and marital status, as well as on-platform engagement in church groups, which are known to facilitate connections across the income distribution (Chetty et al. 2022d). The number of connections to owners is the strongest predictor by far. Moreover, the effect loads almost entirely on the number of connections to owners when all predictors are included in a multivariate regression, suggesting that these individual-level associations with entrepreneurship may be partially mediated by social networks.

Appendix Figure E12 demonstrates that univariate relationship between entrepreneur status and social network composition also holds across time. Recent founding behavior is predicted by friendships that started prior to founding, including friendships with entrepreneurs who started a business more than 10 years ago. Despite the relative homogeneity in the predictive power

of connections across friendship duration and founding year, we note that a meaningful fraction connections among existing entrepreneurs can be explained by behavior around the time an individual founds their firm. Appendix Figure E13 shows that entrepreneurs make friends with other entrepreneurs at the highest rate in the year when the individual founds their firm and the two years following.

II.B.2 Industry-Specific Friendships

Friendships among entrepreneurs are particularly likely within narrowly-defined industries. Using the 68% of the entrepreneurs for whom we measure industry at the 6-digit NAICS level, comprising 1059 distinct industries, the red bars in Figure 4 demonstrate that going from zero to 2 friends in an industry is associated with an increase the relative probability that a given owner works in that industry of 86 times. The green and blue series demonstrate that homophily is far weaker for other industries in the same industry group (4-digit NAICS) and sector (2-digit NAICS), respectively. Appendix Figure E14 presents the distribution of friends in the same industry, industry group, and sector among owners.

II.B.3 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?

Income Gaps Figure 5a demonstrates that the average number of friendships with firm owners increases substantially with income. The pattern is particularly apparent at the top of the income distribution: the average individual above the 90th percentile of the income distribution has 90% more friendships with firm owners than those below the 90th percentile. Figure 5b shows that the relative gap increases to 147% when considering connections to firm owners with greater than $1 million in annual revenue. Appendix Table D3 describes the distribution of owner friends by own socioeconomic status.

These relative differences partly reflect differences in socioeconomic status at birth: Figure

E15 shows that individuals whose parents have incomes above the 90th percentile have 35% more connections with firm owners in adulthood than those from lower-income households. Appendix Table D4 describes the distribution of owner friends by parent socioeconomic status. Differences by income are not associated with patterns in treatment effect magnitudes: quasi-experimental evidence in Section III.A will demonstrate that individuals from high income households do not disproportionately benefit from connections to successful business owners.

Gender Gaps Gender differences in connections to entrepreneurs require a more nuanced analysis. Table 2 provides statistics on the distribution of connections to owners separately by the gender of the individual and the business owner. Men and women are connected to the same share of entrepreneurs as a proportion of their network, at roughly 1.1%. However, the similarity in average exposure disguises heterogeneity by the gender of the entrepreneur. When comparing the mean share of owners in an individual's network between males and females, women have 35% (0.4pp) fewer friendships with female business owners than men have with male owners.

Quasi-experimental evidence in Section III.A will indicate that connections to entrepreneurs have the largest causal effect when the individual and the connected entrepreneur are the same gender. This pattern means that aspiring female entrepreneurs benefit most from connections to female business owners. In this sense, aspiring male entrepreneurs may derive a greater total benefit from their social networks despite men and women having similar shares of owners among their friends.

II.B.4 Geographic Variation in Owner Networking

The extent to which residents of an area form connections with business owners varies substantially across geographic markets in the U.S. Figure 6 plots a county-level map of the average fraction of a person's friends that own a business. The magnitude of the variation across counties is meaningful: 1.0% of friends in a CZ in the bottom 50 percent of (population-weighted) owner networking are owners, compared to 1.6% in a CZ above the 50th percentile when restricting attention to friends of individuals who report a work position. The map also reflects a degree of broader regional variation

in the level of connections, with residents of the Upper Midwest, Pacific Northwest, Northeast, and portions of the Mountain West connecting to entrepreneurs at the highest rates.

While differences in owner networking across places are predicted by the rate of local business ownership, they also reflects differences in rates of friendship formation between the general population and entrepreneurs. Appendix Figure E16, which shows a county-level scatterplot of the average fraction of a person's friends who own a business vs the share of the sample that owns a firm, demonstrates meaningful unexplained variation in friendships with firm owners. Appendix Table D5 provides data on the fraction of firm owners among residents and the fraction of owners among friends of residents in our data for the 30 largest commuting zones.[20]

Previous theory and indirect evidence has suggested that variation in connections to entrepreneurs may be a driving force behind the geographic variation in entrepreneurship rates (e.g., Fluegge 2025). This hypothesis suggests makes a simple prediction: if connections to owners are not an important mediator of geographic variation in entrepreneurship, then two individuals who live in places with different entrepreneurial opportunities but who have the same number of connections to owners should become entrepreneurs at different rates. On the other hand, if local entrepreneurial opportunities are mediated entirely by individual networks, then individuals who have the same number of connections to entrepreneurs should start firms at the same rate regardless of the rate of entrepreneurship among others in their geographic market.

We provide a descriptive test of the hypothesis that connections facilitate local entrepreneurship by running two cross-sectional individual-level regressions of own entrepreneurship status on the in-sample rate of entrepreneurship in the individual's commuting zone, which captures the opportunities for potential entrepreneurs in the individual's geographic labor market. The first regression has no controls, and the second controls for the number of business owners in an individual's network. The results, presented in Appendix Figure E17, indicate that the coefficient on the local entrepreneurship rate falls from 1.00 to 0.31 when connections to other business owners are included in the regression. This substantial decline in the loading on local entrepreneurship rates when holding connections fixed suggests that networks may act as an important mediator of

[20]The average share of friends who are owners differs from the owner rate in the data because owners have more total friends than non-owners.

geographic variation in entrepreneurship.

III Quasi-Experimental Designs and Evidence

Friendships between business owners may reflect causal effects of friendships or non-random network formation. This section develops two complementary quasi-experimental designs aimed at uncovering the causal effect of connections to entrepreneurs on business formation. The first is based on connections formed in college; the second exploits the migration of friends to high-entrepreneurship markets.

III.A College Connections

We first provide direct evidence that connections with entrepreneurs have a causal effect on business formation using plausibly exogenous link formation at the start of the first year of college. The logic behind this design is that early-college connections likely reflect features of the environment that students have little control over, such as dormitory assignments and orientation groups. Long-term connections formed in college are known to be substantially influenced by residential proximity (Marmaros and Sacerdote 2006), and proximity has been specifically identified as the primary determinant of new acquaintances among recently-arrived college students (Newcomb 1961).

Our design exploits this phenomenon by identifying new indirect connections with business owners created by links among students in the first few weeks of college. To operationalize this approach, we compare new students who form more connections to other students who have existing connections to business owners with students who form fewer such connections. Our identification is therefore based on variation in *2nd-degree connections* – that is, differences in the number of business owners among friends-of-friends – generated by early-college friending. Intuitively, our identifying assumption in the baseline specification requires that students with a higher underlying propensity to start a firm more than 4 years later do not disproportionately develop connections to students who know existing business owners at the very start of college. We provide evidence supporting this assumption by showing that early-college friendships which create indirect connections to business owners are unrelated to the number of business owners that were already in an

individual's network, which is the best predictor of an individual's propensity to form new connections to business owners prior to arriving on campus. We also present a within-industry analysis which is valid under narrower identification assumptions, as well as evidence on mechanisms.

III.A.1 Design

The opening weeks of college are among the most social periods of a college-goer's life. Figure 7a 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 U.S. 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 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.

Our quasi-experimental treatment uses characteristics of the 2nd-degree connections (i.e., friends-of-friends) formed during the first 12 weeks of the academic year. The individual-level treatment for focal individual i will be defined as the number of new second degree connections j who satisfy the following conditions. Let k denote an individual who was born in the same year and went to the same college as i.

1. j had started a business prior to the year that i and k turned 18.
2. j and k were friends with each other prior to the year that i and k turned 18.
3. i was not friends with j or k and had no mutual friends with either prior to the year that i and k turned 18.
4. During weeks 33-44 (mid-August to Thanksgiving) of the year i and k turned 18, i becomes friends with k

Define the treatment T_i to be the number of individuals j who satisfy these conditions for focal person i. The distribution of T_i for collegegoers in the sample is presented in Appendix Figure E18.

These conditions ensure that variation in T_i is only generated by links between i and k: j's friending behavior and founding activity are held constant in the analysis. Note also that any new connections which had a mutual friend are dropped from the data, such that connections which may have been formed from introductions among existing friends are not considered.

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 k have little information about each other at the time they form the friendship. Section III.A.3 presents evidence that new 2nd-degree connections during this period are unrelated to the focal student i's characteristics.

III.A.2 Baseline Estimates

Based on the linearity in levels exhibited in Figure 2a, 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 by 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 control vector X_i contains the number of new owner 2nd-degree connections formed between the start of the calendar year until students arrive on campus. This control is included so that the results are not driven by (potentially selected) connections formed with future classmates between admissions decisions and arrival on campus.[21]

OLS estimates of β in are presented in Table 3. The magnitude in our baseline specification in Column (1) implies that one extra 2nd-degree connection to a business owner increases the probability that an individual starts a business after graduation by roughly 0.0047 percentage points. This connection-level effect is relatively modest. However, individuals have many such connections. Scaling the connection-level effect by the p90-p10 difference in the number of 2nd-

[21] Evidence for "early friending" among students prior to the start of the semester can be seen in the divergence between friending rates of college-goers and non-college-goers in March of the year the individual turns 18 shown in Figure 7a.

degree connections to business owners among all first-year college students – a difference of 41 2nd-degree connections – implies a change of 0.19 percentage points in the rate of entrepreneurship after graduation, or roughly 74% of the base rate of entrepreneurship in this sample. Column (2) shows that the estimate does not result from a correlation between overall friendliness and future entrepreneurship, as the coefficient increases slightly to roughly 0.0066 percentage points when controlling for the total number of friends the focal student makes during the treatment window.

The results are not driven by co-founding behavior. Column (3) presents estimates from a model where the outcome is redefined to be starting a firm without co-founders. While the connection-level treatment effect of 0.0038 percentage points estimated in Column (3) is slightly smaller than in the baseline model, going from the 10th percentile to the 90th percentile of the treatment implies a change equal to 84% of the base rate, larger than the relative magnitude of the baseline estimate. Appendix Table D6 demonstrates that the findings are robust to various control vectors, sample restrictions, and outcome definitions, including models in which the outcome is redefined as an indicator for starting a firm with greater than $1 million in annual revenue. We find similar effects as a fraction of the base rate.

There is no evidence that the marginal firm started in response to the treatment is negatively selected relative to the average firm founded by individuals in the sample. To show this, 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 point estimate of 0.32, presented in Column (8) of Table D6, is larger than the share all firms in this sample which reach $1 million in revenue, equal to 0.14. While the 2SLS estimate is somewhat imprecise, we cautiously interpret this result to mean that the marginal entrepreneurs are not negatively selected on ability; rather, the marginal entrepreneurs may have simply lacked the connections required to "get their foot in the door".

Effect Heterogeneity by Gender Effects are strongest for females when they form connections with female firm owners. Table 4 provides estimates from a regression which includes the number of

male and female owner connections separately, interacted with the gender of the student. Columns (1)-(4) present coefficients from the corresponding gender-specific univariate models, and Column (5) presents estimates from a model which includes the interaction between the focal individual's gender and the gender of the connected owner. The estimates in Column (5) imply that for a female student each connection to female owner increases the probability of founding after graduation by 0.012 percentage points, while connections to male owners have no effect on founding behavior.

For male students, each connection to a male firm owner increases the probability of founding after graduation by 0.010 percentage points and each connection to a female owner increases founding by 0.011 percentage points. That is, male students benefit from connections to both male and female entrepreneurs, while female students benefit only from connections to female entrepreneurs.

Effect Heterogeneity by Income Appendix Table D7 examines effects by the predicted parent income of students and the income of the entrepreneur. Each connection to an entrepreneur with income above the 90th percentile of the income distribution increases the probability of founding four or more years later by roughly 0.02 percentage points, while each connection to a less successful entrepreneur appears to have a more modest effect. However, there is little evidence that connections higher-entrepreneurs disproportionately benefit students from different income backgrounds. Column (6) presents estimates from a model which interacts socioeconomic status of the entrepreneur with the parent income rank of the student and find a coefficient which is small in magnitude and statistically indistinguishable from zero.

III.A.3 Falsification Tests

The identification assumption underlying the baseline causal estimates depends on high-ability individuals 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 7b 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 treatment window. However, each of the estimates on the left side of Figure 7b, 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 roughly 0.7SD 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). We therefore conclude that for any unobserved selection to contaminate our treatment effect estimates, the selection must be unrelated to friending behavior in all periods leading up to the start of college.

III.A.4 Within-Industry Design

We next investigate whether the clustering of owner networks within industry that we documented in Section II.B reflects a causal effect. This analysis is useful from two perspectives. First, it substantially reduces the set of plausible threats to identification, as we will show that for a selection effect to explain the results such selection would would need to operate within but not across the

set of narrowly-defined industries. Second, the within-industry analysis provides evidence against a class of possible mechanisms.

The identifying assumption in the within-industry design is that students who have a higher-propensity to found a business in one of 1059 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, an effect operates within-industry would suggest that our design uncovers a causal effect rather than nonrandom network formation. Formal identification results in class-specific models show that effects on closely related industries may be treated as overidentification conditions to distinguish between models of selection (Bell et al. 2019).

To perform the within-industry analysis, our target estimating equation is:

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

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. 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 defined by school $\times$ cohort $\times$ industry $\times$ number of connections.[22] In order to increase the statistical 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 (4) presents estimates of β from the collapsed approach. Overidentification tests in Columns (5) and (6) replace the explanatory variable T_{ik} with the number of connections

[22]For 1 million students and 1000 industries, estimating Equation 2 directly requires running a regression on a student-by-industry-level dataset of 1 billion rows, which exceeds our computational constraints. The same computational approach for estimation in high-dimensional class-specific data is used in (Bell et al. 2019).

in a *different industry but same industry group* (4-digit NAICS) or *different industry but same sector* (2-digit NAICS), similar to the approach in Bell et al. 2019. The effect loads entirely on the same narrowly-defined industry: an extra 2nd-degree connection in industry k has no effect on firm creation in related industries. In order for these patterns to be explained by nonrandom formation rather than by treatment effects, students with a high industry-specific entrepreneurial ability would need to select friends in *only* the industry where they will eventually found, but not in any related industries. Given the granularity of the industry classification, we view such narrow selection as unlikely to be driving our results.

We also view the the within-industry results as inconsistent with a class of explanations for the baseline effects. Existing evidence suggests that many entrepreneurs start firms for the nonpecuniary benefits, such as the agency or flexibility associated with "being one's own boss" (Hurst and Pugsley 2011). These features likely apply to entrepreneurs in many industries. Therefore, if the primary mechanism was communication about nonpecuniary job quality one would expect the treatment effects of network exposure to apply both across and within industries. The lack of effect on entrepreneurship in similar industries provides evidence against this sort of broad, nonspecific spillover. Rather, we view the industry specificity of the effects as suggestive evidence in favor of models where social spillovers transfer specific skills or information about opportunities relevant to a given industry.

III.A.5 Effects of Indirect Connections

One might have expected that the quasi-experimental treatment only matters inasmuch as it shifts the focal individual's direct (1st-degree) connections to entrepreneurs around the time the focal individual actually starts or considers starting their firm. However, several pieces of evidence suggest that the results are not explained by shifting entrepreneurship among direct connections.

First, the results are unlikely to be explained by second-degree connections to entrepreneurs which later become direct connections. Such conversions from 2nd-degree connection to direct friendship are exceedingly 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 period.

Second, our instrument does not induce existing direct connections to start businesses at rates high enough to explain the results. Conditioning on the mutual connection (person k in the formal definition of the treatment) not starting a business does not change the effect magnitudes. The the mutual friend's entrepreneurial behavior does not explain the effect for a simple statistical reason: the baseline rate of entrepreneurship in the sample are low, so dropping every person k who starts a business from the analysis does not appreciably change T_i.[23]

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 both models of agglomeration based on entrepreneurship and the 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 focal individuals, Person A and Person B, both live in Boston. Person A is friends with Person X, who moves to San Francisco in 2016. Person B is friends with Person Y, who moves to San Francisco 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 operationalize and generalize this basic idea by constructing stacked difference-in-differences estimates for every 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 San Francisco for Persons X and Y in the example. Constructing estimates at the level of commuting zone pairs

[23] Note that finding an effect even when the existing connection does not start a business does not imply that entrepreneurship among direct connections has no effect. We find that the effect magnitudes are substantially larger when the mutual friend does start a business.

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, removing any selection from the focal individual or mover's location. We choose commuting zones as the unit of observation in order to capture the local labor markets faced by both nonmovers and movers. Only friendships formed prior to the start of the panel (that is, before 2015) are considered in the design so that endogenous link formation during the sample period 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 individual i has founded a firm by year t and zero otherwise, and 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 E19 provides evidence that moves are 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 8a 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.[24] The estimates in periods $t \geq 0$ indicate that friends of a mover to a city with many successful entrepreneurs, represented in the green series, are 0.03 percentage points more likely to have started a firm 3 years after the move. The patterns are not sensitive to splits based on other definitions of local entrepreneurship, including firm entry or establishments per capita.

Friends of a mover 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 affect 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 the periods prior to the friend moving, implying that individuals exposed to the quasi-experimental treatment earlier do not start firms at different rates prior to their friend's move. The absence of pretrends also suggests that the estimated effects are not driven by selective reporting of entry 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 joined the Facebook platform or was claimed by the founder. Because of this feature of the linked data, any selection effects of friend migration on either existing firms joining the platform

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

or founders claiming ownership of existing firms 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 report numerical 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 $\hat{\beta}_{ndk}$ our estimate of CZ-pair-level treatment effects and $k \geq -1$, 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 created per capita in the commuting zone by firms less than 1 year old. Columns (1) and (2) build up to Column (3) in Table 5, which presents estimates of $\beta_{interaction}$.

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

The patterns of entry among friends of movers to high-entrepreneurship markets 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:

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

where Y_{its} is equals one if individual i has founded a firm in industry s by year t. Appendix Figure E20 plots estimates of β_{ndks} averaged over high-, medium-, and low-entrepreneurship markets at the CZ-by-sector level, calculated from the county-by-sector data in the Business Dynamics Statistics. We find qualitatively similar patterns as in the baseline specification, suggesting that

the industry-specificity found in the College design holds for variation generated by movers as well.

Table 5 reports numerical estimates of CZ-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.

Column (4) of Table 5 provides estimates of Equation 7. 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. 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, analogous to the industry-specific results in the College Design.

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. However, directly comparing movers who make different numbers of friends with entrepreneurs in the destination may be contaminated by 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 E21 demonstrates that friendships with entrepreneurs prior to 2015 is the strongest predictor of the number of entrepreneur friends a person makes post-2015 relative to other individual-level covariates in our data.

Accordingly, Figure 8b 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. We interpret the results as suggesting that effects in the Movers design are driven by movers who have both a history of forming links with entrepreneurs *and* who have the opportunity to form such links after they move.

The result is not driven by movers who themselves become entrepreneurs. Appendix Figure E22 demonstrates the the results are nearly identical when omitting movers who become owners after moving. We interpret this result as evidence that networks transmit information regardless of the entry decision of the mover, similar to the evidence in the College design that conditioning on future entry of classmates did not meaningfully change the effects.

IV Model and Estimation Approach

What can the data tell us how spillover effects interact with policies aimed at promoting entrepreneurship?[25] To answer this question, we start by constructing a peer effects model which incorporates human capital spillovers through social networks. We model human capital as a determinant of entry which helps 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.

The goal of the model is to provide a basis for counterfactual policy simulations, which we describe in Section V. Our counterfactuals will require place-specific estimates of network-based spillovers relevant to firm entry. Section IV.A starts by characterizing entry and spillovers at the level of individuals within a network. Section IV.B details how the model can be aggregated to the level of a local area. Section IV.C describes the approach to place-specific estimation based on

[25]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.

the Movers design at the commuting zones level.

Appendix Section B.2 motivates the focus on firm entry by proving two welfare results under an extension of the baseline model. The first result describes conditions under which policies that produce spillovers will increase aggregate welfare even in the presence of negative business-stealing externalities. The second result shows that the effect of a hypothetical intervention on firm entry proportional to the welfare effect when the intervention 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 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. 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.

We model an individual's human capital E_i as a linear function of the individual's unobserved ability 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. The parameter γ controls the rate at which effects decay through the network. 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 $D_{i,t'} \geq D_{i,t} \;\; \forall \; t' > t$.

Equation 9 specifies a structure which is linear in levels of human capital rather than the mean human capital among an individual's connections. This choice is motivated by the linearity in levels in the descriptive results, where the (predictive) effect of an extra link to a business owner on firm entry does not vary with the number of existing connections. This linearity can be seen in Figure 2a, and is contrasted with the convexity observed when the explanatory variable is the share of friends who own a business in Appendix Figure E8

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

This model implies a unique equilibrium in the vector of individual-level human capital $\vec{E}_t$, which we denote $\vec{E}^*_t$, where Equation 9 is satisfied for all individuals at all points in time.[26] The equilibrium for person i can be expressed as:

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

[26] The equilibrium is unique when γ is less than $\frac{1}{||G||}$ for $||G||$ the largest eigenvalue of the adjacency matrix. Appendix Section C.1.1 provides evidence that this technical condition is satisfied.

where $[G_t^n]_{i,j}$ denotes the (i,j) element of the n-th power of the adjacency matrix G_t. The quantity $[G_t^n]_{i,j}$ can be interpreted as the number of paths of length n between individuals i and j in the network at time 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 $\tau_{i,t}$, is:

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

To keep the notation concise, 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 $B_{i,t} \equiv \frac{1}{K_{i,t}} \sum_{j,n} \gamma^n \left[G_t^n\right]_{i,j} \beta_{c(j,t)}$, where $A_{i,t}$ and $B_{i,t}$ are weighted averages 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 the effect of the intervention on the targeted individual in the absence of any spillovers. We refer to this as the "direct effect" of the intervention. The direct effect depends on the strength of the intervention δ and the barriers to entry in i's market, $\beta_{c(i,t)}$.

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 is 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.

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 C.1 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 connections to the mover to d with friends in n, and let $\overline{\mathcal{C}} \subseteq \mathcal{C}$ denote an arbitrary subset of cities.. 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 commuting zone c'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. We will refer to this quantity as a commuting zone's "Networking Productivity".[27]

[27] The choice of $\overline{\mathcal{C}}$ affects all CZ-level estimates proportionally, such that the constant of proportionality C in Equation 12 and Equation 13 may vary with the choice of $\overline{\mathcal{C}}$.

Direct Effects Appendix Section C.1 demonstrates that the direct effect of an intervention at time t 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 Movers 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. Section IV.C describes how we estimate NP_c, SP_c, and $\mathbb{E}[K_{i,t}]$ using the movers design.

The choice to consider counterfactual policies targeted to individuals with the mean Katz-Bonacic centrality has two advantages. First, fixing centrality at the mean ensures that the results are not driven by any geographic variation in network measurement error. 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 statistical 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 fine subgroups.

The estimator for Katz-Bonacic centrality is described in Section C.1.1.

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 will tend to underestimate spillovers when network size K_{it} and composition B_{it} are positively correlated. In the data, the fraction of friends who are entrepreneurs and the number of friends are correlated 0.16 in the cross section of individuals, suggesting that neglecting the interaction between network size and composition is likely to produce a modest under-estimate of spillover effects. Formal justification for our approach is provided in Appendix Section C.1, which derives conditions under which K_{it} and B_{it} are conditionally mean-independent at the city level.

IV.C Place-Specific Estimation

Our 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 new tools from the network science literature for estimation of Katz-Bonacic centrality in large networks. We discuss the new approach to centrality estimation 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.

Estimating Movers Effects Recall that the Movers effect $\Delta_{n,d}$ is defined as the average annual effect of a friend moving to destination d on a non-mover friend in commuting zone n, and the effect β_{ndk} from Equation 4 denotes the dynamic difference-in-differences effect k periods after a friend moves to a new commuting zone. To estimate $\Delta_{n,d}$, we take the linear trend in β_{ndk} over post-treatment periods by regressing post-treatment β_{ndk} on relative time k:

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

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)

Estimating the Human Capital Correction Commuting zone-specific human capital endowments A_c are estimated in a set of fixed effect regressions performed separately for each commuting zone. Appendix Section C.1.3 derives an estimator for the human capital correction (up to a constant of proportionality), with the following estimating equation:

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

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 $\widehat{\beta}_{loc,c}/\widehat{\beta}_{nonloc,c}$ provides an estimate of $A_c/\overline{A}$ for an unknown constant $\overline{A}$.

Combining Movers Effects and the Human Capital Correction Networking Productivity and Spillover Production are estimated as weighted means over destinations and non-mover commuting zones, respectively:

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

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

where N_{nd} is the sum of treatment and control individuals for each element of the stacked difference-in-differences design used to estimate the movers effect $\widehat{\Delta}_{nd}$ specific to each pair of commuting zones

(n, d). The terms $\frac{\widehat{\beta}_{nonloc,c}}{\widehat{\beta}_{loc,c}}$ provide the correction for local levels of human capital.

IV.C.2 Estimating Katz-Bonacic Centrality

We estimate Katz-Bonacic centrality by extending and implementing a new sampling-based approach from the network science literature (Lin et al. 2021). The sampling estimator provides a method for estimation of individual-level Katz-Bonacic centrality in networks with many nodes, such as the Facebook social graph, where the calculation of centrality using analytical approaches is difficult. The estimation approach and benchmarking is discussed in greater detail in Appendix Section C.1.1.

The estimator uses the fact that individual-level Katz-Bonacic centrality $K_i \equiv \sum_{n=1}^{\infty} \sum_{j \in \mathcal{I}} \gamma^n [G^n]_{i,j}$ is equal to the sum of paths to all nodes in the network, discounted by a decay factor that scales geometrically with the length of a path. By sampling nodes and counting both direct and indirect paths, an estimator can be constructed which appropriately weights sampled indirect connections. The estimator is unbiased for a truncated version of the series representation of Katz-Bonacic centrality up to a given network distance from the focal individual. That is, given a network distance d and decay γ, we can achieve an unbiased estimate of truncated individual-level centrality $\sum_{n=1}^{d} \sum_{j \in \mathcal{I}} \gamma^n [G^n]_{i,j}$ for any individual i. The estimate of truncated centrality is biased downward relative to true centrality, as every term of the sum is positive.

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 (Leskovec and Sosi 2016). Appendix Section C.1.1 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 0.95 at the individual level in the validation sample.

Estimating Network Decay γ In order to provide an estimate of Katz-Bonacic centrality, the estimator requires knowing the decay factor γ, which we defined in Equation 9. 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 the non-mover's other friends. The intuition for this approach is that, under the model, moving to a high entrepreneurship destination affects entry among both the mover's pre-move friends and friends-of-friends in equilibrium. If the friends and friends-of-friends face the same barriers to entry, then the ratio of the effects identifies the decay parameter.

To formally describe this approach, denote the Networking Productivity in place c among pre-move 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. Restricting both the friends and friends-of-friends of the mover to live in the same commuting zone means that they face the same barriers to entry, which divides out in the ratio.

We estimate γ as the ratio of average $NPfof_c$ and NP_c

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

where N_c is the sum of treatment and control individuals used in the each element of the stacked difference in differences estimate of $\widehat{NP}_c$.[28] We derive the approach in Appendix Section C.1.

Centrality Estimation Results We implement this estimator for a sample of 5000 individuals i and all potential connections j using the analysis data and from the Movers design, considering connections up to $d = 8$ degrees of separation away. As in the Movers design, only connections formed prior to 2015 are considered. Using the approach to estimating γ described above, we calculate that the average truncated Katz-Bonacic centrality is equal to 1.25. To assist in interpreting this calculation under the model, a centrality of 1.25 implies that a shock that increases a given individual's entrepreneurial human capital by 1 unit would increase the combined entrepreneurial human capital of all other individuals in the economy by 1.25.

Furthermore, since our estimator allows for term-by-term estimation, we can decompose the

[28] That is, $N_c = \sum_{d \in \overline{c}} N_{c,d}$

overall estimate by network distance from the focal individual. We focus on the share of centrality explained by connections at distance greater than 1 from the focal individual, equal to $1 - \frac{\mathbb{E}[\gamma \sum_{j \in \mathcal{I}} g_{ij}]}{\mathbb{E}[\sum_{n=1}^{d} \sum_{j \in \mathcal{I}} \gamma^n [G^n]_{i,j}]}$. The calculation implies that, at our estimated value of γ, 80% of the estimated average centrality is due to connections at distance greater than 1.

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 decay factor γ may differ when considering other outcomes. With this in mind, Appendix Table D8 describes quantiles of Katz-Bonacic centrality for both our estimate $\hat{\gamma} = 0.0061$ and for alternative values of γ which may be relevant in other contexts. The table also reports the ratio between the mean total Katz-Bonacic centrality and the mean influence among direct connections only. We hope that these additional calculations will help to inform peer effect models in future studies.

V Targeting Policy Counterfactuals

What does the estimated model imply for policy? In this section, we consider how the effects of a hypothetical policy intended to promote firm entry differ depending on the areas and individuals targeted. Our analysis indicates that policies which maximize entry are targeted to places with high spillovers rather than places with the largest effects on treated individuals. We illustrate the quantitative importance of these spillovers, and the tradeoff between maximizing total effects and maximizing effects on treated individuals, by considering a variety of targeting counterfactuals at both the area level and the individual level.

V.A Direct Effects

Figure 9a shows the signal correlation between the direct effects estimated by Equation 16 and commuting zone characteristics. The modest negative correlations between the direct effect and measures of entrepreneurship in the first 3 rows 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. Appendix Table D9 lists the direct effects in the top 10 commuting zones by total effect, where estimates are shrunk

by the signal-to-noise ratio to minimize mean squared error (Chetty and Hendren 2018).

Correlations between direct effects and CZ-level characteristics other than local entrepreneurship rates are sometimes larger than the correlation with entrepreneurship rates. Population, residential rents, income, and levels of schooling more negatively correlated with the treatment effects than local entrepreneurship. Local capital availability as measured by deposits or 7(a) loans are weakly anticorrelated with the treatment effects, while local patent rates exhibit a weak positive correlation. Versions of social capital based on local economic connectedness, measured by connections to high-income users on Facebook (Chetty et al. 2022c; Chetty et al. 2022d) are not strongly systematically correlated with treatment effects on residents who take up treatment. Measurement of CZ-level characteristics is described in Appendix Table D1.

V.B Total Effects and Place-Based Policy

Figure 9b plots the signal correlation between the estimated total effect and commuting zone characteristics, where total effects are estimated using Equation 13. Notably, the predicted total effect from a hypothetical policy is strongly positively correlated with local entrepreneurship rates. This is surprising when compared with the modest negative correlation between the direct effect and entrepreneurship rates found in Figure 9a, 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.

The correlation between total effects and other CZ-level characteristics also exhibit marked differences compared to the correlations with direct effects. Local population and residential rents are strongly positively correlated with total entry, suggesting that agglomerative forces may be particularly important in places with high spillovers among entrepreneurs. The positive correlation between total entry and capital availability, as measured by deposits and 7(a) loans, suggests potential complementarities between financial development and spillover effects. traditional measures of local human capital do not capture total entry effects: income, college attendance, and patent rates are all uncorrelated or weakly anticorrelated with the total treatment effects. Economic connect-

edness is positively associated with with total entry, but the relatively weak correlation suggests that existing measures of social capital do capture the spillovers relevant to entrepreneurs.

Appendix Table D9 lists the top 10 commuting zones by total effect, where estimates are shrunk by the signal-to-noise ratio.

V.B.1 Policy Counterfactuals

To illustrate the strength of spillovers implied by our estimates, we ask how a hypothetical intervention could be targeted to maximize business creation. Since the estimated model produces quantities proportional to the effect of a hypothetical treatment, we normalize the counterfactual estimates so that the hypothetical intervention creates one new firm among individuals directly targeted by the intervention.

We first consider a counterfactual in which our hypothetical policy is targeted to individuals in commuting zones with few successful entrepreneurs. Column (1) of Table 6 shows that assigning the hypothetical policy to randomly selected residents of a commuting zone below the 50th percentile of startup jobs per capita in the population-weighted CZ distribution creates 0.78 new businesses among individuals not directly targeted by the intervention per business created among individuals targeted by the intervention.

Next, we consider a counterfactual in which the policy is targeted to individuals in commuting zones with the largest direct effects. This counterfactual maximizes the number of firms created among individuals directly targeted by the intervention by assigning the treatment to the average person in a commuting zone above the 50th percentile of direct effects in the population-weighted CZ distribution. The results are presented in Column (2). Under this counterfactual, 31% more new firms are created among the targeted population and total entry is 19% higher than the counterfactual which targeted low-entrepreneurship places.

Finally, we consider a counterfactual in which the policy is targeted to individuals in commuting zones with a high total effect, inclusive of spillovers. These commuting zones constitute the places where the total entry from the hypothetical intervention is maximized, again using a split at the population-weighted CZ median. Column (3) shows that targeting the hypothetical intervention to

commuting zones with an above-median total effect creates 16% more new businesses in expectation among individuals who take up the intervention, which is 15 percentage points fewer than if we had targeted the places with the highest direct effects. However, the model predicts that an intervention targeted to high total effect commuting zones creates 53% more new businesses via spillovers, 39 percentage points more than the spillovers created in the commuting zones with high direct effects. Combining the direct and total effects results in 11% higher entry compared to our counterfactual in which places with high direct effects are targeted and 32% higher entry compared to the targeting places with low rates of entrepreneurship.

These results highlight a tradeoff between maximizing direct and total effects. Policymakers who want to maximize entry among a particular subgroup – in this exercises, among residents of low-entrepreneurship places – may reduce the total entry effects of the policy by doing so. In the next section, we show that the tradeoff is muted when the policymaker is allowed to target within commuting zones.

V.C Individual-Level Effects and Targeting

How can networks be used to target interventions within geographic areas?

We first demonstrate empirically that entry effects depend on the individual-level network characteristics of the target. Using the Movers design, we find that Networking Productivity increases substantially with the number of entrepreneurs among an individual's direct connections. This result implies that the direct effects of the hypothetical policy are higher for individuals who already know many business owners, which we interpret as suggesting that network characteristics may help identify individuals who are marginal with respect to treatment.

We then develop an extension of the model which allows for individual-level heterogeneity in treatment effects. We conclude by estimating the extended model with the goal of identifying high-impact policy targets.

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

The simplest way in which network information may allow for 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 is strongly related to treatment effects 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 by tercile of entrepreneur friends among non-movers.

Networking Productivity is highest among nonmovers who have many existing connections to firm owners but live in a low-entrepreneurship place. The estimates suggest that most of the difference between the effect on this "high-impact" group and the average nonmover is explained by the network composition; 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 suggesting that individuals who had either no opportunity or no inclination to befriend entrepreneurs prior to 2015 are likely not marginal; that is, these 'isolated' individuals are unlikely to engage in entrepreneurship without a more substantial intervention.

V.C.2 Estimating the Extended Model for High-Impact Individuals

We next ask whether targeting individuals with large direct effects within places can help to mitigate the trade-off between direct effects and spillovers. To answer this question, we present estimates from an extended model that considers effects which may vary by individual-level characteristics. Appendix Section C.2 develops the extension of the model which allows for subgroup-specific estimation.

We illustrate the value our subgroup-specific causal estimates by estimating the extended model among individuals with high existing exposure to entrepreneurs identified in Section V.C.1. 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"

individuals whose networks over-represent business owners, and consider the spillover effects on business creation among all individuals.

Column (1) of Appendix Table D10 reports the firms created among commuting zones in the top 50% of the population-weighted CZ distribution for high-impact residents in counterfactual simulations from the estimated model. Column (2) reports the results of counterfactual simulations targeting high-impact individuals in CZs with the highest total effect. Comparing Columns (1) and (2), we find that targeting to maximize total effects increases total entry by 5% relative to targeting individuals with the highest direct effect, at the cost of a reduction of 8% in entry among the treated. However, the tradeoff between direct effects and total effects when targeting under the extended model is smaller than the corresponding tradeoff found in the baseline counterfactuals in Table 6, where both direct and total effects differed by 11% when targeting on place alone. We interpret this result as suggesting that such tradeoffs can be meaningfully improved by targeting on individual-level characteristics.

VI Discussion

This paper has presented new data and evidence on the importance of social connections for the creation of new businesses. Our descriptive, quasi-experimental, and model-based results suggest that the diffusion of business-creating spillovers through social networks results in a substantial 'social multiplier' for individual-level entrepreneurship. The importance of social connections for entrepreneurship is largely a result of the scale of social networks; individuals form enough connections over the course of their life that even relatively small effects at the level of a single connection matter in aggregate.

However, despite the high average level of connectedness, substantial gaps in access to valuable social networks exist for groups that are traditionally under-represented in entrepreneurship. We hope that the analysis assists in identifying the most productive markets to focus the myriad of local entrepreneurship-creating programs for both the typical aspiring entrepreneur and those from under-represented populations.

Our analysis also highlights the potential tradeoffs that policymakers face when making policy

targeting decisions in the presence of spillover effects. Policymakers interested in promoting entrepreneurship face a choice between maximizing the total effect of their chosen intervention among the entire population, inclusive of spillovers, and maximizing the effect among the population intended to take up the intervention. On the other hand, the results suggest that new approaches to program design and targeting which directly incorporate social networks, such as eliciting referrals by existing entrepreneurs to business-creating programs, may be effective and worth evaluating.

More broadly, the results provide large-scale evidence on the benefits of targeting interventions towards influential members of a network. 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 generate increasing returns and bolster economic growth. Mapping the structure of social networks and charting the paths of social spillovers in such markets is a particularly promising avenue for future research.

References

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

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

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

Akcigit, Ufuk, Alp, Harun, Pearce, Jeremy, and Prato, Marta (May 2025). *Transformative and Subsistence Entrepreneurs: Origins and Impacts on Economic Growth*. Working Paper 33766. National Bureau of Economic Research. DOI: 10.3386/w33766.

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

Akcigit, Ufuk, Chhina, Raman Singh, Cilasun, Seyit M, Miranda, Javier, and Serrano-Velarde, Nicolas (Mar. 2025). *Credit Card Entrepreneurs*. Working Paper 33618. National Bureau of Economic Research. DOI: 10.3386/w33618. URL: http://www.nber.org/papers/w33618.

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

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

Andrews, RJ, Fazio, Catherine, Guzman, Jorge, Liu, Yupeng, and Stern, Scott (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, Ferguson, Billy, Gentzkow, Matthew, and Schmidt, Tobias (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, Chen, M Keith, and Popov, Anton (2022). *The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley*. Tech. rep. National Bureau of Economic Research.

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

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

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

Babina, Tania, He, Alex Xi, Howell, Sabrina T, Perlman, Elisabeth Ruth, and Staudt, Joseph (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 Howell, Sabrina T (2024). "Entrepreneurial spillovers from corporate R&D". In: *Journal of Labor Economics* 42.2, pp. 469–509.

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

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

Bailey, Michael, Johnston, Drew, Koenen, Martin, Kuchler, Theresa, Russel, Dominic, and Stroebel, Johannes (2024b). "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, Johnston, Drew, Kuchler, Theresa, Stroebel, Johannes, and Wong, Arlene (July 2022b). "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, Johnston, Drew M, Kuchler, Theresa, Kumar, Ayush, and Stroebel, Johannes (2025). *Cross-Gender Social Ties Around the World*. Tech. rep. National Bureau of Economic Research.

Banerjee, Abhijit, Chandrasekhar, Arun G, Duflo, Esther, and Jackson, Matthew O (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, Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, and Van Reenen, John (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, Kass, Tobey, May, Thomas J, McGrattan, Ellen, and Schulz, Evan (2024). *On the nature of entrepreneurship*. Tech. rep. National Bureau of Economic Research.

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

Bonacich, Phillip (1987). "Power and centrality: A family of measures". In: *American journal of sociology* 92.5, pp. 1170–1182.

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

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

Bresnahan, Timothy F and Reiss, Peter C (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 Raider, HJ (2002). "Creating careers: women's paths to entrepreneurship". In: *Unpublished manuscript, University of Chicago*.

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

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

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

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

Chatterji, Aaron K, Chay, Kenneth Y, and Fairlie, Robert W (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, Friedman, JN, Hendren, N, Jones, MR, and Porter, SR (2019). *The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility*. Tech. rep. National Bureau of Economic Research Working Paper w25147.

Chetty, R and Hendren, N (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, Dossi, Gaia, Van Reenen, John, Smith, Matt, Zidar, Owen, and Zwick, Eric (2025). *America's Missing Entrepreneurs: The Determinants of and Returns to Entrepreneurship*. Tech. rep.

Chetty, Raj, Jackson, Matthew O., Kuchler, Theresa, Stroebel, Johannes, Hendren, Nathaniel, Fluegge, Robert, Gong, Sara, Gonzalez, Federico, Grondin, Armelle, Jacob, Matthew, Johnston, Drew, Koenen, Martin, Laguna-Muggenberg, Eduardo, Mudekereza, Florian, Rutter, Tom, Thor, Nicolaj, Townsend, Wilbur, Zhang, Ruby, Bailey, Mike, Barbera, Pablo, Bhole, Monica, and Wernerfelt, Nils (2022c). "Social Capital I: Measurement and Associations with Economic Mobility". In: *Nature* 608.7921, 108–121.

— (2022d). "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, Pekkala Kerr, Sari, Kerr, William R, and Maiden, Louis J (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, Nenov, Plamen T, Santos, Vitor, and Simsek, Alp (2024). *Stock Market Wealth and Entrepreneurship*. Tech. rep. National Bureau of Economic Research.

Choi, Jinkyong, Guzman, Jorge, and Small, Mario L (June 2024). *Third Places and Neighborhood Entrepreneurship: Evidence from Starbucks Cafés*. Working Paper 32604. National Bureau of Economic Research. DOI: 10.3386/w32604. URL: http://www.nber.org/papers/w32604.

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

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

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

Denes, Matthew R, Lagaras, Spyridon, and Tsoutsoura, Margarita (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 Puga, Diego (2023). "Urban growth and its aggregate implications". In: *Econometrica* 91.6, pp. 2219–2259.

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

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

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

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

Fairlie, Robert W, Karlan, Dean, and Zinman, Jonathan (2015a). "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., Kroff, Zachary, Miranda, Javier, and Zolas, Nikolas (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, Jayachandran, Seema, and Pande, Rohini (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, Jayachandran, Seema, Pande, Rohini, and Rigol, Natalia (2016). "Friendship at work: Can peer effects catalyze female entrepreneurship?" In: *American Economic Journal: Economic Policy* 8.2, pp. 125–153.

Fluegge, Robert B (2025). *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, Golub, Benjamin, and Goyal, Sanjeev (2020). "Targeting interventions in networks". In: *Econometrica* 88.6, pp. 2445–2471.

Glaeser, Edward L, Kerr, Sari Pekkala, and Kerr, William R (2015b). "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, Kerr, William R, and Ponzetto, Giacomo AM (2010). "Clusters of Entrepreneurship". In: *Journal of Urban Economics* 67.1, pp. 150–168.

Golosov, Mikhail, Graber, Michael, Mogstad, Magne, and Novgorodsky, David (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, Kovner, Anna, Lerner, Josh, and Scharfstein, David (2010). "Performance persistence in entrepreneurship". In: *Journal of financial economics* 96.1, pp. 18–32.

Gompers, Paul A and Lerner, Josh (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, Hornbeck, Richard, and Moretti, Enrico (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 Helpman, Elhanan (1992). *Innovation and growth: Technological competition in the global economy*.

Gu, Qian, Karoly, Lynn A, and Zissimopoulos, Julie M (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, Pistaferri, Luigi, and Schivardi, Fabiano (2021). "Learning entrepreneurship from other entrepreneurs?" In: *Journal of labor economics* 39.1, pp. 135–191.

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

Haltiwanger, John, Jarmin, Ron S, and Miranda, Javier (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 Nanda, Ramana (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 Aghion, Philippe (1998). "Capital accumulation and innovation as complementary factors in long-run growth". In: *Journal of Economic Growth* 3, pp. 111–130.

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

Hurst, Eric and Pugsley, Benjamin Wild (2011). *What do small businesses do?* Tech. rep. 2, pp. 73–142.

Iaria, Alessandro, Schwarz, Carlo, and Waldinger, Fabian (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 (1970). *The Economy of Cities*. New York: Random House.

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

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

Jones, Jason J., Settle, Jaime E., Bond, Robert M., Fariss, Christopher J., Marlow, Cameron, and Fowler, James H.e (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 Rob, Rafael (1989). "The growth and diffusion of knowledge". In: *The Review of Economic Studies* 56.4, pp. 569–582.

Kalyani, Aakash, Bloom, Nicholas, Carvalho, Marcela, Hassan, Tarek, Lerner, Josh, and Tahoun, Ahmed (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, Kerr, William R, and Nanda, Ramana (2022f). "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 Mandorff, Martin (2023). "Social networks, ethnicity, and entrepreneurship". In: *Journal of Human Resources* 58.1, pp. 183–220.

Kihlstrom, Richard E and Laffont, Jean-Jacques (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 Li, Huiyu (2025). "Entry costs rise with growth". In: *Journal of Political Economy Macroeconomics* 3.1, pp. 43–74.

Klette, Tor Jakob and Kortum, Samuel (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 Johnston, Drew (2024). *Social Ties and Residential Choice: Micro Evidence and Equilibrium Implications*. Mimeo.

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

Kwon, Spencer Y, Ma, Yueran, and Zimmermann, Kaspar (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 Malmendier, Ulrike (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.

Leskovec, Jure and Sosi, Rok (2016). "SNAP: A General-Purpose Network Analysis and Graph-Mining Library". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.1, p. 1.

Lin, Mingkai, Li, Wenzhong, Song, Lynda J, Nguyen, Cam-Tu, Wang, Xiaoliang, and Lu, Sanglu (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, Nathanson, Charles G, and Weyl, E Glen (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 Whinston, Michael D (1986). "Free entry and social inefficiency". In: *The RAND Journal of Economics*, pp. 48–58.

Marmaros, David and Sacerdote, Bruce (2006). "How do friendships form?" In: *The Quarterly Journal of Economics* 121.1, pp. 79–119.

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

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

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

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

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

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

Newcomb, Theodore M. (1961). *The Acquaintance Process*. New York: Holt, Rinehart and Winston.

Oosterbeek, Hessel, Van Praag, Mirjam, and Ijsselstein, Auke (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 (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, Yagan, Danny, Zidar, Owen, and Zwick, Eric (2019). "Capitalists in the Twenty-first Century". In: *The Quarterly Journal of Economics* 134.4, pp. 1675–1745.

Sorenson, Olav and Audia, Pino G (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 Sorenson, Olav (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 Abraham, Sarah (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.

— (2010). *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 (2025). "Entrepreneurial spillovers across coworkers". In: *Journal of Financial Economics* 170, p. 104077.

Table 1: Sample Statistics

	Sample Statistic	All Individuals (1)	Reports Work Experience (2)	All Owners (3)	Founders Since 2015 (4)	Student Sample (5)	Post-Grad Owners (6)
Network Characteristics							
Friends	Mean	545.0	648.7	888.3	937.3	1213.4	1512.2
	Median	376	495	727	778	1135	1447
	Std. Dev	527.4	553.9	630.1	647.3	508.2	577.4
Entrepreneur Friends	Mean	4.7	6.2	21.4	20.9	6.7	21.8
	Median	2	3	14	14	5	16
	Std. Dev	8.2	9.6	23.5	22.9	7.1	20.6
Firm Characteristics							
Years Since Founding	Mean			14.1	6.9		7.0
	Median			10	7		7
	Std. Dev			12.9	2.1		2.5
Share >$1m Annual Revenue				0.18	0.082		0.14
Share with 6-digit NAICS				0.68	0.72		0.56
Share with 2-digit NAICS				0.94	0.91		0.90
Demographics							
Share Attended College		0.29	0.38	0.47	0.46	1	1
Share Female		0.53	0.54	0.46	0.47	0.64	0.55
Birth Cohort	Mean	1984.6	1984.1	1979.6	1981.7	1993.3	1991.9
	Median	1985	1985	1980	1982	1993	1992
	Std. Dev	9.8	9.5	8.1	8.0	2.3	1.9
Zip Code Median Income	Mean	72.6	73.4	80.0	77.7	89.3	84.5
($1000s)	Median	65.8	67.0	73.8	71.7	83.2	79.3
	Std. Dev	30.9	30.6	32.0	30.8	37.9	35.1
Sample Size		128,003,829	71,676,839	627,747	321,544	736,308	1,907

Notes: This table presents summary statistics for the samples used in the analysis. Column (1) includes all individuals in the primary analysis sample, which restricts to users born between 1966 and 2000 with between 30 and 3000 friends. Column (2) additionally restricts to individuals who have self-reported at least one job on their personal profile. Columns (3) and (4) restrict to all business owners and business owners with a business created since 2015, respectively. Column (5) presents statistics on individuals considered in the College design, which includes attendees with at least one job on their profile who turned 18 between 2006 and 2016 and who made at least 5 friends at the same school during the fall of the year the user turned 18. Column (6) 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	4.74	0.01187	129,336,773
Pooled	Male	2.52	0.01483	129,336,773
Pooled	Female	2.17	0.00967	129,336,773
Male	Pooled	4.34	0.01096	59,208,821
Male	Male	2.69	0.01361	59,208,821
Male	Female	1.61	0.00819	59,208,821
Female	Pooled	5.05	0.01099	67,924,458
Female	Male	2.42	0.01414	67,924,458
Female	Female	2.59	0.00931	67,924,458

Notes: This table presents the average number of owner friends and average fraction of owner friends by an individual's own gender and friend gender. The first two columns denote the focal individual's gender and the friend gender considered in that row. The average number of owner friends refers to the number of friends who own a firm and are of the gender in the column denoted "Friend Gender" among focal individuals of the gender in the column "Own Gender". The average fraction of owner friends refers to the fraction of friends of the gender "Friend Gender" who own a firm. The final column provides the total number of focal individuals considered in the row. The sample considers all individuals born between 1966 and 2000 with between 30 and 3000 friends. Summary statistics for the sample are presented in Column 1 of Table 1.

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

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Sole Owner	(4) Fraction Owners	(5) Fraction Owners	(6) Fraction Owners
Owner Friends of Friends (FoFs)	4.67×10^{-5} (2.15×10^{-5})	6.59×10^{-5} (2.66×10^{-5})	3.80×10^{-5} (1.76×10^{-5})			
Owner FoFs in Same 6-Digit Industry				7.41×10^{-6} (9.28×10^{-7})		
Owner FoFs in Same 4-Digit Industry but Different 6-Digit Industry					5.16×10^{-7} (4.42×10^{-7})	
Owner FoFs in Same 2-Digit Industry but Different 4-Digit Industry						2.10×10^{-7} (1.03×10^{-7})
Control for Pre-College Owner FoFs	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
6-Digit Industry Fixed Effects				X	X	X
Unit of Observation	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
Number of Cells	736,973	736,973	736,973	18,822,557	19,469,506	28,899,027
Mean of Dep. Var	0.0026	0.0026	0.0018	1.15×10^{-6}	1.15×10^{-6}	1.15×10^{-6}
SD of Dep. Var	0.051	0.051	0.042	1.89×10^{-4}	1.77×10^{-4}	2.34×10^{-4}
Mean of Ind. Var	3.12	3.12	3.12	0.0017	0.0062	0.12
SD of Ind. Var	3.12	3.12	3.12	0.047	0.092	0.44
p90 - p10 difference in Owner FoFs in year student turns 18	41	41	41			

Notes: This table presents estimates of the effect of connections to firm owners among college students who turn 18 between 2006 and 2016 using the College design. The outcome is an indicator for 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 connections (i.e., 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 typical semester system. Column (1) presents the baseline estimates corresponding to Equation 1. Column (2) adds a control for the number of friendships formed by the student in the treatment period during the first 12 weeks of college. Column (3) redefines the outcome to be whether a student starts a firm with no co-founders. Column (4) estimates a regression 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. Columns (5) and (6) replace the treatment from Column (4) with the number of new connections in a different industry but the same industry group (4-digit NAICS) and in a different industry group but the same sector (2-digit NAICS), respectively. Point estimates from the binned approach are numerically identical to the same regressions run at the individual level. Bins with greater than 10 owner friends-of-friends are dropped in the industry-specific models. Owners constituting the treatment in all columns must have already started a firm and be connected to the intermediate individual by the year that the focal student turns 18, such that variation in the treatment is generated only by friending behavior in the first 12 weeks of college. The College design is described in Section III.A. Standard errors clustered at the birth cohort by college level are reported in parentheses.

Table 4: College Design: Gender-Specific Exposure Effects

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner	(4) Is Owner	(5) Is Owner
Male Owner Friends of Friends		1.21×10^{-5} (4.27×10^{-5})		1.64×10^{-4} (6.54×10^{-5})	-1.22×10^{-5} (4.31×10^{-5})
Female Owner Friends of Friends	9.63×10^{-5} (4.34×10^{-5})		1.70×10^{-4} (6.64×10^{-5})		1.24×10^{-4} (4.40×10^{-5})
Male Student x Male Owner Friends of Friends					1.16×10^{-4} (8.11×10^{-5})
Male Student x Female Owner Friends of Friends					-1.29×10^{-5} (8.16×10^{-5})
Is Male					1.04×10^{-3} (1.89×10^{-4})
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	467,741	467,741	262,577	262,577	730,318
Mean of Dependent Variable	0.0021	0.0021	0.0031	0.0031	0.0025
SD of Dependent Variable	0.046	0.046	0.056	0.056	0.050
Mean of Independent Variable	1.52	1.48	1.36	1.56	
SD of Independent Variable	1.80	1.72	1.69	1.80	

Notes: This table presents gender-specific estimates of the effect of connections to firm owners among college students who turn 18 between 2006 and 2016 using the College design. The outcome is an indicator for firm ownership starting four or more calendar years after the year that the student turns 18. Column (1) estimates a version of Equation 1 among female students, where the treatment is the number of connections to female firm owners. Column (2) repeats Column (1) but replaces the independent variable with the number of male firm owners. Columns (3) and (4) repeat Columns (1) and (2) among male focal students. Column (5) estimates a multivariate regression where firm founding is regressed on exposure to male and female owners, and the interactions with an indicator for whether the focal student is male. Gender is self-reported by users on their public Facebook profile. Owners constituting the treatment in all columns must have already started a firm and be connected to the intermediate individual by the year that the focal student turns 18, such that variation in the treatment is generated only by friending behavior in the first 12 weeks of college. The College design is described in Section III.A. Standard errors clustered at the birth cohort by college level are reported in parentheses.

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
Destination Entrepreneurship by Years After First Move	0.007 (0.0011)	0.0061 (0.0010)	0.0062 (9.85×10^{-4})			
Destination Entrepreneurship In Same Sector by Years After First Move				0.0014 (3.70×10^{-4})		0.0013 (3.69×10^{-4})
Destination Entrepreneurship In Other Sectors by Years After First Move					2.28×10^{-4} (4.63×10^{-5})	2.04×10^{-4} (4.67×10^{-5})
Control for Destination Entrepreneurship Rate		X				
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	Destination CZ by Nonmover CZ by Relative Year	Destination CZ by Nonmover CZ by Relative Year	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	32,128,164	32,128,164	32,128,164
Mean of Dependent Variable	4.71×10^{-5}	4.71×10^{-5}	4.71×10^{-5}	1.81×10^{-6}	1.81×10^{-6}	1.81×10^{-6}
SD of Dependent Variable	0.0047	0.0047	0.0047	9.57×10^{-4}	9.57×10^{-4}	9.57×10^{-4}
Mean of Jobs Variable	0.016	0.016	0.016			
SD of Jobs Variable	0.0047	0.0047	0.0047			
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					0.0046	0.0046

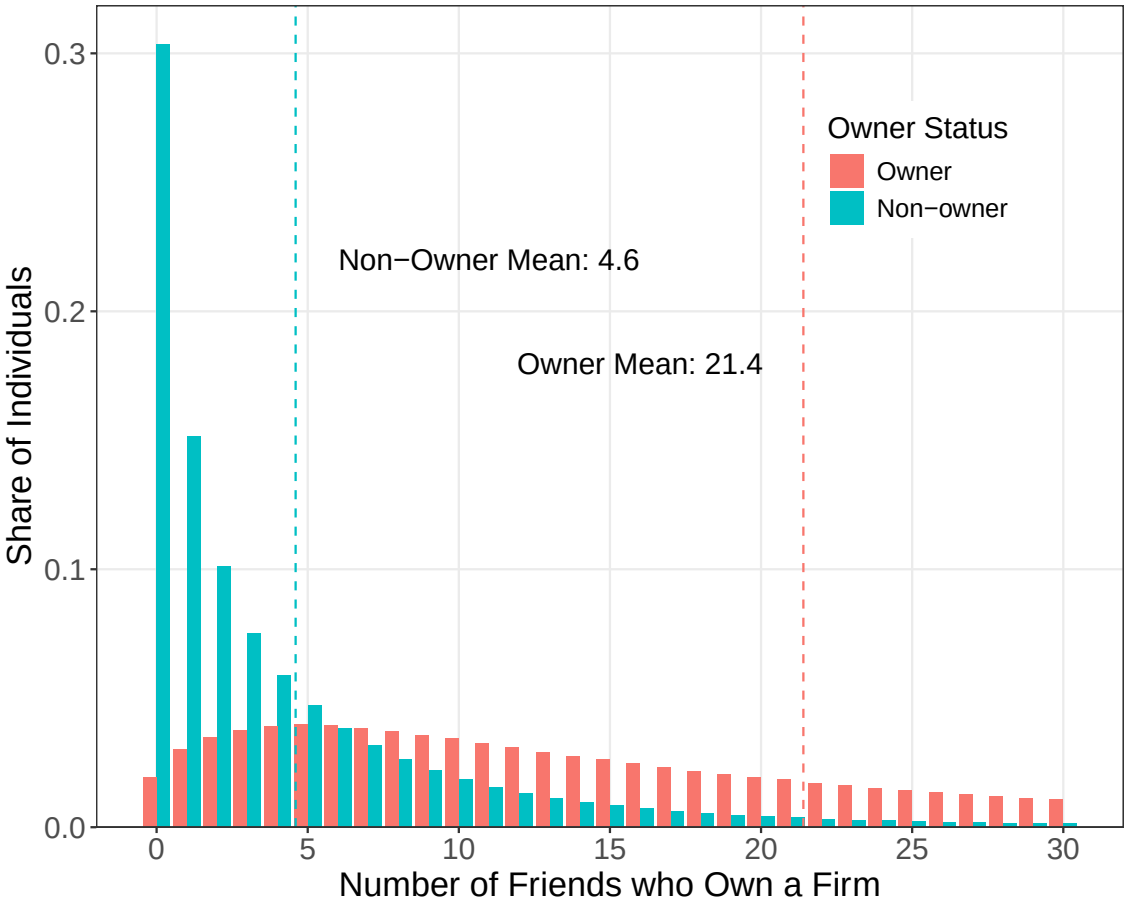
Notes: This table estimates interactions between the destination entrepreneurship rate and the post-treatment trend in the Movers Design using a stacked difference-in-differences design (Sun and Abraham 2021). Column (3) estimates $\beta_{interaction}$ from the baseline specification described in Equation 5. Columns (1) and (2) build up to the baseline specification by estimating Equation 5 without destination CZ-by-nonmover-CZ fixed effects. The outcome in Columns (1)-(3) is an indicator for whether the focal individual has started any business in or before the given year. The treatment is the interaction of years since the first year in which the focal individual has a friend move to a given commuting zone and the rate of successful entrepreneurship in that commuting zone, defined as the number of new jobs per capita created at firms less than one year old. Column (4) presents estimates of $\beta_{interaction}$ from the CZ-by-sector specification of the Movers design presented in Equation 6, where sector is defined using 2-digit NAICS codes. The outcome in Column (4) is defined as an indicator for whether the focal individual has started a firm in a given sector in or before the given year. The treatment is the interaction of years since the first year in which the focal individual has a friend move to a given commuting zone and the sector-specific rate of entrepreneurship in the commuting zone. Column (5) repeats the model in Column (4) but replaces entrepreneurship rate in the interaction term with the rate of entrepreneurship in other sectors. Column (6) presents estimates from a specification which includes both interaction terms simultaneously. The Movers design is described in greater detail in Section III.B.

Table 6: Commuting Zone Targeting under Estimated Baseline Model

	(1)	(2)	(3)	(4)
	Commuting Zones with Below Median Entrepreneurship	Commuting Zones in Top 50% of Direct Effect	Commuting Zones in Top 50% of Total Effect	Ratio of (3)/(2)
Firms Created among Treated Population	1	1.31	1.16	0.89
Firms Created among Individuals Not Directly Treated	0.78	0.8	1.19	1.47
Total Firms Created	1.78	2.11	2.35	1.11

Notes: This table provides counterfactual estimates of the effect of a hypothetical policy on firm entry for different populations under the baseline model, estimated using causal estimates from the Movers design. The effect of the hypothetical policy is normalized such that the policy would create one new firm among directly treated individuals if targeted to individuals who live in commuting zones where the average number of new startup jobs per capita is above the population-weighted median. The first row presents the number of firms created by individuals directly treated by the policy. The second row presents the number of firms created by individuals not directly treated by the policy who are induced to entry through the spillover effects. The third row presents the total number of firms that enter, equal to the sum of the first and second rows. Column (1) considers a policy targeted to individuals living in commuting zones with below median entrepreneurship, such that the first row of Column (1) is equal to unity under our normalization. Column (2) considers a policy targeted to individuals living in commuting zones where the direct effect of the policy is above the population-weighted median. Column (3) considers a policy targeted to individuals living in commuting zones where the total effect of the policy, including spillovers, is above the population-weighted median. Column (4) calculates the ratio of Column (3) to Column (2), which reflects the gains from targeting on total effects relative to targeting on direct effects alone. The model and estimation approach are described in Section IV. The results are discussed in Section V.B.

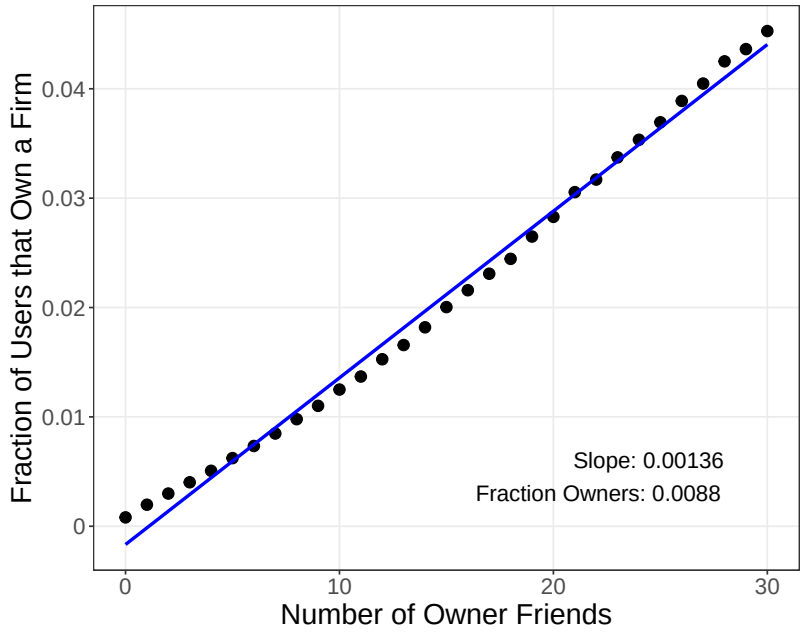
Figure 1: Number of Friends who Own a Business by Owner Status



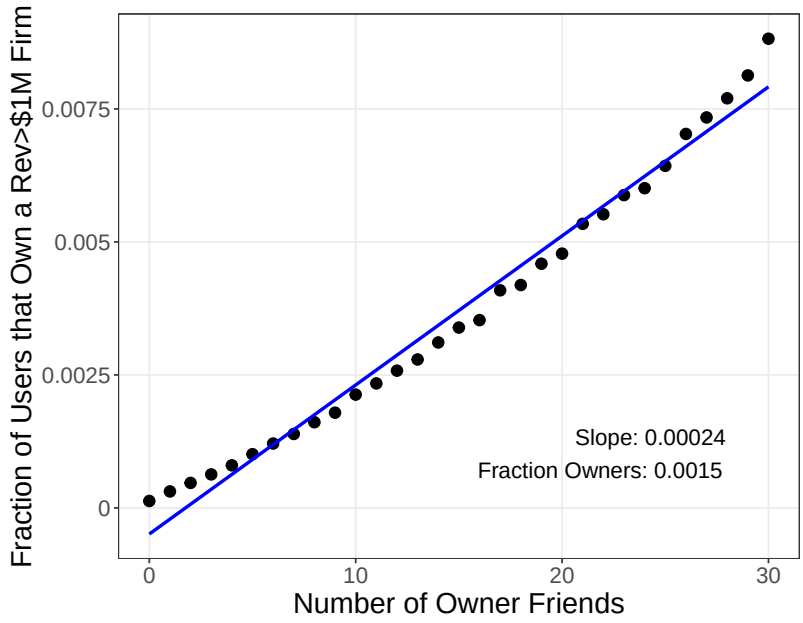
Notes: This figure presents the distribution across individuals in the number friends who own a business, separately for individuals identified as business owners and non-owners. Firms are identified as local business pages on the Facebook platform which can be linked to 3rd party establishment data. Business ownership among individuals and their friends is identified using self-reported employer and job title on public Facebook profiles. We restrict to all individuals born between 1966 and 2000 with between 30 and 3000 friends on Facebook. Summary statistics for the sample are presented in Table 1.

Figure 2: Owner Status by Number of Owner Friends

(a) Fraction Owner by Number of Owner Friends

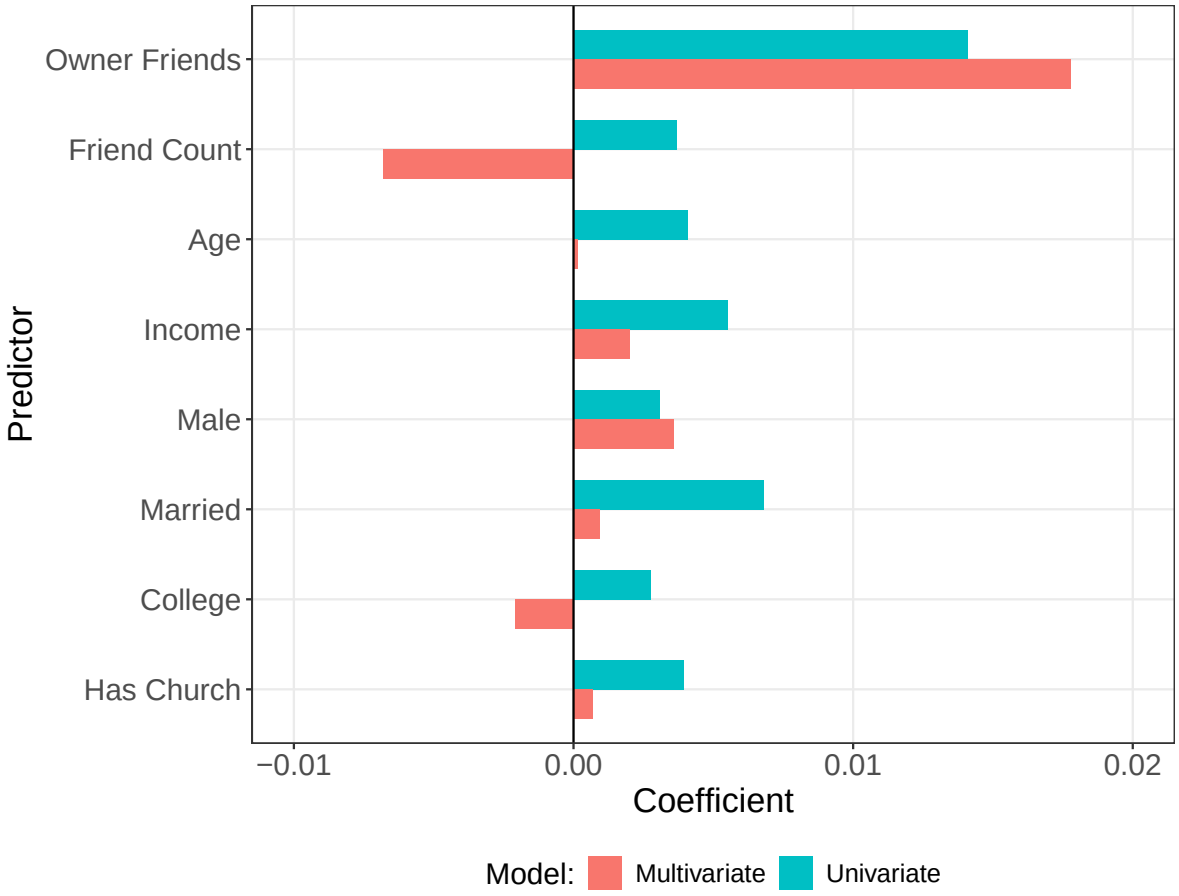


(b) Fraction Owner with Annual Rev. >$1 Million by Number of Owner Friends



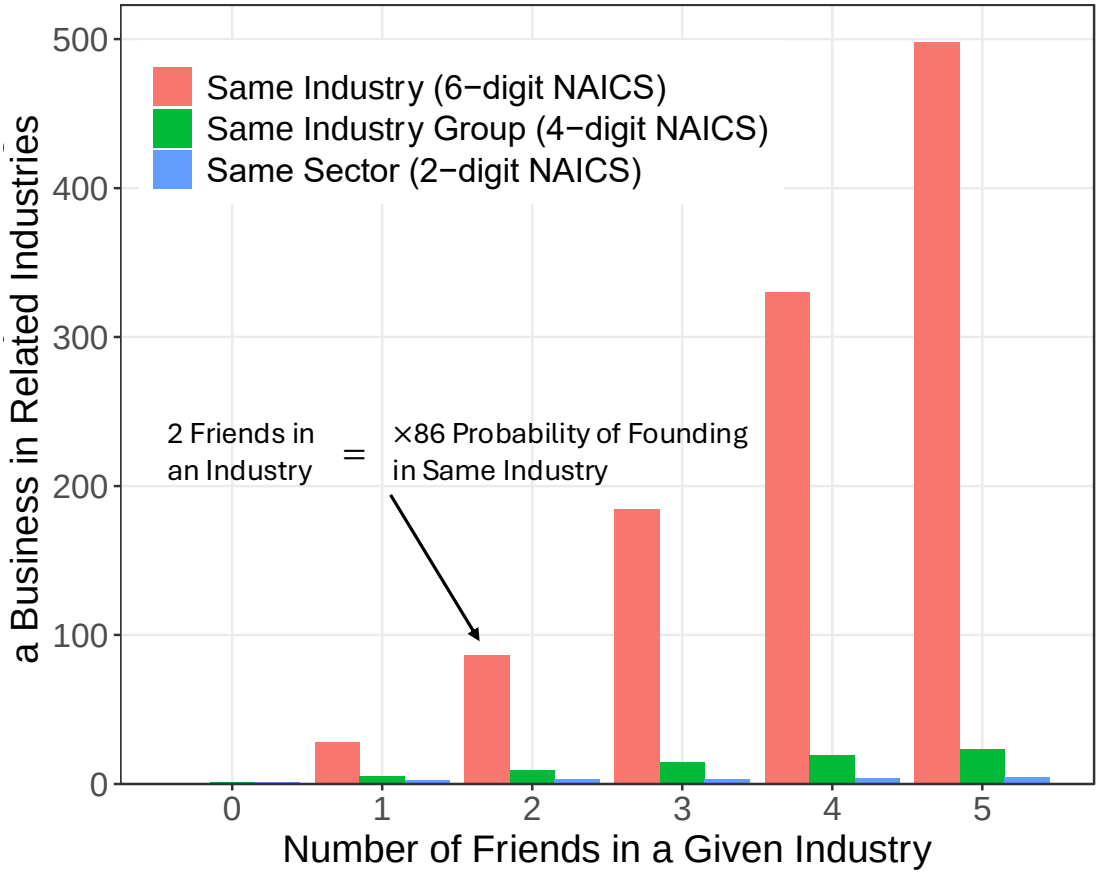
Notes: This figure shows that the number of an individual's friends who own a business predicts whether the individual owns a business among individuals with more than 100 friends. Panel A shows that rates of total business formation increase with the number of friends who own a business. Panel B shows 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 E8 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 3: Predictors of Firm Ownership



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. "Univariate" models correspond to a set of separate univariate regressions for each covariate. The "Multivariate" model presents coefficients from a regression on all covariates simultaneously. Continuous covariates, including the number of owner friends, the number of total friends, age, and income, are standardized across individuals. Dichotomous predictors for gender, college attendance, and relationship status are self-reported. The dichotomous predictor 'Has Church' is equal to one if the individual is a member of at least one Facebook group that corresponds to a church or likes a page corresponding to a church.

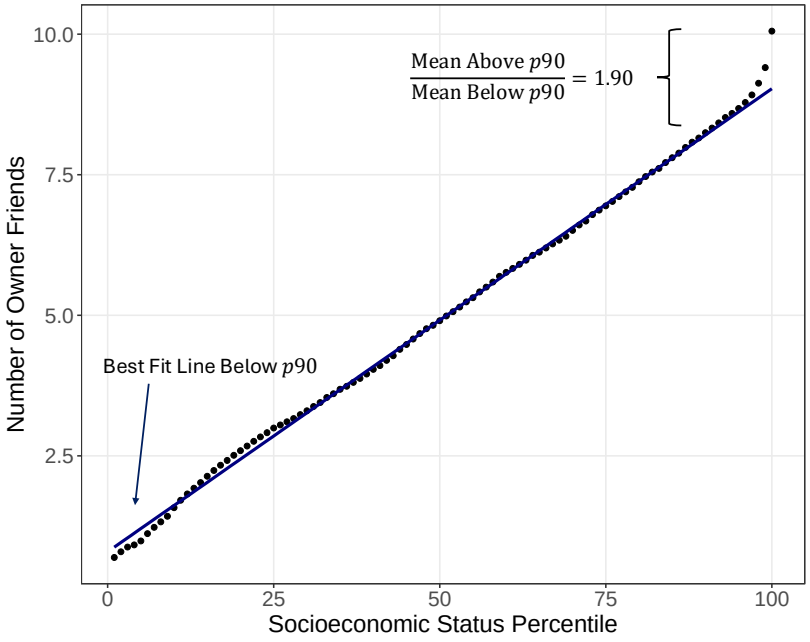
Figure 4: Industry Choice by Friend Industry



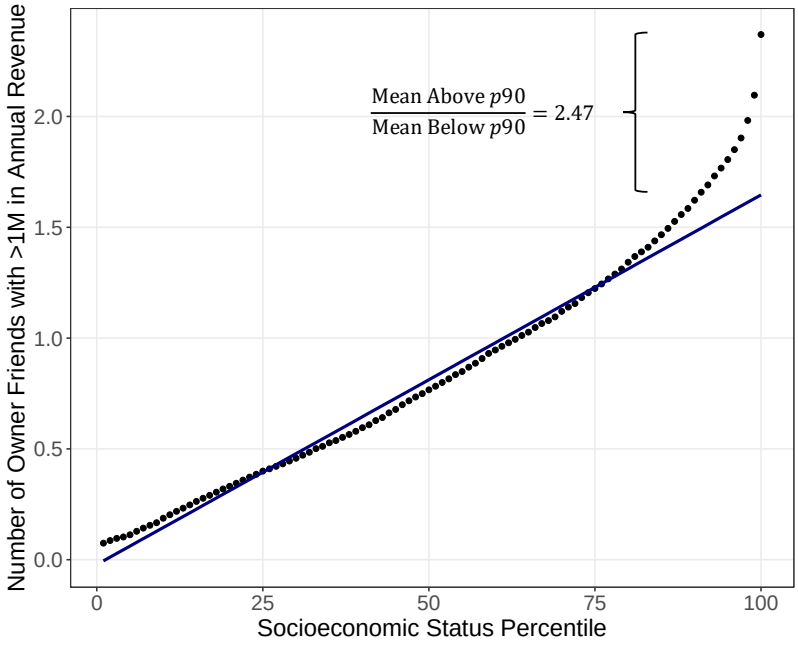
Notes: This figure plots the relative probability of starting a business in a given industry class compared to the baseline probability of starting in that class, separately by the number of friends in the same and related industry classes and estimated among all owners. The relative probability for 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 5: Average Number of Owner Friends by Socioeconomic Status

(a) Average Number of Owner Friends by Percentile of Income (Present)

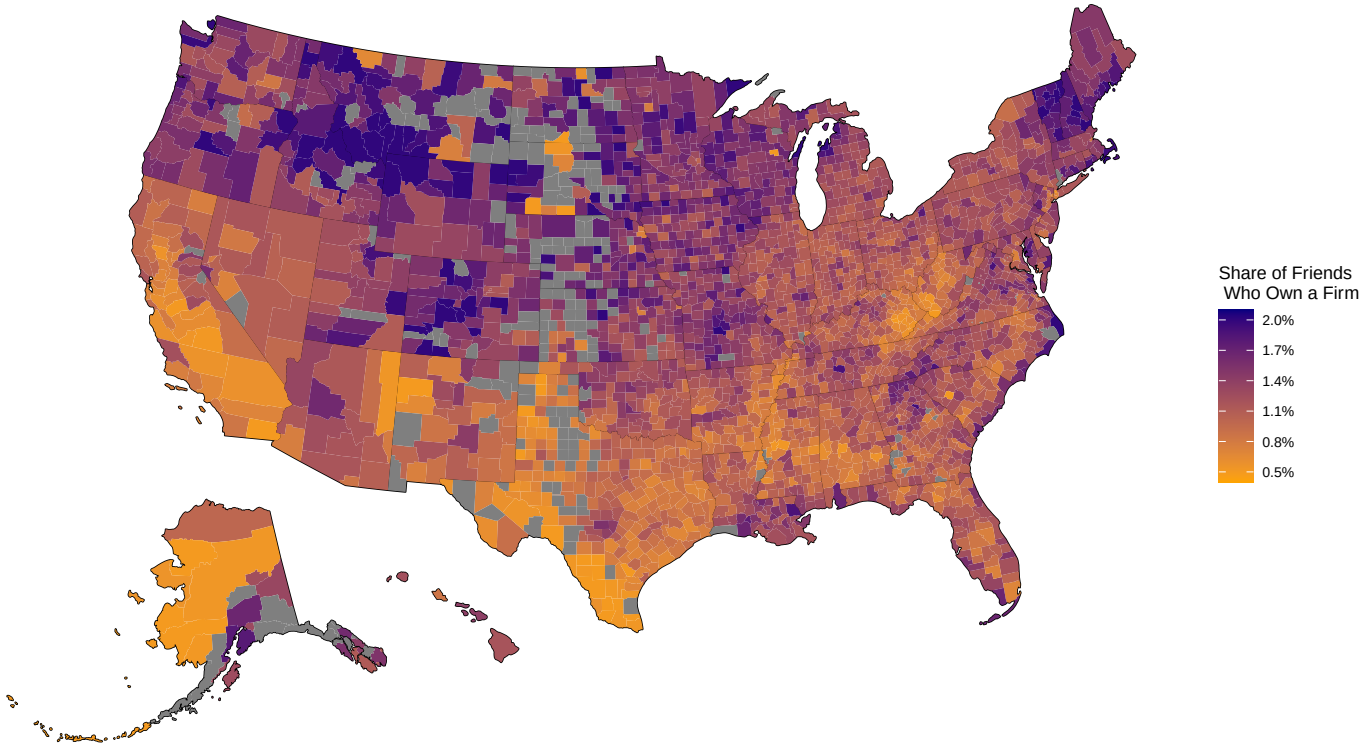


(b) Average Number of Owner Friends with >$1M in Annual Revenue



Notes: This figure plots the average number of friends who have started a firm by the income percentile of the individual. Plotted fit lines are fit only on observations below p90 to highlight the convexity in connections to owners at the top of the income distribution. Panel A presents the mean number of friends who own any firm in our baseline specification, and Panel B plots the mean number of friends who own a firm with annual revenue greater than $1 million. The text reports the ratio of mean number of owner friends for individuals above p90 in the income distribution divided by the mean number of owner friends for individuals below p90. Summary statistics for the sample are presented in Table 1. Income measurement is described in Appendix Section A.1.

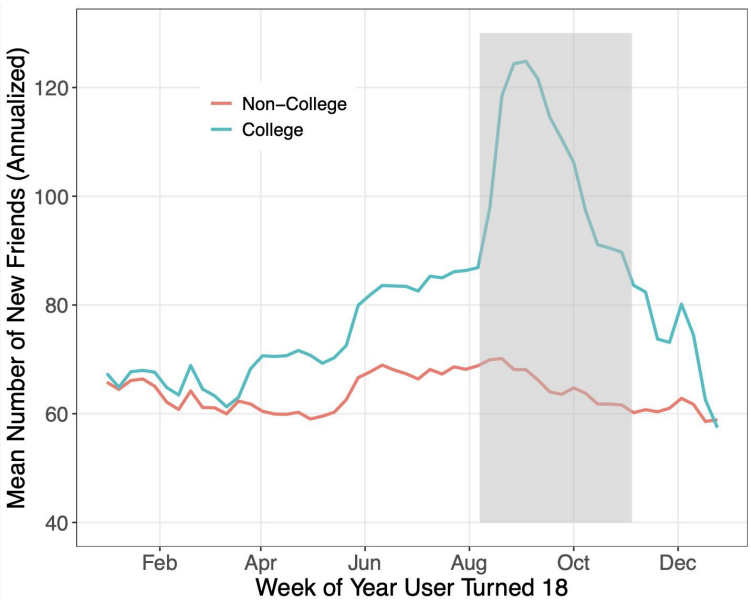
Figure 6: Average Share of Friends Who Own a Firm



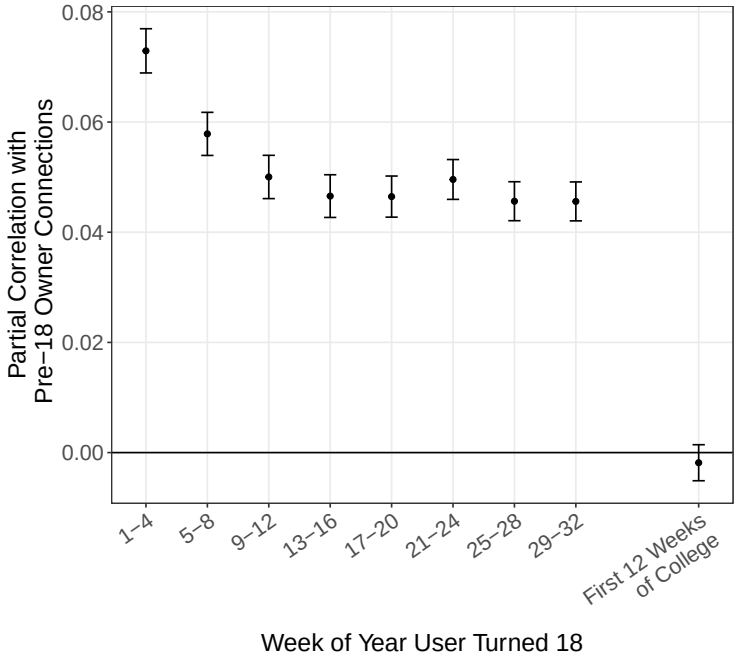
Notes: This figure plots the share of friends that own a business for the average person in each county in the U.S., restricting to counties with more than 1000 users. The share of owners among friends is calculated among users who were born between 1966 and 2000, have between 30 and 3000 friends, and report at least one job title, and friends are restricted to those who report at least one job title. An individual's county is measured using the modal zip code corresponding to their ZIP address in the first 3 weeks of December 2022. Summary statistics for the sample are provided in Table 1.

Figure 7: Patterns of Friendship Formation Among College-goers

(a) Friendship Formation Spikes at the Start of College



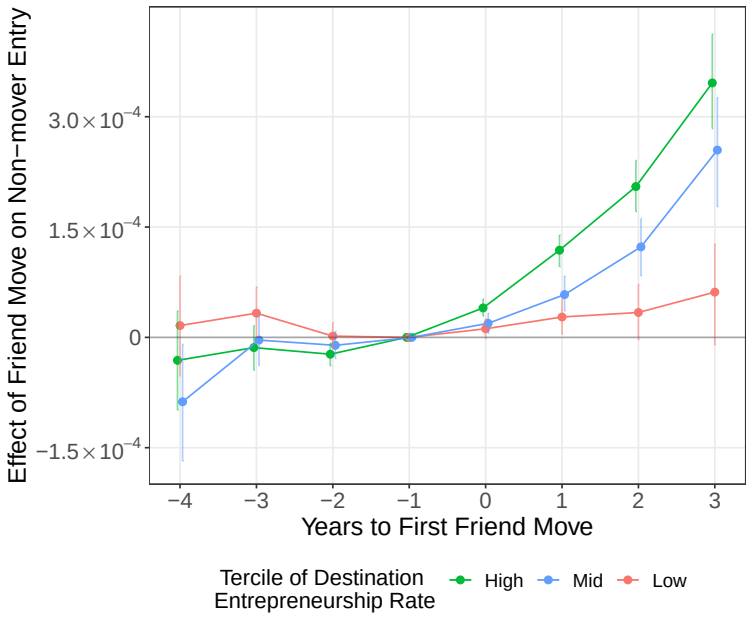
(b) College Network Formation and Existing Network Composition



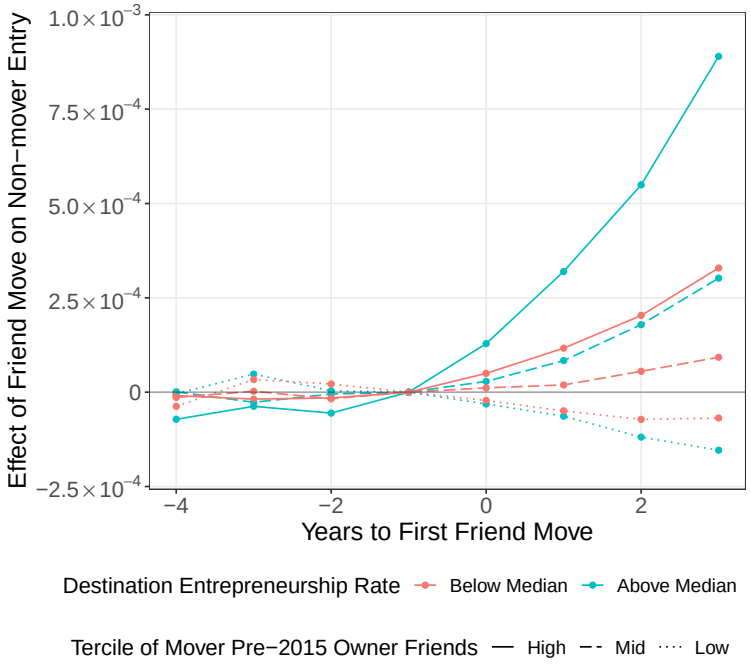
Notes: Panel A presents the average number of new friends made in the week an individual turned 18, separately for users that went to college and did not go to college. College attendance is measured using self-reported college on an individual's Facebook profile. The shaded area corresponds to weeks 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. Panel B 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, and represents only 2nd-degree connections where the connection to an owner was 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. Summary statistics on the sample of college attendees is presented in Table 1.

Figure 8: Effects of Friend Migration on Non-movers by Destination and Mover Characteristics

(a) Effects by Destination Entrepreneurship Rate



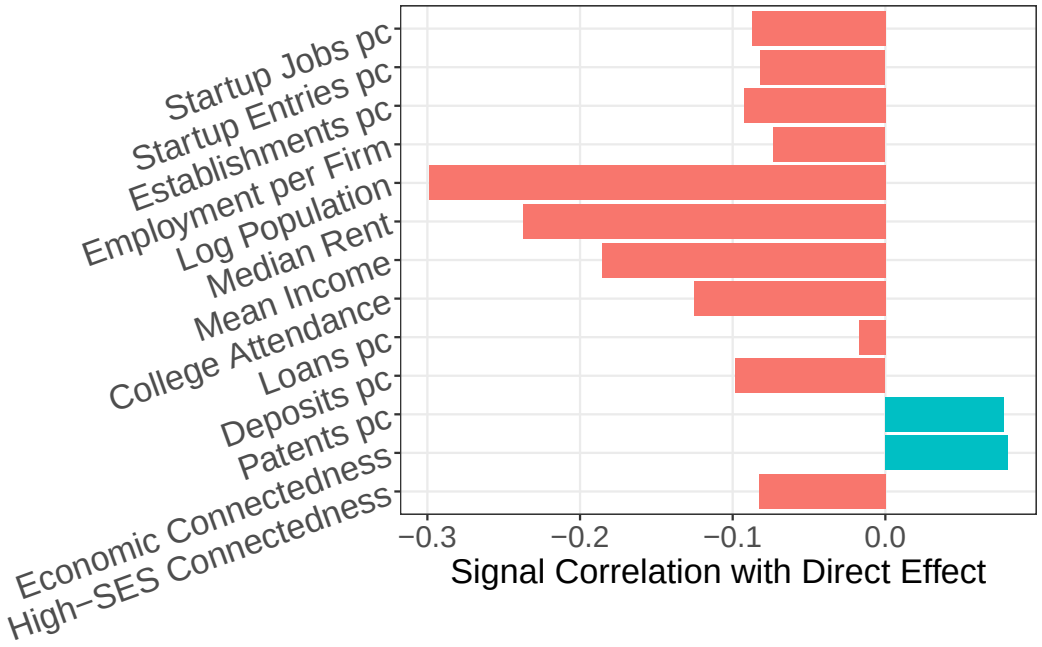
(b) Effects by Destination Entrepreneurship Rate and Mover's Pre-Move Network



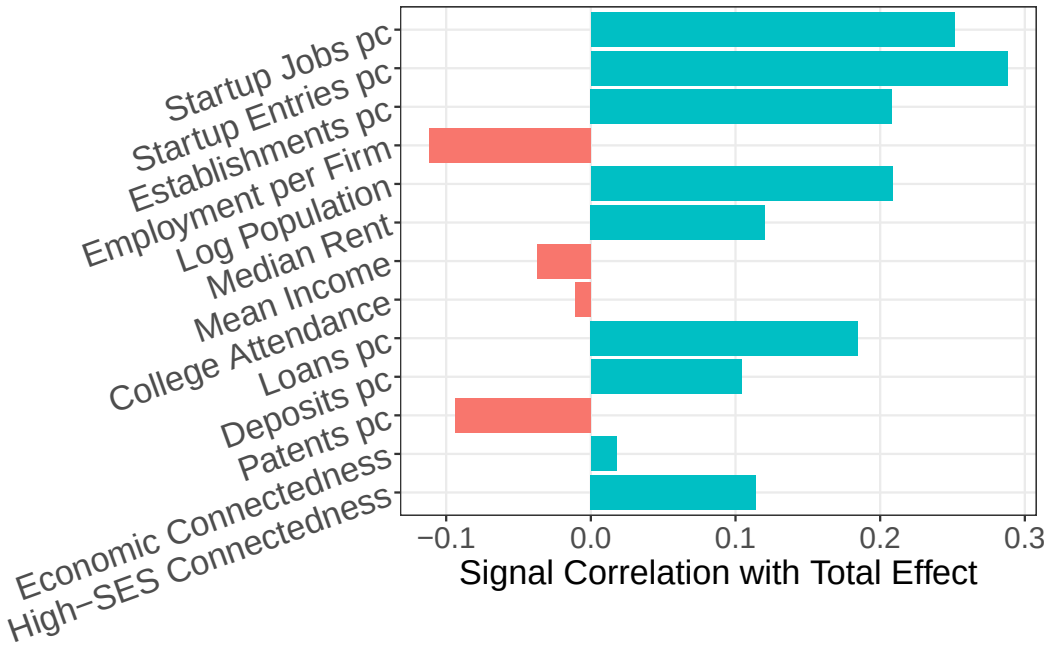
Notes: This figure plots difference in differences estimates from the Movers quasi-experimental design. The outcome is the rate of firm founding among non-movers who are friends with a mover. Panel A plots average effects binned by tercile of the entrepreneurship rate in the mover's destination commuting zone, averaged across all non-mover commuting zones. Panel B plots average effects binned by terciles of 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 lowest tercile of pre-2015 owner friends contains only movers who had zero friendships with firm owners prior to the move. Moves between 2016 and 2019, inclusive, are used as treatment. Control groups include movers between 2017 and 2020. Commuting zone-level 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 in Panel A are calculated via Bayesian bootstrap with 1000 iterations. Details of the estimation strategy are presented in Section III.B. Estimates of the post-move linear trend in the treatment effect interacted with destination entrepreneurship rates are presented in Table 5.

Figure 9: Predictors of Combined Networking Productivity and Spillover Production

(a) Signal Correlation of Direct Effect with Commuting Zone Characteristics



(b) Signal Correlation of Total Effect with Commuting Zone Characteristics



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.

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.

We note that linking the establishment data to Pages appears to cull the establishment data of many entities that are not economically meaningful. Random samples from the Customer Graph data indicate that a substantial proportion of firms in those data which cannot be matched to a Facebook Page are registered to a residential address and have no online presence.

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 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 E19 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 E19 displays the averages disaggregated by year of move.

A.2.2 Parent-Child Linkages

Linkages between parents and children follow the approach in 2022c with technical updates to improve match rates. The approach starts with self-reported linkages between parents and children, then iteratively supplements these linkages using public user-generated wall posts, matching based on age and hashed last names (hashed to protect confidentiality), and parents measured for by self-reported siblings.

A.2.3 Socioeconomic Status

Measurement of individual and parent socioeconomic status follows the approach and data in (M. Bailey 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.

A.2.4 Close Friends

In Appendix Figure E11 we present estimates of homophily among owners using the top 10 closest friends. Close friends are measured on the basis of the frequency of public interactions such as likes, tags, wall posts and comments. First, a model is trained at the user-by-friend level based on internally administered surveys of users asking whether a user is likely to interact with a given friend in-person. Using this model friends are ranked for each person based on the closeness output by the model. The measurement of close friends in this paper is identical to the approach taken in 2022c.

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.

The linear functional form is 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 this result 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

C.1 Baseline Model

The Movers design estimates a set of CZ-pair-level 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 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.1.1 Katz Centrality Estimator and Validation

Our estimate for Katz-Bonacic Centrality uses 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 resulting estimator is unbiased for the truncated series representation of Katz-Bonacic centrality. Since terms of the series representation are always positive, our estimate is biased towards zero for true Katz-Bonacic centrality. This bias can be seen in the resulting validation estimates in Appendix Figure E23.

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 $d=8$, $r=0.01$, and γ is estimated as described in Section IV.C.2. For consistency with the sample used in the Movers design, we consider only connections that formed prior to 2015. We omit nodes with greater than 2400 connections in 2015, equal to roughly the top 0.5% of the degree distribution, due to sensitivity of the estimator to outlier nodes with extremely high degree. This sample restriction is less strict than maximum degree of 3000 set in the primary analysis sample in the sense that the 3000-degree maximum drops roughly the top 1.5% of the degree distribution.

For the validation exercise depicted in Figure E23, we use a publicly available dataset of Facebook connections which contains 4031 nodes (Leskovec and Sosi 2016). We set $d = 8$ and $r = 0.01$ as in the analysis sample. To set γ in the validation exercise, 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, our estimator allows us to directly observe each term of the estimate's series representation. We can inspect these terms to infer the values of γ for which the series begins to diverge. In our simulations, divergence occurs around $\gamma = 0.0071$.

For the validation exercise, we follow the network science literature in holding constant the ratio of "true γ" to the maximum value of γ across both the Facebook data and the validation data. Since our estimate of gamma in the Facebook data is 0.0061 and the inferred maximum is 0.0071, we use $\gamma = \frac{0.0061}{0.0071}\gamma^{max} \approx 0.86\gamma^{max}$ for γ^{max} the maximum possible value of gamma in the validation sample (which can be calculated analytically).

We briefly note that our approach of inferring the maximum attenuation factor by inspection of terms yields a much lower maximum than if we had used well-known asymptotic results. In a wide class of random graphs that result in a 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). Under our sample restrictions the maximum degree in the subgraph used to calculate centrality is 2400, so the asymptotic estimator would imply a maximum value of $\gamma \leq \frac{1}{\sqrt{2400}} \approx 0.020$, much higher than the maximum of 0.0071 we find by inspection. This disagreement between an asymptotic estimator and our inspection of terms suggests that caution should be used when applying asymptotic estimators of maximum eigenvalues to very large networks, as real-world networks may not satisfy the assumptions under which such asymptotic results are derived.

C.1.2 Estimating Network Effect Decay

Calculation of Katz-Bonacic centrality requires knowing the network decay of spillover effects, which we denoted γ. Recall that, under the model, Networking Productivity is equal to:

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

such that the average of Networking Productity across commuting zones can be written as $\mathbb{E}\left[\overline{\alpha}\gamma\beta_n x\phi_c | c \in \overline{\mathcal{C}}\right]$.

Now consider a similar calculation, where the outcome is instead the nonmover, same-CZ friends of a given nonmover exposed to a new CZ via a friend move. In that case, the outcome is measured among individuals that are two degrees removed from the mover. That is, the only difference in the treatment effect between the original focal individual and their nonmover, same-CZ friends is that the treatment is one network step farther for the friends. Since the effect decay by γ at each step, this implies that the average effect among nonmover, same-CZ friends is:

$$\mathbb{E}\left[NPfof(x)\right] = \gamma\mathbb{E}\left[\overline{\alpha}\gamma\beta_n x\phi_c | c \in \overline{\mathcal{C}}\right] \tag{43}$$

Comparing $\mathbb{E}[NP(x)]$ and $\mathbb{E}[NPfof(x)]$, we see that the only difference is a factor of γ. Thus, the ratio of the two estimates γ.

C.1.3 Estimating the Human Capital Scaling Factor

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.

To form an estimate, 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:

$$Y_{i,t} = \beta_{c(i,t)}\alpha_i + \beta_{c(i,t)} \sum_{j,n} \gamma^n [G_t^n]_{i,j}\alpha_j \times 1\{c(i,t) = c(j,t)\} \tag{44}$$
$$+ \beta_{c(i,t)} \sum_{j,n} \gamma^n [G_t^n]_{i,j}\alpha_j \times 1\{c(i,t) \neq c(j,t)\}$$

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:

$$Y_{i,t} = \beta_{c(i,t)}\alpha_i + (t - t_{i,0})\beta_{c(i,t)}a_{c(i,t)}U_i \cdot \frac{d}{dt}K_{i,t,local} + (t - t_{i,0})\beta_{c(i,t)}\overline{a}U_i \cdot \frac{d}{dt}K_{i,t,nonlocal} \tag{45}$$

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:

$$(t - t_{i,0})\beta_{c(i,t)}a_{c(i,t)}U_i/(t - t_{i,0})\beta_{c(i,t)}\overline{a}U_i = a_{c(i,t)}/\overline{a} \tag{46}$$

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

$$Y_{i,t} = \nu_t + \beta_{local}F_{local} + \beta_{nonlocal}F_{nonlocal} + \epsilon_{i,t} \tag{47}$$

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.2 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

$$Y_{it} = \beta_{\theta_i,c(i,t)}E_{it} + \epsilon_{it} \tag{48}$$

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:

$$\tau_{it\theta} = \delta\beta_{\theta_i,c(i,t)} + K_{it\theta}B_{it\theta} \tag{49}$$

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.

In order to model the Movers design in this environment, we start by rewriting the network growth equations such that:

$$\frac{d}{dt}(K_{it\theta}A_{it\theta}) = a_{\theta,c(i,t)}U_{i\theta} \cdot \frac{d}{dt}K_{it\theta} \tag{50}$$

$$\frac{d}{dt}K_{it\theta} = \sum_{s\in\Theta} R_{its}\phi_{s\theta,c(i,t)} \tag{51}$$

$$R_{it\theta} = \psi_{\theta,c(i,t)}V_{i\theta} \tag{52}$$

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:

$$K_{it\theta} = (t - t_{i,0}) \sum_{s,\theta} \phi_{s\theta,c(i,t)}\psi_{s,c(i,t)}V_{is} \tag{53}$$

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

$$\sum_{\eta\in\Theta} \gamma a_{\eta d}\beta_{n\theta} \sum_{s\in\Theta} \phi_{s\eta d}\psi_{sd}\mathbb{E}[U_{j\eta}V_{j\eta}|c(i,t) = n, c(j,t) = d] \tag{54}$$

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

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

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, divide out the average level of human capital as estimated in the baseline estimates. 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
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
Economic Connectedness	Commuting Zone	2022	Share of high-SES friends among low-SES individuals, capturing the extent of cross-class social ties. *Original Source:* 2022c
High-SES Economic Connectedness	Commuting Zone	2022	Share of high-SES friends among high-SES individuals. *Original Source:* 2022c
New Employer Establishments (Sector)	National	2011 – 2020	Number of new employer establishments by sector. *Original Source:* U.S. Census Bureau 2010-2020

Covariate	Geographic Level Data Outside Set	Years Covered	Description
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[29]
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 2010 [30]
Median Rent	Commuting Zone	2015	Median rent for two-bedroom units in renter-occupied housing. *Original Source:* U.S. Census Bureau 2015a[31]
New Firm Registrations	National	2011 – 2016	Proportion of new firm registrations. *Original Source:* Andrews et al. 2022
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:* Andrews et al. 2022

Notes: This table describes the measurement and sources of geographic data. Indicated variables including Mean Household Income, College Graduate Share, and Median Rent are based on U.S. Census data, originally processed and cleaned by R. 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
Female	Female	0.00	0.00	0.00	0.14	1.00	1.33	2.19	3.39	5.32	12.49	6,792,446
Female	Male	0.00	0.00	0.00	0.00	0.88	1.02	2.00	3.05	4.89	12.38	6,792,446
Female	Pooled	0.00	0.00	0.28	1.00	1.82	2.84	4.41	6.54	10.16	23.46	6,792,446
Male	Female	0.00	0.00	0.00	0.00	0.00	0.82	1.00	1.97	3.39	8.93	5,920,883
Male	Male	0.00	0.00	0.00	0.00	0.69	1.00	1.95	3.15	5.29	14.79	5,920,883
Male	Pooled	0.00	0.00	0.00	0.63	1.06	2.05	3.33	5.14	8.62	22.61	5,920,883
Pooled	Female	0.00	0.00	0.00	0.00	0.53	1.00	1.70	2.63	4.47	11.01	12,800,383
Pooled	Male	0.00	0.00	0.00	0.00	0.83	1.03	2.00	3.17	5.15	13.70	12,800,383
Pooled	Pooled	0.00	0.00	0.00	0.98	1.47	2.46	3.83	5.91	9.48	23.11	12,800,383

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 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	0.72	1.00	2.22	7.06	988,425
2	0.00	0.00	0.00	0.00	0.46	1.00	1.51	2.46	4.27	12.03	1,040,506
3	0.00	0.00	0.00	0.39	1.00	1.46	2.29	3.52	5.93	15.46	1,081,010
4	0.00	0.00	0.00	0.97	1.12	2.00	3.01	4.50	7.25	18.23	1,119,902
5	0.00	0.00	0.51	1.00	1.80	2.60	3.76	5.65	8.77	21.13	1,149,337
6	0.00	0.00	0.96	1.41	2.24	3.34	4.68	6.81	10.35	23.93	1,173,809
7	0.00	0.33	1.00	1.94	2.83	4.08	5.60	7.90	11.77	26.17	1,191,963
8	0.00	0.63	1.40	2.35	3.46	4.75	6.56	9.02	13.17	28.53	1,203,178
9	0.00	0.88	1.81	2.83	4.15	5.53	7.45	10.11	14.55	31.18	1,210,197
10	0.17	1.27	2.39	3.55	4.88	6.48	8.49	11.27	15.97	34.36	1,216,889

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 within an individual's birth cohort, while owner friends deciles classify individuals by the number of owner friends. Bins do not have an equal number of observations because income deciles are constructed using all Facebook users, while the analysis sample restricts to individuals with between 30 and 3000 friends.

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

	Owner Friends Decile										
Parent SES Decile	1	2	3	4	5	6	7	8	9	10	Obs per Decile
1	0.00	0.00	0.00	0.00	0.14	1.00	1.30	2.36	4.31	13.31	6,515,160
2	0.00	0.00	0.57	1.00	1.94	2.78	4.15	5.99	9.32	21.69	6,569,317
3	0.00	0.12	1.00	1.63	2.49	3.58	5.12	7.21	10.88	24.38	6,590,885
4	0.00	0.38	1.00	1.99	2.87	4.16	5.66	7.94	11.80	25.90	6,602,197
5	0.00	0.57	1.26	2.16	3.25	4.49	6.16	8.52	12.49	27.19	6,612,654
6	0.00	0.72	1.48	2.41	3.49	4.79	6.58	9.03	13.14	28.30	6,622,536
7	0.00	0.84	1.66	2.61	3.70	5.20	6.93	9.46	13.71	29.25	6,632,028
8	0.00	0.93	1.80	2.77	3.97	5.43	7.22	9.82	14.21	30.23	6,640,917
9	0.02	1.00	1.94	2.98	4.27	5.60	7.59	10.22	14.67	31.28	6,649,932
10	0.10	1.06	2.08	3.20	4.42	5.90	7.86	10.56	15.24	32.79	6,654,007

Notes: This table reports the average number of owner friends by an individual's parent SES decile and owner friends decile. Deciles are constructed within an individual's birth cohort. Bins do not have an equal number of observations because income deciles are constructed using all Facebook users, while the analysis sample restricts to individuals with between 30 and 3000 friends.

Table D5: 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	0.01338	0.01015
Baltimore, MD	0.01233	0.00875
Boston, MA	0.01406	0.01023
Bridgeport, CT	0.01505	0.01193
Chicago, IL	0.01157	0.00800
Cleveland, OH	0.01277	0.00933
Dallas, TX	0.00940	0.00536
Denver, CO	0.01563	0.01254
Detroit, MI	0.01255	0.00904
Fort Worth, TX	0.00977	0.00596
Houston, TX	0.00873	0.00505
Las Vegas, NV	0.01140	0.00769
Los Angeles, CA	0.00827	0.00476
Miami, FL	0.01014	0.00654
Minneapolis, MN	0.01401	0.00986
New York, NY	0.00951	0.00598
Newark, NJ	0.01122	0.00810
Orlando, FL	0.01046	0.00677
Philadelphia, PA	0.01261	0.00932
Phoenix, AZ	0.01414	0.01070
Pittsburgh, PA	0.01345	0.01008
Sacramento, CA	0.00943	0.00599
San Antonio, TX	0.00853	0.00502
San Diego, CA	0.00964	0.00558
San Francisco, CA	0.00851	0.00477
San Jose, CA	0.00734	0.00402
Seattle, WA	0.01346	0.00998
St. Louis, MO	0.01412	0.01006
Tampa, FL	0.01231	0.00796
Washington DC	0.01157	0.00743

Notes: This table reports the average fraction of friends who are firm owners and the average fraction of users who are firm owners for the 30 largest commuting zones. The denominator for each column considers only individuals who have listed at least one job on their public Facebook profile. Sample characteristics for individuals reporting at least one job are reported in Column (2) of Table 1. The data are described in Section II and Appendix Section A.1

Table D6: 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	3.46×10^{-5}	4.71×10^{-5}		1.92×10^{-5}	9.20×10^{-6}	1.48×10^{-5}			
	(2.11×10^{-5})	(2.16×10^{-5})		(1.30×10^{-5})	(1.57×10^{-5})	(8.54×10^{-6})			
Owner Friends of Friends with >$1M in revenue			8.29×10^{-5}				1.24×10^{-5}		
			(4.85×10^{-5})				(1.94×10^{-5})		
Is Owner								0.32	
								(0.19)	
Owner Friends of Friends Among Friends who did Not Start a Firm									7.02×10^{-5}
									(2.26×10^{-5})
Control for Pre-College Owner FoFs	X		X	X	X	X	X	X	X
Flexible Control for Pre-College Owner FoFs		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
School Fixed Effects	X								
Birth Cohort Fixed Effects	X								
Number of Students	736,973	736,973	736,973	736,973	736,973	736,973	736,973	736,973	711,593
Mean of Dep. Variable	0.0026	0.0026	0.0026	8.55×10^{-4}	8.55×10^{-4}	3.53×10^{-4}	3.53×10^{-4}		0.0026
SD of Dep. Variable	0.051	0.051	0.051	0.030	0.030	0.019	0.019		0.051
Mean of Ind. Variable	3.12	3.12	1.04	3.12	3.12	3.12	1.04		2.99
SD of Ind. Variable	3.12	3.12	1.35	3.12	3.12	3.12	1.35		2.97

Notes: This table provides a set of robustness results for the College design. Baseline estimates for the college design can be found in Table 3. Column (1) repeats the baseline design with college and year fixed effects. Column (2) allows the control for the number of pre-college owner friends-of-friends in the year the student turns 18 to vary flexibly by 4-week period. For example, the number of owner friends-of-friends in the first 4 weeks, the second 4 weeks, and so on up to the 33rd week are included separately in the regression. Column (3) replaces the treatment in the baseline specification with the number of owner friends-of-friends who have greater than $1 million in annual revenue. Column (4) replaces the outcome with a narrow measurement of owners using a string matched job title to "Founder". Column (5) repeats Column (4) but includes a control for the number of friends that the focal individual made during the first 12 weeks of college. Column (6) replaces the outcome of the baseline specification with ownership of a firm that makes greater than $1 million in annual revenue. Column (7) repeats column (6) but replaces the treatment with the number of owner friends-of-friends who have greater than $1 million in annual revenue. Column (8) estimates a 2SLS regression model where the outcome is ownership of a firm that makes greater than $1 million in annual revenue and the endogenous regressor is ownership of any firm. The interpretation of Column (8) as reflecting the success of marginal firms is discussed in Section III.A.2. Column (9) replaces the treatment from the baseline regression with the number of owner friends-of-friends where the mutual friend did not start a firm. Details of the sample selection and baseline specification are discussed in Section III.A and the notes to Table 3.

Table D7: College Design: SES-Specific Exposure Effects

Dependent Variable	(1) Is Owner	(2) Is Owner	(3) Is Owner	(4) Is Owner	(5) Is Owner	(6) Is Owner
Owner Friends of Friends Above 90th Percentile Income	2.43×10^{-4}		2.16×10^{-4}	6.82×10^{-5}		1.41×10^{-4}
	(9.02×10^{-5})		(9.20×10^{-5})	(2.76×10^{-4})		(1.55×10^{-4})
Owner Friends of Friends Below 90th Percentile Income		8.00×10^{-5}	5.46×10^{-5}		1.37×10^{-4}	-5.12×10^{-6}
		(4.50×10^{-6})	(4.60×10^{-5})		(1.50×10^{-4})	(2.85×10^{-4})
Parent Income Percentile × Owner FoFs Above p90 Income				1.13×10^{-6}		1.94×10^{-6}
				(3.58×10^{-6})		(3.70×10^{-6})
Parent Income Percentile × Owner FoFs Below p90 Income					-1.22×10^{-6}	-1.50×10^{-6}
					(1.95×10^{-6})	(2.02×10^{-6})
Observations	725,618	725,618	725,618	573,867	573,867	573,867
Mean of Dependent Variable	0.0026	0.0026	0.0026	0.0027	0.0027	0.0027
SD of Dependent Variable	0.0509	0.0509	0.0509	0.0522	0.0522	0.0522
Mean of Independent Variable	0.38	1.21				
SD of Independent Variable	0.73	1.49				

Notes: This table estimates treatment effects in the College design separately by the childhood income of the student and by the income of the connected owner.

Table D8: Distribution of Truncated Katz-Bonacic Centrality Estimates

Attenuation Factor	Mean	Q1	Q2	Q3	Q4	$\frac{\text{Mean Truncated Katz-Bonacic Centrality}}{\text{Mean Influence on Direct Connections}}$
0.0040	0.3927	0.0370	0.1500	0.3457	1.0385	1.893
	(0.0072)					
0.0045	0.5043	0.0439	0.1823	0.4286	1.3630	2.161
	(0.0100)					
0.0050	0.6518	0.0518	0.2207	0.5309	1.8048	2.514
	(0.0144)					
0.0055	0.8509	0.0606	0.2665	0.6591	2.4181	2.983
	(0.0212)					
0.0060	1.1237	0.0706	0.3218	0.8228	3.2805	3.611
	(0.0321)					
0.006187 (estimate)	1.2509	0.0747	0.3454	0.8956	3.6888	3.898
	(0.0376)					
0.0065	1.5014	0.0820	0.3888	1.0344	4.5015	4.454
	(0.0492)					

Notes: This table shows the distribution of truncated centrality estimates across quartiles for different attenuation factors, with standard errors in parentheses. Column (1) estimates the mean truncated centrality across the sample. Columns (2)-(5) estimate the mean truncated centrality in each quartile of the sample. Column (6) reports the ratio of mean truncated centrality divided by mean influence at distance 1 from the focal individual, where influence at distance 1 is calculated as the attenuation factor times degree. Estimates were obtained using a sampling-based SAKE method (Lin et al. 2021) with inverse sampling probability correction. We generate centrality estimates for a random sample of 4999 individuals in the analysis data. We estimate centrality up to a distance of 8 from the focal individual, where connections at a distance ≥ 2 are sampled with with probability 0.01. The estimator and validation exercises are described in Appendix Section C.1.1.

Table D9: Top 10 CZs with the Highest Shrunk Total Effect Among the 50 Most Populated CZs

CZ Name	Total Effect	Direct Effect
Atlanta, GA	2.31×10^{-4}	1.01×10^{-4}
Baltimore, MD	2.40×10^{-4}	9.0×10^{-5}
Charlotte, NC	2.68×10^{-4}	1.05×10^{-4}
Columbus, OH	2.54×10^{-4}	1.28×10^{-4}
Grand Rapids, MI	2.37×10^{-4}	9.0×10^{-5}
Los Angeles, CA	2.55×10^{-4}	7.0×10^{-5}
New York, NY	2.51×10^{-4}	5.8×10^{-5}
San Diego, CA	2.46×10^{-4}	1.17×10^{-4}
St. Louis, MO	2.36×10^{-4}	1.26×10^{-4}
Tampa, FL	2.53×10^{-4}	1.31×10^{-4}

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).

Table D10: Individual-Level Targeting under Extended Model: High-Impact Individuals

	(1)	(2)	(3)
	Commuting Zones in Top 50% of Direct Effect	Commuting Zones in Top 50% of Total Effect	Ratio of (2)/(1)
Firms Created among Treated Population	1	0.92	0.92
Firms Created among Individuals Not Directly Treated	0.51	0.67	1.30
Total Firms Created	1.51	1.59	1.05

Notes: This table provides counterfactual estimates of the effect of a hypothetical policy on firm entry under the extended model. We consider a policy targeted to individuals with many existing connections to entrepreneurs who are not entrepreneurs themselves, and estimate the effects on entry targeted individuals and among all individuals in the economy. The effect of the hypothetical policy is normalized such that the policy would create one new firm among directly treated individuals if targeted to individuals who live in commuting zones where the Direct Effect is above the population-weighted median. The first row presents the number of firms created by individuals directly treated by the policy. The second row presents the number of firms created by individuals not directly treated by the policy who are induced to entry through the spillover effects. The third row presents the total number of firms that enter, equal to the sum of the first and second rows. Column (1) considers a policy targeted to individuals living in commuting zones where the direct effect of the policy is above the population-weighted median, such that the first row of Column (1) is equal to unity under our normalization. Column (2) considers a policy targeted to individuals living in commuting zones where the total effect of the policy, including spillovers, is above the population-weighted median. Column (3) calculates the ratio of (2) / (1). The basic model and estimation approach are described in Section IV. The extension and results are discussed in Section V.C.

Appendix E: Supplemental Figures

Figure E1: Summary of Data Linkage Relationships

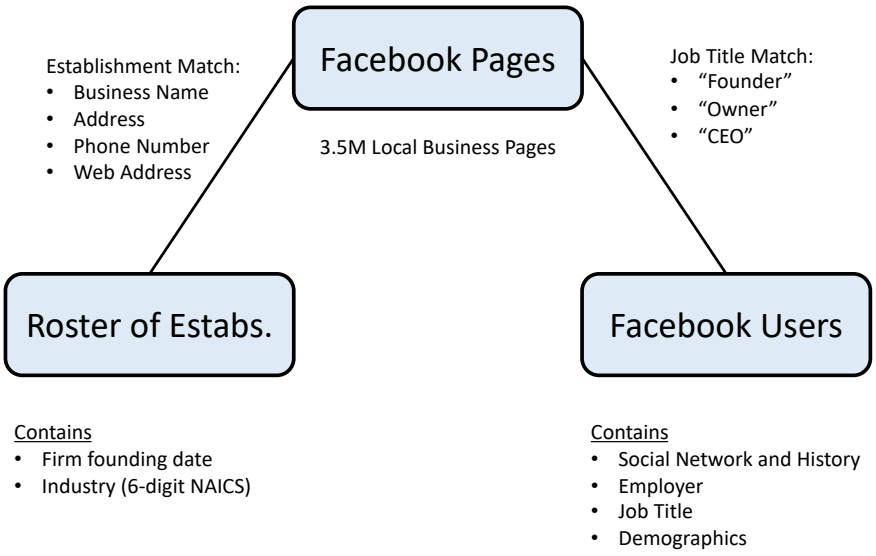
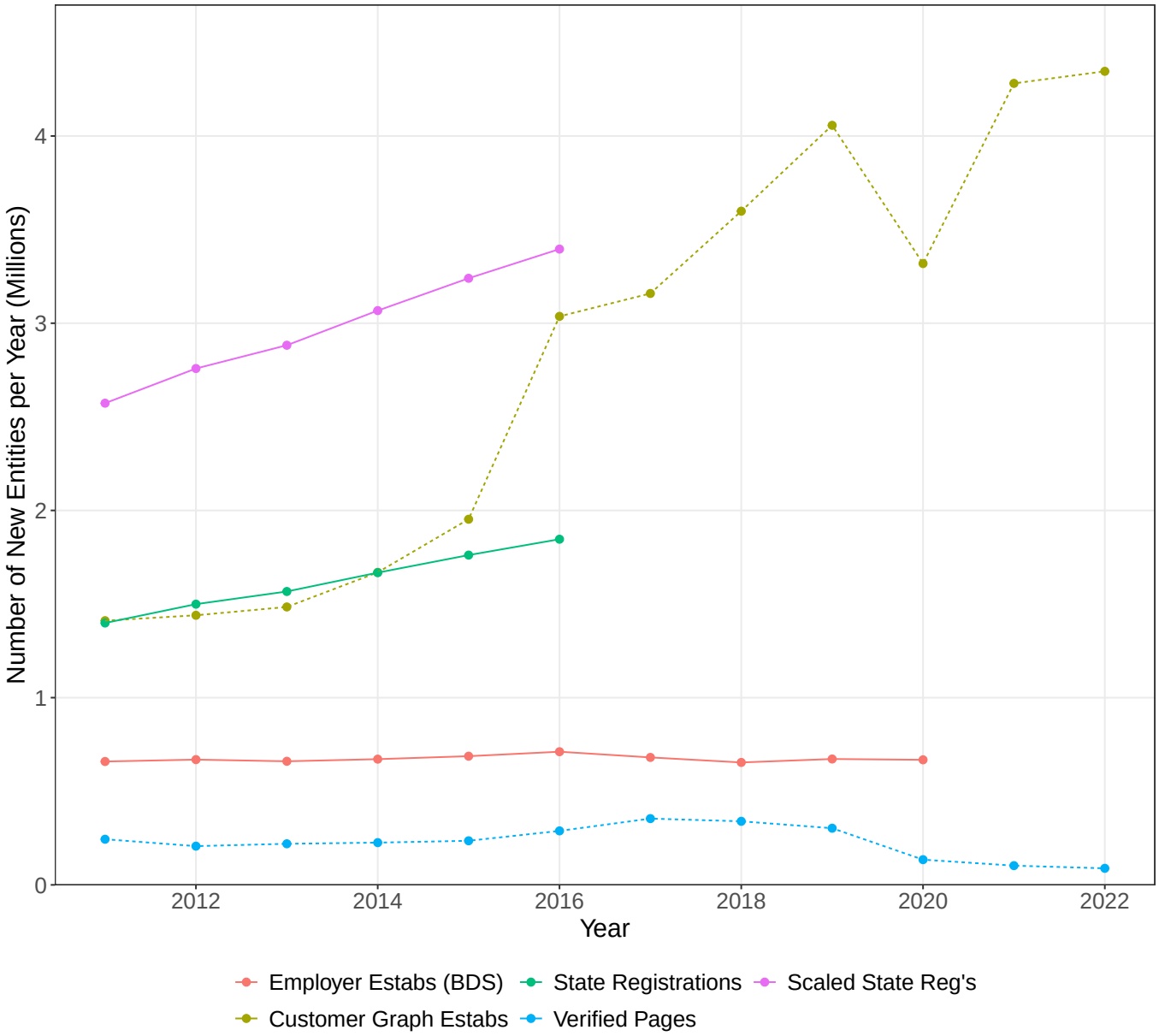
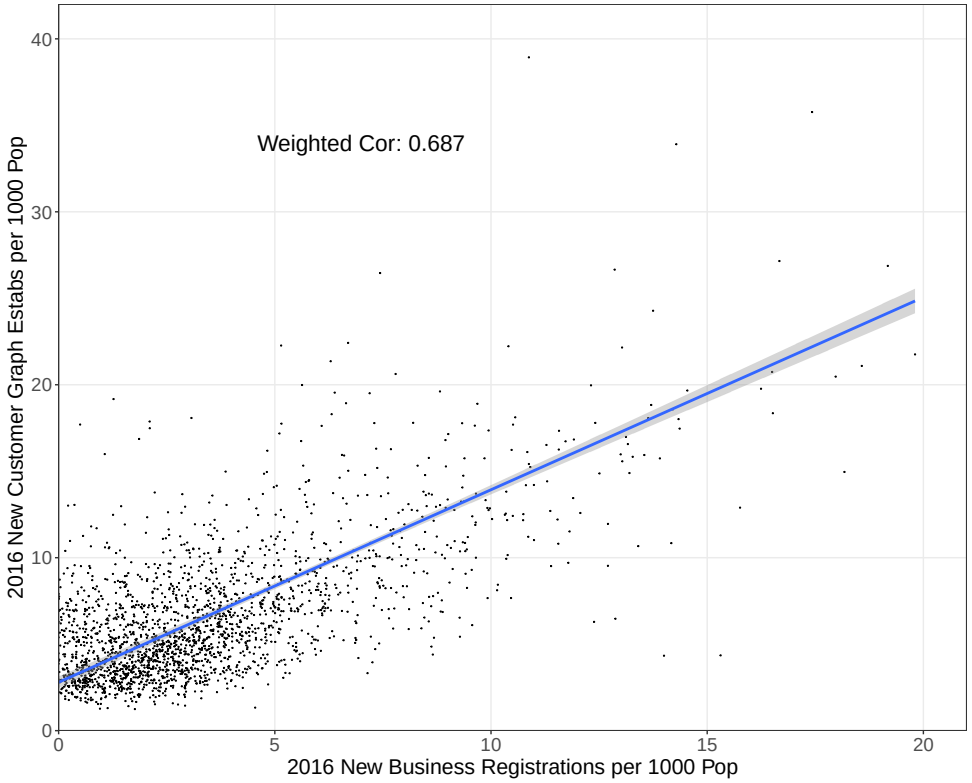


Figure E2: New Establishments by Year: Benchmarks



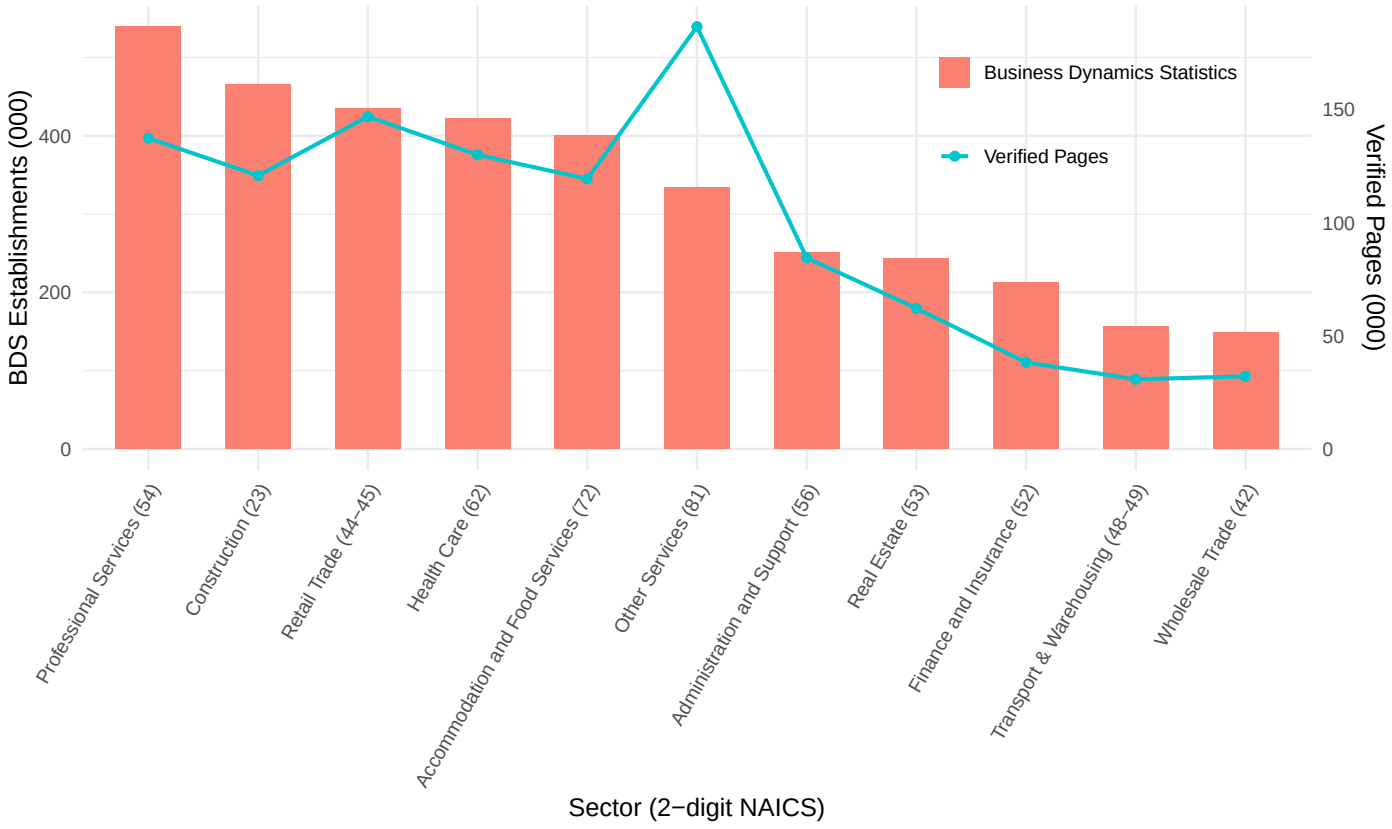
Notes: This figure plots the number of establishments per year from 2011 to 2022 in our sample and in comparable public data. Dashed lines represent out data, and solid lines represent publicly available data on comparison groups. The dashed yellow line represents all new establishments identified in the Customer Graph data by year. The green solid line shows the number of new state business registrations from Andrews et al. 2022, and the pink provides a simple scaling of the green line to adjust for the omission of proprietorships in the registration data. The dashed blue line presents the number of verified Pages by year, which are created by merging Facebook Pages for businesses to the Customer Graph. The solid red line represents the number of new employer establishments in the Business Dynamics Statistics. Further details and analysis are provided in Section II.

Figure E3: Spatial Patterns in Entry per Capita: County-Level Benchmarking



Notes: This figure compares the number of matched Facebook pages per capita to the number of new business registrations per capita in 2016. The reported correlation is weighted by county population. Spatial patterns in business registrations are collected and by Andrews et al. 2022. Facebook pages are matched to 3rd-party data on U.S. establishments using name, location, and publicly available contact information. The data construction and matching process is described in Section II.

Figure E4: New Establishments between 2015-2020 by 2-Digit NAICS Sector: FB vs BDS



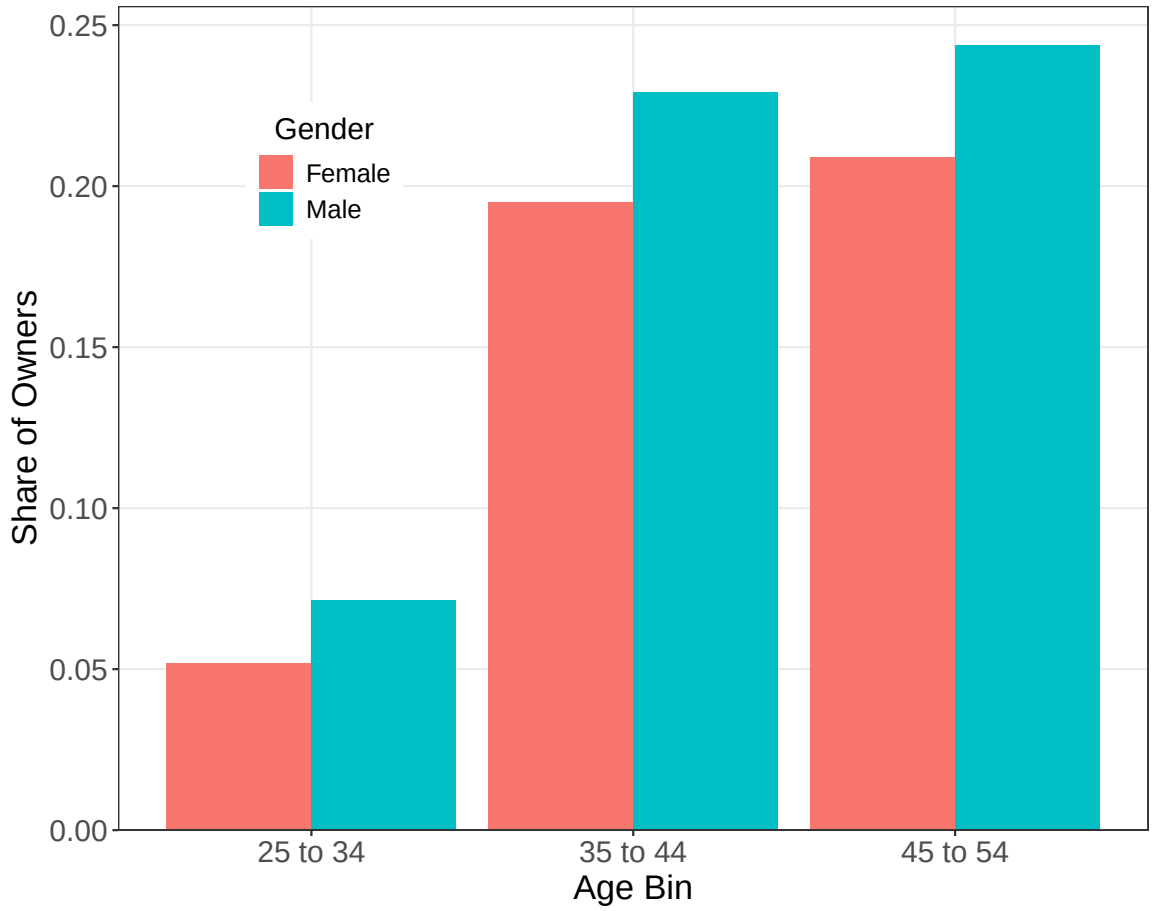
This figure compares the number of verified Facebook pages to the number of new establishments between 2015 and 2020 in all 2-digit NAICS sectors with more than 100,000 new establishments during that period, based on the Business Dynamics Statistics. Facebook pages are verified and matched to third-party data on U.S. establishments using name, location, and publicly available contact information. The data construction and matching process is described in Section II.

Figure E5: Sample Facebook Business Page



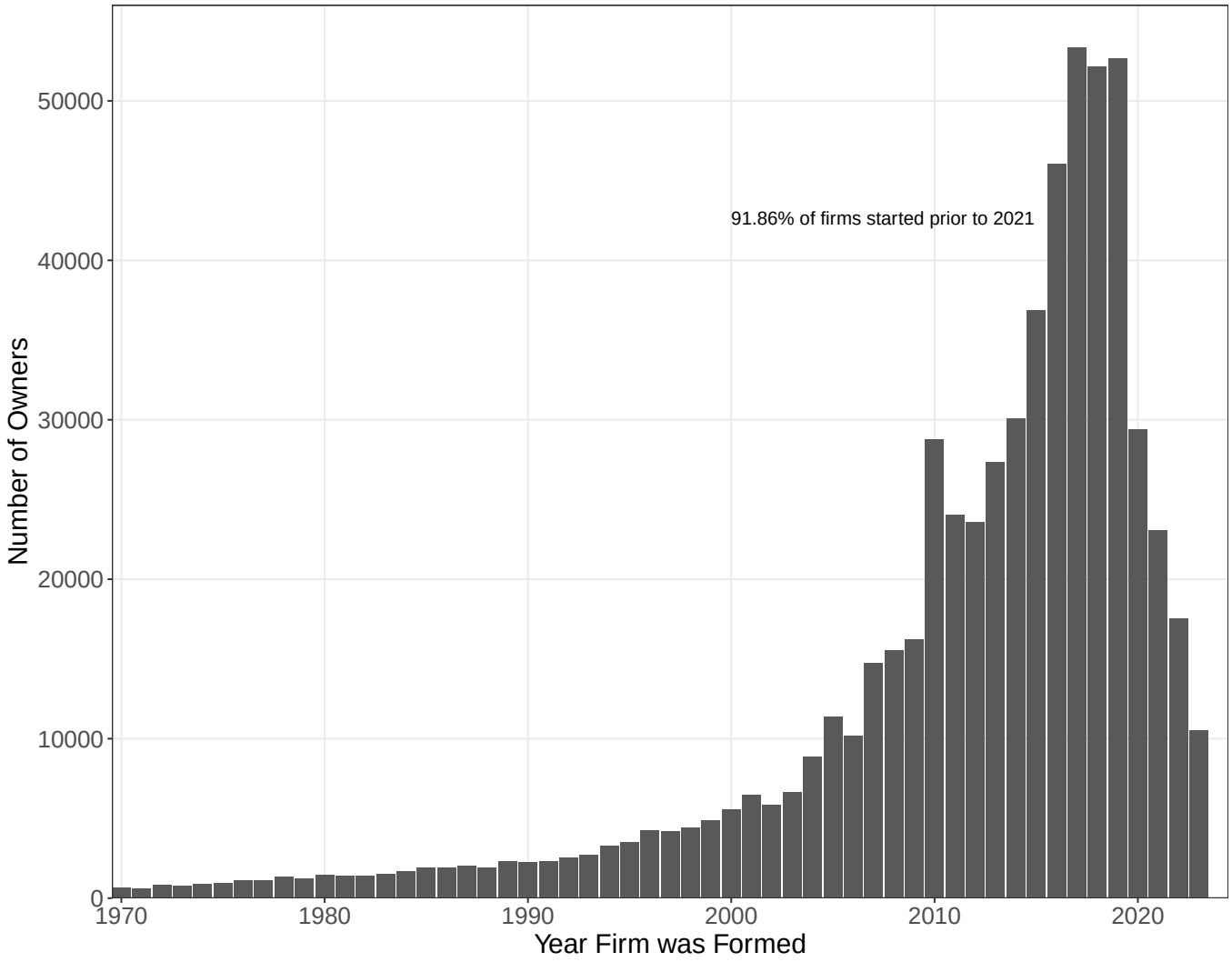
Notes: This figure presents a sample Facebook page for a local business. We use Facebook pages to connect business owners with Facebook profiles to firmographic data. Details of the linking process are presented in Section II.

Figure E6: Age and Gender Breakdown of Owners



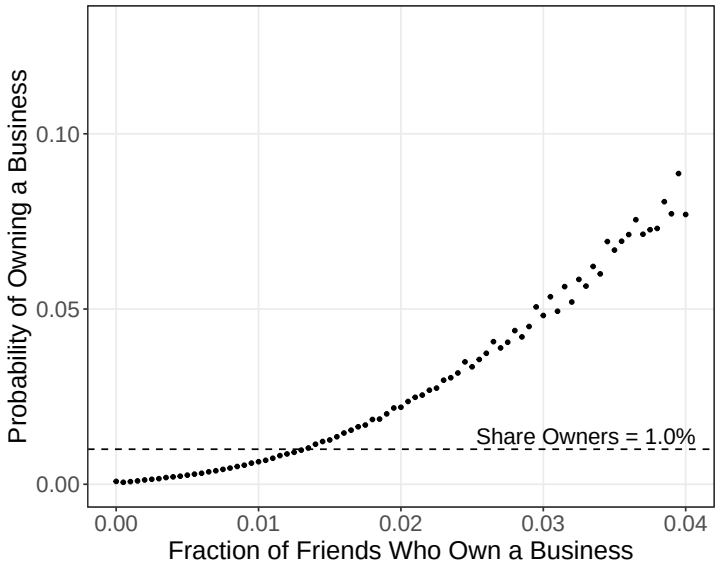
Notes: This figure depicts the age and gender distributions of firm owners identified in our data. Our data contain 630,000 individuals that can be matched to pages in 3rd-party data Owners are matched to pages using the individual's self-reported employer and job title on their public Facebook profile. Details of the matching procedure are described in Section II

Figure E7: Number of Firm Owners on Facebook by Firm Start Year



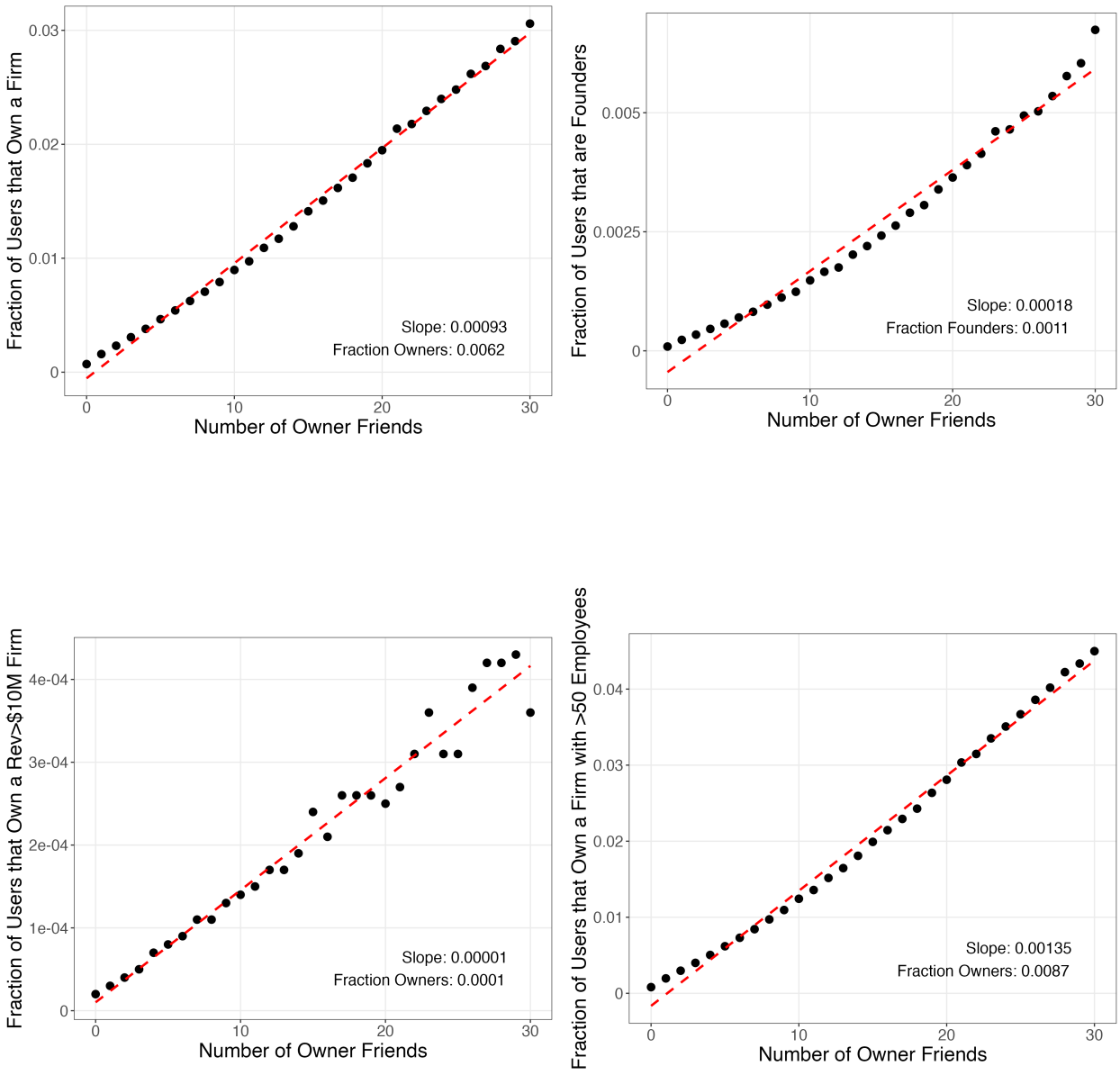
Notes: This figure displays the number of firm owners in the primary sample by the year that they started their first firm. Note that the set of firms for this figure are those that survive the matching process described in Section A.1 where data from the Customer Graph and Establishment Pages are matched. 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 E8: Firm Ownership by Rates of Firm Ownership Among Friends



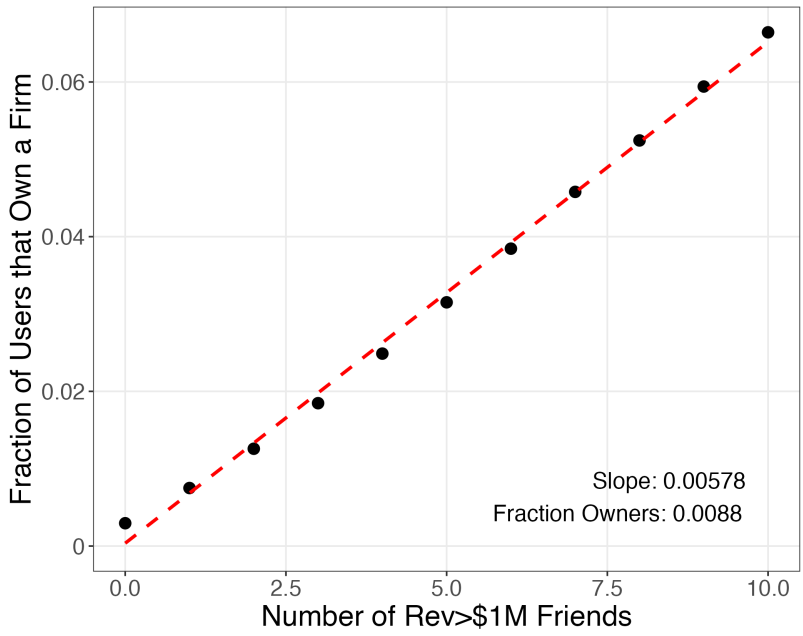
Notes: This figure shows that the fraction of an individual's friends who own a business predicts whether the individual owns a business among individuals with more than 100 friends.

Figure E9: Firm Ownership by Number of Firm Owners Among Friends: Alternative Owner Definitions and Outcomes



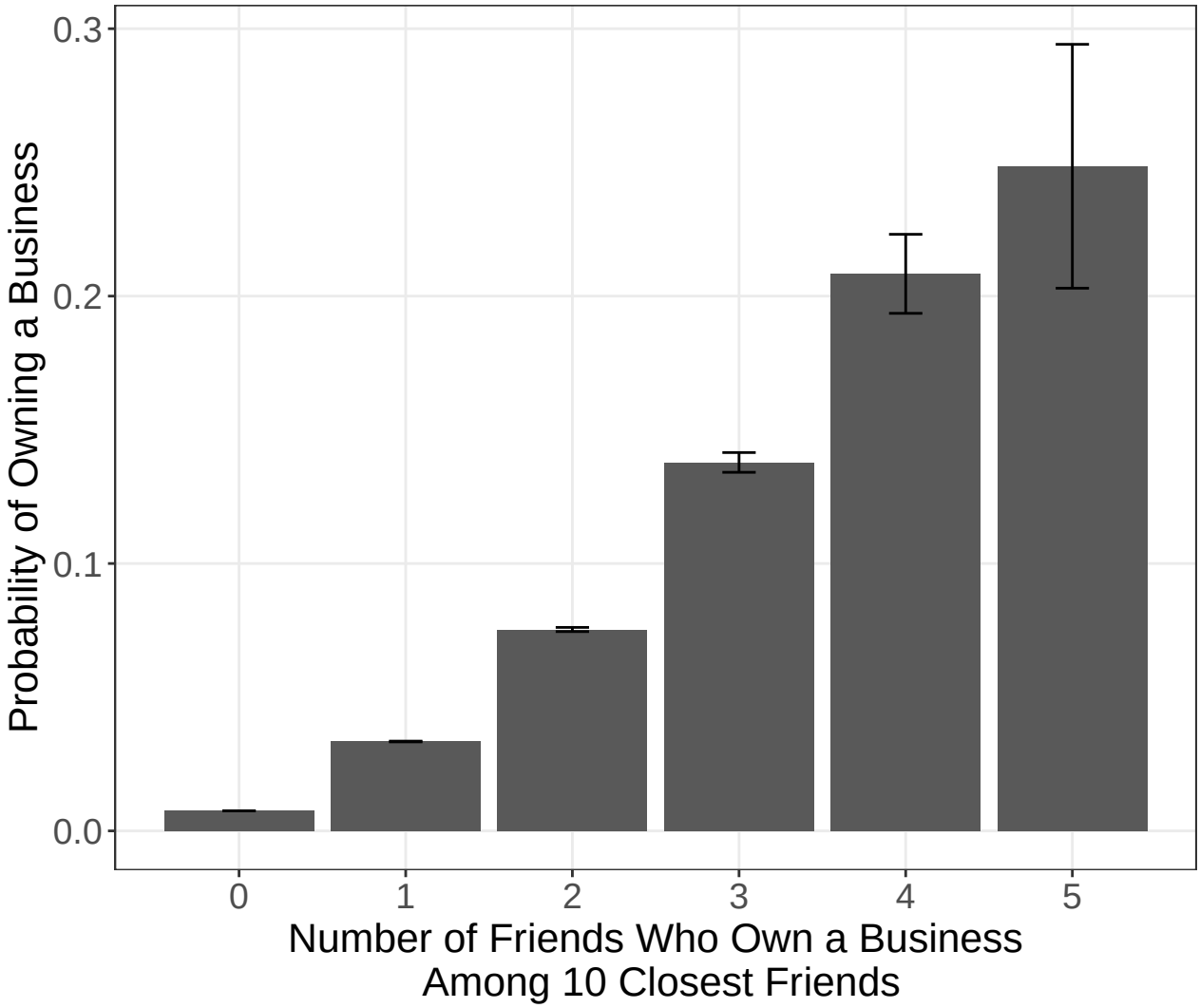
Notes: This figure presents scatterplots of the fraction of individuals who own a firm by the number of friends they have who own a firm for alternative definitions of the outcome. The upper left plot restricts to firm owners that have zero co-founders. The upper right plot restricts the definition of 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 E10: Firm Ownership by Number of Firm Owners with $1 Million in Annual Revenue Among Friends



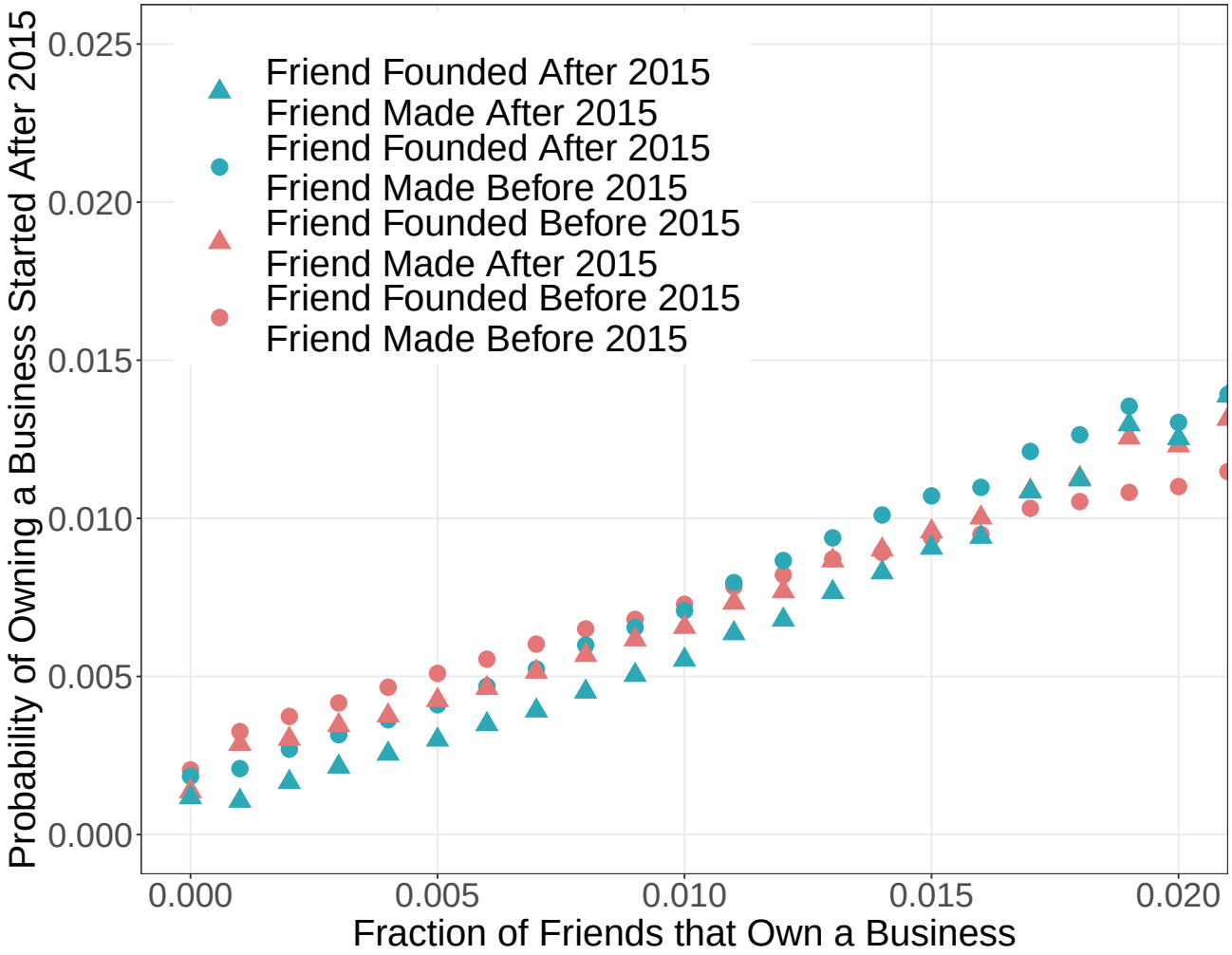
Notes: This figure shows that the number of an individual's friends who own a business with $1 Million in annual revenue predicts whether the individual owns a business.

Figure E11: Probability of Owning a Firm by Number of Firm Owners among 10 Closest Friends



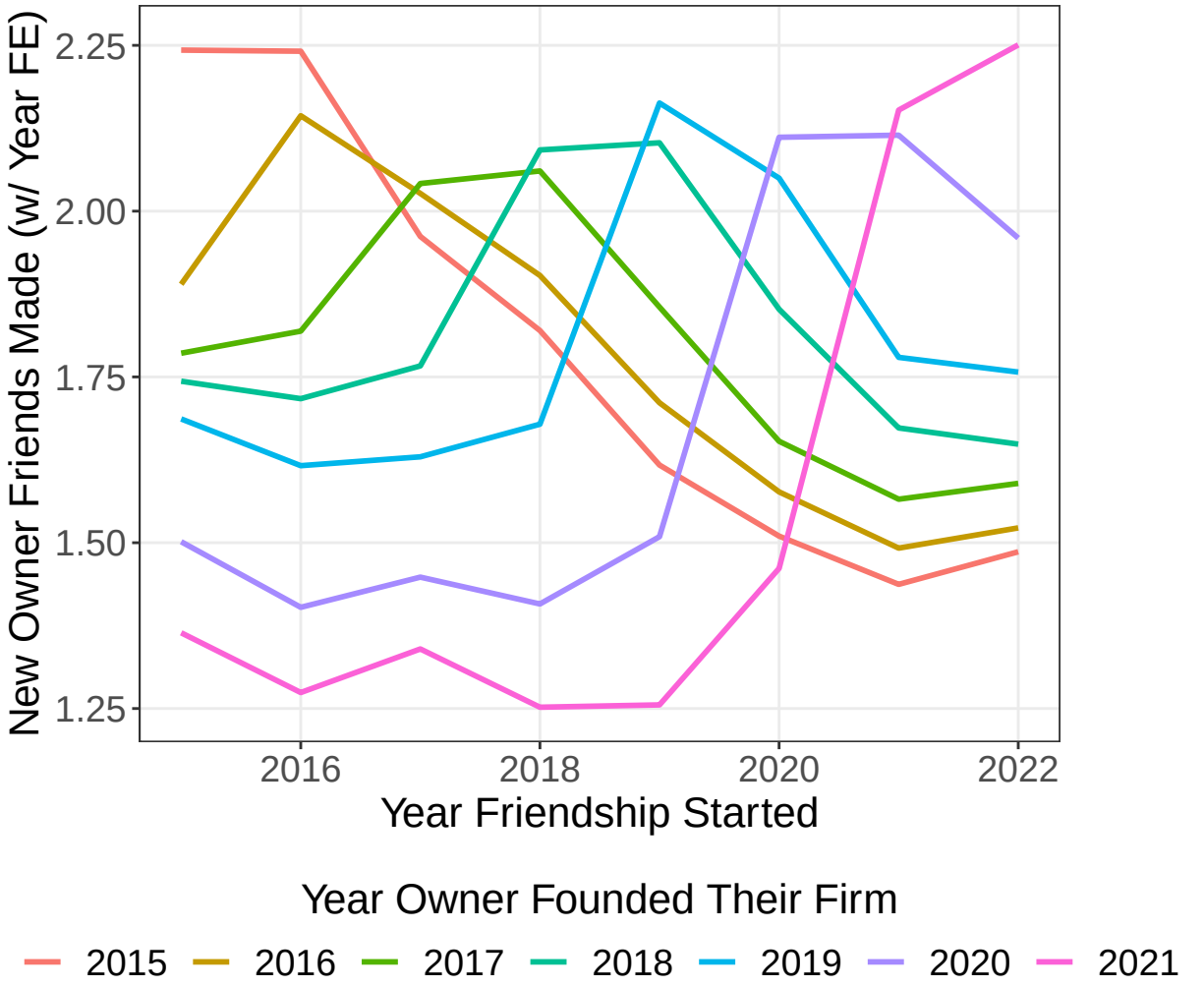
Notes: This figure repeats the analysis performed in Figure 2a, but restricts the dataset to the 10 closest friends. Close friends are identified based on a model which measures the frequency of public interactions such as likes, comments, and tags. Section A.2.4 gives a detailed description of how close friends are identified. This figure shows that even when restricting to only consider the 10 closest friends, the number of an individual's friends who own a business still predicts whether the individual owns a business. Further description is provided in Section II.B.

Figure E12: Founding Rates by Friend Ownership Rates and Friendship Length



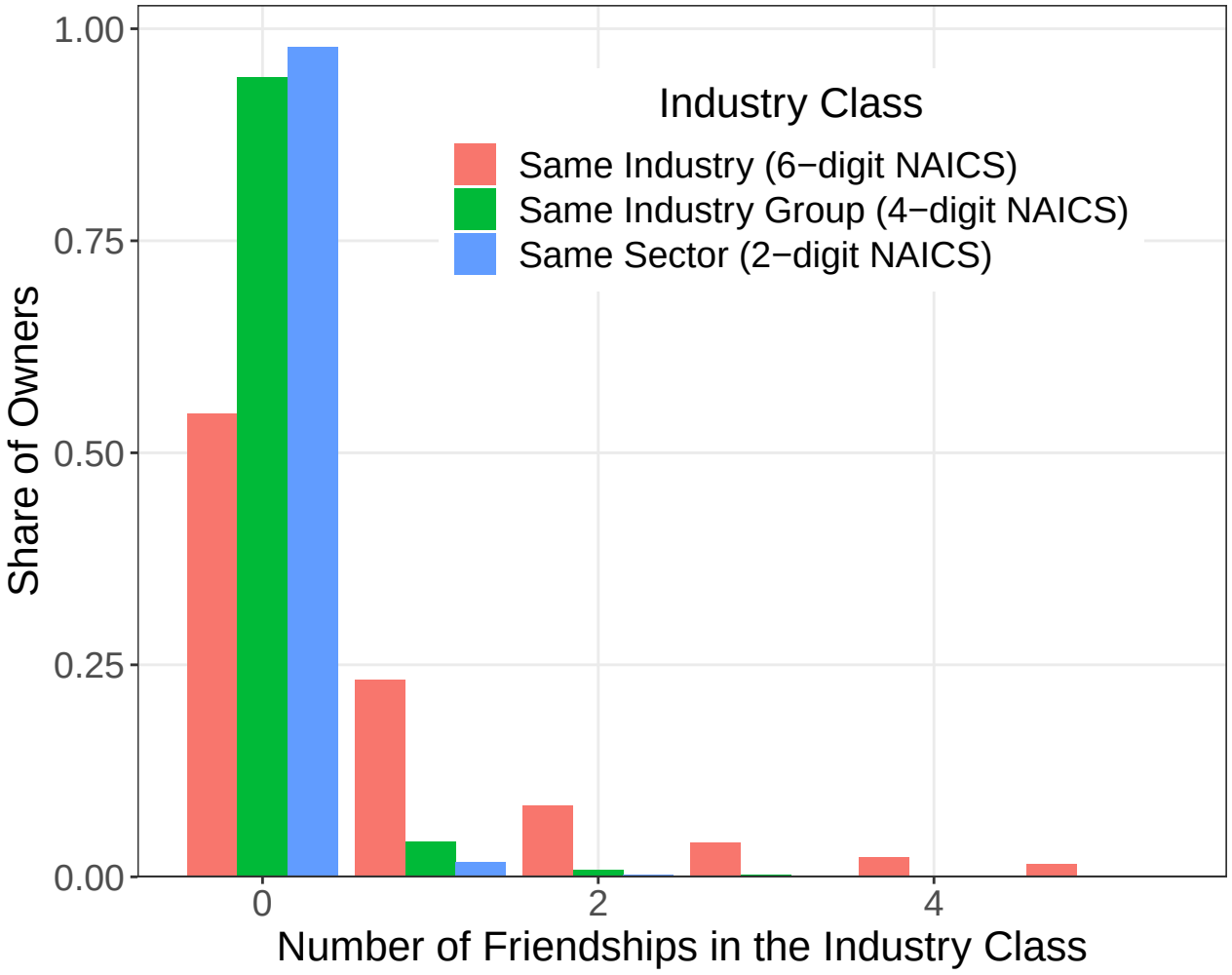
Notes: This figure repeats the analysis from Figure E8 but disaggregated by the year the friendship was formed and the year the friend founded their firm. This figure demonstrates that the relationship between entrepreneurship and friendship holds across both founding date and friendship formation date. Further detail on the data provided in Section II.B.

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



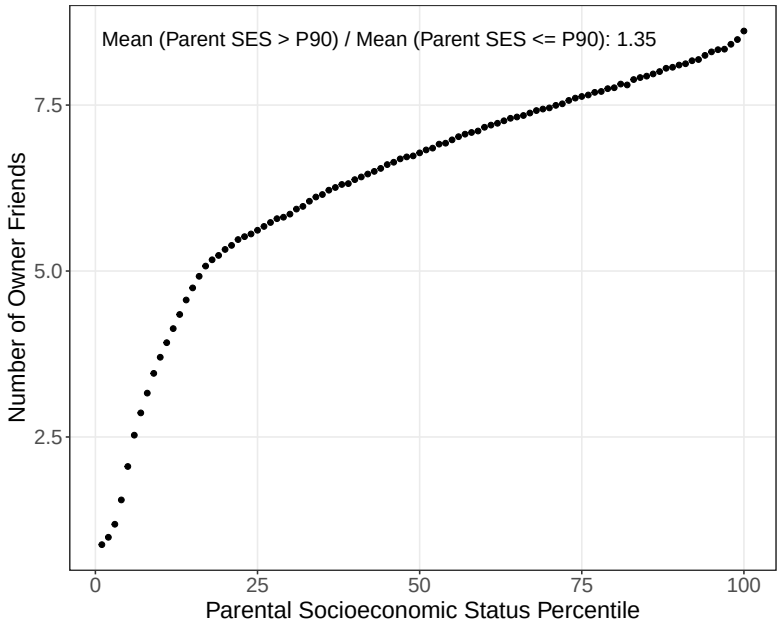
Notes: This figure displays the yearly rates at which owners befriend owners, separated by the year that the focal owner started their firm. This specification includes year fixed effects and adds back the overall mean across all periods. This figure is further analyzed in Section II.B.

Figure E14: Distribution of Owner Friends by Industry Similarity



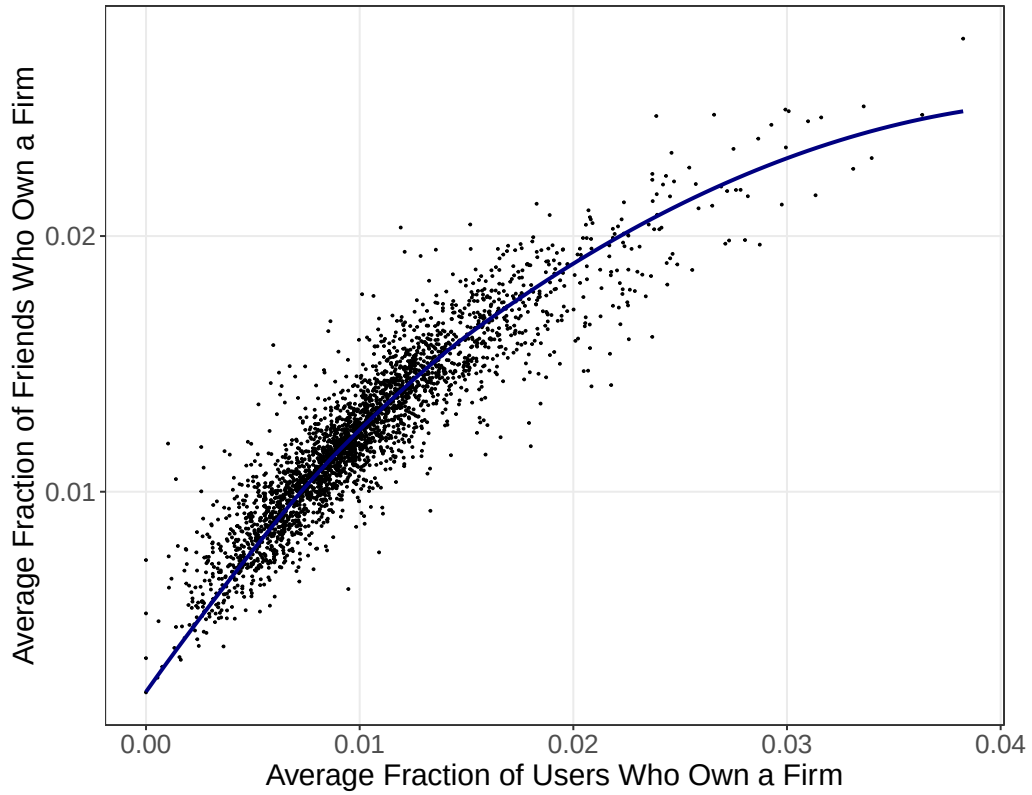
Notes: This figure shows the number of owner friends present at each industry level. Each color is a separate distribution showing the number of owner friends in each industry level separated by NAICS class. So, the red color graphs the distribution of friends in the same 6-digit NAICS industry, the 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 E15: Average Number of Owner Friends by Percentile of Parent Income



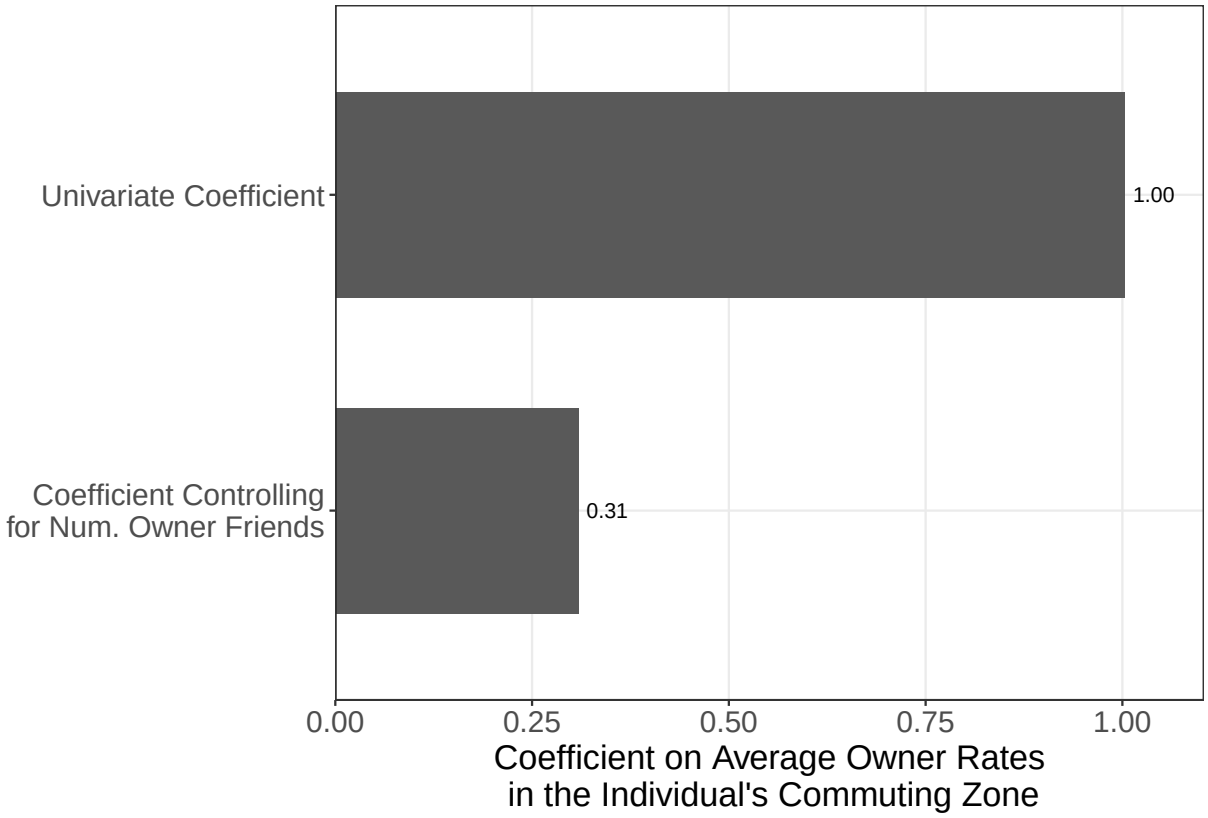
Notes: This figure plots the average number of friends who own a firm at present, by income percentile of the focal individual's parents. Income measurement is described in Appendix Section A.1.

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



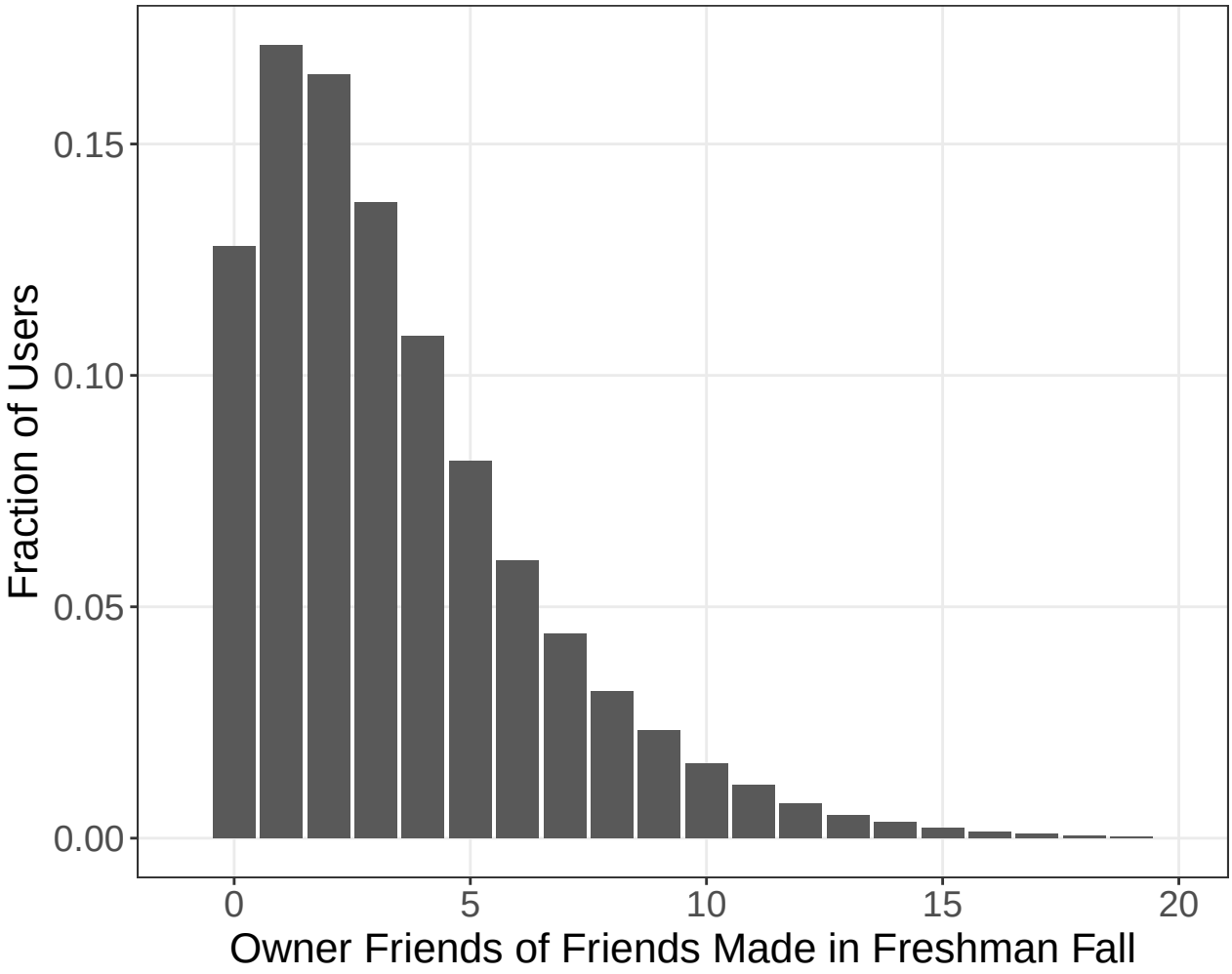
Notes: This figure is a county-level scatterplot comparing the fraction of friends who are owners with the fraction of users who own a firm. The plot includes all counties with Individuals in the analysis sample are born between 1966 and 2000, have between 30 and 3000 friends, and list at least one job title on their Facebook profile. Friends are restricted to those that have at least one job title. Counties with fewer than 1000 users are omitted. The LOWESS best fit is weighted by the number of observations in the county. Summary statistics for the sample are provided in Table 1.

Figure E17: Geographic Variation in Entrepreneurship and Number of Owner Friends



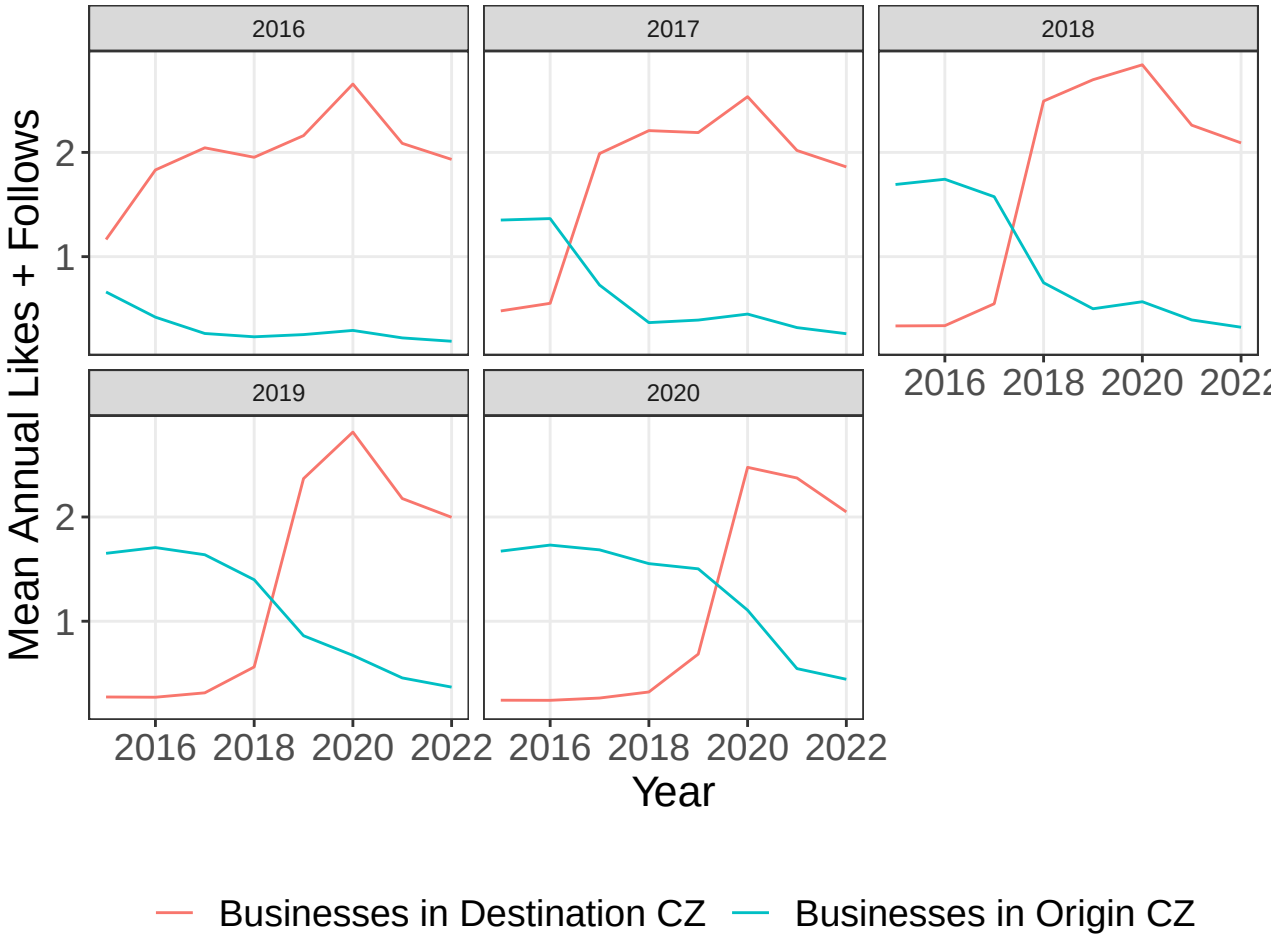
Notes: This figure presents coefficients from two individual-level regressions of entrepreneurship status on the mean rate of entrepreneurship in the individual's commuting zone as measured on Facebook. The first model shows that the coefficient on local ownership rates are equal to one in a univariate regression. The second model shows that the predictive power of local entrepreneurship rates drops dramatically when the individual-level number of owner friends is included as a control. Both regressions are run on all individuals who were born between 1966 and 2000, have between 30 and 3000 friends, and have at least one self-reported job title on their public Facebook profile, equal to 67 million individuals. Summary statistics for the sample are presented in Table 1.

Figure E18: Distribution of Owner Friends-of Friends in College



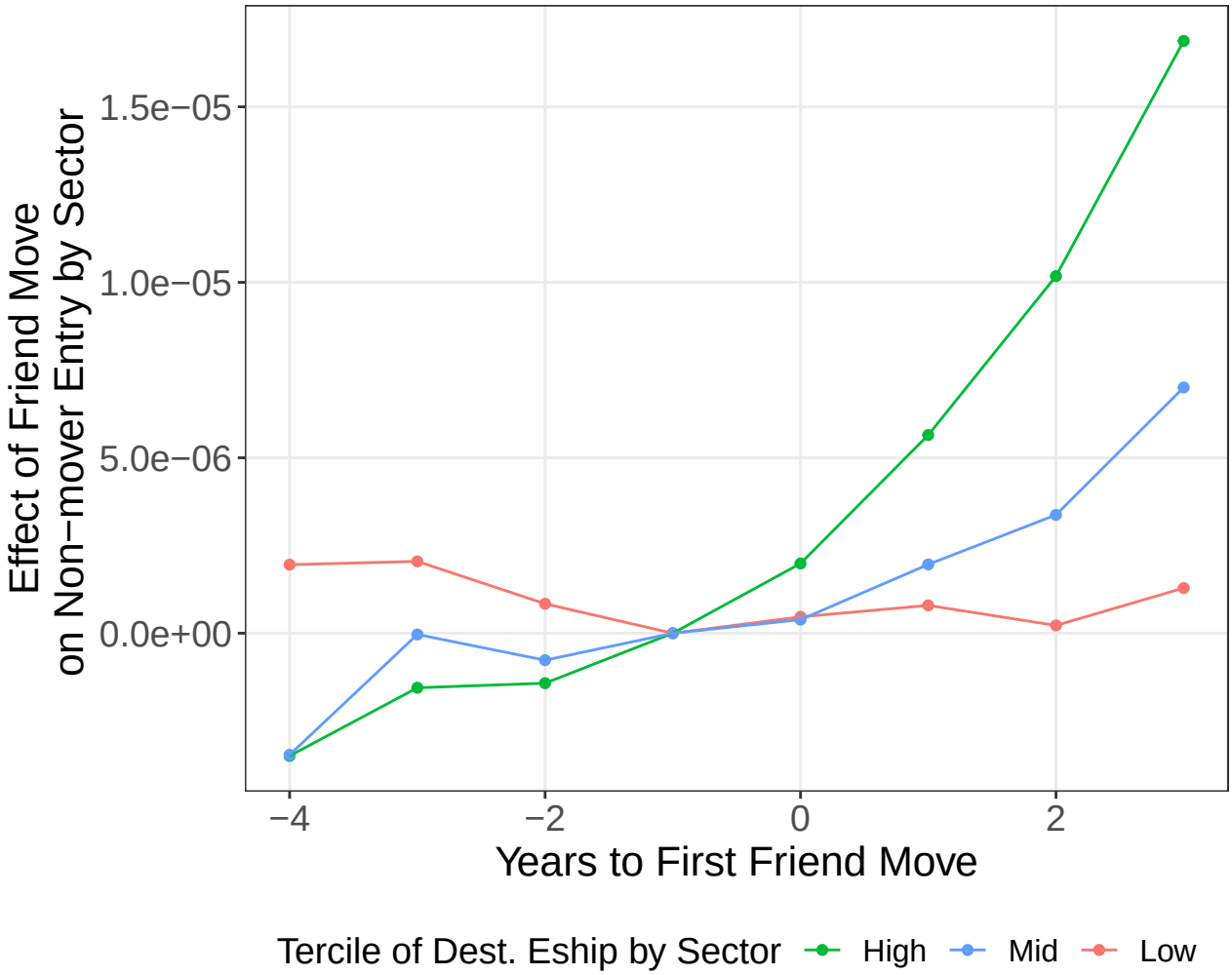
Notes: This figure plots the number of owner friends-of friends in the College Design. For this figure, in order to ensure that only an individual's friending behavior and activity are driving the results, we impose a number of specific characteristics on 2nd-degree connections formed during the first 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 E19: Interactions with Local Business Pages by Year of Move



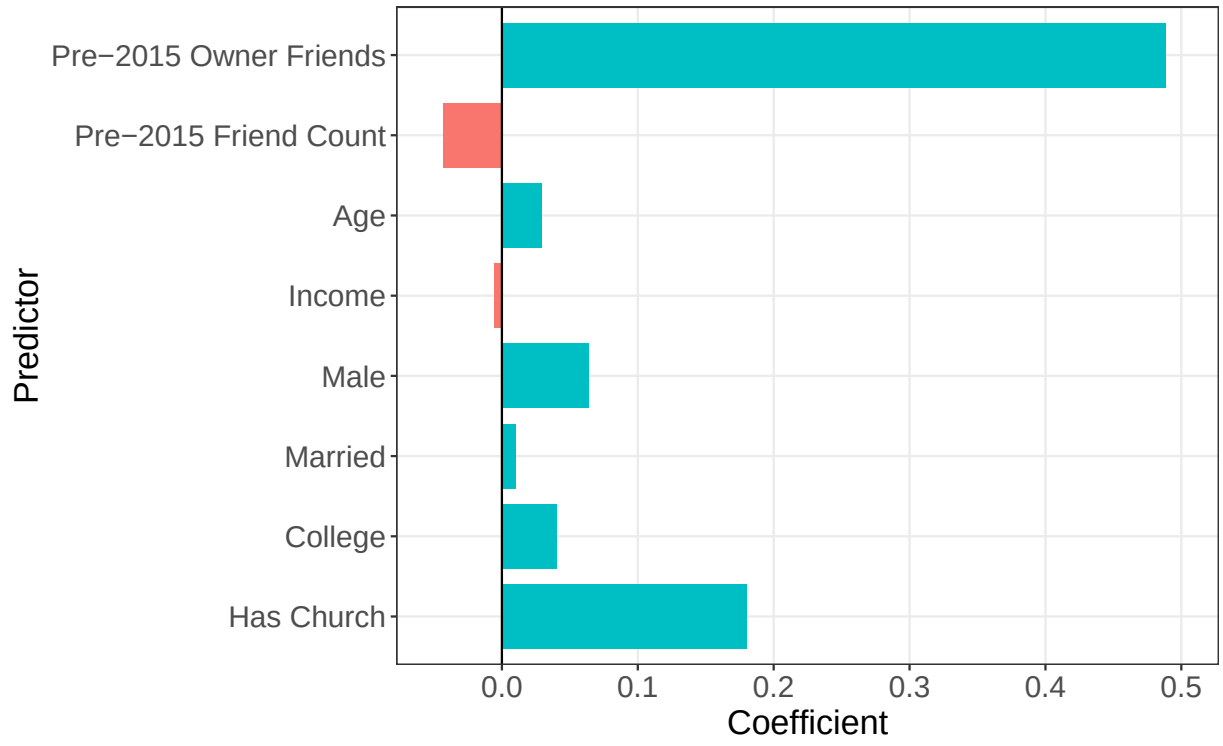
Notes: This figure plots the average mean interactions (likes plus follows) between movers and local business pages in the origin vs. the destination for movers in each year between 2016 and 2020. Further information is provided in Section III.B

Figure E20: Commuting-Zone by Sector Difference in Differences



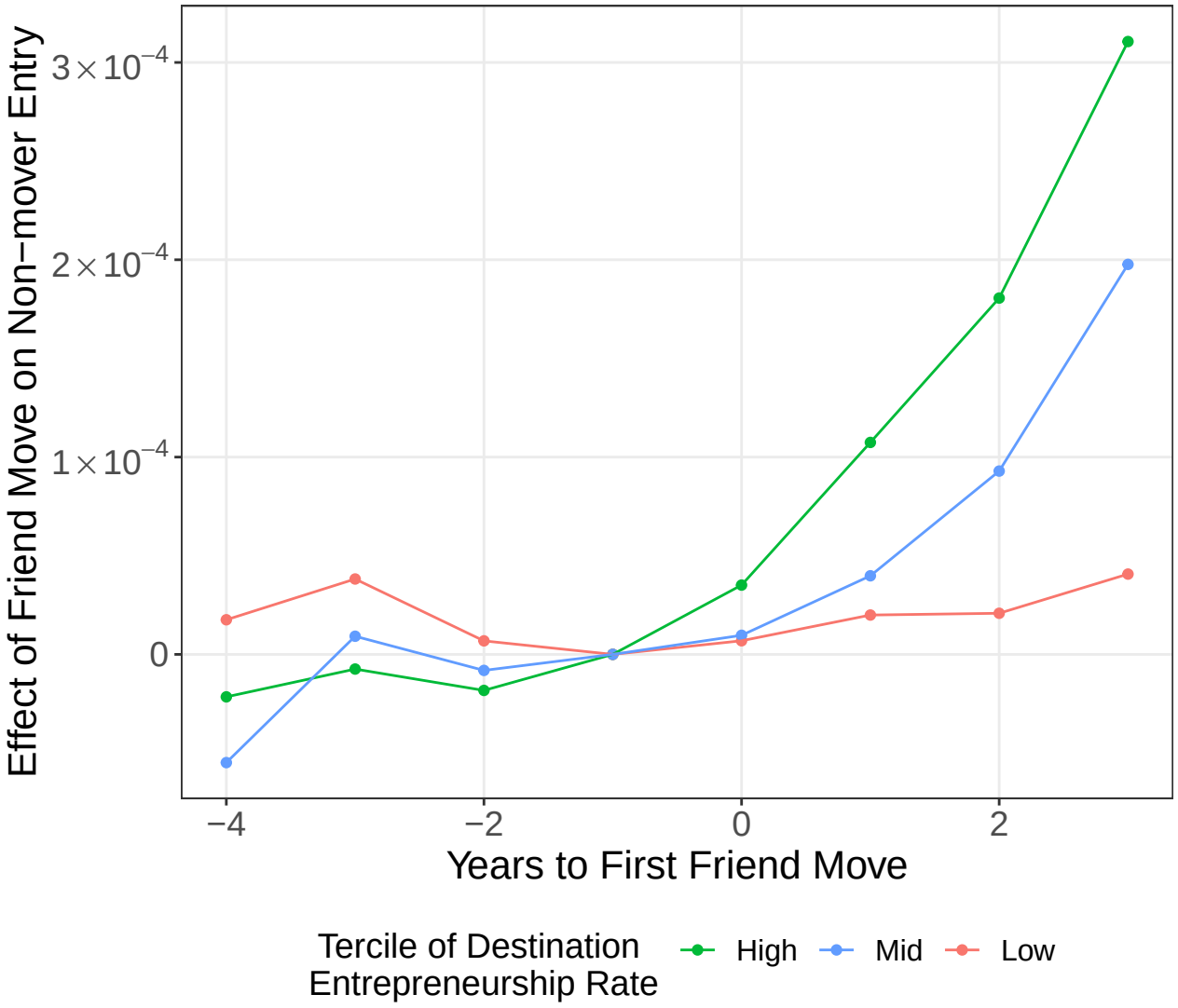
Notes: This figure repeats the analysis in Figure 8a but defines the destination market at the commuting zone-by-sector level. The same technical designs are used from Figure 8a to produce the results in this figure. Further analysis of the figure and comparisons to Figure 8a are given in Section III.B.

Figure E21: Predictors of Fraction post-2015 Friends who are Owners



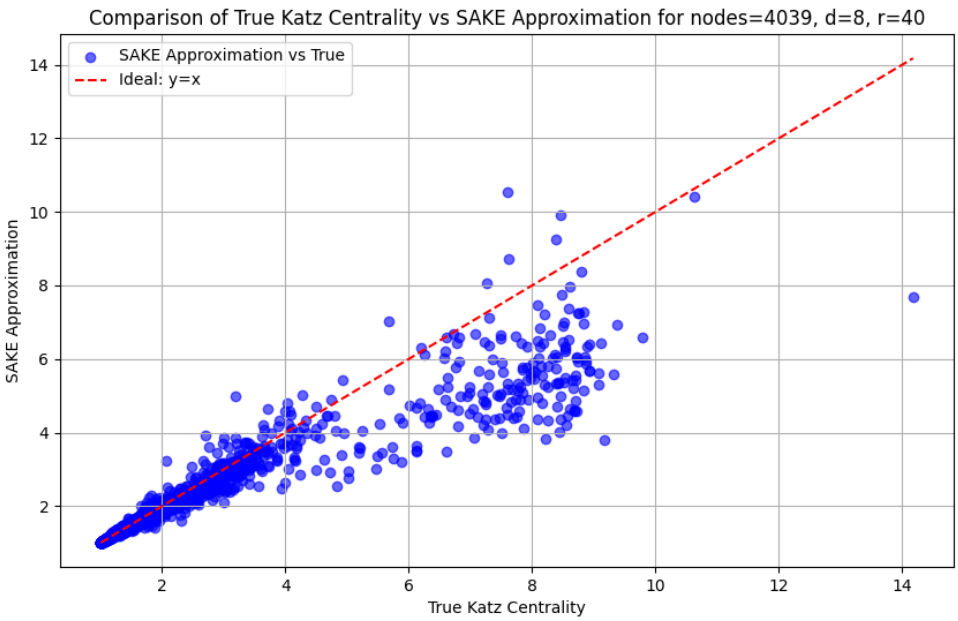
Notes: This figure plots the coefficients obtained from an individual-level regression of each of the listed predictors on the number of owner friends made after 2015. All explanatory variables other than female and college are standardized. Further analysis of the figure is given in Section III.B

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



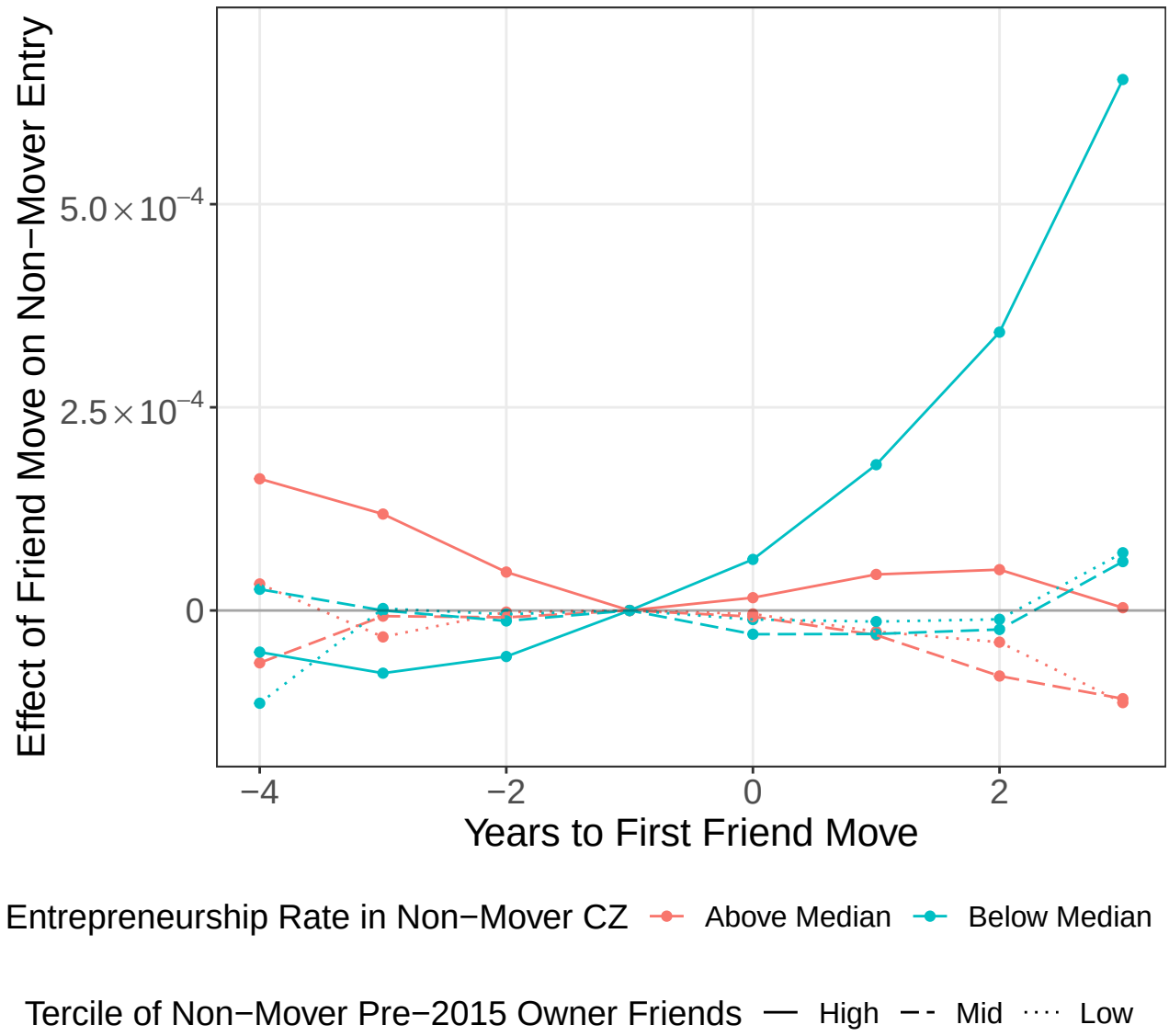
Notes: This figure repeats the analysis in Figure 8a but only considers movers who do not start a business. As with Figure E20, the same technical designs are used from Figure 8a to produce the results in this figure. Further analysis of the figure and comparisons to Figure 8a are given in Section III.B.

Figure E23: Performance of Katz Centrality Estimator in Public Data



Notes: This figure gives an individual-level scatterplot that validates our Katz-Bonacic estimator by comparing true Katz-Bonacic centrality with estimated centrality in a publicly availalbe subgraph of the Facebook network with 4039 nodes ((Leskovec and Sosi 2016)). We estimate truncated centrality up to a network distance of 8, and sample connections at distance greater than 2 with probability 0.01. We implement the sampling estimator described in (Lin et al. 2021). Details on the design of the model and estimator are given in IV.A.

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.