The Local Origins of Business Formation*

Emin Dinlersoz† Timothy Dunne‡ John Haltiwanger§ Veronika Penciakova¶

July 10, 2024

Abstract

What locations generate more business ideas, and where are ideas more likely to turn into businesses? Using comprehensive administrative data on business applications, we analyze the spatial disparity in the creation of business ideas and the formation of new employer startups from these ideas. Startups per capita exhibit enormous variation across granular units of geography. We decompose this variation into variation in ideas per capita and in their rate of transition to startups, and find that both components matter. Observable local demographic, household economic, and incumbent firm characteristics account for a significant fraction of the variation in startups per capita, and more so for the variation in ideas per capita than in transition rate. Income, education, age, and foreign-born share are generally strong positive correlates of both idea generation and transition. Overall, the relationship of local conditions with ideas differs from that with transition rate in magnitude, and sometimes, in sign: certain conditions (notably, the African-American share of the population) are positively associated with ideas, but negatively with transition rates. We also find a close correspondence between the actual rank of locations in terms of startups per capita and the predicted rank based only on observable local conditions – a result useful for characterizing locations with especially high and low startup activity.

*Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau or the Federal Reserve Bank of Atlanta. All results have been reviewed to ensure that no confidential information is disclosed (CBDRB-FY22-022, CBDRB-FY23-055 & CBDRB-FY24-0354). We thank Richard Beem, Shawn Klimek, Emmanuel Yimfor, participants at the Fall 2023 NBER Entrepreneurship Working Group, the 2021 and 2022 Southern Economic Association conference, and seminar participants at the University of North Carolina Kenan-Flagler Business School, and West Virginia University for helpful comments. John Haltiwanger was also a Schedule A employee of the U.S. Census Bureau at the time of the writing of this paper. We thank the Templeton Foundation for financial support.

†U.S. Census Bureau; emin.m.dinlersoz@census.gov

‡University of Notre Dame; tdunne1@nd.edu

§University of Maryland; halt@umd.edu

¶Federal Reserve Bank of Atlanta; veronika.penciakova@atl.frb.org

1 Introduction

Business startups contribute disproportionately to job creation, innovation, and productivity.[1] Yet, the nascent stages of entrepreneurship are not well understood. Characterizing environments conducive to early-stage business activity and entry is critical for assessing the spatial inequality in entrepreneurship across the United States, as the extent of this inequality has implications for economic vitality of locations and policies promoting entrepreneurship.[2] Moreover, to the extent local conditions that influence business entry evolve over time, changes in business dynamism may also be rooted, at least locally, in the evolution of these conditions.

A major impediment to progress in research on the early stages of entrepreneurship and entry has been the absence of systematic data on potential entrants, some of whom ultimately start new employer businesses. Without measures of the volume and types of potential entrants, it is impossible to assess precisely the underlying likelihood of success or entry rate of would-be entrepreneurs. As a result, we know little about what locations attract more potential entrants, and where potential entrants are more likely to start employer businesses. With only data on startups, we cannot ascertain whether spatial variation in startup rates is driven by a lack of business ideas, the difficulty of turning ideas into actual businesses, or both; information on the pool of potential entrants and the nature of their selection into employer startups is needed.

To provide new insights on nascent entrepreneurship, we use unique and comprehensive micro data from the U.S. Census Bureau that contain information on the universe of applications for new businesses in the U.S. and their transition to employer startups over the period 2011-2016. We conduct an in depth analysis of spatial variation in startup activity and nascent entrepreneurial activity. Specifically, we decompose the startup rate in a location into two components of nascent entrepreneurship—entrepreneurial idea generation and the transition rate of ideas to employer businesses. We show that both components help explain observed variation in startup activity across the United States, and document that observable local demographic, household economic, and incumbent firm characteristics help account for a significant fraction of variation in startups, business ideas, and transition rates. We also show that observable local conditions are useful for characterizing locations with both high and low startup activity.

To motivate our decomposition of startup activity and guide our empirical analysis, we

[1]See e.g., Haltiwanger, Jarmin and Miranda (2013), Kerr, Nanda and Rhodes-Kropf (2014), and Decker, Haltiwanger, Jarmin and Miranda (2014)

[2]Glaeser, Rosenthal and Strange (2010a) provides a review of the literature on the spatial dispersion of entrepreneurship and start-up activity.

introduce a simple model with a two-stage entry process where entrepreneurs first consider whether to make a business application based on the quality of their ideas and then decide whether to start an employer business after observing a signal of the value of the potential business. The model highlights the distinct roles of entrepreneurial idea generation and selection that underlie startup activity. We exploit a simple identity that employer startups per capita in a location can be expressed as the product of applications per capita and the transition rate of those applications to startups. Our model highlights that there may be distinct variation in these two components across locations, and guides us to explore how local conditions on a variety of dimensions either work in the same or opposite directions.

We conduct our empirical analysis using the micro data behind the Census Bureau's Business Formation Statistics (BFS) program. BFS integrates administrative data on the universe of applications for Employer Identification Numbers (EINs) from the IRS with the universe of employer businesses in the Longitudinal Business Database (LBD).[3] We measure startup activity in a location as the count of business applications filed in that location that transition into employer businesses within eight quarters of the application date. We proxy the intensity of idea generation in a location by the count of business applications filed there.[4] We calculate the transition rate as the fraction of business applications that turn into employer businesses within eight quarters of the application date. Because our focus is on employer startups, the analysis takes advantage of additional information on the EIN application to identify cases with a more clear intent to become new employer businesses. Specifically, we consider a subset of applications that indicate plans to pay wages as the primary group of applications at risk of transitioning to employer businesses (WBA).[5]

A striking feature of the data is that both variation in applications and transition rates contribute substantially to variation in startups at both the county and tract level.[6] Using publicly available data, Figure 1 shows that startup intensity varies considerably across counties, with counties in the top quintile having more than twice the rate of those in the bottom quintile. Figure 2 shows that when we decompose startups per capita into applica-

[3]For details on the development of the BFS see Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda and Stevens (2018).

[4]Business applications in our data reflect applications for new EINs. All new employer businesses must have an EIN. An application for a new EIN reflecting nascent entrepreneurship activity is consistent with the evidence from the Panel Study of Entrepreneurship Dynamics (PSED—see Reynolds, Carter, Gartner and Greene (2004)).

[5]For robustness, we also consider the entire set of applications (BA) even if the applicant did not indicate planned wages—there is a non-trivial fraction of transitions from applicants that do not explicitly signal an intent to pay wages.

[6]We show in appendix section A.5 that aggregate time series variation in employer startups from WBA is largely accounted for by variation in applications (consistent with evidence in Decker and Haltiwanger (2023)). We leave investigating the difference between aggregate time series variation and spatial variation for future research.

tions per capita and transition rates, the dispersion in each component is also high. Counties with high applications per capita have more than five times or more the application rate as low counties, while high transition counties have more than five times the transition rate as low counties.[7] Yet, counties with high startups per capita are not uniformly characterized by both high applications per capita and transition rates.

Turning to our micro data we confirm that spatial variation in startup activity at the county and tract levels is explained by both spatial variation in idea generation and transition rates. Using a variance decomposition of WBA startups, we find that about 68 percent of the spatial variation across counties in startups per capita is accounted for by variation in applications per capita, 38 percent by transition rates and about -5 percent due to the covariance.

At the tract level, we find much greater variation in startups per capita and its components (e.g., the coefficient of variation in startups per capita is ten times larger at the tract level than county level). Even with the greater variation at the tract level, we find similar fractions of the variation in startups per capita due to applications vs. transitions. About 66 percent of between tract variation in startups per capita is accounted for by applications per capita, 33 percent by transition rates and 2 percent due to the covariance.

These findings help motivate our parallel analysis of the spatial variation in startups and their components at both the county and tract level. By considering both tract and county level analyses, we shed light on the question of how "local" is local in terms of spatial variation in startup activity. In other words, it is an open question whether the spatial variation in startup activity is mostly between markets (we use commuting zones for this purpose), between counties within markets, or between tracts within the counties. Our evidence is there is substantial between and within variation on all of these dimensions. Specifically, we find that about 40 percent of between county variation in startups per capita is accounted for by commuting zone (by year) effects. While that is substantial, most of the variation is within commuting zones. Even more dramatically, only about 11 percent of between tract variation in startups per capita is accounted for by county (by year) effects. Given the enormous between tract variation, the dominant role of within county variation highlights that highly-localized factors that vary across tracts within counties play a critical role in spatial variation in startup activity.

We quantify the fraction of the within variation at both the county and tract level that

[7] The BDS provides annual employer business startups rather than the targeted 8 quarters ahead startups from the BFS. To overcome this limitation we use applications from 2011-16 from the BFS and employer startups from 2013-18 from the BDS. The use of BA rather than WBA also implies some caution in interpretation. We depict these spatial patterns with public domain data to illustrate the patterns on county based maps. Disclosure restrictions prohibit our release of county-level tabulations from our micro data.

is accounted for by local conditions. The motivation for doing this at both the county and tract level is to identify what observable factors help account for the distinct variation at both levels of spatial aggregation. In this parallel analysis, we consider the role of lagged demographic and household economic conditions, as well as incumbent firm characteristics. We also control for commuting zone by year effects to focus on the role of local factors in accounting for the within commuting zone variation. Similarly, at the tract level we control for county by year effects to focus on the role of local factors in accounting for within county variation. All of our analysis also controls for the local industry composition of incumbent businesses. While our analysis does not permit a causal interpretation, our analysis of variance can help facilitate research on specific factors that are associated with different components of early entrepreneurial activity.

At both the county and tract level, we find that observable local conditions account for a larger fraction of spatial variation in startups per capita and applications per capita than transition rates—even for transition rates, we find a number of systematic relationships. Some covariates work in the same direction for applications per capita and transition rates, and hence, startups per capita. This holds for example for the fraction of the local population that has a bachelor's degree (Bachelors+ share). However, for some covariates these factors work in opposite directions. The most dramatic and interesting result along these lines reflects the variation accounted for by the local population that is Black or African American (African American share). Startups per capita are negatively related to the African American share. This pattern reflects offsetting effects of a positive relationship between applications per capita and the African American share and a negative relationship for transitions.

The contribution of local conditions are broadly similar at the county and tract level – including for the two key covariates Bachelors+ share and African American share discussed above. However, there are some notable differences. For example, the fraction of the local population that is Asian has a strongly negative effect on startups per capita and its components at the county level, but a modest positive effect on startups per capita and their components at the tract level.

The micro data also enables us to examine whether there are distinct patterns in the contribution of local conditions across different industries. We know there are differences in objectives of entrepreneurs that vary across industries that are reflected in dramatic differences in the post entry outcomes of startups across industries. For example, high growth young businesses are much more prevalent in the innovative intensive industries (see, e.g., Decker, Haltiwanger, Jarmin and Miranda (2016)). We take a step towards exploring this type of variation by conducting a sub-sample analysis of startups in the innovative intensive sectors (identified by the STEM intensity of workers in the industry)—for shorthand we denote these as the *high tech* industries. We find that qualitative patterns are broadly similar for the high tech startups compared to the all startups sample. However, there are some important quantitative differences. For example, the contribution of Bachelors+ share is much more important for startup activity in the high tech industries – especially in terms of its impact on applications per capita. This increased impact is especially notable using the county level variation.

We use our estimates to conduct a ranking analysis exploring the question of how well observable factors account for the ranking of local areas (counties and tracts) in terms of startups per capita. We find that locations in the top deciles of startups per capita have especially high idea generation (i.e., application intensity) while those in the bottom deciles have especially low transition rates of applications. We find that even though observable local conditions typically account for less than half of the between county and tract variation, they predict a ranking that closely corresponds to the actual ranking. While this finding does not identify causal mechanisms, it highlights the importance of local conditions for differences in entrepreneurship across local areas and is useful in characterizing locations with high startup activity. We also find that spatial variation in startup activity and the covariates underlying this variation are positively related to the social/economic mobility indices of Chetty, Hendren, Kline and Saez (2014). This positive relationship holds especially for transition rates—locations with low social/economic mobility are also locations with low transitions of business applications to new employer businesses.

The paper proceeds as follows. Section 2 provides a review of the literature. Section 3 presents a model to motivate the empirical approach, which is described in section 4. Section 5 provides a description of the data. The decomposition of the variance in startups per capita into its components is presented in section 6. The relationship between these components and local conditions is presented in section 7. Section 8 concludes.

2 Review of literature

Our work is influenced by models of firm entry, selection, and growth. In canonical models of entry, selection, and growth (e.g., Lucas (1978); Hopenhayn (1992)), entrants pay a sunk cost of entry, learn their productivity draw, and then face a profit function with curvature and a fixed cost of operation. Firms with high productivity draws become large, those with low draws stay small, and those with sufficiently low draws exit because of their inability to cover fixed costs. Over time, the literature has introduced additional features and frictions that generate interesting entry and post-entry dynamics. Among them include dynamic learning (Jovanovic, 1982; Ericson and Pakes, 1995), financial frictions (Evans and Jovanovic, 1989; Cagetti and De Nardi, 2006), human capital (Polkovnichenko, 2003; Poschke, 2013), investment risk (Vereshchagina and Hopenhayn, 2009; Bianchi and Bobba, 2013; Choi, 2017), among others (Hurst and Pugsley, 2017; Hombert, Schoar, Sraer and Thesmar, 2020; Vardishvili, 2023). Additionally, models of entry with imperfect competition (Nocke, 2006; Asplund and Nocke, 2006) allow for local competition and market size to effect entry and selection. The insights from the theoretical literature helps guide our choice of the local conditions we consider in our empirical analysis.

In general, most models (at least implicitly) still assume that idea creation and business entry occur at the same time. The empirical literature has more explicitly studied the nascent *phases* of entrepreneurship. One strand of that literature studies survey data. The Panel Study of Entrepreneurship Dynamics (PSED) is one such effort (Reynolds et al., 2004; Reynolds, 2017).[8] Reynolds et al. (2004) describes entrepreneurs spending time and resources in the "conceptual" and "gestational" period prior to the actual commencement of business operations. For our purposes, two key findings from the PSED are of interest. First, nascent entrepreneurs identified the application for an EIN as a critical activity. Second, most individuals engaged in nascent entrepreneurship do not transition to employer businesses. Note, however, that this second finding is influenced by attrition from the PSED sample.

More recently, Bennett and Chatterji (2023) and Bennett and Robinson (2023) conduct a novel survey about nascent entrepreneurship. This survey is internet-based and voluntary but re-weighted to be nationally representative in terms of key demographic characteristics. Bennett and Robinson (2023) find from a sample of about 50 thousand respondents that about 30 percent have considered opening a business. Of those that have considered opening a business, almost one third have taken an important step towards starting a new business—only about one fifth of these include hiring an employee. Especially relevant for our analysis is that these survey-based papers highlight the number of steps involved in the nascent entrepreneurship phase (including the administrative step of applying for a taxpayer ID).

While the BFS does not contain the rich individual data available in a survey, a distinguishing feature of our analysis is the focus on spatial variation in startup activity, idea creation, and transitions which our comprehensive administrative data enables. Specifically, our analysis has three key advantages to the survey data. First, it tracks the universe of applications for new EINs. Second, it contains detailed application information, including industry, location, legal form, and motivation for the application. Importantly, these application characteristics also provide useful proxies for the quality and viability of the underlying

[8]The PSED identified about 5,000 nascent entrepreneurs, defined as individuals who have taken steps within the last 12 months toward creating a venture but have not yet paid employees for more than 3 months. PSED Wave I identified 4,000 and Wave II identified 1,500 individuals that satisfied this criteria.

business idea. Third, by linking the BFS to the universe of employer businesses (LBD), we accurately track the incidence and timing of the transition of ideas to startups.

A second strand of the empirical literature use the state business registries to study the quantity and quality of entrepreneurship (Andrews, Fazio, Liu, Guzman and Stern, 2018; Guzman and Stern, 2020). Guzman and Stern (2020) use high impact outcomes (IPO or a high-profile merger) originating from state business registrations in 32 states between 1988 and 2014 to assess the quality of entrepreneurship, and model these outcomes as a function of a set of registration characteristics. Using data from 8 states and leveraging the surge in business applications during the Covid pandemic, Fazio, Guzman, Liu and Stern (2021) examine the relationship between the growth rate of state business registrations and local conditions across Zipcode Tabulation Areas (ZCTAs).

Our relative contribution to this strand of the empirical literature is multi-fold. First, we focus on the universe of employer startups in the US, and decompose it into business ideas (applications) and transitions. Second, this decomposition allows us to assess the contribution of local conditions on startups into the contribution of local conditions on idea creation and transitions. We are sympathetic to the interest in high-impact outcomes for new businesses. Yet, even for such cases a first critical step is transitioning to an employer business. Moreover, startups are an important source of job creation and economic mobility (i.e, hiring) for local areas.[9]

Finally, our research also contributes to the literature on spatial aspects of entrepreneurship. Glaeser et al. (2010a) provide an overview of the how entrepreneurship has been examined in the urban economics literature. One main line of the literature assesses the impact of entrepreneurship on urban success. A second avenue pursues examinations of the relationship between local characteristics and entrepreneurial activity, attempting to shed light on what factors explain differences in the local supply of entrepreneurs. Research in this line include work by Doms, Lewis and Robb (2010) that examine human capital and entrepreneurial activity and Kerr and Kerr (2020) that looks at the role of immigrants and entrepreneurship in the United States. Rosenthal and Strange (2003) focus on the industrial organization of the local environment and its impact on firm births. This paper adds to the second literature by examining the association between local characteristics and entrepreneurial activity, distinguishing between the business idea generating process and the transitioning of ideas to employer businesses.

[9]We note that the LBD and associated integrated data (e.g., COMPUSTAT) does permit examining the role of local conditions for high-impact businesses – in fact, for a large number of business outcomes, such as startup size, growth rate, or failure rate. We leave that for future work.

3 A model of business ideas and startups

The model highlights pre-entry heterogeneity among potential entrants in the quality of latent business ideas, and explores how the decision to pursue an idea and the idea's transition to an actual startup are related to local conditions potential entrepreneurs face. Startup formation involves two distinct decisions. The first decision is whether to pursue an idea and explore its feasibility further. Additional information about the viability of an idea is revealed in the gestational state during which the potential entrant pursues the idea and takes steps to potentially implement it. Based on the information revealed, the second decision is whether to start an employer business. The model's analysis indicates that local conditions can play distinct roles in an entrepreneur's decision to pursue an idea versus the ultimate decision to start a business.

Consider an economy where economic activity takes place in a large number of locations denoted by the set, $\mathcal{L}$. In each location $l \in \mathcal{L}$ there is a continuum of N_l individuals, each of whom has an idea, $\iota \in [0, \infty)$, for an employer business.[10] Higher values of ι indicate better (or higher quality) ideas in terms of expected return to an employer business, in a sense made more precise below. The distribution of ideas is given by the c.d.f. F_l (with density f_l).

An idea owner has to make an initial investment, $I_l > 0$, to pursue the idea. This investment pertains to various tasks of planning for the potential employer business, including major tasks, such as estimating demand and costs, seeking and securing financing, understanding relevant regulations, socializing the idea and obtaining advice, searching for potential employees and suppliers, as well as more mundane, but necessary tasks, such as applying for an EIN, and obtaining necessary permits and licenses. Many of these tasks involve frictions, such as financial or labor-market related, and can take time and resources to accomplish. Some ideas may thus take longer to transition to an employer business, or not transition at all, due to various local frictions involved.

Over the course of the investment, the idea owner sojourns in a state during which he observes a random signal, $V \in \mathbb{R}$, of the net value of an employer business. V can be thought of as an estimate of the idea owner's net payoff from the business—a scalar index of payoff-relevant factors for his business in location l, such as demand, entry costs (including any startup funding), various fixed and variable costs, the degree of competition, and his own productivity or ability.[11]

[10]We assume that all ideas come from the local population and are aimed for potential businesses in the same location. While business ideas can be aimed at locations other than the entrepreneurs' own locations, empirically we find that the addresses of planned businesses substantially overlap with the addresses of businesses that actually form. See the empirical analysis and the discussion below further on this issue.

[11]The index V can thus include an initial signal of unknown cost or productivity parameter as in Jovanovic (1982) and the idea owner may continue to learn about that parameter after the business starts. Alternatively,

The value of V depends on the quality of the idea, summarized by the conditional distribution $G_l(V|\iota)$. For example, a better idea (higher ι) may correspond to a higher V in a first order stochastic dominance sense. Each idea owner can choose to not pursue the idea and obtain a return of $R_l > 0$ – e.g. income from salary work or a nonemployer business. Importantly, R_l can reflect various labor market frictions. For simplicity, we assume R_l and I_l do not depend on ι, though such dependence is plausible – for instance, individuals with a higher ι may earn higher wages. Higher investment can also improve the idea quality or change G_l, in the spirit of Ericson and Pakes (1995). While these considerations can be incorporated, we proceed with a simpler setup to illustrate our key points.

Each idea owner is "small" with respect to the local economy, and takes as given the local environment $\mathcal{E}_l = \{N_l,\ F_l(\cdot),\ G_l(\cdot|\cdot),\ R_l,\ I_l\}$. We do not study the determination of $\mathcal{E}_l$. In other words, in the background there is a set of spatial equilibrium conditions that ensure businesses and individuals optimize and have no incentive to move across locations, free entry holds in each location, and all markets clear. These equilibrium conditions pin down $\mathcal{E}_l$ for all $l \in \mathcal{L}$. Our focus is on the determination of what ideas are pursued and which ones turn into employer businesses in each location in a spatial equilibrium that generates $\mathcal{E}_l$.[12]

After observing V, an idea owner decides whether to start an employer business. The expected return from pursuing an idea is

$$\mathcal{V}_l(\iota) = E[\max\{V, R_l\}|\iota] = (1 - p_l(\iota))R_l + p_l(\iota)E[V|V \geq R_l; \iota] \tag{1}$$

where $p_l(\iota)$ is the probability that the pursued idea transitions to an employer business

$$p_l(\iota) = P(V \geq R_l|\iota) = 1 - G_l(R_l|\iota). \tag{2}$$

An idea owner will pursue the idea (e.g., make an EIN application, among other steps) if $\mathcal{V}_l(\iota) \geq R_l + I_l$. Assume now that $\mathcal{V}_l(\iota)$ is increasing in ι—better ideas lead to higher expected potential return.[13] As long as $R_l + I_l < \mathcal{V}_l(\iota)$ for some ι, there exists a threshold (marginal

V can depend on a known distribution of productivity as in Hopenhayn (1992), but one that depends on ι—reflecting pre-entry heterogeneity among potential entrants, in contrast to the identical potential entrants in Hopenhayn (1992). Chen, Croson, Elfenbeinand and Posen (2018) also explore learning dynamics in the pre-entry nascent entrepreneurship phase. A distinguishing feature of our model is that we specify that there are explicit costs associated with this pre-entry learning process. We think this is important in accounting for the spatial variation we observe in both idea creation and transitions.

[12] In particular, we do not explicitly study the selection of individuals into a location, which determines the distribution of ideas, F_l. For instance, entrepreneurs may sort into locations based on their ideas and ability as in Nocke (2006). This sorting, however, may not be perfect since amenities and mobility frictions also factor in the determination of F_l.

[13] This is the case, for instance, if higher ι implies higher V in a first order stochastic sense, i.e. $G_l(\cdot|\iota)$ is

idea) $\iota_l^* \in [0, \infty)$, such that all ideas in $[\iota_l^*, \infty]$ are pursued. The marginal idea satisfies

$$\mathcal{V}_l(\iota_l^*) = R_l + I_l, \tag{3}$$

and the mass of pursued ideas (business applications) per capita is

$$A_l = \frac{N_l \int_{\iota_l^*}^{\infty} f_l(\iota) d\iota}{N_l} = \int_{\iota_l^*}^{\infty} f_l(\iota) d\iota = 1 - F_l(\iota_l^*), \tag{4}$$

If $R_l + I_l > \mathcal{V}_l(\iota)$ for all ι, no idea is pursued ($A_l = 0$).

Startups per capita originating from applications is then

$$S_l = \frac{N_l \int_{i_l^*}^{\infty} p_l(\iota) f_l(\iota) d\iota}{N_l} = \int_{i_l^*}^{\infty} p_l(\iota) f_l(\iota) d\iota. \tag{5}$$

There are no startups ($S_l = 0$) if no idea is pursued ($A_l = 0$).

When $A_l > 0$, the (average) transition rate for applications is

$$T_l = \frac{S_l}{A_l} = \int_{\iota_l^*}^{\infty} p_l(\iota) f_l^*(\iota) d\iota = E[p_l(\iota) | \iota \geq \iota_l^*], \tag{6}$$

where $f_l^*(\iota) = \frac{f_l(\iota)}{1 - F_l(\iota_l^*)} = \frac{f_l(\iota)}{A_l}$ is the density of ideas conditional on application. Note that T_l is undefined when $A_l = 0$. By construction, the following holds

$$S_l = \begin{cases} A_l T_l & \text{if } A_l > 0, \\ 0 & \text{if } A_l = 0. \end{cases} \tag{7}$$

Expressions (1)-(7) hold in any spatial equilibrium. Now, we make explicit the fact that in equilibrium, each element of the local environment $\mathcal{E}_l$ will in general be a function of C, the collection of all relevant local characteristics or conditions, C_l, and conditions C_k in other locations $k \in \mathcal{L}$, $k \neq l$. These conditions can pertain to demographics, demand and costs, agglomeration, amenities, industrial composition, labor markets, laws and regulations, etc.[14] The individual elements of $\mathcal{E}_l$ need not depend on all elements of C. For instance, the distribution of ideas, F_l, may only depend on certain demographic characteristics of the population (e.g., age and education) in location l, but the initial investment, I_l, may, in addition, depend on financial frictions. The key variables of interest, A_l, T_l, and S_l, are then

decreasing in ι.

[14]Some of these conditions can be determined in spatial equilibrium (e.g., wages and rents), and some others (e.g., natural amenities) can be exogenous.

functions of the entire set of characteristics, C, because they depend on $\mathcal{E}_l$ by the definitions (3)–(6).[15]

Now consider the change in A_l, S_l, and T_l as a local characteristic or condition $c_l \in C_l$ changes from one location to another. Assume that A_l, T_l, and S_l are differentiable functions. Then, for values of c_l for which $A_l > 0$, we have

$$\frac{dA_l}{dc_l} = -f_l(\iota_l^*)\frac{d\iota_l^*}{dc_l} - \int_0^{\iota_l^*} \frac{df_l(\iota)}{dc_l} d\iota, \tag{8}$$

$$\frac{dS_l}{dc_l} = -p_l(\iota_l^*) f_l(\iota_l^*)\frac{d\iota_l^*}{dc_l} + \int_{\iota_l^*}^{\infty} \left[\frac{dp_l(\iota)}{dc_l} f_l(\iota) + p_l(\iota)\frac{df_l(\iota)}{dc_l}\right] d\iota, \tag{9}$$

$$\frac{dT_l}{dc_l} = -p_l(\iota_l^*) f_l^*(\iota_l^*)\frac{d\iota_l^*}{dc_l} + \int_{\iota_l^*}^{\infty} \left[\frac{dp_l(\iota)}{dc_l} f_l^*(\iota) + p_l(\iota)\frac{df_l^*(\iota)}{dc_l}\right] d\iota. \tag{10}$$

Note that, for values of c_l for which $A_l = 0$, $\frac{dT_l}{dc_l}$ is undefined.[16]

Based on (8), A_l changes as c_l changes because not only the marginal idea ι_l^* changes, but also the distribution of ideas shifts. Similarly, (9) and (10) indicate that S_l and T_l change because the marginal idea, the distribution of ideas, and their transition rates all change. Observe that the sign of $\frac{dA_l}{dc_l}$ is in general unrestricted, and depends on the signs of $\frac{d\iota_l^*}{dc_l}$, and $\frac{df_l(\iota)}{dc_l}$.[17] Similarly, $\frac{dT_l}{dc_l}$ can be positive or negative, and its sign can differ from that of $\frac{dA_l}{dc_l}$. Note that by (6)

$$\frac{dT_l}{dc_l} = \frac{1}{A_l}\left(\frac{dS_l}{dc_l} - T_l\frac{dA_l}{dc_l}\right).$$

Thus, $\frac{dT_l}{dc_l}$ is negatively related to $\frac{dA_l}{dc_l}$, holding S_l constant. But $\frac{dS_l}{dc_l}$ can be positive or negative, and hence, the sign of $\frac{dT_l}{dc_l}$ is not the same as the sign of $\frac{dA_l}{dc_l}$ in general. From (8) and (9), the signs and magnitudes of $\frac{dA_l}{dc_l}$ and $\frac{dS_l}{dc_l}$ depend on how the idea distribution and the marginal idea change as c_l changes. However, the sign and magnitude of $\frac{dS_l}{dc_l}$ depends, in addition, on how the transition probabilities, $p_l(\iota)$, change. Depending on the nature of this change, the

[15] A_l is a function of F_l and ι_l^*, which depends, through (3), on I_l, R_l, and G_l – which itself may be a function of local population (local market size), N_l, that can matter for expected post-entry profit (see, e.g., Nocke (2006) and Asplund and Nocke (2006)). Similarly, S_l is a function of F_l, $p_l(\iota)$, and ι_l^*, which depend on I_l, R_l, and G_l. Finally, T_l is a function of S_l and A_l, and hence, a function of all conditions that the latter two depend on.

[16] The system of partial derivatives (8, 9, 10) should be interpreted as comparative statics with respect to a local characteristic c_l in a cross section of locations in spatial equilibrium. That is, given an equilibrium, we are interested in the change in the key variables attributable to a change in c_l from one location to another.

[17] Using (3), $\frac{d\iota_l^*}{dc_l} = \left[\frac{dI_l}{dc_l} + \left.\frac{dp_l}{dc_l}\right|_{\iota=\iota_l^*}(R_l - E_l) + p_l(\iota_l^*)\left(\frac{dR_l}{dc_l} - \frac{dE_l}{dc_l}\right)\right]\left(\frac{d\mathcal{V}_l}{d\iota_l}\right)^{-1}$, which depends on the rates of change in I_l, R_l, and G_l – the latter through the changes in p_l and $E_l = E[V|V \geq R_l; \iota_l^*]$. While $\frac{d\mathcal{V}_l}{d\iota} > 0$ by assumption, and $R_l < E_l$ by the definition of E_l, the rest of the terms cannot be signed without further restrictions. Similarly, $\frac{df_l(\iota)}{dc_l}$ can be positive or negative.

sign of $\frac{dT_l}{dc_l}$ can be the same as, or different from, the sign of $\frac{dA_l}{dc_l}$.

As an example, consider how A_l and T_l may vary as a local characteristic, such as the average level of educational attainment, changes. Suppose that a higher education is associated with higher idea quality: the idea distribution F_l associated with a higher education level dominates, in a first-order stochastic sense, the distribution associated with a lower education level. All else equal, (4) then implies A_l is higher. Furthermore, (5) implies that S_l also higher, as long as $p_l(\iota)$ is an increasing function – i.e. better ideas transition with higher likelihood. However, a higher education level may also "shift" the transition probabilities up, resulting in a higher $p_l(\iota)$, for every ι – for instance, higher level of education may result in a stochastically higher signal, V, equivalent to a lower value of $G_l(R_l|\iota)$ in (2). Then, (5) implies a higher S_l, but there would be no similar effect on A_l. Finally, the marginal idea, ι_l^*, also changes, since it is a function of G_l. Depending on the signs and magnitudes of these various effects, $\frac{dA_l}{dc_l}$ and $\frac{dT_l}{dc_l}$ may have different signs and magnitudes.

This model is best suited for providing guidance regarding the creation of ideas for potential employer businesses and in turn the factors that influence the transition of such ideas to actual startups. In our empirical analysis, we have applications that have indicated an intent to pay wages at some point, and we observe transitions of such applications into actual startups. This aspect of our empirical work more tightly connects to this model.

4 Empirical approach

The model indicates that business applications (ideas) per capita (A_l), startups per capita (S_l), and transition rate (T_l) are functions of the set, C_l, of relevant local conditions, and in general, conditions in other locations, $k \neq l$. We first examine the variation in S_l over $\mathcal{L}$, the set of all locations at a given level of granularity—we use counties, and tracts, alternatively.

4.1 Variance Decomposition

For the set of locations where $A_l > 0$, (7) implies

$$\log S_l = \log A_l + \log T_l, \tag{11}$$

and the variation in S_l can then be decomposed as follows[18]

$$Var(\log S_l) = Var(\log A_l) + Var(\log T_l) + 2Cov(\log A_l, \log T_l). \tag{12}$$

[18]With a slight abuse of notation, we use the model variables to also refer to their observed counterparts, which are treated as random variables.

Note that the covariance term can, a priori, be positive or negative, depending on how A_l and T_l change across locations. The analysis of the rates of change for A_l and T_l in (8) and (10) indicates that A_l and T_l do not have to move in the same direction as local conditions change from one location to another. Another way to see this is to write the covariance term above in terms of the covariance between S_l and A_l

$$Cov(\log A_l, \log T_l) = Cov(\log A_l, \log S_l - \log A_l) = Cov(\log A_l, \log S_l) - Var(\log A_l).^{19} \tag{13}$$

Now consider the relationship between S_l and A_l in the model. If the distribution F_l and the transition probabilities p_l were the same across locations, A_l and S_l would move in the same direction as idea selection (marginal idea) changes across locations – in this case, a higher A_l would mean higher S_l.[20] However, the variation in F_l and p_l across locations can alter this positive relation. For instance, if transition probabilities are lower or the share of ideas with low transition probability is higher in areas where there are more applications per capita, the positive association between A_l and S_l will be weaker, and can turn negative. If A_l and S_l are negatively correlated, or their covariance is positive but small relative to the variance of A_l, then $Cov(\log A_l, \log T_l)$ will be negative.

Based on (12), we decompose the variation in S_l into the variation in A_l, the variation in T_l, and their covariance. For states and commuting zones, $A_l > 0$ holds for all $l \in \mathcal{L}$, and the decomposition can be carried out for the entire set $\mathcal{L}$. For counties and tracts, we condition on locations for which $A_l > 0$ when implementing this decomposition. This restriction is not severe but still there are counties or tracts that are excluded. In our main empirical analysis of locations considered below, we use a transformation that incorporates zero applications in a location.

4.2 Location-level analysis

We examine the relationship between the three key variables S_l, A_l, and T_l, and local conditions, C_l in a regression framework using panel data across locations $l \in \mathcal{L}$ over years $t = 1, ..., T$. Specifically, we posit the relationships

$$\widetilde{S}_{lzt} = \beta^{S'} C_{lt-k} + f_{zt} + \epsilon^S_{lzt}, \tag{14}$$

$$\widetilde{A}_{lzt} = \beta^{A'} C_{lt-k} + f_{zt} + \epsilon^A_{lzt}, \tag{15}$$

[19] This derivation follows from that fact that $Cov(\log A_l, \log S_l - \log A_l) = E[\log A_l \log S_l - (\log A_l)^2] - E[\log A_l]E[\log S_l - \log A_l] = E[\log A_l \log S_l] - E[(\log A_l)^2] - E[\log A_l]E[\log S_l] + E^2[\log A_l]$.

[20] The model implies $\frac{dA_l}{d\iota_l^*} = -f(\iota_l^*) < 0$, and $\frac{dS_l}{d\iota_l^*} = -p_l(\iota_l^*)f(\iota_l^*) < 0$.

$$T_{lzt} = \beta^{T'} C_{lt-k} + f_{zt} + \epsilon^{T}_{ltz}, \tag{16}$$

where C_{lt-k} is a vector of lagged local characteristics measured in year $t-k$, f_{zt} is area-year fixed effect for a broader geographic area z that contains location l and other economically connected locations, and ϵ^{i}_{lzt} $(i = S, A, T)$ is an error term.

While the identity $S = AT$ suggests a log-linear specification (11), it only holds when $A > 0$ – see (7). At the granular levels of geography we work with (e.g., counties or tracts) there are some location-year observations with no business applications or startups, and the log transformation would not allow us to incorporate these cases in estimation. Instead, the dependent variables $\widetilde{S}$ and $\widetilde{A}$ in (14) and (15) are the Davis, Haltiwanger and Schuh (1996) (DHS) transformations of the variables S and A that represent them in terms of deviations from their grand mean, respectively. Specifically, the transformation we use is $\widetilde{Y} = 2\frac{(Y-\bar{Y})}{(Y+\bar{Y})}$, where $\bar{Y}$ is the grand mean of $Y = S, A$ in the panel used for estimation.[21] This transformation is a second order approximation of the log difference between Y and the grand mean of Y. As we show in the appendix, the implied elasticities of Y with respect to covariates X are easily derived – indeed if X is a log based measure then the elasticity is the estimated coefficient.

We leave the transition rate, $T = S/A$, untransformed in (16), since, by definition, $T \in [0, 1]$. Because T is defined only for $A > 0$, the cases with no applications are not used in the estimation of (16). Hence, the regression (16) is conditional on $A_{lzt} > 0$. Similarly, we relate the DHS-transformed average duration (measured in quarters from the application quarter to transition quarter) to local characteristics, where D_{lzt} is defined only for cases where $T_{lzt} > 0$. This analysis is informative on the duration of the gestational period before a business starts. The duration can be higher because of various frictions, such as those in securing startup funding or hiring an employee, or because of idea complexity and quality, which may lengthen time to implement an idea.

The equations above are estimated separately for two different granular geographic units ($\mathcal{L}$)—counties and tracts. The focus on narrowly defined geographic units reflect our desire to better measure the characteristics of the underlying population and local conditions. At broader levels of geography (e.g., states and commuting zones), the rich variation in conditions, such as demographics, within a geographic unit is substantially masked. When a geographic area is instead very narrowly defined, local conditions can become less informative, as they will depend on conditions in nearby locations (e.g., the existence of nearby

[21]This type of transformation was recommended by Tornqvist, Vartia and Vartia (1985) and also implemented by Davis et al. (1996) for employment growth rates at the establishment-level. We note that this transformation is scale-free avoiding the pitfalls for transformations such as the inverse hyperbolic sine described in Chen and Roth (2023)

competitors or labor market conditions in neighboring locations). While our choice of counties and tracts allow more granular measurement of key local characteristics, we also account for common factors operating at a broader geographic level, by controlling for broader area effects.[22]

In the case of counties, we use a commuting zone containing the county as the broader geographic area (z). The fixed effects f_{lzt} control for time-varying factors that may operate at the commuting zone level, such as agglomeration economies, productivity spillovers, labor market conditions, regulations, etc.—in other words, this specification implicitly assumes that the relationship between the conditions in other locations ($k \neq l$) and the dependent variables works mainly through the overall conditions in the commuting zone that the location l belongs to. Thus, the county-level results will reflect within commuting zone by year variation. In the case of tracts, we use county fixed effects to control for factors at work at the county-level (thus also at broader geographic aggregations of counties, such as commuting zones). The tract-level analysis therefore captures within county by year variation. Furthermore, we check the robustness of our findings by including in C_{lt-k} the characteristics and conditions of the adjacent tracts for each tract, thereby controlling for localized factors operating through neighboring tracts.

The parameters β^S, β^A, and β^T are estimated using panels for locations that span $t = 2011, ..., 2016$. We use k-year lagged values of local characteristics (k varies by c_l) to hedge against the potential simultaneity of the characteristics. The set of local conditions we consider also differs slightly for the county-level and tract-level estimation, as we discuss below.

The estimates of these specifications cannot be interpreted as identifying causal mechanisms. The problem is not reverse causality, especially given that we use lagged local conditions and we have controlled for broad area fixed effects. A host of factors may underlie why a county, after controlling for commuting zone (or a tract, after controlling for county) has idiosyncratic variation in idea creation, transitions and its local demographic, economic and business conditions. These factors may have long run antecedents which offers the potential for identification of those factors. Our objective is a first step: to quantify the extent of variation and in turn the observable covariates that account for this variation.

[22]The approach is similar to Rosenthal and Strange (2003) who model firm births in zip codes controlling for metropolitan area fixed effects. Rosenthal and Strange (2003) emphasize that agglomeration effects may be very local. Our approach largely abstracts from agglomeration effects by controlling for commuting zone by year fixed effects in our county analysis and county by year fixed effects in our tract level analysis. The local nature of agglomeration effects may be captured by our measures of business conditions in the local area (e.g., share of employment from young firms).

4.3 Application-level analysis

Previewing our empirical findings, we find enormous variation in transition rates across counties and tracts, and have only modest success in accounting for that variation with indicators of local conditions. To dig deeper, we also conduct an application-level analysis of transitions. We observe a set of characteristics, x, of a business application, which provide information on the type of business applied for, and may also be informative about the quality of the underlying idea, ι. Using these characteristics, we examine the relationship between the probability of transition $p_l(\iota)$ in (2) and local characteristics C_l in a linear probability (LPM) framework

$$p_{ilzt} = \gamma' x_{ilzt} + \beta^{p'} C_{lt-k} + f_{zt} + \epsilon^p_{ilzt}, \tag{17}$$

where $p_{ilzt} \in \{0, 1\}$ is an indicator of whether application i transitions.[23]

The estimates of β^p are informative about the relationship between the transition rate of an idea and local characteristics, conditional on application characteristics (proxies for the quality of the idea). Because locations differ in the distribution of ideas F_l, it is important to control for the composition of applications in assessing the relationship between transition rate and local characteristics. The LPM in (17) does so by directly incorporating the individual application characteristics. We also compare the estimate of β^p with that of β^T based on (16) to see whether controlling for application characteristics alters the general nature of the relationship between transition rate and local characteristics.

5 Data

5.1 Business applications

We use the administrative micro data underlying the Business Formation Statistics (BFS). The BFS provides high-frequency statistics on entrepreneurial activity based on applications for employer identification numbers (EINs).[24] The Census Bureau receives applications for all new EINs on a weekly basis from the IRS. From the universe of EIN filings, the BFS program constructs a subset of EINs that restrict applications to those are associated primarily with new business formation, as opposed to applications associated with other reasons, such as applications for trusts, estates, and other financial filings.[25] The restrictions are based on

[23] As in (16), the regression above is conditional on $A_l > 0$.

[24] For more information on the publicly available data, visit the BFS website.

[25] In our analysis of the micro data, we also exclude applications for purchasing or a change of ownership type for existing businesses—to avoid using applications from potentially existing employer businesses.

information on the EIN application including reason for applying and type of entity.[26] The details of the micro data are provided in Bayard et al. (2018).

Importantly for our analysis, all employer businesses in the United States are required to have an EIN to file payroll taxes. All new businesses (employer or nonemployer) also file for an EIN if forming a partnership or a corporation. There are some potential business formations that we are not tracking in applications and startups. Specifically, sole proprietor nonemployers do not need to have an EIN, though some choose to obtain one. As discussed in Davis, Haltiwanger, Krizan, Jarmin, Miranda, Nucci and Sandusky (2009), nonemployers with an EIN (including sole proprietors) are about three times as large in terms of revenue than those without an EIN. Many small (in terms of revenue) nonemployer sole proprietors also have other activity (e.g., their main activity as a wage and salary worker). While there are many sole proprietor nonemployers without EINs, they account for a small fraction of aggregate economic activity. Thus, the application micro data we rely on offers nearly full coverage of all economically significant business initiations.

The application form includes the name and address of the applicant and business, application week, business start date, reason for application (hiring, banking, etc.), type of business entity, previous application for an EIN, principal industry, and planned date of initial wage payments—these are potential proxies for the underlying idea quality (ι), which we use as our application-level controls (x) in the estimation of (17). We are especially interested in employer startups. This focus motivates our analysis of applications that indicate a planned date for initial wage payments. Following the naming conventions of the public domain BFS, we refer to all business applications as BA and applications with planned wages as WBA.[27]

A business application includes a mailing address and potentially a business address. The business address is entered only if it is different from the mailing address. We use the business address to assign location, when available, but in most cases the mailing address field is utilized to assign location. Given the nascent stage at the time of the application, the location information can therefore reflect the place of residence. The application addresses are geocoded to the Census county, tract and block-levels.[28]

For the set of business applications that transition, we can evaluate the plausibility of this assumption by comparing the address of the application to the address of the startup.

[26]See Bayard et al. (2018) for the specific set of filters applied to the application data to exclude applications with no business intent based on the application characteristics.

[27]We use data on business applications for business purposes (BA). Business applications with planned wages (WBA) are a subset of BA that indicate a date for (planned) first wage payments.

[28]Over 99 percent of applications are geocoded to the state and county level, while over 85 percent are coded to the tract and block levels, see Bayard et al. (2018).

Specifically, for the set of business applications that transition, in Table B.1.a we compare the address of the location in the application and the address for the startup in the administrative data (LBD). At the county level, over 90 percent of BA and WBA transition in the same county as in the application. This statistic is nearly 80 percent at the tract level. In other words, the application address and the actual address of the business, if it forms, largely coincide at the county and tract levels. These findings imply that the local conditions largely reflect the location of the business. This is important for the interpretation of our results – location variation reflects business location. For the tract level variation there is a larger fraction of applications where business location differs from application location. Since we find that most of between tract variation is within counties, our robustness analysis using neighboring tract characteristics helps address this potential issue.

The public domain BFS includes monthly tabulations of BA and WBA as well as a high-propensity business applications (HBA) at the national, state and 2-digit industry level. The latter are the applications with characteristics that have a higher propensity to become employer businesses. Such characteristics include planned wage payments (WBA), corporation (CBA), and selected detailed industries that have a high propensity to become employer businesses. The public domain BFS also includes the employer startups that emerge from these applications over the next 4 and 8 quarters. These startups in principle can have emerged from any application but in practice are dominated by HBA. In our micro data analysis in this paper, we do not use HBA, but rather take advantage of the characteristics that underlie HBA. In addition, with the micro data we are able to restrict analysis of startups that emerge from a specific type of application (e.g., WBA).

5.2 Business formations

BFS complements the confidential micro data underlying the EIN applications with the Longitudinal Business Database (LBD). The LBD is the longitudinal version of the Census Bureau's Business Register and contains firm and establishment-level information on age, location, industry, number of employees, quarterly payroll, and EIN for the near-universe of employer businesses in the United States. Using the EIN, business applications are matched to the LBD to identify the incidence and timing of transitions to new employer businesses, or startups. In tracking startups, we use the LBD's identification of new firms that do not reflect changes in ownership or M&A activity (that is we focus on transitions to firms with firm age equal to zero). Importantly, we are able to separately identify transitions that stem from BA and WBA.

The startups we focus on are those that occur within 8 quarters of the application date for the cohort of applications in a year. These startups (in per capita terms) are our empirical counterpart for S in the model. We construct the empirical analog of the transition rate, T, in the model as the ratio of the transitions within 8 quarters to applications, both for BA and WBA. Overall, the micro data that we exploit track millions of applications and startups. On an annual basis, we track more than 2.5 million applications and more than 300 thousand actual employer startups that are linked to these applications over the subsequent eight quarters.

Figure 3 provides the motivation for our focus on applications that occur within 8 quarters. Only one-fourth (BA) to one-third (WBA) of applications that transition within 16 quarters, transition in the same quarter as when the application is received. By eight quarters, after which transition rates flatten out, 90 (BA) to 95 (WBA) percent of transitions have occurred.

6 Variation in startups, ideas, and transition rate

6.1 County and Tract Statistics

We now turn to our primary analysis of the variation at the county and tract level using the BFS microdata. Summary statistics for BA and WBA startups, applications, and transition rates at the county and tract level are presented in Table 1 and Table 2. Focusing on the BA statistics first, startup per capita (defined as per 1,000 prime age adults) averaged 1.337 and 1.582, at the county and tract levels, respectively. This reflects applications per capita of 10.48 and 13.24 with transition rates of 0.128 and 0.117, for the county and tract data. The average applications per capita for WBA is 2.149 and 2.284 for counties and tracts with mean transition rates of 0.407 and 0.373. This results in startups from WBA averaging 0.877 and 0.940, respectively. The statistics highlight the fact that while WBA accounts for a relatively small proportion of overall (BA) applications (roughly 15 to 20 percent) they account for (60 to 65 percent) of employer-business startup activity, reflecting relatively high transition rates. While the means of startups, applications and transition rates are quite similar across counties and tracts, the variation across tracts is much larger. The coefficient of variation is roughly eight times larger for tracts compared to counties for startups and applications and twice as large in the case of the transition rate variable.

We use (12) to analyze the relative contribution of applications per capita (A) and transition rates (T) to variation in startup activity (S). As discussed earlier, this identity holds for all cases with nonzero BA and WBA. Our variance decomposition is conditional on county-year cells where there are positive applications – not much of a restriction at the county level. Our tract-level analysis, where zeros are more prevalent, aggregates application and startup activity over time from 2011 to 2016 and then constructs the S, A, and T. This results in significantly fewer zero observations but omits time variation. In the next section, we incorporate cases with zeroes in our analysis of local conditions.

Table 4 and Table 3 present the variance decomposition at the county and tract levels based on (12). It is apparent that variation in both ideas (applications) and transition rates are important in accounting for the spatial variation in startups. This is true for both the county- and tract-level decompositions. Idea variation is somewhat more important for WBA compared to BA, and the covariance between ideas and transition rates is small. Population weighting increases the relative importance of ideas in counties, as seen in the second row of Table 3. Weighted results also generally increase the negative covariance between applications per capita and transitions per capita in the county analysis.

7 The role of local conditions

We use the regression models (14, 15, 16) to assess the contribution of observable variation in local conditions in accounting for the spatial variation in startups, applications, and transition rates. The specifications we consider use panel data with observations at the county by year or tract by year level for years 2011-2016. We use BA and WBA, and the startups and transition rates originating from them.

The BFS micro data are supplemented with measures of the characteristics or conditions of a location (county or tract)—our counterpart for C_l in the model. The choice of local conditions is influenced by factors discussed in the existing literature. The county and tract characteristics are based on measures of local conditions from the American Community Survey (ACS), Bureau of Economic Analysis (BEA), Community Reinvestment Act (CRA), and the Federal Reserve Board (FRB), and the LBD. Table B.1.e found in the Appendix describes all of the local condition variables used for analysis. We separate local conditions, C_l, into four groups:

- **Demographic conditions:** age, education, race, ethnicity, and foreign born.[29]
- **Household economic conditions:** income per capita, employment-to-population ratio, owner occupied housing share, and debt-to-income ratio,.[30]

[29]See, for example, Doms et al. (2010), Fairlie and Miranda (2016), Azoulay, Jones, Kim and Miranda (2020), Azoulay, Jones, Kim and Miranda (2022), Kerr and Kerr (2017), Bennett and Robinson (2023).

[30]See, for example, Evans and Jovanovic (1989), Hurst and Lusardi (2004), Adelino, Schoar and Severino (2015).

- **Incumbent firm characteristics:** small business lending per small business employment (the ratio of loans less than \$1 million made to firms with less than \$1 million in revenue to employment at firms with less than 500 employees), percent employment in young firms, percent of employment in large firms, and average firm size.[31]

- **Commercial share:** ratio of employment in a tract to the employment plus population.

Conditions constructed from the ACS are measured over five year intervals. The other conditions are measured at an annual frequency. All local conditions are measured with a lag, k, with respect to the outcomes. For the ACS based variables, we use $k = 5$. For all other conditions, we use the average across the lags $k = 1, .., 5$.

The county-level analysis includes conditions across the first three groups, omitting the commercial share variable. The commercial share variable is a tract-level variable used to control for the intensity of commercial activity in a tract. We recognize the that some tracts will be commercially oriented, as opposed to primarily residential, due to zoning restrictions and the presence of housing stocks and commercial property and this variable is used as a proxy for these tract-level differences. The tract-level analysis omits the debt-to-income and small business lending per small business employment variables, as these variables are only available at the county level. Table 5 reports the mean, standard deviation and coefficient of variation for all local conditions, startups per capita, applications per capita, and transition rates used in the county-level regression analysis.

Controlling for detailed location by year effects (commuting zone by year for county level analysis and county by year for tract level analysis) undoubtedly mitigates the contribution of some of these covariates. This is by design because we seek to understand how granular the local variation is in entrepreneurial activity and its relationship with local, within-area covariates. Such detailed fixed effects do not enable identification of causal mechanisms, but do provide guidance on the nature of the local variation.

As discussed above, we explore these relationships at the county and tract level in parallel to shed light on the nature of spatial variation in startup activity and its covariates. The summary statistics and variance decompositions already provide prima facie evidence that there is distinct variation at these different levels of spatial aggregation. Our regression analysis provide further insights on this distinction.

[31] See, for example, Nocke (2006), Michelacci and Silva (2007), Glaeser, Kerr and Ponzetto (2009), Glaeser, Kerr and Ponzetto (2010b), Glaeser et al. (2010a).

7.1 Main results

Tables 6 and 7 report the county- and tract-level regression results for WBA.[32] Table 6 reports the variation in county and tract startups, applications and transition rates explained by the model fixed effects (commuting zone by year and county by year). The R^2 is analogous to the share accounted for by the between component in a standard within-between variance decomposition. Two core messages emerge. First, commuting zone by year effects account for a substantial fraction of the county by year variation in startups per capita, applications per capita, and transition rates but well less than half of that variation. Second, county by year effects account for a much smaller fraction of the tract by year variation in all three tract-level models.

We interpret these results as implying that spatial variation in startup activity and its components reflect rich between and within variation on several dimensions. The substantial fraction of between county variation accounted for by commuting zone by year effects highlights that market level conditions matter (markets defined by commuting zones). However, most of the between county variation is within markets. In turn, the contribution of county by year effects to between tract variation implies that it is not just tract specific factors that account for that variation. However, a very large fraction of the between tract variation is within counties highlighting that tract specific factors are important. We recognize that especially at the tract-level that the between tract variation might reflect tract-specific as well as neighboring tract factors. We consider that possibility in robustness analysis below.

Turning to the regression results in Table 7 and focusing on the explained variation, local conditions help account for the enormous spatial variation in these outcomes, even after taking into account model fixed effects. They explain 16 percent and 20 percent of the variation in startups per capita and applications per capita in the county models, respectively, and 22 percent and 26 percent in the tract models. Local conditions are far less relevant for variation in transition rates, explaining less than five percent of the variation across both models. At the individual entrepreneur level, it is not surprising that it is difficult to account for transitions since the quality of the idea likely dominates but there is systematic variation in transition rates that vary across geographies that at first glance local conditions do not provide much guidance.

Turning to specific local conditions and starting with demographic characteristics, we see that locations with a higher share of individuals with at least a bachelors degree, and a higher share of foreign born are positively related to startup and application activity, as is median age for the county analysis. The association between transition rates and foreign born status

[32]The BA results are reported in the Appendix B and discussed briefly below.

differs between county and tract. At the county level, there is a positive association while a negative association appears in the tract-level analysis. Some college is negatively related to all three variables in the tract level model but not at the county level. For race and ethnic groups, the patterns are mixed. Strikingly, a county or tract with a higher African American share of the population has higher applications per capita but lower startups per capita, underlying the pattern are lower transition rates.[33] In interpreting this (and all results), it is instructive to recall that these effects hold after controlling for a wide variety of local household economic conditions and incumbent firms characteristics. Other notable findings are that the share of Asians is negatively associated with startups per capita, applications per capita, and transition rates at the county level. A higher Hispanic share is associated with lower startups and transition rates (county level), but has no significant relationship with applications.

Local household economic conditions provide mixed results. For the tract level analysis, per capita income is positively associated with startups per capita and applications per capita, but not strongly related to transition rates. One variable where there are markedly different estimates in terms of sign across the counties and tracts is the employment-to-population ratio. In this case, we find a positive relationship using the county level data. Counties with higher employment-to-population ratios have higher startup rates and application rates. However, there is a negative relationship between the employment-to-population ratio for all three of dependent variables using the tract level data. This negative relationship at the tract level may reflect that employment opportunities are more robust for workers that reside in higher employment-to-population ratio tracts, so that there is less incentive to purse self or entrepreneurial employment, holding other factors constant.

We next turn to incumbent firm characteristics. These variables are relatively uncorrelated with startup activity and its components in the county-level sample. Alternatively, at the tract level there are clear associations. Young firm employment share is positively related to startup activity and application activity, whereas large firm employment share and average firm size are negatively related to startups, applications and transition rates. In short, locations with larger and more mature firms are less conducive to application or idea generations, transition rates, and subsequently startup activity. Finally, the commercial share variable is positively related to all three measures of startup activity.

To show the quantitative implications of the estimates, we compute the implied percentage change in a dependent variable of interest corresponding to a one standard deviation

[33] Bennett and Robinson (2023) find related evidence that blacks are more likely than whites to consider starting a business but less likely to transition to starting a business. A distinguishing feature of our focus and evidence is that such variation is important in accounting for local spatial variation in employer startup activity.

change in the covariate relative to the mean (in percent). This requires converting the estimates to elasticities as described in Appendix A.Given that the covariates differ significantly in their variation across locations, we quantify their economic significance by taking into account this variation and multiplying each elasticity with the coefficient of variation of the corresponding covariate. This quantification exercise is summarized in Table 8.

Focusing first on the county-level analysis, bachelors+ degree share, median age and the employment-to-population ratio are, in order, the variables associated with the highest positive percentage change for startups, followed by foreign-born share. The highest negative percentage change in startups is associated with the African American share variable. We observe broadly similar patterns when examining applications with one important exception. The African American share variable is positive with a relatively large magnitude. With regards to transition rates, we see that the largest associations are negative values in particular for African American share and Hispanic share. This highlights the fact that the lower startup rates in neighborhoods with higher African American share are driven by low transition rates with a one standard deviation increase in the African American share being associated with a nearly a 10 percent decline in transition rate. Except for the employment-to-population ratio, the magnitudes of the household economic and incumbent firm variables are all relatively modest.

Turning to the tract-level magnitudes, the ordering of the demographic magnitudes shares some similarity to the county analysis but there are differences. Median age estimated magnitudes are relatively small in comparison and bachelor+ degree estimated magnitude in the case of applications per capita is also small relative to the county. What stands out as distinct in the tract model are the per capita income, share of employment in large firms, average firm size, and commercial share. Neighborhoods with one standard deviation higher level of per capita income relative to the mean have 15 percent higher startups and 17 percent higher applications. In the case of firm size, a one standard deviation higher average firm size relative to the mean have 28 percent lower startups, 22 percent lower applications, and 5 percent lower transition rates. While the commercial share magnitudes are the largest in Table 8, our view is that this variable acts as an important control for spatial distribution of commercial activity versus residential activity across tracts within a county but in and of itself is not the focus of our analysis. Still it is in some respects a control for market conditions within the tract.

We also evaluate how important each group of covariates is in explaining the variation in the outcomes of interest. In order to do so, we use the variance decomposition methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024). This decomposition methodology assigns to each covariate the combination of the direct variance contribution along with half of the covariance with each of the other covariates.[34] By construction this method yields a decomposition where all terms (including the residual) sum to one.[35] Table 9 provides the decomposition for the county and tract models. In general, except for commercial share, local conditions explain a higher fraction of the variation in startups and applications when analyzing the county level data. The demographic, household economic conditions and incumbent firm characteristics all contribute to explaining startups and applications. What might be surprising is that incumbent firm characteristics explain a relatively high fraction of variation for the county data, as the regression parameter and magnitude results showed little association. However, industry employment shares (included but not shown in the regression table) are part of the calculation of the incumbent firm characteristics and these controls account for the bulk of the variation explained by this group. With regard to the transition rate results, the local condition variables explain little of the overall variation in the dependent variables. Demographic and incumbent firm characteristics play very modest roles. Again, in the case of incumbent firm characteristics, it is the industry controls that are the most important contributor to the transition rate model and for the demographic variables, it is the African American share variable that provides the largest contribution.

For the tract-level models of startups and applications, the commercial share variable accounts for most of the explained variation attributed to the local conditions variables. Demographic and household economic conditions contribute modestly. Somewhat surprisingly the incumbent firm characteristics contribute negatively to the explained variation. This can happen through the contribution of covariances.[36] In the transition rate, the demographic group provides the largest contribution to an the admittedly-small, explained within variation, with the African American share variable accounting for over 50 percent of the demographic variable group explained variation.

[34]The contribution of a covariate is given by the product of the estimated coefficient, its covariance with the dependent variable and the ratio of its standard deviation to that of the dependent variable.

[35]Moreover, the residual contribution matches the regression results in Table 7 by construction. We note that we use variance decomposition methodology for the observable local conditions. The reported contribution of common market conditions in Table 9 and subsequent decomposition tables is the contribution of commuting zone by year in the county and county by year in the tract level results.

[36]The covariance contribution is allocated evenly across variables in the covariance. The covariance contribution is the product of the estimated regression coefficients and the covariance of the covariates. When the sign of estimated regression coefficients are opposite signs with a positive covariance between the covariates, the covariance contribution can be negative. For the overall contribution of a covariate to be negative, it must that the direct contribution of the covariate (the regression estimate squared times the variance) is small relative to a negative covariance contribution.

7.2 High technology startup activity

Up to this point, we have not exploited any industry-level variation of startup or application activity across locations. In this section, we examine, as an important example of industry-level analysis, how local conditions influence startups in innovation-intensive industries – what we denote as high-tech startups. Each application is assigned to a specific industry and this allows us to identify applications associated with high tech industries. Using the linked LBD-BFS application data from the 14 four-digit NAICS industries identified in Hecker (2005) as high technology industries based on the STEM intensity of workers in the industry, we construct startups per capita, application per capita, and transition rates utilizing the same approach as above, focusing on the subset of high technology WBA. It should be noted that in the high tech sample there are many location-time periods without high tech applications and this substantially reduces the number of observations in our startup and transition rate models.

Table B.1.c reports the summary statistics for the dependent variables used in high tech analysis while the regression results are reported in Table 10. A key result is that parameter estimates on the bachelors or higher share have increased in size, especially in the applications models. Utilizing the regression magnitude approach described above, a one standard deviation increase in bachelors or higher share relative to its mean would be associated with a 18.25 percent increase in high tech startups and a 30.16 percent increase in applications per capita using the county data. At the tract level, a similar exercise would yield a 6.68 percent increase in high tech startups and 17.01 percent increase in high technology applications per capita.[37]

There are other additional findings to note as well. Some college share is negatively associated with high tech startup activity but the source of negative association differs across county and tract. For counties, negative startup-some college relationship is driven by a lower transitions, whereas in the tract-level sample it is driven by fewer applications per capita. In addition in the tract-level analysis, we find that Asian share is positively associated with startups, applications, and transition rates. This contrasts to the WBA results where there was little correlation between Asian share and startups, applications or turnover rates. We find a generally weaker relationship between foreign born share and high technology activity compared to our earlier, more general analysis where we observed

[37]The small increase in the calculated magnitude for high technology startups associated with bachelors degree or higher share compared to the corresponding WBA results is puzzling given the increase in the calculated magnitudes for high technology applications in the tract analysis. We are exploring the source of this result but note that the samples are quite different as the number of tracts with zero applications rises sharply in the high tech sample and these observations are omitted from the startup model but not the application model.

a positive correlation between foreign born share and startups and applications. We also find a somewhat weaker relationship between applications per capita and African American share for the high-tech startups compared to all WBA startups. For the tract level analysis, the coefficient estimates of the commercial share are lower for high-tech startups compared to all WBA startups.

We report the results from the regression decomposition analysis for the high tech analysis in Table B.2.d. Demographic effects contribute substantially more to applications and startups for high tech startups reflecting the larger impact of the Bachelors+ share. The commercial share contributes substantially less to high tech startups – suggesting that high-tech startups are less sensitive to local commercial activity than overall startups.

Overall, the qualitative patterns we detect for high tech startups and their components are broadly similar to the baseline WBA results. These patterns suggest that our findings have similar qualitative implications over a wide range of entrepreneurial activity – from those entrepreneurs that have objectives of being a potentially high growth startup in an innovative intensive industry to those that reflect distinct objectives like being one's own boss. While qualitatively similar, the quantitative patterns differ in notable ways – with the most striking the much larger role of the Bachelors+ share for high tech startups.

7.3 Accounting for application characteristics in transition

A striking finding of our analysis so far is that there is substantial spatial variation in transition rates across local areas but accounting for observable local conditions still leaves substantial variation unexplained. To dig deeper, we consider an application-level analysis of transitions. Our WBA sample includes over 2.3 million individual applications and the BA sample includes over 13.8 million individual applications, providing us with a large set of observations to carry out this analysis. We include the BA sample here in this individual level analysis because we can control for application type directly in the statistical analysis (i.e., we control for the type of application in the analysis, including whether an application is part of WBA – the focus of our analysis earlier).

To explore the variation in transition rates, the linear probability model in (17) is estimated on the underlying application data where the dependent variable is an indicator variable identifying which applications transition to employer status over an eight-quarter window. The control variables include the local condition variables used in the tract-level regressions, as well as a set of variables that control for application characteristics. These characteristics include controls for detailed industry of the application (4-digit NAICS), reason for applying, type of entity applying (e.g, planned wages, incorporation), LLC status, inclusion of a trade name, business start date, and quarter of application.[38] The models include county-year fixed effects, and standard errors are clustered at the county level.

The results are presented in Table 11, where only the coefficients of the local condition variables are presented. The coefficients on the application characteristics are omitted as they involve a large number of estimates which are not our main focus. Overall, the results are consistent with the tract-level regression coefficients reported in Table 7, even after controlling for detailed application characteristics. Local demographic and household economic variables such as African American share, some college share and foreign born share are associated with a lower probability of an application transitioning to employer status, while the share of the population with at least a bachelors degree and owner-occupied share are associated with a higher likelihood of transitioning. With regards to incumbent firm results, locations with higher share of employment in young firms have higher transition rates while those with higher share of employment in big or large average size have lower transition rates. The commercial share variable is positively associated with transition rates. In short, the relationships between local conditions and transition rates observed above are robust when application characteristics are controlled for to account for application heterogeneity across tracts.

The R^2 values in Table 11, especially for WBA, are low in absolute terms and relative to our tract-level results. This is not surprising since idiosyncratic variation in the quality of ideas emphasized in the model in section 3 likely dominates the overall variation.[39] This finding highlights the importance of the nascent entrepreneurship phase (including the creation and the pursuit of an idea) for understanding variation in startups per capita. Note also that the higher R^2 for BA is expected given that a control variable in the linear probability model for BA is an indicator of whether the application is a WBA application (i.e. has planned wage payments). The latter is highly predictive of an application transitioning to an employer business.[40]

[38] See Bayard et al. (2018) for a detailed discussion of estimating transition probabilities based on application characteristics. Our analysis here follows closely that approach with two caveats. The model estimated here includes detailed tract-level characteristics but excludes the detailed interactions among variables. Importantly, our model is estimated pooled across states, whereas in the cited work the models are estimated individually at the state level with a rich set of industry-application characteristic interactions.

[39] Among such factors that some of existing literature has thought about include "home bias" (Dahl and Klepper, 2015) and outside options (Manso, 2016; Choi, 2017; Dillon and Stanton, 2017; Gottlieb, Townsend and Xu, 2022).

[40] See also Bayard et al. (2018) for a similar finding.

7.4 Supplementary Regression Analysis

This section reports on a set of additional models that were estimated along with the core analysis of the WBA models. As mentioned above, we estimated all models and additional analysis using the entire set of business applications (BA). The results are found in Table B.2.a, Table B.2.b and Table B.2.c in Appendix B. The sign and significance of the parameter estimates are very similar to the results reported above in Table 7. The only substantive difference is that in the application regression based on county data there appears more of a role for incumbent firm characteristics. The other main difference between the BA and WBA models is that local characteristics explain more of the overall variation across counties and across tracts for each of the dependent variables. It might seem surprising that the results using the broader based BA are so similar to the WBA results since we know that many BA applications have an objective of becoming a nonemployer business instead of an employer business. We interpret our findings as suggesting that local conditions that are favorable for new employer businesses are also favorable for new nonemployer businesses. Future research investigating the relationship between employer and nonemployer startups in this context would be of considerable interest.

The second set of models estimated are WBA models for the tract-level data that include neighboring-tract characteristics as additional control variables. We use the same set of variables that are included in the base model, only these variables are measured using the respective means across adjacent tracts. Table B.2.e reports the regression results and Table B.2.f reports the regression decomposition. Again, the results are very similar to those reported above. The only substantive difference is found in the applications model where the bachelors or higher degree variable is no longer statistically significant. In addition, the inclusion of neighbor variable controls adds only modestly to the explained variation. We interpret these findings as providing support for the view that there is an important component of spatial variation in startups, applications, and transitions that is truly tract-specific. Put differently, our tract level results are not just capturing the combined influence of own and neighboring tract effects. This finding suggests that spatial variation in startup activity is very local.

The final analysis examines the average length of time it takes an application that transitions within the 8 quarters. The theory section highlighted the fact that the length of time an application takes to become an employer business can also depend on local characteristics, in part due to a variety of local frictions. For each application that transitions we calculate the length of time in the number of quarters it takes to transition and then take the average value across all applications that transition in a location-year cell. We then regress this measure against our local control variables. We report the results of this exercise in Table

B.2.g. We find that there are slower transitions for higher African American share at both the county and tract level. We also find slower transitions for higher Bachelors+ share at the tract level but only a weak relationship at the county level. More generally, the results are less uniform across the county and tract level results than the earlier analysis on startups, applications and transition rates. These patterns suggest to us that variation in duration reflects complex factors. On the one hand, a slower transition may reflect greater frictions for making a transition and the African American share results are potentially consistent with this interpretation. On the other hand, a slower transition may reflect variation in the type of startups being contemplated. It may be, for example, that a business application with a high growth potential takes longer to startup – potentially consistent with the Bachelors+ result.

7.5 Local conditions and ranking of locations

In this section, we conduct a ranking analysis exploring how well observable factors account for the ranking of local areas in terms of startups per capita. We conduct this analysis with both county and tract level variation, focusing on WBA and associated transitions. If the ranking of the locations in startup activity based on the covariates alone captures well the actual ranking, the observable conditions we consider can be used to characterize locations with high versus low startup activity.

To start, we rank all counties and tracts in terms of their startup rates per capita, applications per capita, and transition rate. We then classify each county and tract into the deciles of the startup rates per capita distribution. In turn, we compute the average mean rank of applications per capita and transition rates in each of the deciles of the startup rate per capita distribution.

The left panel of Figure 4 shows the results for this exercise at the county level and left panel of Figure 5 at the tract level. The results are similar so we discuss them together. By construction, the mean of the average rank of startups per capita rises monotonically (and essentially linearly) by startup per capita decile. Both applications per capita and transition rates rankings increase with the startup deciles but with distinct nonlinear patterns. In particular, the top decile counties and tracts are especially characterized as having high rankings of high applications per capita, rather than transition rates. *Superstar* startup areas can thus be characterized as those with high idea creation. In contrast, the bottom decile counties and tracts are especially characterized as having low rankings of transition rates, rather than low rankings of applications per capita.[41]

[41] These inferences are based on the observation that the ranking by applications per capita has a steeper slope at high deciles while the ranking by transitions has a steeper slope at low deciles.

We next evaluate how well our models account for these ranking patterns. The right panel of Figures 4 and 5 provide the results. Here we compute the mean predicted rank of applications per capita and transition rates for each decile from only the fixed effects (commuting zone by year in Figure 4 and county by year in Figure 5) versus from only the observable covariates. Strikingly, the observable covariates yield a predicted ranking pattern that corresponds reasonably closely to the actual ranking pattern. Thus, even though the observable covariates account for less than half of the observed overall spatial variation, they provide substantial guidance with respect to the ranking of counties and tracts in terms of startups per capita.

The observables also capture more of the nonlinearities at the top and bottom deciles discussed above. That is, the observables account for the steeper slope of applications per capita at the top deciles and the steeper slope of transitions at the bottom deciles. For communities and neighborhoods that have especially poor transition rates and associated startup activity, observable factors are highly informative.

Finally, we explore how entrepreneurial activity varies by a measure of social/economic mobility at the tract level. Utilizing information constructed by the Equality of Opportunity project, we replace the rank of startups (the X-axis) with the rank of a variable that measures social/economic mobility – specifically, "the mean household income rank for children whose parents are at the 25th percentile of the national income distribution, where incomes for children were measured as mean earnings in 2014-2015 when they were between 31-37."[42] The layout of Figure 6 mimics that of Figure 5. The left panel shows that the rank of transition rate rises along with social/economic mobility, though in a nonlinear fashion on the lower end, whereas the rank of applications is U-shaped. The right panel shows the relationship between social/economic mobility rank and the predicted ranks based on the regression components. The regression component that corresponds most closely to the social/economic mobility ranking is that based on the predicted transition rate using observable covariates. These findings are consistent with our interpretation that an important component of spatial variation in transition rates reflects frictions that vary across locations that are in turn related to observable factors such as demographics.

This exercise shows that using a relatively parsimonious set of observable characteristics, we can infer the relative position of counties and tracts in their performance with respect to startups per capita. This result is useful in the sense that it identifies some key observable local conditions that can be used to gauge the startup formation potential of locations without having to know detailed information on the volume of applications or their characteristics, or a large set of local characteristics that are hard to come by or measure. For example,

[42] Source: Codebook for Table 1: opportunityinsights.org/data/.

policymakers and local planners can make use of this finding to assess local conditions correlated with startup activity. Designing policies to potentially improve these conditions is also important, to the extent some conditions we consider (such as household economic conditions and incumbent firm characteristics) can be influenced by local policies. However, for such purposes identifying causal mechanisms that generate the observed relationships is crucial – a key task for future work. In fact, some of the conditions we observe may not themselves be causal drivers of entrepreneurship, but they may be the symptoms or results of local policies or other, unobserved factors.

8 Conclusion

Startups play a disproportionate role in aggregate job creation, innovation, and productivity growth. Relatively little is known about the nascent phase of startups both at the aggregate and the micro level. In this paper, we have focused on the spatial variation in the nascent phase of entrepreneurship using novel data that permits us to decompose startups per capita at the local level into idea creation (applications per capita) and transitions of business ideas to employer startups.

We find enormous variation in startups per capita at the granular levels of geography. Much of this variation is *within* in the sense that most of the between-county variation is not accounted for by commuting zone by year effects and most of the tract-level variation is not accounted for by county by year effects. Variation in both applications per capita and transitions contribute substantially to this spatial variation, with both components accounting for about half of the variation in startups per capita.

Local environment, captured by demographic and household economic conditions and incumbent firm characteristics, accounts for a substantial fraction of the within variation at the county and tract level. In general, across both counties and tracts, education, age, and foreign-born share are important local factors in the sense that they have a large positive association with business applications. Interestingly, specific conditions have distinct relationships with idea creation and transitions. For example, at both the county and tract level, we find that local areas with a higher share of African Americans have a lower startup rate per capita, but this reflects an offsetting positive relationship with applications per capita and a negative relationship with transitions.

Local observable conditions account for less than half of the observed spatial variation in startups per capita, applications per capita, and especially, transitions. Nevertheless, we find that the predicted ranking of startups per capita based only on local observable conditions is closely related to the actual ranking. Policymakers and analysts exploring the sources of variation in entrepreneurship can thus use variation in these local observable conditions as a useful indicator to assess the startup potential of an area.

Appropriate caution is needed for the interpretation of our analysis of observable factors in that we are quantifying the contribution of observable covariates without causal inference. Our findings highlight that exploring such causal factors should be a high priority for future research. With variation in startup rates per capita that vary by as much as a factor of five across local areas and accompanying variation in idea creation and transitions, there is enormous disparity in entrepreneurship and its key determinants across local areas (including and especially across neighborhoods in the same county). Understanding the determinants of this variation should be a high priority. Entrepreneurship has important aggregate implications, but it is also a pathway and opportunity for economic mobility both for the entrepreneurs and the workers hired by such firms. In this respect, our finding of enormous spatial variation in entrepreneurship patterns including idea creation and transition rates is related to the findings of Chetty et al. (2014) who emphasized enormous spatial variation in distinct measures of social/economic mobility. Indeed, we show there is a close correspondence between the spatial variation in transition rates and the spatial variation in these social/economic mobility measures.

We regard this analysis as a first step in exploring nascent entrepreneurship with a number of areas open for future research. The relationship between applications, startups, and post-entry dynamics (e.g., high growth outcomes) is a natural area of interest. Another is to explore the transitions to nonemployer businesses. Yet another area of research is the potentially changing relationships of startups, applications, and transition rates in the pandemic at the local level. The surge in overall applications as well as those that are likely employers has received considerable attention (see, Dinlersoz, Dunne, Haltiwanger and Penciakova (2021), Haltiwanger (2022), and Decker and Haltiwanger (2023)). The evidence is just emerging that this surge in applications has yielded a surge in employer startups but there has not been the type of analysis at the local level during the pandemic of the type in the current paper. Part of this analysis will have to await the development of the underlying micro data on transitions of applications for the post-2020 period.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "House Prices, Collateral, and Self-Employment," *Journal of Financial Economics*, 2015, *117*, 288–306.

Andrews, RJ, Catherine Fazio, Yupeng Liu, Jorge Guzman, and Scott Stern, "The Startup Cartography Project: Measuring and Mapping Entrepreneurial Ecosystems," *Research Policy*, 2018, *51* (2), 104437.

Asplund, Markus and Volker Nocke, "Firm Turnover in Imperfectly Competitive Markets," *Review of Economic Studies*, 2006, *73* (2), 295–327.

Azoulay, Pierre, Benjamin Jones, J. Daniel Kim, and Javier Miranda, "Age and High-Growth Entrepreneurship," *American Economic Review: Insights*, 2020, *2* (1), 65–82.

_ , _ , _ , and _ , "Immigration and Entrepreneurship in the United States," *American Economic Review: Insights*, 2022, *4* (1), 71–88.

Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens, "Early Stage Business Formation: An Analysis of Applications for Employer Identification Numbers," 2018. NBER Working Paper No. 24364.

Bennett, Victor and Aaron Chatterji, "The Entrepreneurial Process: Evidence from a Nationally Representative Survey," *Strategic Management Journal*, 2023, *44*, 86–116.

Bennett, Victor Manuel and David T. Robinson, "Why Aren't There More Minority Entrepreneurs?," Technical Report, Duke University 2023.

Bianchi, Milo and Matteo Bobba, "Liquidity, Risk, and Occupational Choices," *The Review of Economic Studies*, 2013, *80* (2), 491–511.

Cagetti, Marco and Mariacristina De Nardi, "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 2006, *115* (5), 835–870.

Chen, Jiafeng and Jonathan Roth, "Log-like? Identified ATEs defined with zero-valued outcomes are (arbitrarily) scale-dependent," Working Paper 2023.

Chen, John S., David C. Croson, Daniel W. Elfenbeinand, and Hart E. Posen, "The Impact of Learning and Overconfidence on Entrepreneurial Entry and Exit," *Organization Science*, 2018, *29*, 989–1009.

Chetty, R., N. Hendren, P. Kline, and E. Saez, "Where is the land of opportunity? the geography of intergenerational mobility in the united states," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Choi, Joonkyu, "Entrepreneurial Risk-Taking, Young Firm Dynamics and Aggregate Implications," Technical Report, University of Maryland 2017.

Dahl, Michael S. and Steven Klepper, "Whom do new firms hire?," *Industrial and Corporate Change*, 2015, *24* (4), 819–836.

Davis, Steven, John Haltiwanger, and Scott Schuh, "Small business and job creation: Dissecting the myth and reassessing the facts," *Small Business Economics*, 1996, *8*, 297–315.

_ , _ , C.J. Krizan, Ron Jarmin, Javier Miranda, Al Nucci, and Kristin Sandusky, "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Non-Employer Businesses," in "in Producer Dynamics: New Evidence from Micro Data," University of Chicago Press, 2009, pp. 329–366.

Decker, Ryan A. and John Haltiwanger, "Surging Business Formation in the Pandemic: Causes and Consequences," *Brookings Papers on Economic Activity*, 2023, *Fall 2023 (forthcoming).*

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "Where has all the skewness gone? The decline in high-growth (young) firms in the US," *European Economic Review*, 2016, *86*, 4–23.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The role of entrepreneurship in US job creation and economic dynamism," *Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

Dillon, Eleanor W. and Christopher T. Stanton, "Self-Employment Dynamics and the Returns to Entrepreneurship," 2017, (23168).

Dinlersoz, Emin, Timothy Dunne, John Haltiwanger, and Veronika Penciakova, "Business Formation: A Tale of Two Recessions," *American Economic Association: Papers and Proceedings*, May 2021.

Doms, Mark, Ethan Lewis, and Alicia Robb, "Local labor force education, new business characteristics, and firm performance," *Journal of Urban Economics*, 2010, *67* (1).

Ericson, Richard and Ariel Pakes, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 1995, *62* (1), 53–82.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta, "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges," *Review of Economic Studies*, 2024.

Evans, David S. and Boyan Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, 1989, *97* (4), 808–827.

Fairlie, Robert W. and Javier Miranda, "Taking the Leap: The Determinants of Entrepreneurs Hiring Their First Employee," *Journal of Economics & Management Strategy*, 2016, *26* (1), 3–34.

Fazio, Catherine, Jorge Guzman, Yupeng Liu, and Scott Stern, "How is COVID Changing the Geography of Entrepreneurship? Evidence from the Startup Cartography Project," Working paper 28787, NBER Working paper 2021.

Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange, "Urban economics and entrepreneurship," *Journal of Urban Economics*, 2010, *67* (1), 1–14.

_ , **William R. Kerr, and Giacomo A.M. Ponzetto**, "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?," *Journal of Economics & Management Strategy*, 2009, *18* (3), 623–663.

_ , _ , **and** _ , "Clusters of entrepreneurship," *Journal of Urban Economics*, 2010, *67*, 150–168.

Gottlieb, Joshua D., Richard R. Townsend, and Ting Xu, "Does Career Risk Deter Potential Entrepreneurs?," *The Review of Financial Studies*, 2022, *35* (9), 3973–4015.

Guzman, Jorge and Scott Stern, "The State of American Entrepreneurship? New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988-2014," *American Economic Journal: Economic Policy*, 2020, *12* (4), 212–243.

Haltiwanger, John, "Entrepreneurship During the Covid-19 Pandemic: Evidence from the Business Formation Statistics," *Entrepreneurship and Innovation Policy and the Economy*, 2022, *1*, 9–42.

_ , **Ron S Jarmin, and Javier Miranda**, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hecker, Daniel E, "High-Technology Employment: a NAICS-Based Update," *Monthly Labor Review*, 2005, *128* (7), 57–73.

Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar, "Can Unemployment Insurance Spur Entrepreneurial Activity? Evidence from France," *The Journal of Finance*, 2020, *75* (3), 1247–1285.

Hopenhayn, Hugo, "Entry, Exit and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60* (5), 1127–1150.

Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Hurst, Erik and Annamaria Lusardi, "Liquidity Constraints, Household Wealth, and Entrepreneurship," *Journal of Political Economy*, 2004, *112* (2), 319–347.

_ **and Benjamin W Pugsley**, "Wealth, Tastes, and Entrepreneurial Choice," in "Measuring Entrepreneurial Businesses: Current Knowledge and Challenges," University of Chicago Press, 2017, pp. 111–151.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670.

Kerr, Sari Pekkala and William R Kerr, "Immigrant entrepreneurship," in "Measuring entrepreneurial businesses: Current knowledge and challenges," University of Chicago Press, 2017, pp. 187–249.

__ **and** __ , “Immigrant entrepreneurship in America: Evidence from the survey of business owners 2007 & 2012,” *Research Policy*, 2020, *49* (3), 103918.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf, “Entrepreneurship as experimentation,” *Journal of Economic Perspectives*, 2014, *28* (3), 25–48.

Lucas, Robert, “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 1978, *9*, 508–523.

Manso, Gustavo, “Experimentation and the Returns to Entrepreneurship,” *The Review of Financial Studies*, 2016, *29* (9), 2319–2340.

Michelacci, Claudio and Olmo Silva, “Why so Many Local Entrepreneurs?,” *The Review of Economics and Statistics*, 2007, *89* (4), 615–633.

Nocke, Volker, “A Gap for Me: Entrepreneurs and Entry,” *Journal of the European Economic Association*, 2006, *60* (4), 929–956.

Polkovnichenko, Valery, “Human capital and the private equity premium,” *Review of Economic Dynamics*, 2003, *6* (4), 831–845.

Poschke, Markus, “Who becomes an entrepreneur? Labor market prospects and occupational choice,” *Journal of Economic Dynamics and Control*, 2013, *37*, 693–710.

Reynolds, Paul D., “When is a Firm Born? Alternative Criteria and Consequences,” *Business Economics*, 2017, *52*, 42–56.

__ **, Nancy M. Carter, William B. Gartner, and Patricia G. Greene**, “The Prevalence of Nascent Entrepreneurs in the United States: Evidence from the Panel Study of Entrepreneurial Dynamics,” *Small Business Economics*, 2004, *23*, 263–284.

Rosenthal, Stuart S. and William C. Strange, “Geography, Industrial Organization, and Agglomeration,” *Review of Economics and Statistics*, 2003, *85* (2), 377–393.

Tornqvist, Leo, Pentti Vartia, and Yrjo O. Vartia, “How Should Relative Changes Be Measured?,” *The American Statistician*, 1985, *39* (1), 43–46.

Vardishvili, Ia, “Entry Decision, the Option to Delay Entry, and Business Cycles,” 2023.

Vereshchagina, Galina and Hugo A. Hopenhayn, “Risk Taking by Entrepreneurs,” *American Economic Review*, 2009, *99* (5), 1808–1830.

Table 1: All startups, applications, and transition rates: summary statistics

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	Startups pc	Applications pc	Transition rate	Startups pc	Applications pc	Transition rate
Mean	1.337	10.480	0.128	1.582	13.240	0.117
SD	0.894	6.109	0.053	9.296	56.870	0.091
CV	0.669	0.583	0.414	5.876	4.295	0.779

Notes: Reports the mean, standard deviation (SD), and coefficient of variation (CV) of BA startups per 1,000 prime-age (20-64 years old) people (startups pc), BA per 1,000 prime-age people (applications pc), and BA transition rate (startups divided by applications) between 2011 and 2016, separately for county- and tract-level data.

Table 2: Wage startups, applications, and transition rates: Summary Statistics

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	Startups pc	Applications pc	Transition rate	Startups pc	Applications pc	Transition rate
Mean	0.877	2.149	0.407	0.940	2.284	0.373
SD	0.628	1.273	0.150	5.207	10.990	0.289
CV	0.716	0.592	0.369	5.540	4.812	0.775

Notes: Reports the mean, standard deviation (SD), and coefficient of variation (CV) of WBA startups per 1,000 prime-age (20-64 years old) people (startups pc), WBA per 1,000 prime-age people (applications pc), and WBA transition rate (startups divided by applications) between 2011 and 2016, separately for county- and tract-level data.

Table 3: Variance decomposition for wage startups per capita

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	Applications pc	Transition rate	2 $\times$ covariance	Applications pc	Transition rate	2 $\times$ covariance
Unweighted	0.676	0.376	-0.052	0.659	0.325	0.016
Weighted	0.919	0.369	-0.288	0.645	0.321	0.034

Notes: Reports the variance decomposition of log(WBA startups pc) into log(WBA pc) and log(WBA transition rate) for the period 2011-2016. The first three columns reports the county-level results, and the last three columns report the tract-level results. county and tract population is used for weighting, as appropriate. Startups are defined as applications that transition to an employer business within eight quarters after application. "pc" refers to per capita (per 1000 prime-age people).

Table 4: Variance decomposition for all startups per capita

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	Applications pc	Transition rate	2 $\times$ covariance	Applications pc	Transition rate	2 $\times$ covariance
Unweighted	0.485	0.518	-0.004	0.579	0.424	-0.002
Weighted	0.983	0.428	-0.410	0.590	0.417	-0.007

Notes: Reports the variance decomposition of log(BA startups pc) into log(BA pc) and log(BA transition rate) for the period 2011-2016. The first three columns reports the county-level results, and the last three columns report the tract-level results. county and tract population is used for weighting, as appropriate. Startups are defined as applications that transition to an employer business within eight quarters after application. "pc" refers to per capita (per 1000 prime-age people).

Table 5: Regression Variable Summary Statistics: Baseline

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	SD	CV	Mean	SD	CV
DHS(WBA startups pc)	-0.178	0.595	-3.338	-0.520	0.977	-1.879
DHS(WBA pc)	-0.121	0.466	-3.847	-0.371	0.782	-2.109
WBA transition rate	0.408	0.150	0.368	0.373	0.289	0.775
log(median age)	3.689	0.130	0.035	3.629	0.200	0.055
bachelors or higher share	0.194	0.087	0.448	0.278	0.184	0.665
some college share	0.295	0.053	0.181	0.287	0.079	0.277
African American share	0.092	0.149	1.620	0.137	0.223	1.628
Asian share	0.012	0.025	2.083	0.045	0.087	1.929
Hispanic share	0.085	0.135	1.588	0.153	0.211	1.376
foreign born share	0.045	0.056	1.244	0.123	0.137	1.114
log(per capita income)	10.030	0.222	0.022	10.140	0.455	0.045
emp-pop ratio	0.550	0.082	0.149	0.582	0.106	0.182
owner-occupied share	0.721	0.078	0.109	0.646	0.227	0.352
log(debt-to-income ratio)	0.436	0.526	1.206	.	.	.
share of emp in young firms	0.120	0.053	0.442	0.175	0.124	0.705
share of emp in large firms	0.131	0.138	1.050	0.063	0.160	2.545
DHS(avg firm emp)	-0.067	0.371	-5.517	-0.287	0.664	-2.309
DHS(SME loans/employment)	-0.212	0.608	-2.875	.	.	.
commercial share	.	.	.	0.179	0.170	0.946

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the mean, standard deviation (SD), and coefficient of variation (CV) of variables used as dependent and control variables in county and tract level regressions. The years covered are 2011-2016. Startups are defined as applications that transition to an employer business within eight quarters after application. "DHS" refers to the transformation based on Davis et al. (1996).

Table 6: Contribution of location fixed effects: WBA

	(1) County	(2) Tract
DHS(WBA startups pc)	0.407	0.111
DHS(WBA pc)	0.513	0.208
WBA transition rate	0.344	0.108

Notes: Reports the share of variance in DHS(WBA startups pc), DHS(WBA pc), and WBA transition rates accounted for by commuting zone by year FE (col. 1) and county by year FE (col. 2) for data at the county-year and tract-year levels in 2011-2016, respectively. “pc” refers to per capita (per 1000 prime-age people).

Table 7: WBA Baseline Regression Results

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
log(median age)	0.795***	0.703***	0.0373	0.0349	0.0897***	0.00803
	(0.109)	(0.0919)	(0.0255)	(0.0317)	(0.0294)	(0.00662)
bachelors or higher share	1.378***	1.170***	0.113**	0.341***	0.0994***	0.0875***
	(0.2)	(0.168)	(0.0473)	(0.0426)	(0.0374)	(0.0107)
some college share	–0.197	0.0735	–0.0835	–0.241***	–0.202***	–0.0417***
	(0.254)	(0.184)	(0.0577)	(0.0577)	(0.0539)	(0.0126)
African American share	–0.347***	0.407***	–0.236***	–0.286***	0.543***	–0.203***
	(0.0906)	(0.0739)	(0.0222)	(0.0404)	(0.0443)	(0.00692)
Asian share	–1.537**	–1.216*	–0.249**	0.0585	0.106	0.00785
	(0.608)	(0.622)	(0.115)	(0.12)	(0.0969)	(0.0231)
Hispanic share	–0.295*	0.00708	–0.188***	–0.164**	–0.064	–0.0198
	(0.175)	(0.14)	(0.0438)	(0.0762)	(0.101)	(0.0144)
foreign born share	0.935***	0.614**	0.175**	0.332***	0.380***	–0.0273**
	(0.349)	(0.306)	(0.0796)	(0.117)	(0.134)	(0.0123)
log(per capita income)	0.0821	0.0698	–0.00535	0.297***	0.336***	0.00487
	(0.103)	(0.0776)	(0.0255)	(0.0178)	(0.0154)	(0.00462)
emp-pop ratio	0.966***	0.949***	0.0228	–0.310***	–0.159***	–0.0235***
	(0.205)	(0.186)	(0.0504)	(0.0435)	(0.0467)	(0.00848)
owner-occupied share	–0.167	–0.244*	0.0106	–0.00261	–0.0327	0.0307***
	(0.18)	(0.134)	(0.0389)	(0.028)	(0.036)	(0.00403)
log(debt-to-income ratio)	–0.0253	–0.0197	–0.00474			
	(0.0196)	(0.0148)	(0.00455)			
share of emp in young firms	0.02	0.249	–0.0689	0.231***	0.198***	0.00576
	(0.2)	(0.171)	(0.0502)	(0.0276)	(0.0203)	(0.00664)
share of emp in large firms	–0.107	0.119	–0.0476**	–0.530***	–0.550***	–0.0243***
	(0.0938)	(0.078)	(0.0212)	(0.024)	(0.0176)	(0.00491)
DHS(avg firm emp)	0.0751	–0.110**	0.0340**	–0.188***	–0.151***	–0.0114***
	(0.0619)	(0.0477)	(0.0138)	(0.00588)	(0.00572)	(0.00143)
DHS(SME loans/employment)	0.0216	0.00162	0.00769*			
	(0.0159)	(0.0128)	(0.00398)			
commercial share				2.666***	2.429***	0.152***
				(0.0295)	(0.0273)	(0.00606)
Ind emp. shares	yes	yes	yes	yes	yes	yes
Observations	17,000	17,500	17,000	398,000	430,000	398,000
Fixed effects	cz x yr	cz x yr	cz x yr	fips x yr	fips x yr	fips x yr
SE clustering	cz	cz	cz	fips	fips	fips
R-squared	0.5023	0.6078	0.3752	0.305	0.416	0.1443
Within R-squared	0.1606	0.1951	0.04781	0.2186	0.2623	0.04071

Notes: County regressions include commuting zone (CZ) $\times$ year FE and tract regressions include county $\times$ year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the CZ level for county regressions and at county level for tract regressions.

Table 8: WBA Regression Magnitudes

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
median age	10.017	8.858	1.147	0.686	1.764	0.431
bachelors or higher share	11.962	10.170	2.419	6.318	1.862	4.322
some college share	-1.050	0.398	-1.086	-1.911	-1.607	-0.886
African American share	-5.184	5.994	-8.586	-6.349	12.047	-12.210
Asian share	-3.749	-3.124	-1.458	0.579	0.965	0.193
Hispanic share	-3.970	0.159	-6.193	-3.440	-1.376	-1.101
foreign born share	5.225	3.483	2.364	4.567	5.236	-1.003
per capita income	1.943	1.659	-0.308	15.385	17.405	0.673
emp-pop ratio	7.912	7.778	0.462	-3.276	-1.693	-0.673
owner-occupied share	-1.308	-1.918	0.207	-0.070	-0.739	1.866
debt-to-income ratio	-1.112	-0.890	-0.534	.	.	.
share of emp in young firms	0.088	1.326	-0.884	2.820	2.467	0.212
share of emp in large firms	-1.470	1.680	-1.575	-8.399	-8.908	-1.018
avg firm emp	2.828	-4.147	3.129	-27.580	-22.152	-4.548
SME loans/employment	2.999	0.273	2.590	.	.	.
commercial share	.	.	.	45.124	41.151	6.906

Notes: Reports the estimated % change in the LHS variable induced by the percent change in the RHS variable equivalent to a one standard deviation multiple of the mean. The LHS variable of the regression is listed in the columns, and each RHS variable is listed in the rows. “DHS” refers to the transformation based on Davis et al. (1996). “pc” refers to per capita (per 1000 prime-age (20-64 years old) people).

Table 9: WBA Regression Decomposition

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Groups						
Demographic	0.056	0.076	0.024	0.029	0.009	0.029
HH economic conditions	0.033	0.047	-0.000	0.029	0.039	0.002
Incumbent firm characteristics	0.072	0.073	0.024	-0.011	-0.020	0.002
Commercial share	.	.	.	0.171	0.235	0.007
Categories						
Local conditions	0.161	0.195	0.048	0.219	0.262	0.041
Common market conditions	0.342	0.413	0.327	0.086	0.154	0.104
Residual	0.498	0.392	0.625	0.695	0.584	0.856

Notes: Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(WBA startups pc), DHS(WBA pc) and WBA transition rate for county- and tract-level analysis. Note that all control variables are included, along with location × fixed effects. "DHS" refers to the transformation based on Davis et al. (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The fourth row is the sum of the contribution of all individual variables (or the sum of the contribution of all grouped variables), and corresponds to the within R^2; the fifth row is the contribution of location × year FE; and the last row is the remaining variation that is unexplained by either local conditions or common market conditions. Note that log(HH debt-to-income) is included in HH economic conditions and DHS(SME loans/employment) is included in incumbent firm characteristics for county level results only.

Table 10: High-tech WBA Regression Results

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
log(median age)	0.429	–0.029	0.165	0.0721	–0.361***	0.016
	(0.34)	(0.179)	(0.12)	(0.0629)	(0.0286)	(0.0194)
bachelors or higher share	2.087***	3.449***	0.087	0.362***	0.922***	0.0958***
	(0.577)	(0.355)	(0.189)	(0.098)	(0.0504)	(0.0312)
some college share	–2.302***	1.212***	–0.795***	–0.281**	–0.216***	0.0213
	(0.796)	(0.4)	(0.266)	(0.136)	(0.0764)	(0.0437)
African American share	–0.156	0.353**	–0.278***	–0.303***	0.0795***	–0.175***
	(0.272)	(0.173)	(0.0844)	(0.0622)	(0.0257)	(0.0205)
Asian share	1.918*	0.209	–0.137	0.429***	0.607***	0.0710**
	(1.122)	(0.773)	(0.3)	(0.147)	(0.212)	(0.0336)
Hispanic share	–0.116	0.535	–0.155	–0.175*	–0.288***	–0.0441
	(0.519)	(0.337)	(0.172)	(0.0977)	(0.0676)	(0.0295)
foreign born share	0.544	0.382	–0.119	–0.114	0.432***	–0.0858**
	(0.977)	(0.565)	(0.304)	(0.105)	(0.1)	(0.0336)
log(per capita income)	0.182	0.286*	–0.122	0.0596*	0.302***	–0.00885
	(0.336)	(0.158)	(0.104)	(0.0348)	(0.0228)	(0.0108)
emp-pop ratio	0.83	0.0192	0.137	–0.0842	–0.0683	0.0595**
	(0.708)	(0.356)	(0.227)	(0.0888)	(0.0475)	(0.0273)
owner-occupied share	0.542	0.101	0.116	–0.00149	–0.0168	–0.00292
	(0.42)	(0.242)	(0.146)	(0.0478)	(0.0351)	(0.0156)
log(debt-to-income ratio)	–0.136*	0.142***	–0.0484**			
	(0.0717)	(0.0281)	(0.0231)			
share of emp in young firms	–0.26	0.746**	0.187	0.109	0.199***	0.00684
	(0.815)	(0.29)	(0.221)	(0.0683)	(0.0202)	(0.0234)
share of emp in large firms	0.363	0.0477	0.0976	–0.124**	–0.190***	0.00393
	(0.347)	(0.15)	(0.124)	(0.0516)	(0.0241)	(0.0169)
DHS(avg firm emp)	–0.538***	0.749***	–0.0551	–0.147***	–0.0330***	–0.0119**
	(0.205)	(0.0838)	(0.0676)	(0.0158)	(0.00659)	(0.00502)
DHS(SME loans/employment)	–0.00421	0.0770***	0.0582**			
	(0.0646)	(0.0259)	(0.0232)			
commercial share				1.110***	1.107***	0.0456***
				(0.0719)	(0.0501)	(0.0173)
Ind emp. shares	yes	yes	yes	yes	yes	yes
Observations	7,200	17,500	7,200	64,000	430,000	64,000
Fixed effects	cz x yr	cz x yr	cz x yr	fips x yr	fips x yr	fips x yr
SE clustering	cz	cz	cz	fips	fips	fips
R-squared	0.5895	0.5403	0.4949	0.1365	0.1922	0.09759
Within R-squared	0.1682	0.251	0.03922	0.04724	0.08963	0.01104

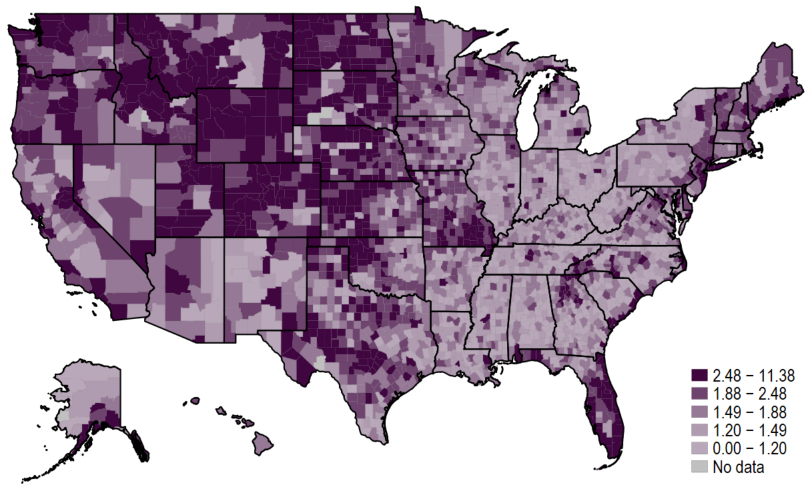
Notes: County regressions include commuting zone (CZ) × year FE and tract regressions include county × year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the CZ level for county regressions and at the county level for tract regressions.

Table 11: Tract-Level Linear Probability Model (LPM) Estimates

	(1) WBA transition	(2) BA transition
log(median age)	0.003	–0.004
	(0.00722)	(0.00343)
bachelors or higher share	0.0644***	0.0210***
	(0.0097)	(0.00429)
some college share	–0.0332***	–0.0155***
	(0.0124)	(0.00433)
African American share	–0.159***	–0.0491***
	(0.00698)	(0.00226)
Asian share	0.012	0.0263***
	(0.0193)	(0.00985)
Hispanic share	–0.022	–0.011
	(0.0156)	(0.00652)
foreign born share	–0.0334**	–0.0188***
	(0.0142)	(0.00685)
log(per capita income)	0.004	0.002
	(0.00376)	(0.00189)
emp-pop ratio	–0.0121*	–0.002
	(0.00664)	(0.00287)
owner-occupied share	0.0236***	0.00800***
	(0.00359)	(0.00149)
share of emp in young firms	0.0145***	0.0102***
	(0.00538)	(0.00187)
share of emp in large firms	–0.0100**	–0.00667***
	(0.00394)	(0.00169)
DHS(avg firm emp)	–0.00811***	–0.00396***
	(0.00108)	(0.000335)
commercial share	0.0794***	0.0347***
	(0.00384)	(0.00204)
Ind emp. shares	yes	yes
Observations	2,355,000	13,840,000
R-squared	0.113	0.2
Within R-squared	0.0098	0.0777

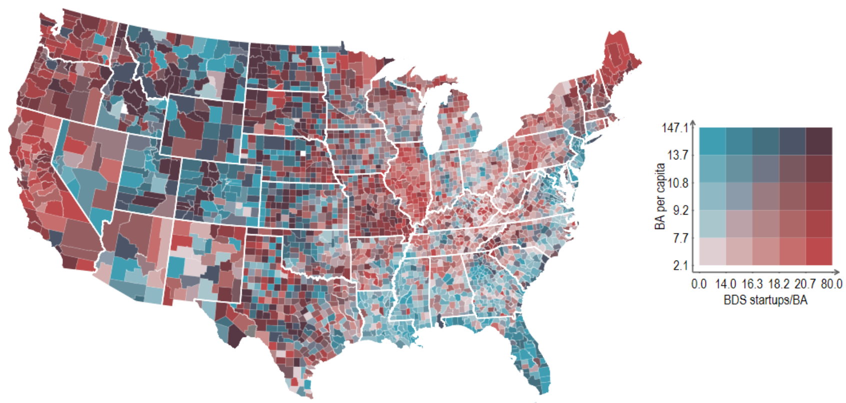
Notes: Regressions include county =× year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the county level.

Figure 1: Startups per capita, by county



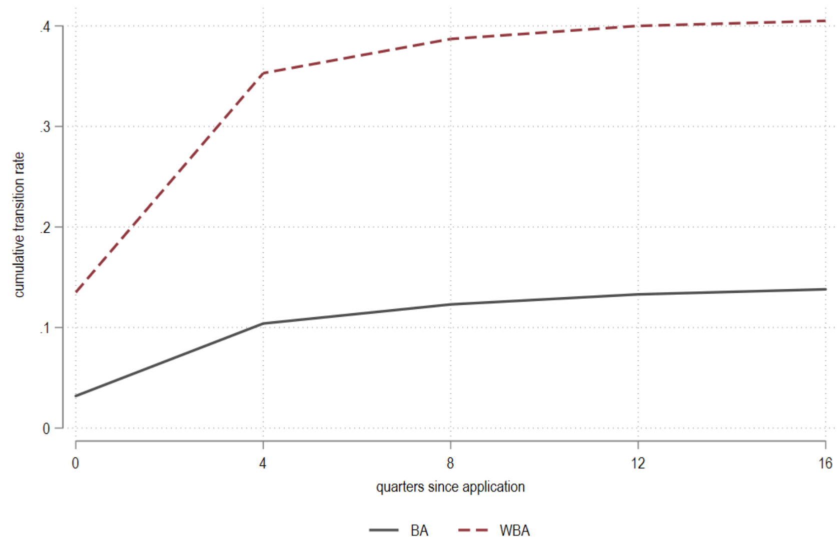
Notes: Depicts average BDS startups per 1,000 prime-age (20-64 years old) people (startups per capita) at the county level between 2012 and 2018. BDS startups are defined as age 0 firms. Note that to account for the time it takes applications to transition into employer businesses (around 8 quarters), we interpret startups between 2012 and 2018 as arising from applications filed between 2011 and 2016.

Figure 2: Applications per capita and transition rates, by county



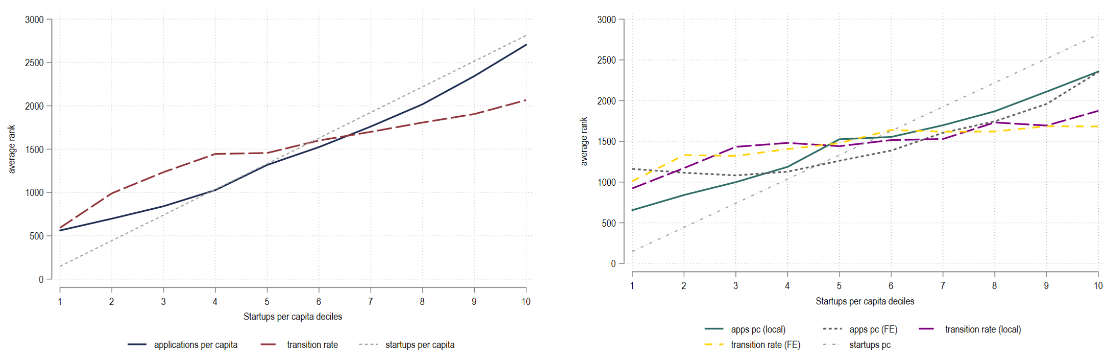
Notes: Depicts a bivariate map of average BA per 1,000 prime-age (20-64 years old) people (applications per capita) and BA divided by BDS startups (transition rates) at the county level between 2011 and 2016. BDS startups are defined as age 0 firms. Note that to account for the time it takes applications to transition into employer businesses (around 8 quarters), we lag the denominator by two years.

Figure 3: Cumulative transition rates for BA and WBA, time Since application



Notes: Depicts the cumulative transition rate of BA and WBA between 0 and 16 quarters after application. Due to disclosure considerations, the cumulative transition rates are calculated only for quarters 0, 4, 8 and 16. Transitions (startups) are defined as applications that transition to an employer business within eight quarters after application.

Figure 4: County-level Rank Analysis (deciles of startups per capita)

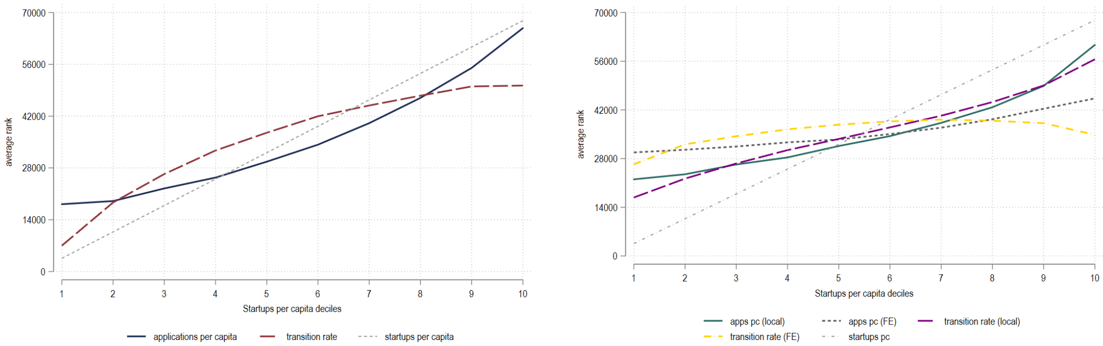


(a) Apps pc & transition rates

(b) Estimated local conditions and FEs

Notes: The rank analysis focuses on WBA and associated transitions. "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). In both figures, the x-axis is the deciles of startups pc (averaged at the county-level over 2011-2016). Figure (a) depicts the average rank, by decile, of applications pc (solid blue), transition rate (dashed red), and startups pc (dotted grey). Figure (b) depicts the average rank based on predicted applications pc from local conditions (solid green), commuting zone fixed effects (tight dash grey), transition rate local conditions (long dashed purple), transition rate commuting zone conditions (long dash-dot yellow), and raw startups pc (short dash-dot grey).

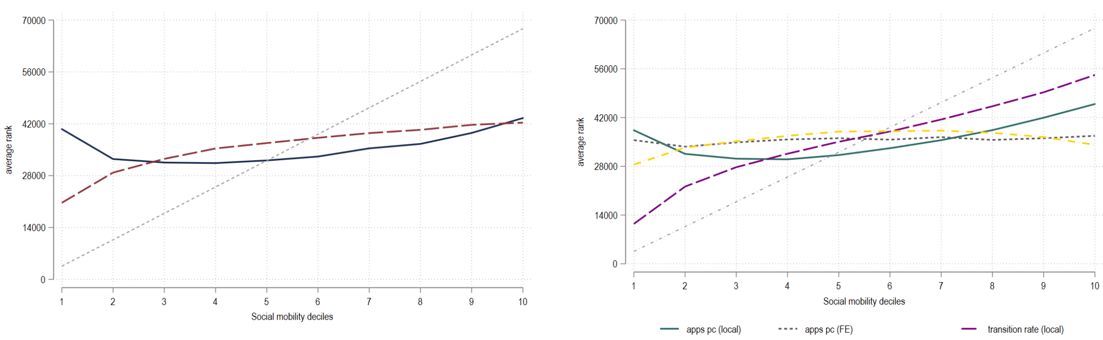
Figure 5: Tract-level Rank Analysis



(a) Apps pc and transition rates (b) Estimated local conditions and FEs

Notes: The rank analysis focuses on WBA and associated transitions. "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). In both figures, the x-axis is the deciles of startups pc (averaged at the tract-level over 2011-2016). Figure (a) depicts the average rank, by decile, of applications pc (solid blue), transition rate (dashed red), and startups pc (dotted grey). Figure (b) depicts the average rank based on predicted applications pc from local conditions (solid green), county fixed effects (tight dash grey), transition rate local conditions (long dashed purple), transition rate county conditions (long dash-dot yellow), and raw startups pc (short dash-dot grey).

Figure 6: Tract-level Social Mobility Rank Analysis



(a) Apps pc and transition rates (b) Estimated local conditions and FEs

Notes: The rank analysis focuses on WBA and associated transitions. "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). In both figures, the x-axis is the deciles of social mobility (Chetty et. al., 2018). Figure (a) depicts the average rank, by decile, of applications pc (solid blue), transition rate (dashed red), and startups pc (dotted grey). Figure (b) depicts the average rank based on predicted applications pc from local conditions (solid green), county fixed effects (tight dash grey), transition rate local conditions (long dashed purple), transition rate county conditions (long dash-dot yellow), and raw startups pc (short dash-dot grey).

Appendices

A Estimated elasticities and magnitudes

A.1 Elasticities

Consider the relationships (14), (15), and (16). The transformation of startups per capita is

$$\widetilde{S}_{lzt} = 2\left(\frac{S_{lzt} - \overline{S}}{S_{lzt} + \overline{S}}\right), \tag{18}$$

where $\overline{S}$ is the grand average over locations l and time t

$$\overline{S} = \frac{1}{N}\sum_{t}\sum_{l\in z} S_{lzt}. \tag{19}$$

We can write the analogous transformation, $\widetilde{A}_{lzt}$, for applications per capita, A_{lzt}. The transition rate, T_{lzt}, is untransformed.

To get the point elasticity of the original variable S_{lzt} with respect to any covariate c_{lt-k} expressed in levels, we proceed by differentiating (18). Note that $\overline{S}$ in (19) is a function of S_{lzt}, which depends on c_{lt-k}. However, the linear model (14) assumes that $S_{kz\tau}$ does not depend on c_{lt-k} for $k \neq l$ and $\tau \neq t$. Thus,

$$\frac{\partial \overline{S}}{\partial c_{lt-k}} = \frac{1}{N}\frac{\partial S_{lzt}}{\partial c_{lt-k}}.$$

Differentiation of (18) yields

$$\begin{aligned}
\frac{\partial}{\partial c_{lt-k}} 2\left(\frac{S_{lzt} - \overline{S}}{S_{lzt} + \overline{S}}\right) &= 2\frac{\frac{\partial S_{lzt}}{\partial c_{lt-k}}\left[(1-\frac{1}{N})(S_{lzt}+\overline{S}) - (1+\frac{1}{N})(S_{lzt}-\overline{S})\right]}{\left(S_{lzt}+\overline{S}\right)^2} \\
&= -\frac{4}{N}\frac{\left(S_{lzt} - N\overline{S}\right)}{\left(S_{lzt}+\overline{S}\right)^2}\frac{\partial S_{lzt}}{\partial c_{lt-k}} \\
&= \beta_c^S.
\end{aligned}$$

The elasticity of the original variable is then

$$\begin{aligned}
\epsilon_c^S(S_{lzt}, c_{lt-k}) &= \frac{\partial S_{lzt}}{\partial c_{lt-k}}\frac{c_{lt-k}}{S_{lzt}} \\
&= \frac{\left(S_{lzt}+\overline{S}\right)^2}{-\frac{4}{N}\left(S_{lzt} - N\overline{S}\right)}\beta_c^S\frac{c_{lt-k}}{S_{lzt}}.
\end{aligned}$$

For N large (which is the case in our application because (l, t) pairs constitute a large sample), we can write

$$-\frac{4}{N}\left(S_{lzt}-N\overline{S}\right)\simeq 4\overline{S},$$

and hence we can approximate $\epsilon_c^S(S_{lzt},c_{lt-k})$ using

$$\epsilon_c^S(S_{lzt},c_{lt-k})\simeq\frac{1}{4}\beta_c^S\frac{\left(S_{lzt}+\overline{S}\right)^2}{\overline{S}}\frac{c_{lt-k}}{S_{lzt}}.$$

Similarly, for applications per capita

$$\epsilon_c^A(A_{lzt},c_{lt-k})\simeq\frac{1}{4}\beta_c^A\frac{\left(A_{lzt}+\overline{A}\right)^2}{\overline{A}}\frac{c_{lt-k}}{A_{lzt}}.$$

The elasticity of the (untransformed) transition rate is

$$\epsilon_c^T(T_{lzt},c_{lt-k})=\beta_c^T\frac{c_{lt-k}}{T_{lzt}}.$$

Using the point elasticities above, the elasticites at the means $(\overline{S},\overline{c})$, $(\overline{A},\overline{c})$, $(\overline{T},\overline{c})$ are given by

$$\begin{aligned}\epsilon_c^S(\overline{S},\overline{c}) &\simeq \frac{1}{4}\beta_c^S\frac{\left(\overline{S}+\overline{S}\right)^2}{\overline{S}}\frac{\overline{c}}{\overline{\overline{S}}}=\frac{1}{4}\beta_c^S\frac{4\overline{S}^2}{\overline{S}}\frac{\overline{c}}{\overline{\overline{S}}}=\beta_c^S\overline{c},\\ \overline{\epsilon}_c^A(\overline{A},\overline{c}) &\simeq \beta_c^A\overline{c},\\ \overline{\epsilon}_c^T(\overline{T},\overline{c}) &= \beta_c^T\frac{\overline{c}}{\overline{\overline{T}}}.\end{aligned}$$

In our analysis, some covariates are expressed in logs or are transformed using (18). For these, we derive the elasticities with respect to the original (untransformed) covariate. For a covariate in logs, we have

$$\begin{aligned}\epsilon_c^S(\overline{S},\overline{c}) &\simeq \frac{1}{4}\beta_c^S\frac{4\overline{S}^2}{\overline{S}}\frac{1}{\overline{\overline{S}}}=\beta_c^S,\\ \overline{\epsilon}_c^A(\overline{A},\overline{c}) &\simeq \beta_c^A,\\ \overline{\epsilon}_c^T(\overline{T},\overline{c}) &= \beta_c^T\frac{1}{\overline{\overline{T}}}.\end{aligned}$$

For a covariate transformed using (18), we have

$$\frac{\partial}{\partial c_{lt-k}}2\left(\frac{S_{lzt}-\overline{S}}{S_{lzt}+\overline{S}}\right)=\frac{\partial}{\partial c_{lt-k}}\beta_c^S 2\left(\frac{c_{lt-k}-\overline{c}}{c_{lt-k}+\overline{c}}\right),$$

which implies

$$\frac{\left(S_{lzt}-N\overline{S}\right)}{\left(S_{lzt}+\overline{S}\right)^2}\frac{\partial S_{lzt}}{\partial c_{lt-k}}\frac{c_{lt-k}}{S_{lzt}}=\beta_c^S\frac{\left(c_{lt-k}-N\overline{c}\right)}{\left(c_{lt-k}+\overline{c}\right)^2}\frac{c_{lt-k}}{S_{lzt}},$$

and

$$\begin{aligned}\epsilon_c^S(S_{lzt}, c_{lt-k}) &= \beta_c^S \frac{\left(S_{lzt}+\overline{S}\right)^2}{\left(S_{lzt}-N\overline{S}\right)} \frac{(c_{lt-k}-N\overline{c})}{(c_{lt-k}+\overline{c})^2} \frac{c_{lt-k}}{S_{lzt}}, \\ &\simeq \beta_c^S \frac{\left(S_{lzt}+\overline{S}\right)^2}{\overline{S}} \frac{\overline{c}}{(c_{lt-k}+\overline{c})^2} \frac{c_{lt-k}}{S_{lzt}}.\end{aligned}$$

Therefore,

$$\epsilon_c^S(\overline{S}, \overline{c}) \simeq \beta_c^S \frac{\left(\overline{S}+\overline{S}\right)^2}{\overline{S}} \frac{\overline{c}}{(\overline{c}+\overline{c})^2} \frac{\overline{c}}{\overline{S}} = \beta_c^S,$$

and similarly

$$\epsilon_c^A(\overline{A}, \overline{c}) \simeq \beta_c^A.$$

Finally,

$$\begin{aligned}\epsilon_c^T(T_{lzt}, c_{lt-k}) &= \frac{\partial T_{lzt}}{\partial c_{lt-k}} \frac{c_{lt-k}}{T_{lzt}} = \frac{\partial}{\partial c_{lt-k}} \left\{\beta_c^T 2\left(\frac{c_{lt-k}-\overline{c}}{c_{lt-k}+\overline{c}}\right)\right\} \frac{c_{lt-k}}{T_{lzt}} \\ &= -\beta_c^T \frac{4}{N} \frac{(c-N\overline{c})}{(c+\overline{c})^2} \frac{c_{lt-k}}{T_{lzt}},\end{aligned}$$

which implies

$$\epsilon_c^T(\overline{T}, \overline{c}) \simeq \beta_c^T \frac{4\overline{c}}{(\overline{c}+\overline{c})^2} \frac{\overline{c}}{\overline{T}} = \beta_c^T \frac{1}{\overline{T}}.$$

The above elasticities can be estimated by replacing the unknown parameters (β's) with their estimates, yielding $\widehat{\epsilon}_c^Y(\overline{Y}, \overline{c})$ for $Y = S, A, T$.

A.2 Quantification of the magnitudes

Note that for any covariate c the estimated percent change in $Y = S, A, T$ induced by the percent change in c equivalent to one standard deviation multiple of the mean is given by

$$\widehat{\Delta} Y_l = \widehat{\epsilon}_c^Y(\overline{Y}, \overline{c})(100 \times s_c/m_c) = \widehat{\epsilon}_c^Y(\overline{Y}, \overline{c})(100 \times CV_c),$$

where s_c, m_c, and $CV_c = s_c/m_c$ denote the sample standard deviation, mean, and the coefficient of variation for the untransformed covariate c, respectively.

A.3 Estimates of elasticities and magnitudes at the county level

We compute the implied percentage change in a dependent variable of interest corresponding to a one standard deviation change in the covariate relative to the mean (in percent). This requires converting the estimates to elasticities as described in Appendix A.[43] Given that the covariates differ significantly in their variation across locations, we quantify their

[43] The elasticities can be found in Table **??** in Appendix B.

economic significance by taking into account this variation and multiplying each elasticity with the coefficient of variation of the corresponding covariate. This quantification exercise is summarized in Table ??.

For startups originating from WBA, bachelors+ degree share, per capita income, and the employment-to-population ratio are, in order, the variables associated with the highest positive percentage change, followed by foreign-born share and median age. The highest negative percentage changes are observed for concentration of employment in incumbent businesses and African American share. Turning to WBA per capita, the variables associated with highest percent changes are per capita income, bachelors+ degree share, and the employment-to-population ratio, followed by African American share and median age. Foreign born share has a lower positive association, compared to the case of startups. Similarly, concentration of employment in incumbent businesses is now associated with a much smaller negative percent change. For transition rate of WBA, foreign born share, bachelors+ degree share, and SME loans are the covariates that are associated with the highest percentage changes. On the negative side, African American share stands out, with a one standard deviation increase in the African American share being associated with a nearly a 10 percent decline in transition rate. Hispanic share is also associated with a high negative percentage change.

The patterns are largely similar when we consider BA and the three outcomes. At first glance, this might seem surprising since BA includes applications with the intent to form a nonemployer business. However, as we have noted EIN based nonemployer businesses are the minority of nonemployer businesses that are substantially larger than sole proprietor nonemployer businesses. The relatively similar results for BA and WBA might imply that the determinants of those pursuing more substantive nonemployer businesses are not so different for those pursuing a new employer business.

A.4 Estimates of the elasticities and magnitudes at the tract level

Table ?? reports the quantification exercise described in Appendix A to assess the impact of a one standard deviation change in a control variable (relative to the mean) on the business formation measures.[44] The share of population with bachelors or graduate degree and median age are associated with relatively large percentage changes, especially in startups and business applications. African American share in a neighborhood is associated with a relatively large percent change in transition rates that also leads to a relatively large negative change in startups. The share of owner-occupied housing is also associated with a large negative change in startup and application activity.

A.5 Decomposition of Aggregate Time Series Variation in WBA Startups

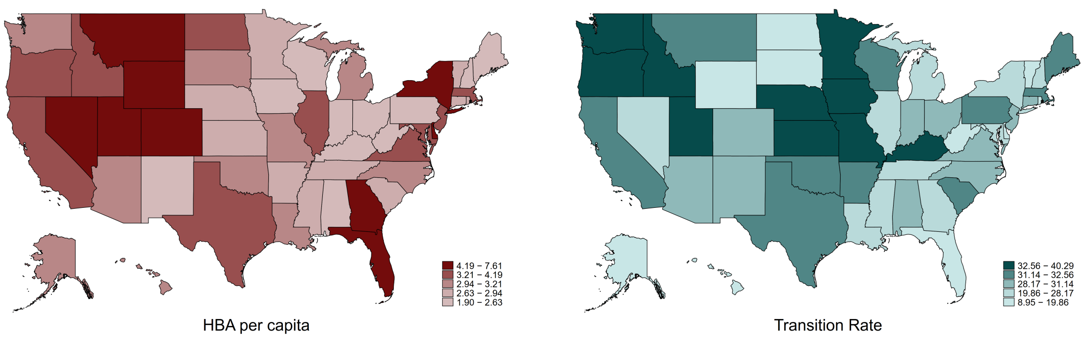
Even though less than 50% of WBA applications transition successfully to employer businesses, most of the time series variation at the aggregate level in employer startups that emerge from WBA is accounted for by time series variation in WBA rather than time series

[44] The elasticities underlying the exercise can be found in Table ?? in Appendix B.

variation in transition rates. Using time aggregated versions of the micro data used in our analysis, Table B.2.i illustrates this showing that slightly more than 100% of the variation in employer startups from WBA is accounted for by variation in WBA.

B Additional tables and figures

Figure B.1: Applications (HBA) Per Capita & Transition Rate, by State



(a) Applications per capita, HBA

(b) Transition rate

Notes: Depicts average HBA per 1,000 prime-age (20-64 years old) people (applications per capita) in the left figure and transition rate (startups divided by applications) in the right figure at the state level between 2011 and 2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

B.1 Additional Statistics and Variable Descriptions

Table B.1.a: Percent of WBA and BA that transition in the same location as application

	(1) WBA	(2) BA
County	91.2	90.3
Tract	79.4	77.7

Notes: Reports the percent of WBA and BA between 2011 and 2016 that transition in the same county or tract as the one in which the application was filed. Transitions (or startups) are defined as applications that transition to an employer business within eight quarters since application.

Table B.1.b: Contribution of location fixed effects: BA

	(1) County	(2) Tract
DHS(BA startups pc)	0.447	0.196
DHS(BA pc)	0.610	0.450
BA transition rate	0.402	0.149

Notes: Reports the share of variance in DHS(BA startups pc), DHS(BA pc), and BA transition rates accounted for by commuting zone by year FE (col. 1) and county by year FE (col. 2) for data at the county-year and tract-year levels in 2011-2016, respectively. "pc" refers to per capita (per 1000 prime-age people).

Table B.1.c: Regression Variable Summary Statistics: Additional Outcome Variables

	County			Tract		
	(1) Mean	(2) SD	(3) CV	(4) Mean	(5) SD	(6) CV
DHS(BA startups pc)	-0.160	0.562	-3.503	-0.484	0.906	-1.872
DHS(BA pc)	-0.085	0.373	-4.377	-0.282	0.599	-2.127
BA transition rate	0.128	0.053	0.414	0.117	0.091	0.779
DHS(high-tech WBA startups pc)	-0.623	1.080	-1.735	-0.843	1.287	-1.526
DHS(high-tech WBA pc)	-0.970	1.287	-1.327	-1.491	1.191	-0.799
high-tech WBA transition rate	0.417	0.363	0.870	0.386	0.450	1.166
average BA transition duration	1.606	0.683	0.425	1.893	1.304	0.689
average WBA transition duration	1.177	0.650	0.553	1.416	1.282	0.905

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the mean, standard deviation (SD), and coefficient of variation (CV) of additional dependent variables considered in county and tract level regressions. The years covered are 2011-2016. Startups are defined as applications that transition to an employer business within eight quarters after application. "DHS" refers to the transformation based on Davis et al. (1996).

Table B.1.d: Regression Variable Summary Statistics: Neighboring Tract Variables

	Neighboring Tracts		
	(1)	(2)	(3)
	Mean	SD	CV
log(median age)	3.643	0.143	0.039
bachelors or higher share	0.278	0.156	0.563
some college share	0.287	0.062	0.215
African American share	0.136	0.197	1.447
Asian share	0.045	0.077	1.689
Hispanic share	0.151	0.191	1.263
foreign born share	0.122	0.124	1.019
log(per capita income)	10.180	0.371	0.036
emp-pop ratio	0.581	0.081	0.139
owner-occupied share	0.652	0.173	0.265
share of emp in young firms	0.172	0.070	0.404
share of emp in large firms	0.066	0.079	1.187
DHS(avg firm emp)	-0.145	0.486	-3.354
commercial share	0.188	0.112	0.593

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the mean, standard deviation (SD), and coefficient of variation (CV) of neighboring tract variables considered as part of robustness regressions. The years covered are 2011-2016. Startups are defined as applications that transition to an employer business within eight quarters after application. "DHS" refers to the transformation based on Davis et al. (1996).

Table B.1.e: Description of Local Condition Variables

Variable	Definition	Source	County-Analysis	Tract-Analysis
log(median age)	log of median age	ACS	✓	✓
BA or higher share	share of pop. with BA or higher degree	ACS	✓	✓
some college share	share of pop. with some college	ACS	✓	✓
African American share	African American pop. share	ACS	✓	✓
Asian share	Asian pop.share	ACS	✓	✓
Hispanic share	Hispanic pop. share	ACS	✓	✓
foreign born share	foreign both pop. share	ACS	✓	✓
log(per capita income)	log of per capita income	ACS	✓	✓
emp-pop ratio	employment to pop. ratio	ACS	✓	✓
owner-occupied share	share of owner-occupied housing units	ACS	✓	✓
log(debt-to-income)	log of household debt to income ratio	FRB	✓	
share of emp in young firms	share of emp. in firms aged 1-5	LBD	✓	✓
share of emp in large firms	share of emp. in firms with 500+ emp	LBD	✓	✓
DHS(average emp)	DHS of the average emp. of firms	LBD	✓	✓
Industry emp. shares	3-digit NAICS employment shares	LBD	✓	✓
DHS(SME loans/emp)	DHS of the SBL to firms with <\$1 mn rev. divided by the emp. in firms with <500 emp.	CRA & LBD	✓	
commercial share	emp. share = emp./ (pop. + emp.)	ACS & LBD		✓

Notes: SBL stands for small business loans. "DHS" refers to the transformation based on Davis et al. (1996), where the deviation is taken from the grand mean. FRB stands for the Federal Reserve Board. CRA stands for the Community Reinvestment Act.

B.2 Supplementary regression analysis

Table B.2.a: BA Baseline Regression Results

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
log(median age)	0.659***	0.658***	0.0167	0.00915	0.125***	0.00286
	(0.107)	(0.0732)	(0.0101)	(0.0391)	(0.0325)	(0.00298)
bachelors or higher share	1.555***	1.118***	0.0549***	0.423***	0.241***	0.0151***
	(0.212)	(0.132)	(0.0167)	(0.0492)	(0.0377)	(0.00438)
some college share	–0.0958	0.139	–0.0282	–0.201***	–0.0658	–0.0250***
	(0.235)	(0.14)	(0.0222)	(0.0557)	(0.0494)	(0.00463)
African American share	–0.224**	0.573***	–0.0807***	–0.239***	0.621***	–0.0656***
	(0.087)	(0.059)	(0.00792)	(0.0672)	(0.0651)	(0.00369)
Asian share	–1.576**	–0.961**	–0.104*	0.103	–0.146	0.0519***
	(0.65)	(0.477)	(0.0541)	(0.154)	(0.0943)	(0.0144)
Hispanic share	–0.249	–0.0533	–0.0560***	–0.179**	–0.114	–0.00234
	(0.162)	(0.112)	(0.0199)	(0.0799)	(0.0923)	(0.00639)
foreign born share	1.182***	0.706***	0.0825*	0.571***	0.438***	0.00149
	(0.366)	(0.255)	(0.0499)	(0.127)	(0.131)	(0.00678)
log(per capita income)	0.168	0.129**	–0.00219	0.363***	0.335***	0.000992
	(0.102)	(0.0596)	(0.00894)	(0.02)	(0.017)	(0.00169)
emp-pop ratio	0.827***	1.201***	0.00218	–0.196***	–0.207***	–0.00960***
	(0.223)	(0.138)	(0.0191)	(0.0525)	(0.0372)	(0.00337)
owner-occupied share	0.0438	–0.143	0.00577	0.0274	–0.0441	0.0105***
	(0.176)	(0.0999)	(0.0141)	(0.0311)	(0.0288)	(0.00227)
log(debt-to-income ratio)	–0.00412	–0.00157	–0.00372**			
	(0.0191)	(0.0105)	(0.00167)			
share of emp in young firms	0.0238	0.360***	–0.019	0.249***	0.164***	0.00859***
	(0.21)	(0.109)	(0.0188)	(0.0253)	(0.0164)	(0.00223)
share of emp in large firms	–0.0231	0.162***	–0.0181**	–0.563***	–0.373***	–0.0275***
	(0.0932)	(0.0512)	(0.00834)	(0.0225)	(0.0136)	(0.00174)
DHS(avg firm emp)	0.0722	–0.132***	0.00771	–0.176***	–0.133***	–0.00839***
	(0.0615)	(0.0314)	(0.00562)	(0.00631)	(0.00444)	(0.00049)
DHS(SME loans/employment)	0.0176	0.00622	–0.000259			
	(0.0152)	(0.00924)	(0.00141)			
commercial share				2.599***	1.783***	0.119***
				(0.028)	(0.0283)	(0.003)
Ind emp. shares	yes	yes	yes	yes	yes	yes
Observations	17,500	17,500	17,500	428,000	430,000	428,000
Fixed effects	cz x yr	cz x yr	cz x yr	fips x yr	fips x yr	fips x yr
SE clustering	cz	cz	cz	fips	fips	fips
R-squared	0.5753	0.7693	0.4422	0.4296	0.6836	0.2127
Within R-squared	0.2323	0.4087	0.06711	0.291	0.4253	0.07462

Notes: County regressions include commuting zone (CZ) × year FE and tract regressions include county × year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the CZ level for county regressions and at county level for tract regressions.

Table B.2.b: BA Regression Magnitudes

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
median age	8.303	8.291	1.638	0.176	2.450	0.470
bachelors or higher share	13.530	9.722	3.718	7.847	4.456	2.394
some college share	-0.507	0.742	-1.176	-1.607	-0.526	-1.690
African American share	-3.402	8.586	-9.396	-5.372	13.838	-12.536
Asian share	-3.958	-2.500	-2.083	0.965	-1.350	3.858
Hispanic share	-3.335	-0.794	-5.876	-3.715	-2.339	-0.413
foreign born share	6.593	3.981	3.608	7.798	6.016	0.223
per capita income	3.982	3.057	-0.403	18.803	17.353	0.414
emp-pop ratio	6.780	9.849	0.134	-2.075	-2.184	-0.874
owner-occupied share	0.349	-1.123	0.360	0.634	-0.986	2.042
debt-to-income ratio	-0.178	-0.089	-1.290	.	.	.
share of emp in young firms	0.133	1.901	-0.796	3.102	2.044	0.916
share of emp in large firms	-0.315	2.205	-1.995	-8.908	-5.854	-3.818
avg firm emp	2.714	-4.976	2.262	-25.819	-19.511	-10.562
SME loans/employment	2.453	0.818	-0.273	.	.	.
commercial share	.	.	.	43.989	30.177	17.217

Notes: Reports the estimated % change in the LHS variable induced by the percent change in the RHS variable equivalent to a one standard deviation multiple of the mean. The LHS variable of the regression is listed in the columns, and each RHS variable is listed in the rows. "DHS" refers to the transformation based on Davis et al. (1996). "pc" refers to per capita (per 1000 prime-age (20-64 years old) people).

Table B.2.c: BA Regression Decomposition

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Groups						
Demographic	0.083	0.165	0.025	0.043	0.046	0.033
HH economic conditions	0.047	0.129	-0.001	0.053	0.097	0.001
Incumbent firm characteristics	0.103	0.115	0.042	-0.010	-0.021	0.002
Commercial share	.	.	.	0.205	0.303	0.039
Categories						
Local conditions	0.232	0.409	0.067	0.291	0.425	0.075
Common market conditions	0.343	0.361	0.375	0.139	0.258	0.138
Residual	0.425	0.231	0.558	0.570	0.316	0.787

Notes: Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(BA startups pc), DHS(BA pc) and BA transition rate for county- and tract-level analysis. Note that all control variables are included, along with location × fixed effects. "DHS" refers to the transformation based on Davis et al. (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The fourth row is the sum of the contribution of all individual variables (or the sum of the contribution of all grouped variables), and corresponds to the within R^2; the fifth row is the contribution of location × year FE; and the last row is the remaining variation that is unexplained by either local conditions or common market conditions. Note that log(HH debt-to-income) is included in HH economic conditions and DHS(SME loans/employment) is included in incumbent firm characteristics for county level results only.

Table B.2.d: High-Tech WBA Regression Decomposition

	County			Tract		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Groups						
Demographic	0.086	0.097	0.013	0.014	0.035	0.008
HH economic conditions	0.018	0.007	-0.002	0.002	0.021	-0.000
Incumbent firm characteristics	0.064	0.147	0.028	0.004	0.008	0.003
Commercial share	.	.	.	0.027	0.026	0.001
Categories						
Local conditions	0.168	0.251	0.039	0.047	0.090	0.011
Common market conditions	0.421	0.289	0.456	0.089	0.103	0.087
Residual	0.411	0.460	0.505	0.864	0.808	0.902

Notes: Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(high-tech WBA startups pc), DHS(high-tech WBA pc) and high-tech WBA transition rate for county- and tract-level analysis. Note that all control variables are included, along with location × fixed effects. "DHS" refers to the transformation based on Davis et al. (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The fourth row is the sum of the contribution of all individual variables (or the sum of the contribution of all grouped variables), and corresponds to the within R^2; the fifth row is the contribution of location × year FE; and the last row is the remaining variation that is unexplained by either local conditions or common market conditions. Note that log(HH debt-to-income) is included in HH economic conditions and DHS(SME loans/employment) is included in incumbent firm characteristics for county level results only.

Table B.2.e: Tract-Level WBA Regression with Neighboring Tract Controls

	(1)	(2)	(3)
	DHS(startups pc)	DHS(applications pc)	Transition rate
log(median age)	0.0211	0.0678***	0.00446
	(0.025)	(0.0231)	(0.00545)
bachelors or higher share	0.319***	0.0285	0.0963***
	(0.0344)	(0.0298)	(0.00957)
some college share	–0.210***	–0.214***	–0.0312***
	(0.0405)	(0.0371)	(0.0112)
African American share	–0.265***	0.346***	–0.156***
	(0.0304)	(0.0242)	(0.00679)
Asian share	0.0393	0.127**	–0.0176
	(0.0603)	(0.0538)	(0.0155)
Hispanic share	–0.0660*	0.0365	–0.00874
	(0.0401)	(0.0405)	(0.0093)
foreign born share	0.204***	0.198***	–0.00241
	(0.0582)	(0.0506)	(0.0118)
log(per capita income)	0.249***	0.303***	0.00507
	(0.0152)	(0.0124)	(0.00408)
emp-pop ratio	–0.281***	–0.123***	–0.0359***
	(0.0348)	(0.0355)	(0.0081)
owner-occupied share	0.0139	0.0133	0.0220***
	(0.0183)	(0.0188)	(0.00429)
share of emp in young firms	0.204***	0.175***	0.00512
	(0.0242)	(0.0184)	(0.00623)
share of emp in large firms	–0.530***	–0.545***	–0.0248***
	(0.0237)	(0.0172)	(0.0048)
DHS(avg firm emp)	–0.175***	–0.142***	–0.00984***
	(0.00558)	(0.00521)	(0.0014)
commercial share	2.642***	2.400***	0.155***
	(0.0305)	(0.0269)	(0.00583)
Ind emp. shares	yes	yes	yes
Neighboring tract controls	yes	yes	yes
Observations	398,000	430,000	398,000
Fixed effects	fips x yr	fips x yr	fips x yr
SE clustering	fips	fips	fips
R-squared	0.308	0.4209	0.1457
Within R-squared	0.222	0.2685	0.04223

Notes: Regressions include county × year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the county level.

Table B.2.f: WBA Regression Decomposition, with Neighboring Tract Controls

	(1) DHS(startups pc)	(2) DHS(applications pc)	(3) Transition rate
Groups			
Demographic	0.025	0.003	0.024
Residual economic conditions	0.025	0.034	0.001
Incumbent firm characteristics	-0.010	-0.019	0.002
Commercial share	0.170	0.232	0.007
Categories			
Local conditions	0.209	0.251	0.035
Neighboring tract conditions	0.013	0.018	0.007
Common market conditions	0.086	0.152	0.104
Residual	0.692	0.579	0.854

Notes: Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(high-tech WBA startups pc), DHS(high-tech WBA pc) and high-tech WBA transition rate for county- and tract-level analysis. Note that all control variables are included, along with location × fixed effects. "DHS" refers to the transformation based on Davis et al. (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The fourth row is the sum of the contribution of all individual variables (or the sum of the contribution of all grouped variables), and corresponds to the within R^2; the fifth row is the contribution of location × year FE; and the last row is the remaining variation that is unexplained by either local conditions or common market conditions. Note that log(HH debt-to-income) is included in HH economic conditions and DHS(SME loans/employment) is included in incumbent firm characteristics for county level results only.

Table B.2.g: Average WBA Transition Duration Analysis

	(1)	(2)
	County	Tract
log(median age)	–0.0279	–0.0556**
	(0.119)	(0.0246)
bachelors or higher share	0.0382	0.342***
	(0.210)	(0.0442)
some college share	–0.00211	0.219***
	(0.222)	(0.0517)
African American share	0.243**	0.230***
	(0.124)	(0.0313)
Asian share	0.327	0.146**
	(0.416)	(0.0672)
Hispanic share	0.322*	–0.0709
	(0.189)	(0.0686)
foreign born share	–0.154	0.225***
	(0.395)	(0.0808)
log(per capita income)	0.258**	0.0373*
	(0.110)	(0.0192)
emp-pop ratio	–0.307	0.0763*
	(0.228)	(0.0399)
owner-occupied share	–0.143	–0.0565***
	(0.153)	(0.0199)
log(debt-to-income ratio)	0.0437**	
	(0.0198)	
share of emp in young firms	0.255	0.0971***
	(0.224)	(0.0294)
share of emp in large firms	–0.0677	0.0112
	(0.102)	(0.0232)
DHS(avg firm emp)	0.269***	0.0109
	(0.0556)	(0.00706)
DHS(SME loans/employment)	0.0451***	
	(0.0161)	
commercial share		–0.146***
		(0.0246)
Observations	17,000	309,000
Ind emp. shares	yes	yes
Fixed effects	cz x yr	fips x yr
SE clustering	cz	fips
R-squared	0.334	0.078
Within R-squared	0.02538	0.003109

Notes: County regressions include commuting zone (CZ) × year FE and tract regressions include county × year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the CZ level for county regressions and at county level for tract regressions.

Table B.2.h: Average BA Transition Duration Analysis

	(1) County	(2) Tract
log(median age)	–0.203	–0.0376
	(0.130)	(0.0236)
bachelors or higher share	0.085	0.388***
	(0.219)	(0.0414)
some college share	0.334	0.269***
	(0.241)	(0.0465)
African American share	0.119	0.139***
	(0.134)	(0.0303)
Asian share	0.368	–0.0543
	(0.429)	(0.0631)
Hispanic share	0.500**	–0.0873
	(0.195)	(0.0559)
foreign born share	–0.475	0.264***
	(0.419)	(0.0686)
log(per capita income)	0.333***	0.0238
	(0.118)	(0.0201)
emp-pop ratio	–0.143	0.187***
	(0.242)	(0.0378)
owner-occupied share	–0.109	–0.0464**
	(0.169)	(0.0194)
log(debt-to-income ratio)	0.0477**	
	(0.0223)	
share of emp in young firms	0.473**	0.0273
	(0.215)	(0.0266)
share of emp in large firms	0.0335	0.0383*
	(0.0964)	(0.0218)
DHS(avg firm emp)	0.177***	0.0128*
	(0.0605)	(0.00662)
DHS(SME loans/employment)	0.0540***	
	(0.0165)	
commercial share		–0.329***
		(0.0273)
Observations	17,000	361,000
Ind emp. shares	yes	yes
Fixed effects	cz x yr	fips x yr
SE clustering	cz	fips
R-squared	0.3352	0.08729
Within R-squared	0.02833	0.004556

Notes: County regressions include commuting zone (CZ) × year FE and tract regressions include county × year FE. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the CZ level for county regressions and at county level for tract regressions.

Table B.2.i: Aggregate variance decomposition of WBA startups

	(1) Applications pc	(2) Transition rate	(3) 2 × covariance
Aggregate	1.076	0.065	-0.141

Notes: Reports the variance decomposition of log(WBA startups) into log(WBA) and log(WBA transition rate) for the period 2011-2016. Startups are defined as applications that transition to an employer business within eight quarters after application.