The Local Origins of Business Formation*

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

November 29, 2023

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, economic, financial, and business conditions accounts 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 high 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). We thank Richard Beem, Shawn Klimek, participants at 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. In particular, we observe neighborhoods with high rates of startups and some with very low rates. Are low startup rates driven by a lack of business ideas, the difficulty of turning ideas into actual businesses, or both? This question cannot be answered with relatively more common data on startups alone. Instead, information on the pool of potential entrants and the nature of their selection into employer startups is needed.

We study the early phases of entrepreneurship using unique and comprehensive micro data from the U.S. Census Bureau that contain information on the universe of applications for new businesses in the United States and their transition to employer startups over the period 2010-2016. The Census Bureau's Business Formation Statistics (BFS) program integrates administrative data on the universe of applications for Employer Identification Numbers (EINs) from the Internal Revenue Service with the universe of employer businesses in the Longitudinal Business Database (LBD).[3] The integrated data is used to construct a measure of nascent entrepreneurial activity based on EIN applications in a location and the transition rate of applications to employer businesses. We think of the count of applications as a proxy for the level of entrepreneurial ideas being generated in a location. While the application data is relatively parsimonious, it does provide sufficient information to create more refined

[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 (2010) provides a review of the literature on the spatial dispersion of entrepreneurship and start-up activity.

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

proxies of business ideas aimed at new employer business formation. We exploit these features of the BFS micro data to refine our measures of nascent entrepreneurial activity.

To guide the empirical analysis, we use 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. These two components exhibit substantial variation across location. This spatial variation and its determinants are the focus of our analysis.

We interpret business applications as part of a nascent entrepreneurship phase when new business ideas are being developed.[4] We calculate the transition rate as the fraction of business applications that turn into employer businesses within eight quarters of the application date. Many applications never transition to employer businesses, as most of the ideas for businesses either do not come to fruition or are destined to become nonemployer businesses. 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. However, we also consider the entire set of applications 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.[5]

Our interpretation of a business application reflecting a nascent entrepreneurship phase is supported by external evidence as well as our findings. Evidence from the Panel Study of Entrepreneurship Dynamics (PSED – see Reynolds, Carter, Gartner, and Greene (2004)) indicates that applying for an EIN is a common signal of nascent entrepreneurial activity. We also find that there is significant delay between applying for an EIN and transitioning to an employer startup, highlighting the uncertainty, time, and resources associated with

[4]Business applications in our data are for businesses with an EIN. Many sole-proprietors do not need an EIN to start a business. Hence, our data does not capture the latter type of business initiation activity. Nevertheless, many sole-proprietors apply for EINs for banking purposes and to establish a formal and credible business entity. Furthermore, in terms of the economic significance, sole-proprietors with EINs tend to be much larger than the ones without (Davis, Haltiwanger, Krizan, Jarmin, Miranda, Nucci, and Sandusky (2009)).

[5]Even when they do not signal they have planned wages, they can signal in other ways. For example, applications for new corporations and in specific industries have a significantly higher likelihood of becoming an employer business. We take advantage of the full set of information on the application form in our analysis below.

the nascent entrepreneurship phase. Finally, our results indicate that spatial variation in both applications and transitions helps account for spatial variation in startups. Moreover, observable factors have distinct relationships with the two components of the decomposition of variation in startup activity.

The public-domain BFS at the national level suggests most of the variation in employer startups is accounted for by variation in applications. We illustrate this in Figure 1 where we depict the variation in startups (measured as log(business formations within 8 quarters)), applications (measured here as log(applications for likely employers)), and the implied transition rate (measured as the log (business formations/likely employer applications)).[6] It is apparent that much of the variation in startups at the national level is accounted for by variation in applications. Using the simple decomposition we use later applied to the national data, 90% of the variation in startups is accounted for by variation in applications.[7]

In sharp contrast, we find that both variation in applications and transition rates contribute substantially to spatial variation in startups. Using the applications for planned wages (WBA), we find that there is enormous variation in startups per capita, applications per capita and transition rates across counties and tracts. The average county has mean startup rate per capita from WBA applications of 0.88 (defined as per 1,000 prime age adults for 2010-2016) with a standard deviation of 0.63. WBA applications per capita have a mean of 2.15 per capita with a standard deviation of 1.27. The mean transition rate is 0.41 with a standard deviation of 0.15. To put these statistics on comparable terms, the coefficients of variation are 0.72, 0.59 and 0.37 respectively for startups per capita, applications per capita, and transitions. Using a variance decomposition, 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.

In examining all startups from all applications (BA in BFS parlance), we find broadly similar patterns especially in terms of the extent of spatial variation as measured by coeffi-

[6] We use likely employers, the high propensity application (HBA) series, for this purpose since HBA includes several characteristics associated with the application leading to an employer startup. In our micro-based data analysis in this paper, we consider these characteristics directly (such as those with planned wages) as well as all applications as noted above. The startups and transitions in the figure include those that arise from all applications but Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (2018) find that only 3% of applications that are not HBA transit to employer businesses. Figure 1 uses national data from 2015:1-2018:12 as startups within 8 quarters are only available in the public domain BFS through this period.

[7] The simple decomposition is that $log(S) = log(A) + log(T)$ where S is startups, A is applications and T is transitions. It is worth noting that even though variation in applications accounts for most of the variation in startups at the national level, the transition rate is only about 30%. There is some variation in the transition rate at the national level but it is dominated by variation in applications.

cients of variation and the relative importance of applications and transition rate components of the variance decomposition of the spatial variation in startups. In addition, moving to the tract level, we find greater variation in all of these components. For example, the standard deviation in startups per capita between tracts is about 64 percent greater than the standard deviation in startups per capita between counties.

What accounts for this enormous geographic variation in startups per capita, applications per capita, and transition rates? Whether we examine WBA or BA, much of the geographic variation reflects within local-area variation in the sense that the between-county variation is mostly not accounted for by broader area (commuting zone) by year effects. Similarly, the between-tract variation is mostly not accounted for by between-county by year variation.

We quantify the fraction of the within variation at both the county and tract level that can be accounted for by local conditions. At the county level, we consider lagged demographic, economic, financial, and business conditions controlling for commuting zone by year effects. At the tract level, we consider lagged demographic and economic conditions we can measure at the tract level while controlling for county by year effects. While our analysis of the contribution of such factors cannot be interpreted as identifying causal factors, the descriptive statistics presented here can facilitate research on such mechanisms.

Although most of the between-county variation is within commuting zone by year, the fraction accounted for by commuting zone by year effects is notable. Using either WBA or BA, commuting zone by year effects account for 35-40 percent of between county variation in startups per capita, applications per capita, and transition rates. Many factors such as regulations, spatial shocks, local market shocks, agglomeration benefits, and market structure likely vary at commuting zone-year level. These factors are undoubtedly important but there is considerable residual variation which is our focus.

In our county-level analysis, we analyze why specific counties within a metro area have variation in startups per capita, applications per capita, and transition rates. We find using either WBA or BA that the observable factors we consider account for about 20 percent of the variation in startups per capita across counties after accounting for the contribution of commuting zone-year effects. Much of this is driven by the finding that observables account for about 39 percent of the variation in the creation of ideas (applications per capita) across counties. The observables are less successful in accounting for variation in transition rates across counties, although we do identify several important covariates.

We find systematic relationships between demographic factors and startups, applications, and transitions. Specifically, applications, startups, and transition rate are increasing in the fraction of the local population that has a bachelor's degree and the fraction foreign born.[8] Startups are negatively related to the share of the local population that is Black or African American (African American share). This pattern reflects offsetting effects of a positive relationship between applications and the African American share and a negative relationship for transitions.[9] There is an inverse relationship between the fraction of the local population that is Asian and startups, applications, and the transition rate. For the fraction of the population that is Hispanic, there is an inverse relationship with startups and transitions, but no significant association with applications.

Local economic conditions are associated more with startups through applications (ideas), but less so with the transition rate. Per capita income and the employment-to-population ratio are both positively related to startups and applications, but not significantly related to transitions. Household financial conditions only have a relationship with the transition rate, while small business lending is associated only with applications. On the business conditions side, a higher local concentration of employment across incumbent businesses is associated with lower startups, applications, and transition rate.

We also explore these patterns at the tract level controlling for county by year fixed effects. The tract-level findings highlight that neighborhood effects are an important source of variation in spatial variation in entrepreneurship. Much of this reflects demographic factors such as age, education, and race and ethnicity. In particular, bachelors degree, median age, median household income, foreign born share, and African American share are key local factors that have a positive association with business applications across tracts. As with the between county analysis, we find that tracts with a higher share of African Americans in the population exhibit lower startups per capita, resulting from the offsetting effects of a positive relationship with applications per capita but a significantly lower transition rate.

Our findings are generally robust to using either all applications or applications with planned wages. There are important quantitative differences across such specifications, but the qualitative patterns are robust. In our discussion below, we highlight the salient differences in the results for the two types of applications.

The overall low contribution of observables to variation in transition rates at the county

[8] For expositional convenience, we often use the abbreviation "startups" for startups per capita and "applications" for applications per capita.

[9] In an analysis of state business registrations from eight states during the Covid pandemic, Fazio, Guzman, Liu, and Stern (2021) also find that the dynamics of state business registrations vary significantly by race and socioeconomic status across neighborhoods. In particular, they observe that areas including a higher proportion of Black residents are associated with higher growth in state business registration rates between 2019 and 2020. In our analysis, we focus on cross-sectional differences in business formation across locations, and we are able to decompose the differences in employer business formation rate across the geography into its two components: business application rate and the transition rate of applications to employer businesses. Investigating the local variation in both applications and transitions during the pandemic is a rich area for future work.

and tract level prompts us to also conduct an application-level analysis of transition rates. This analysis adds to the local conditions we use at the tract-level a rich set of application-specific covariates including reason for applying, legal form of organization, LLC status, timing of application within the year, and detailed industry. We find that the relationship between tract-level covariates and transitions is similar to that in tract-level analysis, even after adding these application-specific controls that proxy for the quality of the underlying business idea. Moreover, we obtain an even lower R^2 at the application-level analysis compared to the tract level, consistent with idiosyncratic factors associated with business idea quality explaining much of the likelihood of transition. Perhaps, at the individual entrepreneur level, it is not surprising that there is much unexplained variation in transition rates since presumably there is large idiosyncratic variation in the quality of ideas that have been generated in the nascent entrepreneurship phase. However, our county, tract, and application level analyses imply that there is systematic variation at the local level in transition rates that is not accounted for by observables.

We regard our county-level and tract-level results as complements rather than substitutes. It turns out that results are quite similar (at least qualitatively) whether we explore between county or between tract variation. However, it was not clear ex ante that this pattern would emerge. The role of local conditions could conceivably be different at the neighborhood (tract) and county level. Moreover, our findings rule out that this similarity in results is driven by tracts within the same county sharing common characteristics. Our tract level results rely on within county variation as we control for county by year effects in the tract level analysis.

To provide perspective on the quantitative implications of our findings, we 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 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.

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

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.

The literature has also 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). Furthermore, models of entry with imperfect competition (Nocke, 2006; Asplund and Nocke, 2006) allow for local competition and market size to effect entry and selection. In general, in all of these models, entry occurs immediately and as long as the present discounted value of expected net profits exceed the costs of entry.

Because most models (at least implicitly) assume that idea creation and business entry occur at the same time, relatively little is known about the nascent phases of entrepreneurship. Entrepreneurs spend time and resources in what Reynolds, Carter, Gartner, and Greene (2004) describe as the "conceptual" and "gestational" period prior to the actual commencement of business operations. The time and resources spent are part of the cost of entry and enable entrepreneurs to learn about the productivity and value of their potential future business.

A key challenge to studying the early stages of business formation is identifying and tracking the population of nascent entrepreneurs (or pool of potential entrants). Through the Panel Study of Entrepreneurship Dynamics (PSED) Reynolds, Carter, Gartner, and Greene (2004) and Reynolds (2017) 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.[10] 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.

The findings of Reynolds, Carter, Gartner, and Greene (2004) and Reynolds (2017)

[10] PSED Wave I identified 4,000 and Wave II identified 1,500 individuals that satisfied this criteria.

help reinforce our interpretation of applications for new EINs as a proxy for business ideas. Though a small step in the nascent phase of entrepreneurship, applying for an EIN, nevertheless, is a universal signal of potential employer businesses. While the BFS does not contain the rich survey responses of the PSED, it does have three key advantages relative to the PSED. 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 business idea. Third, by linking the BFS to the universe of employer businesses (LBD), we accurately track the transition of ideas to startups.

In related work, Andrews, Fazio, Liu, Guzman, and Stern (2018) and Guzman and Stern (2020) use state business registries to study the quantity and quality of entrepreneurship for 32 states over the period 1988-2014. They use high impact outcomes (IPO or a high-profile merger) originating from state business registrations to assess the quality of entrepreneurship, and model these outcomes as a function of a set of registration characteristics. Our relative contribution 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.[11]

The importance of the nascent phase of entrepreneurship has received renewed attention based on the novel survey evidence in Bennett and Chatterji (2023) and Bennett and Robinson (2023). This novel survey is internet-based and voluntary but reweighted to be nationally representative in terms of key demographic characteristics. In Bennett and Robinson (2023), they 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 (including quitting their job, making a sale or hiring an employee) – only about 1/5th 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). A distinguishing feature of our analysis is the focus on spatial variation in startup activity, idea creation and transitions which our

[11] We note that the LBD and associated integrated data (e.g., COMPUSTAT) does permit examining the role of local conditions for high-impact businesses. We leave that for future work.

comprehensive administrative data enables.

Following up on this latter point, our research also contributes to the literature on spatial aspects of entrepreneurship. Glaeser, Rosenthal, and Strange (2010) 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.

3 A model of business ideas and startups

The model highlights pre-entry heterogeneity among potential entrants in the quality of latent business ideas. It explores how the creation and the pursuit of an idea and the idea's success in generating a startup are related to conditions potential entrepreneurs face in a location, as well as the idiosyncratic idea quality. Startup formation involves two distinct decisions: the decision to explore an idea and the decision to start an employer business. Additional information about the viability of an idea is revealed in the gestational state during which the potential entrant pursues the idea. The model's analysis indicates that local conditions can play different 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 $l \in \mathcal{L}$ reside a continuum of N_l individuals, each of whom has an idea, $\iota \in [0, \infty)$, for an employer business.[12] Higher values of ι indicate better (or higher quality) ideas in terms of 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 the major tasks of estimating demand and costs, understanding relevant regulations, seek-

[12] We assume that all ideas come from the local population and are aimed for potential businesses in the same location. See discussion below on this issue empirically.

ing professional advice, looking for potential employees and suppliers, as well as the more mundane, but necessary task of applying for an EIN. Over the course of this 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. The signal 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 cost, fixed and variable costs, the degree of competition, as well as his own productivity or ability.[13]

The value of V depends on the quality of the idea. The dependence of V on ι is summarized by the conditional distribution $G_l(V|\iota)$.[14] Each idea owner can also choose to not pursue the idea and obtain a return of $R_l > 0$, which represents income from salary work or a nonemployer business. For simplicity, R_l and I_l do not depend on ι (and hence, on V).[15]

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 $\mathcal{E}_l$.[16]

After observing V, an idea owner decides whether to start an employer business. Each idea owner secures the return $R_l > 0$ if he does not pursue an employer business. The expected return from pursuing an idea is then

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

[13]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, 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.

[14]For example, a better idea (higher ι) may be associated with a higher V in a first order stochastic dominance sense.

[15]Both R_l and I_l could depend on ι. For instance, individuals with a higher ι may garner higher wages in the labor market, and may incur higher investment to pursue the idea. These dependencies can be incorporated without altering the results below, at the expense of additional notation and conditions.

[16]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).

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.[17] As long as $R_l + I_l < \mathcal{V}_l(\iota)$ for some ι, there exists a threshold (marginal idea) $\iota_l^* \in [0, \infty)$, such that all ideas in $[\iota_l^*, \infty]$ are pursued. The marginal idea then 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

[17] 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 decreasing in ι.

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.[18] The individual elements of $\mathcal{E}_l$ need not depend on all elements of C. For instance, the distribution of ideas, F_l, may depend on certain demographic characteristics of the population (e.g., age and education) that chooses to settle in location l in equilibrium, but the initial investment, I_l, may be independent of these demographics, and instead depend, for example, on laws and regulations pertaining to business initiation. We can then write the key equilibrium variables of interest, A_l, T_l, and S_l, explicitly as functions of the entire set of characteristics, C, i.e. $A_l \equiv A(C)$, $T_l \equiv T(C)$, and $S_l \equiv S(C)$, because they depend on $\mathcal{E}_l$ by the definitions (3)–(6).[19]

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.[20]

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

[18] 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.

[19] 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 the distribution of post-entry profits (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.

[20] 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.

$\frac{df_l(\iota)}{dc_l}$.[21] Similarly, $\frac{dT_l}{dc_l}$ can be positive or negative, and its sign can differ from that of $\frac{dA_l}{dc_l}$.[22] To see the mechanisms at work in more detail, rewrite (6) as

$$T_l = p_l(\iota_l^*) + \frac{\int_{i_l^*}^{\infty} \triangle p_l(\iota) f_l(\iota) d\iota}{A_l},$$

where the transition rate is expressed as the transition probability of the marginal idea and the average excess transition probability, $\triangle p_l(\iota) = p_l(\iota) - p_l(\iota_l^*)$, of supramarginal ideas. Then,

$$\frac{dT_l}{dc_l} = p_l'(\iota_l^*)\frac{d\iota_l^*}{dc_l} + \left.\frac{dp_l(\iota)}{dc_l}\right|_{\iota=\iota_l^*} + \frac{1}{A_l}\int_{i_l^*}^{\infty} \frac{d}{dc_l}\left[\triangle p_l(\iota) f_l(\iota)\right] d\iota - \frac{1}{A_l^2}\left(\int_{i_l^*}^{\infty} \triangle p_l(\iota) f_l(\iota) d\iota\right)\frac{dA_l}{dc_l}.$$

The change in T_l induced by a change in c_l consists of several partial effects. First, the marginal idea and its probability of transition changes, represented by the first and second terms. The excess transition probabilities of supramarginal ideas, and their distribution also change, represented by the third term. Finally, the fourth term is the partial effect of a change in A_l. If $p_l(\iota)$ is increasing, then $\triangle p_l(\iota) > 0$ for all supramarginal ideas.[23] In this case, $\frac{dT_l}{dc_l}$ is negatively associated with $\frac{dA_l}{dc_l}$ based on the fourth term alone. In other words, a change in A_l leads to a change in T_l that goes in the opposite direction, holding constant the marginal idea, its transition probability, and the total excess transition rate of supramarginal ideas. However, the first three terms can have any sign, and hence, the sign of $\frac{dT_l}{dc_l}$ in general differs from the sign of $\frac{dA_l}{dc_l}$. Note also that, by (7) the sign of $\frac{dS_l}{dc_l}$ depends on the signs and magnitudes of $\frac{dA_l}{dc_l}$ and $\frac{dT_l}{dc_l}$.

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. However, as we have noted, there are applications that do not indicate planned wages that in turn make transitions to employer businesses. Given this, we conduct our analysis for both

[21]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 $\frac{d\mathcal{V}_l}{d\iota_l^*} > 0$ 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.

[22]Note that $\frac{dT_l}{dc_l} = \frac{1}{A}\left(\frac{dS_l}{dc_l} - T\frac{dA_l}{dc_l}\right)$. 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.

[23]By (2), $p_l(\iota)$ is increasing if $G_l(\cdot|\iota)$ is decreasing in ι, which would be the case, for instance, if better ideas lead to higher V in a first order stochastic dominance sense.

types of applications. A missing piece is we do not explicitly model or empirically analyze those applications that are intended to become nonemployer businesses. Implicitly, we are treating this outcome as being captured in the reservation value (R_l) in the model. While this is internally consistent, explicit modeling and analysis of transitions to nonemployers would be of interest. We leave that for future work.

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[24]

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

Note that the covariance term can 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

$$Cov(\log A_l, \log T_l) = Cov(\log A_l, \log S_l - \log A_l) = Cov(\log A_l, \log S_l) - Var(\log A_l).^{25} \tag{13}$$

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

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

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

higher A_l would mean higher S_l.[26] 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

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

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

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

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 a zero-mean random 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 "DHS" transformations of the variables S and A that represent them in terms of deviations from their grand mean, respectively.[27] We

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

[27] Specifically, the transformation 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—this type of transformation was recommended by Tornqvist, Vartia, and Var-

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 the subset of locations $l \in \mathcal{L}$ where $A_l > 0$.

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 become less informative, as they will depend on conditions in nearby locations (e.g., the existence of nearby 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.[28]

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.

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. In any event, the contribution of local conditions cannot be interpreted as reflecting causal relationships, but is insightful for quantifying the nature and extent of the enormous

tia (1985) and also implemented by Davis, Haltiwanger, and Schuh (1996) for employment growth rates at the establishment-level.

[28]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).

within area variation in entrepreneurial activity.

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.

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.[29]

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.

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

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).[30] 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.[31] The restrictions are based on information on the EIN application including reason for applying and type of entity.[32] The details of the micro data are provided in Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (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 in-

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

[31]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.

[32]See Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (2018) for the specific set of filters applied to the application data to exclude applications with no business intent based on the application characteristics.

dicate a planned date for initial wage payments. However, as discussed in the introduction, we also consider all applications. Partly this reflects the nascent stage we are capturing. Nascent entrepreneurs with uncertainty about the timing and nature of the activity may not be ready to provide information about plans for hiring. Following the naming conventions of the public domain BFS, we refer to all business applications as BA and applications with planned wages as WBA.[33]

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.[34] Below we explore how the address stated in the application compares with the address of the actual business when it forms. As a preview, we find a close correspondence between the location of the application and the actual business for those that transition to employer businesses.

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

[33]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.

[34]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, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (2018).

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 2 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.

The motivating model assumes that the formation of employer businesses in a location originate from applications made in that same location. 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. Specifically, for the set of business applications that transition, in Table 1 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 also imply that the local conditions largely reflect both the location of the entrepreneur and for successful transitions, the business itself.

6 Variation in startups, ideas, and transition rate across locations

6.1 Motivating Evidence from Public Domain BFS at the State Level

To help provide perspective on the nature of spatial variation in startups, applications and transitions, we provide evidence from the public domain BFS using overall startups, total applications (BA) and transitions at the state-level. Using pooled data for the 2010-16 period, Figure 3 shows that startup intensity varies considerably across states with higher startup per capita states having over twice the rate as low states. States in the West have especially high startup rates per capita. Figure 4 shows that when we decompose startups per capita into applications per capita and transition rates, the dispersion in each component is also high. States with high applications per capita have nearly three times the rate as low states, while high transition states have nearly two times the transition rate as low states.

As shown in Figure 5, states with high startups per capita are not uniformly characterized by both high applications per capita and transition rates. Instead, some high startup states are characterized by relatively low applications per capita and high transition rates, and others by high applications per capita and low transition rates. Some of the variation in applications and transitions by state may reflect different propensities for BA applications to be targeted to be employers vs nonemployers. In appendix figure B.1, we show that patterns illustrated here using all applications (BA) are broadly similar if we use high propensity applications (HBA) for this state-level analysis.

6.2 County-level Analysis

We now turn to our primary analysis using the BFS micro data where we analyze variation at the county and tract level. Summary statistics for startups, applications, and transition rates at the county-level are presented in Table 2. The average county has mean startup rate per capita from WBA applications of 0.88 (defined as per 1,000 prime age adults for 2010-2016) with a standard deviation of 0.63. WBA applications per capita have a mean of 2.15 per capita with a standard deviation of 1.27. The mean transition rate is 0.41 with a standard deviation of 0.15. The coefficients of variation are 0.72, 0.59 and 0.37 respectively for startups per capita, (WBA) applications per capita, and transitions. The mean startup rate per capita from BA applications at the county level is 1.33 with a standard deviation of 0.89, the mean applications per capita from BA is 10.5 with a standard deviation of 6.11, and the transition rate is 0.13 with a standard deviation of 0.41. The coefficients of variation are 0.67, 0.58 and 0.41 respectively for startups per capital, (BA) applications per capita, and transitions.

In comparing the WBA and BA patterns, we find startups per capita from all applications (BA) is, on average, 1.5 times greater than startups per capita from applications with planned wages (WBA), indicating that BA that do not report planned wages are an important source of business formation. This reinforces our interest in considering both BA and WBA. Second, BA per capita is on average about five times WBA per capita suggesting that BA includes applications less targeted to employer startups. Third, consistent with that perspective, the transition rate for WBA is about three times larger than the transition rate for BA. Finally, the variation relative to the mean (coefficient of variation, CV) is highest in startups per capita, followed by applications per capita, and the transition rate. Interestingly, the respective coefficients of variation of the components for WBA and BA are broadly similar in magnitude.

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 which at the county-level is not much of a restriction. In the next section, we incorporate cases with zeroes in our analysis of local conditions.

Table 3 presents the variance decomposition at the county by year level 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. Idea variation is more important for WBA compared to BA, and the covariance between ideas and transition rates is negative and small. Population weighting increases the relative importance of ideas, 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.

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 4 describes all of the local condition variables used for analysis. We separate local conditions, C_l, into three groups:

- **Demographic conditions:** age, education, race, ethnicity, and foreign born.[35]
- **Economic conditions:** income per capita, employment-to-population ratio, and owner occupied housing share.[36]
- **Financial and business conditions:** debt-to-income ratio, small business lending per small business employment (the ratio of loans less than $1 million to employment at firms with less than 500 employees), percent employment in young firms, and concentration of employment (HHI).[37]

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 business and financial conditions, we use the average across the lags $k = 1, .., 5$.

The county-level analysis includes conditions across all three groups, while the tract-level analysis focuses on demographic and economic conditions. Data on the financial conditions is not available at the tract level; and although data on business conditions can be constructed at the tract level using the LBD, these conditions are likely to be more relevant when thinking about counties than tracts (and we control for county by year effects in the tract level analysis). 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

[35] See, for example, Doms, Lewis, and Robb (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).

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

[37] See, for example, Nocke (2006), Michelacci and Silva (2007), Glaeser, Kerr, and Ponzetto (2009), Glaeser, Kerr, and Ponzetto (2010), Glaeser, Rosenthal, and Strange (2010).

covariates. Such detailed fixed effects do not enable identification of causal mechanisms, but do provide guidance on the nature of the local variation.

7.1 County-level results

Table 6 reports the county-level regression results. The results are largely consistent across WBA and BA, and we focus our discussion on WBA results, as the WBA results more closely fit with our model and are easier to interpret. Still, it is interesting that the results are robust to considering the wider class of all applications. We highlight cases where there is discrepancy between the results for WBA and BA.

We begin with a discussion of the R^2, both overall and within, after abstracting from the commuting zone by year fixed effects. Three 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, observable local conditions help account for the enormous spatial variation in these outcomes, even after taking into account commuting zone by year effects. They explain 21 percent and 14 percent for variation in all startups and WBA startups per capita, respectively. Third, local conditions are notably far more important for explaining applications per capita than transition rates. They explain as much as 39 percent of cross-county variation in BA per capita and 18 percent of WBA per capita, but less than five percent of variation in BA and WBA transition rates. 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 counties that at first glance local conditions do not provide much guidance.

Turning to specific local conditions and starting with demographic characteristics, we see that counties with a higher share of individuals with at least a bachelors degree, and a higher share of foreign born are positively related to startups. Meanwhile, the fraction of older individuals is positively associated with startups and applications, but has no significant relationship with the transition rate. For race and ethnic groups, the patterns are mixed. Strikingly, a county with a higher African American share of the population has higher applications per capita but lower startups per capita. Underlying the latter is a lower transition rate.[38] 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

[38] 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.

financial and business conditions. Other notable findings are that the share of Asians is negatively associated with startups per capita, applications per capita, and transition rates. A higher Hispanic share is associated with lower startups and transition rates, but has no significant relationship with applications.

Local economic conditions of households—per capita income and employment-to-population ratio—are positively associated with startups per capita and applications per capita, but not strongly related to transition rates. We next turn to local financial and business conditions. Counties with a higher household debt-to-income ratio have lower startup activity, driven by lower transition rates. Meanwhile small business lending intensity is not significantly associated with WBA startups per capita, applications per capita, or transition rate, but is positively associated with BA per capita. It is not obvious why the partial correlation with BA should be stronger than WBA but may reflect our detailed fixed effects, as well other controls. Young business employment share is not associated with startups per capita. This insignificant relationship appears to arise from a positive relationship with applications per capita (only significant for WBA) being unwound by a negative relationship with transition rates. Finally, startups are inversely correlated with local employment concentration. This effect works through both applications and transition rates.

Even though applications and transitions at the county level tend to be inversely related, several factors are positively associated with both applications and transitions, such as bachelors+ or foreign born shares. Likewise, there are a set of factors that are negatively correlated with both applications and transitions, including Asian share and employment concentration. African American share stands out as having a positive relationship with applications and a negative relationship with transitions. This result is also different in that the negative association with transitions dominates, and African American share has a negative relationship overall with startups per capita.

To provide perspective on the quantitative implications of the estimates, 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.[39] 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 7.

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

[39]The elasticities can be found in Table B.1.a in Appendix B.

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.

We 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 (forthcoming). 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.[40] By construction this method yields a decomposition where all terms (including the residual) sum to one.[41] Table 8 provides the decomposition for the six models estimated. For startup and application models, economic conditions contribute the most, followed by demographic variables. The financial and business conditions contribute the least. These observations apply to the models for both WBA and BA per capita. However, for transition rate models, demographic variables contribute the most, followed by financial and business conditions. [42]

[40]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.

[41]Moreover, the residual contribution matches the regression results in Table 6 by construction.

[42]Table B.1.b includes the contribution of all individual and grouped variables.

7.2 Tract-level results

The county-level analysis shows that start-up activity, application activity, and transition rates are systematically linked to economic and demographic characteristics of the local economy. Furthermore, these characteristics help explain a significant portion of the geographic variation in startup activity across the United States.

We next focus on how startup activity varies across neighborhoods within counties, motivated by the finding that there is substantial variation across tracts within counties. Using the transformed dependent variable defined as the relative variation in application and startup rates per capita, we find that the standard deviation of those variables are 68 and 64 percent higher at the tract level relative to the county level based on wage applications (WBA), respectively. Similar remarks apply to overall applications per capita and startups per capita based on all business applications (BA).

The substantial granular neighborhood variation in the entrepreneurship process is interesting in its own right as it indicates that highly local conditions matter for the creation of ideas and transitions to startups. We now further explore this variation building on our results from the county-level analysis. The unit of measurement is the census tract. The BFS codes over 85 percent of EIN applications to the census tract.

As in the county analysis, application and startup activity are measured on a per capita basis. The adult population variable used comes from the American Community Survey 5-year data, using the midpoint year to match to the application data. The set of controls include demographic characteristics (median age, share of African-Americans or Black, share of Asian, share of Hispanics, and share or foreign born), educational attainment (share of population with some college and share of population with bachelors or graduate degree) and economic characteristics (median household income, the share of owner-occupied housing units and employment-to-population ratio).[43] The tract-level analysis omits variables on lending, debt and market structure due to the lack of availability at the tract level and the fact that we control for county by year effects. These county by year effects will control for differences in variables such as market structure that are likely best represented at a broader geographic level than the census tract.

The main regression results are presented in Table 9. The county by year fixed effects and the control variables explain less of the cross-tract variation for the WBA measures of entrepreneurial activity as compared to the BA measures. In addition, for both WBA and BA the models for startups and applications (14 and 15) explain substantially more of the cross-tract variation as compared to the transition rate model (16). Indeed, the within R^2's

[43] Descriptive statistics for tract-level variables are provided in Table B.2.a.

reported for both the county and tract transition rate models are relatively small, indicating that the neighborhood tract-level observables account for little of the variation in transition rates across neighborhoods within counties. Overall, variation in the outcomes at the tract level is less well-accounted for by observables and fixed effects than the variation at the county level. A greater relative role of idiosyncratic factors at the tract level is likely at work. Still, the observables do account for a substantial fraction of the variation with interesting details to which we now turn.

Looking at the individual coefficients, locations with higher median age and higher levels of human capital, as measured by the share of population at least a bachelors degree, are associated with higher application activity, higher transition rates and higher startup activity. This holds for both application types. Alternatively, a higher share of individuals with only some college is associated negatively with all three dependent variables.

Examining the race and ethnicity covariates, a higher share of African Americans residents in a neighborhood is associated with higher application activity, but lower transition rates and lower overall startup rates. This pattern is consistent with the county analysis and holds for both application types, suggesting that low transition rates are the source of low startup activity for neighborhoods with a higher share of African Americans. Application activity, transition rates and startup activity are not closely linked to the share of Asians in a neighborhood. Application and startup activity is negatively associated with the share of Hispanic residents in a tract, though transition rates are not closely linked to the share of Hispanics in a neighborhood. Finally, tracts with a higher fraction of residents that are foreign born tend to have higher startup and applications per capita for BA.

With regard to the economic variables, higher income neighborhoods tend to have higher applications, higher transition rates, and hence, higher startup activity. The employment-to-population ratio has a mixed relationship with the business formation patterns with no clear cut pattern across the two application groups. Finally, neighborhoods with higher ownership share of occupied housing units are associated with lower application activity, lower transition rates, and lower overall startup activity. This might be surprising if one thought of residential ownership as a potential proxy for wealth. However, it is more likely the negative association reflects the commercial versus residential nature of the neighborhood.

It is interesting to contrast the results on Hispanic and African-American share variables and startup activity. Neighborhoods with higher shares of African American or Hispanics have lower startup activity but the underlying sources of the association is quite different. Hispanic neighborhoods have lower application activity and that is the main source of the negative association between startup activity and Hispanic share. In contrast, African-American neighborhoods have higher application activity but this is countered by lower transition rates and, on net, the negative correlation with transition rates results in an overall negative relationship between African-American share and startup activity. Moreover, this pattern is true for both WBA and BA.

Table 10 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.

Table 11 presents the variance decomposition for the six models with a focus on the grouping of variables into demographic vs economic factors while also providing perspective on the contribution of overall local conditions relative to the residual. Demographic factors dominate the explained variation at this level of aggregation. From the appendix table showing the contribution of individual variables (Table B.2.c), there are three variables that contribute importantly to startup activity from WBA and BA – the share of population with bachelors or graduate degree, median age, and African-American share. The specific mechanism in how these variables contribute to explaining differs though. The share of population with bachelors or graduate degree and median age variables contribute by explaining variation in business applications across tracts within counties. The African-American share variable helps explain the variation in transition rates across tracts within counties. Indeed, the African American share variable is the only covariate that stands out in contributing to the variance in the transition rate model. This variable accounts for over 50 percent of the admittedly-small, explained within variation.

Some of our findings on the connection between business application activity and local demographic and economic conditions also emerge in recent work that uses a different indicator of business initiation: state business registrations. 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), which are larger than tracts on average, but smaller than counties. Their main finding is that areas with higher African-American share are associated with higher growth rates. They also find, albeit weaker evidence, that the growth rate is higher in areas with higher Hispanic share, and lower in areas with higher educational attainment (Bachelors+ degree) and higher share of

[44] The elasticities underlying the exercise can be found in Table B.2.b in Appendix B.

working age population. Our analysis instead considers cross-sectional patterns (not growth rates) in a relatively stable period (2011-2016) before the pandemic, and finds that business applications per capita were generally higher in areas with higher African-American share. However, we also observe that the higher application activity in these areas is offset by the lower transition rate of applications to employer businesses, resulting in lower startup formation per capita. Furthermore, our results indicate that African-American share is only one factor relevant for business applications, and several other local conditions actually have quantitatively stronger association with business applications.

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.

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 neighborhood characteristics used in the tract-level regressions, measured at the tract level, 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, LLC status, inclusion of a trade name, business start date, and quarter of application.[45] The models include county-year fixed effects, and standard errors are clustered at the county level.

The results are presented in Table 12, where only the coefficients of the neighborhood characteristics 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 9, even after controlling for detailed application characteristics. Neighborhood characteristics such as African American share, some college share and owner-occupied share are associated

[45]See Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (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.

with a lower probability of an application transitioning to employer status, while the share of the population with at least a bachelors degree and median household income are all associated with a higher likelihood of transitioning. In short, the relationships between neighborhood characteristics and transition rates observed above are robust when application characteristics are controlled for to account for the differences in the types of applications across tracts.

The R^2 values in Table 12, 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.[46] 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.[47]

7.4 Local conditions and ranking of locations

To provide perspective on the quantitative implications of our findings, 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 6 shows the results for this exercise at the county level and left panel of Figure 7 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

[46]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).

[47]See also Bayard, Dinlersoz, Dunne, Haltiwanger, Miranda, and Stevens (2018) for a similar finding.

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.[48]

We next evaluate how well our models account for these ranking patterns. The right panel of Figures 6 and 7 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 6 and county by year in Figure 7) 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.

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, 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 economic, financial, and business conditions) 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.

[48]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.

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, economic, local business and financial conditions, accounts for a substantial fraction of the within variation at the county and tract level. In general, across both counties and tracts, income, 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, Hendren, Kline, and Saez (2014) who emphasized enormous spatial variation in distinct measures of economic mobility.

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 potential 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 will await the development of the underlying micro data on transitions of applications for the post-2020 period.

References

Adelino, M., A. Schoar, and F. Severino (2015) "House Prices, Collateral, and Self-Employment," *Journal of Financial Economics*, 117, 288–306.

Andrews, R., C. Fazio, Y. Liu, J. Guzman, and S. Stern (2018) "The Startup Cartography Project: Measuring and Mapping Entrepreneurial Ecosystems," *Research Policy*, 51(2), 104437.

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

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

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

Bayard, K., E. Dinlersoz, T. Dunne, J. Haltiwanger, J. Miranda, and J. Stevens (2018) "Early Stage Business Formation: An Analysis of Applications for Employer Identification Numbers," NBER Working Paper No. 24364.

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

Bennett, V. M., and D. T. Robinson (2023) "Why Aren't There More Minority Entrepreneurs?," Discussion paper, Duke University.

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

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

Chen, J. S., D. C. Croson, D. W. Elfenbeinand, and H. E. Posen (2018) "The Impact of Learning and Overconfidence on Entrepreneurial Entry and Exit," *Organization Science*.

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

Choi, J. (2017) "Entrepreneurial Risk-Taking, Young Firm Dynamics and Aggregate Implications," Discussion paper, University of Maryland.

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

Davis, S., J. Haltiwanger, C. Krizan, R. Jarmin, J. Miranda, A. Nucci, and K. Sandusky (2009) "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Non-Employer Businesses," in *in Producer Dynamics: New Evidence from Micro Data*, pp. 329–366. University of Chicago Press.

Davis, S., J. Haltiwanger, and S. Schuh (1996) "Small business and job creation: Dissecting the myth and reassessing the facts," *Small Business Economics*, 8, 297–315.

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

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

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

Dinlersoz, E., T. Dunne, J. Haltiwanger, and V. Penciakova (2021) "Business Formation: A Tale of Two Recessions," *American Economic Association: Papers and Proceedings*.

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

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

Eslava, M., J. Haltiwanger, and N. Urdaneta (forthcoming) "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges," *Review of Economic Studies*.

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

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

Fazio, C., J. Guzman, Y. Liu, and S. Stern (2021) "How is COVID Changing the Geography of Entrepreneurship? Evidence from the Startup Cartography Project," Working paper 28787, NBER Working paper.

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

——— (2010) "Clusters of entrepreneurship," *Journal of Urban Economics*, 67, 150–168.

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

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

Guzman, J., and S. Stern (2020) "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*, 12(4), 212–243.

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

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

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

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

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

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

Hurst, E., and B. W. Pugsley (2017) "Wealth, Tastes, and Entrepreneurial Choice," in *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, pp. 111–151. University of Chicago Press.

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

Kerr, S. P., and W. R. Kerr (2017) "Immigrant entrepreneurship," in *Measuring entrepreneurial businesses: Current knowledge and challenges*, pp. 187–249. University of Chicago Press.

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

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014) "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, 28(3), 25–48.

Lucas, R. (1978) "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523.

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

Michelacci, C., and O. Silva (2007) "Why so Many Local Entrepreneurs?," *The Review of Economics and Statistics*, 89(4), 615–633.

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

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

Poschke, M. (2013) "Who becomes an entrepreneur? Labor market prospects and occupational choice," *Journal of Economic Dynamics and Control*, 37, 693–710.

Reynolds, P. D. (2017) "When is a Firm Born? Alternative Criteria and Consequences," *Business Economics*, 52, 42–56.

Reynolds, P. D., N. M. Carter, W. B. Gartner, and P. G. Greene (2004) "The Prevalence of Nascent Entrepreneurs in the United States: Evidence from the Panel Study of Entrepreneurial Dynamics," *Small Business Economics*, 23, 263–284.

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

Tornqvist, L., P. Vartia, and Y. O. Vartia (1985) "How Should Relative Changes Be Measured?," *The American Statistician*, 39(1), 43–46.

Vardishvili, I. (2023) “Entry Decision, the Option to Delay Entry, and Business Cycles,” .

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

Table 1: 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 2010 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 2: Startups, Applications, and Transition Rates:
WBA and BA Summary Statistics

	WBA			BA		
	(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	1.337	10.480	0.128
SD	0.628	1.273	0.150	0.894	6.109	0.053
CV	0.716	0.592	0.369	0.669	0.583	0.414

Notes: Reports the mean, standard deviation (SD), and coefficient of variation (CV) of startups per 1,000 prime-age (20-64 years old) people (startups pc), applications per 1,000 prime-age people (applications pc), and transition rate (startups divided by applications) at the county level between 2010 and 2016, separately for WBA and BA.

Table 3: Variance Decomposition for County-Level Startups per Capita

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	Applications pc	Transition rate	2 × covariance	Applications pc	Transition rate	2 × covariance
Unweighted	0.676	0.376	-0.052	0.485	0.518	-0.004
Weighted	0.919	0.369	-0.288	0.983	0.428	-0.410

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the variance decomposition of log(startups pc) into log(application pc) and log(transition rate). The underlying data are at the county-year level, covering 2010-2016. County population is used for weighting. Startups are defined as applications that transition to an employer business within eight quarters after application.

Table 4: Description of Local Condition Variables

Variable	**Definition**	**Source**	**County-Analysis**	**Tract-Analysis**
Log(median age)	log of median age	ACS	✓	✓
Bachelors+ degree share	share of pop with Bachelors 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	share of pop foreign born	ACS	✓	✓
Log(per capita income)	log of per capita income	BEA	✓	
Log(median HH income)	median household income	ACS		✓
Owner-occupied share	share of owner-occupied housing units	ACS		✓
Emp-pop ratio	employment to pop ratio	ACS	✓	✓
Log(debt-to-income)	log of household debt to income ratio	FRB	✓	
DHS(SME loans/emp)	value of SBLs to firms with <$1mn revenue divided by the emp. in firms with <500 emp	CRA & LBD	✓	
Share of emp in young firms	share of employment in firms aged 1-5	LBD	✓	
Concentration of emp	Herfindahl index of employment	LBD	✓	

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

Table 5: County-Level Regression Variable Summary Statistics

	(1) Mean	(2) SD	(3) CV
DHS(WBA startups pc)	-0.178	0.595	-3.338
DHS(WBA pc)	-0.121	0.466	-3.847
WBA transition rate	0.408	0.150	0.368
DHS(BA startups pc)	-0.160	0.562	-3.503
DHS(BA pc)	-0.085	0.373	-4.377
BA transition rate	0.128	0.053	0.414
Log(median age)	3.689	0.130	0.035
Bachelors+ degree share	0.194	0.087	0.448
Some college share	0.295	0.053	0.181
African American share	0.092	0.149	1.620
Asian share	0.012	0.025	2.083
Hispanic share	0.085	0.135	1.588
Foreign born share	0.045	0.056	1.244
Log(per capita income)	10.450	0.235	0.022
Emp-pop ratio	0.550	0.082	0.149
Log(debt-to-income ratio)	0.436	0.526	1.206
DHS(SME loans/employment)	-0.212	0.608	-2.875
Share of emp in young firms	0.120	0.053	0.442
Concentration of employment	0.035	0.049	1.400

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 level regressions. The years covered are 2010-2016. Startups are defined as applications that transition to an employer business within eight quarters after application. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996).

Table 6: County-Level Regressions

	WBA			BA		
	(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)	0.468***	0.551***	–0.00417	0.488***	0.607***	0.00321
	(0.0849)	(0.0783)	(0.0204)	(0.0845)	(0.0561)	(0.00792)
Bachelors+ degree share	1.310***	0.942***	0.101***	1.489***	0.962***	0.0471***
	(0.156)	(0.141)	(0.0334)	(0.174)	(0.121)	(0.0138)
Some college share	0.0407	0.0722	–0.0740	0.111	0.0289	–0.0156
	(0.256)	(0.179)	(0.0571)	(0.252)	(0.135)	(0.0233)
African Am. share	–0.418***	0.406***	–0.264***	–0.344***	0.617***	–0.0944***
	(0.0787)	(0.0720)	(0.0192)	(0.0809)	(0.0573)	(0.00702)
Asian share	–1.996***	–1.345*	–0.338**	–1.943***	–0.986*	–0.151**
	(0.640)	(0.690)	(0.131)	(0.750)	(0.532)	(0.0640)
Hispanic share	–0.334*	–0.0406	–0.199***	–0.326*	–0.101	–0.0607***
	(0.188)	(0.142)	(0.0501)	(0.190)	(0.125)	(0.0226)
Foreign born share	1.238***	0.749**	0.203**	1.440***	0.819***	0.0964*
	(0.325)	(0.316)	(0.0925)	(0.337)	(0.285)	(0.0518)
Log(per capita income)	0.379***	0.427***	–0.00433	0.442***	0.438***	0.00117
	(0.0702)	(0.0526)	(0.0153)	(0.0699)	(0.0375)	(0.00696)
Emp-pop ratio	0.976***	0.773***	0.0718	0.927***	0.992***	0.0196
	(0.193)	(0.176)	(0.0456)	(0.214)	(0.130)	(0.0188)
Log(debt-to-income)	–0.0455***	–0.0211	–0.00698*	–0.0173	0.00577	–0.00511***
	(0.0171)	(0.0134)	(0.00382)	(0.0168)	(0.00958)	(0.00143)
DHS(SME loans/emp)	0.0251	0.0160	0.00496	0.0221	0.0255***	–0.000773
	(0.0160)	(0.0123)	(0.00388)	(0.0162)	(0.00895)	(0.00145)
Share of emp in young firms	–0.105	0.290	–0.0950*	–0.206	0.624***	–0.0341*
	(0.200)	(0.188)	(0.0487)	(0.231)	(0.102)	(0.0203)
Concentration of emp	–1.538***	–0.959***	–0.172***	–1.775***	–0.400***	–0.112***
	(0.197)	(0.156)	(0.0472)	(0.209)	(0.113)	(0.0186)
Fixed effects	cz x yr	cz x yr	cz x yr	cz x yr	cz x yr	cz x yr
SE clustering	cz	cz	cz	cz	cz	cz
R^2	.4906	.6001	.3638	.5628	.7622	.4274
Within R^2	.1409	.1793	.0305	.2097	.3907	.0422
N	17,000	17,500	17,000	17,500	17,500	17,500

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports regression results for DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5), and transition rate (cols. 3 and 6), separately for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5, and 6). "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. All regressions include commuting zone (CZ) × year fixed effects. 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 commuting zone level.

Table 7: County-Level Regression Magnitudes

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Median age	5.897	6.943	-0.126	6.149	7.648	0.315
Bachelors+ degree share	11.379	8.198	2.150	12.947	8.378	3.181
Some college share	0.217	0.380	-0.977	0.597	0.163	-0.652
African Am. share	-6.156	5.994	-9.720	-5.184	9.234	-11.016
Asian share	-4.999	-3.333	-2.083	-4.791	-2.500	-2.916
Hispanic share	-4.446	-0.476	-6.511	-4.446	-1.429	-6.352
Foreign born share	6.966	4.230	2.737	8.086	4.603	4.230
Per capita income	10.271	11.572	-0.298	11.978	11.870	0.244
Emp-pop ratio	8.001	6.333	1.445	7.599	8.135	1.252
Debt-to-income	-2.003	-0.934	-0.757	-0.757	0.267	-1.780
SME loans/emp	3.408	2.181	1.636	2.999	3.408	-0.818
Share of emp in young firms	-0.575	1.547	-1.238	-1.105	3.315	-1.414
Concentration of emp	-7.560	-4.760	-2.100	-8.680	-1.960	-4.340

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). 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, Haltiwanger, and Schuh (1996).

Table 8: County-Level Regression Decomposition

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Groups						
Demographic	0.055	0.061	0.025	0.083	0.147	0.028
Economic	0.060	0.094	0.002	0.086	0.213	0.002
Financial & business	0.026	0.024	0.003	0.041	0.031	0.012
Categories						
Local conditions	0.141	0.179	0.030	0.210	0.391	0.042
Commuting zone conditions	0.350	0.421	0.333	0.353	0.371	0.385
Residual	0.509	0.400	0.636	0.437	0.238	0.573

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5) and transition rate (cols. 3 and 6) for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5 and 6) and all control variables are included, along with commuting zone $\times$ year fixed effects. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (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 commuting zone conditions (ie: commuting zone $\times$ year FE); and the last row is the remaining variation that is unexplained by either local conditions or commuting zone conditions.

Table 9: Tract-Level Regressions

	WBA			BA		
	(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)	0.632***	0.711***	0.0334***	0.665***	0.651***	0.0205***
	(0.0375)	(0.0398)	(0.00520)	(0.0512)	(0.0443)	(0.00239)
Bachelors+ degree share	1.046***	0.891***	0.101***	1.208***	0.982***	0.0267***
	(0.0553)	(0.0550)	(0.0102)	(0.0629)	(0.0497)	(0.00451)
Some college share	–0.240***	–0.221***	–0.0368***	–0.222***	–0.0735	–0.0255***
	(0.0678)	(0.0685)	(0.0124)	(0.0686)	(0.0606)	(0.00480)
African American share	–0.609***	0.273***	–0.221***	–0.530***	0.413***	–0.0769***
	(0.0491)	(0.0445)	(0.00715)	(0.0729)	(0.0634)	(0.00382)
Asian share	0.0848	0.107	0.0178	0.107	–0.190**	0.0580***
	(0.131)	(0.0989)	(0.0261)	(0.165)	(0.0924)	(0.0149)
Hispanic share	–0.311***	–0.183*	–0.0222	–0.322***	–0.234**	–0.00395
	(0.0651)	(0.0991)	(0.0167)	(0.0772)	(0.0997)	(0.00678)
Foreign born share	0.221*	0.283**	–0.0306**	0.483***	0.351***	0.000636
	(0.121)	(0.119)	(0.0147)	(0.129)	(0.120)	(0.00801)
Log(median HH income)	0.144***	0.0448**	0.0234***	0.129***	0.0158	0.00637***
	(0.0223)	(0.0225)	(0.00349)	(0.0245)	(0.0206)	(0.00144)
Owner-occupied share	–0.928***	–0.727***	–0.0472***	–0.803***	–0.493***	–0.0321***
	(0.0395)	(0.0365)	(0.00491)	(0.0382)	(0.0285)	(0.00242)
Emp-pop ratio	–0.199***	0.164***	–0.0451***	0.0566	0.128***	–0.0134***
	(0.0523)	(0.0585)	(0.00968)	(0.0620)	(0.0490)	(0.00356)
Fixed effects	cty x yr	cty x yr	cty x yr	cty x yr	cty x yr	cty x yr
SE clustering	cty	cty	cty	cty	cty	cty
R^2	.1960	.2734	.1390	.3072	.5538	.1858
Within R^2	.0960	.0822	.0347	.1389	.1895	.0429
N	398,000	430,000	398,000	428,000	430,000	428,000

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports regression results for DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5), and transition rate (cols. 3 and 6), separately for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5, and 6). "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. All regressions include county (fips) × year fixed effects. 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 10: Tract-Level Regression Magnitudes

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Median Age	12.410	13.961	1.759	13.058	12.783	3.438
Bachelors+ degree share	19.299	16.439	4.999	22.288	18.118	4.207
Some college share	–1.905	–1.755	–0.784	–1.762	–0.584	–1.729
African American share	–13.593	6.093	–13.232	–11.830	9.218	–14.658
Asian share	0.739	0.933	0.416	0.933	–1.657	4.319
Hispanic share	–6.553	–3.856	–1.255	–6.785	–4.930	–0.711
Foreign born share	3.017	3.863	–1.120	6.593	4.791	0.074
Median HH income	7.107	2.211	3.098	6.367	0.780	2.685
Owner-occupied share	–21.093	–16.525	–2.878	–18.252	–11.206	–6.231
Emp-pop ratio	–2.111	1.740	–1.284	0.601	1.358	–1.214

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). 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. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). The LHS variable of the regression is listed in the columns, and each RHS variable is listed in the rows.

Table 11: Tract-Level Regression Decomposition

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Demographic	0.088	0.068	0.034	0.131	0.182	0.043
Economic	0.008	0.014	0.001	0.008	0.008	0.000
Local Conditions	0.096	0.082	0.035	0.139	0.189	0.043
County Conditions	0.100	0.191	0.104	0.168	0.364	0.143
Residuals	0.804	0.727	0.861	0.693	0.446	0.814

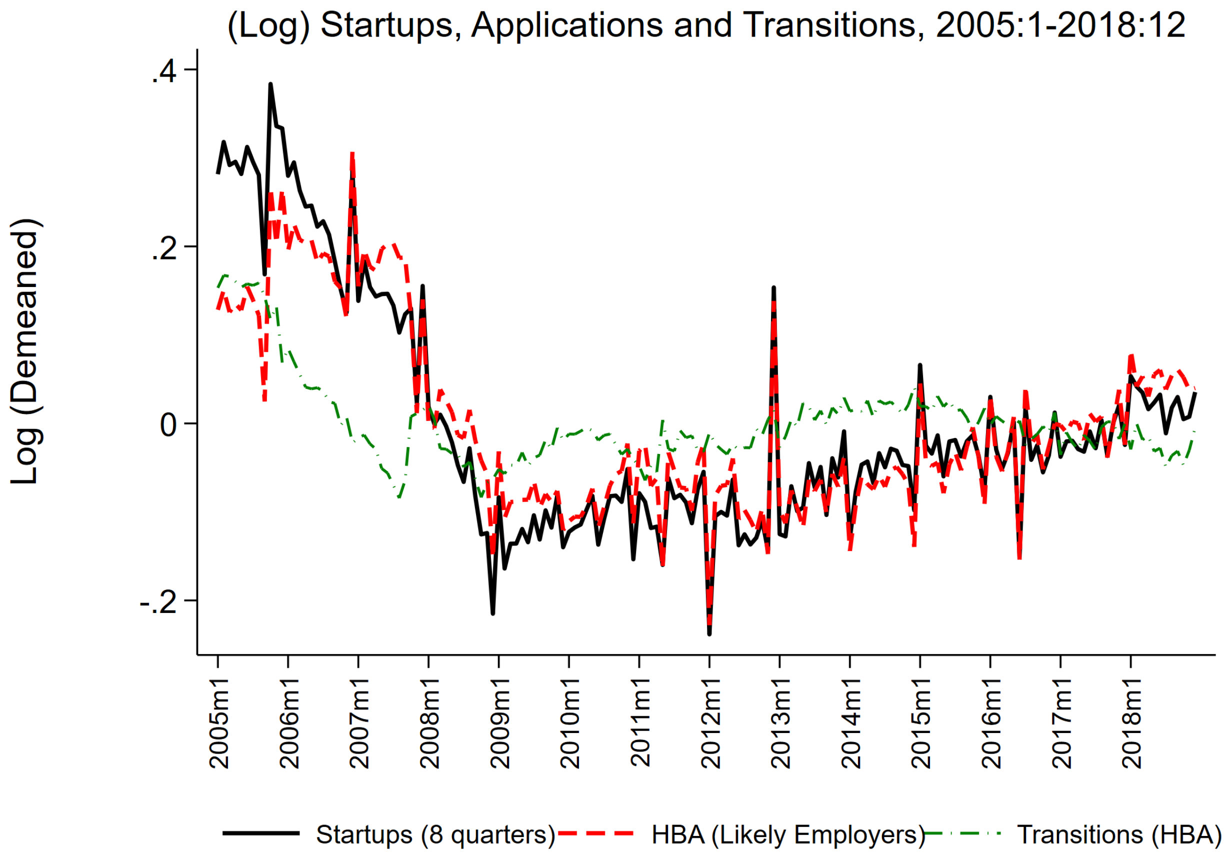
Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5) and transition rate (cols. 3 and 6) for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5 and 6) and all control variables are included, along with county $\times$ year fixed effects. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The Demographic row is the sum of the contribution of the seven demographic variables and the row labeled Economic is the sum of the contribution of the three economic variables. The Local Conditions 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 County Conditions row is the contribution of county-year fixed effects zone conditions; and the last row is the remaining variation that is unexplained by either local conditions or county conditions.

Table 12: Tract-Level Linear Probability Model (LPM) Estimates

	(1) WBA Transition	(2) BA Transition
Log(median age)	0.0198***	0.00235
	(0.00548)	(0.00250)
Bachelors+ degree share	0.0596***	0.0208***
	(0.00762)	(0.00293)
Some college share	–0.0478***	–0.0201***
	(0.0111)	(0.00450)
African American share	–0.170***	–0.0535***
	(0.00737)	(0.00237)
Asian share	0.0190	0.0295***
	(0.0206)	(0.0105)
Hispanic share	–0.0246	–0.0121
	(0.0185)	(0.00738)
Foreign born share	–0.0374**	–0.0204***
	(0.0156)	(0.00732)
Log(median HH income)	0.0208***	0.00679***
	(0.00298)	(0.00118)
Owner-occupied share	–0.0287***	–0.0107***
	(0.00431)	(0.00232)
Emp-pop ratio	–0.0293***	–0.00681**
	(0.00734)	(0.00328)
Fixed effects	cty x yr	cty x yr
SE clustering	cty	cty
R^2	.1127	.1995
Within R^2	.0089	.0774
N	2,355,000	13,840,000

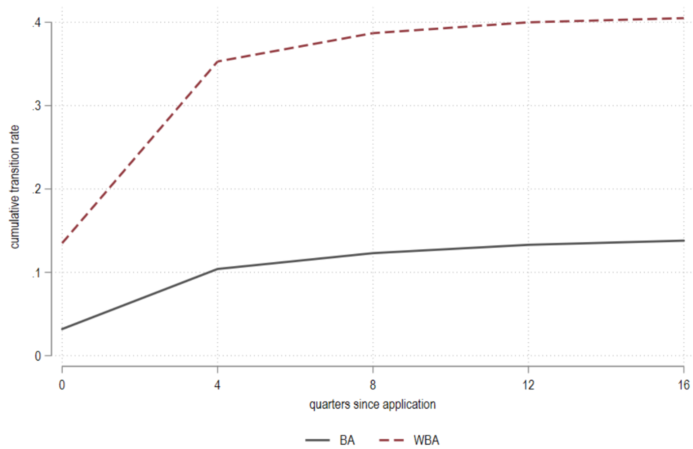
Notes: Reports regression results for transition rate (cols. 1 and 2), separately for WBA (col. 1) and BA (col. 2). All regressions include county (cty) $\times$ year (yr) fixed effects. The observation counts have been rounded for disclosure reasons. ***, **, and * indicate sig. at the 1%, 5%, and 10% sig. Standard errors are clustered at the county level.

Figure 1: Startups, Applications and Transitions: National Data



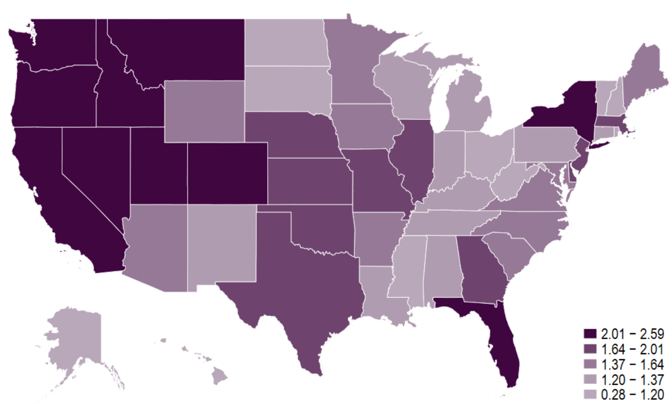
Notes: Depicts log(startups), log(HBA) and log(Transitions) from 2005:1-2018:12. Series are demeaned relative to average over time period.

Figure 2: Cumulative Transition Rates for BA and WBA, Time Since Application



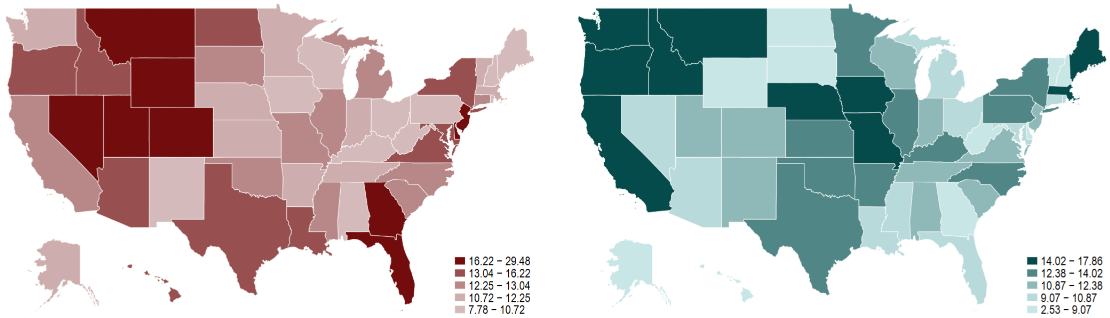
Notes: Depicts the cumulative transition rate of BA and WBA between 0 and 16 quarters after application. Due to disclosure considerations, the cumulative transition rates are calculated only for quarters 0, 4, 8 and 16. Transitions (startups) are defined as applications that transition to an employer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

Figure 3: Startups Per Capita, by State



Notes: Depicts average BA startups per 1,000 prime-age (20-64 years old) people (startups per capita) at the state level between 2010 and 2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

Figure 4: Applications Per Capita & Transition Rate, by State

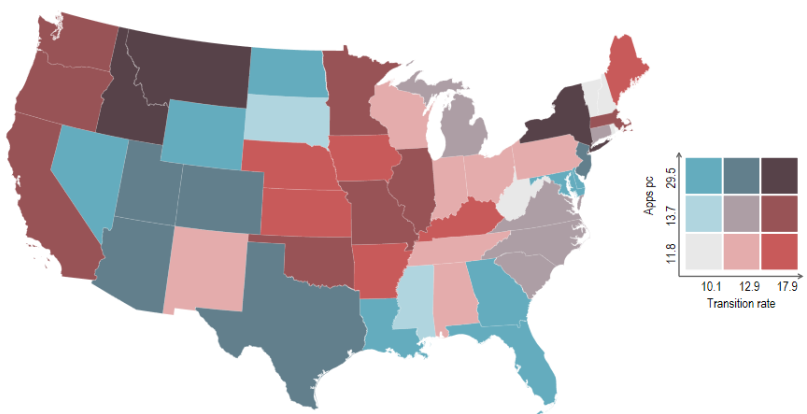


(a) Applications per capita

(b) Transition rate

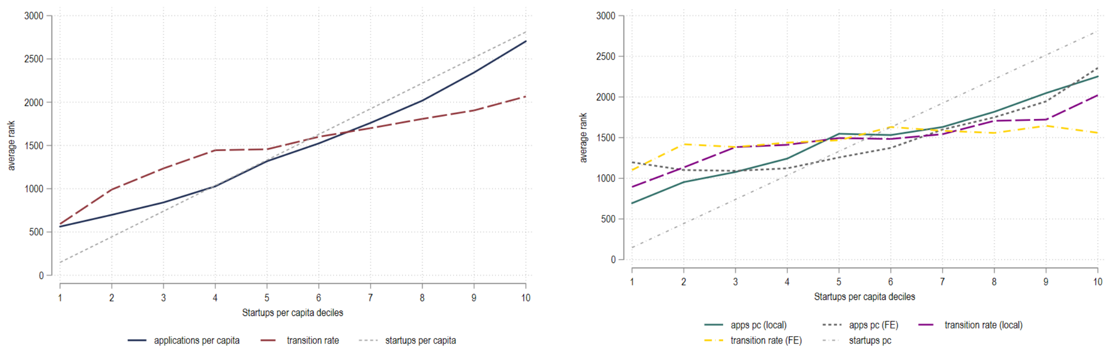
Notes: Depicts average BA per 1,000 prime-age (20-64 years old) people (applications per capita) in the left figure and transition rate (applications divided by startups) in the right figure at the state level between 2010 and 2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

Figure 5: Applications Per Capita vs. Transition Rate, by State



Notes: Depicts average BA per 1,000 prime-age (20-64 years old) people in the left figure versus the transition rate (applications divided by startups) between 2010 and 2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

Figure 6: County-level Rank Analysis (deciles of startups per capita)

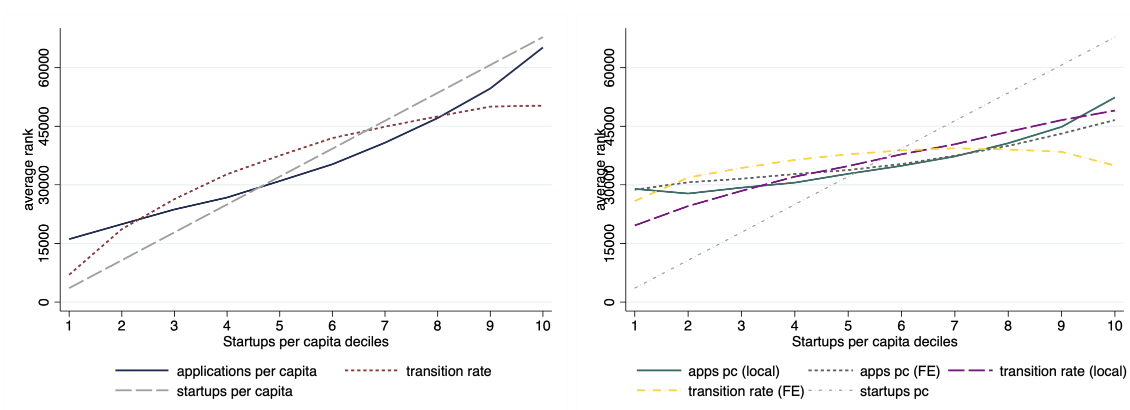


(a) Apps pc & transition rates

(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 2010-2016). 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 7: 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 2010-2016). 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).

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

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

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

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

and

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

Therefore,

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

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

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

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

To provide perspective on the quantitative implications of the estimates, 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.[49] Given that the covariates differ

[49] The elasticities can be found in Table B.1.a in Appendix B.

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 7.

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.

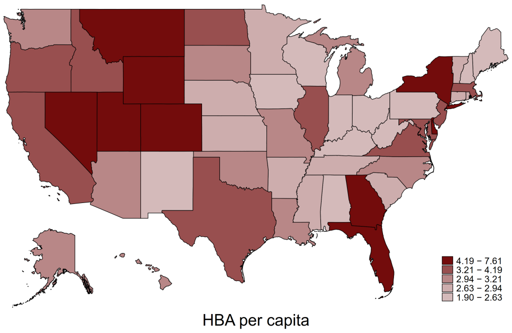
A.4 Estimates of the elasticities and magnitudes at the tract level

Table 10 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.[50] 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.

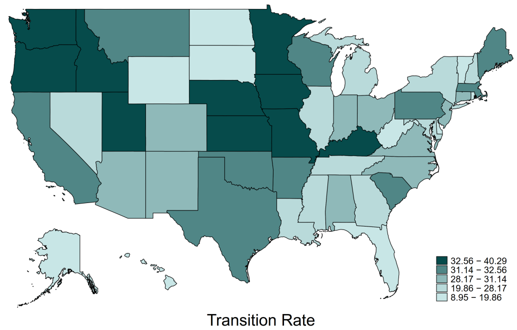
[50]The elasticities underlying the exercise can be found in Table B.2.b in Appendix B.

B Additional tables and figures

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



(a) Applications per capita, HBA

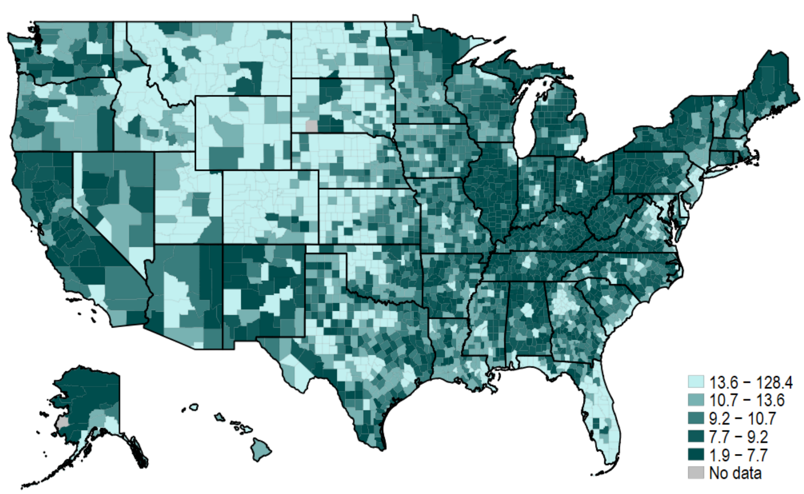


(b) Transition rate

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 2010 and 2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

B.1 County-level analysis

Figure B.2: Applications Per Capita, by County



Notes: Depicts average BA per 1,000 prime-age (20-64 years old) people (applications per capita) at the county level between 2010 and 2016.

Table B.1.a: County-Level Regression Elasticities

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Median age	0.468	0.551	-0.010	0.488	0.607	0.025
Bachelors+ degree share	0.254	0.183	0.048	0.289	0.187	0.071
Some college share	0.012	0.021	-0.054	0.033	0.009	-0.036
African Am. share	-0.038	0.037	-0.060	-0.032	0.057	-0.068
Asian share	-0.024	-0.016	-0.010	-0.023	-0.012	-0.014
Hispanic share	-0.028	-0.003	-0.041	-0.028	-0.009	-0.040
Foreign born share	0.056	0.034	0.022	0.065	0.037	0.034
Per capita income	0.379	0.427	-0.011	0.442	0.438	0.009
Emp-pop ratio	0.537	0.425	0.097	0.510	0.546	0.084
Debt-to-income	-0.045	-0.021	-0.017	-0.017	0.006	-0.040
SME loans/emp	0.025	0.016	0.012	0.022	0.025	-0.006
Emp in young firms	-0.013	0.035	-0.028	-0.025	0.075	-0.032
Concentration of emp	-0.054	-0.034	-0.015	-0.062	-0.014	-0.031

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the regression elasticities for regressions of DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5), and transition rate (cols. 3 and 6) on local conditions (control variables), separately for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5, and 6). "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. All regressions include commuting zone (CZ) $\times$ year fixed effects.

Table B.1.b: County-Level Regression Decomposition (Full Table)

	WBA			BA		
	(1)	(2)	(3)	(4)	(5)	(6)
	DHS(startups pc)	DHS(applications pc)	Transition rate	DHS(startups pc)	DHS(applications pc)	Transition rate
Individual variables						
Log(median age)	-0.001	0.007	0.000	-0.002	0.022	-0.000
Bachelors+ degree share	0.050	0.054	0.004	0.077	0.112	0.007
Some college share	0.000	0.000	-0.000	0.001	0.000	-0.000
African Am. share	0.005	0.002	0.018	0.005	0.009	0.019
Asian share	-0.006	-0.006	-0.000	-0.009	-0.010	-0.001
Hispanic share	0.002	0.000	0.004	0.002	0.003	0.002
Foreign born share	0.006	0.005	-0.000	0.010	0.011	0.001
Log(per capita income)	0.035	0.066	-0.000	0.054	0.138	0.000
Emp-pop ratio	0.025	0.028	0.002	0.031	0.075	0.002
Log(debt-to-income)	0.000	0.000	-0.001	-0.000	0.000	-0.000
DHS(SME loans/emp)	0.003	0.002	0.001	0.003	0.009	-0.000
Share of emp in young firms	-0.000	0.001	-0.000	-0.000	0.012	0.000
Concentration of emp	0.023	0.020	0.003	0.038	0.010	0.012
Grouped variables						
Demographic	0.055	0.061	0.025	0.083	0.147	0.028
Economic	0.060	0.094	0.002	0.086	0.213	0.002
Financial & business	0.026	0.024	0.003	0.041	0.031	0.012
Grouped variables						
Local conditions	0.141	0.179	0.030	0.210	0.391	0.042
Commuting zone conditions	0.350	0.421	0.333	0.353	0.371	0.385
Residual	0.509	0.400	0.636	0.437	0.238	0.573

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the contribution of individual and groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5) and transition rate (cols. 3 and 6) for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5 and 6) and all control variables are included, along with commuting zone $\times$ year fixed effects. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. "Local conditions" 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; "Commuting zone conditions" is the contribution of commuting zone $\times$ year FE; and the last row is the remaining variation that is unexplained by either local conditions or commuting zone conditions.

B.2 Tract-level analysis

Table B.2.a: Tract-Level Regression Variable Summary Statistics

	(1) Mean	(2) SD	(3) CV
DHS(BA startups pc)	-0.484	0.906	-1.871
DHS(BA pc)	-0.282	0.599	-2.127
BA transition rate	0.117	0.0912	0.779
DHS(WBA startups pc)	-0.520	0.977	-1.876
DHS(WBA pc)	-0.371	0.782	-2.109
WBA transition rate	0.373	0.289	0.775
Log(median age)	3.629	0.200	0.055
Bachelors+ degree share	0.277	0.185	0.665
Some college share	0.287	0.079	0.277
African American share	0.137	0.223	1.628
Asian share	0.045	0.087	1.929
Hispanic share	0.153	0.210	1.376
Foreign born share	0.123	0.137	1.114
Log(per capita income)	10.840	0.470	0.043
Owner-occupied share	0.646	0.227	0.352
Emp-pop ratio	0.582	0.106	0.182

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 tract-level regressions. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). The years covered are 2010-2016. Startups are defined as applications that transition to an employer business within eight quarters after application.

Table B.2.b: Tract-Level Regression Elasticities

	WBA			BA		
	(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)	0.632	0.711	0.090	0.665	0.651	0.175
Bachelors+ degree share	0.290	0.247	0.075	0.335	0.273	0.063
Some college share	–0.069	–0.063	–0.028	–0.064	–0.021	–0.062
African American share	–0.083	0.037	–0.081	–0.073	0.057	–0.090
Asian share	0.004	0.005	0.002	0.005	–0.009	0.022
Hispanic share	–0.048	–0.028	–0.009	–0.049	–0.036	–0.005
Foreign born share	0.027	0.035	–0.010	0.059	0.043	0.001
Log(median HH income)	0.144	0.045	0.063	0.129	0.016	0.054
Owner-occupied share	–0.599	–0.470	–0.082	–0.519	–0.318	–0.177
Emp-pop ratio	–0.116	0.095	–0.070	0.033	0.074	–0.067

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the regression elasticities for regressions of DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5), and transition rate (cols. 3 and 6) on local conditions (control variables), separately for BA (cols. 1, 2 and 3) and WBA (cols. 4, 5, and 6). "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. All regressions include county × year fixed effects.

Table B.2.c: Tract-Level Regression Decomposition (Full)

	WBA			BA		
	(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)	0.016	0.022	0.002	0.023	0.055	0.003
Bachelors+ degree share	0.045	0.040	0.008	0.073	0.110	0.006
Some college share	0.002	0.001	0.000	0.002	0.001	0.001
African American share	0.019	–0.001	0.023	0.023	0.004	0.028
Asian share	0.001	0.001	0.000	0.001	–0.002	0.004
Hispanic share	0.006	0.005	0.001	0.009	0.019	0.000
Foreign born share	–0.000	–0.001	0.000	–0.000	–0.005	0.000
Log(median HH income)	0.009	0.002	0.004	0.012	0.002	0.003
Owner-occupied share	0.001	0.012	–0.002	–0.004	0.005	–0.002
Emp-pop ratio	–0.001	0.001	–0.001	0.001	0.001	–0.001
Demographic	0.088	0.068	0.034	0.131	0.182	0.043
Economic	0.008	0.014	0.001	0.008	0.008	0.000
Local Conditions	0.096	0.082	0.035	0.139	0.189	0.043
County Conditions	0.100	0.191	0.104	0.168	0.364	0.143
Residuals	0.804	0.727	0.861	0.693	0.446	0.814

Notes: "pc" refers to per capita (per 1000 prime-age (20-64 years old) people). Reports the contribution of groups of control variables (below *Groups* heading) to total R^2 of regressions where the dependent variables are DHS(startups pc) (cols. 1 and 4), DHS(applications pc) (cols. 2 and 5) and transition rate (cols. 3 and 6) for WBA (cols. 1, 2 and 3) and BA (cols. 4, 5 and 6) and all control variables are included, along with county $\times$ year fixed effects. "DHS" refers to the transformation based on Davis, Haltiwanger, and Schuh (1996). Startups are defined as applications that transition to an employer business within eight quarters after application. The Demographic row is the sum of the contribution of the seven demographic variables and the row labeled Economic is the sum of the contribution of the three economic variables. The Local Conditions 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 County Conditions row is the contribution of county-year fixed effects zone conditions; and the last row is the remaining variation that is unexplained by either local conditions or county conditions.