NO. 7–2023 OCTOBER 2023
FEDERAL RESERVE BANK OF ATLANTA'S *POLICY HUB*

Local Origins of Business Formation

Emin Dinlersoz, US Census Bureau
Timothy Dunne, University of Notre Dame
John Haltiwanger, University of Maryland
Veronika Penciakova, Federal Reserve Bank of Atlanta

Summary:

Using comprehensive administrative data on business applications, we find that startups per capita exhibit enormous variation across counties and tracts in the United States. We decompose this spatial variation into two components: variation in business ideas per capita and in their rate of transition to startups. Both components matter for the variation in startups per capita. Furthermore, local demographic, economic, financial, and business conditions account for a significant fraction of the variation in startups per capita and in its components. In particular, income, education, age, and foreign-born share are strongly and positively associated with idea generation and transition rate. The relationship between local conditions and ideas generally differs in magnitude from the relationship with the rate at which these ideas transition into employer businesses. Interestingly, certain conditions are positively associated with ideas but negatively with transition rates. The predicted rank of locations based only on observable local conditions closely relates to the actual ranking of locations in terms of startups per capita, making it possible to characterize high-startup locations using observable local conditions alone.

Key findings:

1. Startups per capita vary substantially across counties and tracts.
2. The two components of startup formation—business idea creation and the transition rate of ideas to employer businesses—are both important in accounting for the variation in startups per capita.
3. Observable local conditions explain a significant fraction of the variation in startups per capita and its components.
4. The predicted rank of locations based on observable local conditions is similar to the actual rank of locations in terms of startups per capita.

Center Affiliation: Center for Human Capital Studies

JEL Classification: L26, R12, R23

Key words: entrepreneurship, firm entry, business formation, business dynamism, economic geography

https://doi.org/10.29338/ph2023-7

CENTER FOR QUANTITATIVE ECONOMIC RESEARCH

CENTER FOR HUMAN CAPITAL STUDIES

CENTER FOR FINANCIAL INNOVATION AND STABILITY

CENTER FOR HOUSING AND POLICY

ECONOMIC SURVEY RESEARCH CENTER

AMERICAS CENTER

Federal Reserve Bank *of* Atlanta

The Federal Reserve Bank of Atlanta's *Policy Hub*
leverages the expertise of Atlanta Fed economists and researchers to address issues of broad policy interest. Our research centers coordinate this work and seek to influence policy discussions. Areas of interest include: forecasting, fiscal policy, and macroeconomics (Center for Quantitative Economic Research); financial stability, innovation, and regulation (Center for Financial Innovation and Stability); human capital, labor markets, health, and education (Center for Human Capital Studies); and government-sponsored entity reform, mortgage markets, and affordable housing (Center for Housing and Policy). Sign up for email updates at **frbatlanta.org/research/publications/policy-hub**.

Local Origins of Business Formation

Summary: Using comprehensive administrative data on business applications, we find that startups per capita exhibit enormous variation across counties and tracts in the United States. We decompose this spatial variation into two components: variation in business ideas per capita and in their rate of transition to startups. Both components matter for the variation in startups per capita. Furthermore, local demographic, economic, financial, and business conditions account for a significant fraction of the variation in startups per capita and in its components. In particular, income, education, age, and foreign-born share are strongly and positively associated with idea generation and transition rate. The relationship between local conditions and ideas generally differs in magnitude from the relationship with the rate at which these ideas transition into employer businesses. Interestingly, certain conditions are positively associated with ideas but negatively with transition rates. The predicted rank of locations based only on observable local conditions closely relates to the actual ranking of locations in terms of startups per capita, making it possible to characterize high-startup locations using observable local conditions alone.

About the Authors:

Emin Dinlersoz is a principal economist at the US Census Bureau's Center for Economic Studies. His main research areas are firm and industry dynamics, industry evolution, industrial and labor relations, entrepreneurship, and technology adoption.

Timothy Dunne is a professor of the practice at the University of Notre Dame. His research focuses on industrial organization.

John Haltiwanger is a Distinguished University Professor at the University of Maryland and a research associate of the National Bureau of Economic Research. His research focuses on firm dynamics and entrepreneurship with implications for labor markets and productivity.

Veronika Penciakova is a research economist at the Federal Reserve Bank of Atlanta. Her research interests lie in the areas of macroeconomics, firm dynamics, financial frictions, growth, and innovation.

Acknowledgements: Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the US 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 and CBDRB-FY23-055). They thank Richard Beem, Shawn Klimek, and participants at the 2021 and 2022 Southern Economic Association conferences and the University of North Carolina Kenan-Flagler Business School finance seminar for helpful comments. John Haltiwanger was also a Schedule A employee of the US Census Bureau at the time of the writing of this paper. They thank the Templeton Foundation for financial support.

Comments to the authors are welcome at veronika.penciakova@atl.frb.org *or* emin.m.dinlersoz@census.gov.

1 Introduction

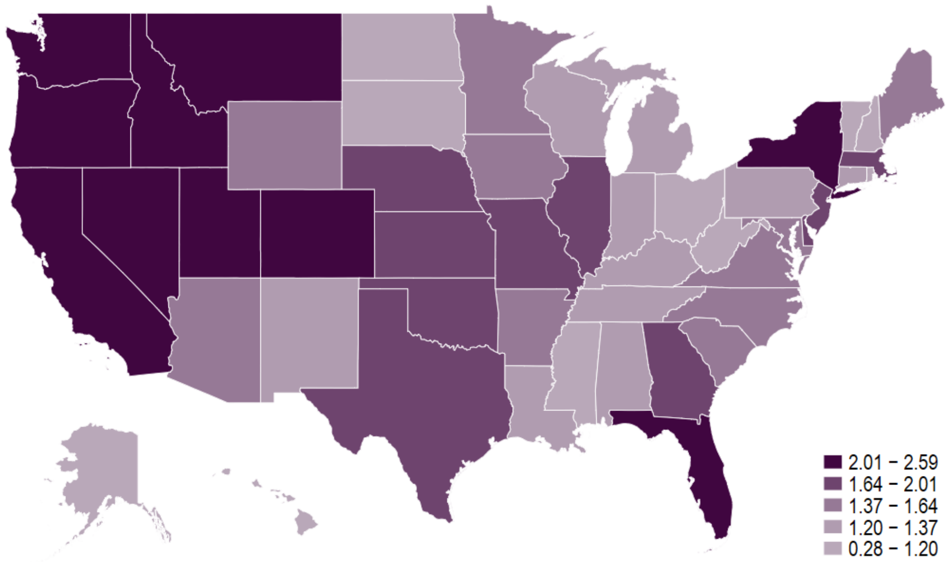
How much spatial inequality exists in early-stage entrepreneurial activity across the United States? Because startups contribute disproportionately to job creation and innovation, the extent of this inequality has implications for local growth. Yet the local conditions associated with the nascent stages of entrepreneurship are not well understood. In a recent working paper (Dinlersoz, Dunne, Haltiwanger, and Penciakova 2023 or henceforth DDHP (2023) for short), we study the spatial variation in early-stage business activity and entry to better understand the characteristics of environments that are conducive to business idea development and the transition of these ideas into startups.

Using micro data from the US Census Bureau, we decompose startup formation into two stages: idea generation and the rate at which ideas transition into employer businesses. Specifically, we use data from its Business Formation Statistics (BFS) program, which contains information on the universe of applications for new businesses and tracks whether, and when, these applications transition to employer startups. Focusing on the period 2010 through 2016, we express startup activity, defined as startups per 1,000 prime age adults (startups per capita, for short); as a product of ideas, defined as business applications (BA) per capita; and the transition rate of ideas (transition rate, for short), defined as the fraction of business applications that become employer businesses.

We find that there is enormous spatial variation in startup activity across the United States. Using public domain data from the BFS program, figure 1 documents the cross-state variation in startups per capita for the period 2010 through 2016. Some states generate less than 0.5 startups per capita, while others generate nearly 2.6 startups per capita. Variation is substantial even across states in the territory covered by the Sixth Federal Reserve District (Alabama, Florida, Georgia, and parts of Louisiana, Mississippi, and Tennessee), where startups per capita range from 1.2 in Mississippi to 2.6 in Florida.

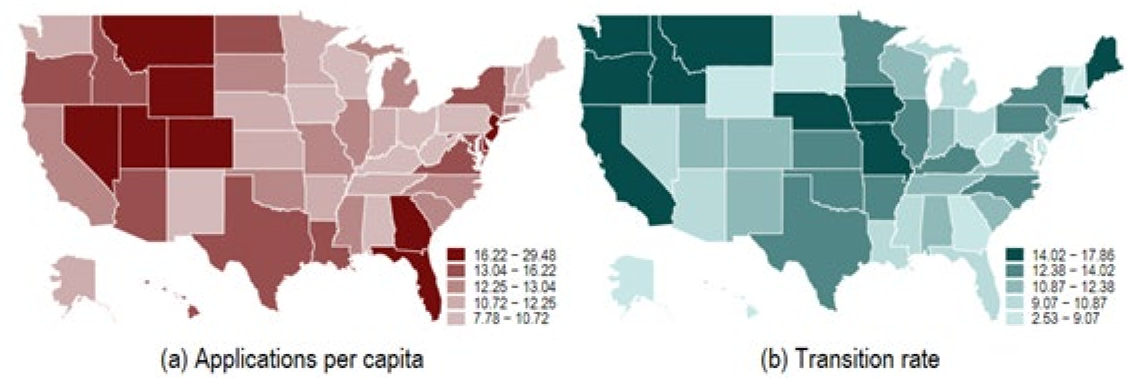
Strikingly, as shown in figure 2, which is also derived from public domain data, there is also substantial spatial variation in how business idea generation versus the transition rate contributes to startup intensity. Across states, similar startup intensity is apparent, with different combinations of idea generation intensity and transition rates. For example, Florida and Washington are characterized by high levels of startup activity, with more than two startups per capita. In Florida, this is driven by high applications per capita (25.4) despite a relatively low transition rate (10.1 percent), whereas in Washington the driver is lower applications per capita (11.9) and a higher transition rate (17.9 percent).

Figure 1: Startups per Capita, by State



Note: Depicts average business application startups per 1,000 prime-age population at the state level between 2010 and 2016. Startups are defined as applications that transition to employer businesses within eight quarters after application.
Source: Public domain Business Formation Statistics (BFS) data

Figure 2: Applications per Capita versus Transition Rate, by State



Note: Data depict (a) average BA per capita versus the (b) transition rate, measured as a percentage, at the state level between 2010 and 2016. Startups are defined as applications that transition to employer business within eight quarters after application.
Source: Public domain Business Formation Statistics (BFS) data.

In DDHP (2023), we delve into what accounts for this geographic variation in startups per capita, applications per capita, and transition rates. First, we show that both the variation in applications per capita and in the transition rates across counties contribute significantly to between-county dispersion in startups. Second, we establish that for all three—startups per capita, applications per capita, and transition rates—common metro area effects, captured by commuting zone by year effects, account for between 33 percent and 42 percent of between-county variation. Consequently, more localized factors explain between 58 percent and 67 percent of this variation. Third, after accounting for common area factors, we find that local observable (county by year) conditions account for about 14 percent of variation in startups per capita across counties. Local observable conditions account for about 18 percent of variation in applications per capita and about 3 percent of variation in transition rates. Fourth, our analysis of local observable conditions reveals local demographic and economic factors are the most important. For example, areas with a higher share of the population with a bachelor's degree systematically have higher startups per capita, applications per capita, and transition rates. Finally, we show that the ranking of counties in terms of startups per capita predicted by the local conditions we consider do well in predicting actual rankings.

2 Data Description

We use administrative micro data underlying the US Census Bureau's BFS program, which contains the universe of Employer Identification Number (EIN) applications. Critical for studying nascent entrepreneurial activity, all employer businesses in the United States are required to have an EIN to file payroll taxes. As such, BFS data offer coverage of most economically significant business initiations.

The application form includes the name and address of the business, application week, business start date, reason for application, type of business entity, industry, and planned date of initial wage payments. We use address information to assign a location (for example, county or tract) to each business idea. Because we are interested in employer startups, our analysis focuses on EIN applications that indicate a planned date for initial wage payments, referred to as wage business applications (WBA, for short).

To determine if and when EIN applications transition to employer businesses, the BFS program uses the Census Bureau's Longitudinal Business Database (LBD). The LBD contains establishment- and firm-level information on age, location, industry, number of employees, payroll, and EIN for nearly all employer businesses in the United States. Using the business applications linked to the LBD by EIN, BFS data identify the incidence and timing of transitions of applications to employer startups. We focus on transitions that occur within eight quarters of the application date because this window accounts for most applications that ever transition to an employer business. On an annual basis, the micro data track more than 2.5 million applications and more than 300,000 employer startups. Additional details of the microdata can be found in Bayard, Dinlersoz, Haltiwanger, Miranda, and Stevens (2018).

The Census Bureau publishes monthly tabulations publishes monthly tabulations of BFS of BA, WBA, and high-propensity business applications (HBA) at the national, state, and two-digit industry levels.[1] The public domain BFS data also tabulate employer startups that emerge from applications within the next four and eight quarters.

While the BFS program is a relatively new data source, it has been used in a number of recent studies that have focused on the aggregate time-series fluctuations in business idea generation. Asturias, Dinlersoz, Haltiwanger, and Hutchison (2023) explore the potential of the BFS program as a leading economic indicator, and Dinlersoz, Dunne, Haltiwanger, and Penciakova (2021) examine application and transitions during the previous two recessions. DDHP (2023) exploits the geographic granularity of BFS data assessing the relationship business idea creation and local characteristics at both the county and census tract level. Here, we summarize only the county-level analysis and results.

3 Spatial Variation in Startups, Ideas, and Transition Rates

The central premise of DDHP (2023) is that startup activity can be decomposed into two margins of nascent entrepreneurship: idea formation (proxied by applications per capita) and the transition of these ideas into employer businesses (proxied by the transition rate). In the working paper, we first test whether these two margins are independently informative about startup activity.

We find that idea origination and transition represent two distinct phases of entrepreneurship. Specifically, only one-third of applications (WBA) that transition within four years (or 16 quarters) of application do so in the same quarter that the application is submitted. Moreover, the vast majority of WBAs that transition to employer businesses do so within the first two years (or eight quarters).

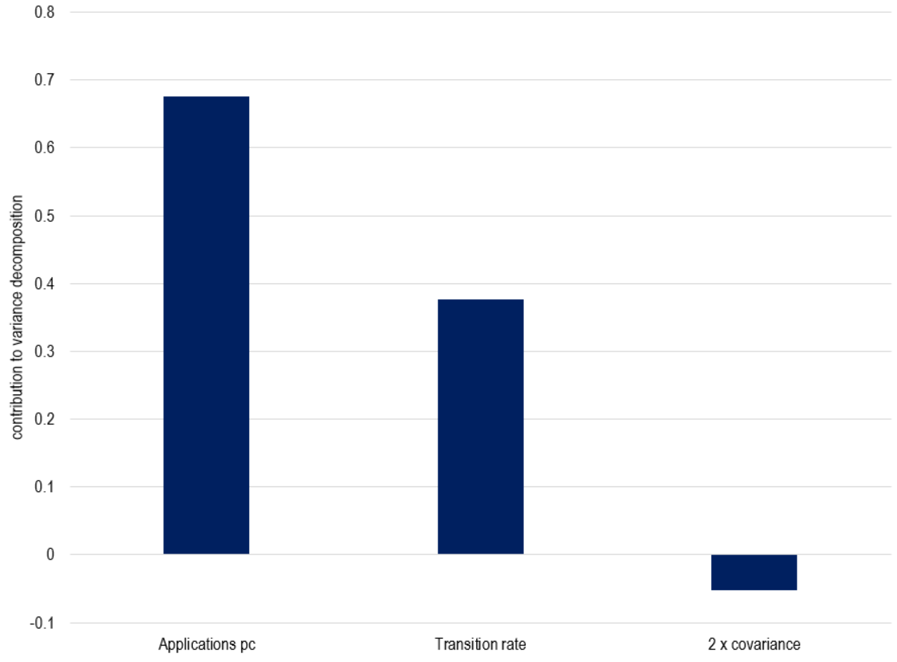
Next, we assess the relative contribution of applications per capita (A) and the transition rate (T) to variation in startup activity (S). Formally, we decompose the variation in S_{ct} as follows:

$$Var(\log S_{ct}) = Var(\log A_{ct}) + Var(\log T_{ct}) + 2Cov(\log A_{ct}, \log T_{ct})$$

where c is a county and t is a year between 2010 and 2016. As depicted in figure 1, about two-thirds of the cross-county variation in startups per capita is explained by applications per capita, with the remaining one-third being explained by the transition rate. Interestingly, the covariance between the two is small but negative, which is consistent with the evidence presented in figures 1 and 2 that similar levels of startups per capita can arise from different combinations of applications per capita and transition rates.

[1] The HBA include applications with characteristics that have a high propensity to become employer businesses.

Figure 3: Variance Decomposition for Startups per Capita



Note: "pc" refers to per capita (per 1,000 prime-age adults). The data report the variance decomposition of log(startups pc) into log(application pc) and log(transition rate), where all variables pertain to WBA and associated transitions. The underlying data are at the county-year level, covering 2010–16. Figures derive from analysis detailed in DDHP (2023).
Source: Administrative BFS and LBD data

4 The Role of Common Market and Local Conditions

In DDHP (2023), we then turn to a regression model to evaluate the contribution of local conditions in accounting for the spatial variation in startups per capita, applications per capita, and transition rates across counties between 2010 and 2016. Specifically, we consider the following regression model:

$$L_{czt} = \mathbf{\Phi} \boldsymbol{C}_{\boldsymbol{lt-k}} + \gamma_{zt} + \varepsilon_{lz}$$

where L is the outcome of interest—startups per capita, applications per capita, or transition rates—in county c that belongs to local labor market (for example, a commuting zone) z in year t; $\boldsymbol{C}_{\boldsymbol{lt-k}}$ is a set (or vector) of lagged local characteristics measured in year $t-k$. γ_{zt} are commuting zone by year fixed effects that account for time-varying factors that operate at the commuting zone level, such as regulations, labor market conditions, productivity spillovers, and agglomeration, among others.

We supplement BFS data with publicly available information on local conditions ($\boldsymbol{C}_{\boldsymbol{lt-k}}$) from a variety of sources including the American Community Survey (ACS), the US Bureau of

Economic Analysis (BEA), the Community Reinvestment Act (CRA), the Federal Reserve Board (FRB), and the LBD.[2] We divide local conditions into three categories:

- **Demographic** includes measures of median age, share of population with a bachelor's degree or higher, share of population with some college education, African American population share, Asian population share, Hispanic population share, and foreign-born population share.
- **Economic** includes measures of per capita income and the employment-to-population ratio.
- **Financial and business** includes measures of the household debt-to-income ratio, bank lending to small businesses as a share of small and medium-sized enterprise employment, share of employment in firms younger than five years, and the concentration of employment (Herfindahl index of employment).

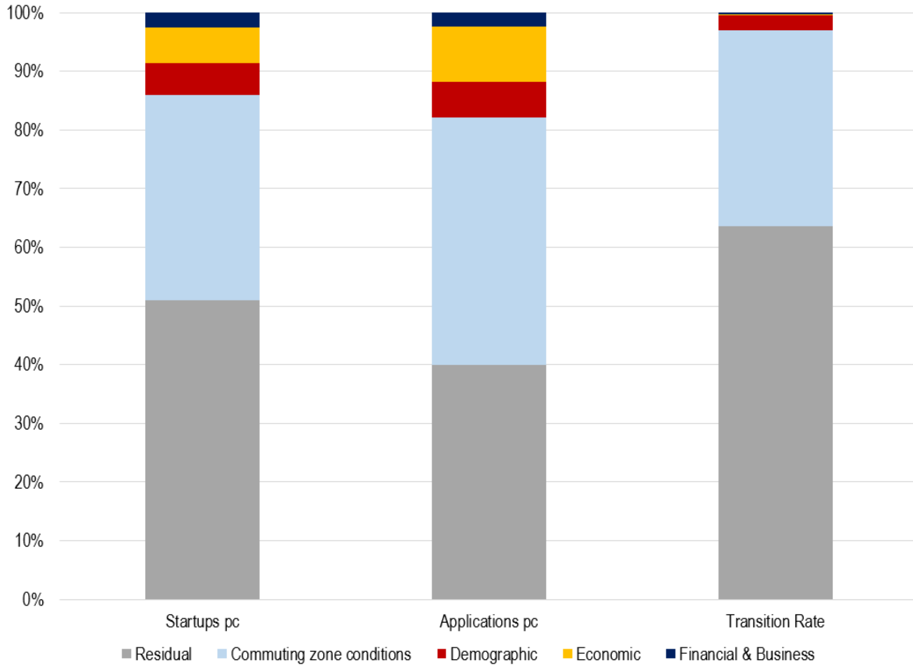
In figure 4, we use a variance decomposition methodology to evaluate the importance of different groups of covariates in explaining the variation in startups per capita, applications per capita, and transition rates.[3] The first notable finding is that across all three outcomes, local market conditions, which are captured by commuting zone conditions, account for between 33 percent and 42 percent of variation, suggesting that although broader regional factors are important in explaining nascent entrepreneurial activity, more localized conditions account for most of the variation (between 58 and 67 percent).

A second key finding from this analysis is that local conditions explain 14 percent of variation in startups per capita, with economic and demographic factors mattering the most. Much of the variation in startups per capita that is explained by local conditions is the result of the same factors explaining 18 percent of the variation in applications per capita. Meanwhile, local conditions explain far less—a mere 3 percent—of the variation in transition rates, with demographic factors being the most important set of observable local factors. It is possible that for transition rates, individual entrepreneur-level factors such as the quality of ideas matter more than local conditions.

[2] Note that local conditions obtained from the ACS are measured over five-year intervals ($k = 5$). Conditions obtained from all other sources are measured at an annual frequency, and therefore we use the average across the lags $k = 1, \dots, 5$.

[3] Please refer to Hottman, Redding, and Weinstein (2016) or Eslava, Haltiwanger, and Urdaneta (2023) for a discussion of the variance decomposition methodology. In summary, the decomposition methodology assigns each covariate the combination of the direct variance contribution plus half of the covariance with each of the other covariates. Consequently, this approach yields a decomposition where all terms (including the residual) sum to one.

Figure 4: County-Level Regression Decomposition



Note: "pc" refers to per capita (per 1,000 prime-age adults). Reports the contribution of groups of variables to total r-squared of regressions where the dependent variables are DHS(startups pc), DHS(applications pc), and transition rate for WBA and all control variables are included, along with commuting zone × 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. Figures derive from analysis detailed in DDHP (2023).
Source: Administrative BFS and LBD data, as well as public domain data from ACS, BEA, CRA, and FRB

An important third finding from the regression analysis is that systematic relationships exist among individual local conditions on the one hand, and startups per capita, applications per capita, and transition rates, on the other. DDHP (2023) provides a detailed discussion of our qualitative and quantitative findings, and here we highlight just a few of the interesting relationships we find.

- **Demographic conditions.** Startups per capita, applications per capita, and transition rates are all positively associated with the share of the population that holds a bachelor's degree or higher, as well as the foreign-born share of the population. Moreover, we find that age is positively associated with startups per capita through its positive association with applications per capita. We also find that the African American population share is negatively related to startups per capita and that this negative relationship arises because the positive relationship between the African American population share and applications per capita is more than offset by its negative relationship with the transition rate.
- **Economic conditions.** Both per capita income and the employment-to-population ratio are positively related to startups per capita via their positive relationship with applications per capita, but they are not significantly associated with the transition rate.

- **Financial and Business conditions.** We find that among this group of variables, the local concentration of employment is negatively associated with startups per capita, applications per capita, and the transition rate.

DDHP (2023) also reports a parallel analysis at the census tract level. The main set of results linking local conditions to startup activity, business idea formation, and transition rates is broadly similar to the county results. The key difference is the tract-level analysis examines the relationship between local conditions and entrepreneurial activity using the variation that occurs across census tracts within counties.

5 Local Conditions and Ranking of Locations

To provide additional perspective on the implications of our findings, in DDHP (2023) we evaluate how well local conditions account for the relative ranking of counties in terms of entrepreneurial activity. We do so by first ranking all counties based on their average startups per capita, applications per capita, and transition rates during the period 2010–16. We next group counties by decile based on their startups per capita rank, such that counties with the lowest startup activity end up in the lowest (1st) decile and those with the highest startup activity end up in the highest (10th) decile. We then calculate the average rank of actual applications per capita, as well as predicted applications per capita based, separately, on commuting zone by year fixed effects and observable local characteristics. We do the same for observed and predicted transition rates.

In table 1, we document that the average rank of both applications per capita and transition rates increases by startup decile, with counties in the lowest (highest) decile in startups per capita also ranking low (high) in applications per capita and transition rates. An interesting feature that emerges is that counties in the top decile of startups per capita are characterized more by their high rank in terms of applications per capita than their rank in terms of transition rates, which can be seen in the third row of the table, where the applications per capita rank is closer to the startups per capita rank than the transition rate rank is to the startups per capita rank. We also find that—despite accounting for only a portion of overall spatial variation in entrepreneurial activity—rankings predicted based on local conditions align closely with the actual ranking of counties in terms of startups per capita.

Table 1: County-Level Rank Analysis (Applications per Capita and Transition Rates)

	Average County Rank		
	(1)	(2)	(3)
Decile	Startups pc	Applications pc	Transition Rate
Bottom (1st)	149	562	593
Middle (5th)	1,332	1,318	1,456
Top (10th)	2,811	2,704	2,066

Note: The rank analysis focuses on WBA and associated transitions. "pc" refers to per capita (per 1,000 adults). The rows of the table are the bottom (1st), mid (5th), and top (10th) deciles of startups pc (averaged at the county level during 2010–16). Columns report the average rank of counties in each of these deciles for startups pc (column 1), applications pc (column 2), and transition rate (column 3).
Source: Administrative BFS and LBD data. Figures derive from analysis detailed in DDHP (2023).

6 Conclusion

In our recent paper, we use novel data on business applications that allow us to observe early stages of entrepreneurial activity and startup formation. Unlike many databases where only information about the incidence and timing of employer startups is available, we observe measures of the incidence and timing of both business idea creation and the transition of an idea into a startup. Not all ideas turn into employer businesses, and the transition typically takes time after the creation of the idea, suggesting that potential entrepreneurs assess the viability of their idea as they receive additional information during this gestational period and make decisions based on that information. Furthermore, proxies for the quality of a business idea are also available in the data, and we can identify business applications with a specific intent to hire employees.

Using this novel dataset, we first document the substantial variation in startups per capita across counties. We then show that both of the components of startup formation—business idea creation and the transition rate of ideas to employer businesses—are important in accounting for the variation in startups per capita across counties. Moreover, county-level demographic, economic, business, and financial conditions account for a significant fraction of the variation in startups per capita and that variation's components. Among the conditions we consider—income, education, age, and foreign-born share—have especially strong positive connection with business applications across counties. At the same time, the local conditions we consider relate to idea creation and transitions in distinct ways, and these relationships differ in magnitude and, in some cases, in sign. For example, one of our notable findings is that counties with a higher share of African Americans have a lower startup rate per capita. However, this association is driven by a positive relationship with applications per capita and a negative relationship with transitions, which overwhelms the former positive association.

Finally, we find that the predicted ranking of startups per capita based only on the parsimonious set of local observable conditions we consider is closely related to the actual ranking. This finding is useful for characterization of localities with high startup activity per capita. Policymakers and analysts exploring the sources of variation in entrepreneurship can thus use variation in these local observable conditions as a useful indicator of the startup potential of an area. However, it is important to emphasize that our results do not imply a causal relationship between local conditions and startup activity; exploring causality is a high priority for future research.

References

Asturias, Jose, Emin Dinlersoz, John Haltiwanger, and Rebecca Hutchison. 2023. Are business applications early economic indicators? *AEA Paper and Proceedings*, volume 113, pp. 151–55. doi.org/10.1257/pandp.20231053.

Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens. 2018. Early-stage business formation: An analysis of applications for Employer Identification Numbers. National Bureau of Economic Research working paper no. 24364. doi.org/10.3386/w24364.

Davis, Steven J., John Haltiwanger, and Scott Schuh. 1996. Small Businesses and Job Creation: Dissecting the Myth and Reassessing the Facts. *Small Business Economics,* volume 8(4), pp. 297–315. doi.org/10.1007/BF00393278.

Dinlersoz, Emin, Timothy Dunne, John Haltiwanger, and Veronika Penciakova. 2021. Business formation: A tale of two recessions. *AEA Paper and Proceedings*, volume 111, pp. 253–57. doi.org/10.1257/pandp.20211055.

———. 2023. The Local Origins of Business Formation. Federal Reserve Bank of Atlanta working paper no. 2023-9. doi.org/10.29338/wp2023-09.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta. 2023. The size and life-cycle growth of plants: The role of productivity, demand and wedges. *Review of Economic Studies*, volume 90(5), p. 2674. doi.org/10.1093/restud/rdad056.

Hottman, Colin, Stephen Redding, and David Weinstein. 2016. Quantifying the sources of firm heterogeneity. *Quarterly Journal of Economics*, volume 131(3), pp. 1291–364. doi.org/10.1093/qje/qjw012.