The Local Origins of Business Formation

Emin Dinlersoz
Bureau of the Census

Timothy Dunne
University of Notre Dame

John Haltiwanger
University of Maryland

Veronika Penciakova
Federal Reserve Bank of Atlanta

December 2023

Disclaimer: Any views expressed are those of the authors and do not represent the views of the institutions the authors are affiliated with. All results have been reviewed to ensure no confidential information is disclosed (DRB numbers: CBDRB-FY22-022 & CBDRB-FY23-055)

Motivation

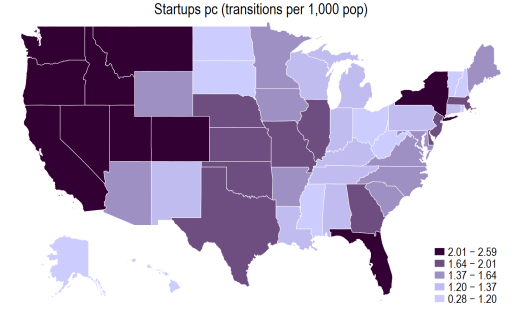
- How much spatial inequality is there in early-stage entrepreneurial activity in the U.S., and what local factors contribute to it?
  - Extent of inequality may have implications for local growth.

- Local conditions associated with nascent entrepreneurship are not well understood.
  - Burgeoning nascent entrepreneurship literature based on survey data (e.g., Bennett and Robinson, 2023).
  - We use comprehensive administrative data to be able to explore local spatial variation.

This Paper:

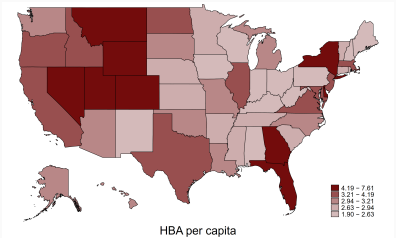
1. Documents enormous spatial variation in startups per capita the United States.
2. Decomposes startups into idea creation and transition rate using novel data (BFS).
3. Explores relationship between local conditions, startup activity and its components

Cross state variation in startups per capita, Avg 2010-16

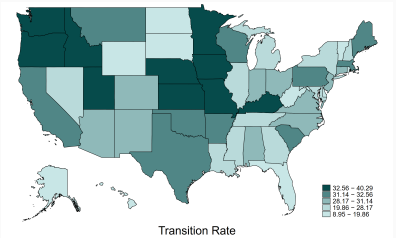


- High startup activity is concentrated in the West, NY, Massachusetts, and Florida.
- Using the microdata underlying the Business Formation Statistics (BFS), we will show that between state variation masks enormous variation within states.

Startups per capita, Ideas per capita, and transition rates, Avg 2010-16



(a) Applications (HBA) per capita



(b) Transition rate

- Startup activity depends both on the volume of ideas and the ability to transition these ideas into employer businesses. county-level
- Startups per capita = Business ideas per capita $\times$ Transition Rate. conceptual framework

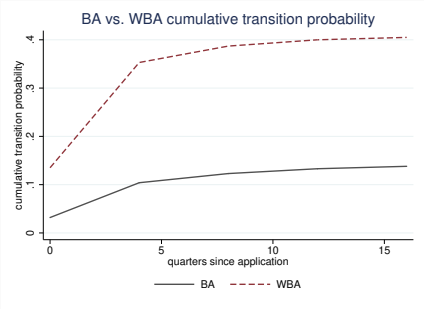
Business Formation Statistics (BFS), LBD, and external data

- **BFS**: SS4 applications for Employer Identification Numbers (EINs):
  - EINs are required for taxes, payroll, and banking.
  - Applications contain info on business characteristics, intent (*wages*), and location (*tract*).
    - BA=All Business Applications, WBA=Planned Wages/Hires, HBA=High Propensity (WBA,CBA, Selected Industries)
  - Interpret as signal about intent to form a business (*idea*)—both employer and non-employer.
  - We focus on WBA given our interest in startups of employer based businesses. Use BA as a cross check (but with risk of commingling applications intended for nonemployers).
- Link BFS to the Longitudinal Business Database (**LBD**):
  - LBD contains firm age, and establishment location, employment, and payroll.
  - Identify applications that transition to employer businesses within 8Q of application.
- External data on local conditions: ACS, CRA, BEA, FRB

Summary statistics for BFS county level data

BA vs. WBA Summary Stats

Variable	BA	WBA
Startups pc	1.34	0.88
Applications pc	10.48	2.15
Transition Rate	0.13	0.41



- Focus on 2010-2016 period at the county (and tract) level.
  - Location assigned based on application (= startup location in 80% – 90% of cases).
- WBA applications are fewer than BA, but have higher transition rates. additional stats
  - BA = business applications. WBA = business applications with intent to pay wages.
- Majority of BA and WBA transition within 8Q of application.

Cross county contribution of ideas pc and transitions rates to startups pc

Simple Variance Decomposition: $S_l = A_l T_l$

Variable	BA	WBA
Applications per 1,000 pop	0.485	0.676
Transition Rate	0.518	0.376
$2 \times$ Covariance	-0.004	-0.052

* all variables are in logs.

- Both ideas and transitions matter for spatial variation in startups.
- Ideas appear to matter relatively more for WBA than BA.
- The covariance between ideas and transitions is small (and negative). weighted

The role of local conditions: regression framework

$$\widetilde{Y}_{lzt} = \beta C_{lt-k} + f_{zt} + \epsilon_{lzt}$$

- Outcome variables ($\widetilde{Y}_{lzt}$): startups p.c., applications p.c., and transition rates. We use the following transformation to accommodate zeros: details
  - $\widetilde{Y} = 2\frac{(Y-\bar{Y})}{(Y+\bar{Y})}$, where $\bar{Y}$ is the grand mean of Y in the panel used for estimation.
  - Approach consistent with Chen and Roth (2023) and Tornqvist, Vartia, and Vartia (1985).
- Lagged local conditions (C_{lt-k}):
  - Demographic: age, education, race/ethnicity, foreign born.
  - Economic: income, e-pop ratio, owner-occupied share (tract).
  - Business: debt-to-income, concentration, young emp. share, small business lending.
- Area-year fixed effects (f_{zt}): CZ-year (county) or county-year (tract).

The role of local conditions: tract level regression decomposition

WBA:	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups			
Demographic	0.088	0.068	0.034
Economic	0.008	0.014	0.001
Categories			
Local Conditions	0.096	0.082	0.035
County-Year Conditions	0.100	0.191	0.104
Residuals	0.804	0.727	0.861

- Demographic conditions matter more for all three outcomes of interest.
- Creation of ideas depends more on observable conditions.
- Idiosyncratic factors (residual) matters more at the tract level.

contrib of county-year conditions only | county decomposition

The role of local conditions: selected tract level quantitative results

WBA (Tract): % Δ in LHS from 1 SD Δ in RHS

	(1) DHS(Startups pc)	(2) DHS(apps pc)	(3) transition rate
log(median age)	12.4	14.01	1.8
Bachelors+ share	19.3	16.4	5.0
African Am. share	−13.6	6.1	−13.2
Asian share	0.7	0.9	0.4
Hispanic share	−6.6	−3.9	−1.3
For Born share	3.0	3.9	−1.1
log(pc income)	7.1	2.2	3.1
emp-pop ratio	−2.1	1.7	−1.3
owner-occ. share	−21.1	−16.5	−2.9

* all regressions include county x year FE.

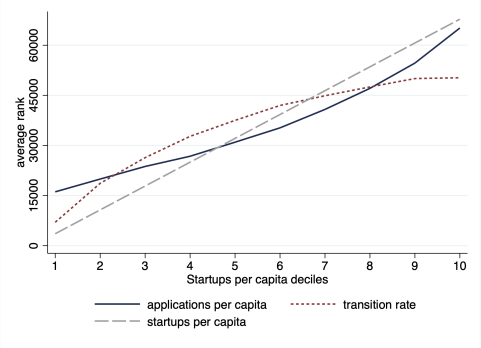
- Some conditions have opposite effects on applications p.c. and transition rates (e.g., African American share).
- Patterns broadly similar at county and tract level. county

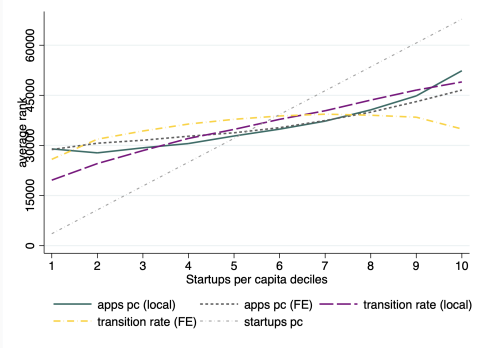
Digging deeper: application-level analysis

	WBA Transitions		
log(median age)	0.0198***	Hispanic share	−0.0246
	(0.00548)		(0.0185)
Bachelors+ degree share	0.0596***	Foreign born share	−0.0374**
	(0.00762)		(0.0156)
Some college share	−0.0478***	log(median HH income)	0.0208***
	(0.0111)		(0.00298)
African Am. share	−0.170***	Owner-occ. share	−0.0287***
	(0.00737)		(0.00431)
Asian share	0.0190	Emp-pop ratio	−0.0293***
	(0.0206)		(0.00734)
	Fixed effects	county x yr	
	R-squared	0.113	
	Within r-squared	0.0089	
	Obs	2,355,000	

- Run LPM of WBA transitions on application characteristics and local conditions.
- Application, tract, and county level correlations are broadly consistent.

Startups pc deciles: relative importance of apps pc and transition rates (tract)





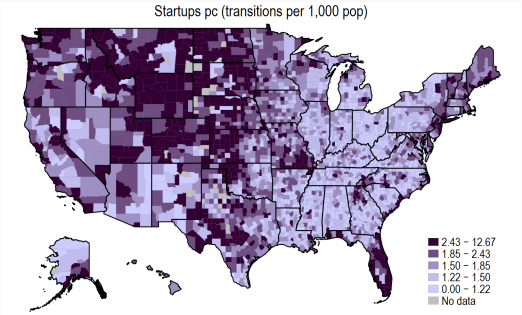
- High startup tracts are characterized more by applications per capita, while low startup per capita tracts are characterized more by transition rates.
- Local conditions are especially informative about applications per capita at the top decile, and transition rates at both the bottom and top deciles. county

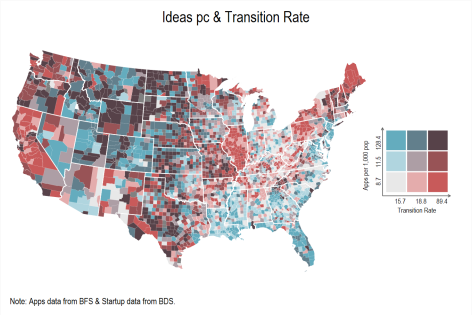
Concluding Remarks

- Little is understood about the nascent stages of entrepreneurship.
  - This paper opens the blackbox of the nascent process by focusing on spatial variation.
  - Decomposes startups into idea creation and transition of ideas into a startup.
  - Exploits BFS microdata and spatial variation to study the local conditions that are conducive to startups pc, ideas pc, and transition rates.
- Key Findings:
  - Enormous spatial dispersion in startups p.c., applications p.c. at the local level.
  - Applications p.c. and transition rates about equally important in accounting for variation.
  - Local conditions account for more variation in ideas than transitions.
    - Application process helps potential entrepreneurs learn about the quality of their ideas.
    - Local conditions impact idea creation vs transitions differentially.
  - Even though large unexplained variation, local conditions help in identifying locations with high startup activity.

Thank You

County level: startups per capita, ideas per capita, and transition rates





- There is substantial county level spatial variation in startups pc, applications pc, and transition rates. return

Conceptual Framework

- Opening up the *black box* of standard (e.g., Hopenhayn (1992)) model of entry process and costs.
  - Critically permit this entry process and cost to reflect local conditions.
- Potential entrepreneurs have ideas drawn from distribution with varying quality.
- Make an investment to learn about quality of idea relative to next best alternative (including taking into account the costs of starting up business).
  - This investment includes making the application.
- After getting signal about net return, ideas with positive net returns yield startups.
- Local conditions influence both the nascent (learning) phase and the startup phase.
- Local conditions may not have the same effect on two phases. Some conditions may favor/impede learning phase but be adverse factors for startup phase.

return full model

Additional summary statistics for BFS county level data

	WBA			BA		
	Startups pc	Apps pc	Transition rate	Startups pc	Apps 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

- Across counties, observe substantial variation in startups, ideas, and transitions.
- Variation across tracts even larger – most between tract variation not accounted for by county effects. return

Contribution of ideas p.c. and transitions rates to startups p.c. (weighted)

Variable	BA	WBA
Applications per 1,000 pop	0.983	0.919
Transition Rate	0.428	0.369
$2 \times$ Covariance	-0.410	-0.288

* all variables are in logs.

- Population weighting increases the relative importance of ideas. return

Measures of relative difference for applications per capita & startups per capita

Dispersion in Transformed Variables

Variable	County	Tract
Startups Per Capita	0.47	0.78
Applications Per Capita	0.59	0.97

- Recall that we accommodate zeros at the county and tract level for startups pc and applications pc, we use the transformation $\widetilde{Y} = 2\frac{(Y-\bar{Y})}{(Y+\bar{Y})}$
- We observe substantially more dispersion across tracts than counties. return

The role of county-year conditions: tract level

WBA:	Include local conditions No	Yes
DHS(startups pc)	0.111	0.100
DHS(apps pc)	0.208	0.191
Transition Rate	0.108	0.104

- Contribution of county-year conditions is very similar with and without controlling for local conditions. return

The role of local conditions: county level regression decomposition

WBA:	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups			
Demographic	0.055	0.061	0.025
Economic	0.060	0.094	0.002
Business	0.026	0.024	0.003
Categories			
Local conditions	0.141	0.179	0.030
Commuting zone conditions	0.350	0.421	0.333
Residual	0.509	0.400	0.636

- Demographic and economic conditions matter more than business conditions.
- Local conditions account for much more of apps. p.c. (17%) than transition rates (3%).
- Perhaps transition rates depend more on idiosyncratic factors (i.e. quality of ideas).

return

The role of local conditions: selected county level quantitative results

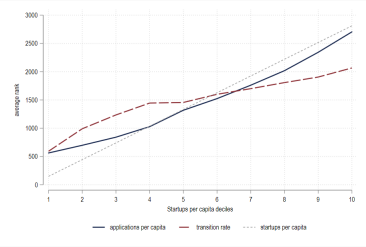
WBA (County): % Δ in LHS from 1 SD Δ in RHS

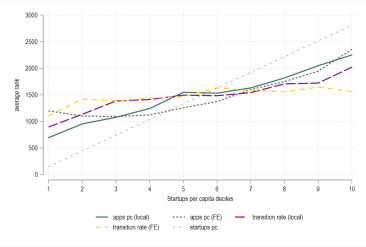
	(1) DHS(Startups pc)	(2) DHS(apps pc)	(3) transition rate
log(median age)	5.9	6.9	-0.1
% college	11.4	8.2	2.2
% pop Afr Am	-6.2	6.0	-9.7
% pop Asian	-5.0	-3.3	-2.1
% pop Hispanic	-4.4	-0.5	-6.5
% pop For Born	7.0	4.2	2.7
log(pc income)	10.3	11.6	-0.3
emp-pop ratio	8.0	6.3	1.4
% emp young firms	-0.6	1.5	-1.2
HHI (emp)	-7.6	-4.8	-2.1
small business lending	3.4	2.2	1.6
log(debt/income)	-2.0	-0.9	-0.8

* all regressions include CZ x year FE.

- Some conditions have opposite effects on applications p.c. and transition rates (e.g., African American Share). return

Startups pc deciles: relative importance of apps pc and transition rates (county)





- Correlates of nascent entrepreneurship in low vs. high startup locations:
  - *Superstar* counties characterized more by applications per capita than transition rates.
  - Local conditions and FEs for apps pc are similarly informative, while observables are more informative than FEs for transition rates. return

Sketch of Model (1)

- In location l ex ante distribution of potential ideas $F_l(\iota)$. To pursue idea must make investment I_l. Pursuing idea yields signal of value of idea V.
- Entrepreneur has reservation value R_l

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

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) if $\mathcal{V}_l(\iota) \geq R_l + I_l$. The marginal idea then satisfies

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

Sketch of the Model (2)

- The mass of pursued ideas (or 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

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

- 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. return