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

3 October 2024

Disclaimer: Any views expressed are those of the authors. They 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 (project number: 7530035, and associated DRB number: CBDRB-FY22-022, CBDRB-FY23-055 & CBDRB-FY24-0354)

Motivation

- What is the extent of spatial variation in early-stage entrepreneurial activity, and what local factors contribute to it?
  - Variation in nascent entrepreneurship has implications for local growth.
- Local conditions associated with nascent entrepreneurship are not well understood.
  - *Nascent entrepreneurship*: early stages when entrepreneurs develop ideas and decide whether to start firms.

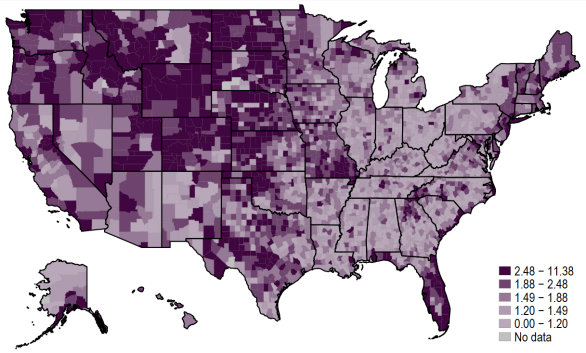
Motivation

- What is the extent of spatial variation in early-stage entrepreneurial activity, and what local factors contribute to it?
  - Variation in nascent entrepreneurship has implications for local growth.
- Local conditions associated with nascent entrepreneurship are not well understood.
  - *Nascent entrepreneurship*: early stages when entrepreneurs develop ideas and decide whether to start firms.

Using administrative data on potential entrants and business entry, this paper:

1. Decomposes startups into idea creation and transition rate using novel data (BFS).
   - Documents contribution of each component to spatial variation in startup activity.
2. Explores relationship between local conditions, startup activity and its components
   - Reports the contribution of local condition to spatial variation in nascent entrepreneurship.

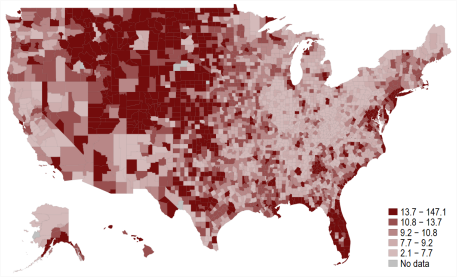
Cross county variation in startups per 1,000 people



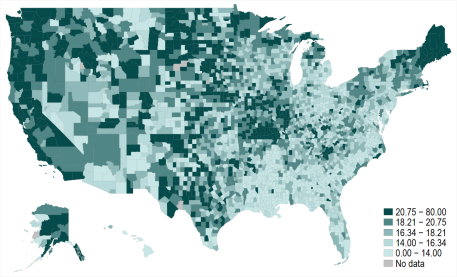
- High startup activity is concentrated in the West, NE corridor, and Florida.

Cross county variation in ideas per capita and transition rates

(a) Business ideas per capita



(b) Transition rate (%)



- Startups per capita = Business ideas per capita $\times$ Transition Rate. conceptual framework
  - West characterized by relatively low ideas pc and high transition rates.
  - South characterized by relatively high ideas pc and low transition rates.

Data Description

- **Business Formation Statistics (BFS)**:

1. Applications for Employer Identification Numbers (EINs) from IRS Form SS4:
    - EINs are required for taxes, payroll, and banking.
    - Applications contain info on business characteristics, intent (*wages*), and location (*tract*).
        - Location assigned based on application (= startup location in 80%-90% of cases).
    - Interpret as signal about intent to form a business (*idea*).
        - BA = business applications. WBA = business applications with intent to pay wages.

Data Description

- **Business Formation Statistics (BFS)**:

1. Applications for Employer Identification Numbers (EINs) from IRS Form SS4:
   - EINs are required for taxes, payroll, and banking.
   - Applications contain info on business characteristics, intent (*wages*), and location (*tract*).
     - Location assigned based on application (= startup location in 80%-90% of cases).
   - Interpret as signal about intent to form a business (*idea*).
     - BA = business applications. WBA = business applications with intent to pay wages.
2. Link EINs 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. details

Data Description

- **Business Formation Statistics (BFS)**:
  1. Applications for Employer Identification Numbers (EINs) from IRS Form SS4:
     - EINs are required for taxes, payroll, and banking.
     - Applications contain info on business characteristics, intent (*wages*), and location (*tract*).
       - Location assigned based on application (= startup location in 80%-90% of cases).
     - Interpret as signal about intent to form a business (*idea*).
       - BA = business applications. WBA = business applications with intent to pay wages.
  2. Link EINs 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. details
- **External data** on local conditions: ACS, CRA, FRB.

BFS summary statistics

	BA	WBA
Startups per capita (1,000 people)	1.58	0.94
Applications per capita	13.24	2.28
Transition rate	0.12	0.37

- Focus on 2011-2016 period at the tract levels.
  - Use tract-level data to proxy for potential entrepreneur characteristics, and local consumer, product, and regulatory conditions.
- WBA applications are fewer than BA, but have higher transition rates.
  - We focus on WBA given our interest in startups of employer based businesses.

Contribution of ideas pc and transitions rates to startups pc: unweighted WBA

Variance Decomposition: Startups pc = Ideas pc $\times$ Transition Rate

	BA	WBA
Applications per 1,000 pop	0.58	0.66
Transition rate	0.42	0.33
2 $\times$covariance	-0.002	0.02

* all variables are in logs.

- Both ideas and transitions help explain spatial variation in startups.
- The covariance between ideas and transitions is very small.
- Decomposition is similar for BA and WBA. weighted

The role of local conditions: regression framework

$$\widetilde{Y}_{lzt} = f_{zt} + \mathbf{\Phi C^o_{lt-k}} + \epsilon_{lzt}$$

- Outcome variables ($\widetilde{Y}_{lzt}$): startups p.c., applications p.c., and transition rates. Note that we use a transformation to accommodate zeros. details
- Fixed effects (f_{zt}): county-year FE.
- Lagged own-tract local conditions (C^o_{lt-k}):
  - Demographic: age, education, race, ethnicity, foreign born.
  - Household economic conditions: income pc, emp-to-pop, owner occupied housing share.
  - Incumbent firm characteristics: share emp. in young firms, share emp. in large firms, avg. firm size, industry emp. shares.
  - Commercial share: emp/(emp + pop)

The role of local conditions: regression framework

$$\widetilde{Y}_{lzt} = f_{zt} + \mathbf{\Phi C^o_{lt-k}} + \mathbf{\Lambda C^n_{lt-k}} + \epsilon_{lzt}$$

- Outcome variables ($\widetilde{Y}_{lzt}$): startups p.c., applications p.c., and transition rates. Note that we use a transformation to accommodate zeros. details
- Fixed effects (f_{zt}): county-year FE.
- Lagged own-tract local conditions (C^o_{lt-k}):
  - Demographic: age, education, race, ethnicity, foreign born.
  - Household economic conditions: income pc, emp-to-pop, owner occupied housing share.
  - Incumbent firm characteristics: share emp. in young firms, share emp. in large firms, avg. firm size, industry emp. shares.
  - Commercial share: emp/(emp + pop)
- Lagged neighboring-tract local conditions (C^n_{lt-k}): same set of covariates as in $\mathbf{C^o_{lt-k}}$.

The role of local conditions: WBA regression decomposition

	Own Tract Only			Own & Neighboring Tract		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups						
Demographic	0.029	0.009	0.029	0.025	0.003	0.024
HH economic	0.029	0.039	0.002	0.025	0.034	0.001
Incumbent firm	-0.011	-0.020	0.002	-0.010	-0.019	0.002
Commercial share	0.171	0.235	0.007	0.170	0.232	0.007
Categories						
Own local conditions	0.219	0.262	0.041	0.209	0.251	0.035
Neighboring local conditions	.	.	.	0.013	0.018	0.007
Fixed effects	0.086	0.154	0.104	0.086	0.152	0.104
Residual	0.695	0.584	0.856	0.692	0.579	0.854

* follow methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024).

- Own local conditions: demographic and HH economic conditions matter most.

The role of local conditions: WBA regression decomposition

	Own Tract Only			Own & Neighboring Tract		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups						
Demographic	0.029	0.009	0.029	0.025	0.003	0.024
HH economic	0.029	0.039	0.002	0.025	0.034	0.001
Incumbent firm	-0.011	-0.020	0.002	-0.010	-0.019	0.002
Commercial share	0.171	0.235	0.007	0.170	0.232	0.007
Categories						
Own local conditions	0.219	0.262	0.041	0.209	0.251	0.035
Neighboring local conditions	.	.	.	0.013	0.018	0.007
Fixed effects	0.086	0.154	0.104	0.086	0.152	0.104
Residual	0.695	0.584	0.856	0.692	0.579	0.854

* follow methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024).

- Own local conditions: demographic and HH economic conditions matter most.
- Neighboring local conditions: contribute relatively little.

The role of local conditions: WBA regression decomposition

	Own Tract Only			Own & Neighboring Tract		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups						
Demographic	0.029	0.009	0.029	0.025	0.003	0.024
HH economic	0.029	0.039	0.002	0.025	0.034	0.001
Incumbent firm	-0.011	-0.020	0.002	-0.010	-0.019	0.002
Commercial share	0.171	0.235	0.007	0.170	0.232	0.007
Categories						
Own local conditions	0.219	0.262	0.041	0.209	0.251	0.035
Neighboring local conditions	.	.	.	0.013	0.018	0.007
Fixed effects	0.086	0.154	0.104	0.086	0.152	0.104
Residual	0.695	0.584	0.856	0.692	0.579	0.854

* follow methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024).

- Own local conditions: demographic and HH economic conditions matter most.
- Neighboring local conditions: contribute relatively little.
- Fixed effects: most important for ideas.
- Residual: most important for transition rate.

The role of local conditions: selected quantitative results

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

	DHS(startups pc)	DHS(apps pc)	Transition rate
median age	0.686	1.764	0.431
bachelors+ share	6.318	1.862	4.322
some college share	-1.911	-1.607	-0.886
African American share	-6.349	12.047	-12.210
Asian share	0.579	0.965	0.193
Hispanic share	-3.440	-1.376	-1.101
foreign born share	4.567	5.236	-1.003
per capita income	15.385	17.405	0.673
emp-pop ratio	-3.276	-1.693	-0.673
owner-occupied share	-0.070	-0.739	1.866

* all regressions include county x year FE.

- Some conditions have reinforcing effects (e.g. bachelors+ share)
- Others have opposite effects (e.g. AA share and foreign born share).

application level regression | duration regression

Example of industry variation: high-tech regression decomposition

- Anticipate cross-industry differences in the relevance of potential entrepreneur characteristics and local consumer, product and regulatory conditions.

Example of industry variation: high-tech regression decomposition

	All Industries			High-tech Industries		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
Groups						
Demographic	0.029	0.009	0.029	0.014	0.035	0.008
HH economic	0.029	0.039	0.002	0.002	0.021	-0.000
Incumbent firm	-0.011	-0.020	0.002	0.004	0.008	0.003
Commercial share	0.171	0.235	0.007	0.027	0.026	0.001
Categories						
Own local conditions	0.219	0.262	0.041	0.047	0.090	0.011
Fixed effects	0.086	0.154	0.104	0.089	0.103	0.087
Residual	0.695	0.584	0.856	0.864	0.808	0.902

* follow methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024).

- Anticipate cross-industry differences in the relevance of potential entrepreneur characteristics and local consumer, product and regulatory conditions.
- High-tech sector application: demographic characteristics are especially important.
  - *However, local conditions account for less variation in high tech nascent entrepreneurship.*

Example of industry variation: high-tech county level regression results

	All industries			High-tech		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
log(median age)	0.0349	0.0897***	0.00803	0.0721	-0.361***	0.016
	(0.0317)	(0.0294)	(0.00662)	(0.0629)	(0.0286)	(0.0194)
bachelors or higher share	0.341***	0.0994***	0.0875***	0.362***	0.922***	0.0958***
	(0.0426)	(0.0374)	(0.0107)	(0.098)	(0.0504)	(0.0312)
Asian pop share	0.0585	0.106	0.00785	0.429***	0.607***	0.0710**
	(0.120)	(0.0969)	(0.0231)	(0.147)	(0.212)	(0.0336)
share of emp in large firms	-0.530***	-0.550***	-0.0243***	-0.124**	-0.190***	0.00393
	(0.024)	(0.0176)	(0.00491)	(0.0516)	(0.0241)	(0.0169)
Observations	398,000	430,000	398,000	64,000	430,000	64,000
Fixed effects	county x year	county x year	county x year	county x year	county x year	county x year
R-squared	0.305	0.416	0.1443	0.1365	0.1922	0.0976

* regressions include all variables; table reports subset of coefficients

- For high tech sample:
  - Stronger correlation: bachelors+ share and Asian pop share.
  - Weaker correlation: emp share in large firms and median age.

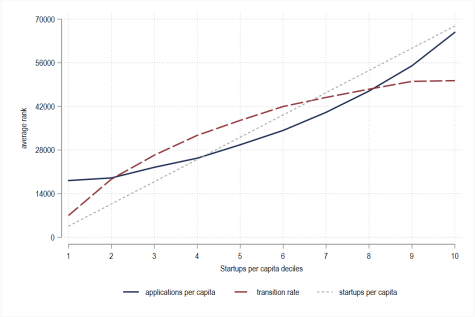
Example of industry variation: high-tech county level regression results

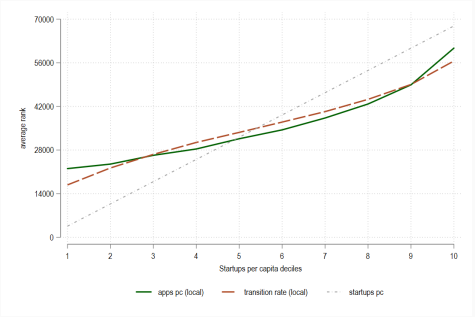
	All industries			High-tech		
	DHS(startups pc)	DHS(apps pc)	Transition rate	DHS(startups pc)	DHS(apps pc)	Transition rate
log(median age)	0.0349	0.0897***	0.00803	0.0721	-0.361***	0.016
	(0.0317)	(0.0294)	(0.00662)	(0.0629)	(0.0286)	(0.0194)
bachelors or higher share	0.341***	0.0994***	0.0875***	0.362***	0.922***	0.0958***
	(0.0426)	(0.0374)	(0.0107)	(0.098)	(0.0504)	(0.0312)
Asian pop share	0.0585	0.106	0.00785	0.429***	0.607***	0.0710**
	(0.120)	(0.0969)	(0.0231)	(0.147)	(0.212)	(0.0336)
share of emp in large firms	-0.530***	-0.550***	-0.0243***	-0.124**	-0.190***	0.00393
	(0.024)	(0.0176)	(0.00491)	(0.0516)	(0.0241)	(0.0169)
Observations	398,000	430,000	398,000	64,000	430,000	64,000
Fixed effects	county x year	county x year	county x year	county x year	county x year	county x year
R-squared	0.305	0.416	0.1443	0.1365	0.1922	0.0976

* regressions include all variables; table reports subset of coefficients

- For high tech sample:
  - Stronger correlation: bachelors+ share and Asian pop share.
  - Weaker correlation: emp share in large firms and median age.
- Other sectors (in progress): non-tradables; Hurst & Pugsley sectors.

Startups pc deciles: importance of apps pc and transition rates





- High startup pc tracts: characterized more by high apps pc.
  - At the top, local conditions are especially informative about apps pc.
- Low startup pc tracts: characterized by low transition rates
  - At the bottom, local conditions are informative about transition rates. mobility deciles

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 distinctly important in accounting for variation.
  - Local conditions account for more variation in ideas than transitions.
    - Local conditions impact idea creation vs transitions differentially.
  - Local conditions help in identifying locations with high startup activity and social mobility.

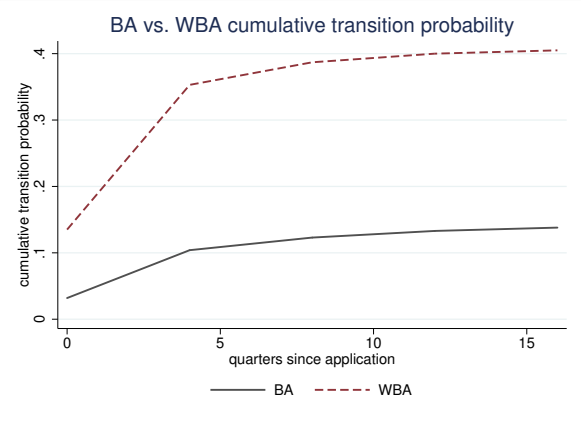
Thank You

Conceptual Framework

- Open the *black box* of standard model of entry process and costs (Hopenhayn, 1992).
  - Critically permit this entry process and cost to reflect local conditions.
- Key elements of the framework:
  1. Potential entrepreneurs have ideas drawn from distribution with varying quality.
  2. Make an investment to learn about quality of idea relative, including taking into account the costs of starting up business.
  3. After getting signal about net return, ideas with positive net returns yield startups.
- Role of local conditions:
  - 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 the learning phase but impede the startup phase.

return full model

BFS cumulative transition probabilities



- WBA applications are fewer than BA, but have higher transition rates.
- Majority of BA and WBA transition within 8Q of application. return

Contribution of ideas pc and transitions rates to startups pc: weighted

Simple Variance Decomposition: $S_l = A_l T_l$

	BA	WBA
Applications per 1,000 pop	0.59	0.65
Transition rate	0.42	0.32
2 $\times$covariance	-0.007	0.03

* all variables are in logs.

- Weighting does not alter tract results (expected given similarity in size across tracts).

return

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

Dispersion in Transformed Variables

Variable	Tract
WBA Startups per capita	0.977
WBA per capita	0.782

- Recall that we accommodate zeros for startups pc and applications pc, we use the transformation $\widetilde{Y} = 2\frac{(Y-\bar{Y})}{(Y+\bar{Y})}$
- We observe substantial dispersion across tracts. return

Application-level analysis

WBA Transitions			
log(median age)	0.003	log(per capita income)	0.004
	(0.00722)		(0.00376)
bachelors or higher share	0.0644***	emp-pop ratio	-0.0121*
	(0.0097)		(0.00664)
some college share	-0.0332***	owner-occupied share	0.0236***
	(0.0124)		(0.00359)
African American share	-0.159***	share of emp in young firms	0.0145***
	(0.00698)		(0.00538)
Asian share	0.012	share of emp in large firms	-0.0100**
	(0.0193)		(0.00394)
Hispanic share	-0.022	DHS(avg firm emp)	-0.00811***
	(0.0156)		(0.00108)
foreign born share	-0.0334**	commercial share	0.0794***
	(0.0142)		(0.00384)
	Ind emp. shares	Yes	
	Observations	2,355,000	
	R-squared	0.113	
	Within R-squared	0.0098	

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

Digging deeper: duration analysis at the tract level

WBA duration: tract level			
log(median age)	-0.0556**	log(per capita income)	0.0373*
	(0.0246)		(0.0192)
bachelors or higher share	0.342***	emp-pop ratio	0.0763*
	(0.0442)		(0.0399)
some college share	0.219***	owner-occupied share	-0.0565***
	(0.0517)		(0.0199)
African American share	0.230***	share of emp in young firms	0.0971***
	(0.0313)		(0.0294)
Asian share	0.146**	share of emp in large firms	0.0112
	(0.0672)		(0.0232)
Hispanic share	-0.0709	DHS(avg firm emp)	0.0109
	(0.0686)		(0.0071)
foreign born share	0.225***	commercial share	-0.146***
	(0.0808)		(0.0246)
	Observations	309,000	
	Ind emp. shares	yes	
	Fixed effects	fips x yr	
	R-squared	0.078	
	Within R-squared	0.003109	

- Local determinants of WBA duration (= avg. # of quarters from app. to transition). return

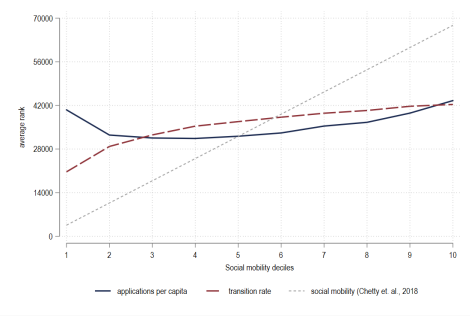
Digging deeper: high-tech regression detailed decomposition

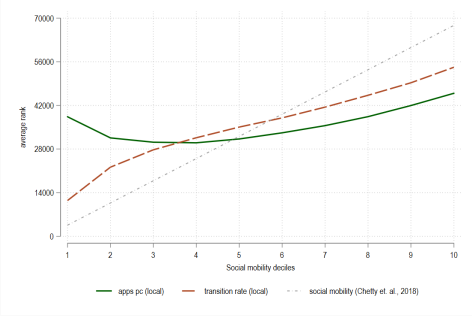
	County			Tract		
	DHS(startups pc)	DHS(apps pc)	Transition Rate	DHS(startups pc)	DHS(apps pc)	Transition Rate
All industries						
Demographic	0.056	0.076	0.024	0.029	0.009	0.029
HH economic	0.033	0.047	0.000	0.029	0.039	0.002
Incumbent firm	0.072	0.073	0.024	-0.011	-0.020	0.002
Commerical share				0.171	0.235	0.007
High tech						
Demographic	0.086	0.097	0.013	0.014	0.035	0.008
HH economic	0.018	0.007	-0.002	0.002	0.021	0.000
Incumbent firm	0.064	0.147	0.028	0.004	0.008	0.003
Commerical share				0.027	0.026	0.001

* follow methodology in Hottman, Redding and Weinstein (2016) and Eslava, Haltiwanger and Urdaneta (2024).

- Demographic and incumbent firm characteristics explain more variation in high tech apps pc than overall apps pc. return

Mobility deciles: importance of apps pc and transition rates—tract level





- Low social mobility tracts (Chetty et. al., 2014) characterized by low transition rates.
- Across social mobility deciles, local conditions are informative about 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