Startup Location, Local Spillovers and Neighborhood Sorting

Rolando Campusano

University of Toronto (Rotman)

October 4, 2021

Agglomeration spillovers

- Firms co-locate partly due to agglomeration spillovers

  Ellison and Glaeser (1999), Duranton and Overman (2005), and Greenstone, Hornbeck, and Moretti (2010)

- Some of these spillovers stimulate the birth and success of startups

  Porter (1998), Henderson, Kuncoro, and Turner (1995), Chatterji (2009), and Chatterji, Glaeser, and Kerr (2014)

  Glaeser, Kerr, and Kerr (2015)

- And evidence show that these effect are very local and decay rapidly

  Rosenthal and Strange (2003), Arzaghi and Henderson (2008), and Catalini (2018)

Do they apply to Startups?

- Location is a difficult choice entrepreneurs face

  Figueiredo, Guimarães, and Woodward (2002) and Audretsch, Lehmann, and Warning (2005)

- Public and private efforts recognize it while aiming for co-location of new firms

  'Big Push'(Murphy, Shleifer, and Vishny 1989), SBIR matching programs (Lanahan and Feldman 2015),
  Startup X,Y,Z, Co-working and Acceleration labs

- There is still more to know about if (and if so, how) these spillovers affects startups

  Glaeser, Kerr, and Ponzetto (2010), Chatterji, Glaeser, and Kerr (2014), and Kerr and Kominers (2015)

This paper: **How critical are location attributes for startups?**

Ideal Experiment

Twin startups in *random* locations

This paper: **How critical are location attributes for startups?**

However, in reality we only observe *chosen* locations

This paper: **How critical are location attributes for startups?**

Possible approaches

- Find real or synthetic counterfactuals
  Greenstone, Hornbeck, and Moretti (2010), or Heblich et al. (2019) and Qian and Tan (2021)
- Use within geography variation
  Arzaghi and Henderson (2008), Liu, Rosenthal, and Strange (2017), and Roche (2020)

This paper: **How critical are location attributes for startups?**

Possible approaches

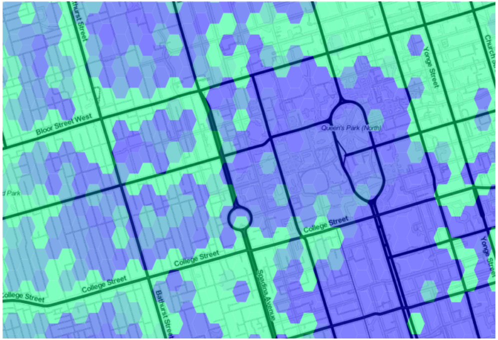
- Find real or synthetic counterfactuals
  Greenstone, Hornbeck, and Moretti (2010), or Heblich et al. (2019) and Qian and Tan (2021)
- Use within geography variation
  Arzaghi and Henderson (2008), Liu, Rosenthal, and Strange (2017), and Roche (2020)

This paper mixes both: distinction between neighborhoods and blocks

- Zoom IN: Block-level spillovers are identified within neighborhoods
- Zoom OUT: Sorting into neighborhoods based on entrepreneurs' utility maximization

Zoom IN: Local (block) spillovers are identified within neighborhoods

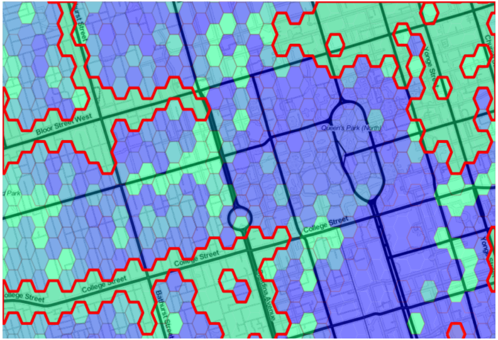
- algorithm *cluster* blocks as a collection of *adjacent counterfactual* locations
  extension of (Ward 1963) to include adjacency constraints
- based on propensity score using historical data for each city and industry
  spatial version of (Rosenbaum and Rubin 1984)



Zoom IN: Local (block) spillovers are identified within neighborhoods

- exclusion restriction: choices orthogonal to within variation of block attributes

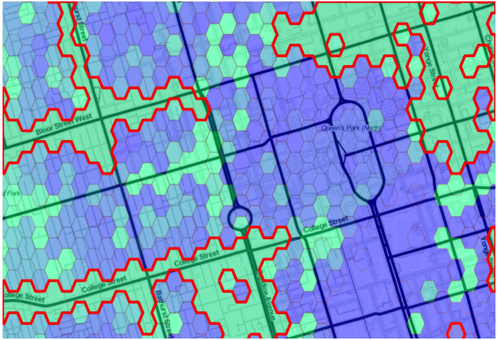
  Bayer, Ross, and Topa (2008), Liu, Rosenthal, and Strange (2018), and Roche (2020)



Zoom IN: Local (block) spillovers are identified within neighborhoods

Spoiler Alert! - local (block) spillovers are important for startups

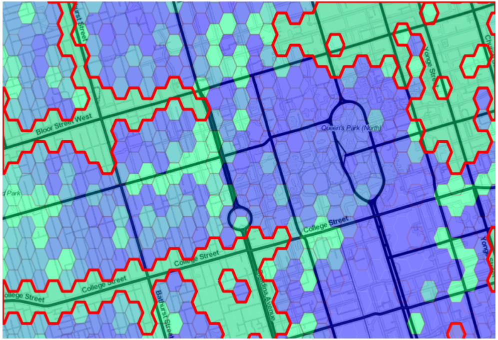
- elasticity revenue to average employment is .12 and to average revenue is .03
  - higher survival rates (.02) and lower moving rates (-.15)
  - rapid spatial decay is consistent with smaller estimates in other contexts



Zoom IN: Local (block) spillovers are identified within neighborhoods

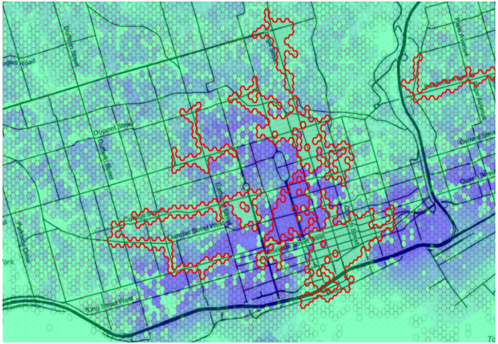
Spoiler Alert! - local (block) spillovers are heterogeneous across...

- *industries*: e.g. emp. effects in manufacturing (.8) / revenue in finance (.05)
- *type of exposure*: e.g. emp effects with occ similar (.14) / revenue with upstream (.10)



Zoom OUT: local spillovers corrected by sorting across neighborhoods

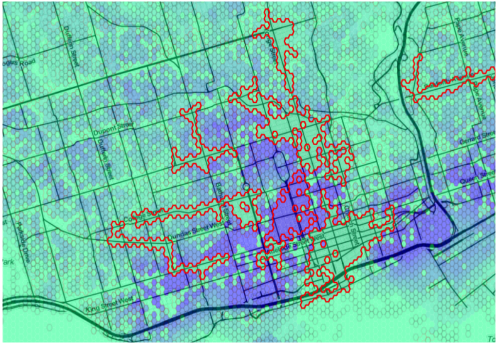
- develop an extended Roy model of entrepreneurs neighborhood selection
  - correction a la Heckman: Dahl (2002), Mazza and van Ophem (2018), and Ransom (2021)



Zoom OUT: local spillovers corrected by sorting across neighborhoods

- <u>exclusion restriction</u>: entrepreneur's personal location preferences

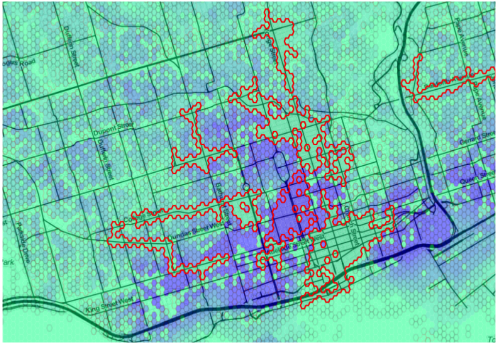
  Michelacci and Silva (2007), Dahl and Sorenson (2012), and Rosenthal and Strange (2011)



Zoom OUT: local spillovers corrected by sorting across neighborhoods

Spoiler Alert! - Sorting into neighborhoods matters

- chosen neighborhood in 1st quintile, residence neighborhood in 5th quintile
- block spillovers are bias by ~20-30% => emp downward to .15-.18 but rev upward to .016-.019



Contribution

- **Entrepreneurship**: location attributes and startup outcomes
  Figueiredo, Guimarães, and Woodward (2002) and Delgado, Porter, and Stern (2010);Dahl and Sorenson (2012);
  Chatterji, Glaeser, and Kerr (2014);Guzman (2019)

- **Urban**: fine geographies, spillovers vs. location sorting
  Arzaghi and Henderson (2008), Greenstone, Hornbeck, and Moretti (2010), and Rosenthal and Strange (2011)
  Glaeser, Kerr, and Ponzetto (2010), Baum-Snow, Gendron-Carrier, and Pavan (2021), and Heblich et al. (2019)
  Duque, Ramos, and Suriñach (2007);Arribas-Bel, Garcia-López, and Viladecans-Marsal (2019) and Galdo, Li, and Rama (2019)

- **Methods**: ML in economics: clustering and neural nets
  Athey and Imbens (2019) and Mullainathan and Spiess (2017);
  Stambaugh, Yu, and Yuan (2015) and David, Saynisch, and Smith-McLallen (2018)

Agenda

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Outline

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Main data is EE with labor tracking, ownership structure and location

Statistics Canada - Canadian Employer Employee Dynamics Dataset (CEEDD)

- Universe of incorporated firms and workers for 4 major cities in 2001-2017
- Detailed location
- Firm labour tracking
- Ownership structure and owners characteristics

Every new firm is a startup?

- Measuring entrepreneurship is difficult (Decker et al. 2014)
- Distinction between institutional entrants and startups
- Startups are
  1. Active
     - With revenues, employment, and/or costs
  2. Really new
     - Not result of restructuring, new subsidiary or M&A
     - Controlled by few individual(s)
  3. With industry and location information
     - Single Location
     - No NAICS 1,2, 9

Startups are smaller and founded by more young females/immigrants

- Smaller than incumbents and other type of new firms

End of first year	**Incumbent**	**Startup**	**Inst. Entrant**
Revenue (2000 mm CAD)	2.245	0.158	1.704
Payroll (2000 mm CAD)	0.648	0.055	0.287
Employment	6.73	0.75	3.39
Alive within 5 years	0.834	0.415	0.333

- Relatively founded by more young female or immigrant entrepreneurs

	Incumbent	**Startup Entrant**
Female Controller	0.195	0.262
Immigrant Controller	0.284	0.467
Age of Controller	52.10	42.35
Industry Experience	2.25	1.33
Number of businesses	1.72	1.92

Startups open in a residence or where incumbents are

- They open at residence of one owner (home-bias), or where incumbents in their industry are located

Block characteristics	**Incumbents**	**Startups**	**Inst. Entrant**
At home*	44%	63%	82%
Firms in Same Industry	6.83	5.25	10.22
Total Employment Same Industry	88.19	64.59	189.54
Total Revenue Same Industry (mm)	43.7	23.4	115
Average Employment Same Industry	13.3	6.5	23.1
Average Revenue Same Industry (mm)	2.8	1.1	5.5
Alive within 5 years	86%	42%	35%
Move within 5 years	55%	69%	66%
Move and alive within 5 years	53%	30%	31%

▸ Locations Explained

Full Sample of Startups

New firms are spread across many industries

	Calgary	Montreal	Toronto	Vancouver	# Startups	# Incumbents-Year	#Incumbents
3	943	2,724	4,885	1,834	10,386	377,085	58749
41	999	2,882	6,377	2,815	13,073	518,074	74979
44	3,043	8,303	15,200	5,148	31,694	657,431	110609
48	4,262	6,275	22,646	2,980	36,163	517,711	86829
51	627	1,547	4,244	2,124	8,542	214,552	36934
52	3,092	5,925	10,066	4,494	23,577	786,241	110614
53	3,313	7,082	10,480	6,010	26,885	858,064	124684
54	19,479	14,090	46,022	14,418	94,009	1,725,843	265428
55	1,122	1,848	4,345	1,708	9,023	439,544	55964
56	2,968	5,590	6,901	2,734	18,193	453,339	68837
71	557	1,114	2,196	1,061	4,928	142,489	22137
72	2,169	6,141	10,044	4,021	22,375	391,635	72688
8	2,870	4,147	8,826	2,915	18,758	791,668	133474
# Startups	45,444	67,668	152,232	52,262	317,606		
# Incumbents-Year	920,051	1,916,827	3,621,351	1,415,447		7,873,676	
# Incumbents	150,779	285,337	563,512	222,298			1,221,926

Outline

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Identification Problem: Location Sorting

Consider the following outcome equation for **startup** i in industry$j$ **starting operations** in location l at time t

$$y_{lt}^{(i,j)} = x_t^{(i,j)}\beta + X_{lt}^{(j)}\theta + \lambda_t^{(j)} + \epsilon_{lt}^{(i,j)} \tag{1}$$

s$x_t^{(i,j)}$ firm characteristics,$X_{lt}^{(j)}$ block industrial characteristics,
$\lambda_t^{(j)}$ time industry fixed-effects,$\epsilon_{lt}^{(i,j)}$ error term

The main identification problem is location sorting

Identification Problem: Location Sorting

Consider the following outcome equation for **startup** *i* in industry*j* **starting operations** in location *l* at time *t*

$$y_{lt}^{(i,j)} = x_t^{(i,j)}\beta + X_{lt}^{(j)}\theta + \lambda_t^{(j)} + \epsilon_{lt}^{(i,j)} \quad (1)$$

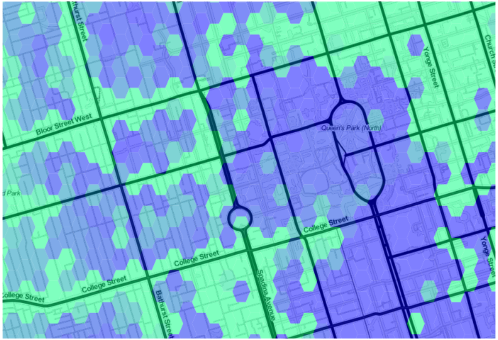
s$x_t^{(i,j)}$ firm characteristics,$X_{lt}^{(j)}$ block industrial characteristics,
$\lambda_t^{(j)}$ time industry fixed-effects,$\epsilon_{lt}^{(i,j)}$ error term

The main identification problem is location sorting

1. **ZOOM IN:** no sorting into blocks within neighborhoods
   - exclusion restriction: no sorting within neighborhoods
     Bayer, Ross, and Topa (2008), Liu, Rosenthal, and Strange (2018), and Roche (2020)
2. **Zoom OUT:** sorting across neighborhoods

Economic Neighborhoods

Propensity score captures the likelihood that a block is suitable for a new startup (as in Heblich et al. (2019) and Qian and Tan (2021))



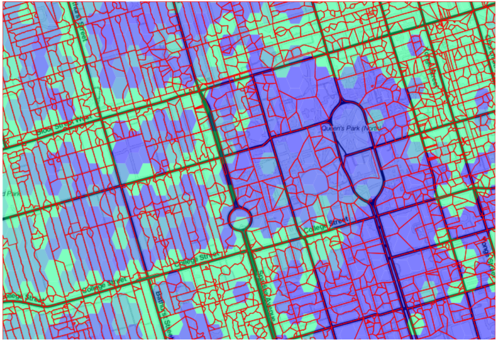
Distribution of Propensity Score (Pooled)

Downtown Toronto

▸ Propensity Score Specification based on ex-ante (2002-2006) startup block choices and block characteristics

Economic Neighborhoods

Propensity score captures the likelihood that a block is suitable for a new startup (as in Heblich et al. (2019) and Qian and Tan (2021))



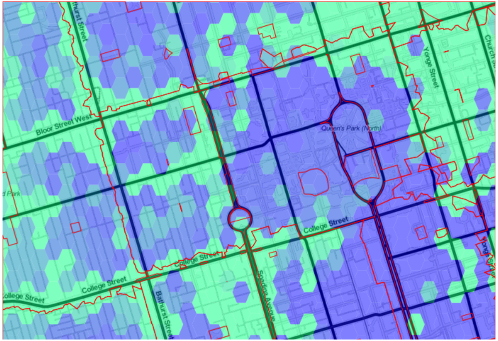
Distribution of Propensity Score (Pooled)

Downtown Toronto (Overlay Local Delivery Units)

▸ Propensity Score Specification based on ex-ante (2002-2006) startup block choices and block characteristics

Economic Neighborhoods

Propensity score captures the likelihood that a block is suitable for a new startup (as in Heblich et al. (2019) and Qian and Tan (2021))



Distribution of Propensity Score (Pooled)

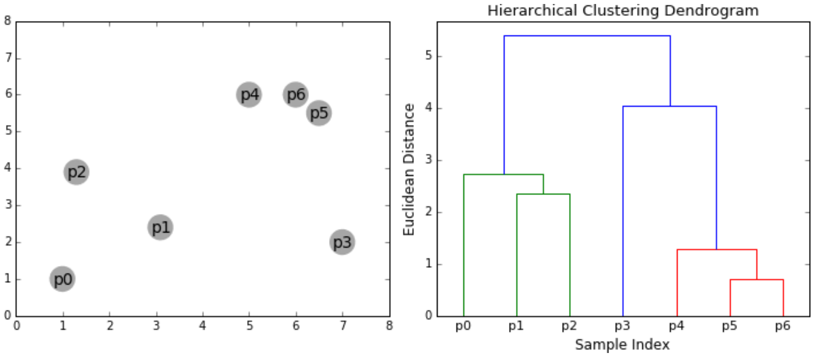
Downtown Toronto (Overlay Forward Sorting Areas)

▸ Propensity Score Specification based on ex-ante (2002-2006) startup block choices and block characteristics

Economic Neighborhoods

Use ML to groups blocks into economic neighborhoods at the industry level

- Similar in spirit to Rosenbaum and Rubin (1984)'s propensity score stratification
- Economic neighborhoods as spatial propensity score stratum Campusano (2021) ▸ Algorithm Details

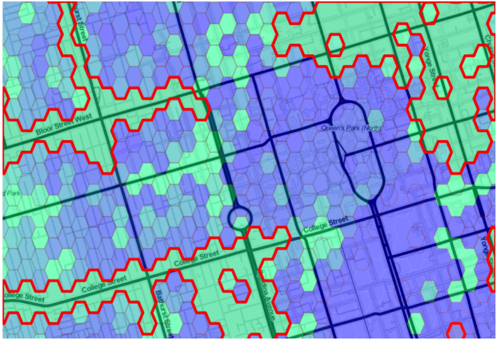


▸ Propensity Score Specification based on ex-ante (2002-2006) startup block choices and block characteristics

Economic Neighborhoods

Use ML to groups blocks into economic neighborhoods at the industry level

- Similar in spirit to Rosenbaum and Rubin (1984)'s propensity score stratification
- Economic neighborhoods as spatial propensity score stratum (Campusano 2021)



Some Neighborhoods (Pooled)

▸ By Industry ▸ Identification Test and Threshold

▸ Propensity Score Specification based on ex-ante (2002-2006) startup block choices and block characteristics

Outline

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Zoom IN: Local Environment and Startups

Providing random allocation within neighborhoods, adding neighborhood-year fixed effects allow us to identify the causal effects of block level environment θ within neighborhoods

$$y_{bnt}^{(i,j)} = x_t^{(i,j)}\beta + X_{bnt}^{(j)}\theta + \lambda_t^{(j)} + \delta_{nt}^{(j)} + \epsilon_{bnt}^{(i,j)} \tag{2}$$

- $y_{bnt}^{(i,j)}$: revenue, employment, exit, move
- $X_{bnt}^{(j)}$: average revenue, average employment
- $x_t^{(i,j)}$: number of owners, corporate partner, years of experience, number of businesses
- controls for number of firms in the block
- Estimation: HDFE and many zeroes => PPML (Correia, Guimarães, and Zylkin 2019)

Estimating Sample

Sample: **sample only considers startups subject to random block**

- Opening away from residence
- In neighborhoods where startups are in different blocks

Startups in Different Blocks

	Calgary	Toronto	Vancouver	Total
3	74	741	223	1,434
8	371	1,163	310	2,643
41	114	873	469	1,876
44	589	3,259	1,292	7,171
48	556	3,387	451	5,404
51	65	574	205	1,001
52	506	1,666	840	3,853
53	536	1,823	1,129	4,913
54	2,439	7,775	2,753	15,209
55	146	772	343	1,510
56	313	993	312	2,577
71	81	394	77	667
72	547	2,465	944	5,447
Total	6,337	25,885	9,348	53,705

▸ All Neighborhoods

Zoom IN: Local spillovers are positive

- Positive effects of average employment in a block
- Non conclusive effects of average revenue in a block

(incumbents same industry)	Revenue	Employment	Alive First Year	Alive 2017	Move Out First Year	Total Moves
Log (Average Employment)	**0.120****	**0.218*****	0.000245	0.0177**	-0.159***	-0.0304***
	(0.0523)	**(0.0442)**	(0.00224)	(0.00625)	(0.0199)	(0.00545)
Log (Average Revenue)	**0.0273****	**-0.0257***	0.000503	0.0000189	0.000953	0.00348**
	(0.0129)	**(0.0139)**	(0.000632)	(0.00191)	(0.00479)	(0.00149)
Number of Startups	30,318	20,675	29,621	28,989	19,714	28,528

Controls: Ownership structure, number of incumbents, dummy for zero activity.
FE: neighborhood year, industry-year, city-year
Std. Errors clustered at Neighborhood-Year level

▸ Full Table

Zoom IN: Local spillovers are positive, but very local

These effects are very local

- decay quickly one block away

	Same Block	1st Ring 150m	2nd Ring 225m	3rd Ring 300m
End of Year Revenue				
Log (Average Employment Same Industry)	0.118**	0.0621*	-0.0230	0.0154
	(0.0489)	(0.0350)	(0.0508)	(0.0280)
Log (Average Revenue Same Industry)	0.0259**	0.00739	-0.0168*	-0.00138
	(0.0123)	(0.0133)	(0.0101)	(0.00802)

Each panel is one regression. Coefficients correspond to measure of variable in the first column

▸ Decay Explained

Zoom IN: heterogenous across industries

- Employment intensive and industries benefit more from higher employment
- Knowledge intensive industries benefit more from higher average revenue

End of Year Revenue	Information / Financial Services	Manufacturing	Professional / Business Services	Retail, Leisure and Hospitality	Transport / Wholesale Trade
Log (Average Employment Same Industry)	-0.227**	0.812***	0.0739	0.209***	0.340**
	(0.106)	(0.189)	(0.114)	(0.0459)	(0.126)
Log (Average Revenue Same Industry)	0.0530**	-0.0802*	0.0293*	0.00145	0.0143
	(0.0187)	(0.0442)	(0.0159)	(0.0129)	(0.0185)

Each panel is one regression. Coefficients correspond to a dummy for a group of industries interacted with the variable in the first column.

▸ Employment

Zoom IN: and type of industry exposure

Exposure not only to same industry

- Use of input-output weights (StatsCan 2001)
- Use of occupational similarity weights (BLS 2001)

	Same	All	Downstream	Upstream	Occ. Similarity
End of Year Revenue					
Log (Average Employment)	0.120**	0.186***	0.0437**	0.0407**	0.141***
	(0.0523)	(0.0296)	(0.0147)	(0.0149)	(0.0261)
Log (Average Revenue)	0.0273**	0.0319**	0.0907***	0.100***	0.0637***
	(0.0129)	(0.0145)	(0.0130)	(0.0135)	(0.0131)

▸ Employment

Zoom IN: In Summary

In summary

- **Spillovers** Mostly positive effects
- **Learning** Moving less in ST and surviving more in MT
- **Heterogeneity** Across industries and type of exposure
- **Decay** rapidly in the space => benefits of hyper concentration

Outline

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Zoom OUT: Sorting across Neighborhoods

- Neighborhood-Year fixed effects
  - control for unobserved differences across blocks within boundaries
  - without considering unobserved heterogeneity across neighborhoods
- Roy model: to correct estimates block-level spillovers for self-selection into neighborhoods

Zoom OUT: Sorting across Neighborhoods

An entrepreneur expected utility V_n^i for neighborhood n is a function of two components

$$V_n^i = \overline{y}_n^i + t_n^i \tag{3}$$

where t_n^i is a vector of personal preferences and $\overline{y}_n^i$ is the expected outcome of the startup,

$$\overline{y}_n^i \equiv \mathrm{E}_n^i \left[y_{bn}^i \mid x^i \right] \quad = \quad x^i \beta + \overline{X}_n^i \theta + \mathrm{E}_n^i \left[\nu_{bn}^i \mid x^i \right] \tag{4}$$

in which y_{bn}^i correspond to outcome of startup i in block b of neighborhood n

$$y_{bn}^i = x^i \beta + X_{bn} \theta + \nu_{bn}^i \tag{5}$$

Zoom OUT: Sorting across Neighborhoods

Expected utility V_n^i can be written in terms of population averages and an error component

$$V_n^i = V_n + e_n^i \tag{6}$$

where $V_n^i = \mathrm{E}\left[y_{bn}^i \mid x^i\right] + \mathrm{E}\left[t_n^i \mid z^i, d_n^i\right]$, and $e_n^i = \varepsilon_n^i + \mu_n^i$ in which

$$\begin{aligned} \overline{y}_n^i - \mathrm{E}\left[y_{bn}^i \mid x^i\right] &= \varepsilon_n^i \\ t_n^i - \mathrm{E}\left[t_n^i \mid z^i, d_n^i\right] &= \mu_n^i \end{aligned}$$

leading to the following neighborhood selection rule

$$y_{bn}^i \qquad \text{is observed iff} \quad \max_{k \neq n} \left(V_n^i - V_k^i + e_n^i - e_k^i\right) \leq 0 \tag{7}$$

(average deviations ε_n^i from expected outcomes analogous to neighborhood-year fixed effects)

Zoom OUT: Sorting across Neighborhoods

- Traditional selection correction methods require imposing severe restrictions in the sorting process (Lee and Salanié 2018)
- And become quickly untractable as the choice set increases (Vella 1998)
- Using (7) we can correct for sorting across neighborhoods by including all sub-utilities

$$y^i_{bn} = x^i\beta + X_{bn}\theta + M^i_n \times \Omega_n\left(V^i_1 - V^i_n, V^i_2 - V^i_n \ldots, V^i_N - V^i_n\right) + \mu^i_{bn} \quad (8)$$

$\Omega_n(\bullet) = \mathrm{E}_n\left[e^i_n | V_1 - V_n, V_2 - V_n \ldots, V_N - V_n\right]$

μ^i_n error term with mean zero in the conditional sample

M^i_n is a dummy variable equal one if i chooses neighborhood n

Zoom OUT: Sorting across Neighborhoods

- As in Dahl (2002), Mazza and van Ophem (2018), and Ransom (2021), I use an index sufficiency assumption

$$
\begin{aligned}
& g_n\left(\varepsilon_n^i, \max_{k\in N}\left(V_n^i - V_k^i + e_n^i - e_k^i\right) \mid V_1 - V_n, V_2 - V_n \dots, V_N - V_n\right) \\
&\qquad = g_n\left(\varepsilon_n^i, \max_{k\in N}\left(V_n^i - V_k^i + e_n^i - e_k^i\right) \mid \mathrm{P}_n^i\right) \qquad (9)
\end{aligned}
$$

- That leads to the following estimating equation

$$
y_{bnt}^{(i,j)} = x^i\beta + X_{bnt}^{(j)}\theta + \delta_{nt}^{(j)} \times \lambda_{nt}\begin{pmatrix} {}^{star}\mathrm{P}_{nt}^{(i,j)}, & {}^{home}\mathrm{P}_{nt}^{(i,j)} \end{pmatrix} + \omega_{bn}^{(i,j)} \qquad (10)
$$

Zoom OUT: Sorting across Neighborhoods

Two step procedure

1. Estimate location choice probabilities for all possible neighborhoods (and home)
   - Choice cells are too many => not feasible with current methods
   - Conditional choice model estimated using the Adam method for stochastic optimization (Kingma and Ba 2017)
   - Exclusion restriction: **owners have preference for location that do not affect outcomes**
2. Use probabilities to correct for selection into neighborhoods
   - Now we can include all startups and variation across neighborhoods

Zoom OUT: Step 1 - Neighborhood Choice Probabilities

1. Chosen neighborhoods are in first quintile of options
2. Home neighborhoods rank lower, even when they are chosen

Table: Descriptive Statistics Neighborhood Choice Probabilities

Variable	**Mean**	**Std.Dev.**	**Min**	**Max**
Number of Neighborhoods	189.6	178.1	1.0	573
Rank of Chosen Neighborhood	31.29	67.48	1.0	573
Probability of Chosen Neighborhood	0.245	0.345	0.0	1.0
Distance of Chosen Neighborhood	168.8	686.5	0.0	4570
First Best Probability	0.575	0.286	0.04	1.0
Probability of Home	0.691	0.461	0.0	1.0
Rank of Home	100.3	91.34	1.0	287

Zoom OUT: Step 2 - Corrected Estimates: Revenue

- Estimates of average employment are downward biased ~25-50%
- While estimates of average revenue are upper biased ~30-40%

(End of Year Revenue)	Uncorrected		Distance to First Residence		Distance to Prev. Residence	
	Benchmark Sample	Full Sample	Using $P^{star(i,j)}_{nt}$	Using $P^{home(i,j)}_{nt}$	Using $P^{star(i,j)}_{nt}$	Using $P^{home(i,j)}_{nt}$
Log (Average Employment Same Industry)	0.120**	0.169***	0.152***	0.176***	0.184***	0.172***
	(0.0523)	(0.0197)	(0.0323)	(0.0311)	(0.0389)	(0.0362)
Log (Average Revenue Same Industry)	0.0273**	0.0381***	0.0299**	0.0183**	0.0165**	0.0198**
	(0.0129)	(0.0100)	(0.0109)	(0.00699)	(0.00767)	(0.00909)
Number of Startups	30,318	247,594	125,676	125,676	125,676	125,676

FE: $\delta^{(j)}_{nt} \times P^{(i,j)}_{nt}$, Industry-Year FE and City-Year FE.
Same controls as benchmark regression.

Employment | Benchmark Sample

Zoom OUT: In summary

- Large percentage of startups are founded in place of residence
- Suggest sorting might be important
  - Unless important benefits of residence
  - Evidence suggest the opposite
- Sorting into lower ranked places (including home) bias estimates
  - Benefits of concentration and hyper concentration

Outline

Introduction

Data

Identification

Zoom IN: Local (Block) Spillovers

Zoom OUT: Local Spillovers and Neighborhood Sorting

Conclusions

Conclusions

- Novel way to identify spillovers while accounting location sorting
  - Within: Propensity Score + Ward (1963) = Economic neighborhoods
  - Across: Roy Model + Dahl (2002) + Neural networks = semi-parametric correction
- Spillovers are generally positive but VERY local
  - Positive effects of average employment and revenue
  - Rapid decay in the space and vary across industries and type of exposure
- Location sorting matters
  - Chosen neighborhoods are not the first best, but belong to first quantile
  - 60% chooses home but the ranking of that choice is, generally, below the median
  - Sorting downward bias estimates of employment

Startup Location, Local Spillovers and Neighborhood Sorting

Rolando Campusano

University of Toronto (Rotman)

October 4, 2021

Appendix

References I

- Arribas-Bel, Daniel, M. À. Garcia-López, and Elisabet Viladecans-Marsal (Nov. 30, 2019). "Building(s and) Cities: Delineating Urban Areas with a Machine Learning Algorithm." In: *Journal of Urban Economics*, p. 103217. ISSN: 0094-1190.
- Arzaghi, Mohammad and J. Vernon Henderson (Oct. 1, 2008). "Networking off Madison Avenue." In: *The Review of Economic Studies* 75.4, pp. 1011–1038. ISSN: 0034-6527.
- Athey, Susan and Guido Imbens (Mar. 24, 2019). *Machine Learning Methods Economists Should Know About*. arXiv: `1903.10075 [econ, stat]`.
- Audretsch, David B., Erik E. Lehmann, and Susanne Warning (Sept. 2005). "University Spillovers and New Firm Location." In: *Research Policy* 34.7, pp. 1113–1122. ISSN: 00487333.
- Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan (2021). "Local Productivity Spillovers." In: *mimeo*, p. 25.
- Bayer, Patrick, Stephen L Ross, and Giorgio Topa (2008). "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." In: *Journal of Political Economy* 116.6, pp. 1150–1196.

References II

Campusano, Rolando (2021). "Delineating Neighborhoods Using Location Choices."

Catalini, Christian (Sept. 2018). "Microgeography and the Direction of Inventive Activity." In: *Management Science* 64.9, pp. 4348–4364. ISSN: 0025-1909, 1526-5501.

Chatterji, Aaron, Edward L Glaeser, and William R Kerr (2014). "Clusters of Entrepreneurship and Innovation." In: *Innovation Policy and the Economy*. Josh Lerner and Scott Stern. Vol. 14. University of Chicago Press.

Chatterji, Aaron K. (2009). "Spawned with a Silver Spoon? Entrepreneurial Performance and Innovation in the Medical Device Industry." In: *Strategic Management Journal* 30.2, pp. 185–206. ISSN: 1097-0266.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin (Aug. 2, 2019). *Ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects*. arXiv: `1903.01690 [econ]`.

Dahl, Gordon B. (2002). "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." In: *Econometrica* 70.6, pp. 2367–2420. ISSN: 0012-9682.

References III

- Dahl, Michael S. and Olav Sorenson (June 2012). "Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures." In: *Management Science* 58.6, pp. 1059–1071. ISSN: 0025-1909, 1526-5501.
- David, Guy, Philip A. Saynisch, and Aaron Smith-McLallen (May 1, 2018). "The Economics of Patient-Centered Care." In: *Journal of Health Economics* 59, pp. 60–77. ISSN: 0167-6296.
- Decker, Ryan et al. (Sept. 2014). "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." In: *Journal of Economic Perspectives* 28.3, pp. 3–24. ISSN: 0895-3309.
- Delgado, Mercedes, Michael E. Porter, and Scott Stern (Jan. 7, 2010). "Clusters and Entrepreneurship." In: *Journal of Economic Geography* 10.4, pp. 495–518. ISSN: 1468-2702, 1468-2710.
- Duque, Juan Carlos, Raúl Ramos, and Jordi Suriñach (July 1, 2007). "Supervised Regionalization Methods: A Survey." In: *International Regional Science Review* 30.3, pp. 195–220. ISSN: 0160-0176.

References IV

- Duranton, Gilles and Henry G. Overman (Jan. 10, 2005). "Testing for Localization Using Micro-Geographic Data." In: *The Review of Economic Studies* 72.4, pp. 1077–1106. ISSN: 0034-6527, 1467-937X.
- Ellison, Glenn and Edward L Glaeser (1999). "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" In: *American Economic Review* 89.2, pp. 311–316.
- Figueiredo, Octávio, Paulo Guimarães, and Douglas Woodward (Sept. 2002). "Home-Field Advantage: Location Decisions of Portuguese Entrepreneurs." In: *Journal of Urban Economics* 52.2, pp. 341–361. ISSN: 00941190.
- Galdo, Virgilio, Yue Li, and Martin Rama (Dec. 24, 2019). "Identifying Urban Areas by Combining Human Judgment and Machine Learning: An Application to India." In: *Journal of Urban Economics*, p. 103229. ISSN: 0094-1190.
- Glaeser, Edward L., Sari Pekkala Kerr, and William R. Kerr (May 2015). "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines." In: *Review of Economics and Statistics* 97.2, pp. 498–520. ISSN: 0034-6535, 1530-9142.

References V

- Glaeser, Edward L., William R. Kerr, and Giacomo A.M. Ponzetto (Jan. 2010). "Clusters of Entrepreneurship." In: *Journal of Urban Economics* 67.1, pp. 150–168. ISSN: 00941190.
- Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." In: *Journal of Political Economy* 118.3, pp. 536–598.
- Guzman, Jorge (Mar. 1, 2019). *Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley*. SSRN Scholarly Paper ID 3175328. Rochester, NY: Social Science Research Network.
- Heblich, Stephan et al. (May 21, 2019). *Industrial Clusters in the Long Run: Evidence from Million-Rouble Plants in China*. 19/712. School of Economics, University of Bristol, UK.
- Henderson, Vernon, Ari Kuncoro, and Matt Turner (Oct. 1, 1995). "Industrial Development in Cities." In: *Journal of Political Economy* 103.5, pp. 1067–1090. ISSN: 0022-3808.

References VI

- Kerr, William R. and Scott Duke Kominers (2015). "Agglomerative Forces and Cluster Shapes." In: *Review of Economics and Statistics* 97.4, pp. 877–899.
- Kingma, Diederik P. and Jimmy Ba (Jan. 29, 2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs].
- Lanahan, Lauren and Maryann P. Feldman (Sept. 2015). "Multilevel Innovation Policy Mix: A Closer Look at State Policies That Augment the Federal SBIR Program." In: *Research Policy* 44.7, pp. 1387–1402. ISSN: 00487333.
- Lee, Sokbae and Bernard Salanié (2018). "Identifying Effects of Multivalued Treatments." In: *Econometrica* 86.6, pp. 1939–1963. ISSN: 1468-0262.
- Liu, Crocker H., Stuart S. Rosenthal, and William C. Strange (2017). *Building Specialization, Anchor Tenants and Agglomeration Economies*.
- — (July 2018). "The Vertical City: Rent Gradients, Spatial Structure, and Agglomeration Economies." In: *Journal of Urban Economics* 106, pp. 101–122. ISSN: 00941190.

References VII

- Mazza, Jacopo and Hans van Ophem (2018). "Separating Risk from Heterogeneity in Education: A Semiparametric Approach." In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181.1, pp. 249–275. ISSN: 1467-985X.
- Michelacci, Claudio and Olmo Silva (Nov. 2007). "Why So Many Local Entrepreneurs?" In: *Review of Economics and Statistics* 89.4, pp. 615–633. ISSN: 0034-6535, 1530-9142.
- Mullainathan, Sendhil and Jann Spiess (May 2017). "Machine Learning: An Applied Econometric Approach." In: *Journal of Economic Perspectives* 31.2, pp. 87–106. ISSN: 0895-3309.
- Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1989). "Industrialization and the Big Push." In: *Journal of Political Economy* 97.5, pp. 1003–1026. ISSN: 0022-3808.
- Porter, Michael E (1998). *Clusters and the New Economics of Competition*. Vol. 76. 6. Harvard Business Review Boston.
- Qian, Franklin and Rose Tan (2021). "The Effects of High-Skilled Firm Entry on Incumbent Residents."

References VIII

- Ransom (2021). "Selective Migration, Occupational Choice, and the Wage Returns to College Majors." In: *Annals of Economics and Statistics* 142, p. 45. ISSN: 21154430.
- Roche, Maria P. (Dec. 1, 2020). "Taking Innovation to the Streets: Microgeography, Physical Structure, and Innovation." In: *The Review of Economics and Statistics* 102.5, pp. 912–928. ISSN: 0034-6535.
- Rosenbaum, Paul R. and Donald B. Rubin (1984). "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." In: *Journal of the American Statistical Association* 79.387, pp. 516–524. ISSN: 0162-1459.
- Rosenthal, Stuart S. and William C. Strange (May 2003). "Geography, Industrial Organization, and Agglomeration." In: *Review of Economics and Statistics* 85.2, pp. 377–393. ISSN: 0034-6535, 1530-9142.
- — (Apr. 5, 2011). "Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch." In: *The Review of Economics and Statistics* 94.3, pp. 764–788. ISSN: 0034-6535.

References IX

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan (2015). "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle." In: *The Journal of Finance* 70.5, pp. 1903–1948. ISSN: 1540-6261.

Vella, Francis (1998). "Estimating Models with Sample Selection Bias: A Survey." In: *The Journal of Human Resources* 33.1, p. 127. ISSN: 0022166X.

Ward, Joe H. (Mar. 1, 1963). "Hierarchical Grouping to Optimize an Objective Function." In: *Journal of the American Statistical Association* 58.301, pp. 236–244. ISSN: 0162-1459.

*Estimating Sample I

Startups

	Calgary	Toronto	Vancouver	Total
3	358	2,218	907	4,733
8	1,153	4,250	1,564	9,095
41	371	2,611	1,318	5,465
44	1,401	7,240	2,701	15,898
48	1,006	5,832	967	9,343
51	175	1,347	908	2,934
52	1,124	3,877	2,501	9,498
53	1,194	4,219	3,076	11,382
54	4,501	13,142	5,812	27,770
55	442	1,873	985	4,057
56	755	2,284	1,089	5,997
71	180	819	500	1,943
72	1,229	5,440	2,413	12,897
Total	13,889	55,152	24,741	121,012

*Estimating Sample II

Back to Estimating Sample

*Hexagon Grid Assignment

- The city is divided in a 75-meter hexagon grid cell
- Each firm is assigned to this grid cell

Figure: Hexagon Grid Cell and Rings Around

Back to Data: Where Back to Decay Results

*Propensity Score Specification

- DMTI Data on Points of Interest
- The propensity score captures the probability that a given incumbent firm chooses block cell b based on its characteristics

$$\#\text{Startups}_{bjt} = \beta + \sum_{POI} \beta_{POI}\text{POI}_{bt} + \sum_{POI} \beta_{\text{MA}POI}\text{MA\_POI}_{bt} + \sum_{POI} \beta_{LAND}\text{LAND}_{bt} + \sum_{LAND} \beta_{\text{MA}LAND}\text{MA\_LAND}_{bt} + \sum_{POI} \beta_X X_{bjt} + \sum_{POI} \beta_{\text{MA\_}X}\text{MA\_}X_{bjt} + \epsilon_i$$

  - *POI* are points of interest, *LAND* are land uses
  - X_{bjt}: attributes and composition of nearby firms and workers with
    - *Up*, *Down*, *Eq*, *Oc* based on Input-Output (StatCan) and Occupational Similarity (BLS 2001) weights
  - $MA(*)$ are measures within 1 km with distance exp decay (rho=1)
- Sample: 2002-2006

back

*Adjacent Hierarchical Clustering Algorithm

The loss of information when grouping blocks into a neighborhood $N \subset \mathcal{P}$

$$I(N) = \sum_{\mathrm{P}_t^u} \parallel \mathrm{P}_t^u - \overline{\mathrm{P}}_N \parallel^2$$

where $\overline{\mathrm{P}}_N = n^{-1} \sum_{u=1}^{n} \mathrm{P}_t^u$ is the *centre of gravity* of N and n is the number of blocks in the neighborhood.

Starting from a partition $\{N_1, ..., N_l\}$ of $\mathcal{P}$, the loss of information when merging two neighborhoods N_u and N_v is quantified by:

$$\delta\left(N_u, N_v\right) = I(N_u \cup N_v) - I(N_u) - I(N_v)$$

That, when minimized, it is equal to minimizing the variation of *within-cluster sum of squares* after merging two clusters(Ward 1963)

back

*Adjacent Hierarchical Clustering Algorithm

For a given $\mathcal{P} = \{\mathrm{P}_t^u\}_{u=1}^{\mathbf{B}}$ set of all block-level probabilities to be clustered.

1. Initialize with set of neighborhoods to be $\{N_1, ..., N_B\}$ where $N_u = \{\mathrm{P}_t^u\}$ for all $u = 1, ..., B$

◂ back

*Adjacent Hierarchical Clustering Algorithm

For a given $\mathcal{P} = \{\mathrm{P}_t^u\}_{u=1}^{\mathbf{B}}$ set of all block-level probabilities to be clustered.

1. Initialize with set of neighborhoods to be $\{N_1, ..., N_B\}$ where $N_u = \{\mathrm{P}_t^u\}$ for all $u = 1, ..., B$
2. Compute the dissimilarity between all pairs, that is, compute $\delta\left(N_u, N_v\right)$ for all $u < v \in \{adjacent_u\}$

◂ back

*Adjacent Hierarchical Clustering Algorithm

For a given $\mathcal{P} = \{\mathrm{P}_t^u\}_{u=1}^{\mathbf{B}}$ set of all block-level probabilities to be clustered.

1. Initialize with set of neighborhoods to be $\{N_1, \ldots, N_B\}$ where $N_u = \{\mathrm{P}_t^u\}$ for all $u = 1, \ldots, B$
2. Compute the dissimilarity between all pairs, that is, compute $\delta\left(N_u, N_v\right)$ for all $u < v \in \{adjacent_u\}$
3. While there is more than one neighborhood in the original set:
   3.1 Merge a pair which have minimal dissimilarity

   $$\delta\left(N_{u'}, N_{v'}\right) = \min_{u'<v'} \delta\left(N_u, N_v\right)$$

   set $N_{u'} = N_{u'} \cup N_{v'}$ and remove $N_{v'}$ from the set of neighborhoods

   3.2 Compute dissimilarity between $N_{u'}$ and the remaining neighborhoods in original set

back

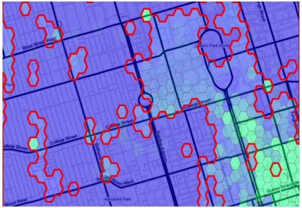
*Adjacent Hierarchical Clustering Algorithm

For a given $\mathcal{P} = \{\mathrm{P}_t^u\}_{u=1}^{\mathbf{B}}$ set of all block-level probabilities to be clustered.

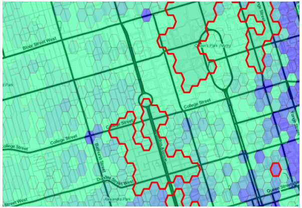
1. Initialize with set of neighborhoods to be $\{N_1, \ldots, N_B\}$ where $N_u = \{\mathrm{P}_t^u\}$ for all $u = 1, \ldots, B$
2. Compute the dissimilarity between all pairs, that is, compute $\delta(N_u, N_v)$ for all $u < v \in \{adjacent_u\}$
3. While there is more than one neighborhood in the original set:
   3.1 Merge a pair which have minimal dissimilarity
   $$\delta(N_{u'}, N_{v'}) = \min_{u' < v'} \delta(N_u, N_v)$$
   set $N_{u'} = N_{u'} \cup N_{v'}$ and remove $N_{v'}$ from the set of neighborhoods
   3.2 Compute dissimilarity between $N_{u'}$ and the remaining neighborhoods in original set
4. The final set of neighborhoods $\{N\}$ is defined as the subset of $\mathcal{P}$ in which the dissimilarity within (across) neighborhoods is minimized (maximized)

back

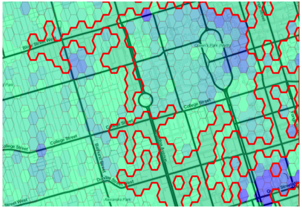
*Neighborhoods by industry



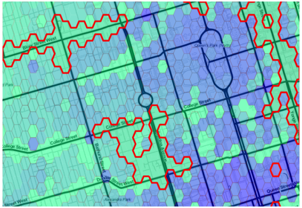
(a) Manufacturing



(b) Transportation



(c) Professional Services



(d) Retail, Accommodation and Food Services

back

*Testing Neighborhoods

- Each cell is coefficient of regression of number of startups on the score
- Conditional on neighborhood fixed effects using 0.01 caliper

	Number of Startups			
Propensity Score	**Toronto**	**Vancouver**	**Montreal**	**Calgary**
3	768.6	-233.6	982.8**	858.8
	(0.55)	(-0.32)	(2.86)	(1.31)
41	-1789.4	-1683.4	411.1	1151.1
	(-1.02)	(-1.60)	(1.01)	(1.74)
44	1358.1	-2529.3	991.0	4255.4***
	(0.63)	(-1.20)	(1.81)	(3.69)
48	7878.5	-1075.6	972.3	951.3
	(1.93)	(-0.97)	(1.46)	(0.38)
51	2503.0	719.2	655.0**	-316.6
	(1.71)	(0.77)	(2.62)	(-0.59)
52	1686.3	998.6	1859.7**	-61.76
	(0.96)	(0.91)	(3.23)	(-0.06)
53	6893.2**	-184.3	2569.1***	-737.4
	(3.27)	(-0.11)	(4.01)	(-0.52)
54	5154.8	-2067.6	4080.9***	-1220.2
	(0.67)	(-0.77)	(3.31)	(-0.19)
55	1867.9	564.6	661.9**	759.6
	(1.83)	(1.08)	(2.72)	(1.28)
56	3753.4*	247.1	1268.9*	1203.6
	(2.33)	(0.27)	(2.00)	(0.75)
71	2580.0*	1083.6*	172.4	-216.7
	(2.37)	(2.32)	(0.83)	(-0.44)
72	2352.0	2783.3*	336.8	761.9
	(1.20)	(2.30)	(0.75)	(0.84)

back

*Zoom IN: Local Environment and Startups: Revenue

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Average Employment Same Industry)	0.120**	0.171***	0.137***	0.157**	0.119**	0.0530	0.143**	0.0921**	0.108**	0.120**
	(0.0523)	(0.0416)	(0.0229)	(0.0569)	(0.0517)	(0.0673)	(0.0515)	(0.0411)	(0.0418)	(0.0526)
Log (Average Revenue Same Industry)	0.0273**	0.0351*	0.0365**	0.0175**	0.0142*	0.0527**	0.0282**	-0.00245	0.0345**	0.0273**
	(0.0129)	(0.0187)	(0.0134)	(0.00817)	(0.00759)	(0.0210)	(0.0128)	(0.00744)	(0.0123)	(0.0127)
Log (# Incumbents Same Industry)	-0.114*	-0.00935	0.00371			-0.121	-0.114**	0.0357	-0.00624	-0.114*
	(0.0599)	(0.0332)	(0.0305)			(0.130)	(0.0572)	(0.0412)	(0.0387)	(0.0602)
No Incumbent Economic Activity	0.108	0.299*	0.431**			0.684**	0.116	-0.0586	0.226**	0.108
	(0.134)	(0.174)	(0.166)			(0.241)	(0.133)	(0.0663)	(0.115)	(0.137)
Previous Ownership Experience	0.0490***	0.0104***	0.0175***		0.0490***	0.0363*	0.0384***	0.102***	0.0481**	0.0490***
	(0.0121)	(0.00286)	(0.00395)		(0.0122)	(0.0192)	(0.00846)	(0.0157)	(0.0157)	(0.0120)
Previous Industry Experience	0.0732***	0.0672***	0.0674***		0.0733***	0.0801***	0.0389***	0.0606***	0.0617***	0.0732***
	(0.00700)	(0.00772)	(0.00502)		(0.00706)	(0.0164)	(0.00848)	(0.00657)	(0.00493)	(0.00688)
Number of Owners	-0.0697	-0.00356	-0.0492		-0.0693	-0.0751	-0.143**	0.0379	-0.0137	-0.0697
	(0.0586)	(0.0629)	(0.0397)		(0.0599)	(0.0952)	(0.0554)	(0.0760)	(0.0417)	(0.0604)
Corporate Partner	0.246	0.248	0.131		0.247	-0.184	0.202*	0.549**	0.806***	0.246*
	(0.152)	(0.173)	(0.277)		(0.154)	(0.520)	(0.116)	(0.220)	(0.153)	(0.146)
Family Control	0.470**	0.482***	0.556***		0.470**	0.213	0.631***	0.547***	0.744***	0.470**
	(0.149)	(0.120)	(0.0974)		(0.149)	(0.254)	(0.146)	(0.0821)	(0.0868)	(0.149)
Number of Startups	30,318	41,112	81,319	30,318	30,318	5,363	37,944	81,586	142,465	30, 318
Sample	Away	Away	Away	Away	Away		Inst. Entrants	Home	Home/Away	Away
Neighborhood Year FE	YES	NO	FSA	YES	YES		YES	YES	YES	NO
Cluster Level	NeighYear	Year	NeighYear	NeighYear	NeighYear		Neigh Year	NeighYear	NeighYear	Neigh

Industry-Year FE and City-Year FE

Back to main results

*Zoom IN: Local Environment and Startups: Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Average Employment Same Industry)	0.218***	0.333***	0.244***	0.252***	0.217***	0.178*	0.220***	0.209***	0.279***	0.218***
	(0.0442)	(0.0443)	(0.0221)	(0.0459)	(0.0460)	(0.0932)	(0.0439)	(0.0383)	(0.0269)	(0.0449)
Log (Average Revenue Same Industry)	-0.0257*	-0.0348**	-0.0238**	-0.0170**	-0.0150**	-0.00887	-0.0259*	-0.0135**	-0.0212***	-0.0257*
	(0.0139)	(0.0131)	(0.00758)	(0.00776)	(0.00751)	(0.0298)	(0.0139)	(0.00613)	(0.00597)	(0.0136)
Log (# Incumbents Same Industry)	-0.0305	-0.117**	-0.00839			-0.0290	-0.0314	-0.00939	-0.0207	-0.0305
	(0.0613)	(0.0437)	(0.0225)			(0.0896)	(0.0605)	(0.0286)	(0.0316)	(0.0622)
No Incumbent Economic Activity	-0.194	-0.233*	-0.145			0.133	-0.192	-0.0584	-0.130**	-0.194
	(0.171)	(0.134)	(0.0918)			(0.398)	(0.172)	(0.0677)	(0.0645)	(0.171)
Previous Ownership Experience	0.00138	0.00141	0.00716*		0.00124	0.0950**	0.00214	0.0233**	0.00337	0.00138
	(0.00551)	(0.00234)	(0.00419)		(0.00531)	(0.0458)	(0.00605)	(0.0103)	(0.00533)	(0.00549)
Previous Industry Experience	0.0836***	0.0692***	0.0746***		0.0836***	0.0590***	0.0775***	0.0780***	0.0782***	0.0836***
	(0.00685)	(0.00704)	(0.00407)		(0.00684)	(0.0154)	(0.00628)	(0.00542)	(0.00405)	(0.00715)
Number of Owners	0.228***	0.177***	0.117***		0.229***	0.105	0.200***	0.0842**	0.174***	0.228***
	(0.0521)	(0.0518)	(0.0295)		(0.0524)	(0.110)	(0.0483)	(0.0358)	(0.0258)	(0.0543)
Corporate Partner	0.632***	0.565***	-0.0247		0.627***	0.657	0.539***	0.931*	0.966***	0.632***
	(0.166)	(0.156)	(0.292)		(0.165)	(0.816)	(0.150)	(0.512)	(0.184)	(0.162)
Family Control	0.114	0.114	-0.0153		0.103	0.390	0.0995	-0.0844	-0.0202	0.114
	(0.181)	(0.153)	(0.100)		(0.181)	(0.465)	(0.150)	(0.113)	(0.0860)	(0.184)
Number of Startups	20,675	41,570	66,678	20675	20675	3174	25088	60,740	109,411	20,675
Sample	Away	Away	Away	Away	Away		Inst. Entrants	Home	Home/Away	Away
Neighborhood Year FE	YES	NO	FSA	YES	YES		YES	YES	YES	NO
Cluster Level	NeighYear	Year	NeighYear	NeighYear	NeighYear		Neigh Year	NeighYear	NeighYear	Neigh

Industry-Year FE and City-Year FE

Back to main results

*Zoom IN: Local Environment and Startups: Future

	(1) Alive First Year	(2) Alive 2017	(3) Yrs Alive / (2017 - Found. Year)	(4) Move Out First Year	(5) Total Moves	(6) Move and Alive
Log (Average Employment Same Industry)	0.000245	0.0177**	0.00971**	-0.159***	-0.0304***	-0.0954***
	(0.00224)	(0.00625)	(0.00385)	(0.0199)	(0.00545)	(0.0127)
Log (Average Revenue Same Industry)	0.000503	0.0000189	0.000681	0.000953	0.00348**	0.000804
	(0.000632)	(0.00191)	(0.00114)	(0.00479)	(0.00149)	(0.00328)
Number of Startups	29,621	28,989	29,621	19,714	28,528	30,318

Neighborhood-Year FE, Industry-Year FE and City-Year FE. Sample: Away. Controls include: number of incumbent firms, dummy of no economic activity, owners previous industry experience, owners previous entrepreneurship experience, total number of owners, dummy of corporate partner, and dummy of family control.

◂ Back to main results

*Zoom IN: Local spillovers are positive, but very local

	Same Block	1st Ring 150m	2nd Ring 225m	3rd Ring 300m
End of Year Employment				
Log (Average Employment Same Industry)	0.212***	0.0280	0.0239	0.0118
	(0.0428)	(0.0332)	(0.0275)	(0.0242)
Log (Average Revenue Same Industry)	-0.0255*	-0.0262***	-0.000222	0.00649
	(0.0139)	(0.00641)	(0.0107)	(0.00765)

Each panel is one regression. Coefficients correspond to measure of variable in the first column

▸ Back

*Zoom IN: Local spillovers are positive, heterogenous across industries

End of Year Employment	Information / Financial Services	Manufacturing	Professional / Business Services	Retail, Leisure and Hospitality	Transport / Wholesale Trade
Log (Average Employment Same Industry)	0.000781	0.722***	0.200	0.202***	0.202*
	(0.234)	(0.175)	(0.131)	(0.0385)	(0.115)
Log (Average Revenue Same Industry)	-0.00499	-0.0731**	-0.0370*	-0.0268*	-0.0102
	(0.0283)	(0.0344)	(0.0215)	(0.0137)	(0.0200)

Each panel is one regression. Coefficients correspond to a dummy for a group of industries interacted with the variable in the first column.

▸ Back

*Zoom IN: Local spillovers are positive, and type of industry exposure

	Same	All	Downstream	Upstream	Occ. Similarity
End of Year Employment					
Log (Average Employment)		0.209***	0.0359**	0.0379**	0.173***
		(0.0331)	(0.0137)	(0.0137)	(0.0280)
Log (Average Revenue)		-0.0101	0.0418**	0.0435**	0.000121
		(0.0173)	(0.0202)	(0.0205)	(0.0199)

Back

*Zoom OUT: Step 2 - Corrected Estimates: Employment

(End of Year Employment)	Uncorrected		Distance to First Residence		Distance to Prev Residence	
	Benchmark Sample	Full Sample	Using $P^{star(i,j)}_{nt}$	Using $P^{home(i,j)}_{nt}$	Using $P^{star(i,j)}_{nt}$	Using $P^{home(i,j)}_{nt}$
Log (Average Employment Same Industry)	0.218***	0.316***	0.319***	0.313***	0.312***	0.294***
	(0.0442)	(0.0167)	(0.0358)	(0.0396)	(0.0428)	(0.0400)
Log (Average Revenue Same Industry)	-0.0257*	-0.0159**	-0.0260**	-0.0223**	-0.0310**	-0.0229**
	(0.0139)	(0.00532)	(0.00919)	(0.0100)	(0.00986)	(0.0111)
Number of Startups	20,675	249,938	88,687	87,760	89,666	86,325

FE: $\delta^{(j)}_{nt} \times P^{(i,j)}_{nt}$, Industry-Year FE and City-Year FE.
Same controls as benchmark regression.

◂ Back to Revenue Results

*Zoom OUT: Step 2 - Corrected Estimates: Restricted Sample

(End of Year Revenue)	Uncorrected		Distance to First Residence		Distance to Prev Residence	
	Benchmark Sample	Full Sample	Using $P_{nt}^{star(i,j)}$	Using $P_{nt}^{home(i,j)}$	Using $P_{nt}^{star(i,j)}$	Using $P_{nt}^{home(i,j)}$
Log (Average Employment Same Industry)	0.163***	0.169***	0.198***	0.163**	0.136**	0.163***
	(0.0441)	(0.0449)	(0.0565)	(0.0629)	(0.0414)	(0.0487)
Log (Average Revenue Same Industry)	0.0288*	0.0249	0.0221	0.0281	0.0313**	0.0356**
	(0.0165)	(0.0157)	(0.0172)	(0.0182)	(0.0142)	(0.0175)
Number of Startups	29,250	29,250	29,250	29,250	29250	29,250

FE: $\delta_{nt}^{(j)} \times P_{nt}^{(i,j)}$, Industry-Year FE and City-Year FE.
Same controls as benchmark regression.

Back to Revenue Results

*Zoom OUT: Step 2 - Corrected Estimates: Restricted Sample

(End of Year Employment)	Uncorrected		Distance to First Residence		Distance to Prev Residence	
	Benchmark Sample	Full Sample	Using $P_{nt}^{star(i,j)}$	Using $P_{nt}^{home(i,j)}$	Using $P_{nt}^{star(i,j)}$	Using $P_{nt}^{home(i,j)}$
Log (Average Employment Same Industry)	0.278***	0.262***	0.239***	0.223***	0.310***	
	(0.0388)	(0.0430)	(0.0430)	(0.0340)	(0.0509)	
Log (Average Revenue Same Industry)	-0.0279**	-0.0155	-0.0319**	-0.00536	-0.0433**	
	(0.0126)	(0.0145)	(0.0129)	(0.0119)	(0.0190)	
Number of Startups	17,998	17,576	17,768	16,932	17,824	

FE: $\delta_{nt}^{(j)} \times P_{nt}^{(i,j)}$, Industry-Year FE and City-Year FE.
Same controls as benchmark regression.

Back to Revenue Results