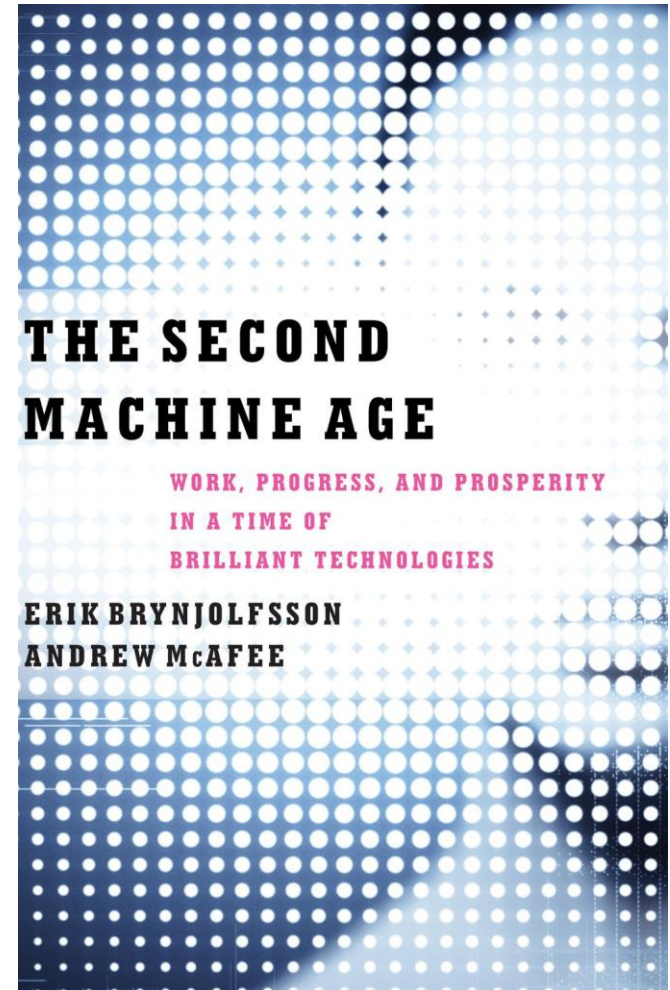# Nowcasting and Placecasting Growth Entrepreneurship

Jorge Guzman, MIT
Scott Stern, MIT and NBER
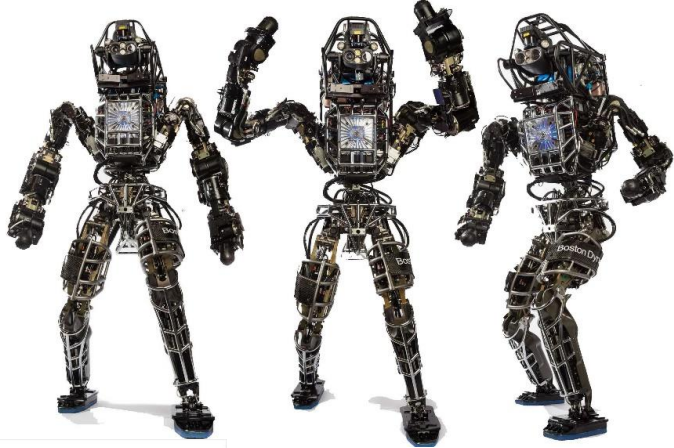
MIT Industrial Liaison Program, September 2014

The future is already here…it's just not evenly distributed…

Quote from William Gibson, Big Dog from Boston Dynamics

The Boston entrepreneurial ecosystem seems to be playing a central role in this emerging entrepreneurial cluster

But we do not understand how to measure and track entrepreneurial clusters in a reliable way....

How can we capture emerging entrepreneurial clusters robotics in real time and at different levels of granularity?

The Entrepreneurship Measurement Challenge

- Lots of interest by academics, policymakers and practitioners in measuring "growth" entrepreneurship
  - Understand the origins and dynamics of start-up firms that are commonly believed to be a key driver of economic growth and job creation
  - Be able to evaluate the role of institutions, regional ecosystems, and economic and social factors in shaping both the creation and dynamics of stat-up firms
  - Be able to forecast and measure real-time changes in the nature and location of growth entrepreneurship
- However, little consensus on what exactly is meant by growth entrepreneurship or what data might be useful
  - Traditional measurement of broad-based entrepreneurship is based on surveys (such as the Global Entrepreneurship Monitor) of randomly selected individual.
  - Much academic research *conditions* on a certain level of growth, such as the receipt of VC

MIT REAP

REGIONAL ENTREPRENEURSHIP
ACCELERATION PROGRAM

**MIT
Initiative on the
Digital Economy**

Nowcasting and Placecasting Growth Entrepreneurship

- Our research agenda introduces a novel approach to the measurement of growth entrepreneurship
  - **Business Registration**. We take advantage of the fact that nearly all growth activity requires some form of incorporation or business registration. Comprehensive and consistent over time and place.
  - **Predicting Entrepreneurial "Quality."** We use information available at the time of registration to predict the "quality" of every business registrant.   Model relates meaningful growth outcomes (e.g., IPO or high-value acquisition) to information observable about the start-up at the time of incorporation (its name, patents and copyrights, etc)
  - **Placecasting**. Creating an entrepreneurial quality index for firms in a given location for a given start-up cohort (at any level of granularity)
  - **Nowcasting**. Identifying firms or areas on a real-time basis that display high entrepreneurial quality (perhaps with information related to particular technologies or industries)

Key Findings

- Business Registration data turns out to be a rich (and essentially unused) resource that has been largely digitized and can be exploited for detailed understanding of business activity

- **Prediction.** There is a meaningful relationship between the growth outcome of start-ups and publicly available information at the time of registration (or just after)
  - 74% of growth is from top 5% of start-up quality with 53% in the top 1%

- **Entrepreneurial Quality Rather than Entrepreneurial Quantity. By focusing on "Quality," we break through the inconsistencies of prior research and develop a novel characterization of entrepreneurial clusters such as Silicon Valley and Boston**

- **Placecasting.** We track the migration of innovation in the Boston Area from Route128 to Cambridge as well as the location of individual firms.

- **Nowcasting.** Results suggest the ability to offer a real-time tool that provides detailed insight into how to use incorporation data for policy and practitioner forecasting

Outline

- The Measurement Challenge

- Data Overview

- Methodology Overview

- Where is Silicon Valley?

- Nowcasting Growth Entrepreneurship

- Predicting Employment Growth

The long-time data challenge

- Analyses of entrepreneurship must include *successful and failed* entrepreneurs.
- But failed entrepreneurs are not in data:
  - Not in venture capital data:
    - Might not raise venture capital
    - VCs might not recognize them
  - Not in innovation data:
    - Might never file a patent

- But seeing these firms is surely critical to understand entrepreneurship dynamics

If only there were a single, comprehensive and real-time source for data on all start-up activity….
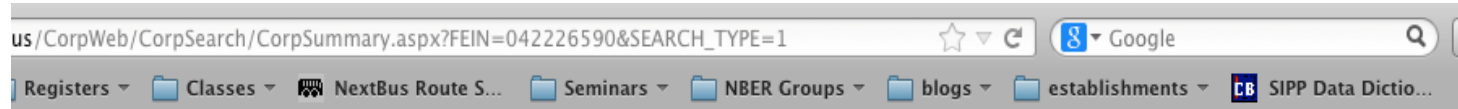
Business registration records offer a benefit above current datasets

- They are public records and can be accessed by anyone.
  - No special relationships
  - No security clearances
- They are free or very cheap to request depending on the region.
  - $50 in Massachusetts, $200 in California.
- They have the full population of firms that register for business.
  - No selection on employment, VC funding, patenting etc.
- They have panels that cover a very long period of time.
  - Often all the way back to the 1800's.

Examples of Business Registration

us/CorpWeb/CorpSearch/CorpSummary.aspx?FEIN=042226590&SEARCH_TYPE=1 ☆ ▽ C Google

Registers ▾   Classes ▾   NextBus Route S...   Seminars ▾   NBER Groups ▾   blogs ▾   establishments ▾   SIPP Data Dictio...

Corporations Division
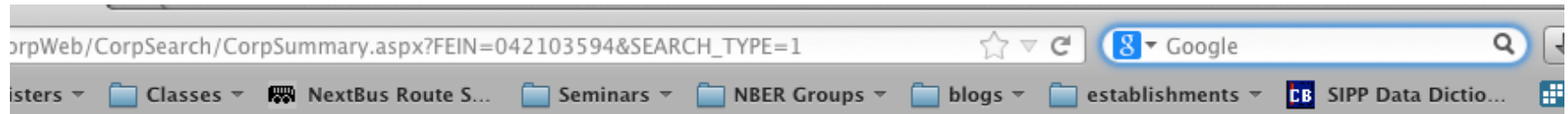
Business Entity Summary

ID Number: 042226590   Request certificate   New search

Summary for: **DIGITAL EQUIPMENT CORPORATION**

The exact name of the Domestic Profit Corporation: DIGITAL EQUIPMENT CORPORATION	
Merged into COMPAQ COMPUTER CORPORATION on 12-31-1999	
Merged with MAYNARD DEVELOPMENT CO., INC. on 06-27-1974 **Merged with MAYNARD INDUSTRIES, INC. on** 06-27-1974 **Merged with APL SOFTWARE SYSTEMS, INC.(PA) on** 06-27-1975 **Merged with DEC REALTY TRUST(MA TR) on** 08-13-1981 **Merged with COMPAQ MERGER, INC. on** 06-11-1998	
Entity type: Domestic Profit Corporation	
Identification Number: 042226590	
Date of Organization in Massachusetts: 08-23-1957	
	Last date certain:
Current Fiscal Month/Day: 12/31	**Previous Fiscal Month/Day:** 06/30
The location of the Principal Office: Address: 40 OLD BOLTON RD. City or town, State, Zip code, Country: STOW, MA 01775 USA	

Examples of Incorporation

Browser address bar: orpWeb/CorpSearch/CorpSummary.aspx?FEIN=042103594&SEARCH_TYPE=1 — Google

Bookmarks: isters | Classes | NextBus Route S... | Seminars | NBER Groups | blogs | establishments | CB SIPP Data Dictio...

Business Entity Summary

ID Number: 042103594   Request certificate   New search

Summary for: MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The exact name of the School: MASSACHUSETTS INSTITUTE OF TECHNOLOGY	
Entity type: School	
Identification Number: 042103594	**Old ID Number:**
Date of Organization in Massachusetts: 04-10-1861	
	Last date certain:
Current Fiscal Month/Day: 06/30	**Previous Fiscal Month/Day:** 01/01

The location of the Principal Office in Massachusetts:

Address: 77 MASSACHUSETTS AVENUE

City or town, State, Zip code, Country: CAMBRIDGE,   MA   02139   USA

Examples of Business Registration

Corporations Division

Business Entity Summary

ID Number: 710933087

Summary for: KIVA SYSTEMS, INC.

The exact name of the Foreign Corporation: KIVA SYSTEMS, INC.	
The name used to transact business in Massachusetts: KIVA	
The name was changed from: DISTROBOT SYSTEMS, INC. **on** 01-10-2008	
Converted into KIVA SYSTEMS LLC on 08-09-2012	
Entity type: Foreign Corporation	
Identification Number: 710933087	**Old ID Number:** 000860142
Date of Registration in Massachusetts: 02-04-2004	
Date of Conversion: 08-09-2012	**Last date certain:**
Organized under the laws of: State: DE **Country:** USA **on:** 01-10-2003	
Current Fiscal Month/Day: 12/31	
The location of the Principal Office:	
Address: 300 RIVERPARK DRIVE	
City or town, State, Zip code, Country: NORTH READING, MA 01864 USA	

Our dataset includes ~350,000 observations per year

Number of local firms registered per year
5 year periods
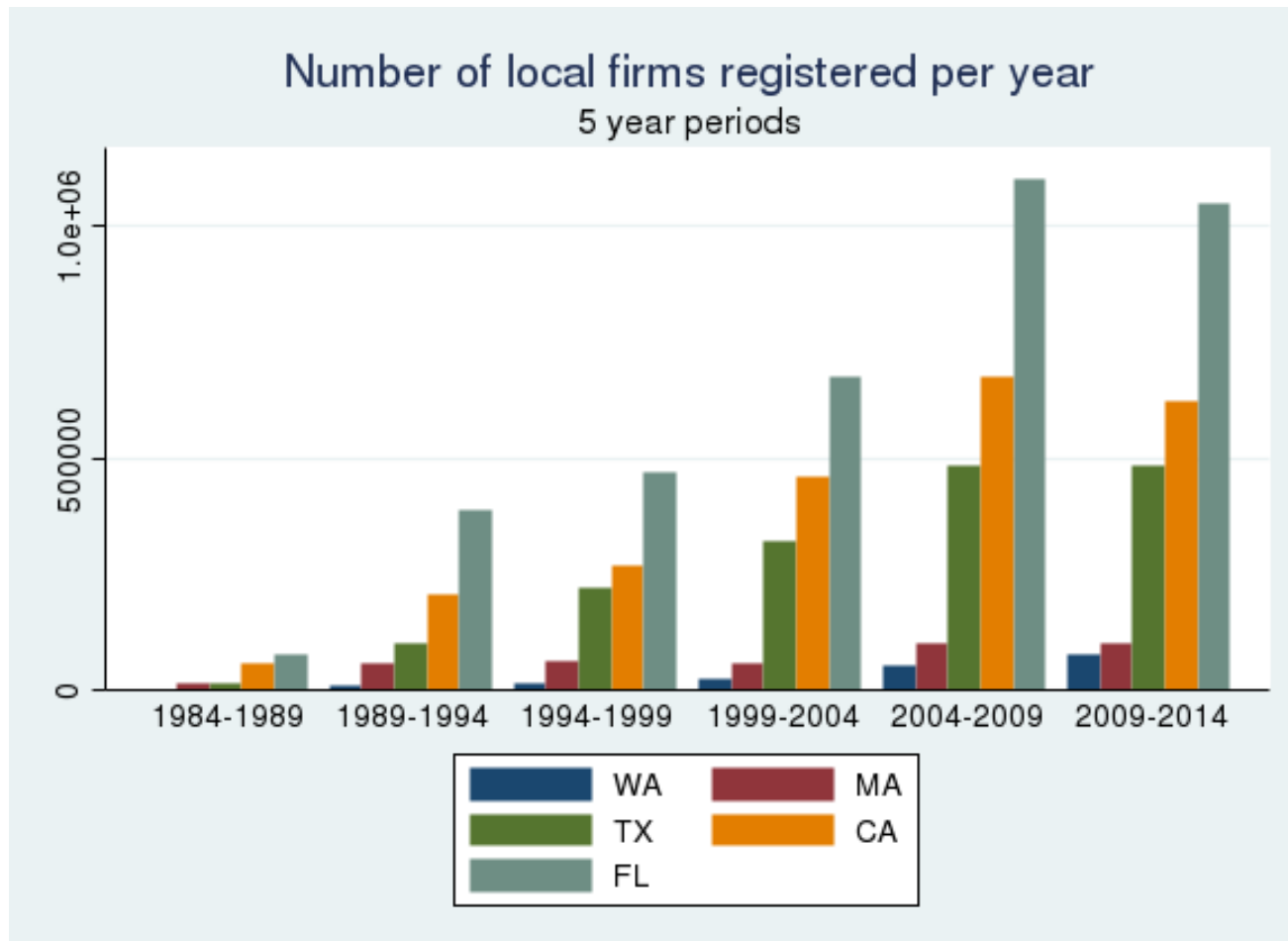
Our methodology

- Stacked logit regression:

$$P(growth_{i,t+k} | X_{i,t}, Z_{i,t}) = \alpha + \beta' X_{i,t} + \gamma' Z_{i,t}$$

- ***growth*$_{i,t+k}$**: is a binary growth outcome (today IPO or high value acquisition, but could be others)

- ***X*$_{i,t}$ *and Z*$_{i,t}$**: are early characteristics from business registration data and other sources

- **k:** a specific and constant time window to achieve the outcome (6 years)

Creating an entrepreneurial quality estimate

- After running the regression we *predict* the probability of growth on *all* firms using only information observable at founding or close to it.

- This probability of growth is their estimate of entrepreneurial quality.

APPLICATION #1:
WHERE IS SILICON VALLEY?

Guzman and Stern 2014a

The puzzle: According to rankings, Montana is the most entrepreneurial region in the US
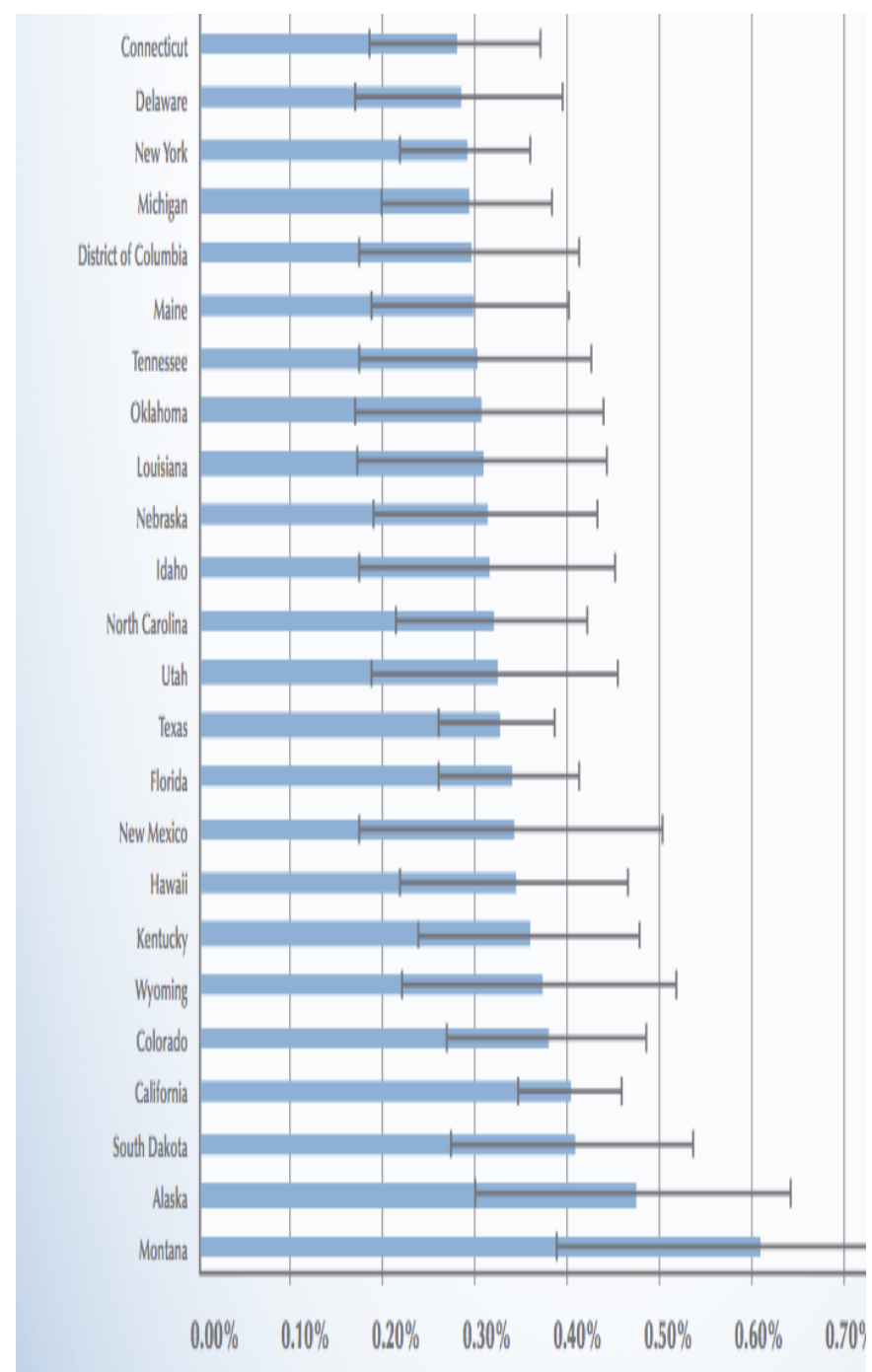
Source: 2013 Kauffman Index of Entrepreneurial Activity

Perhaps we should look at something else than quantity of firms

- Highly innovative locations like California, Massachusetts, or New York do not come out on top.

- One possible reason is that the indexes look for the number of new firms, not their quality.

- Accounting for quality is hard, and selecting proxies (e.g. through VC funding or patenting firms) can produce other biases.

Our approach: build a probability of growth

We can use our dataset to build a measure of entrepreneurial quality that includes all firms and allows them a potential for growth.

1. Stacked logit regression:

$$P(growth_{i,t+k}|X_{i,t}, Z_{i,t}) = \alpha + \beta'X_{i,t} + \gamma'Z_{i,t} + \varepsilon_{i,t}$$

- *growth*$_{i,t+k}$: is a binary growth outcome (IPO or acquisition over $10M)
- $X_{i,t}$ and $Z_{i,t}$: are early characteristics
- **k:** a specific and constant time window (6 years)
- Train with all California firms from 2001 to 2006

2. Predict for new firms:
- Consider the estimated Prob(growth) of new firms as their growth potential
- On all firms registered in California in 2009 or 2011

Logit Regression: Regressors

- Internal Measures: Information included within a business registration form
  - Delaware Jurisdiction
  - Corporation / LLC or Partnership
  - Eponymy (firm named after the founder)
  - Local Industry (restaurant, pizza, cleaners, etc)
  - Tech (Robotics, Dynamics, etc)

- External Measures: Data Observable at the Time of Founding and Matched to Bus Reg Data
  - Patent (in first year)
  - Trademark application in first year

- For years 2001 to 2006, train on 70% of the sample and test with 30%. For years 2008 to 2011, build predictive results.

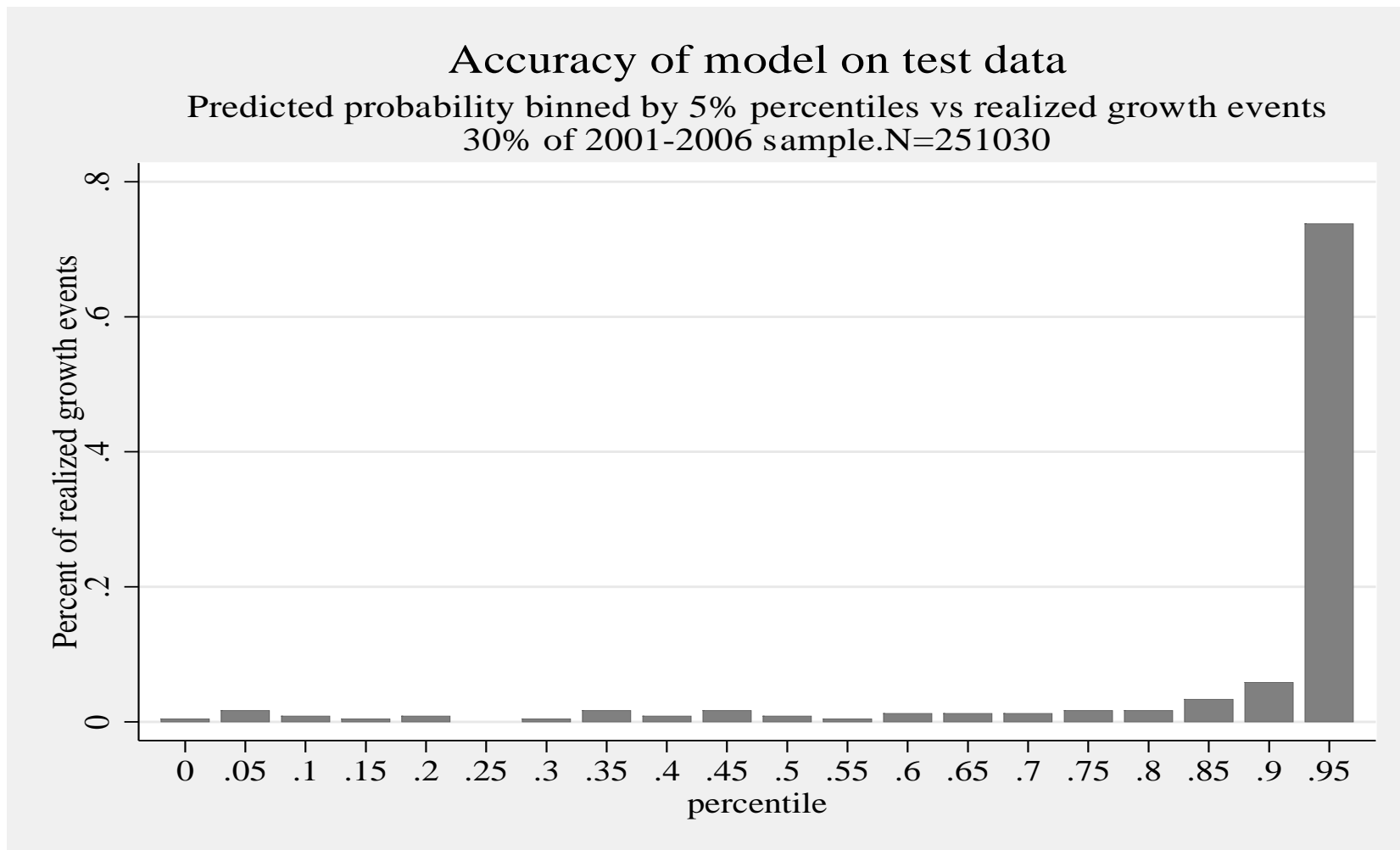| Growth Probability (Combined Odds Ratios)	
Eponymous	0.261**
	[0.10]
Local	0.188+
	[0.13]
Technology	1.812**
	[0.22]
Short Name	1.985**
	[0.23]
Corporation	4.915**
	[0.75]
Delaware	
Jurisdiction	12.82**
	[1.71]
Patent	8.028**
	[1.25]
Trademark	12.12**
	[1.79]
Constant	0.0000814**
	[0.000013]
Observations	584916
Pseudo-R²	0.31

Robust standard errors in brackets. + p<0.05 * p < 0.01 ** p < .001. Dependent variable is binary equal to 1 if a firm achieves an IPO or is acquired.

What this means

- Each coefficient is how the chance for a growth outcome (IPO or acquisition) changes depending on characteristics observable at or near the time of incorporation

- All coefficient are relative to 1.0

- For example, firms named after their founders are ~74% less likely to achieve a growth outcome than other firms, all else equal

- On the other hand, a firm with a trademark has a 1200% higher chance for a growth outcome (IPO or acquisition) than a randomly selected business registrant

- A Delaware technology-based corporation, with both early patents and trademarks, is about **20,000 times more likely to grow** than a local LLC.

Evaluating the predictive accuracy of our index

We separate 30% of our training data to do testing without bias and over-fitting. Our model's predictive accuracy ranks very well.

Accuracy of model on test data
Predicted probability binned by 5% percentiles vs realized growth events
30% of 2001-2006 sample.N=251030

Result: The top 5% of the test distribution accounts for 74% of all Growth outcomes, 53% for the "top 1%").

How to measure regional entrepreneurial quality?

- City quality = mean entrepreneurial quality per 1000 firms.

- This can be aggregated at any level.
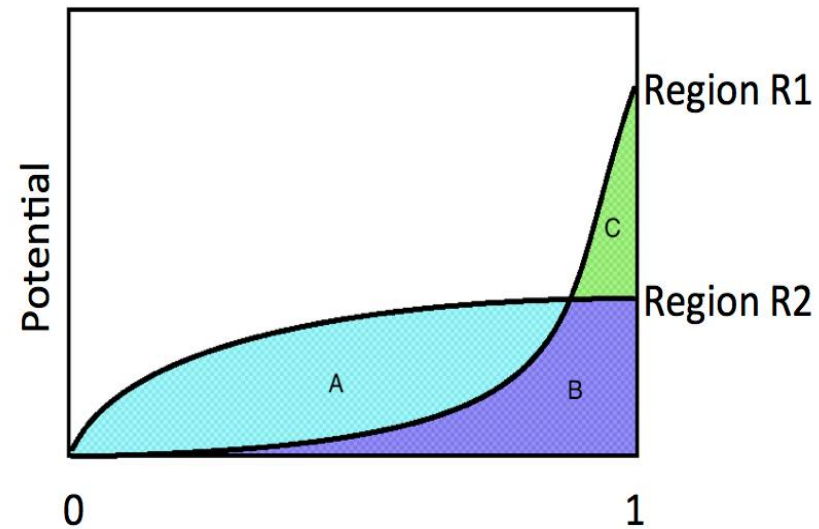
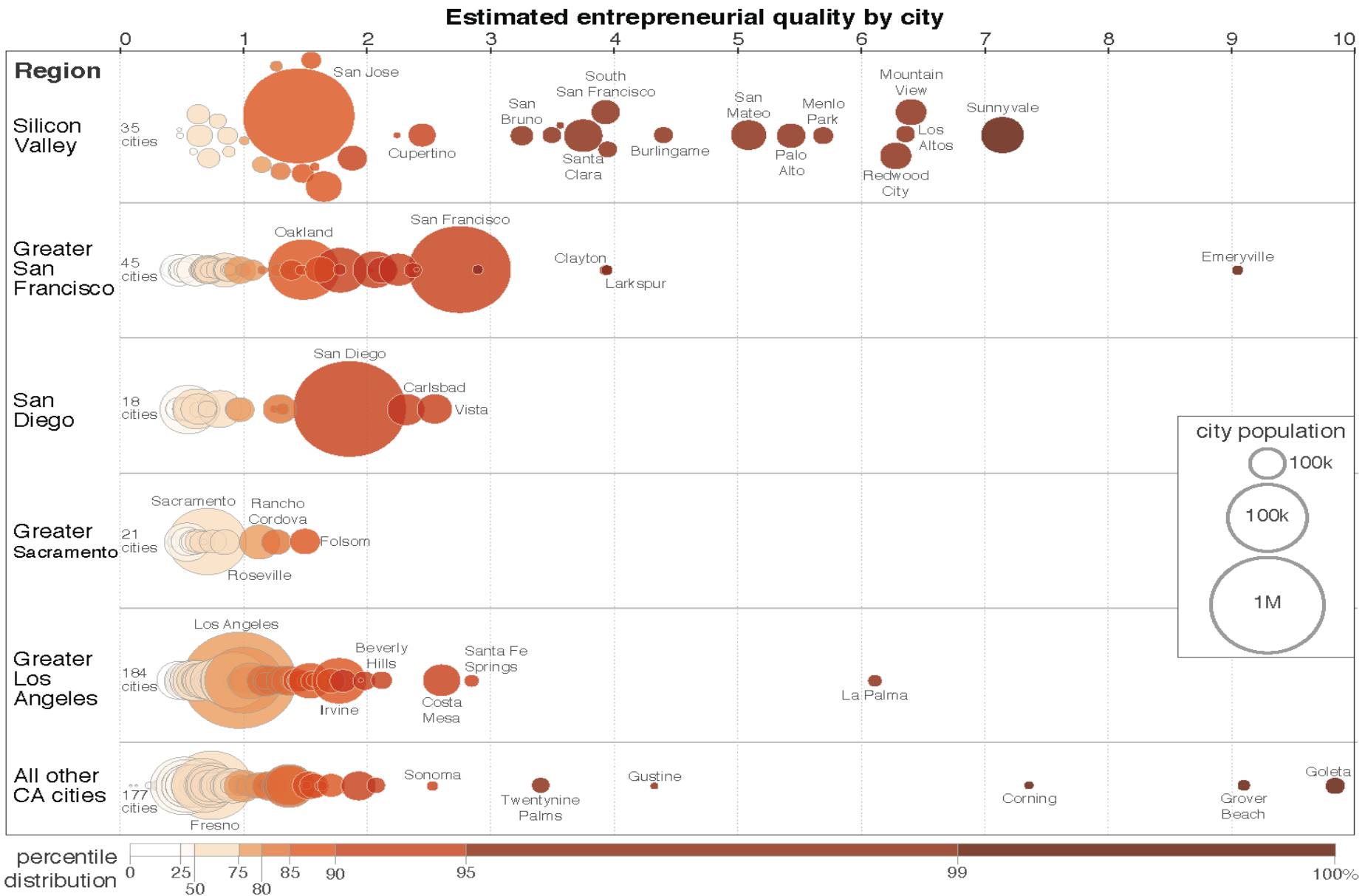$$r_r = 1000 \times \int_0^1 G_r(X)\,dg$$

Fig. 3. Estimated entrepreneurial quality by city. Estimated entrepreneurial quality by city showing all cities and towns in California listed in the Census 2010 Incorporated Places and Minor Civil Subdivisions dataset. Population of the city is a continuous measure that changes the size of each bubble. It is easy to see Silicon Valley separated in quality from other regions

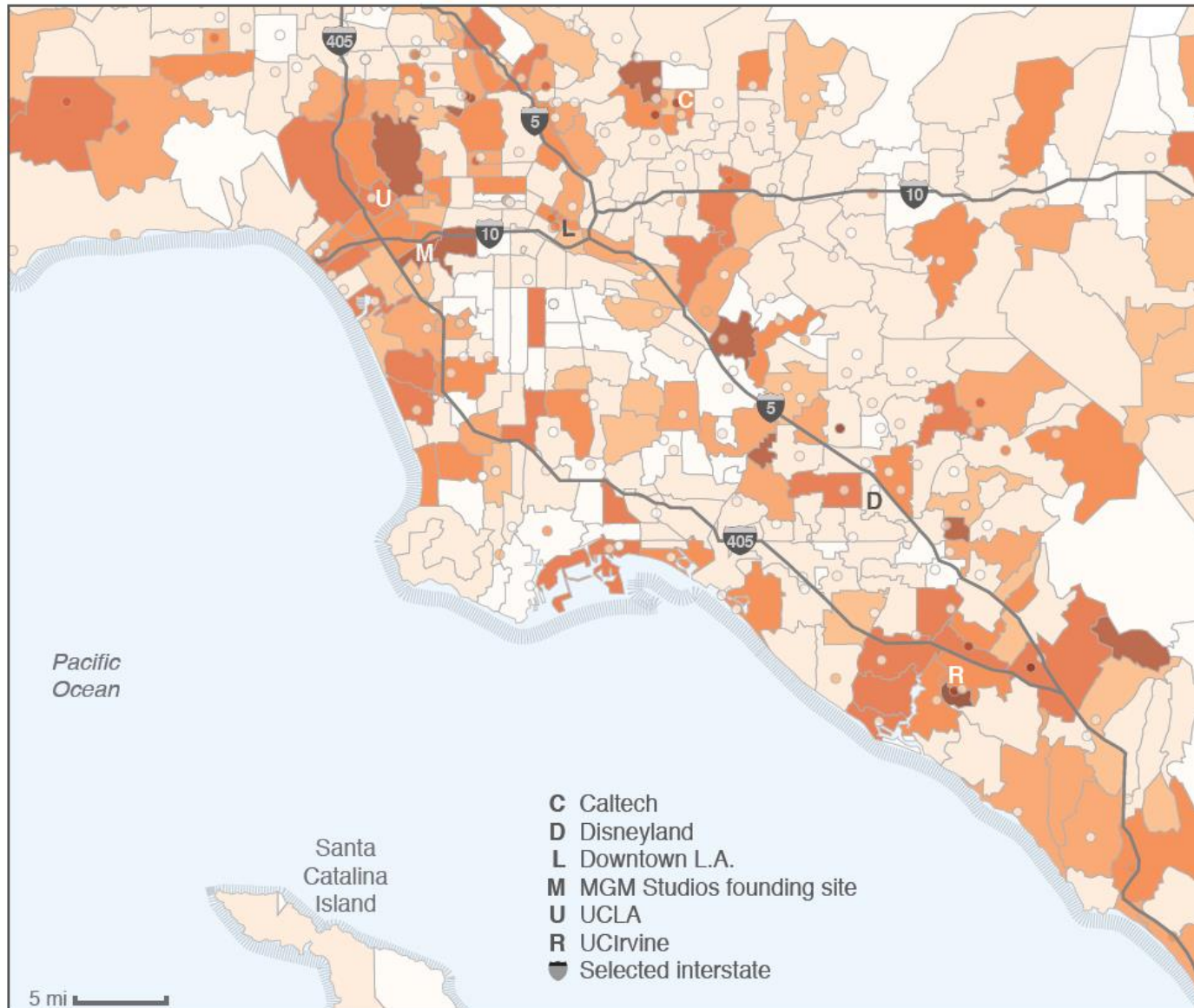**Results: Innovation of all cities in California**

Entrepreneurial Quality in the SF Bay Area
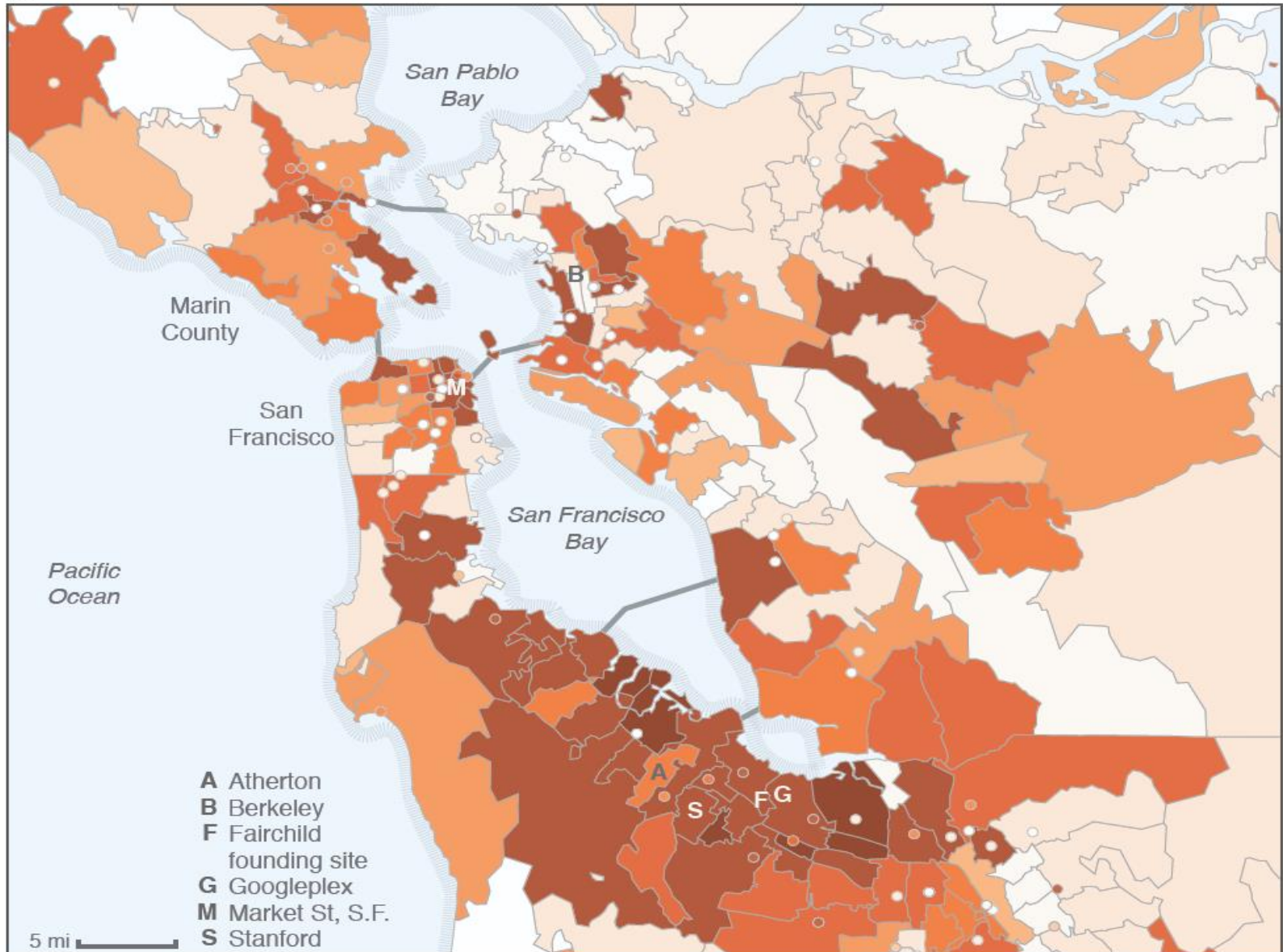
Entrepreneurial Quality in the LA Basin
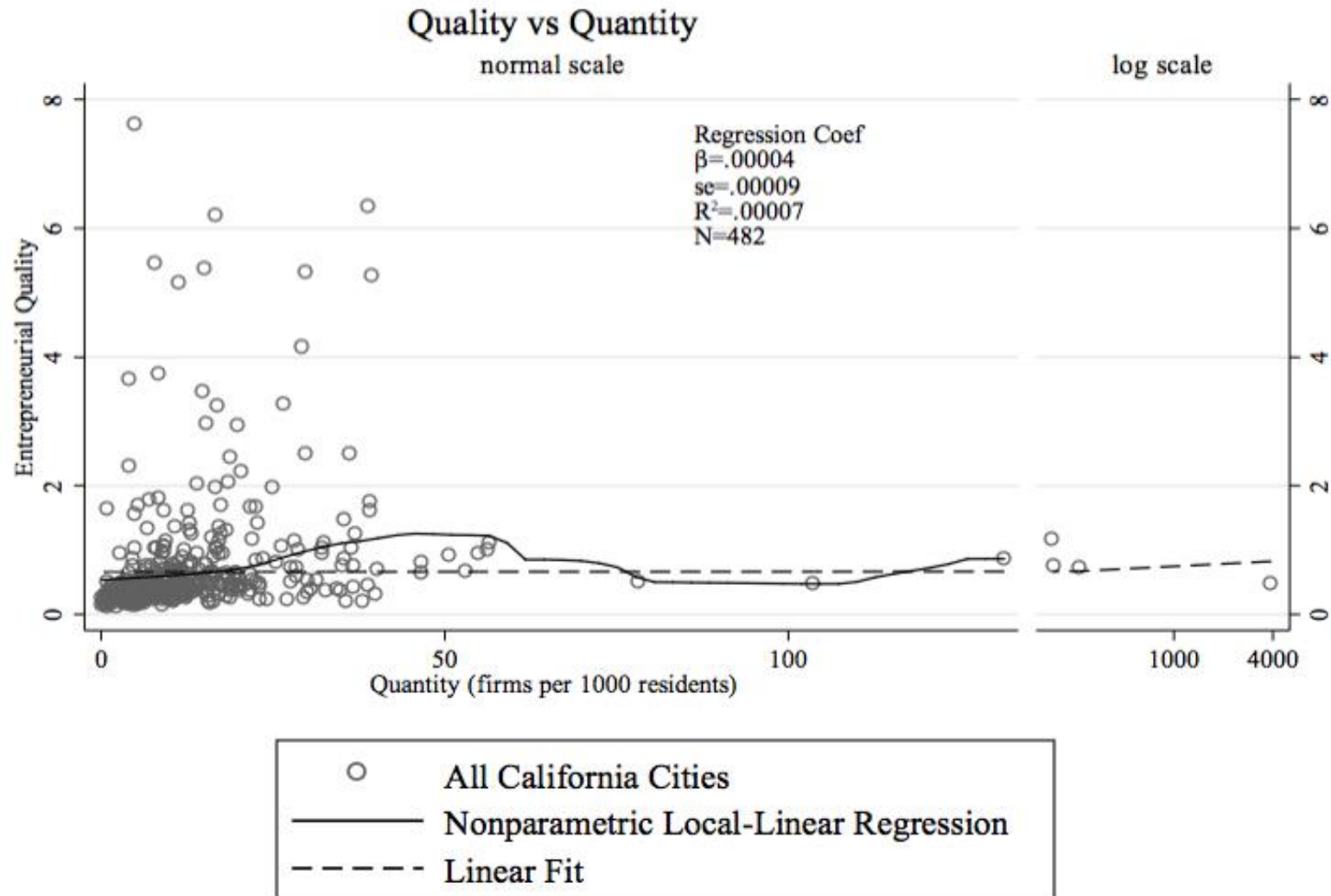
Quantity (number of start-ups per capita) and Quality (growth probability of start-ups) are mostly unrelated

Quality vs Quantity

normal scale — log scale

Regression Coef
$\beta=.00004$
$se=.00009$
$R^2=.00007$
$N=482$

Entrepreneurial Quality

Quantity (firms per 1000 residents)

○ All California Cities
— Nonparametric Local-Linear Regression
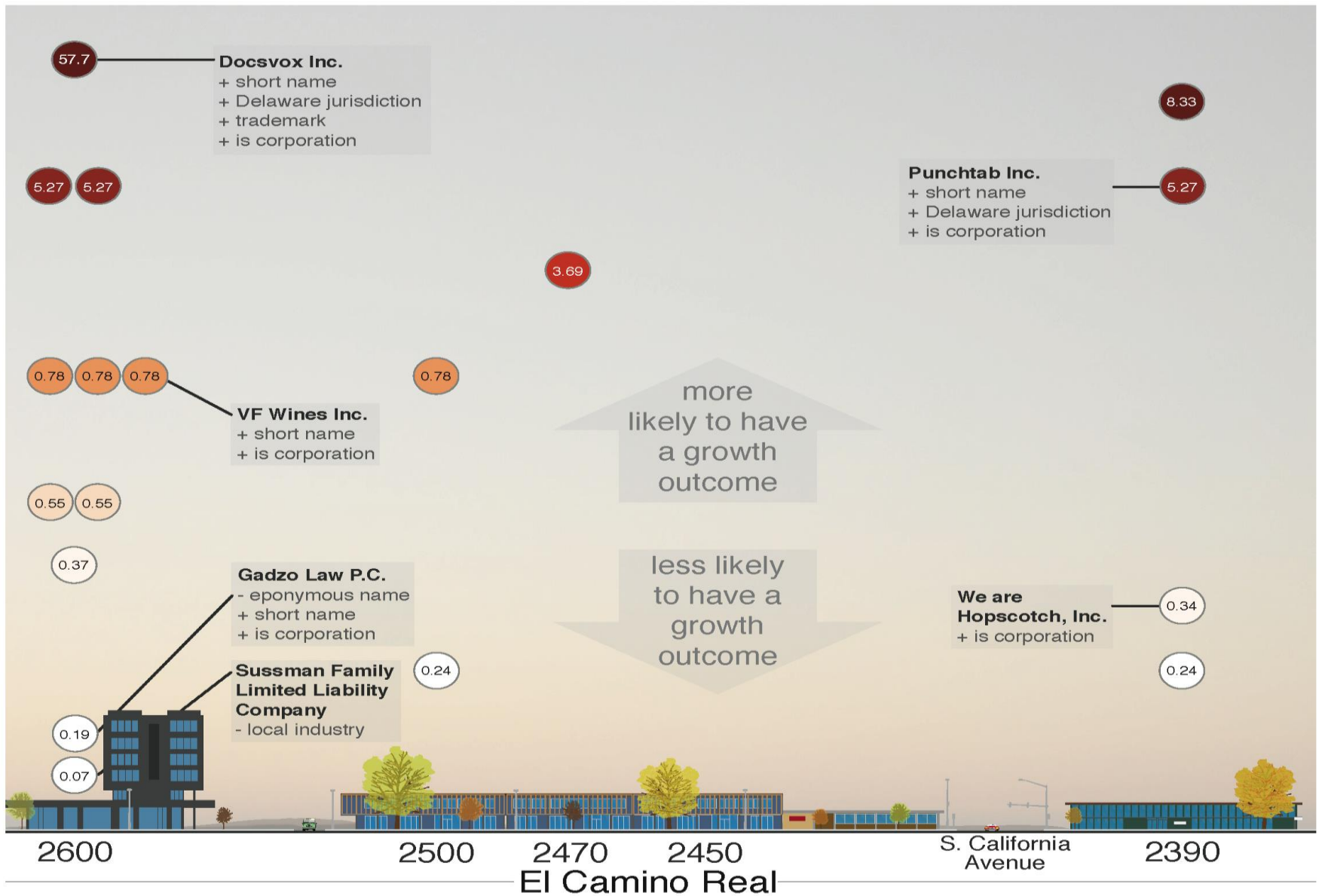- - - - - Linear Fit

Fig. 2. Section of El Camino Real, Palo Alto. This is an exemplar representative street in Palo Alto, the heart of Silicon Valley. Represented is the full set of firms registered between 2010-2011 for this block of El Camino Real with their entrepreneurial quality scores (multiplied by 1000). The color of each score is consistent with the other figures, and placed above the building the firm is associated with. The vertical positioning of the scores is ordinal, not cardinal. The block was selected due to a high number of new firms in the same location that span the full range of estimated values. High-tech businesses (Docsvox, Punchtab) rank at the top with our model; the middle level contains other traded businesses, in this case VF Wines is a wine wholesaler; lower scores contain local and professional firms, Gadzo Law and We Are Hopscotch (a graphic design consultancy); at the lowest level we find family firms.

Key takeaways

- A methodology that can be applied to any level of aggregation.

- The quality of entrepreneurship ranks Silicon Valley as the most entrepreneurial location in California

- Quality and quantity are unrelated: We need better, not more entrepreneurs.

- Quality and quantity are distinct attributes which have often been confounded and lead to vastly .

APPLICATION #2: NOWCASTING GROWTH ENTREPRENEURSHIP

Guzman and Stern 2014b

Digging into Massachusetts

1. Look at the historic migration of growth entrepreneurship in the Boston Area. Can we track the movement from Route128 to Cambridge?

2. Look at specific *individual* firms and their locations.

- These are illustrative examples on a methodology paper.

Incorporations in Massachusetts (1995-2014)

Domestic Profit Entities	Count	% of Total
Domestic Limited Liability Company (LLC)	163,027	34.2%
Domestic Limited Partnership (LP)	8,031	1.7%
Domestic Profit Corporation	179,189	37.6%
Professional Corporation	7,543	1.6%
Other Domestic Entities		
Nonprofit Corporation	29,174	6.1%
Registered Domestic Limited Liability Partnership (LLP)	1,310	0.3%
Religious (Chapter 180)	3,093	0.6%
Voluntary Associations and Trusts	2,662	0.6%
Foreign Entities		
Foreign Corporation	28,916	6.1%
Delaware firm in MA	26,192	5.5%
Foreign Limited Liability Company (LLC)	25,037	5.3%
Foreign Limited Partnership (LP)	2,222	0.5%
Total	**476,396**	**100%**

Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Industry Realtor	481809	0.0541376	0.2262893	0	1
Industry Restaurant	481809	0.0089724	0.0942971	0	1
Industry Law	481809	0.0063511	0.0794402	0	1
Industry Dental	481809	0.0026068	0.0509907	0	1
IPO Date	480	14553.82	2191.158	10995	19766
Merger Date	6462	16146.53	2499.635	10975	19788
Employees	39578	10.9951	60.17641	1	5000
Trademark in 6mo	481809	0.0120546	0.1091297	0	1
Trademark in 6-12mo	481809	0.0016957	0.0411439	0	1
Patent in 6mo	481809	0.007393	0.0856641	0	1
Patent in 6-12mo	481809	0.00165	0.0405871	0	1
Innovativeness in Name	447471	0.1012467	0.205025	2.91E+15	1
Delaware Firm	481809	0.1152967	0.31938	0	1
Eponymous	481809	0.070117	0.2553444	0	1
Is Corporation	481809	0.5401041	0.4983896	0	1
Inc Date	481809	16540.69	1962.854	12784	19723
Inc Year	481809	2004.82	5.359871	1995	2013
log(Innovativeness in Name)	447471	13.27737	3.497498	15.05149	18.42068

Results: IPO or M&A as growth

Dependent Variable: Dummy with 1 if IPO or merger > 10M within six years

Sample: Massachusetts, years 1995 to 2005, all firms

|	(1)		(3)	
	Logit Model	Marg. Effects	Logit Model	Marg. Effects
Delaware Jurisdiction	1.497***	0.0223***	1.212***	0.0136***
	(0.0409)	(0.000939)	(0.0462)	(0.000782)
Is Corporation	0.752***	0.00540***	0.599***	0.00375***
	(0.0489)	(0.000309)	(0.0505)	(0.000288)
Name innovativeness	0.196***	0.00155***	0.154***	0.00104***
	(0.0169)	(0.000135)	(0.0195)	(0.000133)
Eponymous	-1.611***	-0.00731***	-1.483***	-0.00599***
	(0.150)	(0.000345)	(0.151)	(0.000322)
Patent in 6mo			0.810***	0.00834***
			(0.105)	(0.00157)
Patent in 6-12mo			0.552*	0.00497
			(0.220)	(0.00257)
Trademark in 6mo			2.820***	0.0936***
			(0.0718)	(0.00652)
trademark in 6-12mo			0.842**	0.00886
			(0.307)	(0.00474)
Industry: Realtor			-0.594***	-0.00314***
			(0.146)	(0.000583)
Industry: Restaurant			-0.871**	-0.00399***
			(0.331)	(0.000965)
Industry: Law			-0.470	-0.00256
			(0.408)	(0.00175)
Industry: Dental			-0.724	-0.00351
			(0.721)	(0.00240)
N	251726	251726	251726	251726
Base Probability		0.00796		0.00683

Regional Patterns:
Separating High-Growth Firms

- Our goal is to see if high-growth entrepreneurship has moved from Route128 to the Cambridge area

- In this case, we simply define high-growth firms as those at the top 5% of the distribution of firms.

Quantity of entrepreneurship does not show any "shift" from Route 128 to Cambridge

Firms in Cambridge vs Route128
by cohort, all firms

(min) incyear

(sum) cambridge   (sum) route128
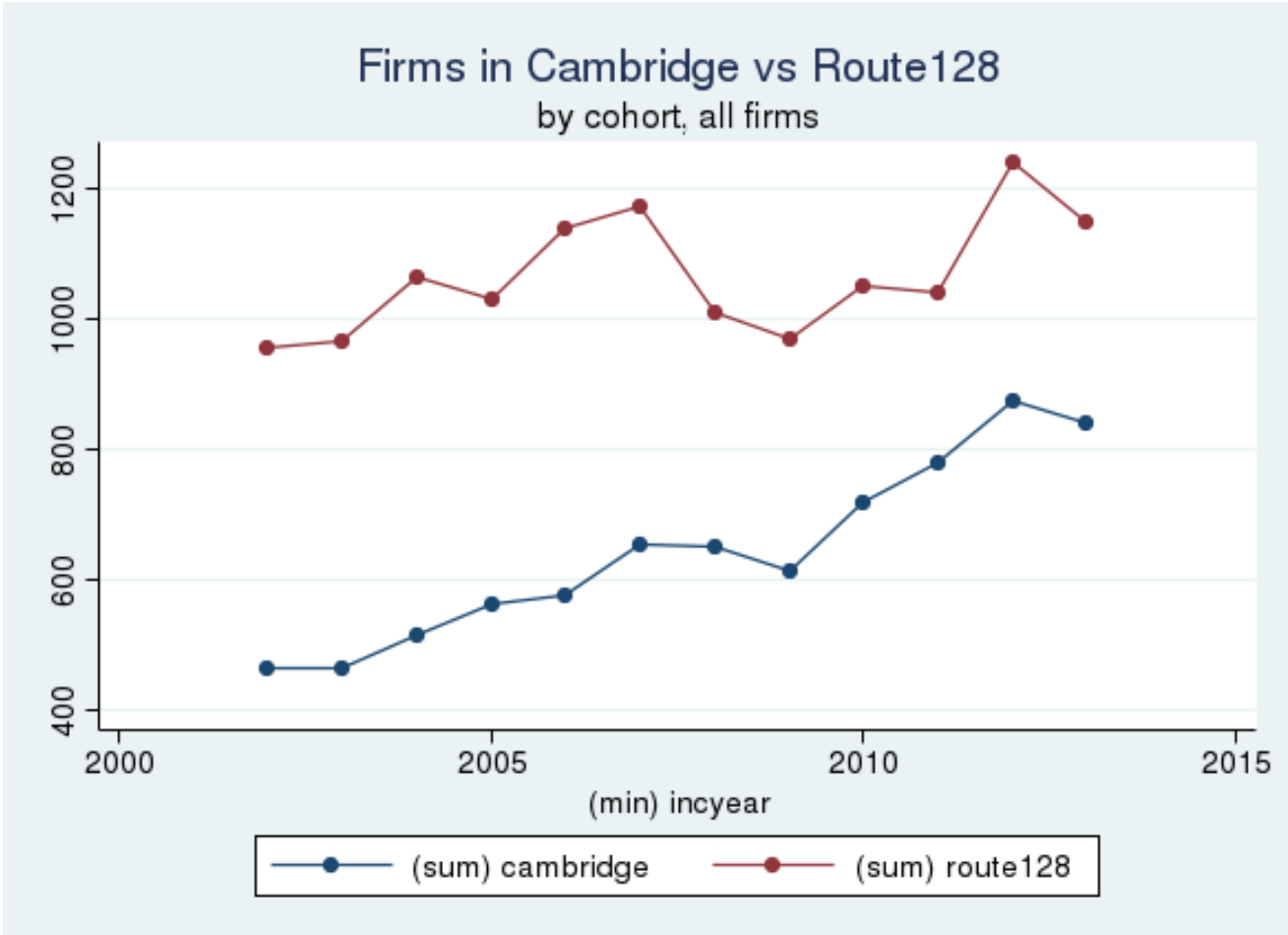
Looking at entrepreneurial quality, decline in Route 128 and surge in Cambridge

Firms in Cambridge vs Route128
by cohort, high-growth firms only
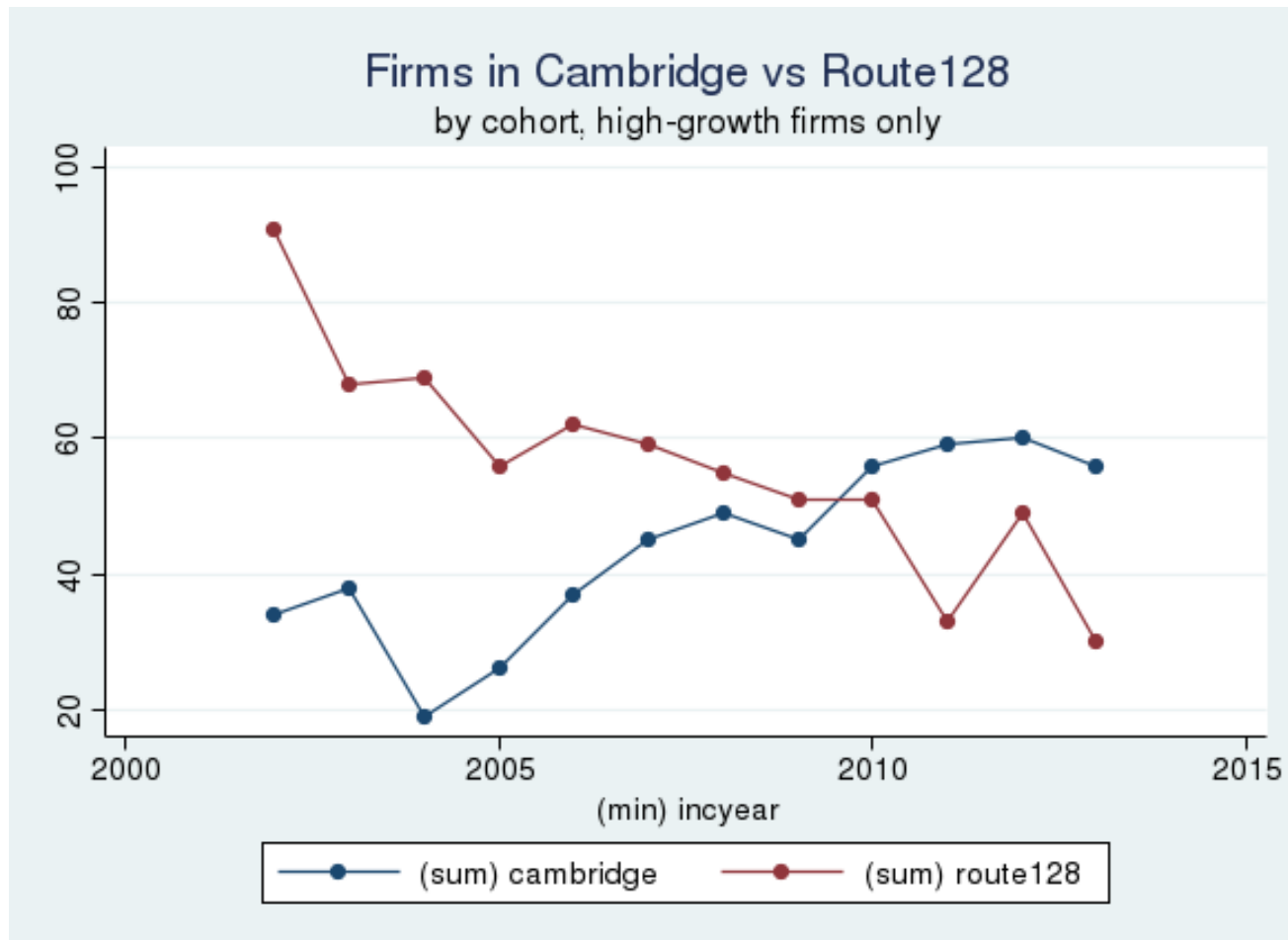
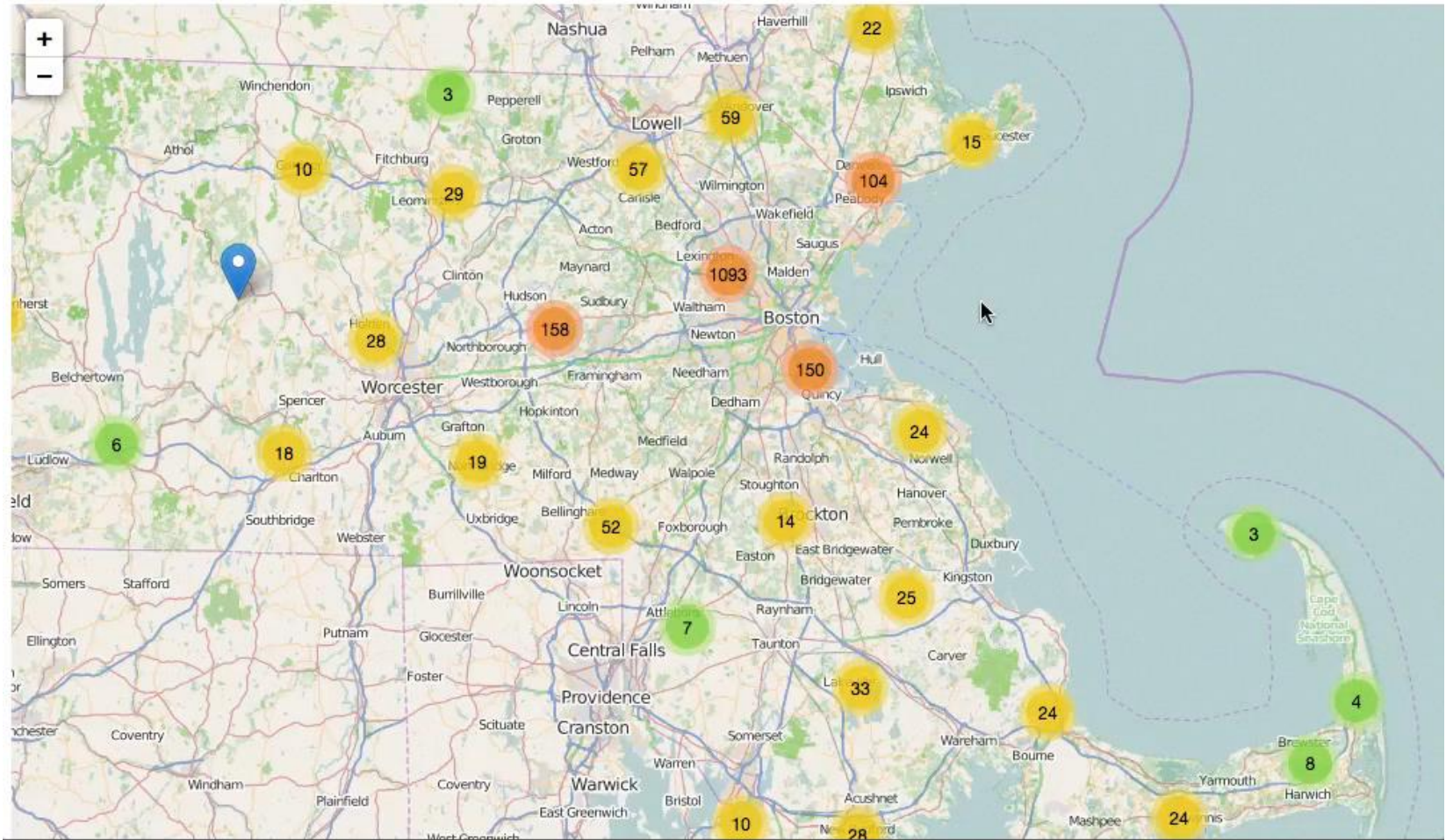# The Rise of Kendall Square
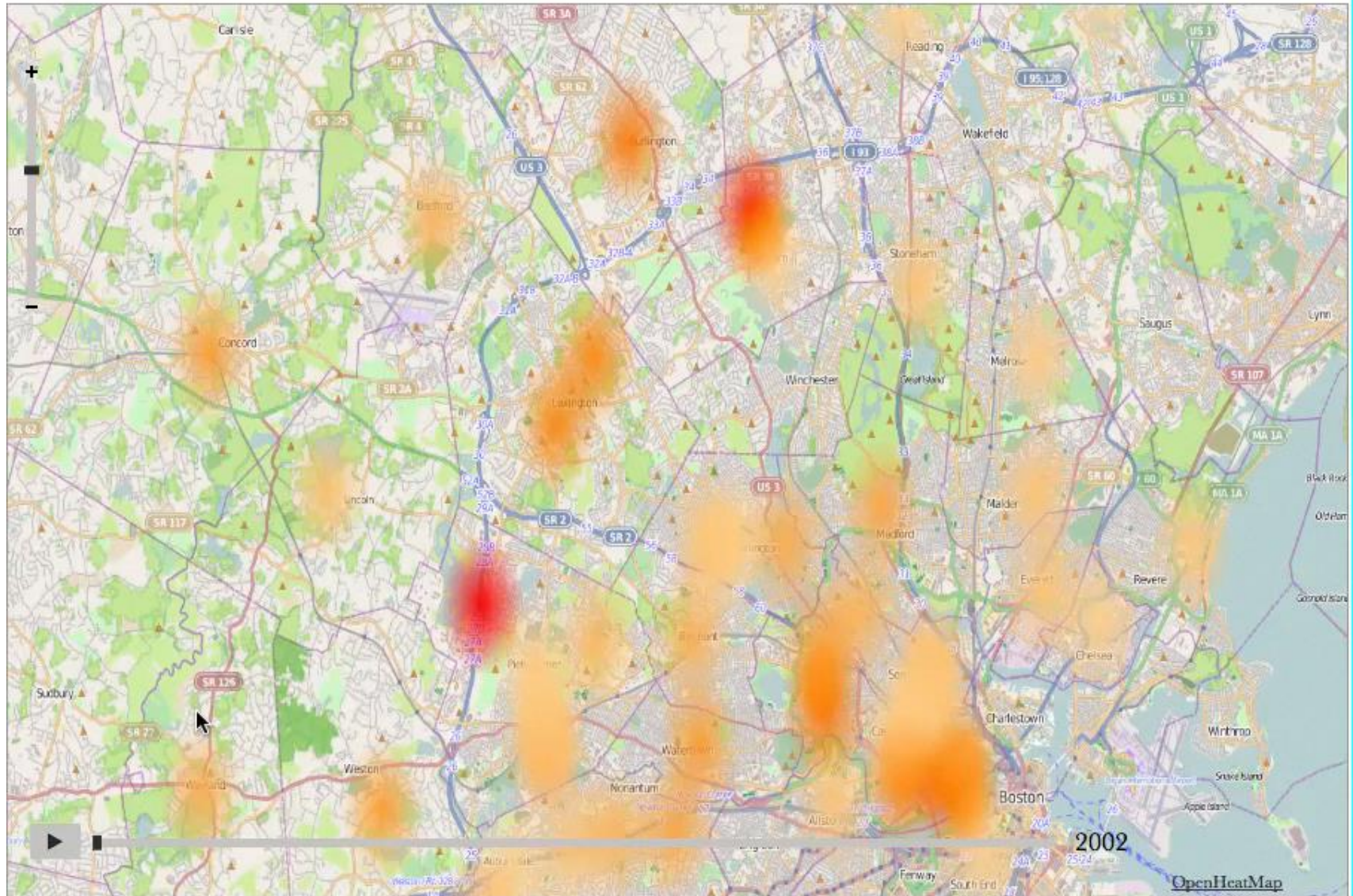
The Cambridge Innovation Center

In the year of 2011

We can also trace patterns inside the city
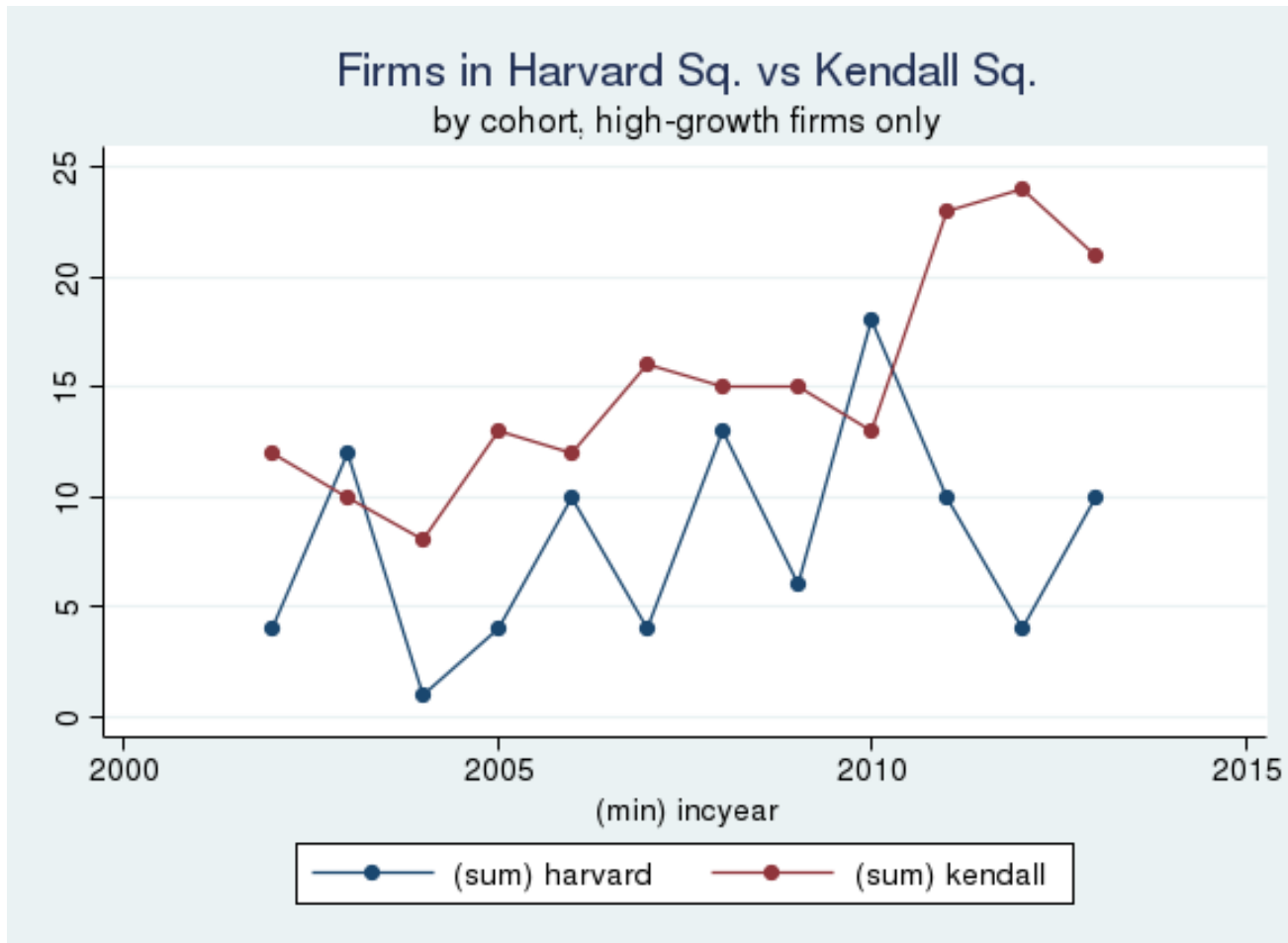
Firms in Harvard Sq. vs Kendall Sq.
by cohort, high-growth firms only

Parting Thoughts

- We have developed a new approach for measuring not simply the quantity but also the *quality* of entrepreneurship
  - Systematic approach using business registration records and predictive model provides more robust foundations than prior approaches

- Suggests that we should not be focused simply on more entrepreneurs but on encouraging *better entrepreneurs*

- Tool for the MIT Regional Entrepreneurship Acceleration Program (MIT REAP) as a way for policymakers and practitioners to track, evaluate, and target selected interventions into accelerating their regional entrepreneurial ecosystem

Using Big Data to Find Where the Future Has Already Arrived….

THANK YOU!

SSTERN@MIT.EDU
SCOTT-STERN.COM

REAP.MIT.EDU