Using Websites to Understand Factors Associated with Growth in US Green Goods Companies

Global TechMining Conference

Jan Youtie, Philip Shapira, Sanjay Arora

Georgia Institute of Technology

September 25, 2013

Background

- Green goods companies attracted attention in economic downturn
- These firms have low patenting, publication rates
  - Publications: 10%
  - Patents: 19%
- Can we use websites to understand factors leading to these firms' growth?
  - Triple Helix (Etzkowitz & Leydesdorff 1997, Ogden et al 2008)
  - Absorptive Capacity (Cohen & Levinthal 1990)
  - Regional (Uzzi 1996, 1997) v. Supra-regional (Saxenian 2005, Knight & Cavusgil, 2004)

What are Green Goods Companies?

- Selection of 300 US GGC using Dun & Bradstreet Million Dollar Database
    1. Established 2003-2007
       US headquarters
       NAICS code = manufacturing
       # of employees ≤ 250
    2. Key words applied to "Line of business" field in Dun & Bradstreet
       →2505 firms, ~700 with websites
    3. 4-point scale coding of relevance by 2 blind coders to discern "manufacturing" and "greenness"
       → ~300 resulted (a few subsequent duplicates, out-of-business)

Renew-ables, 28

Emer-ging low carbon, 94

Environ-ment, 178
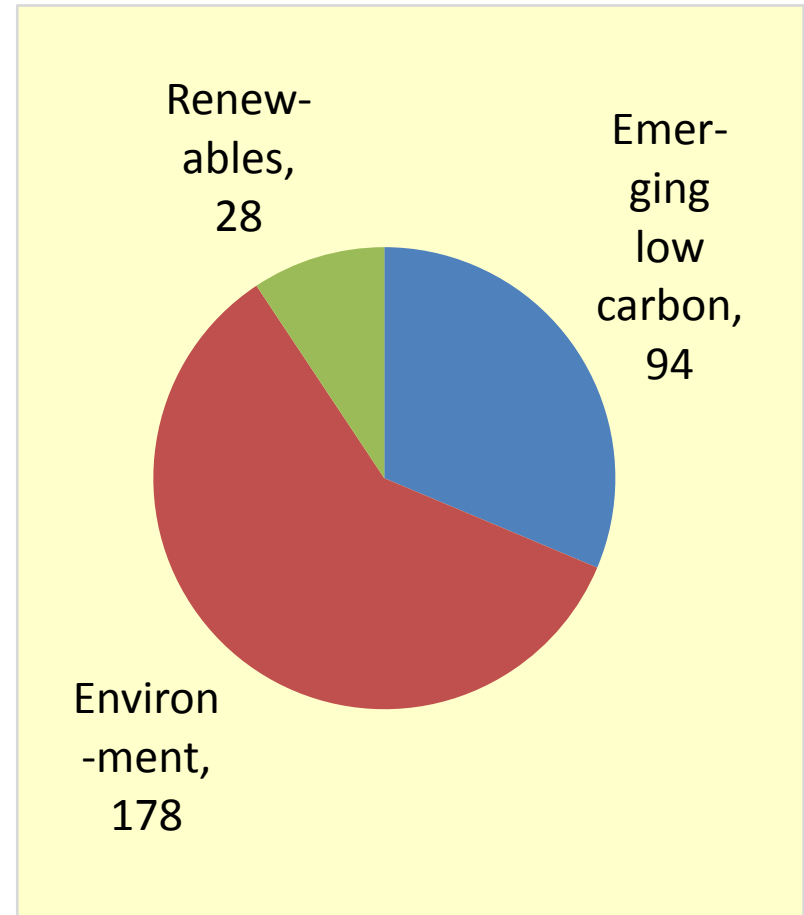
Concentration and Heterogeneity in GGC Locations

Green Goods Companies in the US
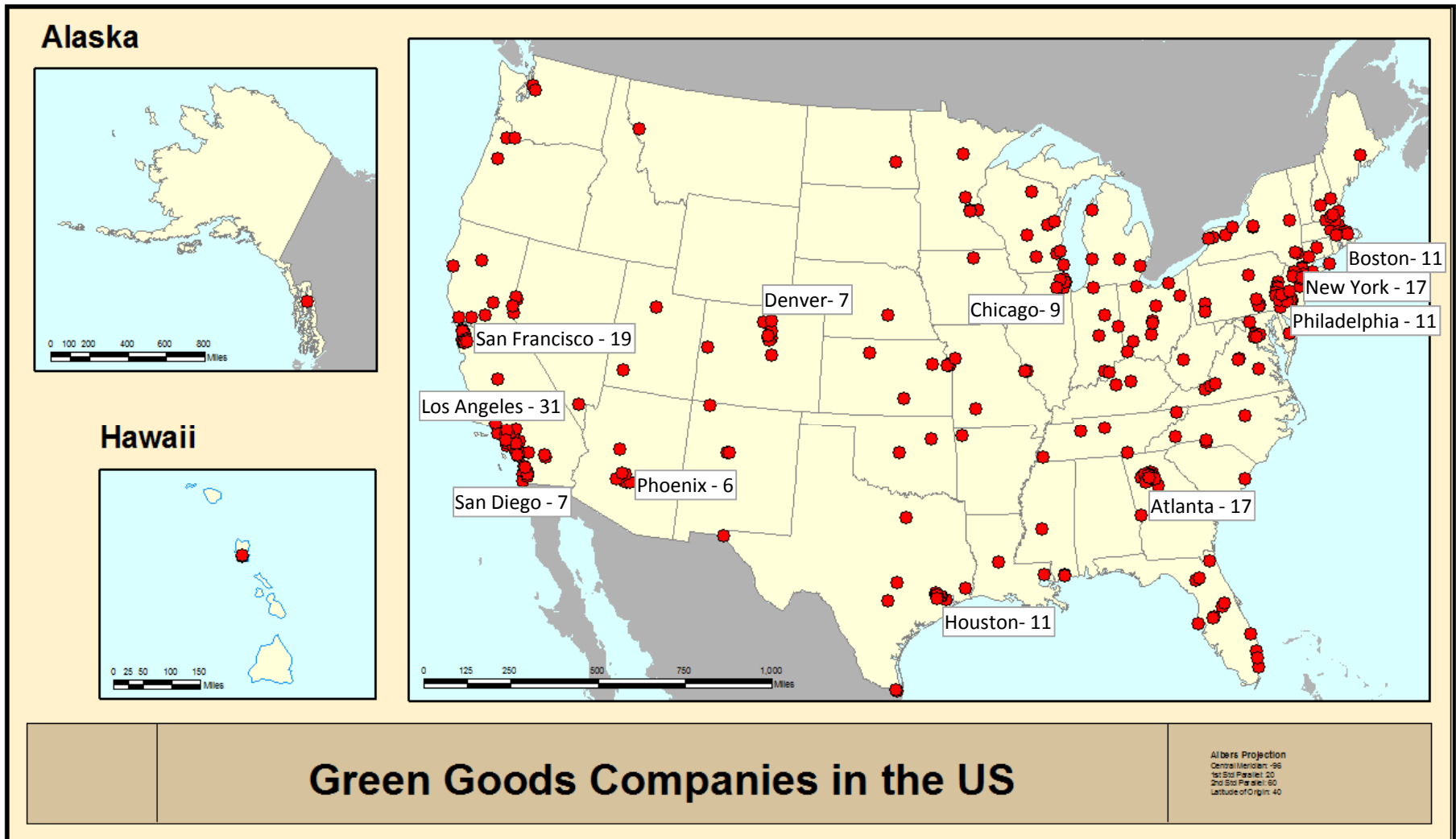
Research Propositions

- GGCs with triple helix connections more likely to have faster growth
  - Firm R&D mediates this relationship
  - Alternative: triple helix connections lead to less growth (e.g., coordination costs)
- GGCs with more regional connections more likely to have faster growth
  - Alternative: regional connections lead to less growth (e.g., limited markets)

Change in sales growth =
f(Local links, Links with universities, government, industry, Triple helix interactions, R&D, Controls)

<u>Where</u>
- Change in sales 2010-2012, dummy variable
- Links with government: **sam.gov contracting (dummy)**
- Links with university: **key terms, 2008-10, std. # pgs.**
- Links with industry (finance): **key terms, 2008-10, std. # pgs.**
- Triple helix: government, university, industry 3 way, all 2 way interactions
- Geographic links: local, national, international: **geographic names, 2008-10, std. # pgs.**
- R&D: **key terms, 2008-10, std. # pgs.**

<u>Controls</u>
- Employment logged
- D&B based classification of firms by "technology focus" (emerging low carbon, environmental, renewables)
- Size of web presence: average number of words per page per year

<u>Websites 2008-10</u>

INTERNET ARCHIVE
WayBackMachine
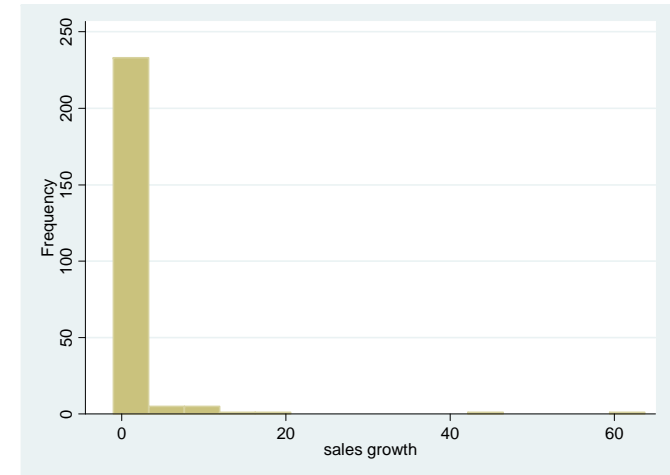
Source: archive.org

Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Sales Growth	249	0.97	5.33	-0.94	63.71
Sales Growth Dummy	249	0.53	0.50	0.00	1.00
Govt	298	0.52	0.50	0.00	1.00
R&D	261	0.68	1.19	0.00	9.76
Local	261	0.21	0.50	0.00	3.66
National	261	1.18	1.63	0.00	9.56
University	261	0.16	0.35	0.00	2.13
Industry Index	261	0.44	1.32	0.00	15.33
University*Industry	261	0.29	2.11	0.00	32.69
Govt*Industry	261	0.27	1.11	0.00	15.33
Govt*University	261	0.11	0.32	0.00	2.13
Govt*University*Industry	261	0.24	2.09	0.00	32.69
R&D*University	261	0.38	1.61	0.00	20.81
Employment (logged)	282	2.79	1.62	0.00	8.09
Low Carbon	298	0.31	0.46	0.00	1.00
Renewable Energy	298	0.09	0.29	0.00	1.00
Words per page (logged)	261	5.66	0.76	2.25	7.70

Web variables are in red.

Model Results

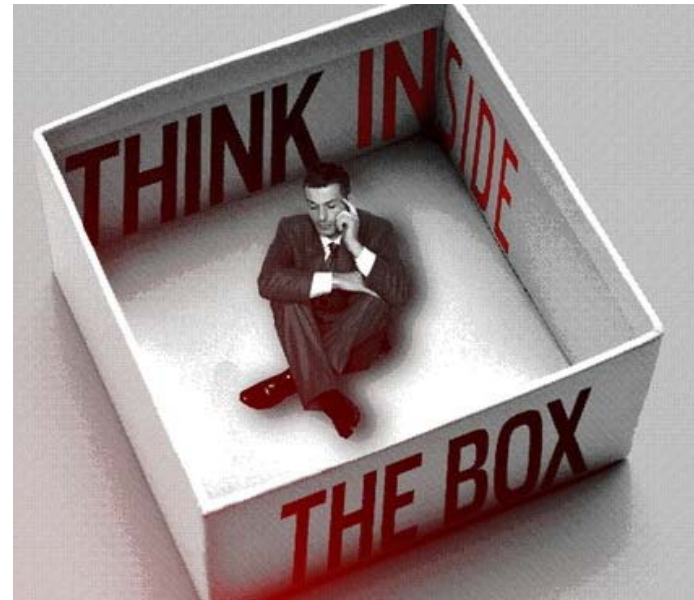| DV: Sales Growth (Dummy)	
R&D	-0.31
Govt	-1.00***
Local	0.33
National	-0.24*
University	-4.94***
Industry Index	-1.02
University*Industry	3.12
Govt*Industry	2.40**
Govt*University	2.76*
Govt*University*Industry	-4.29**
R&D*University	1.40**
Employment (logged)	0.51***
Low Carbon	-0.22
Renewable Energy	-0.36
Word per page (logged)	0.70**
Constant	-4.10***
Observations	209
R-squared	0.17
Correctly Classified	72%
Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1	

Further Interpretations

- Triple Helix variables
  - "Govt", "university", "industry index" do not have positive impact on growth by themselves.
  - **"Govt*Industry" has significant positive impact on growth,** indicating growth when there is **public-private joint investment**
  - **"Govt*University" significant positive impact on growth,** suggesting public support in **commercializing academic research**
  - **"Govt*University*Industry" has negative impact on growth.**
    - Multiple linkage might incur **high coordination costs** to the firm, distracting it from commercial pursuit.
    - Both public and private investment + university connections might happen to be **research-oriented ventures**, which place less focus on commercialization.
  - University*Industry: no effect – each needs Govt separately?

Further Interpretations

- Absorptive capacity variables
  - "R&D" and "university" are not linked to firm growth individually.
  - "R&D*university" has **significant, positive impact** on sales growth, confirming **role of absorptive capacity** to absorb and commercialize technology transferred from the university
- Geographic variables
  - Not significantly or positively linked to firm growth
  - May need finer grained coding to measure geographic impacts

Issues in Using Web Variables

- Limits in how far back in time Web variables are useful in Wayback
- Not a quick process
  - Data cleaning and integrity at each phase critical
  - Messy results for people, **places**, organizations
  - More useful with theory v. exploration

WSJ, 15 June 2013