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

American Evaluation Association, Washington DC
Jan Youtie, Philip Shapira, Sanjay Arora
Georgia Institute of Technology
October 14-20, 2013

Research for this working paper was supported by the Project on Sustaining Growth for Innovative New Enterprises, sponsored by the UK Economic and Social Research Council. The opinions and conclusions presented in the paper are those of the authors and should not be attributed to the project sponsor.
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)
What are Green Goods Companies?

• Selection of 300 US GGC using Dun & Bradstreet Million Dollar Database
     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)
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(Links with universities, government, industry, Triple helix interactions, R&D, Local links, Controls)

Where

• 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.

Controls

• 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

Source: archive.org
Explanatory Variables Operationalized

**Government**
- dummy for whether the firm registered in sam.gov for government contracts

**University**
- University
- College
- Institute
- Academy

**Industry**
- Financial
- Venture capital

**R&D**
- lab*
- R* & D*
- Researcher
- Scientist

**Local/National/International**
- Count of geographic Names

All web variables standardized by # web pages, 2008-10
## Descriptive statistics

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

*Web variables are in red.*
## Model Results

**DV: Sales Growth (Dummy)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;D</td>
<td>-0.31</td>
</tr>
<tr>
<td>Govt</td>
<td>-1.00***</td>
</tr>
<tr>
<td>Local</td>
<td>0.33</td>
</tr>
<tr>
<td>National</td>
<td>-0.24*</td>
</tr>
<tr>
<td>University</td>
<td>-4.94***</td>
</tr>
<tr>
<td>Industry Index</td>
<td>-1.02</td>
</tr>
<tr>
<td>University*Industry</td>
<td>3.12</td>
</tr>
<tr>
<td>Govt*Industry</td>
<td>2.40**</td>
</tr>
<tr>
<td>Govt*University</td>
<td>2.76*</td>
</tr>
<tr>
<td>Govt<em>University</em>Industry</td>
<td>-4.29**</td>
</tr>
<tr>
<td>R&amp;D*University</td>
<td>1.40**</td>
</tr>
<tr>
<td>Employment (logged)</td>
<td>0.51***</td>
</tr>
<tr>
<td>Low Carbon</td>
<td>-0.22</td>
</tr>
<tr>
<td>Renewable Energy</td>
<td>-0.36</td>
</tr>
<tr>
<td>Word per page (logged)</td>
<td>0.70**</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.10***</td>
</tr>
</tbody>
</table>

Observations | 209
Pseudo 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