Draft – preliminary version, subject to revision

Regional Agglomerations and Innovation Dynamics:

Is Location Important for Green Goods Companies?

Marianne Sensier[a], Abdullah Gök[b], Philip Shapira[b,c] and Jan Youtie[c]

May 2013

Working Paper 2013/04
Project on Sustaining Growth for Innovative New Enterprises
Manchester Institute for Innovation Research
Manchester Business School
University of Manchester, Manchester, UK

Author affiliations and addresses:

[a] Department of Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK; marianne.sensier@manchester.ac.uk

[b] Manchester Institute for Innovation Research, Manchester Business School, University of Manchester, Manchester M13 9PL, UK; pshapira@mbs.ac.uk; abdullah.gok@mbs.ac.uk

[c] School of Public Policy, Georgia Institute of Technology, Atlanta GA 30332-03345, USA; pshapira@gatech.edu; jan.youtie@innovate.gatech.edu

Acknowledgements: 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.

Regional Agglomerations and Innovation Dynamics:

Is Location Important for Green Goods Companies?

Abstract: The green goods sector has received much attention for its potential to re-generate the economy in addition to its energy and environmental benefits. At the same time, the promise of "green jobs" for example, has not necessarily lived up to the hopes for economic renewal through the green economy. To understand what is behind the innovation dynamics of the green sector, this study looks at a subset of small and medium sized enterprises (SMEs) involved in manufacturing green goods. The paper uses an explicit definition of green goods that is keyword based, building on prior definitions while going beyond traditional government industry classifications and patent-oriented definitions. The work is guided by a set of hypotheses that above average growth in green goods companies is positively associated with the extent of linkages to other industries, universities, and government agencies and research organizations. We further posit that these linkages will be more significant to the extent that they occur in the same local region as the green goods company. The results suggest that local government linkages are positively associated with growth of green goods companies, a finding which emphasizes the importance of demand-led innovation by regional government entities. On the other hand, a substantial role for international linkages with companies and universities outside the local region is noted in the data. This finding suggests that green goods companies must be both local and global in their orientation and linkages with government, universities, and businesses if they are to be successful.

Keywords: Innovation dynamics, agglomeration, green goods sector, web-scraping

1. Introduction

There is a growing interest in innovation for sustainability and the development of green industries and green jobs. The green sector is defined in the UK Government's report on skills for a green economy (HM Government, 2011) "as one in which value and growth are maximised across the whole economy, while natural assets are managed sustainably. Such an economy would be supported and enabled by a thriving low-carbon and environmental goods and services sector. Environmental damage would be reduced, while energy security, resource efficiency and resilience to climate change would all be increased".

To understand what is behind the innovation dynamics of the green sector our work is guided by a set of hypotheses that above average growth in green goods companies is positively associated with the extent of linkages to other industries, universities, and government agencies and research organizations. We further posit that these linkages will be more significant to the extent that they occur in the same local region as the green goods company. Existing studies on the green economy mainly work with aggregate public data (UNEP, 2011; OECD, 2011; Spencer and Arwas, 2013; PCT, 2013) and these mix manufacturing, service, and government sectors, making it difficult to parse out the value-added contribution of green goods manufacturing and what factors lead to innovation and growth in this sector. Moreover, because green goods are an emerging technology and an engineering-driven industry (rather than being purely science driven), reliance on traditional measures of innovation such as scientific publications and patents do not do a good job of capturing how the sector is innovating and growing. In this paper we analyse a subset of small and medium sized enterprises (SMEs) involved in manufacturing green goods and utilise web-scraping methods to collect data from company web-sites which compliments traditional measures of innovation. As the green goods sector's growth is related to some extent to regional or national government policies such as renewable portfolio standards or environmental regulations as well as being tied in some cases to university research and industry collaboration, we propose the relevance of a modified triple helix model to understand the innovation attributes of this sector.

The paper uses an explicit definition of green goods that is keyword based, building on prior definitions while going beyond traditional government industry classifications and patent-oriented definitions. The results suggest that local government linkages are positively associated with growth of green goods companies, a finding which emphasizes the importance of demand-led innovation by regional government entities. On the other hand, a substantial role for international linkages with companies and universities outside the local region is noted in the data. This finding suggests that green goods companies must be both local and global in their orientation and linkages with government, universities, and businesses if they are to be successful.

The paper proceeds as follows. In the next section, we discuss literature and concepts about green goods, enterprise growth, and regional location factors. We draw on this to develop propositions about the locational factors that might influence green goods enterprise growth. This is followed by a section where we summarise our methods for identifying green goods companies. We analyse some descriptive statistics on our firm data set from the United Kingdom (UK) and the United States (US). We also explain how we have web-scraped additional data on firms in our data set. Section 4 discusses how we operationalise and model this data. Results are presented in Section 5 on our initial estimates from the green goods SMEs. The final section offers some concluding remarks.

2. Literature and key concepts

In countries around the world, the financial crises of the late 2000s and their ongoing aftershocks have stimulated renewed debate about the importance of economic rebalancing (Harrop, 2013; Stijn et al, 2011). While there is much variation as to how the objectives and processes of rebalancing are described, common themes can be discerned. In the UK, there has been heightened discourse about strengthening manufacturing and shifting towards low-carbon and greener production and consumption so as to counterbalance an overreliance on financial services, foster regional growth outside of the financial capital of London, promote exports, and ensure a more resilient path for economic recovery and sustainable growth (HM Treasury, 2011).

Determining the spatial location of green goods firms and their linkages both locally and globally seems to be key in understanding the innovation dynamics of a green firm. We investigate if green firms exist within a cluster or agglomeration; are they well connected to their local network of private firms and public research organisations and do these in turn help to link them to the global market place? A new but growing literature has investigated the relationship between innovation and space and found that there is more at play than just local innovation dynamics. Shearmur (2011) specifically argue that it is accessibility to the infrastructure, information and markets available in larger cities that influences the geography of innovation, and that different types of industries will locate differently with respect to cities depending on their interaction requirements. Kauffeld-Monz and Fritsch (2013) analyse 18 German regional innovation networks and they demonstrate that public research organisations (especially universities) are involved in the knowledge exchange process and have more central broker positions than private firms. They suggest that the public research organisations have a "gatekeeper" function.

Shearmur (2012) has sought to distinguish agglomeration economies from local innovation dynamics in a study of Knowledge Intensive Business Services in Montreal. He actually finds that the most intense innovators locate away from the high employment clusters. Agglomeration economies occur when the concentration of economic activities leads to the emergence of positive externalities, which are transmitted both within and between firms through channels such as technological spillovers, an increasingly skilled labour market and enhanced firm-supplier networks. We build upon previous work in Sensier, Curran and Artis (2011) were we found that over recent years, both industry and service sectors have increased productivity in city regions (although for industry this has been accompanied by the cost of falling employment in this sector), where they benefit from the growing demand that a relatively affluent daytime population brings. The long term decline in traditional manufacturing, caused by increased competition in the global market place, also means industries are less likely to cluster together. However, we find that the financial services sector does benefit from locating near companies in the same sector. Davis (2011) explains that traditional manufacturing in the UK were dispersed geographically mainly due to natural resources – i.e. the concentration of steel works in Wales due to the availability of pig iron and coal – but now with the rise of the knowledge industries, and the greater need for face-to-face contact, people cluster in particular regions which has lead to the regional imbalances we see today.

Papers that have sought to determine a link between manufacturing intensity and innovation include the papers of Coad and Rao (2011) and Coad and Rao (2008). Here the authors apply quantile regression to US high-tech manufacturing firms to determine if rates of growth (in employment and sales) differ with the size of the firm. In both papers the authors extract an index of innovation from R&D and patent data using principal components analysis to filter out extremes. Coad and Roa(2008) establish that innovation is of greater importance for the fastest growing firms (by sales) in their sample. Coad and Rao (2011) analyse the growth of firms by employment size and find the firms that experience the fastest employment growth have invested previously in innovation they also find that larger firms fare better.

In the next section we will describe the concepts being tested and what they measure. We will operationalise these concepts in our modified triple helix model to understand the innovation attributes of the sector with regression models, comparing OLS to logistic models.

3. Methods and data

This section discusses the approach used for identifying and developing data on green goods enterprises in the UK and US. In Shapira, et al (2013) we have developed a key word based method for identifying a set of green goods companies. The novelty of our approach comes from adding to our explanatory variables with information that is collected from firm's web-sites. This "web-scraping" data collection approach allows us to build a unique dataset that can give us an invaluable insight into innovation dynamics of the green goods firms. As well as providing data to construct variables on the characteristics of the companies such as the importance of green goods in their product portfolio, membership to relevant organisations, awards on green manufacturing, venture capital involvement, investments and government contracts, web-scraping can reveal important information about the spatial links including relationships with universities and other research institutions, partnerships with other firms and any other regional and extra-regional links. We also look into the evolutionary dynamics of these characteristics and relationships thanks to the historic website analysis enabled by the Wayback machine archives.

We summarize our method for identifying firms in the green goods sector in Section 3.1. We explain how we have web-scraped additional data on firms in our data set in Section 3.2 and in Section 3.3 we describe our variables selection process.

3.1 Identifying firms in the Green Goods Sector

While encouraging the green economy has become a priority target for policy makers throughout the world[1], it is also apparent that defining and identifying green industries, companies, products and jobs is not trivial. As a recent ILO (2012) study points out, these categories do not always overlap. For example, energy conserving products may be marketed to consumers as green but may not necessarily be made of materials or processes that are themselves green. Similarly, green jobs in such occupations as waste reduction engineer may not be located in industries or companies designated as green. In Shapira, et al (2013) we have developed a key word search-based method for identifying green goods companies. The Department for Business, Innovation and Skills (BIS) has produced a number of reports that quantify the Low Carbon and Environmental Goods and Services (LCEGS) sector (Innovas in 2009 and BIS in 2010, 2011 and 2012) and the key word search is based on information in classifying the green goods and services sector in these reports. The reports offered a typology

[1] The UK Government has recently announced the Green Deal to create 1000 new apprenticeships and help support 65,000 jobs in the construction sector. Also the Green Investment Bank has been set up in Edinburgh to provide finance to invest in UK projects which are both green and commercial.

of green goods and services that includes 24 categories in three broad areas: environmental, renewable energy, and emerging low carbon. We focused on the terms related to green technology manufacturing firms (hence in our own application did not include terms related to environmental consultancy or carbon finance, although these could be added if needed). We then drew on other sources to enrich the search approach. We added terms in general categories related to green manufacturing technologies and also introduced specific terms included in the Derwent Manual Codes for Green Technology, including biological treatment, electrochemical processes, and batteries. We selected terms that had relevance to green goods sectors and which could be disambiguated by adding Boolean operators such as 'AND' and 'OR' in order to optimize precision and recall.

We identified our relevant panel of green goods of around 300 firms for each of the US and the UK small and mid-size enterprises (SMEs). The search procedure was applied to the FAME database (Bureau van Dijk, 2004, this stores financial data supplied by the firms to Companies House) for UK firms and Dun & Bradstreet Million Dollar database was searched for the US. The underlying research rationale for developing these search terms was to identify SMEs in green goods industries and to examine the factors that cause some of these firms to grow particularly rapidly over time. FAME allows us to search not only for companies by year of incorporation, employment size, or industrial sector but also in the text of fields related to trade description and business lines. We applied the search goods search terms to these fields for companies that were SMEs when incorporated within our target years (1995-2007), then undertook manual reviews (by two separate coders) of the companies so identified using the FAME record and web searches. We collected data on 304 UK firms from FAME and have expanded the dataset by adding in further financial information supplied by Experian.

The search process identified a broad range of enterprises reporting that they were engaged in green goods manufacturing. By broad categories, the leading sectors for the 304 sample UK enterprises were green building technologies (24%), alternative vehicles and fuels (24%), renewables and water treatment (20%), pollution control (8%), and batteries (4%). The underlying data set records enterprises by units of legal business incorporation. Of our UK green goods sample, almost half (46.7%) were independent entities, about one half (49.0%) were subsidiaries of parent companies (enterprises owned by other firms), and the balance (4.3%) comprised holding companies (enterprises that owned other firms). Almost one-third (31.6%) of the enterprises were subsidiaries of non-UK companies. This mix of ownership forms reflects the nature of the UK small and mid-size enterprise sector, where many SMEs are ultimately owned, or have been taken over, by larger parent companies and where there is a relatively high degree of foreign ownership including in the UK small and mid-size manufacturing sector.

By employment size, small enterprises with 50 or fewer employees comprised 38.2% if the UK green goods sample, while 38.5% were enterprises with 51-250 employees. Some 21 enterprises (6.9% of the sample) employed more than 251 employees in 2010. (In our selection process, enterprises were included if they had up to 500 employees between 2002 and 2007, with the possibility that some might have grown beyond this by 2010). There were 50 enterprises (16.4%) with missing data in FAME. These are probably very small firms exempt from returning accounts data, hence are not fully captured in FAME.

We examined the sales and employment growth rates of our enterprise sample by sector. Table 1 shows the change in nominal turnover, we see that not all green sectors grew over the last decade or so. The highest 2004-2011 sales growth rate was seen in the renewables and water treatment sector, with relatively good growth rates in alternative vehicles/fuel and pollution control. However, over this period, the battery sector saw negative growth, particularly for 2007-2011. In Table 1 we also present employment change, the building technologies sector on average lost labour for 2004-2011, with a more dramatic decline over the 2007-11 period covering the recession. The fastest employment growth was in the renewables and water treatment sector. All sectors had smaller employment growth rates for 2007-2011 compared with 2004-2007.

3.2 Web scraping

For firms identified as a green goods company in FAME/ D&B, we have access to the business data collected by this database on that firm. In addition, we developed additional data on company business chronologies and strategies, technologies and products, and relationships as reported on their current and archived pages from their web sites (this approach was piloted on nanotechnology firms in Youtie, et al, 2012 and Arora, et al, 2013).

We developed a set of multiple keywords were used to proxy foundational Triple Helix concepts of university, industry, and government for subsequent use in models that gauge their relationship to firm growth. University connections were measured by references to the following keywords: university, college, institute, academy. Three dimensions of company linkages were adopted: partnership (through terms such as partner, stakeholder*, distributor, collaborat*, alliance, joint venture, agreement, supplier, parts manufacturer), investment/venture capital (through terms such as venture capital, invest, IPO), and links to other firms through membership organizations (through terms such as member, membership, affiliation, association, club, participation). In the case of government, we discovered that keywords did not do a good job of measuring this concept, hence we wound up using a database of government contractors (sam.gov in the US and a list from the Technology Strategy

Board in the UK). We also considered intra-regional and extra-regional geographic terms to understand the role of local versus national versus global connections in firm growth. Local terms were defined as references to cities within the combined statistical area, metropolitan area, or micropolitan area of the firm's headquarters location.

These concepts were operationalized through counts of keywords extracted from information on company websites. Website pages and subsets with complex catalogues or considerable use of PDFs and Flash were not included in the analysis, but otherwise we captured the full content of each company's website. IBM Content Analytics (ICA) was employed as a software application for web extraction, because of its ability to allow the user to define keyword definition sets, structure parsing rules with these definitions, and deploy such parsers to collections of structured sets of firm URLs (those excluded due to heavy use of Flash, etc.) Collections were extracted from current company websites and from past company websites in the 2004-2012 timeframe using the Wayback Machine (archive.org); these archived company websites in the 2004-2008 timeframe are the primary data source for this analysis, although the data from current sites and more recent sites were used to gauge the quality of past website data (in the 2004-2008 period) and perform checks of any missing information. After scraping of current and past collections was conducted and parsing rules were deployed on the data set, ICA delivered data concerning the amount of times the keyword terms within the model appeared. XML files of these data were exported from IBM Content Analytics to a proprietary Java-based program to transform these counts into variables for analysis. In addition, we used the IBM Content Analytics automated parser to count the number of intra-regional and extra-regional geographic references. The pipeline for our web-scraping process is illustrated in Figure 1. Less historical information will be available historically due to broken links during the crawl of the Wayback Archives (this is illustrated in Figure 2 for the data collected on US green firms).

3.3 Variable Selection

A list of the variables selected from financial and web-based sources are given in Table 2 for the US along with the sources noted. Initially we have experimented with the change in employment in 2011 and 2009 weighted by current sales (2011), from D&B, as our dependant variable expressed as follows (see a cross plot of these variables in Figure 3):

$$SWE\_per2 = \left(\left(SALES_{2011}.EMP_{2011}\right) - \left(SALES_{2011}.EMP_{2009}\right)\right) / \left(SALES_{2011}.EMP_{2009}\right) \qquad (1)$$

Dummy variables are included in the model to control for employment size of the firm (small 0-50, medium 51-250 or large >250 (this is based on the last year of the sample); the region that the firm is located in; counts of technology focus (emerging low carbon, environmental, renewables), standardised by number of words/sites, 2004-8; if the firm has received a government grant or not.

Descriptive Statistics of our non-web variables are presented in Table 3. Here we can see that the number of employees in our data set ranges from 1 to 16,410, with sales varying between $2,000 and $4.65 billion. The dependent variable calculated in equation (1) above ranges from -0.95 to 155.5. We experimented with a logistic model and the variable growth2 takes on the value of 1 with swe_per2>0 or 0 otherwise.

Descriptive Statistics of our web variables are presented in Table 5. The minimum value of 0 signifies there were no counts of that work on web-pages for some firms, with the maximum number showing the greatest number of word counts found on web-pages standardised by the number of web pages searched over.

4. Model

The form of the model that we are analysing has the growth rate of sales or employment as the dependent variable with web and non-web variables in the regressor set:

$$\Delta EMP \text{ or } \Delta SALES = f\begin{pmatrix} \text{local links, national links, international links, links} \\ \text{with universities, government, industry, controls} \end{pmatrix}$$

We compare OLS to logistic models where the dependent variable is a binary series that will be 1 if there is a positive change in sales or employment and 0 otherwise. Other financial variables collected in FAME that can be tested for the UK include the firms R&D expenditure; overseas turnover; liquidity; gearing; net tangible assets; shareholders funds. Location variables that will be tested come from web-scraped data. UK variables: Technology Strategy Board funding, patents and publications of firms in our dataset.

Later we will filter our web-scraped variables using principal components analysis similar to Coad and Rao (2011). Here the authors use principle component analysis to draw out an “innovation index” from patent information and R & D expenditure from US firms and focus in on 4 sectors in the high-tech manufacturing industry (SIC 35 Machinery & computer equipment; SIC 36 electric/electronic equipment; SIC 37 transportation equipment; SIC 38 measuring instruments). They use the innovation index within a quantile regression to explain how different sized firms are affected by innovation and how larger firms with more innovation grow faster than smaller firms.

Using our unique data set we test the hypothesis of how embedded a green goods sector firm is to its local region and whether this offers a competitive advantage to the firm in compared with those firms that are not so embedded. The alternative hypothesis is that connections outside the region may be more important (example, born global). To test the extent of regional "embedded-ness" versus connections outside the region, is measured by counting the number of times a firms mentions its surrounding cities on its web-site (the local city is counted X-1 times as this will be listed on the contact us page or any firm literature on the web-site).

5. Results for the US Green Goods Enterprises

We test the validity of our model by comparing OLS to logistic model and then check the residuals for heterogeneity. Our preliminary results for the US green goods enterprises are presented in Table 7. The first two columns of results with the log(swe2c) variables are the OLS models with the growth2 columns showing the logistic model results.

We can see from the OLS results that only a couple of variables are significant, these are "invnvc" (web-count of mentions of financially oriented or venture capital keywords). Also the log of employment is slightly significant. In the logistic models the variables "industryindex" and "grant_2008" along with regional dummies are significant. Initial reflections suggest that financial backing of the firms is important and how well connected they are to their industrial supply chain. Also the stimulation provided by government grants seems to be useful indicator for future growth.

The regional variables point to the enterprise being located in the south of west of the US as being beneficial to growth, firms here may well be more connected to local clusters.

6. Conclusions

The results suggest that local government linkages are positively associated with growth of green goods companies, a finding which emphasizes the importance of demand-led innovation by regional government entities. On the other hand, a substantial role for international linkages with companies and universities outside the local region is noted in the data. This finding suggests that green goods companies must be both local and global in their orientation and linkages with government, universities, and businesses if they are to be successful.

Table 1: Average Sales and Employment Growth Rates by Sector for UK firms

Sector	**Average annual growth rate (percent) over period**					
	Sales			**Employment**		
	2004 -2007	**2007 -2011**	**2004 -2011**	**2004 -2007**	**2007 -2011**	**2004 -2011**
Pollution control	39.7	6.5	11.8	17.9	0.6	3.7
Building technologies	8.9	3.6	4.3	-0.6	-6.8	-4.7
Battery	10.2	-16.4	-17.2	9.7	5.5	1.8
Alternative vehicle/fuel	34.8	-5.2	13.6	11.8	0.6	7.9
Renewables and water treatment	24.5	19.9	23.0	30.8	4.6	12.9
Other green goods	17.3	0.3	7.3	5.1	-2.6	-2.2
Total	21.2	4.7	10.2	10.5	-1.3	2.3

Source: Application of green goods search approach (Shapira, et al 2013) to FAME business database (BvD 2012). 304 UK green goods enterprise sample (see Table 2). Averages over annual nominal turnover and employment growth rates in specified years, from FAME.

Table 2: Description of US Variables and Source

Variables	Source	Measure
Employment: growth2, growth1, Employment_current, employment_2, employment_3, size_micro, size_medium, size_large	D&B	• Current (assume 2011), 2 years prior • 2011 number categorized as dummy
Sales: sales, swe_per2, swe_per1	D&B	Current
Region/Geography: local, national, intl	ICA, Places	City-> CSA/MSA/Micropolitan area thesaurus, then coded into local, non-local national or international
University: university	ICA	Count of terms university, college, institute, academy
Technology Focus: lo_carbon, rnl_energy, environment, general, carbon, renewable, environ	D&B, web scan, ICA	• Nominal: Product description, website scan • Count: terms for low carbon, environmental, renewable
Government: grant, grant2008	sam.gov	Active or Expired after 2004, 2008
Industry—Member: **industryindex**	ICA	Count of member, membership, affiliation, association, club, participation
Industry—Partner: industryindex	ICA	Count of partner, stakeholder*, distributor, collaborat*, alliance, joint venture, agreement, supplier, parts manufacturer
Industry—Investment: invnvc	ICA	• Count of financially oriented keywords • Count of venture funding keywords
Region: **northeast, midwest, south, west**	D&B	Dummies for company location in Census Regions (Northeast, Midwest, South, West)

Note: D&B is Dun and Bradstreet; ICA IBM Analytics web-variables

Table 3: Descriptive Statistics for US non-web Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
swe_per2	235	2.80	12.92	-0.95	155.5
growth2	235	0.37	0.48	0	1
swe_per1	280	3.55	35.89	-0.91	588
growth1	280	0.24	0.43	0	1
Employment_current	298	151.58	932.76	1	15000
employment_2	281	134.31	1,018.46	1	16410
employment_3	235	140.77	1,016.60	1	14848
sales	296	37,800,000	302,000,000	2,000	4,650,000,000
size_micro	299	0.71	0.45	0	1
size_medium	299	0.22	0.41	0	1
size_large	299	0.07	0.26	0	1
grant	299	0.52	0.50	0	1
northeast	299	0.20	0.40	0	1
midwest	299	0.18	0.39	0	1
south	299	0.29	0.45	0	1
west	299	0.32	0.47	0	1
lo_carbon	299	0.31	0.47	0	1
rnl_enrgy	299	0.09	0.29	0	1
environment	299	0.59	0.49	0	1

Table 4: Weight Variable for US firms

growth[t-2]			
	Population	Companies with Websites	Weight (expansion)
Growth	85	74	1.15
Same	132	98	1.35
Decline	17	16	1.06
	234	188	

Table 5: Descriptive Statistics for US web Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
general	235	0.02	0.14	0	1.43
carbon	235	3.01	6.70	0	49.85
renewable	235	0.48	1.61	0	19.50
environ	235	1.79	3.20	0	26.00
products	235	2.81	2.74	0	17.95
demo	235	0.31	0.50	0	3.16
trial	235	0.37	0.58	0	3.64
rndd	235	0.62	1.16	0	11
manufacturing intensity	235	1.62	1.88	0	13.29
custom	235	0.16	0.31	0	1.90
vc	235	0.09	0.37	0	4.58
investments	235	0.27	1.10	0	15.17
currency	235	0.76	1.45	0	10.21
university	235	0.13	0.33	0	2.75
members	235	0.15	0.36	0	4.00
greenness	235	0.45	0.87	0	7.50
awards	235	0.19	0.34	0	3.00
customer	235	0.52	0.66	0	5.33
local	235	0.19	0.35	0	2.87
national	235	1.08	1.62	0	11.67
intl	235	0.86	2.50	0	24.63

Table 6: Pairwise Correlations for US Model Variables

	growth2	lg_swe2c	grant 2008	local	national	intl	industry index	uni	low car	renew	logem	ne	south	west
growth2	1													
lg_swe2c	0.17	1												
grant 2008	0.25	-0.03	1											
local	0.01	0.10	-0.01	1										
national	-0.07	-0.01	0.03	0.20	1									
intl	-0.07	-0.03	-0.04	0.18	0.36	1								
industry index	0.11	0.13	0.09	0.15	0.39	0.10	1							
uni	0.11	0.17	0.14	0.17	0.30	0.03	0.66	1						
low car	0.04	-0.03	0.03	-0.06	-0.07	0.05	0.03	-0.09	1					
renew	0.18	0.22	0.15	0.05	0.08	-0.01	0.19	0.28	-0.21	1				
logem	0.36	0.18	0.09	0.01	0.10	0.09	0.10	0.08	0.09	0.19	1			
ne	-0.05	0.11	0.11	0.05	-0.01	0.08	0.09	0.01	0.01	0.02	-0.02	1		
south	0.05	-0.14	0.00	-0.03	0.06	-0.07	0.01	-0.05	-0.02	-0.13	-0.04	-0.32	1	
west	0.16	0.11	0.01	0.05	-0.12	-0.05	-0.04	0.15	0.00	0.18	0.05	-0.35	-0.45	1

Table 7: OLS & Logistic Model Results for US Green Goods Enterprises

	Log(swe2c)	Log(swe2c)	growth2	growth2
local	0.22	0.21	0.32	0.27
national	-0.04	-0.03	-0.26	-0.28
intl	-0.02	-0.02	-0.08	-0.08
invnvc	0.07**		0.24*	
industryindex		0.02		0.80*
grant_2008	-0.19	-0.19	1.06***	1.08***
university	0.11	0.28	-0.17	-0.31
lo_carbon	-0.08	-0.05	-0.67	-0.69
rnl_enrgy	0.49	0.54	0.72	0.80
logem	0.09*	0.09*	0.66***	0.66***
northeast	0.21	0.23	0.10	0.09
south	-0.21	-0.20	1.24**	1.25**
west	0.13	0.12	1.43**	1.48***
Constant	-0.15	-0.17	-3.65***	-3.74***
Observations	173	173	189	189
R-squared	0.136	0.13	0.24	0.24
Robust standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Note: dependent variables log(swe2c) are OLS regressions and growth2 are logistics models.

Figure 1: The web-scraping process

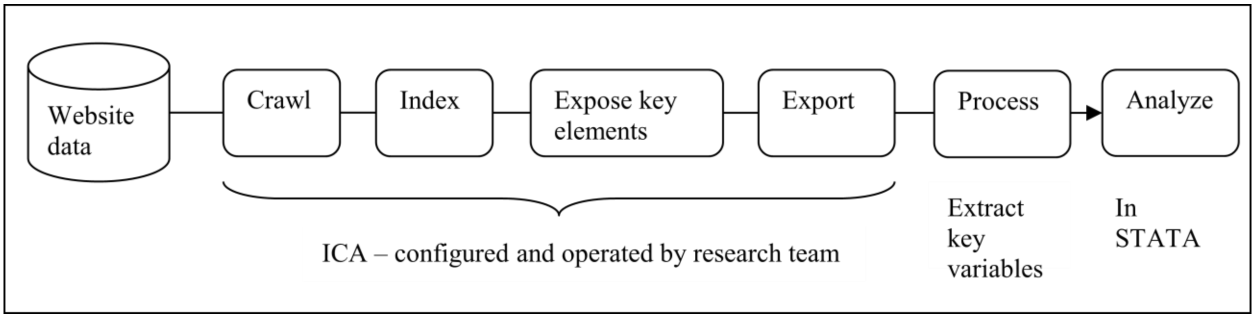


Figure 2: Number of US Companies with words in Wayback by Year

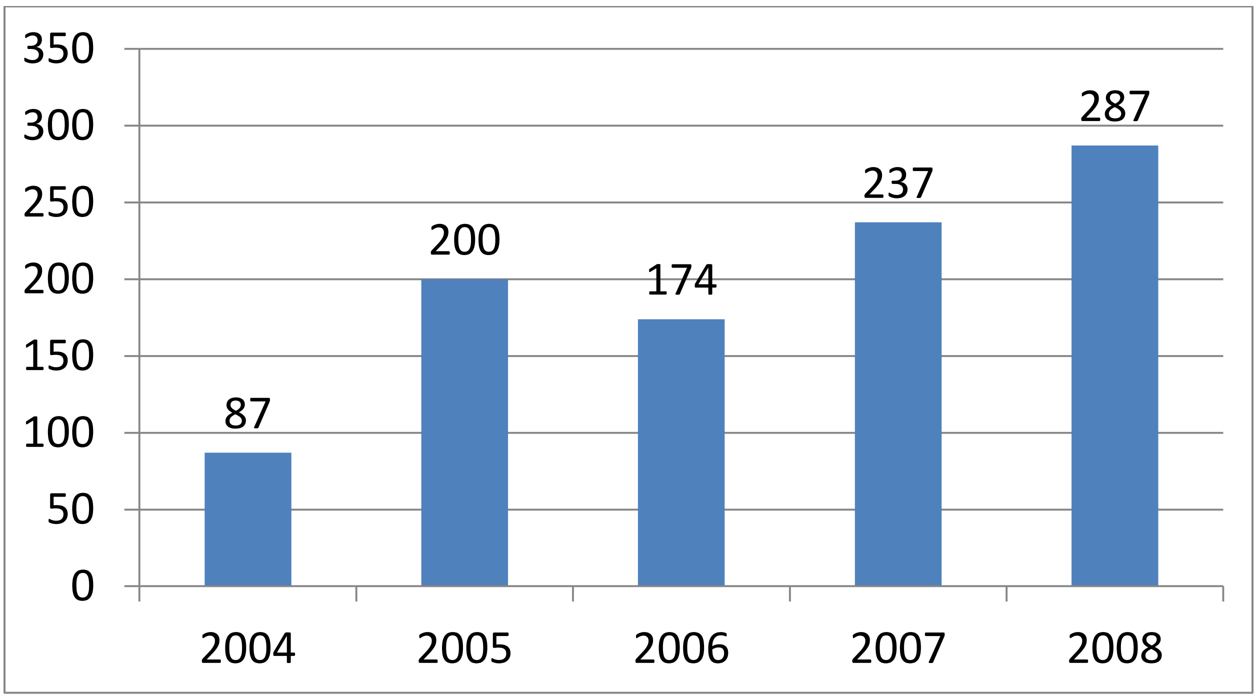
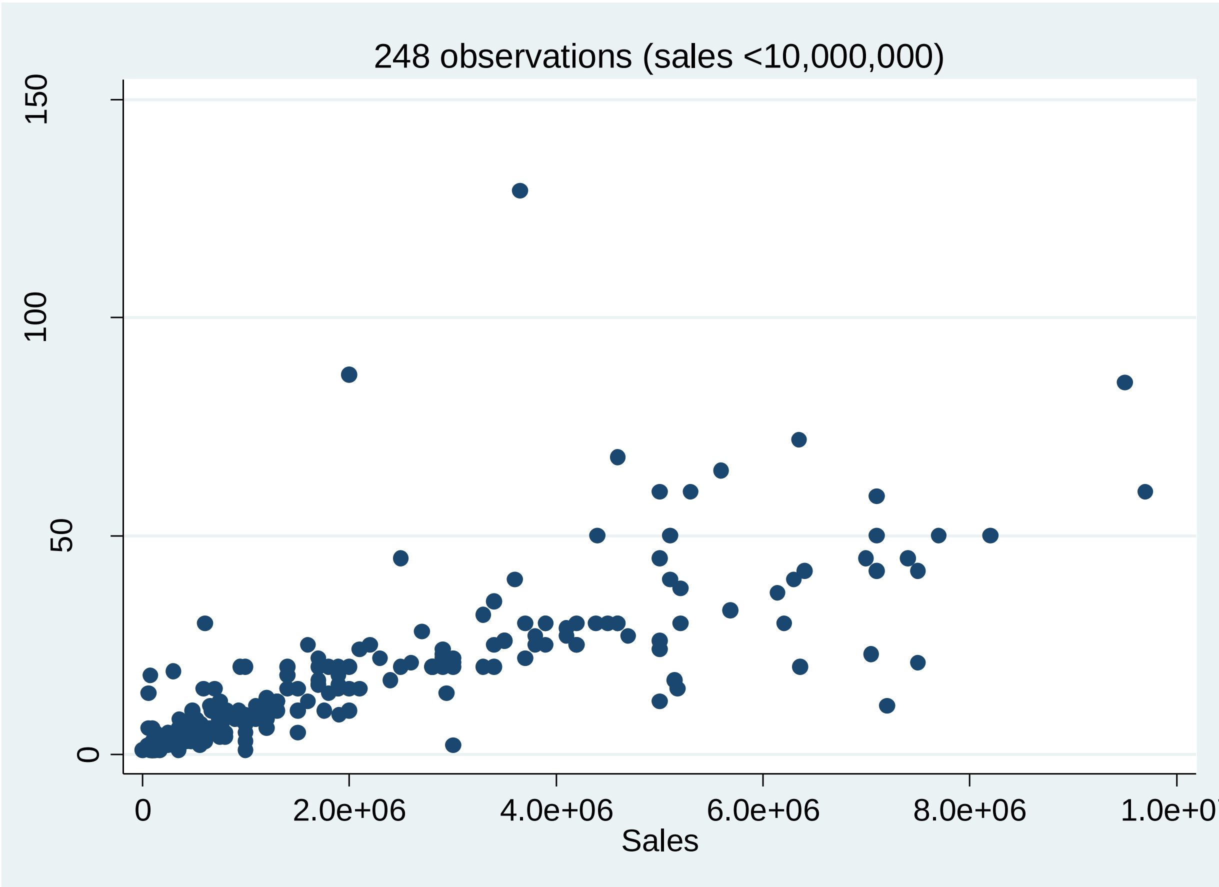


Figure 3: Current Employment by Current Sales, US firms



7. References

Arora S.K., Youtie, J., Shapira, P., Gao, L. and Ma, T.T. (2013). "Entry strategies in an emerging technology: a pilot web-based study of graphene firms", Scientometrics.

BIS (2010) "Low Carbon and Environmental Goods and Services: an industry analysis: Update for 2008/09", accessed on 16th January 2012 from http://www.bis.gov.uk/assets/biscore/business-sectors/docs/10-795-low-carbon-environmental-goods-analysis-update-08-09.pdf.

BIS (2011) "Low Carbon and Environmental Goods and Services: report for 2009/10", accessed on 10th January 2012 from http://bis.ecgroup.net/Publications/BusinessSectors/LowCarbonBusinessOpportunities/11992x.aspx

BIS (2012) "Low Carbon and Environmental Goods and Services: Report for 2010/11", accessed on 15th July 2012 from http://www.bis.gov.uk/assets/biscore/business-sectors/docs/l/12-p143-low-carbon-environmental-goods-and-services-2010-11.pdf.

Bureau van Dijk, (2004). Advanced Internet Quick Guide. Bureau van Dijk Electronic Publishing.

Coad, A. and Rao, R. (2011). "The firm-level employment effects of innovations in high-tech US manufacturing industries", Journal of Evolutionary Economics, vol. 21, pp. 255-283.

Davis, E. (2011). "Made in Britain".

Harrop, A. (2013) "The Great Rebalancing: how to fix the broken economy", The Fabian Society.

Head. K. and Mayer, T. (2004). "Market potential and the location of Japanese investment in the European Union", Review of Economics and Statistics, 86: 959-72.

HM Government (2011) "Skills for a Green Economy: A Report on the Evidence", accessed on 23rd November 2011 from http://www.bis.gov.uk/assets/biscore/further-education-skills/docs/s/11-1315-skills-for-a-green-economy.

HM Treasury (2011). Budget 2011. London: The Stationary Office (March). http://cdn.hm-treasury.gov.uk/2011budget_complete.pdf

ILO (2012). Working Towards Sustainable Development: Opportunities for decent work and social inclusion in a green economy. Geneva: International Labour Organization.

Innovas (2009) "Low Carbon and Environmental Goods and Services: an industry analysis", accessed on 24th November 2011 from http://www.berr.gov.uk/files/file50253.pdf.

Kauffeld-Monz, M. and Fritsch, M. (2013). “Who Are the Knowledge Brokers in Regional Systems of Innovation? A Multi-Actor Network Analysis”, Regional Studies, 47:5, 669-685.

OECD (2011). Towards Green Growth. Paris: Organisation for Economic Cooperation and Development.

PCT (2013). Advantage America: The US-China Clean Energy Technology Trade Relationship in 2011. Philadelphia, PA: The Pew Charitable Trusts.

Sensier, M., Curran, D. and Artis, M.J. (2011). Investigating Agglomeration Economies in a Panel of European Cities and Regions, Spatial Economics Research Centre discussion paper number SERCDP0078.

Shapira, P., Gök, A., Klochikhin, E. and Sensier, M (2013), Probing “Green” Industry Enterprises in the UK: A New Identification Approach”, Manchester Institute of Innovation Research, Sustaining Innovation ESRC Project Working Paper Series, number 2 (http://bit.ly/UIcTKy).

Shearmur, R. (2011) Innovation, regions and proximity: from neo-regionalism to spatial analysis. Regional Studies, vol. 45(9), pp. 1225-43.

Shearmur, R. (2012). "The Geography of Intrametropolitan KIBS Innovation: Distinguishing Agglomeration Economies from Innovation Dynamics"? Urban Studies, 49(11), 2331-56.

Spencer, M. and Arwas, P. (2013). “Nurturing UK cleantech enterprise: four steps to improve low carbon innovation”, The Green Alliance.

Stijn C., Evenett, S., and Hoekman, B. (2011) Rebalancing the Global Economy: A Primer for Policymaking, CEPR: London.

UNEP (2011). Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication. Nairobi: United Nations Environment Programme.

Youtie, J., Hicks, D., Shapira, P., & Horsley, T. (2012). “Pathways from discovery to commercialization: Using web sources to track small and medium-sized enterprises strategies in emerging nanotechnologies”, Technology Analysis and Strategic Management, 24(10), 981–995.