Profiling Green Goods Sector Enterprises in China: An Exploratory Alibaba Web Mining Analysis

Working Paper · January 2015

DOI: 10.13140/RG.2.2.22416.81921

3 authors, including:

Alec Waterworth
The University of Manchester
6 PUBLICATIONS **116** CITATIONS
SEE PROFILE

Philip Shapira
The University of Manchester
259 PUBLICATIONS **5,714** CITATIONS
SEE PROFILE

MANCHESTER 1824

The University of Manchester

Profiling Green Goods Sector Enterprises in China: An Exploratory Alibaba Web Mining Analysis

Working Paper: Project on Sustaining Growth for Innovative New Enterprises

December 2014

Alec Waterworth

Abdullah Gök

Philip Shapira

Manchester Institute of Innovation Research
Manchester Business School, University of Manchester
http://research.mbs.ac.uk/innovation/

Working Papers

Working papers of the project represent pre-publication work in progress and are subject to review and change. Comments are welcome to the corresponding author: alecwaterworth@gmail.com

This work was supported by the Economic and Social Research Council [grant number ES/J008303/1] as part of the project *Sustaining Growth for Innovative Enterprises.*

The project *Sustaining Growth for Innovative Enterprises* probes the growth strategies of innovative small and medium-size enterprises (SMEs). Our research focuses on emerging green goods industries that manufacture outputs which benefit the environment or conserve natural resources, with an international comparative element involving the UK, the US, and China.

The authors appreciate the assistance of Jie Ren (Beijing Institute of Technology) in the identification and selection of the dataset of Chinese firms included in the analysis.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the project sponsors or authors' institutions.

Recommended Citation

Alec Waterworth, Abdullah Gök, and Philip Shapira. *Profiling Green Goods Sector Enterprises in China: An Exploratory Alibaba Web Mining Analysis*. Working Paper, Project on Sustaining Growth for Innovative Enterprises, Manchester Institute of Innovation Research, University of Manchester, Manchester, 2014 (December).

Table of Contents

List of Figures

List of Tables

1 Introduction

This working paper presents findings from the web mining analysis of Chinese green goods firms, conducted as part of the ESRC-sponsored Project on Sustaining Growth for Innovative New Enterprises. The analysis draws on the data collected from the English-language Alibaba (www.alibaba.com) web pages of a sample of 300 Chinese small and medium-size enterprises (SMEs). In this study, SMEs are enterprises with 1000 employees or less, established in 2002 or later, who are involved in the production of manufactured 'green goods'. The focus on enterprises established in 2002 or later is consistent with the project objective of examining the growth trajectories of relatively recent start-up firms.

Green goods industries are identified as those that manufacture outputs which benefit the environment or conserve natural resources (for a detailed discussion of the definition used by the project, see Shapira et al., 2014). Green goods firms in China were identified through a comprehensive set of sector-spanning green goods search terms (Shapira et al., 2014), resulting in an initial list of 588 firms. This list was then refined to eliminate those firms established before 2002 or with more than 1000 employees, or simply those that upon inspection of their websites were not green goods. This resulted in a list of 300 firms for which the Alibaba website addresses were also recorded.

The web mining approach used here was previously used in a similar analysis of 300 UK green goods SMEs and is discussed in Gok et al., 2014. Webpages for the years 2006-2013 were accessed through the Internet Archive Wayback Machine (web.archive.org), which is online archive of historical website content. Whereas the previous UK analysis had focussed on company websites, for the Chinese firms, we chose instead to mine the Alibaba webpages for these companies. This decision was initially made due to a lack of English-language webpages for our sample of 300 firms. However, a unique opportunity quickly arose from this. Whereas a selection of 300 company websites will vary enormously in the content available and how that content is organised and arranged across the website's many pages, Alibaba webpages are in a regularised format, with a set of standardised fields. See Figure 1 for an example.

Following an inspection of the Alibaba webpages for several companies across our timeframe (2006-2014), we were able to anticipate which fields would appear on which webpages (e.g. home page, company profile, etc.), in which years. Therefore, once the web mining process was complete, we were able to isolate these fields and extract the data therein for our sample of companies. It is these fields, hereafter variables, which are the focus for this analysis.

2 Results

2.1 Descriptive Statistics

2.1.1 Number of observations in each year

As shown in Figure 2, the number of firms successfully mined rose steadily from 2006 to 2013, with a significant increase between 2011 to 2012 (from 123 firms to 241), followed by a decrease in 2014 (255 in 2013, down to 213). The increase from 2006 to 2013 can be explained two-fold. Firstly, many of the 300 firms in our sample will not have existed in the earlier years, and of those that did, many more will not have been listed on Alibaba (or indeed, may not have had a web presence at all). Secondly, the coverage of the Wayback Machine is sparser the further back in time one goes. As a result, fewer of these firms' Alibaba webpages will have been captured by the Wayback Machine and therefore were not available for us to mine.

As noted, we found a decrease in the number of Chinese firms in our sample in 2014 as evidenced by their available web pages. These webpages were crawled in Alibaba and not through the Wayback Machine, hence it cannot be an issue of coverage in the Wayback archive. As these firms were initially identified in 2012, it may be that by late 2014, some are no longer operational, or have, for some reason, withdrawn from Alibaba. As there is no further data available on these firms through the web mining process, it is hard to identify the reason for this.

2.1.2 Year Established and Year Joined (2014 data only)

The data for the year the company was established shows a fairly even distribution across the years (2002-12; see Figure 3). Significant growth in green goods industries can be seen around 2004, when significantly more of our sample's firms were established (from 7 in 2002, to 13 in 2003, to 22 in 2004). The sudden decrease in firms established after 2010 is likely just due to the fact that we selected these firms in 2012, therefore making it less likely that a sample firm had been in operation from 2011 or later.

Although Alibaba was founded in 1999, of our sample of 300 firms, none were listed on Alibaba before 2004 (see Figure 4). Two significant increases are seen over the following ten years: firstly in 2007-8 and then again in 2009-10, whereby the frequencies roughly double. It is this latter time period that can be considered the peak for joining Alibaba: almost half of our entire sample joined Alibaba in these two years. This can also be seen in the steady decline in joining Alibaba after this time.

2.1.3 Product Offerings

An examination of the number of products offered in 2014 by the sample firms shows a fairly even spread across the categorised variable (0-499, 500-999, 1000-1499, 1500-1999, 2000-2499, 2500-4999, 5000-10000; see Figure 5). Over half of our sample offer more than 1500 products, with over a quarter offering over 2500, and 12% of firms offering more than 10000. This is an indicator of the growth experienced by a significant section of our sample firms since they were selected in 2012. Reliable means for the product offering variable are only available for years 2011-14 but these certainly confirm a growth (an accelerating growth, in fact) in this variable over the 4 year period:

- 2011: mean of 1803 products per firm (from 122 observations);

- 2012: mean of 2178 products per firm – a 21% increase on the previous year (from 222 observations);
- 2013: mean of 3075 products per firm – a 41% increase on the previous year (from 111 observations);
- 2014: mean of 4915 products per firm – a 60% increase on the previous year (from 209 observations).

Overall, the firms have a rather high number of product offerings. Our review of web pages suggests that in many cases, these product offerings are variations in specifications (e.g. similar products, but specified for different voltages, sizes, or countries).

2.1.4 Product Price Range (2014 data only)

Whilst this is a very broad variable, with only three possible responses, it does offer insight into where most firms price their product offering to the market. As shown in Figure 6, almost half of the firms for whom we have observations define their price range as Average. A further 30% offer High/Average priced products, with the remaining 21% offering Low/Average priced products.

2.1.5 Contract Manufacturing (2014 data only)

The 2014 data for the contract manufacturing services offered by the sample firms shows that 30% do not offer any contract manufacturing services at all. Of those that do, around two-thirds offer Original Equipment Manufacturing, 56% offer Design Service and just under half offer Buyer Label (with many firms offering a combination of these). Refer Figure 7.

2.1.6 Factory Size (2014 data only)

This variable offers similar insights as the product offering variable. There is a fairly even distribution across most of the categories (Under 1,000m^2, 1,000-3,000m^2, 3,000-5,000m^2, 5,000-10,000m^2, 10,000-30,000m^2: which account for 85% of the firms). Only 8% of the sample firms have factory sizes of under 1,000m^2 and roughly the same proportion have production facilities of over 50,000m^2 (see Figure 8).

2.1.7 Annual Revenue (2014 data only)

This financial variable is self-reported by the firm and is categorised broadly by Alibaba. Nevertheless, there is an insight to be gained from examining the data collected (see Figure 9). From 181 observations, two-thirds of the sample firms report annual revenues under US$10m. 15% of firms report annual revenues of over US$100m, which coincides with the number of firms that offer a very large number of products, although, as seen above, this is not reflected to the same extent in the Factory Size variable.

2.1.8 Number of Employees

Whilst our definition of an SME led us to remove firms with more than 1000 employees, as of 2014, only 10 of the 211 firms for which we have observations report an employee base of more than 500 people (see Figure 10). This is somewhat surprising, especially given that two years have passed since these firms were selected as SMEs, and that a much more significant proportion of firms have reported both very large annual revenues and product offerings. When the data is reviewed across the years for which we have a reasonable number of observations for this variable, roughly the same

proportion of firms (between 5 and 9%) are seen to have more than 500 employees each year. It may be that these firms have grown without having to significantly expand their employee base, or that the green goods industries as a whole have generally not experienced growth, although clearly despite a significant increase in the number of products offered.

In 2014, around 60% of firms have 10 or fewer R&D employees, with a further 24% between 11 and 20 (see Figure 11). Around two-thirds of firms have 10 or fewer trade employees, with a further 21% between 11 and 20 (see Figure 12).

2.2 Export Markets

The 2014 data highlights the extent to which the sample firms are focussed on export markets: almost 75% of firms have an export percentage of over 50%. Most striking however is that a third report an export percentage of over 80%. See Figure 13.

Firms also report the markets to which they export and the extent to which they export to that market as a percentage (e.g. Eastern Asia 20%, Western Europe 10%, etc.). When the mean of these percentages is taken across all firms (as shown in Figure 14), it is clear that for Chinese green goods SMEs, Europe and Asia are the pivotal export markets. A mean percentage of around 10% is consistent across North America, Oceania, Africa and South America. A mean of three times that is reported for Europe and Asia.

For historical data, the percentage is not available but the region names are still listed and can therefore be processed as keywords, with a mean for the number of instances each region is listed across all firms and all years. When examined across the twelve year timeframe (divided into three three-year eras), the dominance of these two export markets is consistent throughout this timeframe. See Figure 15. The lesser focus on North America and Asia in the 2006-08 is likely to be an artefact of the data, where much fewer observations were available during this era. The similarity between the export market focus in 2009-11 compared to 20012-14 is quite striking and supports the finding that Europe and Asia have long been the focus for export activity from green goods firms.

When the export percentage variable is reviewed across these three eras, a similar consistency is found across the twelve years (see Figure 16): around 75% of firms report an export percentage of 50% or more. This suggests that whilst the focus regions for exporting activity have been consistent since many of these companies were incepted, the extent to which this industry is export-focussed has also remained equally consistent.

What is most interesting is that there is no significant correlation between the numerous measures of the size of a firm (number of products offered, number of employees, annual revenue and factory size) and how much they export. There is nothing to suggest those who charge more for their products also export more. There is also not a significant correlation between the export percentage and where a company is based: somewhat surprising since, whilst most of the sample companies are based on the coast, many are not. Finally, we find no significant relationship between a company being more innovative (assessed through the extent to which they refer to R&D keywords or by the number of R&D personnel they employ) and exporting more. These relationships were all tested using Pearson's chi-squared tests, the results of which can be found in

2.3 Location

The 2014 data highlights a number of provinces of particular importance to the Chinese green goods sector (see Figure 17). Guangdong is the most popular location of the sample firms, with 25% located there. This is followed by Shandong, Zhejiang and Jiangsu, each with around 15%. These four provinces alone account for over two-thirds of the companies' locations and notably are all located on the coast of mainland China. The proportion of firms based in Shanghai (5%) and Beijing (4%) is certainly lower than may be expected. As these two locations are also on the coast, it cannot be for reasons of exporting, which is an assumed benefit to a firm based in Guangdong, Shandong, Zhejiang or Jiangsu. At least to some extent, it may be that these places are expertise bases for green goods technologies, and as such, companies choose to locate there.

The historical data informs this discussion further. By again splitting the twelve years into three eras, we can see that the regional distribution amongst the firms in 2006-08 was much like it is in 2014, with one exception (see Table 9-Table 11. Shandong – the second most popular location in 2014 – hosts just one firm across these three years. While this may be an artefact of the data given the low coverage of this variable during this era (only 39 observations across three years), it may indicate that Shandong has more recently become a hub for green goods firms in China. By 2009-11, Shandong has risen to fourth most common location. Further, by isolating just the Shandong firms in the 2014 data, we can see from the Year Established variable that 24 of the 33 firms based in Shandong were established in 2007 or later. By contrast, of Guangdong's 52 firms, less than half (24) were established after 2006, and Zhejiang and Jiangsu follow the same pattern, with around two-thirds of their firms established in 2007 or later (17 of 30 and 17 of 29 respectively). The Pearson's chi-squared test for the Location and Year Established shows a perfect correlation. See Table 12.

Other variables have been seen to be influenced by where a company is located. A Pearson's chi-squared test revealed that location does correlate with product offering (see Table 13). Although not a strong correlation, the data does reveal an interesting relationship between the two variables. Most striking here is that whilst Guangdong has a considerable number of firms producing a high number of products (29 of 51 firms are producing over 2500 products), Zhejiang – the third most popular location for green goods SMEs in our firms – have a notably smaller product portfolio (15 of the 30 firms offer fewer than 1000 products, with 10 of these offering less than 500 products).

Similarly, the Pearson's chi-squared test for Location and Factory Size also shows a strong correlation (see Table 14). As with the number of products offered by a firm, Guangdong is the key location for those firms with the largest factory size: of the 14 firms with a factory size over 100,000m^2, 6 are in Guangdong. By contrast, 19 of Zhejiang's 25 firms have factories smaller than 5000m^2. There is, however, no significant correlation between a company's location and their annual revenue or the extent to which they refer to R&D keywords.

In summary, the correlation, or lack thereof, between numerous other variables suggests that a disparity in the size of firms – not as measured in annual revenue but in production capacity and product offering – across the sample depending on location. This suggests that whilst Guangdong may be the more established hub, with firms with more products and larger factories, the other provinces are emerging centres for this industry. The historical data supports this, indicating

development in regions that were not reported on prior to 2012: whilst the era 2009-11 has firms in 14 provinces, the 2012-14 data shows 22 provinces across China with at least one green goods firm.

2.4 Keyword Variables

This analysis used two keyword sets to assess R&D intensity (as outlined in Gök et al., 2014) and greenness. Taking the extracted text from the Alibaba webpages, the keywords were extracted and then normalised by the total number of phrases for each firm's webpages. Both keyword sets went through several iterations and are considered comprehensive. The means of each from the 2014 data suggest that a greater emphasis is placed on the innovativeness of a company and its products than on their greenness. The mean of the normalised greenness keywords is less than half of that of the R&D keywords. Given that the source webpages are largely used for marketing purposes, this suggests that innovativeness is generally regarded as a more important attribute for a company to have than greenness – even in the green goods sector.

In both cases, no significant correlations were found between R&D keywords or greenness keywords and its annual revenue, as well as the aforementioned non-association with export percentage or location. Further, there is no correlation between the use of R&D keywords and the number of R&D personnel a company employs. There is also no significant relationship between these two variables and how many products a company offers. See Table 15Table 17.

What is interesting, however, is that across the years (2006-14), whilst the use of greenness keywords remains fairly consistent, the use of R&D keywords varies tremendously. See Figure 18. The data shows a sharp increase in the use of R&D keywords between 2006 and 2009, peaking in 2010, before seeing a drastic decrease in their use after 2012. This is difficult to explain, especially in light of the consistent use of greenness keywords across the same time frame, but one possibility is that in the years leading up to 2009, the value of R&D keywords for marketing purposes increased and with it their use on a company's website: especially one such as Alibaba which is aimed primarily at a firm's customer base. The value of these terms may have since decreased and so their use has fallen out of favour.

3 Key Findings

From the observations made above, the following key findings are derived:

- There is insight to be gained from in mining Alibaba webpages. Whereas web mining of company websites typically leads to a very large, unstructured dataset, with Alibaba the data is spread across a smaller number of pages for each firm. The data also has a structure to it, with consistent web addresses for all firms (e.g. www.alibaba.companyname.com/ companyprofile.html) and fields therein. This aligns web mining of Alibaba with more traditional innovation studies quantitative methods, such patent analysis and bibliometrics (i.e. semi-structured approaches).
- The green goods sector is highly export-focussed (a third of firms report an export percentage of 80% or more) and consistently has been as far back as our data goes (2006). During this time, Asia and Europe have remained pivotal export markets for the sector. Most striking is that the extent to which a firm exports is not determined by where it is located, the size of the firm (measured by annual revenue, size of product offering, number of employees, size of production facility), the price range at which their products are offered, nor how innovative they are (measured by R&D keywords and the number of R&D employees).
- Guangdong is central to the green goods sector in China, accounting for the location of 25% of the sample firms. This is followed by Shandong, Zhejiang and Jiangsu (around 15% each). The latter two, along with Guangdong, have long held their foremost position, although Shandong has more recently emerged as an important region for the sector. The data shows the sector is becoming more regionally dispersed, with 14 provinces reported in 2009-11 and 22 provinces in 2012-14. Beijing and Shanghai are notable for their small share of the firm locations. Whilst some key provinces are more-developed, and their prominence longer-standing, considerable development is evident in other regions.
- The data shows that green goods firms put a greater emphasis on their innovativeness, as shown through the use of R&D keywords, than their greenness (i.e. the use of greenness keywords). The occurrence of such keywords is not significantly correlated with where a company is located, the size of the firm, or extent to which they export their products. What is most striking in the data is that whilst the use of greenness keywords is fairly consistent across years 2006-14, a rapid increase in the use of R&D keywords can be observed between 2006 and 2009, followed by a sudden decrease in their use from 2012 onwards. One possible explanation is that the value of R&D keywords for marketing purposes has risen and fallen considerably during this relatively short space of time and so too has their use on Alibaba webpages.

4 Appendix A – Figures

Figure 1 Example of Alibaba Webpage Content

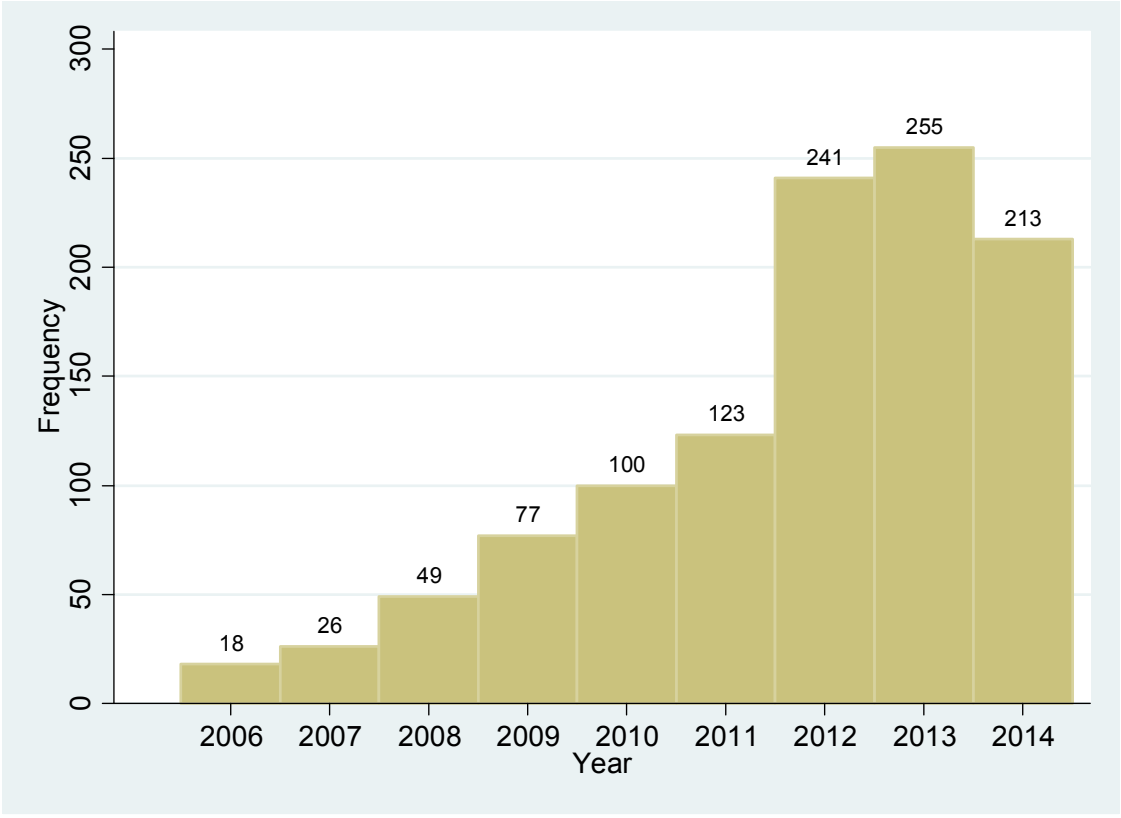

Figure 2 Histogram of Observations across 2006-14

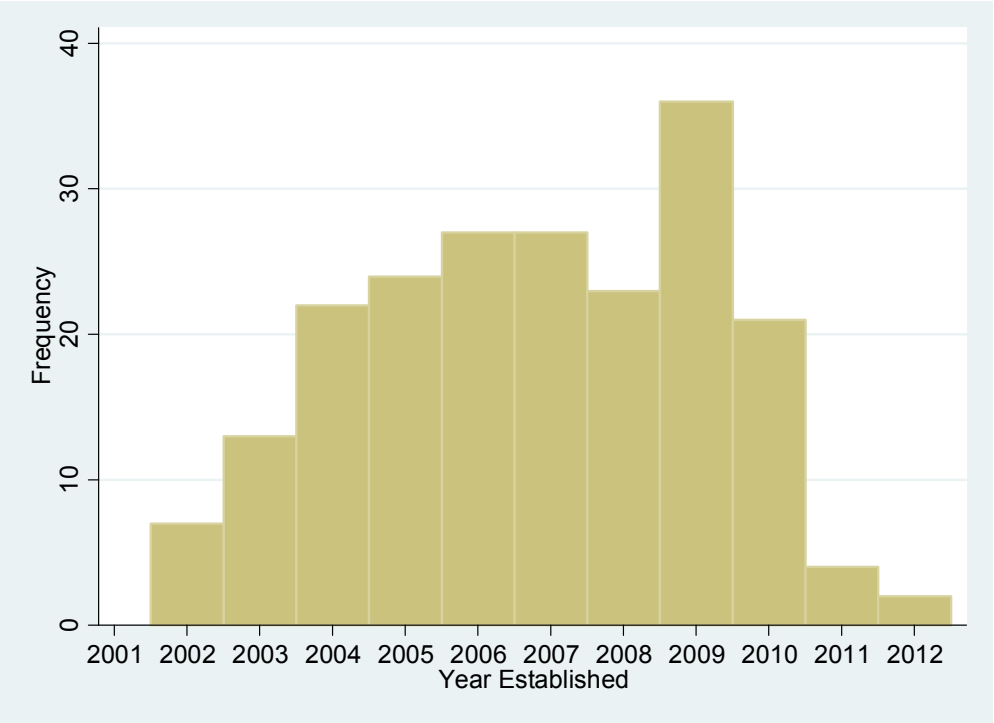

Figure 3 Histogram of frequency of Year Established

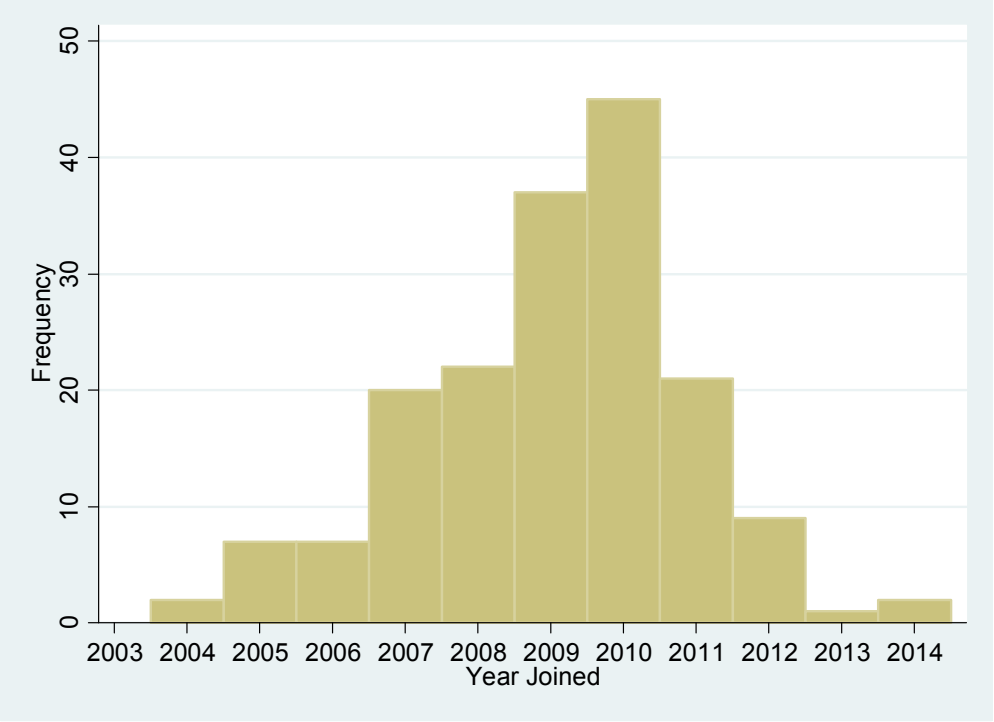

Figure 4 Histogram of frequency of Year Joined

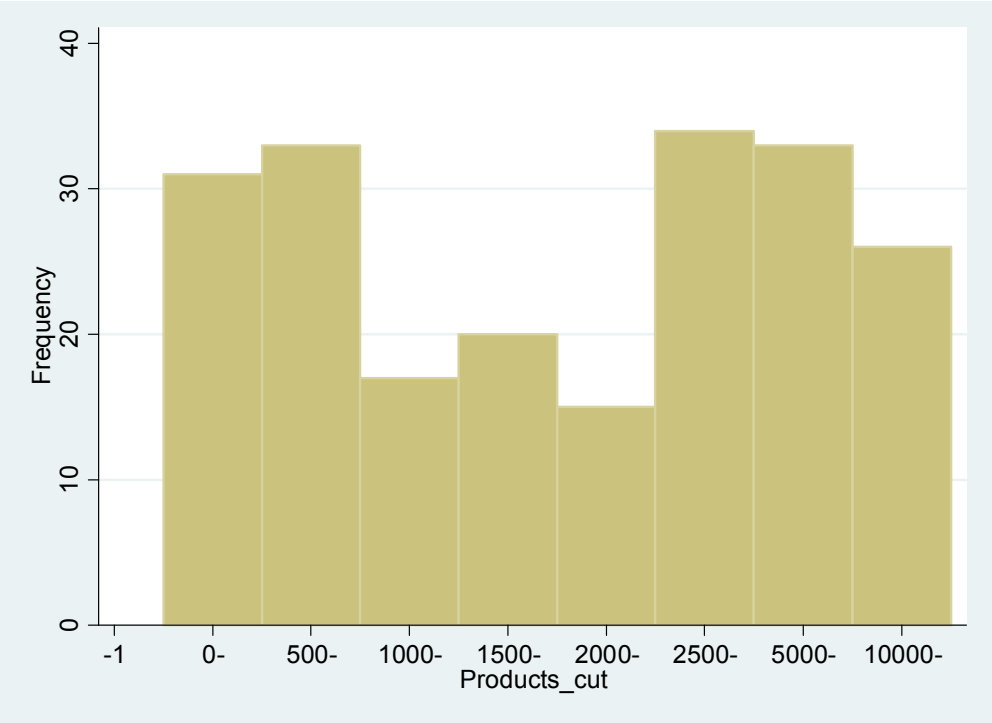

Figure 5 Histogram of Number of Products (categorised)

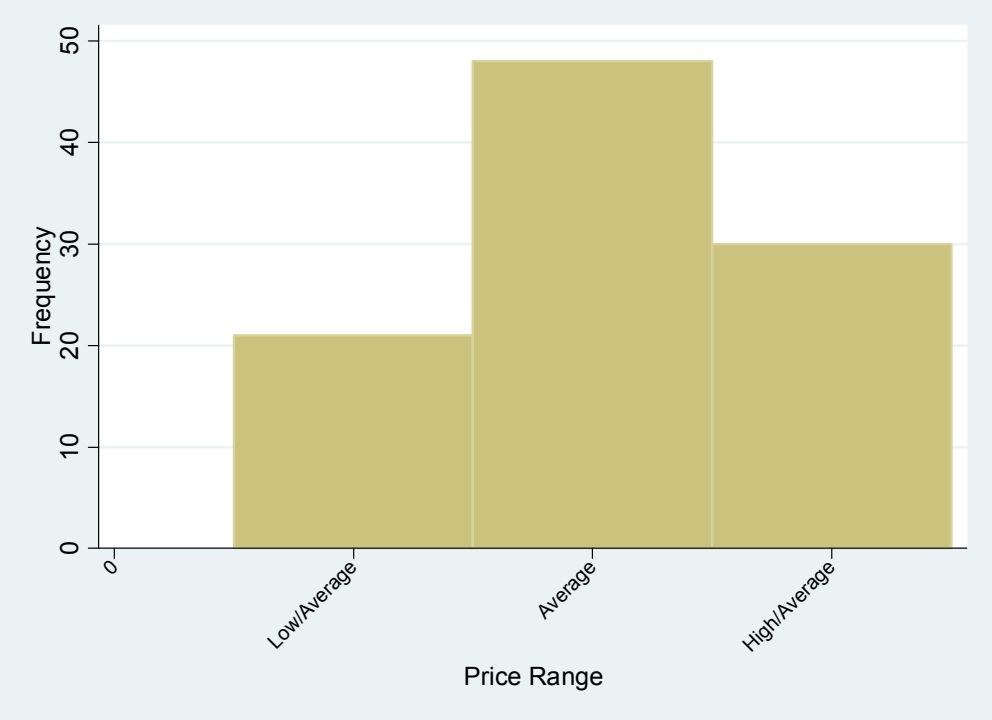

Figure 6 Histogram of Price Range

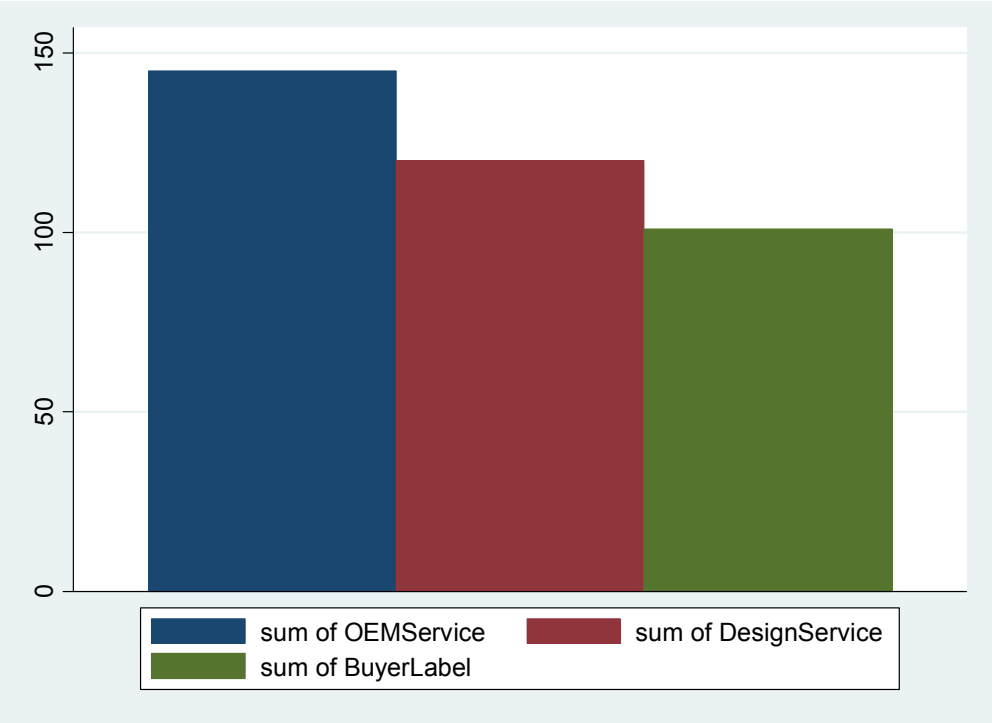

Figure 7 Bar Graph of Contract Services Offered

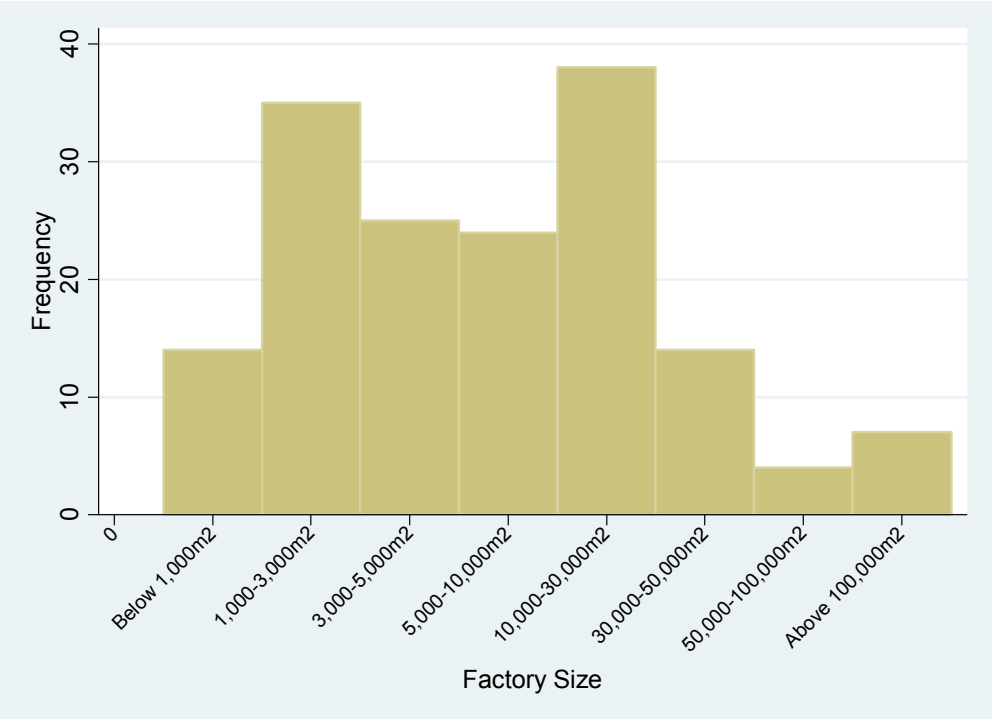

Figure 8 Histogram of Factory Size

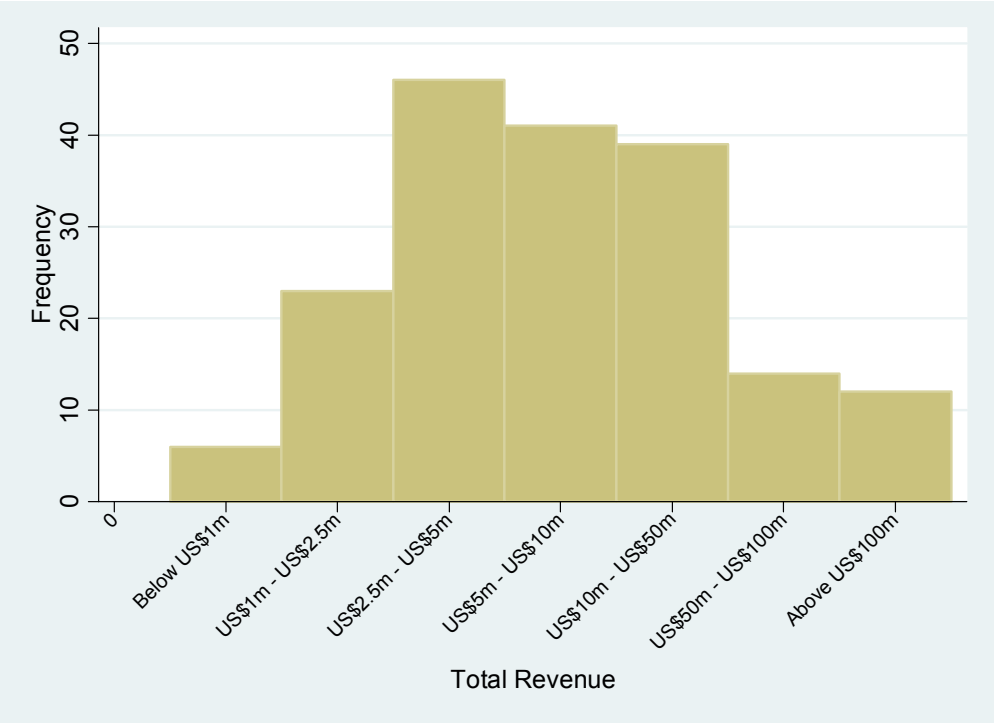

Figure 9 Histogram of Annual Revenue

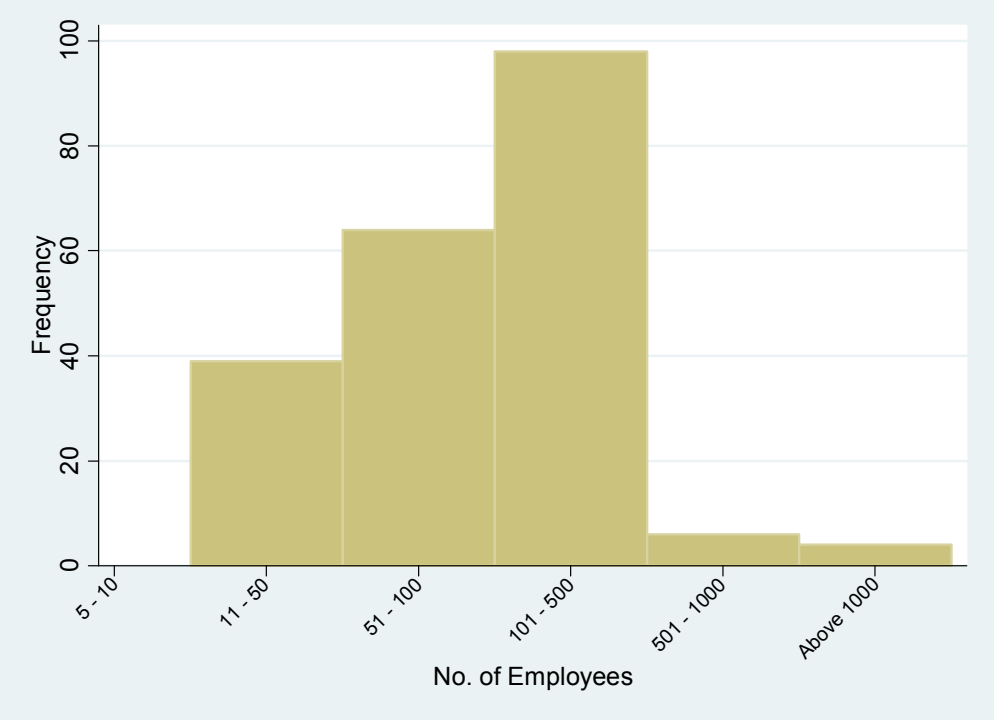

Figure 10 Histogram of Number of Employees

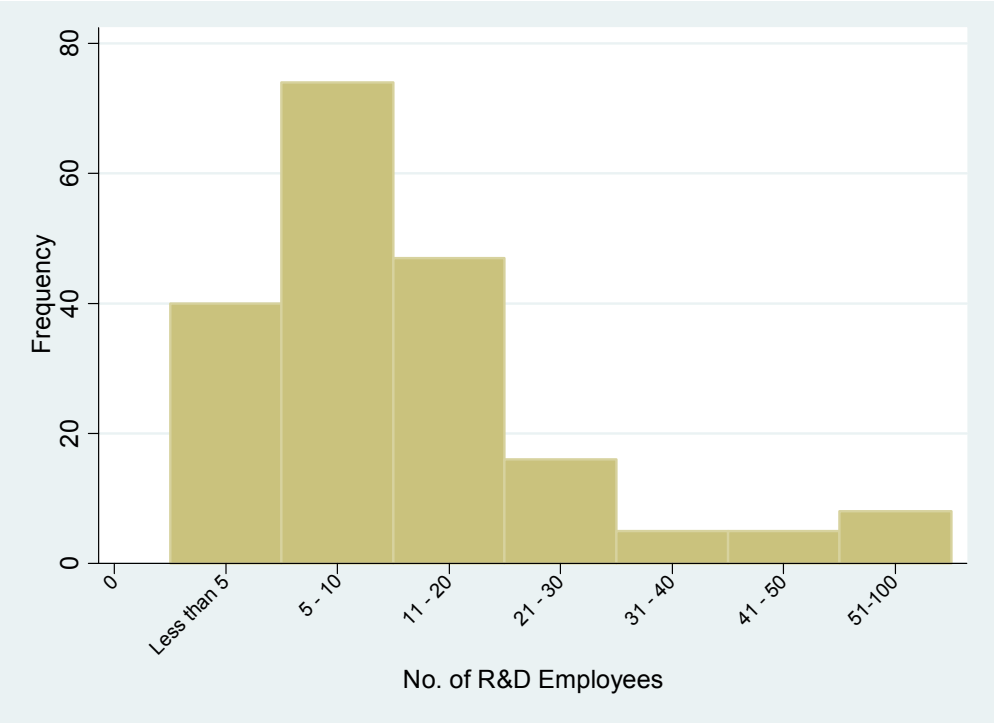

Figure 11 Histogram of Number of R&D Employees

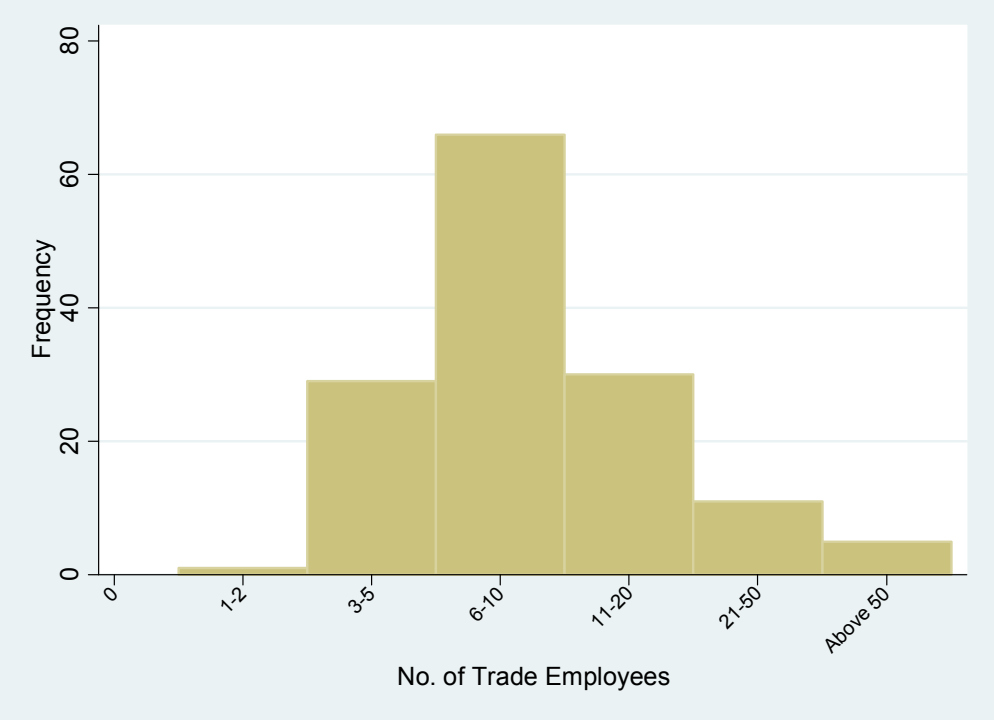

Figure 12 Histogram of Number of Trade Employees

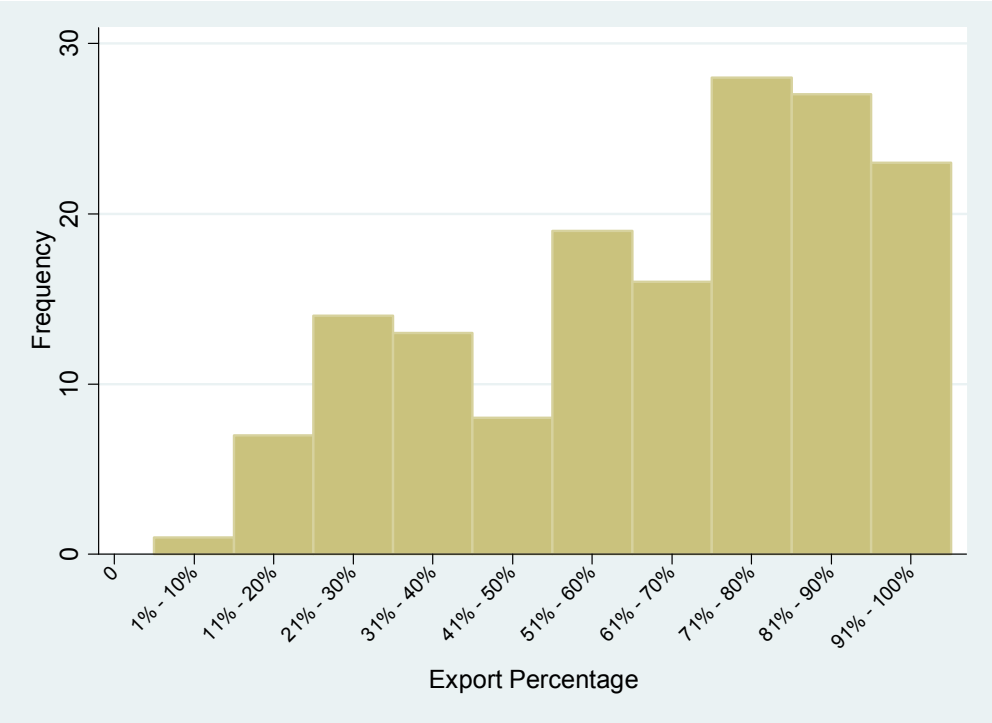

Figure 13 Histogram of Export Percentage

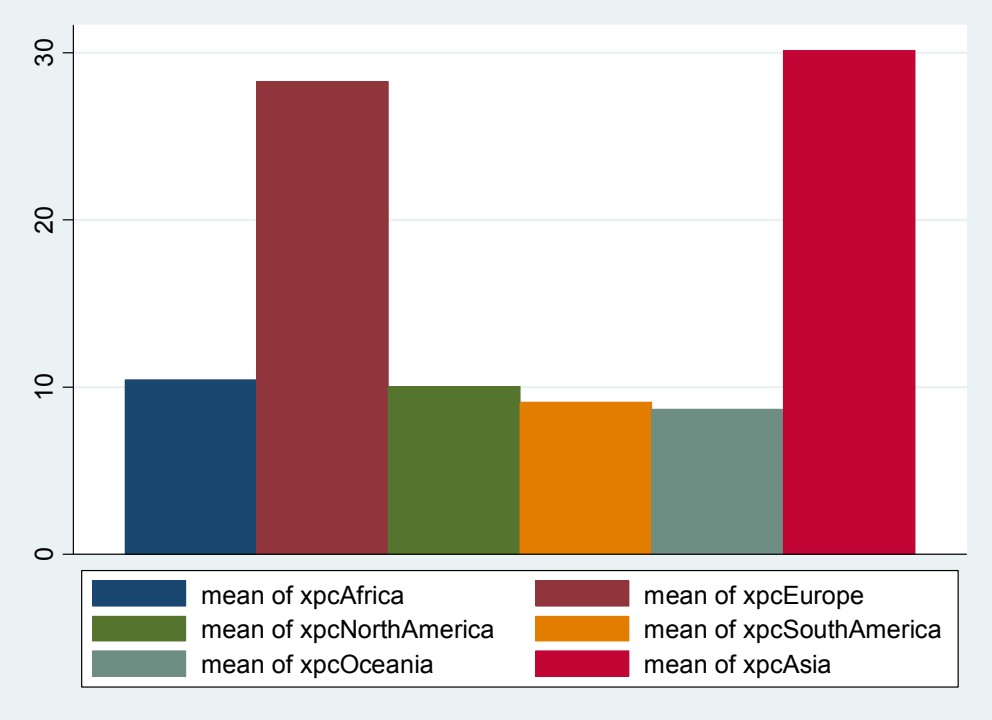

Figure 14 Bar Graph of Mean Percentage in Export Markets

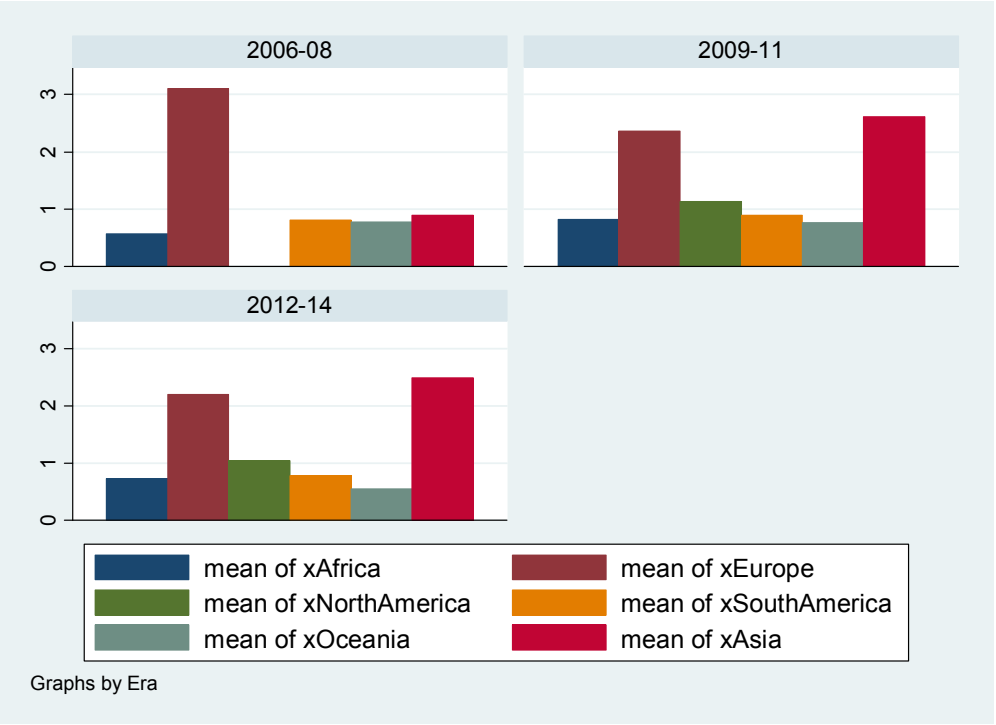

Figure 15 Bar Graph of Mean in Export Markets across Eras

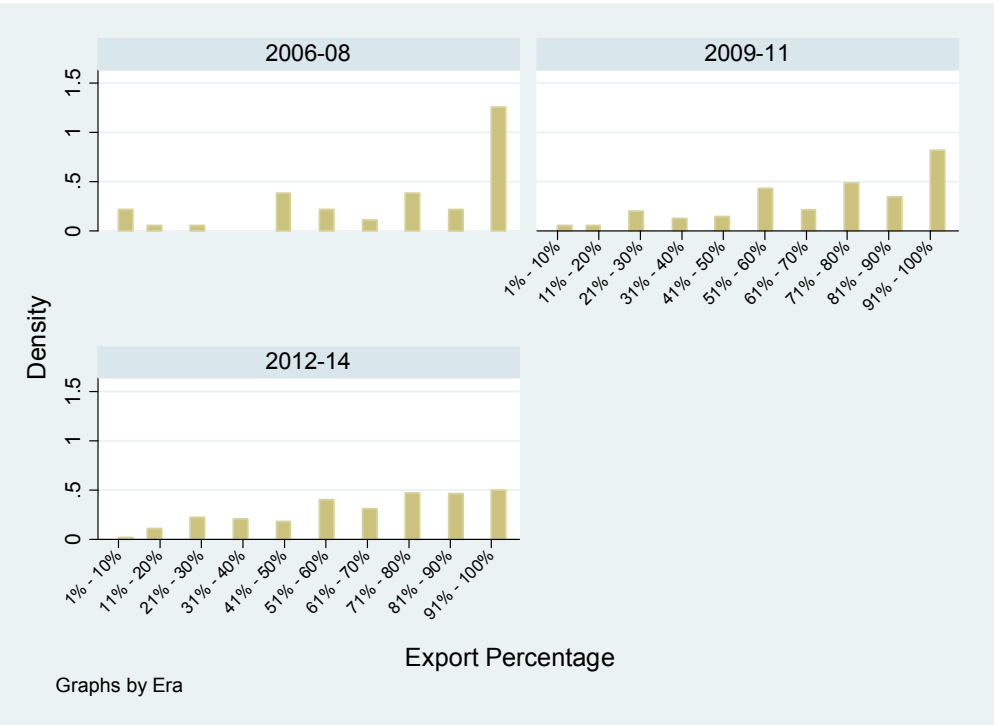

Figure 16 Histogram of Export Percentage across Eras

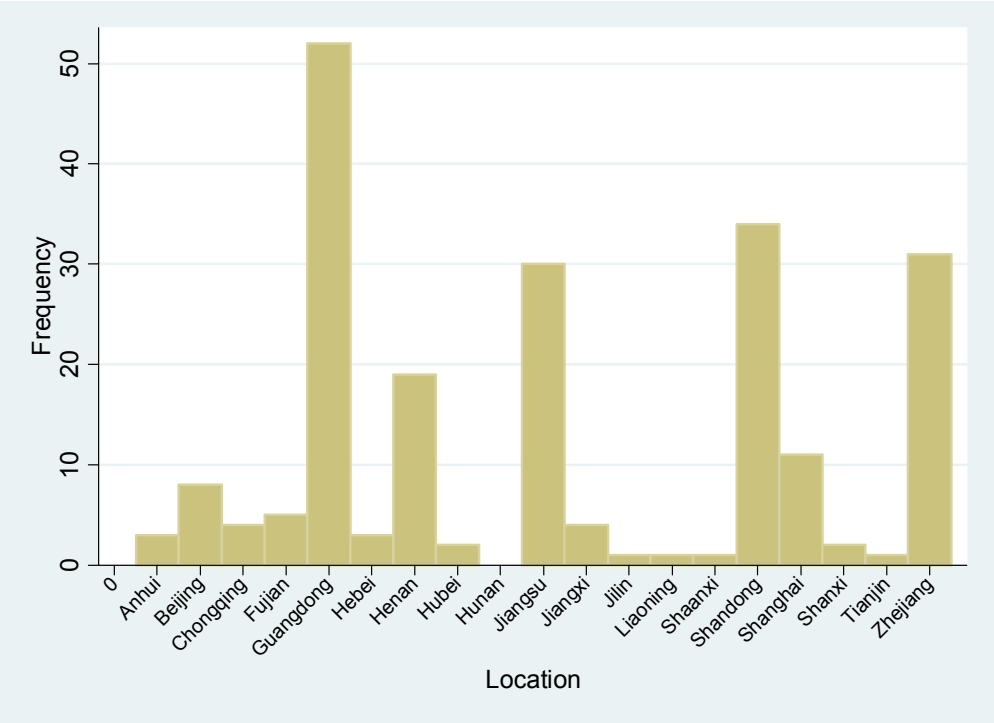

Figure 17 Histogram of Location

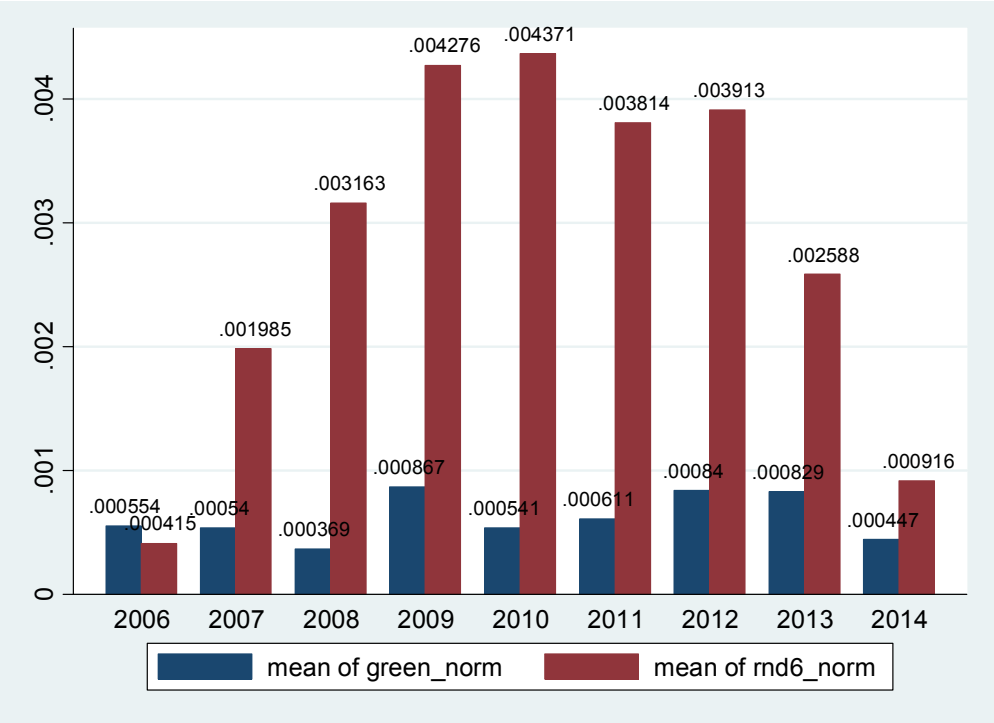

Figure 18 Bar Graph of Mean R&D and Greenness Keywords across Years

5 Appendix B – Tables

Table 1 Pearson's chi-squared test of Export Percentage by Number of Products

Export Percentage	Products_cut 0-	500-	1000-	1500-	2000-	Total
1% - 10%	1	0	0	0	0	1
11% - 20%	3	0	1	0	1	7
21% - 30%	2	3	2	1	1	12
31% - 40%	0	2	0	2	2	13
41% - 50%	0	1	1	2	0	8
51% - 60%	2	2	1	1	1	19
61% - 70%	2	4	1	1	1	16
71% - 80%	3	4	1	2	2	27
81% - 90%	4	3	3	2	3	27
91% - 100%	4	4	1	2	1	23
Total	21	23	11	13	12	153

Export Percentage	Products_cut 2500-	5000-	10000-	Total
1% - 10%	0	0	0	1
11% - 20%	0	1	1	7
21% - 30%	1	1	1	12
31% - 40%	4	2	1	13
41% - 50%	4	0	0	8
51% - 60%	4	4	4	19
61% - 70%	2	2	3	16
71% - 80%	4	4	7	27
81% - 90%	2	6	4	27
91% - 100%	4	6	1	23
Total	25	26	22	153

Pearson chi2(63) = 50.8122 Pr = 0.865

Table 2 Pearson's chi-squared test of Export Percentage by Location

Export Percentage	Location Anhui	Beijing	Chongqin	Fujian	Guangdon	Total
1% - 10%	0	0	0	0	0	1
11% - 20%	0	0	0	0	2	7
21% - 30%	0	1	1	0	3	14
31% - 40%	0	1	0	0	1	13
41% - 50%	0	0	0	0	2	8
51% - 60%	1	1	0	1	7	19
61% - 70%	0	2	0	1	2	16
71% - 80%	0	0	0	0	7	28
81% - 90%	1	0	0	0	11	27
91% - 100%	0	0	0	3	4	23
Total	2	5	1	5	39	156

Export Percentage	Location Hebei	Henan	Hubei	Jiangsu	Jiangxi	Total
1% - 10%	0	0	0	1	0	1
11% - 20%	0	1	0	3	0	7
21% - 30%	0	1	1	3	1	14
31% - 40%	0	5	0	4	0	13
41% - 50%	0	1	0	2	0	8
51% - 60%	1	0	0	0	1	19
61% - 70%	0	2	0	1	1	16
71% - 80%	0	1	0	2	0	28
81% - 90%	0	3	1	2	0	27
91% - 100%	0	2	0	4	0	23
Total	1	16	2	22	3	156

Export Percentage	Location Jilin	Liaoning	Shaanxi	Shandong	Shanghai	Total
1% - 10%	0	0	0	0	0	1
11% - 20%	0	0	0	0	0	7
21% - 30%	0	0	0	2	1	14
31% - 40%	0	0	0	0	0	13
41% - 50%	0	0	1	1	0	8
51% - 60%	1	0	0	3	2	19
61% - 70%	0	0	0	2	1	16
71% - 80%	0	0	0	8	3	28
81% - 90%	0	0	0	2	2	27
91% - 100%	0	1	0	3	0	23
Total	1	1	1	21	9	156

Export Percentage	Location Shanxi	Zhejiang	Total
1% - 10%	0	0	1
11% - 20%	0	1	7
21% - 30%	0	0	14
31% - 40%	0	2	13
41% - 50%	0	1	8
51% - 60%	0	1	19
61% - 70%	0	4	16
71% - 80%	1	6	28
81% - 90%	0	5	27
91% - 100%	1	5	23
Total	2	25	156

Pearson chi2(144) = 149.7745 Pr = 0.354

Table 3 Pearson's chi-squared test of Export Percentage by Number of Employees

Export Percentage	No. of Employees					Total
	11 - 50	51 - 100	101 - 500	501 - 100	Above 100	
1% - 10%	0	0	1	0	0	1
11% - 20%	0	2	4	1	0	7
21% - 30%	1	7	5	1	0	14
31% - 40%	2	3	8	0	0	13
41% - 50%	1	4	3	0	0	8
51% - 60%	3	7	8	1	0	19
61% - 70%	2	6	8	0	0	16
71% - 80%	7	7	14	0	0	28
81% - 90%	6	8	12	1	0	27
91% - 100%	5	3	12	1	2	23
Total	27	47	75	5	2	156

Pearson chi2(36) = 30.1777 Pr = 0.741

Table 4 Pearson's chi-squared test of Export Percentage by Annual Revenue

Export Percentage	Total Revenue					Total
	Below US$	US$1m - U	US$2.5m -	US$5m - U	US$10m -	
1% - 10%	0	1	0	0	0	1
11% - 20%	1	0	2	0	3	6
21% - 30%	0	1	6	0	3	13
31% - 40%	0	1	3	3	3	12
41% - 50%	0	1	1	2	1	6
51% - 60%	1	4	2	3	5	19
61% - 70%	0	1	0	6	7	14
71% - 80%	0	4	8	7	1	24
81% - 90%	1	2	8	4	3	22
91% - 100%	2	2	4	4	6	21
Total	5	17	34	29	32	138

Export Percentage	Total Revenue		Total
	US$50m -	Above US$	
1% - 10%	0	0	1
11% - 20%	0	0	6
21% - 30%	1	2	13
31% - 40%	2	0	12
41% - 50%	1	0	6
51% - 60%	3	1	19
61% - 70%	0	0	14
71% - 80%	4	0	24
81% - 90%	1	3	22
91% - 100%	1	2	21
Total	13	8	138

Pearson chi2(54) = 62.7515 Pr = 0.194

Table 5 Pearson's chi-squared test of Export Percentage by Price Range

Export Percentage	Price Range Low/Avera	Average	High/Aver	Total
1% - 10%	0	1	0	1
11% - 20%	1	2	3	6
21% - 30%	3	0	2	5
31% - 40%	2	6	0	8
41% - 50%	0	3	3	6
51% - 60%	1	4	4	9
61% - 70%	3	3	5	11
71% - 80%	4	9	2	15
81% - 90%	3	11	4	18
91% - 100%	3	4	3	10
Total	20	43	26	89

Pearson chi2(18) = 20.6410 Pr = 0.298

Table 6 Pearson's chi-squared test of Export Percentage by Factory Size

Export Percentage	Factory Size Below 1,0	1,000-3,0	3,000-5,0	5,000-10,	10,000-30	Total
1% - 10%	0	0	0	0	1	1
11% - 20%	0	2	0	1	2	7
21% - 30%	1	2	0	3	3	11
31% - 40%	1	1	4	1	4	13
41% - 50%	0	1	1	1	3	7
51% - 60%	0	2	1	2	3	13
61% - 70%	3	3	2	2	4	16
71% - 80%	1	9	2	3	4	21
81% - 90%	2	6	6	5	4	25
91% - 100%	2	4	4	2	4	16
Total	10	30	20	20	32	130

Export Percentage	Factory Size 30,000-50	50,000-10	Above 100	Total
1% - 10%	0	0	0	1
11% - 20%	1	0	1	7
21% - 30%	0	1	1	11
31% - 40%	1	1	0	13
41% - 50%	0	0	1	7
51% - 60%	5	0	0	13
61% - 70%	1	0	1	16
71% - 80%	2	0	0	21
81% - 90%	1	0	1	25
91% - 100%	0	0	0	16
Total	11	2	5	130

Pearson chi2(63) = 60.5860 Pr = 0.563

Table 7 Pearson's chi-squared test of Export Percentage by Number of R&D Employees

Export Percentage	No. of R&D Employees: Less than	5 - 10	11 - 20	21 - 30	31 - 40	Total
1% - 10%	0	1	0	0	0	1
11% - 20%	1	2	2	1	0	7
21% - 30%	3	3	5	0	0	13
31% - 40%	1	7	3	2	0	13
41% - 50%	1	3	2	1	0	8
51% - 60%	5	6	4	2	0	18
61% - 70%	4	2	5	4	0	16
71% - 80%	7	13	5	1	0	27
81% - 90%	4	10	7	2	2	27
91% - 100%	5	8	6	1	1	22
Total	31	55	39	14	3	152

Export Percentage	No. of R&D Employees: 41 - 50	51-100	Total
1% - 10%	0	0	1
11% - 20%	1	0	7
21% - 30%	2	0	13
31% - 40%	0	0	13
41% - 50%	0	1	8
51% - 60%	1	0	18
61% - 70%	0	1	16
71% - 80%	0	1	27
81% - 90%	0	2	27
91% - 100%	0	1	22
Total	4	6	152

Pearson chi2(54) = 46.3877 Pr = 0.760

Table 8 Pearson's chi-squared test of Export Percentage by R&D Keywords (quartiles)

Export Percentage	rnd_norm_cut4: 0	1	2	3	Total
1% - 10%	0	1	0	0	1
11% - 20%	1	3	3	0	7
21% - 30%	2	10	2	0	14
31% - 40%	0	12	0	1	13
41% - 50%	1	5	2	0	8
51% - 60%	3	13	2	1	19
61% - 70%	2	11	2	1	16
71% - 80%	4	18	6	0	28
81% - 90%	3	19	5	0	27
91% - 100%	2	16	5	0	23
Total	18	108	27	3	156

Pearson chi2(27) = 18.1014 Pr = 0.900

Table 9 Frequency Table of Location in 2006-08

Location	Freq.	Percent	Cum.
Guangdong	15	38.46	38.46
Zhejiang	11	28.21	66.67
Jiangsu	4	10.26	76.92
Beijing	2	5.13	82.05
Chongqing	2	5.13	87.18
Tianjin	2	5.13	92.31
Fujian	1	2.56	94.87
Jiangxi	1	2.56	97.44
Shandong	1	2.56	100.00
Total	39	100.00	

Table 10 Frequency Table of Location in 2009-11

Location	Freq.	Percent	Cum.
Guangdong	40	29.85	29.85
Zhejiang	26	19.40	49.25
Jiangsu	21	15.67	64.93
Shandong	19	14.18	79.10
Beijing	7	5.22	84.33
Chongqing	7	5.22	89.55
Shanghai	3	2.24	91.79
Tianjin	3	2.24	94.03
Fujian	2	1.49	95.52
Liaoning	2	1.49	97.01
Hebei	1	0.75	97.76
Henan	1	0.75	98.51
Hunan	1	0.75	99.25
Taiwan	1	0.75	100.00
Total	134	100.00	

Table 11 Frequency Table of Location in 2012-14

Location	Freq.	Percent	Cum.
Guangdong	132	25.14	25.14
Zhejiang	86	16.38	41.52
Shandong	84	16.00	57.52
Jiangsu	76	14.48	72.00
Henan	41	7.81	79.81
Beijing	18	3.43	83.24
Shanghai	18	3.43	86.67
Fujian	14	2.67	89.33
Chongqing	9	1.71	91.05
Hebei	9	1.71	92.76
Jiangxi	8	1.52	94.29
Anhui	6	1.14	95.43
Hubei	5	0.95	96.38
Shanxi	4	0.76	97.14
Liaoning	3	0.57	97.71
Tianjin	3	0.57	98.29
Hunan	2	0.38	98.67
Jilin	2	0.38	99.05
Shaanxi	2	0.38	99.43
Ningxia	1	0.19	99.62
Taiwan	1	0.19	99.81
Yunnan	1	0.19	100.00
Total	525	100.00	

Table 12 Pearson's chi-squared test of Location by Year Established

Location	Year Established 2002	2003	2004	2005	2006	Total
Anhui	0	0	0	0	1	3
Beijing	0	1	1	1	0	6
Chongqing	0	1	0	0	0	3
Fujian	0	0	0	0	1	5
Guangdong	3	2	3	8	12	52
Hebei	0	0	0	0	1	3
Henan	2	1	1	4	1	19
Hubei	0	0	1	1	0	2
Jiangsu	0	2	6	3	1	29
Jiangxi	1	1	1	0	0	4
Jilin	0	0	0	0	0	1
Liaoning	0	0	0	0	0	1
Shaanxi	0	0	1	0	0	1
Shandong	1	4	3	0	3	33
Shanghai	0	0	2	3	0	11
Shanxi	0	0	0	1	0	2
Tianjin	0	0	0	1	0	1
Zhejiang	0	1	3	2	7	30
Total	7	13	22	24	27	206

Location	Year Established 2007	2008	2009	2010	2011	Total
Anhui	1	0	1	0	0	3
Beijing	1	0	1	1	0	6
Chongqing	0	1	1	0	0	3
Fujian	1	0	2	1	0	5
Guangdong	6	4	9	5	0	52
Hebei	0	1	1	0	0	3
Henan	4	2	1	1	2	19
Hubei	0	0	0	0	0	2
Jiangsu	3	4	4	4	2	29
Jiangxi	0	0	1	0	0	4
Jilin	0	0	0	0	0	1
Liaoning	1	0	0	0	0	1
Shaanxi	0	0	0	0	0	1
Shandong	6	4	8	3	0	33
Shanghai	2	0	4	0	0	11
Shanxi	0	1	0	0	0	2
Tianjin	0	0	0	0	0	1
Zhejiang	2	6	3	6	0	30
Total	27	23	36	21	4	206

Location	Year Established 2012	Total
Anhui	0	3
Beijing	0	6
Chongqing	0	3
Fujian	0	5
Guangdong	0	52
Hebei	0	3
Henan	0	19
Hubei	0	2
Jiangsu	0	29
Jiangxi	0	4
Jilin	1	1
Liaoning	0	1
Shaanxi	0	1
Shandong	1	33
Shanghai	0	11
Shanxi	0	2
Tianjin	0	1
Zhejiang	0	30
Total	2	206

Pearson chi2(170) = 241.7235 Pr = 0.000

Table 13 Pearson's chi-squared test of Location by Number of Products

Location	Products_cut 0-	500-	1000-	1500-	2000-	Total
Anhui	0	0	2	0	0	3
Beijing	1	1	1	1	1	8
Chongqing	2	1	0	0	0	4
Fujian	0	3	0	1	0	5
Guangdong	5	7	2	4	4	51
Hebei	0	1	0	1	1	3
Henan	0	1	2	4	1	19
Hubei	1	0	0	0	0	2
Jiangsu	8	2	4	1	2	30
Jiangxi	0	0	1	1	1	4
Jilin	0	0	0	0	0	1
Liaoning	0	0	0	0	0	1
Shaanxi	0	0	1	0	0	1
Shandong	1	11	1	3	1	33
Shanghai	3	0	1	2	1	11
Shanxi	0	0	0	0	0	2
Tianjin	0	1	0	0	0	1
Zhejiang	10	5	2	2	3	30
Total	31	33	17	20	15	209

Location	Products_cut 2500-	5000-	10000-	Total
Anhui	0	1	0	3
Beijing	1	2	0	8
Chongqing	1	0	0	4
Fujian	0	1	0	5
Guangdong	10	10	9	51
Hebei	0	0	0	3
Henan	3	4	4	19
Hubei	1	0	0	2
Jiangsu	3	6	4	30
Jiangxi	1	0	0	4
Jilin	1	0	0	1
Liaoning	0	1	0	1
Shaanxi	0	0	0	1
Shandong	7	5	4	33
Shanghai	0	1	3	11
Shanxi	1	1	0	2
Tianjin	0	0	0	1
Zhejiang	5	1	2	30
Total	34	33	26	209

Pearson chi2(119) = 134.5627 Pr = 0.156

Table 14 Pearson's chi-squared test of Location by Factory Size

Location	Factory Size: Below 1,0	1,000-3,0	3,000-5,0	5,000-10,	10,000-30	Total
Anhui	0	0	0	0	2	2
Beijing	2	0	2	0	0	4
Chongqing	0	0	0	0	0	2
Fujian	1	1	0	0	0	3
Guangdong	6	13	6	8	4	41
Hebei	0	0	0	0	1	2
Henan	0	1	5	2	4	16
Hubei	0	0	1	0	0	2
Jiangsu	2	3	3	5	8	22
Jiangxi	0	0	0	2	0	3
Liaoning	0	0	0	1	0	1
Shaanxi	0	1	0	0	0	1
Shandong	0	8	4	3	9	29
Shanghai	2	2	0	1	0	7
Shanxi	0	0	0	0	1	1
Zhejiang	1	6	4	2	9	25
Total	14	35	25	24	38	161

Location	Factory Size: 30,000-50	50,000-10	Above 100	Total
Anhui	0	0	0	2
Beijing	0	0	0	4
Chongqing	1	1	0	2
Fujian	1	0	0	3
Guangdong	3	1	0	41
Hebei	1	0	0	2
Henan	3	1	0	16
Hubei	0	0	1	2
Jiangsu	0	0	1	22
Jiangxi	0	0	1	3
Liaoning	0	0	0	1
Shaanxi	0	0	0	1
Shandong	3	1	1	29
Shanghai	1	0	1	7
Shanxi	0	0	0	1
Zhejiang	1	0	2	25
Total	14	4	7	161

Pearson chi2(105) = 137.0310 Pr = 0.020

Table 15 Pearson's chi-squared test of R&D Employees by R&D Keywords (quartiles)

No. of R&D Employees	rnd_norm_cut4: 0	1	2	3	Total
Less than 5	6	29	4	1	40
5 - 10	8	54	11	1	74
11 - 20	7	27	11	2	47
21 - 30	1	11	4	0	16
31 - 40	0	2	3	0	5
41 - 50	0	3	2	0	5
51-100	1	6	1	0	8
Total	23	132	36	4	195

Pearson chi2(18) = 15.4101 Pr = 0.634

Table 16 Pearson's chi-squared test of Number of Products by R&D Keywords (quartiles)

Products_cut	rnd_norm_cut4				
	0	1	2	3	Total
0-	2	20	8	1	31
500-	2	27	3	1	33
1000-	2	12	3	0	17
1500-	4	14	2	0	20
2000-	2	8	5	0	15
2500-	6	22	5	1	34
5000-	9	15	8	1	33
10000-	6	17	3	0	26
Total	33	135	37	4	209

Pearson chi2(21) = 20.9099 Pr = 0.464

Table 17 Pearson's chi-squared test of Number of R&D Employees by Number of Products

No. of R&D Employees	Products_cut					
	0-	500-	1000-	1500-	2000-	Total
Less than 5	2	5	5	4	1	39
5 - 10	9	13	6	7	5	74
11 - 20	8	11	3	4	6	46
21 - 30	4	1	0	1	3	16
31 - 40	0	1	0	0	0	5
41 - 50	3	0	0	0	0	5
51-100	2	1	1	1	0	8
Total	28	32	15	17	15	193

No. of R&D Employees	Products_cut			
	2500-	5000-	10000-	Total
Less than 5	8	9	5	39
5 - 10	12	11	11	74
11 - 20	5	4	5	46
21 - 30	2	3	2	16
31 - 40	1	3	0	5
41 - 50	1	0	1	5
51-100	2	1	0	8
Total	31	31	24	193

Pearson chi2(42) = 42.2312 Pr = 0.461

6 References

Gök, A., Waterworth, A., & Shapira, P. (2014). Use of web mining in studying innovation. Scientometrics. Available online: http://dx.doi.org/10.1007/s11192-014-1434-0.

Shapira, P., Gök, A., Klochikhin, E., & Sensier, M. (2014). Probing ‘‘green’’ industry enterprises in the UK: A new identification approach. Technological Forecasting and Social Change, 85, 93–104.