Watson
Innovation analytics

D2.1 Specifications of SMEs market segmentation based on secondary research

Deliverable due date: March 31, 2018	Actual submission date: May 31, 2018
Start date of project: January 1, 2018	Duration: 18 months
Lead beneficiary for this deliverable: INVENTYA	Revision: 1.1

Nature: R	Dissemination Level: PU
R = Report E = Ethics P = Prototype D = Demonstrator O = Other W = Website, patents, filling, etc.	PU = Public PP = Restricted to other programme participants (including the Commission Services) RE = Restricted to a group specified by the consortium (including the Commission Services) CO = Confidential, only for members of the consortium (including the Commission Services)

This project has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement No. 777439

Contributors	
Name	**Organization**
Martynas Kiskis	Inventya

Peer Reviewers	
Name	**Organization**
Ausra Lingyte	Inventya
Kate Lowes	Inventya
Stephen Bourner	Kapitalise
Ralf Martin	Imperial College London

Revision History		
Version	**Date**	**Modifications**
1.0	March 27, 2018	First draft
1.1	May 30, 2018	Addressed peer review comments

Executive Summary

Identifying different segments of innovation, however important, has been proved to be a very complicated process. Several researchers have attempted to measure innovations in a systematic manner and, while each methodology presents researches with valuable insights, none of them are flawless. Moreover, even though Joseph Schumpeter has outlined the theory of innovation back in the 1930s, up until 2018 there is no universally accepted technique to segment SMEs in terms of their innovative activities. At best, innovation economists employ several methodologies in order to understand what makes some companies more innovative than others.

Perhaps the most popular way of measuring innovations is by reviewing patents. Patents deal with exclusively new ideas - the patent databases are accessible worldwide and are easy to use and contain a detailed description of innovative ideas. However, despite these favourable traits, patent analysis has its flaws.

While patents address inventions, it is often hard to assess the economic value of these inventions. Hence, not every patented idea is worthy to be called an innovation. Yet, there is no way to evaluate the economic value solely by performing a patent analysis. In addition, the patent analysis will likely produce results with a substantial lag of up to 5 years in relation to current market trends.

Nevertheless, no one could argue that patents lead to some form of innovation and patent analysis could be one of the best tools for SMEs innovation segmentation.

Another popular tool for innovation research is the EU Community Innovation Survey (CIS). The survey is conducted in every European Union Member State to collect data on innovation activities in enterprises, i.e. on product innovation (goods or services) and process innovation (organisational and marketing aspects). CIS questionnaires, however, have been evolving each year since the inception of the survey and thus the data can sometimes be difficult to compare. Also, not every member state has reported CIS data throughout the years and several data gaps exist.

Innovation counts is an interesting methodology that the U.S. Small Business Administration (SBA) has come up with in the 1980s. The SBA has manually reviewed multiple trades and scientific journals identifying thousands of innovations across selected industries. While this database is now completely obsolete, it could be possible to reproduce the innovation counts methodology semi-automatically by pulling out data from online databases such as Angel List and Crunch Base where tens thousands of SMEs self-report innovative products and services in order to attract talent and investors.

Finally, firms' growth rate has been proven to be a decent yet imperfect indicator of high-quality innovation. According to NESTA in the UK, truly innovative firms tend to grow twice as fast as traditional businesses (4% vs 2% on average). This should not come as a surprise as one of the best ways to gain a competitive advantage is through improving one's products and/or services. NESTA did a great job compiling its own database and designing the methodology, however the organisation used datasets available solely in the UK and it might be rather difficult to replicate their success in at the wider European level due to reporting disparities between member states.

Other methodologies based on innovation inputs (R&D counts, R&D expenditures) have been also discussed in this document. However, inputs-based segmentations are often unreliable as it is impossible to accurately determine the relationship between R&D capacity and innovation outputs. There are many historical examples of highly funded companies going bankrupt despite having spent millions of Euros into building their R&D capacity.

Table of Contents

1 Methodology & Objectives of the Report

1.1 Research Methodology

Secondary research consists of desk-based research using a variety of sources to provide an initial exploration into the SME innovation segmentation research topic. Sources include specialist publications, official statistics, government reports, industry publications, technical and scientific journals, patent databases and competitor websites in order to:

- Analyse existing sources of information to identify and review the segmentation methodology and key variables that are used to differentiate between innovation driven SMEs and traditional businesses.
- Provide insight how social, economic and environmental trends have influenced the definition of an innovative SME;
- Identify and analyse additional variables that could be used to identify innovative SMEs more efficiently and effectively.

1.2 Objectives

The objective of this report is to review and define existing segmentation methods of SMEs innovation, identify their shortcomings and gaps.

Specifically, the key objectives of this report are to deliver the following:

1. Review of existing segmentation methods used for innovation (e.g. Gazelles and high growth, service portfolio; geography, smart specialisation strategy based) and their short comings;
2. Identification and review of existing datasets, identification of potential gaps to the datasets;
3. Analysis of current variables used in existing datasets;
4. Identification of additional variables required to gather relevant intelligence on innovating SMEs and develop new segmentations;

2 Definition of Innovation

This report will further rely on Joseph Schumpeter's view on innovation and entrepreneurship. Schumpeter's view that entrepreneurship gives birth to innovation has never seemed so appropriate as nowadays, and especially when analysing SMEs' approach to creating new, and often radical, ideas.

In a broad sense, Schumpeter defines innovation as "new combinations [of existing knowledge and inputs]" that entrepreneurs use to create economic growth. Specifically, Joseph Schumpeter argued that innovation may occur due to the following factors:

1. the launch of a new product or a new species of the already known product;
2. application of new methods of production of sales of a product (not yet proven in the industry) – or in other words – *business model innovation*;
3. the opening of a new market for which a branch of the industry was not yet represented;
4. acquiring new sources of supply of raw material of semi-finished goods;
5. new industry structure such as the creation or destruction of a monopoly position.

These five factors can be further used to classify innovations into *product innovations*, *process innovations*, and *product-process innovations* that transform entire value chain.

If the innovation involves new of significantly improved characteristics of a product or a service offered to customers, we will classify it as a **product innovation**. If the innovation involves improved methods, equipment and skills used to perform a service or a process – it shall be classified as a **process innovation**, and finally if the innovation involves significant improvement in both the characteristics of the service and/or product and in the methods, equipment and skills used to deliver the product or the service to market, it should be classified as a **product-process innovation**.

These definitions are also used in the EU Community Innovation Survey questionnaire, which has been the gold standard for innovation studies and is also used in many empirical models.

The Community innovation survey, abbreviated as CIS, is conducted in every European Union (EU) Member State to collect data on innovation activities by enterprises.

Despite the general consensus on the definition of innovation, measurement of innovation is still a strongly debated issue in the economic literature.

3 Review of Existing Segmentation Methods

3.1 Differences Between Innovators and Traditional SMEs

At a very high level, Aulet et al. (2013) from the Kauffman Foundation discussed different types of entrepreneurship, specifically typical Small-Medium Enterprises (SMEs) differ from similarly sized innovation-driven enterprises (IDEs). It is important to mention that Aulet et al. (2013) in their paper also rely on Schumpeter's innovation definition. However, the authors stress in particular that innovation does often and increasingly not involve new technology but rather new business models and approaches.

3.1.1.1 Geographic scope of activities

Apparently, geographic scope of firm's activities could also be a great differentiating factor that helps to distinguish IDEs from traditional businesses. IDEs typically pursue global opportunities based on bringing their innovative products and services to global customers. On the other hand, traditional SMEs that build their business on well-understood business ideas will typically serve local markets.

Not all IDEs will be focused on addressing global markets from the very beginning. Often these companies start out trying to solve a problem in regional or niche market. This approach works as a testbed for a broader deployment of their idea. Hence, if the initial experiment has been proven successful – IDEs will try to scale their services across borders resulting in increased levels of export for the region.

3.1.1.2 Growth pace

A company which is based on some sort of innovativeness (product, process, business model) is likely to grow faster than a similarly sized company that has been incorporated based on a traditional business idea.

In many cases growth and scale are not of concern for traditional SMEs and neither are they easily attainable since these SMEs have little to no competitive advantage in global markets due.

Innovation-driven enterprises, on the other hand, often have no other choice but to scale their operations or fail as they need to compensate for the negative cash flows during extensive R&D periods.

3.1.1.3 External capital

Innovation-driven enterprises are more likely to rely on external capital. External capital is one of the hallmarks of innovation-driven enterprises when compared to traditional SMEs. Typically, IDEs will require different levels of investment at different stages. Some investment will be required to develop the product or service, another round might be used to access global markets at scale.

In 2018 IDEs have access to a very broad range of external capital sources – from traditional venture capital, crowdfunding, angle funds, to initial coin offerings (ICOs) where firms distribute virtual tokens that investors may later redeem for products and services. In fact, according to

CNBC, new types of funding surpassed traditional early-stage VC funding for the first time in 2017[1].

3.1.1.4 Access to the global job market

Many innovation-driven enterprises approach the hiring process very differently from traditional SMEs. While traditional businesses rely on the locally available workforce, IDEs employee more and more people remotely – from marketers to whole remote teams of engineers.

In addition, IDEs are more likely to employee people with a very diverse set of skills and a higher level of education (PhDs, MBAs, MDs). However, as innovation-drive enterprises pass their infancy, they also create employment that requires lower skills such – manufacturing staff, laboratory technicians, customer support specialists and other jobs.

According to Enrico Moretti, innovation-driven enterprises create five auxiliary jobs for every direct IDE job[2]. London Tech City area, for example, has been completely transformed by direct and indirect employment opportunities thanks to a cluster of innovation-driven enterprises.

3.1.1.5 Ownership structure

Innovation-driven enterprises often exhibit a more diverse ownership structure due to a wide availability of external capital providers. While empirical evidence linking ownership structure to innovation output is not definitive, traditional SMEs are more likely to remain family business or lifestyle companies as attracting external capital is much more difficult for these types of companies.

3.1.1.6 Growth rate

Innovative SMEs are a lot more likely to follow the "hockey stick" growth pattern than traditional business. Innovation-driven firms start up by initially *burning* money to develop their ideas and markets. Only then, if innovation is accepted by markets, IDEs transition to a rapid growth rate.

On the other hand, traditional companies typically grow with a more linear growth rate.

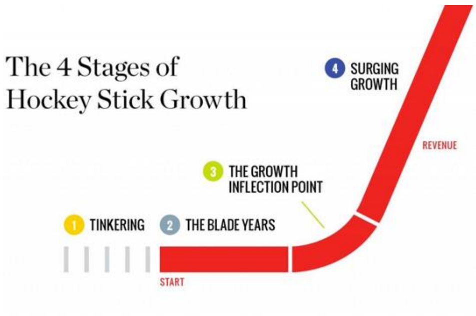

Figure 1 The four stages of the Hockley Stick growth pattern. Source: Forbes (2016).

[1] https://www.cnbc.com/2017/08/09/initial-coin-offerings-surpass-early-stage-venture-capital-funding.html
[2] Enrico Moretti, The New Geography of Jobs (New York: Houghton Mifflin Harcourt, 2012).

Table 1 Summary of IDE vs. SME differentiating factors. Source: Kauffman foundation (2013).

Factor	SMEs	IDEs
Market approach	Focus on local and regional markets.	Focus on global markets.
Competitive advantage origin	Competitive edge is maintained by adjusting capital and/ or labour inputs.	Competitive advantage is based on innovation (technology, process, or business model).
Talent source	Local talent pool. The company relies on local workforce.	Global talent pool. Jobs don't have to be performed locally.
Ownership structure	Traditional ownership structure (e.g. family businesses). Little to no-external capital.	Diverse ownership structure due to a wide choice of external capital providers.
Growth trajectory	Linear growth track. Business responds quickly to increased capital and/ or labour inputs.	Growth rate resembles a "hockey stick". The company grows fast upon completion of a product or a service (i.e. firm must first invest in R&D before any sales can be expected).

3.2 Innovation Inputs

3.2.1 R&D Expenditure

There have been several attempts to measure innovation by assessing how much companies are spending on research and development activities. This can be done either by measuring total R&D expenditure or R&D expenditure as a percent of revenues.

The underlying assumption for the effectiveness of such methodology is simple – larger inputs would produce more innovation. This is a risky assumption since there is plenty of empirical evidence where highly research-intensive firms created unpopular products and/ or services and went out of business. The major limitation of trying to measure innovation through R&D expenditure is caused by ignoring efficiency between inputs and outputs.

In fact, finding direct statistical significance between been R&D expenditure and innovation is nearly impossibly complex. Hall et. all (2009) have reviewed 50 years of economic research around the private rate of return to investing in R&D, however the researchers conclude that the stochastic nature of R&D outcomes means that there is nothing like a single private "rate of return" that is close to a cost of R&D capital.

In spite of the revealed complexity of the problem, Hall et. all (2009) provide evidence that in general, the private returns to investing in R&D are positive and often higher than those for other means of capital. Social R&D returns are strongly positive as well, however variable and often imperfectly measured.

3.2.2 R&D Personnel

R&D capacity especially the number of qualified R&D personnel directly affects firm's capabilities to innovate due to a simple fact that experimental development intensity requires significant effort and commitment from the firm's employees. In this perspective of input-based methodology, Zhang et al. (2009) argue that the ratio of R&D personnel is especially important for small firms' innovation performance[3].

In addition, R&D personnel mobility seems to contribute to innovation performance even more than the firms R&D personnel ration. The explanation for this variable stems from the assumption that labour mobility increases the employer-employee matching quality within the R&D department as well as net inflows of R&D knowledge.

However, there seems to be an optimum churn rate of R&D personnel. If churning of R&D personnel is very high, costs may exceed the benefits and the relationship between churning and innovation is more likely to be inverse u-shaped[4].

3.2.3 Venture Capital

Venture Capital (VC) is defined as "independent, professionally managed, dedicated pools of capital that focus on equity or equity-linked investments in privately held, high growth companies" (Gompers et al, 1999).

While often venture capital is used to accelerate the proliferation of innovative products and services once the product-market fit has been confirmed, it is also uncommon to use VC money to fund research and development activities that over time translate into more innovation for these VC-funded firms.

Hence, venture capitalists invest in technology firms where growth and returns are expected to be significantly higher than other industries (Rossi et al, 2011; Rossi et al, 2013). This information can be effectively used to track which industries should be considered innovative.

[3] http://ieeexplore.ieee.org/document/5344279/?reload=true

[4] http://ftp.zew.de/pub/zew-docs/dp/dp10032.pdf

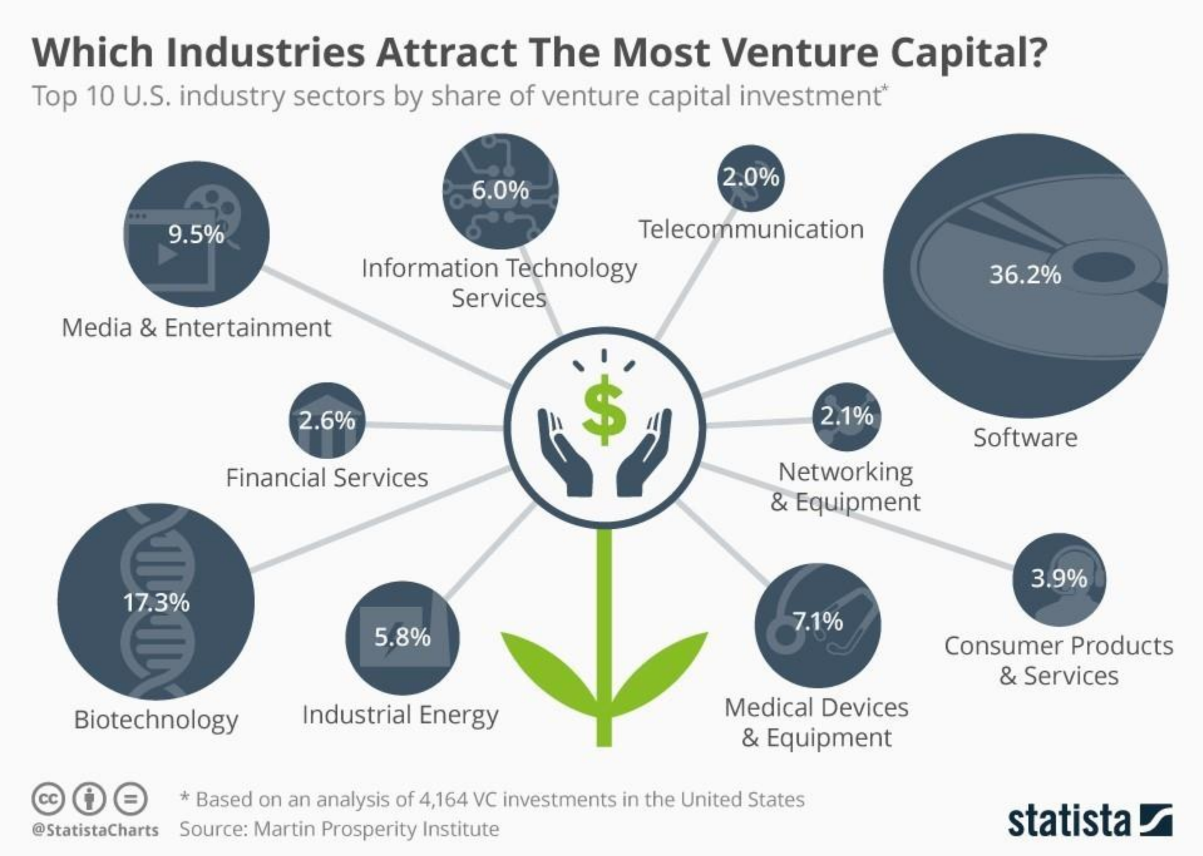

Figure 2 Industries that attract more venture capital than others inevitably end up producing more innovation.

Software investment accounts for 36.2% of US VC funding, while biotechnology comes second with 17.3%. Media and entertainment rounds off the top three with 9.5%.

Other areas of growth and innovation include information technology services, medical devices and equipment and industrial energy. Lower levels of investment are in financial services, consumer products and telecommunications.

3.3 Innovation Outputs

3.3.1 Patents

Due to data accessibility, patent analysis is often researchers' preferred way of measuring innovation. While caveats exist, patent analysis represents one of the best sources of both historical and current technical information that is available due to the following reasons:

1. patents deal exclusively with innovative ideas;
2. patents included a detailed description of the patented invention;
3. patent analysis can give early signals of technological change across industries.

Consequentially, a patent analyst has been an integral part of measuring innovation in numerous studies.

Katila et al. (1999) have proposed a way to distinguish between incremental and radical innovations using a more advanced patent analysis methodology. Evidently, patents involving radical innovations tend to be cited later than patents involving incremental innovations. Consequentially, short citation lags may not be suitable to capture the performance of radical innovations.

While citation-weighted patent analysis provides insight into the innovation quality, it is still difficult to compare the innovation performance among companies or business units solely by patent analysis, one of the reasons being that a significant timeframe for patent citations to accumulate is needed for the analysis to be statistically significant. According to Katila et al. (1999), reviewing patents that are between five and six years old could be optimal. Which from the business perspective is a very long time during which the competitive landscape will likely change completely.

3.3.2 Innovation Counts

The U.S. Small Business Administration compile innovation count data by collecting new product announcements by more than 100 technology, engineering, and trade journals. This turned out to be a very good list focused on new products—only a segment of innovations. However, because it was compiled just once, in 1982, it clearly is outdated now.

Despite the data is now being obsolete, the methodology used to compile innovation counts could work as a reference for the new segmentation methodology. Instead of relying on industry magazines to compile this data, a novel innovation counts methodology could tap automatically tap into databases such as CrunchBase and/or Angel List that both offer data for accessible via open APIs.

Table 2 Innovation counts per industry based on the data collected by the U.S. Small Business Administration. While this data is obsolete, the methodology could inspire how new segmentations could be created.

Sector	Total Innovations	Small-Firm Innovation Share	Small-Firm Employment Share
Food	3.739	0.228	0.220
	(1.977)	(0.052)	(0.175)
Textiles	0.333	0.133	0.325
	(0.062)	(0.161)	(0.156)
Apparel	0.576	0.099	0.565
	(0.605)	(0.159)	(0.220)
Lumber	0.647	0.088	0.564
	(0.338)	(0.228)	(0.264)
Furniture	5.539	0.365	0.596
	(0.422)	(0.074)	(0.264)
Paper	3.588	0.161	0.234
	(0.103)	(0.022)	(0.199)
Printing	1.471	0.191	0.524
	(0.368)	(0.048)	(0.245)
Chemicals	17.929	0.313	0.135
	(0.014)	(0.041)	(0.163)
Petroleum	4.800	0.400	0.113
	(2.400)	(0.200)	(0.163)
Leather	0.546	0.273	0.422
	(0.776)	(0.230)	(0.182)
Stone,	2.185	0.219	0.347
Clay, and Glass	(0.036)	(0.013)	(0.264)
Primary Metals	2.846	0.276	0.227
	(1.631)	(0.043)	(0.167)
Fabricated Metal	9.556	0.388	0.524
Products	(4.301)	(0.038)	(0.192)
Machinery	25.886	0.485	0.289
(nonelectrical)	(1.355)	(0.014)	(0.199)
Electrical	20.946	0.411	0.164
Equipment	(1.824)	(0.008)	(0.155)
Transportation	9.000	0.149	0.075
Equipment	(1.000)	(0.029)	(0.325)
Instruments	56.615	0.506	0.221
	(16.344)	(0.017)	(0.121)

[a]The sector means are weighted averages (by 1977 value-of-shipments) of the four-digit SIC industry values. The small-firm innovation share is defined as the percentage of total innovations contributed by small firms. The small-firm employment share is defined as the percentage of industry employment accounted for by firms with fewer than 500 employees. The data come from the U.S. Small Business Administration, Small Business Data Base, 1982. Standard deviations are listed in parentheses.

3.4 Geographic Segmentation

The underlying assumption for geographic clustering of innovative companies is based on the geographic concentration of innovation inputs, including university R&D, industrial R&D, the presence of related industry and the presence of specialized business services (Feldman, 2014). In addition, there could be other, deeper reasons for clustering that originate from the inherent nature of people – new ideas are easier to realise when people work closer together.

The evidence collected by Feldman (2014) clearly demonstrate that the process of introducing new products to the market is facilitated by the firm's geographic location. Kauffman Foundation reveals that innovative SMEs are not as geographically concentrated as venture capital investors, however they are way more unequally distributed than the general population. However, Kaufmann Foundation's findings contradict with Feldman as they suggest that there is little relationship between high growth companies and traditional innovation inputs as such (incl. research universities, R&D funding schemes). In contrast, Kauffman suggests that the most important input for innovative companies to emerge is the presence of the highly skilled labour force.

Finally, the Kaufmann Foundation study argues that geographic polarization comes in a cycle of twelve to thirty years (the study has been done in the U.S.). And again, more innovation-driven enterprises emerge in areas that have been able to provide better living conditions to highly-skilled professionals. For example, Utah and the District of Columbia in the U.S. have seen a significant increase in high-growth innovation-driven SMEs, despite venture capital being concentrated in Silicon Valley and New York area.

3.5 Growth as an Innovation Indicator

According to Nesta in the UK, high-growth firms are major innovators. The organisation's research points out that innovative firms grow twice as fast, both in employment and sales, as firms that fail to innovate. These findings could be used as a very important indicator in order to come up with a better segmentation for innovation driven enterprises.

However, these findings reflect earlier investment by high-growth companies in innovative resources and in development of R&D capacity. Hence, this methodology to discern innovative SMEs might leave out early-stage startup companies that are relying on lean management models, despite having very innovative ideas.

Nesta supports their claim that growth trajectory is a good indicator of innovation with actual numbers. Innovation-driven SMEs (those that have introduced a novel product or a service during the observation period) grew at an average 10% in sales, compared to 5.8% which was a sales growth average for traditional companies in the UK. This trend translates into employment growth too – innovative SMEs experienced 4.4% growth in employment versus to the 2% average growth displayed by traditional companies. It is important to note that Nesta conducted this research between 2002 and 2004 which was a while ago. Actual numbers will always depend on many more macro and micro economic factors.

Innovation is the underlying force behind rapid growth pace. It provides the innovative firms with enough competitive advantage to outperform their traditional competitors. In addition, this relationship is bidirectional – innovative companies continue to invest into innovation capacity even after the initial period of growth, whereas slow-growing companies do not invest enough or invest into traditional growth creators such as hiring more sales people, buying more of the same ads and other that often will have diminished returns.

Table 3 Relationship between growth rates and firms' innovation statuses. *Source: Nesta (2002–04)*

Firms engaging in:	Product innovation		Process innovation		Wider innovation	
	Yes	No	Yes	No	Yes	No
Average annual growth rate, 2004-07						
Employment growth, 2004-07	4.44	2.02	4.79	2.22	4.15	2.02
	(0.43)	(0.22)	(0.54)	(0.21)	(0.42)	(0.21)
n =	4058	8198	2858	9398	4604	7652
Turnover growth, 2004-07	10.09	5.81	11.36	5.97	10.56	5.22
	(0.81)	(0.47)	(1.02)	(0.44)	(0.86)	(0.42)
n =	4056	8196	2856	9396	4601	7651

Many other factors including differences in industry sizes could have affected the results. Nevertheless, Nesta claims that their methodology took out possible biases including the survivor bias (Nesta's methodology in detail is described in Appendix 1). Hence, it can be said with confidence that innovation drives growth, especially for top performing SMEs and it makes one of the best explanatory variables for future segmentations.

4 Review of Existing Datasets

4.1.1 Community Innovation Survey

The Community Innovation Survey (CIS) based innovation statistics are part of the EU science and technology statistics. Surveys are carried out with two years' frequency by EU member states and number of ESS member countries. Compiling CIS data is voluntary to the countries, which means that in different surveys years different countries are involved.

The CIS is a survey of innovation activity in enterprises. The harmonised survey is designed to provide information on the following aspects:

- innovativeness of sectors by type of enterprises;
- the distinct types of innovation;
- various aspects of the development of an innovation, such as the objectives, the sources of information, the public funding, the innovation expenditures etc.

The CIS provides statistics broken down by countries, type of innovators, economic activities and size classes.

New microdata release normally takes place two and half years after the end of the survey reference period. CIS microdata can be accessed via CD-ROMs (scientific-use files) and in the Safe Centre (SC) at Eurostat's premises in Luxembourg.

Table 4 Community Innovation Survey data will vary each year since every year different countries are involved. *Source: Eurostat (2006-14).*

	CIS 2014		CIS 2012		CIS 2010		CIS 2008		CIS 2006		CIS 4		CIS 3	
	CD	SC	CD	SC	CD	SC	CD	SC	CD	SC	CD	SC	CD	SC
BE		x		x							x		x	x
BG	x	x	x	x	x	x	x	x	x	x	x	x	x	x
CZ	x	x	x	x	x	x	x	x	x	x	x	x	x	x
DK										x		x		x
DE	x	x	x	x	x	x	x	x			x		x	x
EE	x	x	x	x	x	x	x	x	x	x	x	x	x	x
IE						x	x	x						
EL	x	x							x	x	x	x	x	x
ES	x	x	x	x	x	x	x	x	x	x	x	x	x	x
FR		x		x		x		x				x		x
HR	x	x	x	x	x	x								
IT		x		x		x	x	x			x	x		x
CY	x	x	x	x	x	x	x	x	x	x				
LV	x	x		x		x	x	x	x	x	x	x	x	x
LT	x	x	x	x	x	x	x	x	x	x	x	x	x	x
LU		x		x		x		x		x		x		x
HU	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MT								x		x				
NL						x		x						x
AT														
PL														
PT	x	x	x	x	x	x	x	x	x	x	x	x	x	x
RO	x	x	x	x	x	x	x	x	x	x	x	x	x	x
SI			x	x	x	x	x	x	x	x	x	x		x
SK	x	x	x	x	x	x	x	x	x	x	x	x	x	x
FI		x		x		x	x	x		x		x		x
SE		x		x		x		x		x		x		x
UK														
NO	x	x	x	x	x	x	x	x	x	x	x	x	x	x
IS												x	x	x

4.1.2 Patents

In addition to the methodological strength of patent data in innovation studies, the availability of such data further motivates researchers to measure innovative activities this way.

Internet-access to patent data through for example European Patent Office (EPO), the U.S. Patent and Trademark Office, PATSTAT has significantly improved its use in both industrial and academic research (Pavitt, 1988; Walker, 1995). PATSAT specifically could be is particularly important for the WATSON project since this database allows research to know if companies that are registered in the EU have applied for patent protection in other important markets.

Leaving the U.S. Patent and Trademark Office data aside since this project focuses on the European SME segmentation, EPO provides patent data as raw data for companies, commercial providers or individuals who have, or wish to develop, their own data search and analysis tools.

The European Patent Register data backfile provides complete historical data for an EPO patent application at any given time. The European Register data is a snapshot of the database on the date specified, and includes not only the information listed in Rule 143 EPC but also additional data such as legal status, attorney details, etc.

Rule 143

Under the Rule 143 entries in the European Patent Register include the following variables:

1. number of the European patent application;
2. date of filing of the application;
3. title of the invention;
4. classification symbols assigned to the application;
5. the Contracting States designated;
6. particulars of the applicant for or proprietor of the patent;
7. family name, given names and address of the inventor designated by the applicant for or proprietor of the patent;
8. particulars of the representative of the applicant for or proprietor of the patent;
9. priority data (date, State and file number of the previous application);
10. date of publication of the application and, where appropriate, date of the separate publication of the European search report;
11. date of filing of the request for examination;
12. date on which the application is refused, withdrawn or deemed to be withdrawn;
13. date of publication of the mention of the grant of the European patent;
14. date of lapse of the European patent in a Contracting State during the opposition period and, where appropriate, pending a final decision on opposition;
15. date of filing opposition;
16. date and purport of the decision on opposition;
17. dates of stay and resumption of proceedings;
18. dates of interruption and resumption of proceedings;
19. date of re-establishment of rights;
20. the filing of a request for conversion;
21. rights and transfer of such rights relating to an application or a European patent where these Implementing Regulations provide that they shall be recorded;
22. date and purport of the decision on the request for limitation or revocation of the European patent;
23. date and purport of the decision of the Enlarged Board of Appeal on the petition for review.

4.1.3 Innovation Counts

Acs and Audretsch (1988) described in detail how the innovation counts data has collected and processed.

According to Acs et al (1988), innovation counts database has been compiled in the U.S. by the U.S. small business administration and, therefore, not available in the European context. In addition, this data has been compiled in the 1980s and now is absolutely outdate.

Nevertheless, the methodology how this dataset has been created in the first place offers valuable insights.

The data consists 8,074 innovations introduced into the United States in 1982. Of these innovations, 4,476 were identified as occurring in manufacturing industries. The Small Business Administration constructed this data base by examining over 100 technology, engineering, and

trade journals, covering each manufacturing industry. From the sections in each trade journals listing innovations and new products, a data base consisting of the innovations by four-digit SIC industry was formed. The entire list of trade journals used to compile these data is available from the author.
When compiling the list, the U.S. small business administration looked for processes that begin with an invention, processes that involved a development of the invention, and results where new inventions were introduced in a form of a process, a service, or a product.
The data were also checked for duplication. In fact, 8,800 innovations were actually recorded, but it was subsequently found that 726 of them appeared either in separate issues of the same journal or else in different journals. Thus, double counting was avoided.
The innovation data were classified according to the industry of origin based on the SIC code of the innovating enterprise. The data were then classified into innovations by large firms, defined as firms with at least 500 employees, and innovations by small firms, defined as firms with fewer than 500 employees. For example, an innovation made by a subsidiary of a diversified firm would be classified by industry according to the SIC industry of the innovating subsidiary(enterprise) and not by the SIC industry of the parent firm. However, the innovation would be classified by size according to the size of the entire firm and not just by the size of the subsidiary.

Modern approach to innovation counts methodology

Obviously, both the data and the exact methodology is currently obsolete. However, while reviewing magazines and trade journals might be counterproductive in 2018 – alternatives do exist. Many innovative SMEs self-report their innovation activates online through tools like the CrunchBase and the Angel List. This way SMEs seek to attack talented employees as well as raise funding. Moreover, not only their innovations are listed but often external capital, number of employees, location, industry, and in sometimes even investor details. This information can be accessed through open APIs in a potentially automated process. Hence, a modern approach to innovation counts methodology should be considered by the members of the consortium. Especially knowing that both LSE and the Imperial College London that are among the members of the consortium have the expertise of getting innovation indicators via social media data.

4.1.4 SMEs' Growth Data

Unfortunately, there are few universal datasets that keep track of individual SMEs growth rate across the European Union. One of possible candidate databases for such analysis could be ORBIS which contains information on over 275 million companies worldwide, with an emphasis on private company information.

A good example methodology of using SME's growth data as an innovation indicator has been derived by the National Endowment for Science, Technology and the Arts (NESTA) organisation in the UK. Specifically, NESTA has performed an extensive study on a relationship between business growth and innovation in the UK specifically.
The new dataset has been compiled from the 45 UK city regions which account for just under 80% of total UK employment.

NESTA used the following variables in their methodology:

1. Mean firm growth rate (employment-weighted)
2. Employment share of high-growth firms
3. 90th percentile firm growth rate (employment-weighted)
4. New-start firms per 1,000 working age population

This dataset has been built up from multiple sources including the Local Authority District (LAD)-level data derived from a number of sources such as NOMIS, the Annual Survey of Hours and Earnings and the Business Structure Database. Hence, it would be unfeasible to compile a similar database at the European level due to major difference in regional databases, unless databases similar in structure to the Business Structure Database in the UK are identified in each geography.

The Business Structure Database effectively covers all known firms in the UK. It is derived primarily from the Inter-Departmental Business Register (IDBR), which is a live register of data collected by HM Revenue and Customs via VAT and Pays as You Earn (PAYE) records.

The IDBR data are combined with data from ONS business surveys. If a business is liable for VAT (turnover exceeds the VAT threshold) and/or has at least one member of staff registered for the PAYE tax collection system, then the business will appear on the IDBR (and hence in the BSD). In 2004 it was estimated that the businesses listed on the IDBR accounted for almost 99 per cent of economic activity in the UK. Only very small businesses, such as the self-employed were not found on the IDBR[5].

The detailed methodology which explains the relationship between firms' growth and innovation using the variables discussed above is displayed in Appendix 1 on page 20.

[5] https://discover.ukdataservice.ac.uk/catalogue?sn=6697

5 Discussion

Innovations are intangible, take many forms, and are scattered across a variety of industries. Consequentially, innovations are extremely difficult to measure and still there is no general consensus on how to measure innovative activities.

Segmenting innovative SMEs based on innovation outputs seems to work better than relying on analysis based on inputs such as R&D personnel or R&D expenditure. Patent analysis, innovation counts, the Community Innovation Survey in the EU are some of the more popular methods used in innovation studies.

One recent study by NESTA in the UK explored the relationship between innovation and firms' growth. The study results confirm such relationship with high statistical significance, however this methodology might be difficult to test in a greater European context as data required for such study would be difficult to obtain due to many discrepancies amongst the member states. ORBIS database currently seems the best candidate where the WATSON project members could obtain data containing private company information at a large scale.

In addition to patent analysis, and the Community Innovation Survey, *innovation counts* methodology resulted in a valuable segmentation that researchers have been discussing for multiple after it has been compiled. While innovation counts data is currently obsolete, the methodology is worth exploring in a modern context.

Bibliography

Acs Zoltán J. "High-impact Firms: Gazelles Revisited." Global Entrepreneurship, Institutions and Incentives: 542-83.

Acs Zoltán J., David Audretsch B., "Innovation in Large and Small Firms: An Empirical Analysis", *The American Economic Review*, Vol. 78, No. 4 (Sep., 1988), pp. 678-690

Aulet William, and Fiona E. Murray. "A Tale of Two Entrepreneurs: Understanding Differences in the Types of Entrepreneurship in the Economy." *SSRN Electronic Journal*, 2013.

Brychan Thomas, Christopher Miller, and Murphy Lyndon. Innovation and Small Business. 1st ed. Vol. 2.

Catherine Co. "Evolution of the geography of innovation: evidence from patent data." Growth and Change no. 33 (4):393–423. 2002

Feldman Maryann P. "An Examination of the Geography of Innovation." Industrial and Corporate Change 2, no. 1 (1993): 451-70.

Geoff Mason, Kate Bishop and Catherine Robinson. "Business Growth and Innovation The wider impact of rapidly-growing firms in UK city-regions", *NESTA,* 2009.

Griliches Zvi. "Patent Statistics as Economic Indicators: A Survey." 1990.

Henriksen Magnus, and Dan Johansson. "Gazelles as job creators: a survey and interpretation of the evidence." *Small Business Economic*s No. 35 (2):227–244. 2010.

Hall, Bronwyn H. and Mairesse, Jacques and Mohnen, Pierre, Measuring the Returns to R&D (January 1, 2010). CIRANO - Scientific Publications 2010s-02.

Mccann Philip, and Les Oxley. "Innovation, Entrepreneurship, Geography and Growth." 2012

Motoyama Yasuyuki, Danley Brian. "The Ascent of America's high-Growth companies: An Analysis of the Geography of Entrepreneurship". *Ewing Marion Kauffman Foundation*. 2012.

National Science Foundation. "Science and engineering indicators". Washington, D.C.: National Science Foundation. 2011

Schumpeter Joseph. Theory of Economic Development". Leipzig: Duncker and Humblot. 1912

Appendix 1

Empirical models of firm growth and innovation by NESTA (2002-2004)

In order to explore the links between innovation inputs, innovation success and firm growth, we first estimate an equation taking a measure of innovation success as a dependent variable, with measures of firms' investments in the development of innovation-related capabilities in the current year and the previous two years entered as independent variables:

$$(2.1)\ NPS_{it} = \beta_0 + \beta_1 \sum_j INN_{ijt^*} + \beta_2 \sum_k X_{ikt} + \varepsilon_{it}$$

where i denotes firms; NPS_t is the new products share of turnover in year t; INN_{t^*} is a vector of j innovation inputs and capabilities developed in the time period t-2 to t; X_t is a vector of k firm-specific characteristics such as employment size, age, sector, region and geographic market focus; and ε is an error term.

Subsequently, the predicted values of innovation success generated by this regression are entered as independent variables in a second equation which estimates the determinants of firm growth, measured successively as growth in employment and turnover (sales):

$$(2.2)\ FG_{(t+3)it} = \beta_0 + \beta_1 NPS_{it} + \beta_2 \sum_k X_{ikt} + v_{it}$$

where $FG_{(t+3)it}$ is a measure of firm growth between years t and t+3, and v is an error term.

By evaluating the relationship between innovative success in year t and firm growth in a subsequent three-year period, we attempt to control for problems of simultaneity, i.e. two-way interdependence between firm growth and innovation in the same time period. Furthermore, using the predicted value of the dependent variable in the first equation as a regressor in the second equation has the advantage of addressing other concerns about endogeneity (reverse causality) since the predicted values of the innovation success measure are not correlated with the error term in the second equation as might be expected if we used the actual values of this measure.

Equation 2.1 is estimated by Tobit methods which are preferred when the dependent variable is left- and right-censored (as it is in this case, ranging between 0-100). Equation 2.2 is first estimated using Heckman two-step OLS (Ordinary Least Squares) regression methods with firm growth conditioned on a prior selection equation estimating the probability of firm survival from 2004 until 2007. The identifying variables in this selection equation relate to market conditions, namely, the extent to which markets are dominated by established firms and/or are subject to uncertain demand. This produces estimates of the 'average' effect of innovation success on firm growth. Subsequently, we carry out quantile regressions which enable us to explore the relationship between innovation success and firm growth at different points on the firm growth distribution.