Analysis of Estonian startups in an international context

by

Andres Kütt

Master of Business Administration
Estonian Business School, 2006

Submitted to the System Design and Management Program in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2013

©2013 Andres Kütt. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author: Andres Kütt

System Design and Management Program
January 8, 2013

Certified by: Christopher L. Magee

Professor of Practice, Engineering Systems
Thesis Supervisor

Accepted by: Patrick Hale

Director, System Design and Management Program
Analysis of Estonian startups in an international context

by

Andres Kütt

Submitted to the System Design and Management Program on January 8, 2013 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management

Abstract

As more and more countries reach the stage of development where economic growth is rooted in innovation, the question of assessing the quality and development of national innovation systems becomes increasingly more important. In this thesis, the author analyzes the startup ecosystem of a small Northern European country, Estonia. Firstly, a context is laid out in the form of a body of literature relating to innovation systems and their measurement, including bibliometry and co-author graphs. With this background in mind, established metrics as well as an analysis of other available sources of data are used to assess the state of innovation in Estonia. The results are compared to the same metrics for other countries and to predictions made in previous researches. As a result to this analysis, interesting discrepancies emerge between what major research organizations have said about the Estonian innovation economy, the secondary indicators which indirectly reflect the Estonian startup ecosystem and the conclusion of this thesis. Lastly, a modified approach for assessing innovation ecosystems of small open economies is developed and used to make tentative policy recommendations.

Thesis Supervisor: Christopher L. Magee
Title: Professor of the Practice, Engineering Systems
Contents

List of figures 4

List of tables 6

1 Introduction 8
   1.1 Goal and scope 8
   1.2 Methodology and data 9

2 Overview of literature 13

3 Comparison with other countries 17
   3.1 Economic development 18
   3.2 Innovation economy 20
   3.3 Aggregate indexes 26
   3.4 National culture 29
   3.5 Summary of the comparison 32

4 Analysis of the Estonian innovation ecosystem 33
   4.1 Estonian startups in an international context 33
   4.2 Additional quantitative analysis 39
      4.2.1 Co-author graph 39
      4.2.2 Network fragmentation 44
      4.2.3 Subgraphs 48
   4.3 Analysis of the patent graph 49
   4.4 Summary of the graph analysis 60

5 Implications of the research on policy development 61
   5.1 Current goals and policies 61
   5.2 Recommendations based on the current research results 63

6 Conclusion & further study 65

References 67
List of Figures

1. Percentage changes in GDP per capita. Source: International Monetary Fund (2012); Luo et al. (2012)

2. GDP per capita in dollars based on Purchasing-Power-Parity (PPP). Source: International Monetary Fund (2012)


4. USPTO patent grants per million people. Source: USPTO

5. Engineering journal articles per million people. Source: Luo et al. (2012), Compendex

6. Percentage of all patents filed at USPTO. Data for South Korea and Taiwan was not available. Source: WIPO

7. The 12 pillars of competitiveness. Source: Sala-i Martin et al. (2007)

8. Hofstede’s five dimensions for the benchmarked countries. Information was not available for Latvia, the temporal dimension was not available for Estonia and Israel. Source: Hofstede (2012)

9. Entrepreneurship level indicators for Estonia

10. Percentage of all patent applications where the first named inventor is from a particular country. Source: WIPO, authors calculations

11. Relative contribution to the number of Estonian patent applications from major organizations. Source: WIPO, Author’s calculations

12. An example of a co-author network used in the thesis

13. Algorithm for generating the co-author graph of a given country

14. Algorithm for generating a random co-author graph

15. Coefficient $C_1$ for a patent graph with overlap probability $p$ and for an equivalent random graph. Source: author’s calculations

16. Relationships between fragmentation coefficients $C_1$, $C_2$, overlap probability and cluster size. Source: author’s calculations
<table>
<thead>
<tr>
<th></th>
<th>Relationships between fragmentation coefficient $F$ and key graph parameters. Source: author’s calculations</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Relationships between number of components and probability of cluster overlap. Source: author’s calculations</td>
<td>50</td>
</tr>
<tr>
<td>19</td>
<td>Percentage of foreign entities among patent applicants and inventors in Estonia, Latvia and Israel. Source: Author’s calculations</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>Estimated cumulative distribution function for inventors and applicants per patent for Estonia, Latvia and Israel. Source: Author’s calculations</td>
<td>54</td>
</tr>
<tr>
<td>21</td>
<td>Estimated cumulative distribution function of patents for Estonia, Latvia and Israel. Source: Author’s calculations</td>
<td>55</td>
</tr>
<tr>
<td>22</td>
<td>Metrics of the co-author graphs for Estonia, Latvia and Israel. Source: Author’s calculations</td>
<td>56</td>
</tr>
<tr>
<td>23</td>
<td>Distribution functions of the component sizes for Estonia, Latvia and Israel. Source: author’s calculations</td>
<td>58</td>
</tr>
</tbody>
</table>
List of Tables

1 Correlations between GDP growth rates of the countries. Source: Harrell Jr. (2012), International Monetary Fund (2012), Author’s calculations ........... 21
2 Relative rankings of countries based on Knowledge Economy Index and Global Competitiveness Report. GCR data for 1995 is not available, Estonia and Latvia were not covered by GCR in 2000, Taiwan was not covered by GII in 2012. Source: The World Bank (2012), Dutta et al. (2012), Porter et al. (2000) .......... 29
3 Summary of the co-author graphs of the relevant countries 51
4 Largest clusters and their dominant applicants ........ 59
Acknowledgments

This thesis is dedicated to my wife, Maria. Without her, none of this would have become a reality. When there is a "why" every "how" becomes possible.

The author wishes to thank professor Christopher L. Magee for direction and patience and mr. Jianxi Luo for sheer inspiration.
1 Introduction

1.1 Goal and scope

As we find ourselves in the middle of a global financial and economic crisis, being able to rely on human ingenuity, rather than on the tried and tested, has become increasingly more important. This is especially true for small countries that lack significant natural or human resources. Estonia, being such a country, has done relatively well during the crisis and is often depicted as the poster child of the former Soviet Union. Investigating the reality behind this seemingly well-established myth is the underlying motivation for this research.

The goal of this thesis is to either confirm or renounce the following hypotheses:

1. Compared to other countries, the past economic development and policies have resulted in the Estonian startup ecosystem being in generally good shape
2. Theories put forth in academic literature and commonly recognized indexes adequately reflect the state of the Estonian startup ecosystem and innovation economy

The first hypothesis stems from the need to validate the overall impression that Estonia has been remarkably successful in establishing a high-tech culture and fostering innovation (Rooney, 2012; Business Insider, 2012). At the same time, most of the articles on the subject refer to one or two main successes that, although admittedly substantial, do not necessarily reflect the state of the ecosystem as a whole. Hence, it is vital to adequately assess the current state of the country in order to change course, if necessary.

The second hypothesis is related to the first one in the way that the success stories of e.g. Skype and Playtech have been big enough to have distorted the overall understanding of the country’s economic success. Estonia has one of the most open (Miller et al., 2012), yet
smallest economies in the world, which means that the currently recognized theories of innovation may not be applicable in this case. Due to this, for example, the country can be considered incapable of strong R&D investments which, combined with the openness of the economy, may lead to a loss of knowledge resources instead.

These two hypotheses define the scope of the present thesis.

To test the first hypothesis, the author compares a number of countries with Estonia in section 3 and seeks to find explanations for possible discrepancies. The framework of this thesis does not include defining a new set of metrics, as it chooses to focus on generalizing and aggregating the metrics instead.

In order to test the second hypothesis, the author analyzes some aspects of the Estonian innovation economy and startup ecosystem in more detail, and applies established models of economic development and innovation to those aspects. The main focus is on applicants and inventors mentioned on patent applications. This work is not intended to be comprehensive in nature, but rather reflective on the most recognized and applicable theories in the field as well as analytical on a relatively narrow aspect of the subject at hand. Should the existing methods for describing the research subject exhibit deficiencies, the thesis plans to include definitions of new metrics and basic validations of these. Application of these new metrics to a wider setting, however, as well as the development of a thorough theoretical background, lies outside the scope and purpose of this thesis.

### 1.2 Methodology and data

For the data related to the Estonian economy and population, information from the World Bank and IMF was used, with the dataset extended and supplemented with information from the Estonian Board of Statistics. Where applicable, data gathered and kindly made available by Luo et al. (2012) was used to maintain the comparability of the research.
As previously stated, social networks were chosen as means of studying the Estonian startup community. The purpose for this lies in the fact that the goal of this thesis is to assess the dynamic capabilities of the Estonian ecosystem in general, as opposed to studying the behavior of individual organizations or output variables of the entire ecosystem.

Traditionally, investigations of social networks have been carried out through field studies (Newman, 2001). Although GEM\textsuperscript{1} and INSEAD have executed field studies with a global scope and ambition (Bosma et al., 2012; Dutta et al., 2012), their data has focused on providing specific measures of innovation and entrepreneurship rather than information on the social structures of a startup ecosystem. These studies have also been subject to issues related to limited statistical accuracies and the errors caused by the subjective nature of both the questions and answers used in the researches, as has been pointed out by Newman. While at least some of the social networks under analysis in this thesis can be considered relatively small or even undefinable, they cannot be considered to be exhibiting the kind of behavior that would be distinguishable from the general economic patterns of their respective societies. Therefore, the inclusion of secondary measures of these networks is likely to prove more beneficial. Furthermore, since this paper deals with dynamic rather than static systems, the network of interest for this thesis needs to be observable over long periods of time in a standardized fashion.

In short, the network that would fit the needs of this thesis would have to be globally and temporally stable as well as readily accessible in terms of data.

It can be argued that affiliation networks are more likely to fit the above-mentioned criteria as opposed to conventional one-mode social networks. However, methods for the analysis of such networks have been less well developed (Wasserman and Faust, 1994, p. 292). Al-

\textsuperscript{1}Global Entrepreneurship Monitor, http://www.gemconsortium.org/
though entrepreneurs often belong to a number of affiliation networks (events defined as founding a company, being on the same board of directors or attending various social events which the startup community has no shortage of, etc.), they do not commonly fit the criteria outlined above - either due to the geographically dispersed availability of data or their temporal stability. The two types of affiliations that do fit the criteria are authors of scientific papers and inventors, as well as applicants of patent applications. Since the nature of the goal of this thesis is more economical than bibliometrical, it chooses to largely rely on patent information as one of the few readily available and comparable affiliation datasets.

While there are numerous organizations offering access to a wide range of patent information, gathering a comprehensive database on patents related to a particular country is difficult. One of the main reasons for this is that despite the fact that businesses have realized the benefits of protecting intellectual property globally, the actual legal and regulative space remains fragmented due to dominant USPTO, JPO and EPO agencies, followed by a number of smaller local agencies. Only few of the patent databases that are available feature interfaces that allow for complex queries and automated processing of the results. However, the ones that allow for that, the NBER project, for example\(^2\), tend to either have limited time series or focus on patent grants rather than applications. Also, the data quality is generally low, with patents missing authors and titles, same entities being featured under marginally different names due to filing or technical issues, and so forth. Despite USPTO being the default source for patent information for most researches (Benson and Magee, 2012; He and Ho-sein Fallah, 2009; Ter Wal, 2011), in addition to aggregated reports like the GCR and KEI, it was the WIPO dataset that was chosen for the purposes of this thesis. The main reason for this was based on the home advantage effect discussed in length by Criscuolo (2006). The

\(^2\)https://sites.google.com/site/patentdataproject/
author claimed that due to this effect, companies are more likely to seek patent protection from their local patent offices - an assumption that the research also confirmed. Since USA was not part of the group of countries to be analyzed, it was felt that a wider coverage (WIPO, 2012) of patent offices would be necessary for this thesis, hence the choice naturally spoke in favor of WIPO.

Within the WIPO database, for each observable country (Estonia, Latvia and Israel) the query "ARE:<country_code>" was issued. However, since variable latency between a patent application being filed and a patent being granted would introduce temporal difficulties in the analysis phase, patent application was used instead of patent grant data.

Upon analyzing the patent data for each of the three countries used in this thesis (Estonia, Latvia and Israel), the author of this thesis found significant problems with the quality of the data, some of which (variations in the spelling and visualizations of names) have also been noted in Newman (2001). The problem is further exaggerated by a linguistically diverse dataset, as companies have used both translated and native names, while the lack of support for Unicode has additionally led to variability in the spelling of names. For example, Ivars Kalvins appeared in the WIPO dataset as both Ivars Kalvins and Ivars Kalvinsh. For Estonia and Latvia, the dataset was manually cleaned using visualizations to identify potentially similar nodes. This process did not significantly alter the nature of the results due to the aggregative nature of the analysis. Since the dataset for Israel was relatively large, it was left untouched and instead the assumption was made that its size would decrease the impact of individual nodes, thereby bringing the impact the quality of the data has further down.

Firstly, R (R Core Team, 2012) was used for the data analysis, while for a more complex data analysis and the acquisition of the patent data, Python scripts were utilized. For the purposes of the latter, the Beautifulsoup library was used (Richardson, 2011) as the
data needed to be extracted from the web pages that were not intended to be machine-readable. Finally, for generating and analyzing the graphs, Pygraph (Matiello et al., 2012) proved to be invaluable for this thesis.

2 Overview of literature

Below, an overview of literature covering entrepreneurial social networks and an assessment of various innovation economies is given.

As the focal points of interest of this thesis are startup ecosystems and entrepreneurship in general, it is necessary to define what is meant by entrepreneurship and ecosystem. Chosen from a range of definitions for the first term, the one introduced by Whittaker et al. (2009) is particularly relevant for this thesis: "Entrepreneurship (is a) process in which opportunities are discovered or created, and turned into market outcomes by organizational means". Here, the relevancy can be seen in the way that this definition emphasizes the relationship between an entrepreneur and its context, thereby signifying the impact that entrepreneurs, in general, can have on a society. The term "ecosystem" shares this consideration for context of the subject entity. Tansley (1935) states: "when we are trying to think fundamentally we cannot separate [the organisms] from their special environment, with which they form one physical system". Willis (2008), having aggregated a large body of literature on this subject, defines ecosystem as "a unit comprising a community (or communities) of organisms and their physical and chemical environment, at any scale, desirably specified, in which there are continuous fluxes of matter and energy in an interactive open system". Based on this definition, an entrepreneurship ecosystem can be defined as a unit comprising a community (or communities) of entrepreneurs and their social and economic environment, at any scale, desirably specified, in which there are continuous fluxes of resources in an interactive open system. The definition of
innovation ecosystem can be derived in a similar manner.

As the above concept of resource exchange and interaction within a social context implies, entrepreneurs are frequently embedded in social networks which play a critical role in the entrepreneurial process (Aldrich and Zimmer, 1986). Thus, studying such a network is of great importance when the overall state of an entrepreneurial community is to be assessed. The properties of a social network are also significant for benchmarking purposes, as the ways in which these networks are developed and utilized can vary from country to country (Whittaker et al., 2009, p. 119-131). Hoang and Antoncic (2003) state three aspects which are critical for the theoretical and empirical research of these social networks among entrepreneurs (subsequently referred to as "actors"): the nature of the content that is exchanged between actors, governance mechanisms in these relationships and the network structure created by the crosscutting relationships between these actors. The current thesis is mainly concerned with how these crosscutting relationships between actors relate to the governance mechanisms within these relationships.

In order to understand the structure of the relationships between the actors within such a social network, at least three different views need to be combined - social, economic and mathematical. Various linguistic and scientific tools used in these three fields of science lead to the research of entrepreneurial social networks being fragmented due to conflicting terminology and different approaches to researching the same topic. For instance, a good example where two very distinct lines of research have emerged in past studies is the issue of detecting and characterizing a social network's structure. E.g. one of the lines of research on this subject was driven by computer scientists with applications in parallel computing, to name but one, while another line of research was led by a group of sociologists (Newman, 2006). The fact that these different approaches to the same problem use a wide variety of non-overlapping terminology is indicative of the situation.
Adjacent research is often not integrated. For example, Watts and Strogatz and Wasserman and Faust have used two different terms for the idea of measuring the "clumpyness" of a given graph, but they both arrive at a mathematically identical construct (Watts and Strogatz, 1998; Wasserman and Faust, 1994). On the other hand, Newman uses similar terminology but a different mathematical construct (Newman, 2001).

In the following, the three afore-mentioned views - the social, economic and mathematical - necessary for understanding entrepreneurial social networks, are discussed along with key texts within the three fields.

What can be considered as the seminal work on the topic from a social perspective is that of Wasserman and Faust (1994), since it contains a bibliography and index of names and terminology, which both serve as a highly useful point of reference for this thesis. The book functions as a core source of information for a comprehensive overview of key mathematical terms and the graph theory in general. While seeking to form a bridge between the specific and well-defined realm of mathematics and the metaphorical and often undefinable domain of social sciences, it provides explicit, formal statements along with measures of social structural properties that may otherwise be only defined in metaphorical terms (Wasserman and Faust, 1994, p. 17).

The mathematical approach to understanding the social networks in question can be claimed to be led by Newman (Newman et al., 2000; Newman, 2004, 2006) who has also ventured into its applications and done key work in the area of co-author networks (Newman, 2001). Watts and Strogatz have supplemented that research by having defined fundamental ideas on the internal structures of different entrepreneurial social networks. This was done by introducing the concept of 'small-world' networks and the key methods for further research within the subject (Watts and Strogatz, 1998).
For understanding the economic perspective of entrepreneurial networks, Hoang and Antoncic (2003) is a valuable critique combined with a strong, integrated effort but which unfortunately lacks a unified theoretical understanding of the view. Shane (2003) states that the reason for this lies in the division among the field’s researchers – one set of researchers focuses on the entrepreneur as an individual, while another focuses exclusively on external forces surrounding the entrepreneur. This general sentiment has been echoed by Wilken (1979). For the purposes of this study, it is important to note that despite the lack of a common theoretical foundation, this thesis sees entrepreneurship as a process which combines factors for production and is thereby heavily dependent on the ability of the entrepreneur to build and maintain both formal and informal relationships with its peers. Asheim and Isaksen (2002) specifically emphasize the need for international relationships between entrepreneurs. There is no consensus regarding how the nature of these relationships should affect entrepreneurs, but several sources (Salavisa et al., 2012; He and Hosein Fallah, 2009) have pointed to the need for integrative approaches to analyzing formal and informal networks, using both quantitative and qualitative methods. After all, both kinds of networks have been found to be mobilized in terms of accessing resources and they are often found to overlap as formal relationships lead to informal ones and vice versa (Salavisa et al., 2012).

An additional important line of research focuses on the relationships between various levels of entrepreneurship and the economic output of a society. At least two studies (Anokhin and Wincent, 2011; Acs et al., 2005) have found a U-shaped development model where entrepreneurship levels fail to provide economic growth below a certain threshold. Other authors (Whittaker et al., 2009) argue that the two concepts lack a clear cause and effect relationship, forming a feedback loop.

As there is no commonly shared theory of entrepreneurship, em-
pirical studies and benchmarking with other economies have resorted to assessing the state of a particular economy instead. As a result of this, a large body of literature exists focusing on such studies. These studies can be divided into two - firstly, there are periodic, comparative large-scale studies which arrange countries based on a specific aggregate index. This type of researches are discussed in detail in section 3.3. Secondly, there are numerous studies focusing on analyzing the innovation economy of a particular country (for example Kristapsons et al. 2011; Eljas-Taal 2011; Kurik et al. 2002; Reid et al. 2011) or a group of countries (Whittaker et al., 2009). Commonly, studies related to smaller countries like Latvia and Lithuania are aggregative in nature and build on previous research instead. Section 3.2 provides an application of these studies to the current problem of entrepreneurship ecosystems.

3 Comparison with other countries

In order to better understand the relative position of Estonia, and test the hypothesis on the country’s economic development, it is compared to a group of countries with similar macroeconomic constraints in terms of their size and population, while still having a proven track record in converting innovation into GDP growth. That group consists of Finland, Israel, Taiwan, South Korea and Taiwan. Latvia is added to the comparison due to its relatively similar geopolitical position to Estonia. Furthermore, the innovation economies of Estonia and Latvia have been found to have been almost the same at one point (Arundel and Hollanders, 2005; Allik, 2008). Thus a comparison of the innovation economies of the above-listed countries should outline the effectiveness of the different policies which have been enforced in each country. Finally, Singapore is used as a role model due to its high GDP per capita and universally recognized success in monetizing its innovation economy.
Most of the countries previously outlined have enjoyed higher than average and relatively uniform economic growth for the most part of the past three decades (figure 1). Only the exception of the economic crisis in 2008-2009 shows a notable deviation from the usual pattern. In terms of absolute numbers, as can be expected, three distinct groups emerge based on the selection of countries (figure 2): Singapore shows the highest GDP per capita, all of the countries in-between, and Estonia and Latvia positioning at the bottom of the ranking. The reason for why Singapore shows a significantly higher economic output than the other countries can be attributed to the country’s success in establishing itself as a logistics and finance hub (Luo et al., 2012). But these factors, in turn, are likely to stem from the investment that has been made into nurturing and attracting the necessary talent. The single decisive factor that made for Singapore’s development was the ability
of its ministers and the high quality of civil servants who supported them” (Lee, 2000, p. 664).

One way to explain the position of Estonia and Latvia as the countries with the lowest GDP per capita is by using Michael E. Porter's model of stages of competitive development (Porter, 1990, p. 545). It has been argued that Estonia, and the same is likely to hold for Latvia, has since the early 90s been in a rapid transition from a factor-driven stage, where availability and prices of basic factors are the sources of advantage, to an investment-driven stage, where success is determined by the willingness and ability of the nation and the companies it forms to start investing aggressively. (Kurik et al., 2002, p. 18). Due to the small size of Estonia and Latvia, as well as the very short time period that the two countries have attempted to master this transition, the amount of investments have not had enough time to accumulate. Both countries show a very similar growth trend compared to other countries within the group, who are around 15-20 years further down the line, indicating that the relatively late establishment of independence in both Estonia and Latvia has been a game-changing factor.

All of the countries under review show stable growth throughout the figure, while still remaining sensitive to global economic trends. This becomes especially evident around the year 2008. Thus, differentiation is necessary to be able to quantify any possible similarities in the trends that are not immediately obvious in figure 1. Table 1 contains the correlation coefficients and p-values for the GDP growth rates of each country. Only Estonia and Latvia seem to follow a similar economic growth pattern (author’s note: important figures are highlighted in bold) thereby confirming the literature-based assumptions made in section 3. The relationship between the rest of the countries, however, is either not strong or significant enough to be highlighted here.
3.2 Innovation economy

In the previous section, the state of a number of economies was reviewed in terms of their economic outputs. These numbers, however, have little to say about the ability of these countries to be globally competitive or even innovative. To gain a better understanding of this, various indicators of innovation inputs and outputs in the named countries are benchmarked in the following section.

Here, it is important to note that based on the theory put forth by Porter (1990), countries in different stages of economic development have different needs in terms of inputs. The GCR index (Schwab et al., 2011) positions Finland, Taiwan, Israel, Korea and Singapore in a group of innovation-driven economies, while Estonia and Latvia are positioned as being in a transition phase from an efficiency-driven to an innovation-driven stage. Therefore, it can be argued that innovation plays an important role in driving growth in all of these coun-
tries, but its role is much less significant for the transition economies of Estonia and Latvia.

As differences in patterns of GDP development would allow to assume, the intensity with which countries invest in R&D differs considerably. When comparing figure 3a to figure 2 on page 20, two main differences emerge. Firstly, the distribution of the amount of investments in all of the countries is much more uniform, with only Latvia appearing as an outlier by spending much less than the other countries and showing no long-term trend in the growth of investments. Secondly, the relative intensity of Singapore’s R&D expenditure is rather low. While in the case of Latvia, it is likely to be the result of conscious economic policies, the latter is likely due to Singapore’s ability to absorb investments. The GDP per capita of Singapore is at least two times larger than that of the other countries, which means that in absolute numbers per capita, Singapore’s investment is on par with the top spenders in the field.

Figure 3b shows how a number of countries have 60-60% of their R&D spending done privately, while Estonia is a clear outlier with its spending ratio having only recently approached 40 per cent. A further analysis in section 4.3 confirms the finding that most of the R&D spending in Estonia is driven by the public sector, notably the universities, but does not offer an explanation as to why this is so. If

<table>
<thead>
<tr>
<th></th>
<th>EE</th>
<th>FI</th>
<th>TW</th>
<th>IL</th>
<th>LV</th>
<th>KR</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>1.00</td>
<td>0.85</td>
<td>0.45</td>
<td>0.35</td>
<td>0.95</td>
<td>0.24</td>
<td>0.40</td>
</tr>
<tr>
<td>FI</td>
<td>0.85</td>
<td>1.00</td>
<td>0.43</td>
<td>0.44</td>
<td>0.86</td>
<td>0.37</td>
<td>0.45</td>
</tr>
<tr>
<td>TW</td>
<td>0.45</td>
<td>0.43</td>
<td>1.00</td>
<td>0.56</td>
<td>0.36</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>IL</td>
<td>0.35</td>
<td>0.44</td>
<td>0.56</td>
<td>1.00</td>
<td>0.41</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>LV</td>
<td>0.95</td>
<td>0.86</td>
<td>0.36</td>
<td>0.41</td>
<td>1.00</td>
<td>0.27</td>
<td>0.25</td>
</tr>
<tr>
<td>KR</td>
<td>0.24</td>
<td>0.37</td>
<td>0.71</td>
<td>0.52</td>
<td>0.27</td>
<td>1.00</td>
<td>0.68</td>
</tr>
<tr>
<td>SG</td>
<td>0.40</td>
<td>0.45</td>
<td>0.77</td>
<td>0.50</td>
<td>0.25</td>
<td>0.68</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(a) Correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>EE</th>
<th>FI</th>
<th>TW</th>
<th>IL</th>
<th>LV</th>
<th>KR</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>0.00</td>
<td>0.06</td>
<td>0.16</td>
<td>0.00</td>
<td>0.34</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>0.16</td>
<td>0.01</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.08</td>
<td>0.27</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>KR</td>
<td>0.34</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>0.10</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

(b) P-values
anything, it can be argued that the numbers could be even lower since Skype would appear as a statistical outlier. Israel was omitted from this comparison due to the defense sector being a prominent source of funding.

When it comes to the innovation output of the countries, the differences are remarkable. When the numbers of the USPTO patents per million people are compared, Taiwan emerges as a clear leader (see figure 4) while Singapore leads the way in terms of engineering articles (figure 5). Estonia and Latvia, however, barely register in terms of patent grants and simply lag behind with their number of engineering articles. To put this into perspective – over the past decade, Estonia has been granted around three times as few patents in total as Singapore received in 2010 alone. This is likely to be an indication of the relatively immature intellectual property market present in Estonia.
combined with the academia-led R&D investments, where patenting is considered less important than, for example, publishing.

Figure 4: USPTO patent grants per million people. Source: USPTO

This raises an important question on the impact that intellectual property policies have on statistics, especially when it comes to smaller countries and more dominant organizations, where a change in patenting strategy at one institution alone can have a significant impact on the statistics of the country. In order to confirm the hypothesis that patenting patterns differ significantly from country to country, the percentage of all of the patents filed at USPTO was calculated, with the results depicted on figure 6. Two conclusions can be drawn from the data found. Firstly, the percentage of the patents filed at USPTO varies considerably among countries. Latvia, for example, filed no more than 10 per cent of all of the patents at USPTO,

---

3 The way these questions impact the choice of data sources for this thesis is discussed in 1.2. Also, see section 4.3 for a more thorough analysis.
while Israel had already exceeded 50 percent by 2002. Secondly, there are significant differences in patenting policies in each country, as is evident in the case of comparing Singapore to Estonia. Therefore, it can be concluded that the number of patents filed at USPTO alone is not an adequate measure of the innovation activity within a country.

Despite the discrepancies found in the patent data, the trends evident for Estonia and Latvia are distinct enough from the other countries to warrant a separate and detailed analysis for their case. The rest of the countries have already been thoroughly benchmarked in Luo et al. (2012).

In a 2005 study commissioned by the EU, thorough cluster analysis (Arundel and Hollanders, 2005) found Estonia and Latvia to be very similar in terms of a range of innovation indicators, describing both of the countries’ innovation inputs and outputs. The two countries
were found belonging to a group called "the laggards" together with Greece, Poland and Portugal. By 2011, however, Estonia had moved to a group termed "innovation followers" by having demonstrated innovation levels close to the average of the EU27 countries (Acheson et al., 2011, p. 1). This change was attributed to the country's linear growth of investments into R&D (Eljas-Taal, 2011), which was further supported by the distinct trend present in both Latvia and Estonia in terms of their R&D investments (figure 3). Estonia does seem to follow the right pattern, as increasing investments into R&D is in accordance with transitioning from an efficiency-driven to an innovation-driven economy, which is by nature characterized by high investments into R&D from both public and private sectors (Anokhin and Wincent, 2011). The weak position of Latvia, however, was attributed to structural issues within the country (Kristapsons et al., 2011), while Estonia was likely to have benefitted from their extensive reforms and structural changes in this field, thereby making the effectiveness of their R&D spending much less obvious.
A similar conclusion was drawn by Allik (Allik, 2008) upon analyzing the scientific outputs of Estonia, Latvia and Lithuania, thereby pointing out the influence that policy changes with a large temporal lag can have.

### 3.3 Aggregate indexes

The analysis of different economic factors which lead to a country’s ability to be innovative should be anything but cursory. There are, however, organizations which aggregate a wide range of metrics based on a standardized approach and then publish the results. One of them is the Knowledge Economy Index (KEI) of the World Bank. This index is based on four main sources or pillars and represents a country’s or region’s overall preparedness for competing in the Knowledge Economy (The World Bank, 2012):

1. Economic Incentive and Institutional Regime
2. Innovation and Technological Adoption
3. Education and Training
4. Information and Communications Technologies (ICT) Infrastructure

Interestingly enough, in their latest published report, Singapore and Israel ranked relatively low, while Nordic countries such as Sweden, Finland, Denmark and Norway claimed four of the top five positions. This was explained by the span of factor categories the index encompasses in which the top five countries excel in and where, apparently, the R&D spending and patent output, as analyzed previously, has a relatively minor role to play.

Another well-established report, which is published more frequently than the Knowledge Management Index, is the Global Competitiveness Report (GCR) by the World Economic Forum. The theoretical foundations of this report are the classic works by Porter (for example, Porter (1990)). The report computes the rankings of countries
based on twelve different factors or pillars (see figure 7), where the weights of different factors differ based on the development stage of the country being analyzed.

![Diagram of 12 pillars of competitiveness]

Figure 7: The 12 pillars of competitiveness. Source: Sala-i Martin et al. (2007)

The well-known business school INSEAD has also been publishing a Global Innovation Index (GII) report since 2009, with the latest edition published in 2012 (Dutta et al., 2012). It offers the shortest time span compared to the last three indexes, but it sets itself apart with an approach that relies on seemingly secondary indicators - such as monthly Wikipedia edits and YouTube uploads, but also dedicated polls executed in the country - and much less on publicly available and generally recognized metrics like patent counts and GDP rates.

Lastly, the GEM consortium conducts an annual survey on the rate and profile of the entrepreneurial activity around the globe Bosma et al. (2012). Although the annual report does not assign a specific index to a country, the data it provides is detailed enough to allow for international comparisons.

Upon analyzing the main reports - GEM, GCR and GII - three observations can be made. Firstly, all three encompass very different
data frames and the differences in their publishing frequency make comparisons between the indexes rather difficult. Secondly, the geographical coverage of countries varies significantly. For example, Taiwan is not covered by the latest GII report, while Estonia and Latvia were absent from GCR for as late as the year 2000. Finally, the factors under consideration and their relative weights differ to the point that the indexes are almost incomparable beyond the significant differences that present themselves in relative rankings.

Since the absolute index numbers, and comparisons with countries which are not part of the group under review in this thesis, add little to the benchmarking of these countries, table 2 attempts to summarize those countries’ relative rankings in the 2012 and 2002 GCR reports, the 2012, 2000 and 1995 KEI reports, and the 2012 GII report for a further comparison. Latvia consistently ranks as the last country in all of the reports, while Finland and Taiwan are leaders in the KEI and GCR reports, and Singapore appears to be first in the 2012 GCR report. Estonia has moved up three places in the KEI index, thereby surpassing Israel, South Korea and Singapore. This change is partly supported by the GII index that places Estonia ahead of Korea, but behind the rest of the countries. The reason for why these countries have been outpaced by Estonia in the KEI report can only be explained by the computational structure of the index, as Singapore, Israel and Korea, contrary to Estonia, all appear in the top ten of one or more categories. It appears that the index favors uniform performance in all of the assessed areas.

To sum up, while aggregating a large number of factors into a single ranking does offer a less complex way of comparing countries, the diverse methodologies and reporting periods of the different index reports make the results hard to interpret. The change of Estonia’s index ranking in KEI, for example, was not reflected in the GCR report due to a lack of data, while Singapore ranked fourth in the KEI report, but first in the GCR. To some extent, the summary reports,
Table 2: Relative rankings of countries based on Knowledge Economy Index and Global Competitiveness Report. GCR data for 1995 is not available, Estonia and Latvia were not covered by GCR in 2000, Taiwan was not covered by GII in 2012. Source: The World Bank (2012), Dutta et al. (2012), Porter et al. (2000)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Taiwan, China</td>
<td>2</td>
<td>3</td>
<td></td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Singapore</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Israel</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Estonia</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Korea, Rep.</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Latvia</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

which can complement these indexes, make up for these shortcomings by providing a breakdown of the data for each country, thereby helping to explain some of the discrepancies.

Thus, it can be concluded that while these three reports are important sources of analysis and can be used to determine trends for any given country, they are not quite as reliable for benchmarking countries against each other.

3.4 National culture

When comparing innovation economies, national culture needs to be taken into account as it has an impact on the creative and innovative processes within a country (Westwood and Low, 2003). Wilken (1979) points out the relationship between national cultures and entrepreneurship via the concept of social mobility. Curiously, the author provides several contradictory views on how this relationship can function, the only common motif to these views being that the rela-
There are numerous ways of defining "culture" - based on the field of research (anthropology, sociology and management each have their own perspectives on the term), scope of the definition (national or organizational cultures, for example) or even semantics (the word "culture" in the context of the English language versus the term "culture" in a scientific context). In a business context, the definition for "culture" as provided by Hofstede is probably the most widely used. He defines culture as "the collective programming of the mind which distinguishes the members of one human group from another" (Hofstede, 1980, page 25). Hofstede complements the definition with a quantitative framework for assessing the various aspects of a national culture from the perspective of five different dimensions:

**Power distance** The degree to which the less powerful members of the society accept unequal distribution of power

**Individualism versus collectivism** Location of a society on an axis between where individuals are expected to take care of themselves and where a group looks after its members in exchange of loyalty

**Masculinity versus femininity** Position of a nation on the scale between masculine and feminine values

**Uncertainty avoidance** The degree to which members of the society feel uncomfortable when faced with uncertainty and ambiguity

**Long-term versus short-term orientation** The ability of a nation to adapt traditions to changed conditions

To these five dimensions, Hofstede later added a 6th dimension of "Indulgence versus Restraint" which measures the extent to which a society allows free gratification of basic and natural human drivers related to enjoying life and having fun.
By design, the numeric measures for these dimensions are ordinal in nature and only become useful when used to compare countries with each other. Figure 8 compares the countries previously listed along with Hofstede’s dimensions.

![Figure 8: Hofstede’s five dimensions for the benchmarked countries. Information was not available for Latvia, the temporal dimension was not available for Estonia and Israel. Source: Hofstede (2012)](image)

Unfortunately, the figure does not aid in making clear conclusions on the relationships between cultural dimensions and innovation, which is in line with what has been claimed in subject-related literature. Westwood and Low (2003) describe the relationship between culture and innovation as "nuanced" and not one to be considered "universally, simplistically or un reflexively". A good example of that is a very low uncertainty avoidance score for Singapore which has nevertheless not prevented the country from developing a successful innovation economy. In addition to that, it is evident that Asian countries in the group at hand show much stronger signs of collectivism than Finland or Israel, there is a significant difference in power dis-
tances in Singapore and Israel and so forth. Based on this data, there seems to be no clear relationship between Hofstede’s dimensions and the innovation success of a country. This conclusion is rather counterintuitive, as one would at least expect dimensions such as uncertainty avoidance or power distance to have an effect on the results. Thus, while Estonia does show a strong cultural similarity to Finland, it should not be taken as a sign of Estonia’s potential as an innovation economy.

This empirical finding is in line with previous researches that have also found no obvious relationship between any of the cultural variables and entrepreneurial activity in general (Hunt and Levie, 2003). There is, however, a research that seems to point to the contrary (Thomas and Mueller, 2000). The reason for the differences in these findings could be that Hunt and Levie correlate absolute, while Thomas and Mueller use relative values by measuring cultural distance from a base point (United States). And yet, the notion prevails that some nations, cultures or subcultures may be more entrepreneurial than others. A good example of this is the work by Senor and Singer (2011) which attributes much of Israel’s success to a particular type of national culture. For example, quoting Simon Peres: "The greatest contribution of the Jewish people in history is dissatisfaction. That’s poor for politics but good for science" (Senor and Singer, 2011, p. 228). This is in line with the statement that there seems to be a dominant and commonly shared belief that national culture and successful entrepreneurship are connected, but the conclusive nature of that relationship has remained elusive.

3.5 Summary of the comparison

The benchmarking of the Estonian economy with that of Finland, South Korea, Taiwan, Singapore, Latvia and Israel yielded few unexpected results. The smaller and less developed countries such as Latvia and Estonia are clearly behind in terms of all of the metrics
which were reviewed in this comparison, while the rest of the countries form a rather tight-knit group with few outliers. Estonia and Latvia, as was expected, are indeed similar in terms of their strongly correlated GDP growths (see table 1) and innovation outputs. However, the differences in the innovation outputs and positions in the aggregated index analysis, along with a much larger relative R&D expenditure in Estonia, point to a difference in relevant policies. Here it is important to note that these differences are not reflected in the GDP change patterns, which confirms that neither of the two countries are as of yet an innovation-driven economy.

4 Analysis of the Estonian innovation ecosystem

4.1 Estonian startups in an international context

The analysis in the previous sections of this paper seems to indicate that Estonia is lagging behind in many areas such as GDP rate or innovation output, while still showing potential for growth and improvement. It is not clear, however, how much of the economic and innovation-related growth in the recent years can be attributed to the Estonian startup ecosystem, the corporate entrepreneurs and the academic institutions.

Anokhin and Wincent (2011) argue that the relationship between the levels of entrepreneurship and innovation change when a country moves from one stage of development to the next. In an early development phase, the relationship is negative (higher startup activity rates lead to lower innovation rates), but turns positive in later stages. This finding was supported by Wennekers et al. (2005) and Acs et al. (2005) who found a U-shaped relationship between entrepreneurship
levels and per capita income. Based on both cited models, the levels of entrepreneurship in Estonia should continue to drop over the next few years, as the country moves towards the bottom of the curve. This is in opposition to what can be concluded from Schwab et al. (2011) which places Estonia in a transition phase\(^4\) with an innovation-driven economy. This should mean rapidly increasing entrepreneurship levels for Estonia, as incumbent innovators decentralize and new organizations emerge.

Unfortunately, there is little information available on this nascent Estonian entrepreneurship level, as the country has not participated in the GEM project or in any other similar endeavors. The European Commission does execute a Community Innovation Survey on a regular basis, but this survey is focused on the ways that existing companies deal with innovation and does not encompass the concept of a startup as a nascent organization. Furthermore, the latest CIS survey, covering the period of 2006-2008, does not even consider organizations with less than ten employees (Eurostat, 2011).

There are two secondary sources, however. Firstly, the Estonian Board of Statistics is able to extract the number of employees in companies that are younger than 3 years in any given year. This is an approximation of the definition of a startup used by the GEM (Bosma et al., 2012) and while not directly comparable, it should still allow for identifying certain trends. Secondly, one can look at the number of applicants of patents with Estonian inventors. This is, again, not a direct measure, as patents can only be seen as an indirect evidence for innovation activity, and since the patent dataset used contains non-resident entities (see section 4.3 for discussion).

Both data sources are summarized in figure 9. While the number of patent applicants show a clear growth trend \((Pr(> |t|) < .01)\), the percentage of the population employed by companies that could

\(^4\)Curiously, i Martin (2009) defines Estonia as an innovation-driven economy, with later citings moving the country back to a transitional stage. The glitch is likely to have been caused by an unsustained GDF growth prior to the crisis in 2008
be considered entrepreneurial is showing a slight downward trend \((Pr(> |t|) = 0.245)\) if anything. This is in contrast with Anokhin and Wincent, who predicted that there would be a clear upward trend (which the employment figures do not support), and with Wennekers et al. (2005), who assumed that there would be a strong downward trend in both metrics.

Thus, it can be concluded that Estonia does not exhibit the kind of behavior that would fit the established economic development models.

The situation proves to be even more peculiar by a significant body of evidence for the Estonian startup ecosystem to have rapidly grown in the past ten years. Former Skype employees alone have established numerous startups (Transferwise, Candycane, Marinexplore to name a few), Starkell (2012) found over a hundred startups and entrepreneurs on the popular microblogging service Twitter. Garage48 is exporting its rapid prototyping events globally, while GrabCAD, Sportlyzer, Campalyst and ZeroTurnaround have all secured interna-
tional financial backing. At the Seedcamp London 2011, four out of twenty finalists were from Estonia, with one of them moving on to becoming one of the winners. There is also remarkable coverage on the success of the Estonian startups in globally influential media outlets (Rooney, 2012; Business Insider, 2012). In short, there are definitive signs of an entrepreneurial spirit being on the rise in Estonia. However, that trend does not seem to be perceptible in any of the usual metrics.

There seem to be strong discrepancies between established academic theory of entrepreneurship, application of that theory by various indexing organizations, circumstantial evidence and empirical data. This controversy is too evident to be explained by weaknesses in the theory or its application. Thus it is likely that the metrics that link empirical data and entrepreneurship theory might not be adequate or even applicable in the case of Estonia. Thus, the author suggests two hypotheses for potentially explaining these discrepancies.

Firstly, Estonia spends only a small proportion of its relatively low GDP on R&D (see figures 2 and 3a on pages 20 and 22). This may lead to a critical mass of investment not being achieved in many areas of research. This, combined with an open economy and few restrictions on international academic collaboration, may lead to Estonian researchers starting to contribute to the work of larger groups of researchers in other countries, as opposed to establishing the groups themselves. If this were to happen, the proportion of primary and secondary inventors in Estonia would be significantly different from that of a country with a much larger investment in R&D per capita. Indeed, figure 10 shows that once the patent volume picks up, Israel starts to exhibit a consistently higher percentage of first named inventors than, for example, Estonia or Latvia, both of whom have much lower patent volumes in comparison. Since patent statistics commonly use the country of the first named inventor to attribute the patent to, then this means that in 2005 about 84% of the patents that Latvian
citizens worked on did not end up contributing to the official innovation output of their native country.

Figure 10: Percentage of all patent applications where the first named inventor is from a particular country. Source: WIPO, authors calculations

Hence, it can be said that the hypothesis that Estonia has not been able to achieve critical mass in many research areas seems to hold.

Secondly, the large volatility in the Estonian and Latvian first named inventor percentages leads to the hypothesis that the size of the R&D community makes the outputs sensitive to small fluctuations of inputs. A clear outlier, such as Skype, is able to contribute to the innovation output indicators disproportionally to its overall effect on the economy. Figure 11 summarizes the impact the three biggest members of the R&D investment community - Skype, University of Tartu and Tallinn University of Technology - have had on the number of filed patent applications in the last decade. It becomes evident that the impact has been significant - the rapid growth in the number of patent applications between 2006 and 2007 has to a large extent been influenced by Skype and University of Tartu. Without
their contribution, the growth would have been around 48%, but was 152% instead. In 2009, Skype contributed 36% to the total number of patent applications in Estonia. Thus, it can be said that the hypothesis of sensitivity of the metrics should not be renounced at this point, as relatively insignificant decisions made by a small number of organizations alone can heavily influence the overall innovation output indicators of a country. The problem of overly dominant patent applicants is further accentuated by the relatively low data quality of different patent data sources as well as the discrepancies in intellectual property policies (see figure 6). Further analysis of the patent dataset supported the sensitivity hypothesis for Estonia, and exposed an even larger dependence on a single organization for patent applications in Latvia (see table 4 on page 59).

![Figure 11: Relative contribution to the number of Estonian patent applications from major organizations. Source: WIPO, Author's calculations](image)

It must be emphasized that the fact that a few companies can have a significant impact on a country’s output indicators is not limited to patent application numbers. In small economies, indicators
ranging from education to R&D spending are much more dependent on individual organizations than they are in larger economies.

The shortcomings of patents as indicators for innovation activity have been outlined before, most notably by Archibugi and Planta (1996). However, they do not include the issues pointed out above - the problem of a sample size eroding the trustworthiness of statistics, and the impact structural differences in the patent body have on an innovation economy. Therefore, it can be concluded that these shortcomings are not widely recognized and have not been taken into account when assessing or benchmarking countries against each other.

In addition to the previously outlined factors, there are others which are not as easily quantifiable. Many of them are related to globalization and the limitations set upon by the socioeconomic size of Estonia. The globalization factor implies that many Estonian startups have the opportunity to choose to group together where the intellectual property policies or access to resources such as capital, customers or talent are more favorable. In addition to that, the limitations set upon by the small socioeconomic size of Estonia means that more startups are likely to pick the afore-mentioned route, as the home market can only support a limited amount of enterprises with a limited customer base. This can lead to a dilution of the Estonian startup ecosystem, as companies start to become increasingly more globally-oriented in their course of action.

4.2 Additional quantitative analysis

4.2.1 Co-author graph

It has now been exemplified that while the number of issued patents can generally be considered a good proxy for assessing innovation activity levels (Archibugi and Planta, 1996), simply counting the patent applications can still be misleading (see page 38 for a more detailed discussion). Also, as the fragility of an innovation economy exhibits
little that can be observed using conventional econometric means, a look into the social structure of an entrepreneurial social network may yield more informative results. Since most of the ways in which these entrepreneurial social networks exist do not allow for readily accessible data, information on patent applicants and inventors of patent applications is, while flawed, the most beneficial insight into the subject-matter. The rationale behind choosing this type of social network for further analysis is more thoroughly discussed in section 1.2.

The classical approach to constructing a network based on patent data results is to construct a co-author network consisting of authors linked by the events of article co-authorship (He and Hosein Fallah, 2009; Ter Wal, 2011). In the current setting, however, this would be insufficient as the goal is to understand the commercial co-operation structure of inventors as well as the relationships between inventors. Therefore, the network to be studied has to encompass all of the actors within a given startup ecosystem. This includes the creative force behind a patent, the inventors and the entities that apply for a patent, meaning the applicants themselves.

The second important consideration is the question of graph weights. A natural way of assigning weights to graphs would be to use a temporal parameter combined with the number of relationships between the nodes that a dataset generates. Another approach would be to use PageRank (Page et al., 1999) or an analogous algorithm which would assign weights to the nodes based on the information on the edges. The goal of the current task, however, is not to gather information on the importance of individual nodes but on the structure of a graph in general. Therefore, a decision was made not to use weighted edges and to represent any number of relationships with one unweighted edge instead.

Figure 12 shows a sample of such a graph. It depicts three different patents. One with inventors I5, I4 and applicant A1, with inventors I1, I2, I3 and applicant A1, another with applicants A2, A3 and inventors
The algorithm laid out above is more formally described in figure 13. It should be noted that this is a simple implementation of the algorithm of the complexity $O(pn)$, where $p$ is the total number of patents and $n = |I \cup A|$ is the cardinality of a set of all inventors and applicants. Algorithm 13 is thus usable for relatively small datasets. Newman (2001) provides a more optimal algorithm which, for the sake of simplicity, is not used in this thesis.

In the context of an analysis of a social network, the graph $G_P$ resulting from algorithm 13 is a two-mode affiliation network encompassing both ecosystem actors and information on patenting (Wasserman and Faust, 1994, p. 40).

This algorithm constructs a graph that consists of a number of fully connected components (i.e. all the inventors and applicants of a patent) which at times have overlapping nodes, as inventors can apply for patents while working for different companies and organizations can have multiple associated research teams etc.
The definition of an applicant within the context at hand emerges from this construct - an applicant is an entity that appears on the applicant list of a patent application but does not appear on the inventor list. It is important to make this distinction, as the inventors also often appear as applicants for a particular patent, despite their primary focus being on innovating not the organization they work for.

Another important term derived from this definition is the overlap probability (hereafter denoted as $p$), meaning that the probability with which a particular node can in an application-generated subgraph of inventors and applicants appear in another similar subgraph.

Hence, the two main characteristics - the generating parameters - of a co-author graph are the distribution of the sizes of fully connected components and the overlap probability.

In order to develop metrics for assessing such graphs, a specific method is necessary for constructing graphs with similar properties but in a predictable manner. For this purpose, algorithm 13 is modified to generate a random set of patents with a fixed size for a set of...
actors, resulting in the algorithm in figure 14. It creates a graph with two main parameters - the size of the clusters and the overlap probability. The algorithm simulates an ecosystem that only has patent applications with a fixed number of inventors and applicants, and where any party part of the application can, with a fixed probability, participate in a different application.

**Data:** Size of patent clusters $c$, probability of nodes overlapping between clusters $p$, size of the resulting graph $S$

$G \leftarrow$ new graph

**while** $|G| < S$ **do**

$V_s \leftarrow$ new set

**while** $t < c$ **do**

| if **probability** $p$ **then**
| | $n \leftarrow$ random element of $V(G)$
| else
| | $n \leftarrow$ new vertex
| **end**

| add $n$ to $G$
| add $n$ to $V$
| $t = t + 1$
| **end**

add a fully connected subgraph consisting of nodes in $V$ to $G$

**end**

**Figure 14:** Algorithm for generating a random co-author graph

As figure 15 shows, graphs generated in this manner are "small-world" graphs, as defined in Watts and Strogatz (1998) - the $C_1$ figures for a generated graph with a given overlap probability $p$ are much higher than the ones for a random graph with the same node and edge count.\(^5\)

---

\(^5\)Since the generated graph is disconnected for majority of $p < 1$, its characteristic path length is $\infty$ which is always larger than the finite result of a random graph.
4.2.2 Network fragmentation

Although the generating parameters can be assessed based on the affiliation matrix of a graph (Wasserman and Faust, 1994, p. 298), the assessment would need to be aggregative in nature and neglect the distribution of both parameters. It would therefore be desirable to develop easily interpretable metrics which would allow for directly quantifying the properties of co-author graphs that correlate to the generated parameters.

Watts and Strogatz introduce the concept of 'small-world' networks and define it by using a so-called clustering coefficient Watts and Strogatz (1998). The same concept is named “subgraph density”
by Wasserman and Faust (Wasserman and Faust, 1994, p102). The local clustering coefficient for an individual node \( n_i \) with the neighborhood \( N(n_i) \) and degree \( d_i = d(n_i) \) is defined as

\[
C_i = \frac{\Gamma_i}{d_i(d_i - 1)}, \Gamma_i = |E(N(n_i))|
\]

(1)

Effectively, \( C_i \) shows what proportion of all of the possible edges in the neighborhood of a vertex are actually there. From these local coefficients a global clustering coefficient can be calculated as follows:

\[
C_1 = \frac{1}{N} \sum C_i
\]

(2)

As it is clear from 1, \( C_i \) is not defined for cases where \( d_i < 2 \). Newman et al. (2000) alleviate this problem by using individual sums for both parts of the fraction:

\[
C_2 = \frac{\sum \Gamma_i}{\sum d_i(d_i - 1)}
\]

(3)

Since this approach still underrepresents leaves and isolated nodes, Kaiser suggested adjusting \( C_1 \) by the fraction of nodes that have only one or even zero neighbors, \( \theta \) (Kaiser, 2008):

\[
C' = \frac{1}{1 - \theta} C_1 = \frac{1}{N(1 - \theta)} \sum C_i
\]

(4)

Since by definition, the graph \( G_R \) consists of a number of fully connected subgraphs that are relatively sparsely connected among themselves, coefficients \( C_i \) for individual nodes are likely to exhibit little variability and would be unsuitable for assessing key properties of a graph. In order to test this assumption, a large number of graphs were generated using the algorithm in figure 14 by applying to it the clustering coefficients \( C_1 \) and \( C_2 \). The results in figure 16 show that neither \( C_1 \) nor \( C_2 \) are suitable metrics for co-author graphs as they

\[C' \] was eventually omitted since the algorithm already utilizes a constant cluster size, \( C' = C_1 \).
show a strong non-monotonic relationship to node overlap probability. In the current setting, clustering coefficients are thus less than ideal metrics for characterizing the properties of co-author graphs or their generating parameters. Cluster sizes seem to have little impact on the coefficients, which can be argued is not surprising, considering that the components are much smaller than the actual graph size, thereby minimizing the impact the actual size these particular clusters can have.

Figure 16: Relationships between fragmentation coefficients $C_1$, $C_2$, overlap probability and cluster size. Source: author’s calculations

This result confirms the hypothesis that the nature of $G_R$ makes these metrics unsuitable for assessing a graph’s properties. Based on
the definition of the graph given in algorithm 13, it may prove to be more beneficial to use the number of edges on the outside of the neighborhood instead of the number of edges within the neighborhood. A ratio of the number of edges a neighborhood misses due to the nodes on the outside to the number it could potentially have is a good measure of local fragmentation. Formally:

\[
F_i = 1 - \frac{\Theta_i}{d_i(N - d_i - 1)}, \Theta_i = \sum_{n_j \subseteq N(n_i)} |E(N(n_j) - N(n_i))| - d_i \tag{5}
\]

The total fragmentation is then calculated as an average of local values:

\[
F = \frac{1}{N} \sum F_i = \frac{1}{N} \sum \left(1 - \frac{\Theta_i}{d_i(N - d_i - 1)}\right)
= 1 - \frac{1}{N} \sum_{n_j \subseteq N(n_i)} \frac{|E(N(n_j) - N(n_i))| - d_i}{d_i(N - d_i - 1)} \tag{6}
\]

Figure 17: Relationships between fragmentation coefficient F and key graph parameters. Source: author’s calculations
Figure 17 shows the application of the newly defined metrics to the dataset generated in the same manner as in figure 16. Here, the results are much more definite, as the relationship between the two does seem to be monotonic. Again, there is no apparent correlation between cluster size and a fluctuating mean. It can thus be concluded that the fragmentation metric defined by the equation 6 can be considered suitable for comparing co-author networks of different countries. Since the graph generated by the algorithm in figure 14 is not directly dependent on the algorithm 13 and can thus generate a generic two-mode affiliation network, the result can be applicable for a more generic case as well.

The main problem with the fragmentation coefficient $F$ is that it barely deviates from 1 up to probability .5 where most of the actual data is likely to be located. In fact, the distribution of $F$ seems to be close to the formula $y = 1 - ae^{b(x-1)}$ which is depicted in figure 17a where $a = 0.2$ and $b = 10$. Further study in this direction may lead to an improvement of the coefficient but at this stage this course of action lies outside the scope of this thesis.

### 4.2.3 Subgraphs

In addition to being aware of the relative fragmentation level of a graph as a whole, it would be desirable to structure a graph into subgraphs where, based on a random chance, fewer relationships than expected would exist between the subgraphs (Newman, 2006). There are numerous algorithms for achieving such a modularization (see Newman (2004) for a review). In the case of the graph in question, however, the problem is more complex as it is inherently built of graphs which fit the definition of Newman, i.e. the network structure of the subgraph is already known. Thus, we are not interested in the optimal subgraph division but the structure of the hierarchy of the subgraphs. In addition to this, since the probability for node overlap is likely to be small in this case, the graph to be analyzed would have to consist
of a large number of components where only a few of which would exhibit a significant internal structure (see section 4.3 for the results of the data analysis).

Therefore the components, meaning the subgraphs which do not connect to any additional vertices within the supergraph, of the co-author graphs can be used as good proxies for subgraphs. This can be understood as the number of groups of people that have a relationship with each other via patentable work. Since this number would be dependent on the size of the graph, it would need to be normalized. Normalization, however, makes the resulting metric hard to interpret and introduces an inconvenient dependence on the size of the graph, as for meaningful graphs $N$ grow faster than the number of clusters when nodes are added.

Formally, the metric for graph $G$ is defined as:

$$N_C = |G_i|, G_i = (V, E) \in G, \forall v = \{x, y\} \in V, x, y \in E \quad (7)$$

When the same method of graph generation is applied as described in section 4.2.2, one sees that the relationship between the component count and overlap probability is monotonous and appears to be declining exponentially (see figure 18).

Since the numbers are not normalized, it is not be possible to draw conclusions directly based on the numbers. However, comparing the relationship dynamics between different countries should provide a more informative insight.

### 4.3 Analysis of the patent graph

Before any type of analysis can be done on the patent graph, the question of temporal mapping of the graphs needs to be solved. A solution where patents applied for in each given year are mapped would most likely be an unsuitable approach due to the nature of the process of applying for a patent. With this approach, however, one would lose
the information available on people moving from one organization to another. On the other hand, aggregating the entire patent set of a given country to one graph would undermine the dynamics behind the mapping. In the following, a combination of the two approaches is used, where an annual graph is constructed by using information on all of the patents that were applied for throughout the entire given year, thereby effectively reflecting the state of the startup ecosystem at a given point in time. It would be preferable to use the moving window concept for this process, much like it was done by Ter Wal (2011). However, in the case of Estonia and Latvia, the resulting decrease in the size of their datasets is likely to be more harmful for the overall results than the omission of the windowing concept.

The fundamental goal of this section of the thesis is to use this type of graph to determine whether the structures of the startup communities of two less successful countries, Estonia and Latvia, are different.
Table 3: Summary of the co-author graphs of the relevant countries

<table>
<thead>
<tr>
<th></th>
<th>Estonia</th>
<th>Latvia</th>
<th>Israel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patents involved</td>
<td>437</td>
<td>352</td>
<td>24,885</td>
</tr>
<tr>
<td>Edges</td>
<td>7,790</td>
<td>6,594</td>
<td>23,327</td>
</tr>
<tr>
<td>Nodes</td>
<td>1,140</td>
<td>8,213</td>
<td>36,742</td>
</tr>
</tbody>
</table>

Table 3 summarizes the main characteristics of the individual graphs, dated 31st December 2011. Immediately, it becomes evident that Israel’s graph contains two more orders of magnitude than Estonia and Latvia. It can be argued that the cause for this running theme throughout the entire analysis to follow lies behind the much smaller variability of the Israeli data. At the same time, the other two countries are very close in graph size, with Estonia edging slightly ahead of Latvia.

Figure 19: Percentage of foreign entities among patent applicants and inventors in Estonia, Latvia and Israel. Source: Author’s calculations
Another important question, beyond the basics, is the question of community boundaries. It is important to note here that the definition of a patent applicant given in section 4.2.1 speaks of all of the patents having an inventor from a particular country. This raises the question of to what extent and in what roles are foreign entities involved in a patent application process. Asheim and Isaksen (2002) offer a clear correlation between the amount of foreign relationships and corporate longevity, therefore a difference here could be important for the success of the ecosystem as a whole. Also, while patent applications contain information on the national affiliation of a person or a company, this information is not verified and is furthermore subject to corporate policies. Thus, in this context, the term "foreign" is likely to mean "not from a country as determined by the entity", as opposed to pointing to a legal or geographical position as it would in a normal setting.

Figure 19 shows the dynamics between local inventors and applicants for the three countries. The most apparent and surprising result is the extent to which the communities are not based in their respective countries. In 1994, up to 90% of the applicants in the case of Estonia, and in 1997, all of the inventors for Latvia were not of local origin. Indeed, back then there were very few patent applications from either country (2 and 6 respectively), in addition to the fact that it would take relatively little to either have a local scientist participate in a large foreign research group or a foreign group of inventors to choose a small country as the location for their intellectual property. For larger sample sizes, like in the case of Israel, such anomalies even out by themselves, but for small samples, such as in the case of Estonia and Latvia, they end up disrupting the results. Throughout the period reviewed, the average percentage of foreign applicants and inventors in the Latvian and Estonian communities is around 50-60%. Interestingly enough, the proportion of foreign inventors in Israel has been trending up for more than a decade, while the numbers for ap-
Applicants in Israel have been relatively stable or even showing a slight downward trend. This seems to indicate that more and more of the patenting work in Israel is done via researches that cooperate with foreign entities, while the nature of business cooperation has not changed significantly. The latter trend is in contrast with Estonia and Latvia where the percentage of foreign inventors starts very high after slowly trending down (see figure 19b).

All in all, it seems that the degree of foreign cooperation does not differ significantly between successful and less successful countries. This raises an important question on the validity of the subsequent analysis done in this thesis. When more than half of the entities involved are not local, how much can this thesis effectively conclude on a particular country? While it is a valid concern, it illustrates the nature of the chosen countries and the basic problem in counting patents more than it undermines the quality of the conclusions that can be drawn from the analysis. Any attempts to discard the foreign entities from the analysis would lead to loss of data about the country. If the foreign inventors would be removed from the picture, local inventors would appear disproportionately influential, while some patents would lose all of their inventors, not to mention the loss of information on important relationships between inventors. Therefore, one should accept the fact that in the case of small countries, the innovation ecosystem extends itself far beyond the borders of the country itself. Hence, in this globalized world, Merz and Skype are as much members of their respective local communities as are the inventors or applicants born and raised in that country.

Before having to turn to more advanced methods, a simple frequency analysis can provide equally as useful insights, especially, as was previously discussed, in the context of an international cooperation. In this case, one would want to investigate how the number of inventors and applicants per patent, the number of patents per inventor and the number of patents per applicant compare between the
Figure 20: Estimated cumulative distribution function for inventors and applicants per patent for Estonia, Latvia and Israel. Source: Author's calculations

three countries. For this type of a comparison, the cumulative distribution functions were estimated for the three metrics and the three countries. Here, applicant is understood as a person or a company that does not appear on the inventor list of a particular patent but is present on the applicant list.

Figure 20 shows the distribution of inventors and applicants per patent, where it becomes apparent that Latvia is in that sense quite different from Israel and Estonia. The latter two show a neat adherence to a power curve while Latvia does not, thereby pointing to a much more substantial proportion of teams larger than 2-3 people. Figure 21a shows all three countries following a clean power curve.
with visibly different parameters. Estonian inventors have the least patents on average, followed closely by Latvia and Israel. Curiously though, Estonia’s curve is almost the exact match to Lotka’s law with the exponent -2. It is interesting that -2 was the original exponent estimated by Lotka, while Newman found significantly different numbers upon analyzing data on co-authors for publications in the field of medicine and computer science (Newman, 2001). Why Estonia performs remarkably similarly to predictions made in terms of a law in scientific productivity is unclear at this stage. It can be speculated that the reason for this lies in a relatively high proportion of academic research and the dominance of a classic university (see table 4) in the Estonian startup ecosystem, but it is here that the hypothesis remains untested. Finally, figure 21b depicts the numbers for patents per applicant, were Latvia is once again the outlier. Firstly, the shape of

---

7 The matching curves would be even more similar if foreign inventors were discarded from the analysis.
Latvia’s curve is further removed from the power curve than is the case for the other two countries. However, while the curves of Israel and Estonia start to flatten beyond 20 applicants, thereby pointing to a long tail of applicants with a large number of patents, it is Latvia whose most prolific applicant in the dataset has about 25 patents.

The above analysis does not seem to hint at major structural differences between the countries, as the relatively small deviations on the part of Latvia can be explained by its smaller dataset. There are, however, clear differences in terms of the parameters of the structure of the networks. In order to understand these differences, a more thorough analysis of the properties of the graphs needs to be undertaken. The new metric F, which was developed previously, seems to be suitable for assessing the overlap probability of an affiliation network, as it can be used to interpret the measure of fragmentation in a given community.

Figure 22: Metrics of the co-author graphs for Estonia, Latvia and Israel. Source: Author’s calculations

Figure 22 shows the calculated F values and N_C change rates for
the selected countries. Again, the usual pattern emerges - the variability of data for smaller countries is much larger than for the bigger countries. Still, all of them seem to behave differently. The fragmentation coefficient for Israel shows a clear and steady upward trend, as the chances for people and companies having worked with people from outside their group decrease over time. This points to a constant stream of new companies and inventors who are not necessarily brought in by already established companies or inventors. This finding is diametrically opposed to the commonly utilized logic (Reid et al. 2011 specifically points out the effort being put into making the community tighter and increasing cooperation) that a tightly coupled innovation ecosystem is more desirable than a fragmented one. It is interesting that the growth rates for a number of the components are constantly on a decrease, with Israel showing a much lower growth rate than, for example, the other two countries. This seems to signal that there can only be an optimal amount of research groups in a given country, after which the chances for newcomers to create their own groups become poor.

In addition to the amount of components, the size of the components is also of interest here, as a small number of large groups surrounded by a number of smaller ones points to a community that may be sensitive to the changes caused by a small number of organizations. Figure 23 attempts to illustrate the situation. A vast majority of the subgroups in all three countries are small, mainly consisting of 2-3 people. This means that there is only one company with one or two people producing the patents. While the graph for Israel is relatively smooth, the graphs for Latvia and Estonia exhibit curious differences around the cluster sizes of 6 and 8. This may be caused by the dominance of universities in the ecosystems and can be related to the size of the research teams. However, the verification of this hypothesis, once again, lies outside the scope of this thesis. The estimated cumulative distribution functions (CDFs) show a much clearer
distinction between the countries. The curve for Israel is that of an obvious power curve. The curve for Estonia, on the other hand, is much flatter, while the one for Latvia is flatter still, all of which indicates a strong bias towards larger research groups. 10% of Latvian CDFs is larger than 12 people, the same figure for Israel, for example, is 6. This finding confirms the large fragmentation in the Israeli innovation ecosystem.

As the graphs for all three countries contain single large clusters, and since the cumulative numbers say little about the outliers, it is logical to look into them in more detail. To do this, the three largest components of each country were identified, and for each the applicant with most patents was found. Table 4 summarizes the results.

The first observation one can make is that of all the three countries containing a large cluster which is by far the most dominant one. The emergence of a "giant component" is a common phenomenon known from the theory of random graphs (Newman, 2001). In the case of

---

Figure 23: Distribution functions of the component sizes for Estonia, Latvia and Israel. Source: author's calculations

---

(a) Estimated distribution function

(b) Estimated CDF
Table 4: Largest clusters and their dominant applicants

<table>
<thead>
<tr>
<th>Country</th>
<th>Cluster size</th>
<th>Name</th>
<th>Type</th>
<th>Corporate home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Israel</td>
<td>26 077</td>
<td>Yissum Research Development Company of the Hebrew University of Jerusalem Ltd</td>
<td>Technology transfer</td>
<td>Israel</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>First Solar, inc.</td>
<td>Solar energy</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>Wave Group ltd</td>
<td>Technology holding</td>
<td>Israel</td>
</tr>
<tr>
<td>Latvia</td>
<td>252</td>
<td>Merz Pharma GmbH &amp; Co. KGAA</td>
<td>Pharmaceuticals</td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>Cytos Biotechnology AG</td>
<td>Pharmaceuticals</td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Codexis, inc.</td>
<td>Industrial biotech</td>
<td>Global</td>
</tr>
<tr>
<td>Estonia</td>
<td>381</td>
<td>University of Tartu</td>
<td>National university of Estonia</td>
<td>Estonia</td>
</tr>
<tr>
<td></td>
<td>101</td>
<td>Skype</td>
<td>Telecommunications</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>Nordbiochem</td>
<td>Biotech</td>
<td>Estonia</td>
</tr>
</tbody>
</table>

Israel, its largest component is two orders of magnitude larger than the second biggest one. This, along with the previous findings, seems to suggest a well-balanced innovation system where a well-structured and state funded source of institutional innovation exists. It can be said that this system is further balanced out by a large number of smaller, startup-like companies supplying their share of patent applications. For Latvia, the difference in components is smaller but still substantial. Estonia can be considered an exception here with Skype being four times smaller than University of Tartu. This deviation from a widely recognized pattern can indicate that Skype acts as an outlier and holds the power to significantly skew the overall statistics for Estonia.

The second interesting observation to make is that of the nature of the companies leading the clusters. Israel’s results offer a balanced
combination of local and global companies as well as a wide spectrum of industries. The Estonian list is similar in structure but narrower in terms of the variety of industries. Latvia, on the other hand, is dominated by German pharmaceutical companies, closely followed by a global biotech company. Although this signals to the existence of a globally competitive R&D potential in the country, it also makes Latvia heavily dependent on the intellectual property policies of these two companies. If the list were to be extended to the top five components for Latvia, it would also include AstraZeneca and Technical University of Riga. Such a dominance of a single industry is likely to cause issues in scaling the ecosystem, as it becomes increasingly difficult to have the growing number of researchers be cutting edge in their field. What can furthermore be considered alarming is the dominance of individuals over companies. Table 4 names the largest components based on the dominant applicant - an entity that is not mentioned among inventors. When that constraint is removed, the results do not change for Estonia or Israel, as only the small components are affected. In the case of Latvia, however, the picture changes dramatically, as in all of the top three components there is just one person at the top of the list who is more connected than even the most connected company. The most notable example for this is the case for the largest component which is largely dominated by Ivars Kalviņš, the Director and Head of the Department of Medicinal Chemistry at the Latvian Institute of Organic Synthesis (Latvian Academy of Sciences, 2012). Such a remarkable dependence on one relatively narrow industry, as well as on one person alone, is unlikely to be sustainable in the long run.

4.4 Summary of the graph analysis

What may be considered as not much of a surprise is that Latvia and Estonia exhibit a very different behavior compared to Israel, both in terms of static and dynamic measures. What is unexpected, however,
is the extent of the differences between the countries, as the differences have proven to be both structural and numerical. Especially in the case of Estonia, which researchers have positioned to being very close to an innovation economy, as one would then expect to find differences that stem from the size of the ecosystem, the number of research institutions and so forth. What the research shows, however, is that only in two areas (internationalization and fragmentation) are there significant structural differences - the curves for Estonia’s internationalization and graph fragmentation are not just located at a different place but also have a different shape. The results of the analysis for Latvia have even more reason to cause concern, as the country exhibits almost no similarity to Israel or Estonia. The key information to take away from this analysis in the context of section 3.2 is that both Latvia and Estonia have significant outliers among their patent applicants. This heavily distorts the picture of the actual innovation output while also pointing out that Latvian innovation is unsustainably dependent on one person alone, as part of a single industry.

5 Implications of the research on policy development

5.1 Current goals and policies

At this stage, the direction that Estonia has for research and development include the following main goals (supported by specific metrics), as set forth by Engelbrecht et al. (2007):

1. the competitive quality and increased intensity of research and development

2. innovative entrepreneurship creating new value in the global economy

3. an innovation friendly society aimed at long-term development
In order to be able to track the implementation of the above-outlined strategy, Estonia has for a number of years now been participating in the Community Innovation Survey, initiated by the European Commission. In addition to this, there is an established tradition of publishing comprehensive reports based on the results of these surveys (Reid et al., 2011; Kurik et al., 2002). These reports provide an analysis of the gathered data but also make a point to recommend improvements to policy makers and companies alike. In the latest report, the following recommendations were made regarding innovation policies by Reid et al. (2011):

1. Analyze the possibility of directing financial support for innovation more towards a small number of key organizations that participate in global value chains
2. Continue supporting nascent companies with high added value in new and developing sectors and increase the investment capability of such companies
3. Increase knowledge transfer to private sector by establishing a flexible body, independent of the research institutions, which would proactively connect companies, experts, higher education and research institutions based on the needs of a specific company
4. Increase support for establishing positions of innovation managers at companies

In light of the results of this study, the recommendations seem to have been based on a number of assumptions. The first recommendation, in particular, suggests increasing support for established organizations with an existing potential for growth. Here, the underlying assumption seems to be that of the U-shaped (Wennekers et al., 2005; Acs et al., 2005) development curve where, in the case of Estonia, the extent of entrepreneurial activity is trending down and the focus is on large, already established players. This model, however,
is not supported by the GCR reports (Schwab et al., 2011; i Martin, 2009) which position Estonia in a phase where the focus should shift to smaller, less well established players. Furthermore, the research conducted by the European Commission has found no evidence for the Estonian levels of entrepreneurship actually dropping.

The assumption behind the second recommendation seems to suggest that an increased knowledge transfer from universities to companies is crucial for achieving the overall goal. Given the vague nature of the strategic goals, this assumption is hard to back up with specific research results. However, in terms of specific indicators, this seems to directly contribute to only one or two out of 18 metrics listed in the strategy document referenced above. There is also indirect evidence for the fact that, in this respect, Estonia already qualifies as a successful country, with its largest patent clusters being led by universities (see table 4 on page 59). Estonian universities also seem to exhibit very different intellectual property policies, hence it is hard to see the grounds for this assumption.

5.2 Recommendations based on the current research results

The first recommendation that can be made, all the while considering the current strategy for Estonia, existing innovation policy analysis and the research done in this thesis, is that a shift in current strategic goals may be necessary. Present goals and their metrics focus on either inputs (R&D investment) or outputs (growth in patent numbers), but little has been said about the structure of the innovation ecosystem itself. Since it is clear that Estonia exhibits structural differences in its research community compared to at least one successful country, it is probably important to focus on achieving this structural change. In the current setting, increases in inputs like R&D investments are not likely to yield an increase in outputs.
The second recommendation is based on the discrepancies found in the empirical data and theories of innovation and entrepreneurship. Current research clearly shows that small, open economies with nascent innovation ecosystems are vulnerable to outliers, heavily dependent on foreign cooperation and suffer from variability in data, all caused by their small sample sizes. There is, however, little academic research available which focus on the issues in these kind of economies or which attempt to understand how small, open economies differ from countries with a different kind of a geopolitical profile. Since the subject of the innovation and entrepreneurship strategy with regards to smaller economies is not very well researched, it may prove to be beneficial to invest more time in further exploration of the field. There are steps being taken in this direction with the Estonian GEM team being established at the time this thesis was being written, but a more fundamental academic research in this area may still be necessary.

If one were to make the assumption that Estonia is about to become an innovation-driven economy, the structure of the innovation policies needs to change as well. On one hand, little is known about the Estonian entrepreneurship ecosystem, as it does not leave many readily observable traces in established systems. This is because, by definition, entrepreneurs as disruptors use different operational principles. On the other hand, not a significant amount is done to understand or further support this ecosystem. Amongst eight different innovation policy measures, only one is directed at supporting nascent entrepreneurship (Reid et al., 2011, table 5.1.1.). If indeed such companies decide to play a crucial role in the Estonian economy, much more effort needs to be directed into understanding and nurturing the economy.
6 Conclusion & further study

As a result, it can be said that the main conclusion of the present study is that based on the research done in this thesis, the two hypotheses presented in section 1.1 cannot be confirmed.

The first hypothesis was about the positive state of the Estonian startup ecosystem and innovation economy in general. Firstly, it proved to be very difficult to assess it separately from the innovation ecosystem; simply because the best available sources of data and patents covered both areas. Secondly, it was discovered that the Estonian innovation economy compares rather poorly with the other countries used in the present analysis. For example, Latvia was the only country that Estonia consistently had an advantage over. Moreover, a drastically improved position in one report, KEI, is more likely a statistical error than a sign of considerable success. Majority of the observable innovation outputs for Estonia exist either thanks to state or EU funded universities, or the one outlier - Skype. This points to the existence of an Estonian innovation ecosystem but also to a distinct absence of an innovation economy.

The second hypothesis was related to academic literature and generally recognized indexes adequately reflecting the state of the Estonian innovation and startup ecosystems. This turned out not to be the case. Different theories and indexes predict different types of behavior, while empirical data seems to contradict the majority of it. It was also found that at least one metric, the number of USPTO patents, which is frequently used in these indexes, is subject to inflation and is generally a very poor indicator of the innovation output of a country.

There were, however, several other interesting results that were uncovered while attempting to confirm these hypotheses.

Firstly, a clear structural difference was found to exist between the patenting ecosystems of Estonia, Latvia and Israel. This finding is significant for pointing out that not only is the patent count a deceptive measure of an innovation output, the level of its deceptive-
ness also varies between countries. In addition to this, small, open economies with nascent innovation ecosystems seem to be most prone to this effect which, as a result, can significantly alter the countries’ statistics.

Secondly, a metric was introduced to assess the fragmentation level of affiliation networks. It was shown to be behaving better than existing metrics, and although the metric still requires some further research and refinement, it was successfully used to detect differences between the patent co-author networks of the three countries.

Thirdly, it was found that the Israeli patent co-author graph not only displayed higher fragmentation than the one of Estonia and Latvia, but that the fragmentation also increased. There seems to be a commonly held belief that the opposite is true, meaning that companies with a wider social network are believed to be more successful, while the present study could clearly point to the contrary.

These three findings can also be used as the main body for further research, as the affects that were discovered in this paper need to be applied to a wider set of countries to be able to determine their validity and exact properties. Visual renderings of the patent graphs created to aid the writing of this thesis, however, seem to indeed be usable for analyzing the behavior of actors within a network. In the case of the Estonian startup ecosystem, for example, University of Tartu finds itself in the center of the graph with a wide range of spinoffs, research teams and companies radiating from it towards the edges of the visualization. At the same time, Tallinn University of Technology has a much smaller, but more tightly knit graph, indicating a tendency to patent directly rather than to push the management of the intellectual property to patent via privately held companies.
References


68


70


