



Identifying Conducive Institutional Determinants of Entrepreneurial and Intrapreneurial Activity using Robust Elastic Net Regression

Bachelor Thesis

With the supervision of dr. W. Liebrechts

Frank Ritchi

1025402 (TU/e), 929981 (TiU)

f.a.ritchi@student.tue.nl

Eindhoven University of Technology / Tilburg University

Jheronimus Academy of Data Science

15th of June 2020



ABSTRACT

By applying Robust Elastic Net regression analysis, this study aimed to identify institutional determinants of both intrapreneurial (GEM EEA) and entrepreneurial (GEM TEA) activity and assess whether these differ and whether they differ from determinants found in previous research with a more limited scope. A relatively large set of 154 independent variables have been included in the dataset, consisting of GEM National Expert Survey variables and WEF Global Competitiveness Index variables, that together define the social, cultural, political and economic context as defined by the GEM. By combining GEM data from five years, 2011–2015, a total of 180 observations could be included. Expecting different determinants for different levels of country development level, distinct models have been developed for two subsets; developing and developed countries. In total four models have been fitted (high-HDI TEA, high-HDI EEA, low-HDI TEA and low-HDI EEA).

In general, the found determinants in the four models have limited overlap with each other, indicating that determinants for EEA and TEA are indeed different, on both development levels and with determinants found in previous research. Interpretation of individual variables must be done with care, given the, considering the model, still limited dataset and capriciousness of the Elastic Net procedure, despite the added robustness. However, some overlap with extant research has been found, in particular signs that cultural indicators associated with individualism and (female) autonomy.

Hard conclusion cannot be drawn based on solely this research, but the study shows that applying machine learning techniques on large multivariate data can be a promising addition, or in some cases, replacement for more traditional research techniques and can help to consolidate an, as described in an influential systematic review, highly fragmented research field.

Keywords: Intrapreneurship, Entrepreneurship, Intrapreneurial activity, Entrepreneurial activity, Corporate Entrepreneurship, Robust Elastic Net. GEM, Global Entrepreneurship Monitor, APS, NES, TEA, EEA, GCI, HDI,

TABLE OF CONTENTS

ABSTRACT.....	i
TABLE OF CONTENTS.....	ii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	iv
1. INTRODUCTION.....	1
1.1 Research Rationale and Scientific and Societal Relevance	1
1.2 Problem Statement and Research Questions.....	2
2. LITERATURE REVIEW.....	3
2.1 Entrepreneurial Activity.....	3
2.1.1 Entrepreneurial Activity.....	3
2.1.2 Intrapreneurial Activity.....	3
2.2 Institutional Conditions and Entrepreneurial Activity.....	4
2.2.1 Institutional Conditions and Entrepreneurial Activity.....	4
2.2.2 Institutional Conditions and Intrapreneurial Activity.....	6
3. METHODOLOGY.....	7
3.1 Determining Variables and Data Sources.....	7
3.1.1 Dependent Variables: Entrepreneurial Activity.....	7
3.1.2 Independent Variables: Institutional Determinants.....	8
3.1.3 Country Development Indicator.....	10
3.2 Dataset Preparation.....	11

3.3 Modeling.....	13
3.3.1 Modeling Considerations.....	13
3.3.2 Elastic Net Regression.....	15
3.3.3 Robust Elastic Net Regression.....	16
3.4 Reproducibility.....	17
4. RESULTS.....	17
4.1 Results.....	17
4.2 Validation.....	22
5. DISCUSSION.....	24
5.1 Implications.....	24
5.2 Limitations and Improvements.....	25
6. CONCLUSION.....	26
7. REFERENCES.....	27
8. APPENDICES.....	31
8.1 Appendix A: Overview of Independent Variables.....	31
8.2 Appendix B: Correlation Matrix of Independent Variables.....	34
8.3 Appendix C: Composition of Dataset.....	35
8.4 Appendix D: Conducive Variables per Model.....	37

LIST OF TABLES

CHAPTER 3: METHODOLOGY

Table 1: NES Components and GCI Pillars.....	9
--	---

CHAPTER 4: RESULTS

Table 2: Top 5 Conducive Variables per Model.....	18
Table 3: Conducive Variables on Multiple Levels.....	19
Table 4: Entire Dataset vs. Subsetting Model Comparison.....	22

LIST OF FIGURES

CHAPTER 3: METHODOLOGY

Figure 1: TEA and EEA rates per HDI-cohort.....	10
Figure 2: Correlation Matrix of Independent Variables.....	13
Figure 3: Regular OLS fit vs Robust Fit.....	16

CHAPTER 4: RESULTS

Figure 4: Parameter Optimization	18
Figure 5: RMSE Comparison.....	20
Figure 6: Fitted Values vs Residuals	23

1. INTRODUCTION

1.1 Research Rationale and Scientific and Societal Relevance

Ever since Joseph Schumpeter developed his first theory on entrepreneurship, Mark I, and coined the term *Unternehmersgeist*, or 'entrepreneurial spirit', to refer to the drive for innovation and technological change of 'wild spirited' entrepreneurs, research into the effects of entrepreneurship on economies and societies and determinants for entrepreneurship has been plentiful and has demonstrated the beneficial effects associated with increased entrepreneurial activity (e.g. Wennekers & Thurik, 1999; Bjørnskov & Foss, 2016). While there has been a significant amount of research on individual factors that characterize entrepreneurs and despite the interest in the beneficial impact of entrepreneurship, from both scientific and policy fields, research into institutional determinants for entrepreneurial activity is still rather limited, as is consensus (Bjørnskov & Foss, 2016). There has been research into institutional determinants of entrepreneurial activity (e.g. Bowen & de Clercq, 2008; Crnogaj & Hojnic, 2016), but most of them have a rather limited scope, focusing on distinct variables instead of a broad set of possible determinants.

Coming back to Schumpeter, it was not until many years later, in his Mark II theory, described in his later work *Capitalism, Socialism and Democracy* (1942) that he argued that (large) companies are agents of innovation, productivity and hence economic growth as well, as they possess large capital resources to invest in R&D that entrepreneurs lack. It has since been argued that both theories are complementary (Fontana, Nuvolari, & Shimizu, 2012), but it can be argued that ever since the actual conception of entrepreneurship theory that the importance of entrepreneurship *within* firms, intrapreneurship, has been rather overlooked.

Research into personal characteristics that are conducive for intrapreneurial activity has been undertaken (e.g. Woo, 2018; Marchiori, Madeira, & Dinis, 2018) but research into determinant of intrapreneurial activity is still almost succinct. Bosma, Stam, & Wennekers, (2011) for example did look at differences between entrepreneurial and intrapreneurial activity on a macro level but did not investigate the impact of (possible different) determinants. As the positive effects of increased entrepreneurship likely also apply to intrapreneurship more insights into institutional determinants of intrapreneurial activity would be relevant and beneficial for researchers and policy makers alike. Considering the fact that research in this field is still limited, the results of this study could be a foundation for further research into this, so far underexposed field.

Furthermore, it is also interesting to see whether institutional determinants for entrepreneurial and intrapreneurial activity differ. Differences could imply that different policies might be needed for stimulating intrapreneurial activity than for stimulating entrepreneurial activity. Determining different determinants for entrepreneurial and intrapreneurial activity could also spur new research in this field.

Besides covering an underexposed part of the research field, this study also aims to contribute to the field by applying a novel research method. While most studies in this field look only at one, or a limited number of, well-defined, hypothesized, variables, this study aims to capitalize on the potential of applying machine learning techniques to handle large datasets with many, possible correlating and non-relevant variables. The main advantage of such methods is that fewer and less detailed assumptions must be made regarding which variables to use as input. As such, new dependencies might be discovered that would likely not be found with traditional research methods. Another advantage is that such techniques also enable researching data over multiple years, adding an important and relevant dimension.

Significant discrepancies between previously found institutional determinants of entrepreneurial activity and results in this study could have significant implications: Besides that this study could provide new insights into determinants of entrepreneurial activity, it could also accelerate the use of similar research methods in new economic/entrepreneurial research as well as challenge previous research, especially research concerning multidimensional data.

1.2 Problem Statement and Research Questions

Several problems and opportunities can be distilled. Firstly, there is little to no previous research into finding institutional determinants for intrapreneurial activity. Related to this, there is little to no previous research into whether institutional determinants for intrapreneurial and entrepreneurial activity differ. Furthermore, traditional research methods are rather limited regarding finding conducive variables in multidimensional data and the opportunities of applying machine learning methods to this field of research are yet to be fully utilized. Considering the societal relevance and impact, it is worthwhile to address these problems and opportunities. This paper aims to contribute to the research field in several manners:

- 1) Find conducive institutional determinants for intrapreneurial activity
- 2) Investigate whether the determinants for intrapreneurial and entrepreneurial activity differ
- 3) Do so by applying a novel advanced machine learning method
- 4) Investigate whether found determinants for entrepreneurial activity differ from those found in previous research.

Based on these research aims the following research questions have been formulated:

- 1) Which institutional determinants are conducive for entrepreneurial activity?
- 2) Which institutional determinants are conducive for intrapreneurial activity?
- 3) How do the institutional determinants for entrepreneurial and intrapreneurial activity differ?
- 4) How do the institutional determinants for entrepreneurial activity differ from those found in previous research?

2. LITERATURE REVIEW

2.1 Entrepreneurial Activity

2.1.1 Entrepreneurial Activity

Defining entrepreneurial activity is no easy task. Many definitions of entrepreneurship have been formulated over the years and has been a topic of debate in the literature (Shane & Venkataraman, 2000). Some of these definition are of a more theoretical nature (Sharma & Chrisman, 1999; Shane & Venkataraman, 2000), while other definitions are stricter and more applicable to be used in quantitative research, such as self-employment or the number of newly registered businesses (Justo, De Castro, & Maydeu-Olivares, 2008); Lassmann & Busch, 2015). The disadvantage of such stringent definitions is that they seldom cover the entire range of entrepreneurship. Later in this research, a definition and methodology will be introduced and discussed that over the past years has become a benchmark in entrepreneurship research, the GEM.

Entrepreneurship has a societal and economical importance; There is ample research indicating that entrepreneurship contributes to economic growth, at least in higher-income countries (Carree & Thurik, 2008; Minniti & Lévesque, 2010; Valliere & Peterson, 2009). According to Baumol (1990) though, a distinction should be made between productive, unproductive and destructive entrepreneurship as only the former contributes to economic growth. Another distinction should be made in this respect between necessity-based and high expectation/growth-potential entrepreneurship according to Valliere and Peterson (2009) and Wong, Ho and Autio (2005).

2.1.2 Intrapreneurial Activity

While Schumpeter might be “the most relevant predecessor in the study of the ‘entrepreneurial businessman’” (Quesada, Onaindia, & Ochoa Laburu, 2011), research into intrapreneurship only took off seriously in the 1980’s. The term intrapreneurship itself was first coined by Pinchot (1985). Other terms include ‘corporate entrepreneurship’ and ‘corporate venturing’. Over the years, intrapreneurship has become a more important and growing part of entrepreneurship research and numerous definitions have been formulated, with one of the most eloquently and all-encompassing being: .

“Intrapreneurship is the set of strategies and practices which a company undertakes to promote, cultivate, and manage the entrepreneurial competencies in the organization to create the context conditions that make feasible the development of new ideas and business projects or the renewal of key ideas upon which the company had been founded.”(Bierto, 2008)

Other definitions vary, but three characteristics seem to be a common factor; proactiveness, innovativeness and risk-taking propensity (Stefanovici, 2012). These are not coincidentally also the most common character traits contributed to individual intrapreneurs (Stefanovici, 2012). According to Hornsby et al. (1993) individual characteristics of intrapreneurs have a main influence on the success of intrapreneurship.

Besides studying individual traits also factors relating to intrapreneurial businesses are being studied. Kuratko et al. (1990) found that the three most important internal factors relevant for the success of intrapreneurship within a company are management support, resources (including time) and organizational structure. Zahra (1991) suggests furthermore that business strategy and external environmental factors like dynamism, heterogeneity and industry competitiveness play decisive roles. Antoncic and Hisrich (2004) large endorse these findings.

Implementing and stimulating intrapreneurship accelerates organizational growth, profitability and renewal as well as helps secure the organization's survival in business (Zahra, 1991). As such it is, like 'traditional' entrepreneurship a driver for growth. This makes the field of intrapreneurship not only relevant to people and business involved in intrapreneurship, but also for society as a whole, and therefore policy makers. However, so far research into drivers of intrapreneurship has been largely focused on either the entrepreneurial employee, business or industry level, while research into possible institutional determinants is until now still nearly succinct. Considering the importance of intrapreneurship, and hence the importance of understanding its drivers, this seems like a missed opportunity.

2.2 Institutional Conditions and Entrepreneurial Activity

2.2.1 Institutional Conditions and Entrepreneurial Activity

According to Terjesen, Hessels, & Li, (2016) in their recent systematic review, research into institutional conditions is the most employed theory in comparative institutional entrepreneurship (CIE) and is applied in at least 24 studies in the last three decades. The approach of the majority of these studies is in line with the description of formal and informal institutions, as formulated by North (1994). The second most employed theory is culture theory, dominated by Hofstede's (2001) five cultural dimensions.

Studies grounded in culture theory in particular have found significant associations between culture indicators and entrepreneurial activity. Autio, Pathak, & Wennberg, (2013) for example found that societal institutional collectivism practices negatively associated with entrepreneurial activity. In line with this, a consistent finding is that individualism is conducive for entrepreneurial activity (Baughn & Neupert, 2003; Mitchell et al., 2002). Studies on institutional (legal, political, economic) conditions provide mixed results (Terjesen et al., 2016). Bankruptcy laws (S.H. Lee, Peng, & Barney, 2007) and taxes and business regulations (van Stel, Storey, & Thurik, 2007) have been identified as antecedents for entrepreneurial

activity, as is “perception of administrative complexities” (Grilo & Irigoyen, 2006). Studies on political institutional determinants have ambiguous results (e.g. Ireland, Tihanyi and Webb, 2008). Unemployment rates negatively impact male, but not female entrepreneurship (Verheul, van Stel, & Thurik, 2006)

Although they see a growing body of evidence regarding country-level antecedents, Terjesen et al. (2016) state that “comparative international entrepreneurship research is highly fragmented, with substantial knowledge gaps related to content, theory and methodology” (pp. 324) and that research into this subject usually considers only one or a few types of institutions. As such, they suggest future research “should examine different types of institutional influences” (pp. 316). Despite the “limited insights into the mechanisms driving and resulting from differences” (pp. 309) between countries as a result of the described fragmentation regarding definitions and institutions, they do however indicate that the field has considerably “matured and progressed” over the last few years. This in line with the findings of Bjørnskov and Foss’s comparative analysis (2016). Another drawback that Terjesen et al., (2016) highlight is that antecedents CIE studies take a static perspective considering institutions and hence are not able to capture institutional development over the years. They therefore argue for more multi-year research into antecedents of entrepreneurial activity.

To a certain extent, their wished are already partly being met, as the number of quantitative country-level CIE studies has increases significantly increased in the last years. An important driver of this is the Global Entrepreneurship Monitor, that collects reliable and consistent survey data of over 80 countries. However, according to Terjesen et al., (2016) only fifteen percent of the studies assess more than one potential determinant. This contributes to the fragmentation in the research field. The GEM has however indeed made it easier to conduct country-level studies with multiple informants. An example is Crnogaj and Hojnik (2016), who have conducted a multivariate analysis based on GEM data and found that greater economic freedom (in the institutional context) positively affects entrepreneurial activity. Another study by Bowen and de Clercq (2007) found that financial and educational activities targeted at entrepreneurship positively affect high-growth entrepreneurship, while corruption has a detrimental effect.

In accordance with Terjesen et al. (2016), this study aims to aid in establishing a broader research framework and to reduce the fragmentation in the research. Furthermore, again in accordance to recommendations by Terjesen et al. (2016), this study will be multi-year. Not discussed by either Terjesen et al. (2016) or Bjørnskov and Foss (2016) is the fact that more traditional statistical analyses cannot cope well with large multivariate data. This research therefore also aims to contribute to the research field in the form of introducing a novel research method in this context. Even without determining concrete and definitive conducive determinants for entrepreneurial and intrapreneurial activity, this research might be a valuable contribution as it could demonstrate advantages of applying machine learning into similar research, either in the form of finding so-far unknown determinants or by aiding to reduce fragmentation

in the research field by helping to identify overarching constructs and as such bridging the gaps between different 'fragments'.

2.2.2 Institutional Conditions and Intrapreneurial Activity

Although there has been research into the effect of external factors on intrapreneurship (e.g. Antoncic & Hisrich, 2003), there is still very little research into institutional determinants of intrapreneurship. A reason for this is that it is a difficult challenge to measure the entrepreneurial activity of established firms (Bjørnskov & Foss, 2016), as most developed measures (e.g. number of patents or R&D expenditures) do not entirely capture all intrapreneurial activity. On the other hand, more abstract definitions, such as Bierto's, (2008) are hard to quantify.

Since 2011, The GEM provides a partial solution, as it since then includes variables in their survey with the aim to measure intrapreneurship. Furthermore, inclusion in the GEM dataset that also aims to measure entrepreneurship in a similar way makes it possible to perform comparative analyses concerning their respective determinants. Bosma et al., (2011) is the first international comparative study of intrapreneurship and independent entrepreneurship. It has found that intrapreneurship is about twice as high in high income countries as compared to low income countries. And that there is a negative correlation between intrapreneurial and entrepreneurial activity. This might be attributed to lesser necessity-based entrepreneurship in higher income countries, as well as the presence of larger firms and positive impact of education on intrapreneurship, as opposed to a negative impact on entrepreneurship.

One of the few studies that found a conducive institutional determinant for intrapreneurship is Morris, Davis, & Allen (1994), that found in a cross-cultural, empirical investigation that the level of individualism has a positive effect on the level of intrapreneurship. This is in line with the impact of individualism on entrepreneurship. While this is an indication that determinants for entrepreneurship and intrapreneurship overlap at least partially, the different distributions of TEA and EEA (GEM's proxies for entrepreneurship and intrapreneurship respectively) indicate that there is also a difference.

Considering the evident gap in the research field is evident, the relevance of the topic, the possibilities that the GEM data provides as well as the possibilities that the novel research method provides, this paper aims to set first steps to identify conducive determinants for intrapreneurial activity. While identified determinants on entrepreneurship might be compared with those found in extant research, this is not the case with those found for intrapreneurial activity. Although it is also not expected that identified variables can be considered conducive with certainty, this research can be valuable as an exploratory analysis upon which further research can build upon. Identifying conducive variables is worthwhile in itself, but particularly when these differ for entrepreneurship and intrapreneurship.

3. METHODOLOGY

3.1 Determining Variables and Data Sources

3.1.1 Dependent Variables: Entrepreneurial Activity

To conduct an analysis on the effect of institutional determinants on entrepreneurial and intrapreneurial activity proxies are needed to act as dependent variables in the analysis. In the past, several measures have been used to measure entrepreneurial activity, such as self-employment or the number of newly registered businesses (Justo et al., 2008; Lassmann & Busch, 2015). It has been argued though that such measures do not serve as good proxies for entrepreneurship (Bjuggren, Johansson, & Stenkula, 2010). Furthermore, relevant to the nature of this research, they do not incorporate a methodology that makes it possible to compare (determinants of) entrepreneurship and intrapreneurship.

The Global Entrepreneurship Monitor (GEM) is an annual assessment of entrepreneurial activity, aspirations and attitudes of individuals across a wide range of countries and is the largest ongoing study of entrepreneurial dynamics in the world (Bosma, Jones, Autio, & Levie, 2008) and provides proxies for both entrepreneurship and intrapreneurship, based on a thorough methodology and large data collection. GEM is a networked consortium of national country teams primarily associated with top academic institutions and carries out survey-based research on entrepreneurship and entrepreneurship ecosystems around the world¹. GEM defines entrepreneurship as any attempt at new business or new venture creation such as self-employment, a new business organization or the expansion of an existing business, by an individual, a team of individuals or an established business².

Its main instrument to measure the level and nature of entrepreneurial activity around the world is the Adult Population Survey (APS). This yearly conducted survey is administered by GEM National Teams to a representative national sample of at least 2000 respondents that answer questions on both an individual and national level. The APS is conducted by independent surveyors, selected by GEM National Teams whose raw data is sent directly to the GEM data team for review, quality check and uniform statistical calculations³. GEM measures entrepreneurial activity by identifying *nascent entrepreneurship* and *owners-managing a new firm*. The sum of the rates of these is defined as Total early-stage Entrepreneurial Activity Rate (TEA rate)⁴.

Since 2011, the APS also includes several measures of intrapreneurship, which it refers to as Employee Entrepreneurial Activity (EEA). The EEA rate is measured as the percentage of people between 18 and 64 years old involved in entrepreneurship in a leading role, either as a percentage of the population or of the workforce. A further distinction is made between current intrapreneurial activity and intrapreneurial activity within the last three years, leading to a number of four EEA-related variables. For this research, the variable 'IPACLTD_ALL' has been selected as the proxy for intrapreneurial activity. This variable

¹ Retrieved from <https://www.gemconsortium.org/about/gem/5>. last accessed on June 09th 2020

² Retrieved from <https://www.gemconsortium.org/wiki/1146>. last accessed on June 09th 2020

³ Retrieved from <https://www.gemconsortium.org/wiki/1141>. last accessed on June 09th 2020

⁴ Retrieved from <https://www.gemconsortium.org/wiki/1149>. last accessed on June 09th 2020

includes all activity within the past three years, as that is arguably a more stable score than the current rate, as a percentage of the total population, to be in line with the TEA rate which is also a proportion of the total population.

It is important to note that by implementing the GEM APS provided TEA and EEA rates as proxies for entrepreneurial activity researchers by default have to adopt the definitions of entrepreneurial activity on which these proxies are based. Although these definitions and the assumptions on which the proxies are based are, as any other definitions and assumptions, up to debate, there are a number of considerations as to why using the GEM proxies can be considered a preferred approach. First of all, the methodology of GEM has been thoroughly validated and is constantly being refined and expanded ⁵. Combined with the fact that the GEM data collection has been vast and consistent over a large number of years this has resulted in GEM becoming one of the leading sources for data on research into entrepreneurial activity (Bjørnskov & Foss, 2016). As such, GEM's definitions and proxies have been applied in plentiful research in this field ⁶. Furthermore, the incorporation of proxies for both entrepreneurial and intrapreneurial activity in one comprehensive dataset makes the GEM data uniquely applicable in research into the differences between them, as in this research.

Based on the above-mentioned considerations the Total early-stage Entrepreneurial Activity Rate and the variable 'IPACLTD_ALL' (*% of 18-64 years in the population that has been involved in entrepreneurship in a leading role in the past three years*) have been selected as proxies for entrepreneurship and intrapreneurship. These will be referred to as TEA rate and EEA rate respectively from hereon.

3.1.2 Independent variables: Institutional Determinants

One of the advantages of machine learning analysis is that it can generally better cope with many, possible correlating or non-relevant, variables compared to more traditional statistical methods. As such, less assumptions need to be made when it comes selecting possible determinants for entrepreneurial and intrapreneurial activity and a detailed rationale for selected determinants becomes less relevant and might even be unnecessarily limiting. As discussed earlier, the advantages of this approach are the possibility to find determinants that might not have been selected in a more traditional approach, or that might be obscured by correlating determinants.

The GEM does not only provide good proxies for entrepreneurial activity, but also provides an overview of, and data on, the social, cultural, political and economic context relating entrepreneurial activity, and as such provides a comprehensive and relevant overview of candidate determinants of entrepreneurial activity. In its latest conceptual framework (version 3, 2015-present) ⁷. GEM divides this context into 'Entrepreneurial Framework Conditions' and National Framework Conditions'.

⁵Retrieved from <http://gem-consortium.ns-client.xyz/about/wiki>, last accessed on June 09th 2020

⁶For an non-exhaustive overview of papers that have used GEM data, see PDF: <https://www.gemconsortium.org/research-papers-export%2Ffull&usg=AOvVaw188wwlwYKMMdjWU4f5Vwlc>

⁷For the full GEM conceptual framework, see: <https://www.gemconsortium.org/wiki/1148>, last accessed on June 09th 2020

Entrepreneurial Framework Conditions are conditions that are considered to be conducive (or a hinder) to entrepreneurial activity. According to Bosma et al., (2008). they constitute “the necessary oxygen of resources, incentives, markets and supporting institutions for the creation and growth of new firms”. GEM aims to capture these framework conditions, and other related topics, in its National Expert Survey which, as the name suggests, is a survey among experts rather than a representative population sample. As the APS, the NES collects data in a harmonious and consistent manner, making the gathered data useful for research over time and between different countries. In order to create an as comprehensive as possible dataset, the NES consists of nine components that each aim to capture one of the Entrepreneurial Framework Conditions, in a total of 62 variables ⁸.

Besides these Entrepreneurial Framework Conditions there are many other, formal and informal, variables that might impact entrepreneurial and intrapreneurial activity. GEM refers to these as National Framework Conditions, that are defined by the variables of the World Economic Forum’s Global Competitiveness Index (GCI). The GCI assesses competitiveness in 144 economies and provides insight into their productivity drivers and is considered the most comprehensive worldwide assessment of national competitiveness (Levie et al., 2015). The index consists of three sub-indices containing twelve ‘pillars of competitiveness’ to assess a broad variety of aspects of competitiveness. In total 110 variables are included in the Global Competitiveness Index ⁹.

An overview of the nine components of the NES and 12 pillars of the GCI can be found in Table 1. An overview of all independent variables can be found in Appendix A.

Table 1: NES Components and GCI Pillars

National Expert Survey Components	Global Competitiveness Index Pillars
1. Entrepreneurial Finance	1. Institutions
2. Government Policy	2. Infrastructure
3. Government Entrepreneurship Programs	3. Macroeconomic Environment
4. Entrepreneurship Education	4. Health and Primary Education
5. R&D Transfer	5. Higher Education and Training
6. Commercial and Legal Infrastructure	6. Goods Market Efficiency
7. Entry Regulation	7. Labor Market Efficiency
8. Physical Infrastructure	8. Financial Market Development
9. Cultural and Social Norms	9. Technological Readiness
	10. Market Size
	11. Business Sophistication
	12. R&D Innovation

The social, cultural, political and economic context captured in the combination of the WEF GCI and the GEM NES data provides a comprehensive and balanced overview of relevant variables but does not necessarily capture all possible conducive determinants. This research will only assess the GEM NES and WEF GCI data, but it should be kept in mind that with the used research method, other possible

⁸ For a more detailed description of the components, see: <https://www.gemconsortium.org/wiki/1142>, last accessed on June 09th 2020

⁹ For a more detailed description of the WEF GCI composition, <http://www3.weforum.org/docs/GCR2017-2018/04Backmatter/TheGlobalCompetitivenessReport2017-2018AppendixA.pdf>

determinants from other source could be added to the analysis, even if they are not relevant of (highly) correlating. This might be especially interesting and worthwhile if analysis of the context as defined by the GEM points to determinants that might be underexposed in the GCI and NES variables.

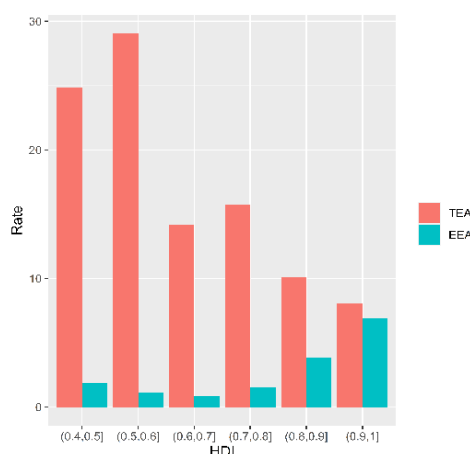
3.1.3 Country Development Indicator

Intuitively, it seems plausible that determinants may differ between different levels of development and it is interesting to see whether some variables are conducive on only some levels of development and some are conducive irrespective of development level. To determine the effect of country development level on the conduciveness of institutional variables on TEA and EEA rates, an indicator for countries' level of development has been added to the data.

The WEF's Global Competitiveness Report has its own country development indicator, as it classifies countries by their stage of development (factor-driven, efficiency-driven and innovation-driven economies). This metric however is solely based on GDP per capita. The Human Development Index, developed by the United Nations Development Programme, however is a more comprehensive measure of a country's level of development as it incorporates not only a country's per capita income, but life expectancy and level of education as well. It is therefore considered an appropriate proxy for a country's level of development (Dervis & Klugman, 2011) and as such it will be used as a proxy for level of development in this research as well ¹⁰.

Exploratory data analysis seems to support the hunch described above. Figure 1 shows relatively high TEA rates for countries with a (very) low level of development and that this rate declines sharply as countries become more developed. It is important to note that the TEA rate encompasses all levels of entrepreneurial activity, not only entrepreneurship with an innovative nature.

Figure 1: TEA and EEA rates per HDI-cohort



¹⁰ For a detailed description of the HDI methodology, see: <https://www.gemconsortium.org/wiki/1142>, last accessed on June 09th 2022 <http://hdr.undp.org/sites/default/files/oc12.pdf>

A likely factor for high TEA rates in developing countries is that people in countries with a low development level have to be more self-reliant and likely work mainly in the primary sector, as the economies of these countries mostly have a poorly developed secondary and (especially) tertiary sector. This is in line with Hessels, van Gelderen and Thurik (2008) and McMullen, Bagby and Palich (2008) who have found that GDP per capita is negatively associated with necessity-motivated entrepreneurship.

In highly developed countries, with a more sophisticated tertiary sector and diversified economy, entrepreneurial activity seems to be born less out of necessity, but more out of available economic opportunities. Mersha, Sriram, and Hailu (2010), discuss the differences between these forms of entrepreneurship. Their findings hint to different institutional determinants for TEA.

On the other hand, EEA rates seem to only grow significantly once a relatively high level of development has been achieved. It is interesting to see which changing conditions correlate with this and facilitate EEA rate growth. Furthermore, it is interesting to see whether the determinants of EEA remain constant or change depending on development level.

Adding the HDI scores as a control variable would capture some of the variance, but the model would then output one set of coefficients for all countries and as such it could not be determined whether the sets of conducive determinants differs between different levels of country development.

The dataset has therefore been split on HDI-scores into two subsets; developing and developed countries. The splitting value has been set on 0.800. This is a rather arbitrary value, but it seems appropriate based on the data in Figure 1. Splitting on more levels might result in an even better understanding of how institutional determinants of TEA and EEA change over time, but the dataset was deemed too small to create more subsets.

3.2 Dataset Preparation

As discussed above, data used in the analysis has been retrieved from several sources :

- GEM Adult Population Survey – National Level
- GEM National Expert Survey – National Level
- WEF Global Competitiveness Index
- UNDP Human Development Index

The aim in the creation of the dataset has been to include as many instances and variables as possible, given the available data. There have been two limiting factors regarding the retrieval of the dependent variables TEA and EEA. Firstly, full GEM datasets only become publicly available after several years. The latest APS and NES datasets that could therefore be used were those of 2015. Secondly, GEM only started

measuring EEA from 2011 onwards. As such the dataset covers five years, from 2011 to 2015. Because EEA has only been an optional variable, a relatively large set of instances (78 out of 322) did not have a corresponding EEA value, rendering these instances useless, making discarding them inevitable.

Next, the HDI data was merged with the APS data. HDI data is available for a large amount of countries, but two disputed countries, Kosovo and Taiwan, and an unincorporated territory of the United States, Puerto Rico, that were part of the GEM data were not included in the HDI data. These instances were removed, leaving 237 instances.

Although the GEM NES data covers mostly the same countries as the GEM APS, merging these datasets led to a loss of six instances, as six countries present in APS 2011 data were not present in the NES data of 2011, leaving 231 instances. NES variables A07 and A08 were only available in data of one year though, so these were discarded as well. The remaining 52 variables were added to the dataset.

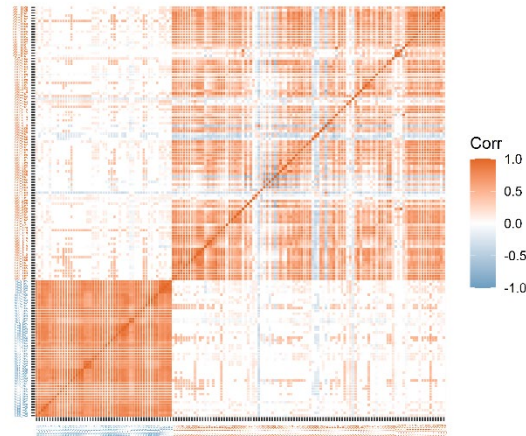
Merging GCI data was more cumbersome due to the presence of 2259 missing values. After identifying several rows and columns with a high number of missing values, all rows that contained many missing values (more than 20% of total values) were removed, leaving only 103 missing values. Several imputation handling strategies have been considered, but without assumptions to base the imputation on, these missing values were ultimately discarded as well.

Clear and consistent labeling of variables and country names increases the interpretability of the results and prevented accidental exclusion or inclusion of duplicates into the dataset.

The final dataset consists of 180 instances with in total 69 countries. With 52 NES-variables and 102 GCI-variables, there are a total of 154 independent variables included in the dataset. 31 countries have been classified as developing ($HDI \leq 0.800$), 34 as developed ($HDI > 0.800$) and 4 moved from developing to developed over the course of the period. In total 85 instances fall in the 'Low-HDI' subset whereas 95 instances fall in the 'High-HDI' subset.

As it seems plausible that many of the included variables are correlated, the correlation between the predictors has been investigated by developing a correlation matrix. This matrix (Figure 2) shows all (significant) correlations. Interestingly, but perhaps not surprisingly, the plot clearly identifies the two distinct parts of the data; the NES variables in the lower left quadrant and the GCI variables in the upper right quadrant. There is a large amount of significant correlations within these quadrants, but much less between the GCI and NES datasets. This seems to be in line with the GEM framework, that considers the Entrepreneurial Framework Conditions and the National Framework Conditions to be complementary. Despite the presence of some highly correlating variables, no variables have been removed based on this high correlation, considering that this research applies an approach that is deemed to cope well with highly correlating variables. As such, removing variables could only have a detrimental effect on the results. A full-scale version of the matrix can be found in Appendix B.

Figure 2: Correlation Matrix of Independent Variables



A full and detailed description of the used data and data preparation steps, including the full code, can be found in the accompanying *Research Compendium*. The full dataset can be accessed via a link in the compendium to a GitHub repository. A list of countries / instances, including corresponding TEA, EEA and HDI scores can be found in Appendix C.

3.3 Modeling

3.3.1 Modeling Considerations

Considering that the dataset consists of many numeric predictors and two numeric dependent variables, the most appropriate candidate analysis was deemed a form of advanced multivariate regression analysis, as such a method can identify and quantify the impact of independent variables on entrepreneurial activity and predict continuous valued output. As the aim of this research is to find conducive variables, the applied method needs to be able to offer feature selection, i.e. be able to set coefficients to zero.

Another consideration is the fact that the dataset consists of a large number and variety of independent variables, with many of them correlating. Multicollinearity might therefore be a factor. Although multicollinearity does not reduce the predictive power (e.g. RMSE) of the model, it does affect the calculations of the effects of individual predictors. While a collinear model might still have significant predictive power, it might not give valid results about any of the individual predictors. Hence, applying a method that can handle possible multicollinearity well is imperative.

Furthermore, the number of independent variables (154) rivals the number of instances (180) and adding just one other year of data to the dataset would lead to a dataset where the number of predictors exceeds the number of observations ($n \ll p$). Regular linear regression cannot properly handle such 'high-dimensional / low observations' datasets, as there are insufficient degrees of freedom to estimate the full regression model. Besides this issue, adding many independent variables to a linear regression model leads to a continuous increase in variance, regardless of the significance and size of the true coefficients of the added predictors. A model that utilizes some form of regularization and could shrink some less important or non-significant variables would introduce extra bias, but would most likely also decrease variance by a larger degree, decreasing the overall error of the model and reducing the risk of overfitting.

A regression method that can cope well with multicollinearity due to regularization is Ridge regression. Whereas the standard OLS loss function is:

$$(\beta_0, \beta_1, \dots, \beta_p)^T = \underset{\beta_0, \beta_1, \dots, \beta_p}{\operatorname{argmin}} \left[\frac{1}{2N} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 \right] \quad (1)$$

Ridge regression adds a penalty function (L2 penalty) to OLS.

$$(\beta_0, \beta_1, \dots, \beta_p)^T = \underset{\beta_0, \beta_1, \dots, \beta_p}{\operatorname{argmin}} \left[\frac{1}{2N} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right] \quad (2)$$

The penalty function penalizes large feature weights values by penalizing the squares of the coefficients. Cross validation is used to determine the most optimal regularization parameter. However, although it minimizes the impact of irrelevant variables on the trained model, it will not shrink the coefficients to zero and will therefore not completely remove them from the model. As such, Ridge regression it is not very suitable for feature selection, which considering the aim research is a significant drawback.

Lasso regression applies a different penalty term (L1 penalty), that penalized the absolute size of the coefficients, rather than the square. The Lasso loss function is as follows:

$$(\beta_0, \beta_1, \dots, \beta_p)^T = \underset{\beta_0, \beta_1, \dots, \beta_p}{\operatorname{argmin}} \left[\frac{1}{2N} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right] \quad (3)$$

The difference with the Ridge regression is that the Lasso regression penalty function can set coefficients to exactly zero, meaning that these features will be removed from the model. A stronger penalty leads to more coefficients being pushed to zero. As Lasso regression can eradicate features completely from the model it is suited for dimensionality reduction and feature selection and can deal well with datasets containing variables that might not have an impact on the dependent variable at all, as is the case in this analysis. However, it does not cope well with multicollinearity which, considering the

many correlating variables in the dataset, is a significant drawback. Another drawback is that Lasso does not cope well with $n \ll p$ datasets.

Thus, both methods have significant drawbacks; Ridge cannot shrink coefficients to zero and Lasso cannot cope well with correlating variables and $n \ll p$ datasets. A mixture of both methods, Elastic Net regression, combines the advantages of both methods while simultaneously reducing their respective drawbacks.

3.3.2 Elastic Net Regression

Elastic Net is a rather novel regularization and variable selection method, proposed by Zou and Hastie (2005), that uses a hybrid of both aforementioned regularization methods (mixed-L1-L2-regularization) as it consists of the standard OLS loss function and both Lasso (L1) and Ridge (L2) penalties:

$$(\beta_0, \beta_1, \dots, \beta_p)^T = \underset{\beta_0, \beta_1, \dots, \beta_p}{\operatorname{argmin}} \left[\frac{1}{2N} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \left(\frac{1-\alpha}{2} \beta_j^2 + \alpha |\beta_j| \right) \right] \quad (4)$$

Thus, Elastic Net penalizes a mix of both absolute and square sizes of the coefficients. As such it can better cope with correlated variables than Lasso can while still be as useful for feature selection as it too can set coefficients to zero, in contrast to Ridge regression.

The Elastic Net penalty is a convex combination of the Ridge and Lasso penalties. In addition to the regularization parameters λ_1 and λ_2 , the Elastic Net function therefore includes another parameter, α . This tuning parameter determines the relative size of the two penalty terms. With $\alpha = 1$ the Elastic Net model is similar to a Ridge regression while it is similar to a Lasso regression when $\alpha = 0$. Alphas between 0 and 1 result in a hybrid form. In simplified form, this objective function looks as follows:

$$OLS + \alpha * \text{Ridge penalty} + (1 - \alpha) * \text{Lasso penalty for } \alpha \in [0,1]$$

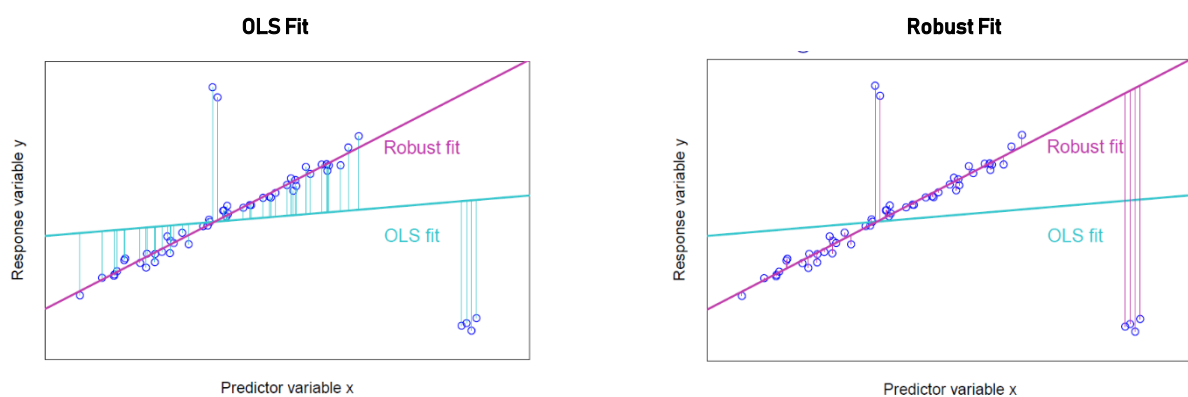
As with the regularization parameters, the optimal tuning parameter α is found by applying cross validation. Real world data and a simulation study show that the Elastic Net often outperforms the Lasso, while enjoying a similar sparsity of representation (Zou & Hastie, 2005).

As Elastic Net regression seems to be the most appropriate approach given the modeling consideration and its described potential, it has been selected for this research.

3.3.3 Robust Elastic Net Regression

Although Elastic Net regression can handle multicollinearity and $n \ll p$ data quite well, regression analysis on a relatively small number of instances and large number of predictors remains challenging. Besides multicollinearity and the $n \ll p$ issue, outliers and uninformative variables can have a relatively large impact. A more robust estimation would be less dependent on outliers and other noise and instead of the best fit through all data points aim to have an optimal fit through only most of the data points, accepting larger residuals on the outliers for the benefit of a model that fits the majority of the data better. A schematic overview of a robust fit can be found in Figure 3.

Figure 3: Regular OLS fit vs Robust Fit



Kurnaz, Ortner and Filzmoser (2017) have developed a fully robust version of the Elastic Net estimator, which is based on the idea of repeatedly applying the non-robust classical estimators to data subsets. Outlier-free subsets can be identified efficiently by applying a sparse least trimmed squares regression as developed by Alfons, Croux and Gelper (2013) after which appropriate tuning parameters for the Elastic Net penalties can be selected. A final reweighting step improves the efficiency of the estimators. According to Kurnaz, Ortner and Filzmoser (2017), simulation studies show good performance in real data examples and superiority in comparison to non-robust and other robust estimators.

An accompanying `|R|`-package, `enetLTS`, utilizes the developed procedure. This `|R|`-package has been used in this analysis as it has been considered to be the most appropriate approach¹¹. The procedure includes cross-validation for parameter optimization and repeats this cross-validation for more stable results. The reweighting step to improve the efficiency of the estimators is the same procedure as developed by Alfons et al., (2013).

For the procedure to produce meaningful results, the independent variables need to be standardized. Standardization is included in the `enetLTS` procedure. Furthermore, it should be noted that, whilst the regularization parameter for the L1 and L2 regularization, λ_1 and λ_2 , could be different, the `enetLTS`

¹¹For documentation of the `enetLTS` `|R|`-package, see: <https://cran.r-project.org/web/packages/enetLTS/enetLTS.pdf>

procedure identifies one single value for both. This is in line with other Elastic Net packages, such as `glmnet` and `elasticnet`¹²¹³.

The dataset has been split into a training set, used to fit the models and a testing set, used to assess the performance of the models on new data, hence enabling validation of the results. A random sample of 70% of the dataset has been assigned to the training set, the remainder of 30% has been assigned to the testing set. All models have been trained and tested on equal training and testing sets.

3.4 Reproducibility

For scientific research to play a relevant role in the scientific community it is of utmost importance that the findings of the research can be validated, challenged and scrutinized by scientific peers, for example by comparing the results of the used research with other research methods or by applying the same research method on new data. Even in the case of a validated and accepted research method, research reproducibility can be used to broaden the research by including new observations, something that is especially relevant in a research like this one where the number of variables are larger than the number of observations (when split on development level).

It is therefore imperative that replicability and reproducibility play an integral role in the entire research process. As such, this research has been conducted with reproducibility constantly in mind. To ensure full reproducibility all steps needed to achieve this need to be documented. The procedures used in this research can be found in the accompanying *Research Compendium*. The compendium includes a link to a GitHub repository containing all raw data sources and |R|-code.

4. RESULTS

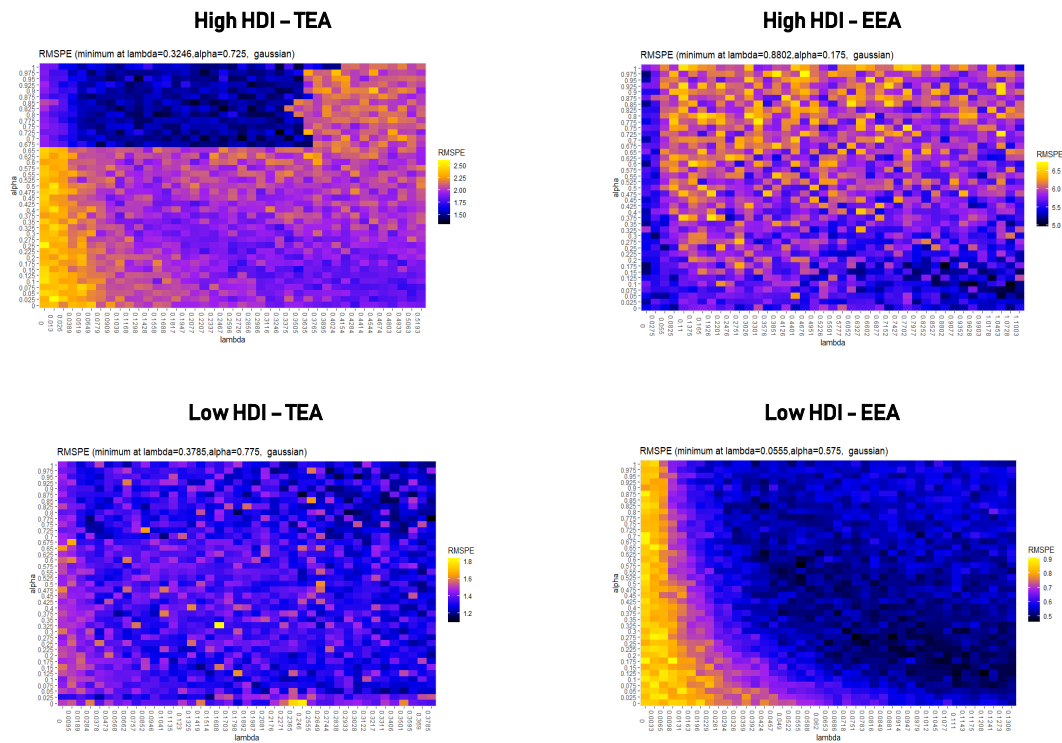
4.1 Results

A total of six models have been fitted; A model on the entire dataset and two models for both levels of country development for both dependent variables. The models on the entire dataset have been part of the validation process and will therefore be discussed in the validation section. Each model outputs a list of significant coefficients with corresponding values. The performance metric used is the root mean square error (RMSE). As discussed, optimal parameters must be determined for regularization parameter *lambda* and tuning parameter *alpha*. Figure 4 shows the performance of all assessed combinations of lambdas and alphas for each of the four models (metric: root mean square prediction error, or RMSPE).

¹²For documentation of the `glmnet` |R|-package, see: <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>

¹³For documentation of the `elasticnet` |R|-package, see: <https://cran.r-project.org/web/packages/elasticnet/elasticnet.pdf>

Figure 4: Parameter Optimization



All four models have a different optimal combination of alphas and lambdas. Furthermore, the plots indicate clearly that small variations in parameter values can lead to vastly different outcomes. Hence, the number of cross-validation folds has been increased from the default value of 5 to 10.

The full lists of conducive variables for each of the four models can be found in Appendix D. In total, 18 variables have been determined to be conducive for TEA, and 17 for EEA in the developed countries subset. In the developing countries subset, these numbers are 34 and 13 respectively. Below, in Table 2, the top five conducive variables of each model are listed, in descending order of their absolute values.

Table 2: Top 5 Conducive Variables per Model

High HDI – TEA		
Code	Variable Description	Coefficients
GCI_9.03	FDI and technology transfer	0.8791185479
NES_I03	The national culture encourages entrepreneurial risk-taking	0.6209032959
NES_B05	The amount of taxes is NOT a burden for new and growing firms	- 0.4908398789
GCI_2.05	Quality of air transport infrastructure	- 0.4655406759
GCI_1.21	Strength of investor protection	- 0.3437431756

High HDI – EEA

Code	Variable Description	Coefficients
GCI_11.09	Willingness to delegate authority	0.846635329
GCI_6.08	Agricultural policy costs	0.322325235
GCI_12.01	Capacity for innovation	0.255388870
GCI_9.03	FDI and technology transfer	0.218281439
GCI_5.06	Internet access in schools	0.114413640

Low HDI – TEA

Code	Variable Description	Coefficients
GCI_12.06	Availability of scientists and engineers	- 1.8328420997
GCI_6.01	Intensity of local competition	1.7916711822
GCI_7.01	Cooperation in labor–employer relations	1.1784710324
GCI_6.15	Degree of customer orientation	0.8904379021
GCI_2.03	Quality of railroad infrastructure	- 0.7307355453

Low HDI – EEA

Code	Variable Description	Coefficients
GCI_7.10	Female participation in labor force	0.2805800153
GCI_6.02	Extent of market dominance	- 0.1309822159
GCI_10.02	Foreign market size index	- 0.0982966754
GCI_12.01	Capacity for innovation	- 0.0874052501
GCI_8.05	Venture capital availability	- 0.0599826620

Besides finding the most conducive variables per model, it is also valuable to determine which variables are conducive for both TEA or EEA, irrespective of development level, or are conducive for both TEA and EEA on a single development level. Table 3 shows the variables that have been identified as conducive on multiple levels.

Table 3: Conducive Variables on Multiple Levels

Conducive Variables for TEA on both Development Levels			
Code	Variable Description	Coef. High HDI	Coef. Low HDI
NES_D05	Business and mgmt. education provide adequate preparation for starting & growing new firms	- 0.2228536408	0.1409262606
GCI_1.21	Strength of investor protection	- 0.3437431756	0.2161870427
GCI_2.05	Quality of air transport infrastructure	- 0.4655406759	- 0.0096620077
GCI_4.03	Business impact of tuberculosis	0.0223444060	- 0.0008455736
GCI_5.01	Secondary education enrollment rate	- 0.0052848694	- 0.0091658913
GCI_7.02	Flexibility of wage determination	0.2197546626	0.3179739311
GCI_9.06	Internet bandwidth	0.0002325376	- 0.0062910886

Conductive Variables for EEA on both Development Levels

Code	Variable Description	Coef. High HDI	Coef. Low HDI
GCI_2.08	Mobile telephone subscriptions	- 0.002738772	- 0.0006568882
GCI_12.01	Capacity for innovation	0.255388870	- 0.0874052501

Conductive Variables for TEA and EEA for Developed Countries

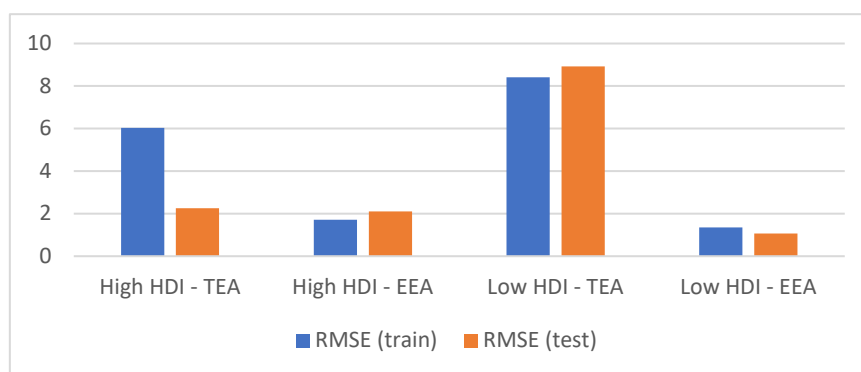
Code	Variable Description	Coef. TEA	Coef. EEA
GCI_2.05	Quality of air transport infrastructure	- 0.4655406759	- 0.099337612
GCI_5.06	Internet access in schools	0.0511977501	0.114413640
GCI_9.03	FDI and technology transfer	0.8791185479	0.218281439

Conductive Variables for TEA and EEA for Developing Countries

Code	Variable Description	Coef. TEA	Coef. EEA
GCI_2.08	Mobile telephone subscriptions	- 0.0032073222	- 0.0006568882
GCI_6.02	Extent of market dominance	- 0.1243480969	- 0.1309822159
GCI_7.02	Flexibility of wage determination	0.3179739311	0.0128614034
GCI_7.10	Female participation in labor force	0.4290972606	0.2805800153
GCI_9.06	Internet bandwidth	- 0.0062910886	- 0.0004120142
GCI_10.02	Foreign market size index	- 0.1889320903	- 0.0982966754

After training the models have been validated on a test set. In an ideal situation, the RSME's for the training set and testing set are close to equal. As can be seen in Figure 5, this applies to three of the four models, the model for determining conducive variables for TEA in developed countries has a significant discrepancy between the two values, surprisingly the RSME for the test set is lower. No explanation has been found for this discrepancy, but the rather limited size of the dataset (the test set for developed countries consists of 29 instances) might play an important factor.

Figure 5: RMSE Comparison



On both levels of HDI, the most important variables differ for TEA as well as EEA. Variable GCI 9.03 (*FDI and technology transfer*) is the only variable that is present in both high-HDI top-5's (TEA and EEA). Another variable that appears in two top-5's is variable GCI 12.01 (*Capacity for innovation*). This variable however is positive for high-HDI EEA and negative for low-HDI EEA. Interestingly, only two NES-variables are included in the four top-5's, indicating that the national framework conditions might be a better indicator for entrepreneurial activity than the entrepreneurial framework conditions.

Several informal institutions appear in the top-5's. *Willingness to delegate authority* was found to be the most conducive variable for high-HDI EEA, with a large positive coefficient. *The national culture encourages entrepreneurial risk-taking* was found to be the most conducive variable for high-HDI TEA, also with a large positive coefficient. *Female participation in labor force* was found to be the most conducive variable for low-HDI EEA, and in fact the only one with a positive coefficient. *Female participation in labor force* has also been found to be conducive for low-HDI TEA. Other variables in the respective top-5's are deemed to be either too small to draw conclusions upon or are hard / impossible to interpret (e.g. *Agricultural policy costs*).

Table 3 shows variables that are deemed to be conducive on at least two models (incl with a different direction). An issue arises when trying to assess which variables to discuss; some variables are in line with extant literature and intuition (e.g. FDI and technology transfer being conducive for both high-HDI TEA and EEA) while many other variables are much less intuitive (e.g. *Quality of air transport infrastructure* having a negative impact on high-HDI TEA and EEA). As such, apart from the ones discussed above, individual variables will not be interpreted and discussed in this study, as that would likely result in being forced to make unverified assumptions and 'cherry-picking' of plausible variables in order to interpret the distinct variables. As a result, these variables will also not be compared to variables of previous research as comparing them would not be a responsible approach, considering the uncertainty of the found results.

In line with extant literature, cultural aspects seem to play indeed important roles. *Willingness to delegate authority* and *Encouragement of entrepreneurial risk-taking* could be in line with granting and taking individual responsibility. As discussed earlier, individualism seems to be positively related to both entrepreneurial and intrapreneurial activity in extant research (Baughn & Neupert, 2003; Mitchell et al., 2002; Morris et al., 1994). Furthermore, *Female participation in the labor force* being conducive for both TEA and EEA in low-HDI countries seems to match with Verheul, van Stel, & Thurik's (2006) finding that unemployment rates negatively impact male, but not female entrepreneurship. These determinants strengthen the indications in the extant literature that cultural aspects play an important role on all types of entrepreneurial activity and on all levels of development.

Without going into too much detail regarding individual determinants, considering the uncertainty, it does seem clear that determinants differ significantly per model. This is an indication that determinants indeed differ for TEA and EEA, and for both levels of development.

4.2 Validation

The developed Elastic Net model has been applied not only to the developed and developing countries subsets, but also to the entire dataset in order to assess the decision to split on country development model. Summary data of this comparative analysis can be found in Table 4. The differences between the models' RSME values are relatively small, just as the discrepancy between the RSME's on the training and testing sets. The number of outliers, instances not used in fitting the model and therefore with relatively high residuals are also similar (considering the complete dataset is as big as the high-HDI and low-HDI datasets combined). The fact that the RMSE's for the three TEA models and the three EEA models are close together gives confidence that the two subsets have not become too small. The only factor that hints to a too small dataset is the fact that the testing RMSE of the high-HDI, TEA model is quite smaller than the training RMSE. On the other hand, the models on the entire dataset produce a very large number of coefficients, the TEA model no less than 104, rendering interpretation impossible. As such, it is considered that subsetting on HDI level is a valid approach. A larger dataset would however lessen concerns of a too small dataset.

Table 4: Entire Dataset vs. Subsetting Model Comparison

Model	Coefficients (n)	RMSE (train)	RMSE (test)	Outliers
All HDI - TEA	104	6.81	6.97	26
All HDI - EEA	33	1.41	1.54	5
High HDI - TEA	18	6.03	2.25	14
High HDI - EEA	17	1.72	2.10	6
Low HDI - TEA	34	8.41	8.93	12
Low HDI - EEA	13	1.36	1.06	9

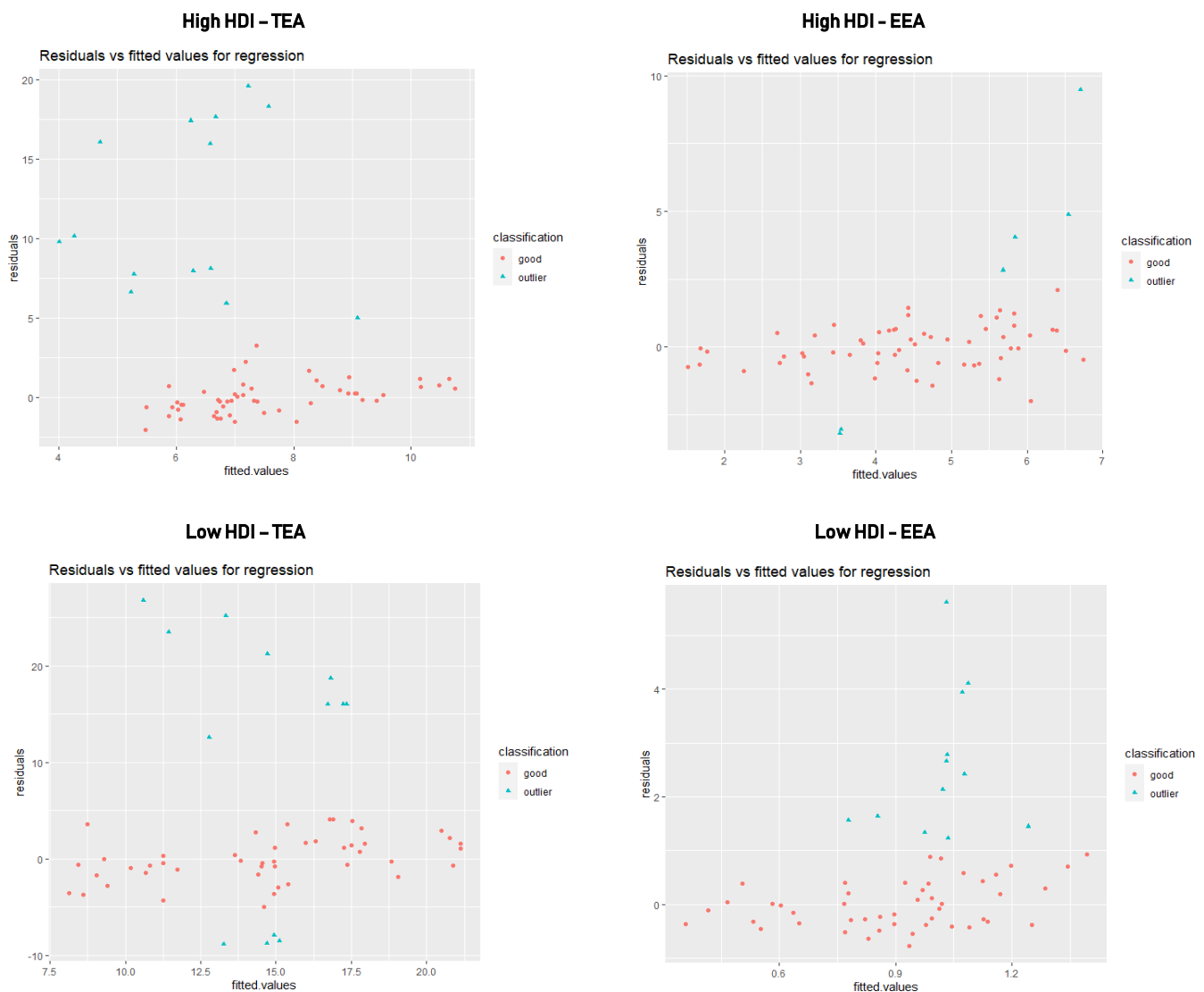
The Elastic Net procedure only outputs a list of coefficients deemed significant and the RMSE of the model. Contrary to standard regression methods the Elastic Net procedure (and other regularized regression methods) does not offer standard errors and derived indicators of significance like p-values, confidence intervals and R^2 -value, as standard errors are not meaningful for strongly biased estimators such as the ones obtained with penalized regression methods as variance only explains just a small part of the total error. Bootstrap-based confidence intervals could be obtained, but these too would be based on assessing the variance, rather than bias. A significance test for regularized regression has been proposed by a research group including the inventor of the Lasso method (Lockhart, Taylor, Tibshirani, & Tibshirani, 2013) and has been developed into an R-package, `selectiveInference`¹⁴, but the assumptions on which the procedure is based is so stringent that such a method has not been considered.

¹⁴ For documentation of the selectiveInference R-package, see: <https://cran.r-project.org/web/packages/selectiveInference/selectiveInference.pdf>

As the Elastic Net procedure is capable of selecting significant and robust variables nonetheless, the lack of significance testing options is not relevant regarding the validity of the outcomes. However, predictors chosen by the Elastic Net procedure are, as with any feature selection method, highly dependent on the dataset, especially in cases of highly correlating predictors and large grouping effects. Due to the instability in feature selection this causes, the interpretation of the results should be done with care as some of the selected variables might act as a proxy for another, or several other, correlating variables.

In Figure 6, the fitted values are plotted against the residuals. The effects of the robustness procedure can be clearly seen as the models aim to fit only a subset of the data (min. 75%), accepting larger deviations for the other instances. That this procedure is successful can be derived from the distribution of the instances classified as 'good' (red values), that shows a constant variance, as the values are all in a relatively narrow horizontal band, centered around $y = 0$. Because of the added robustness, the results are more resistant to variations in the data and better applicable to new data as a non-robust model,

Figure 6: Fitted Values vs Residuals



5. DISCUSSION

5.1 Implications

This study has for the first time applied Elastic Net regression for feature selection in the field of entrepreneurship research. While the procedure has produced validated models with a good fit and a relatively limited number of variables deemed conducive, and the robustness of the method strengthens the quality of the findings, interpreting the variables, and hence assess the differences between determinants for TEA and EEA, is difficult.

An issue with penalized regression methods, and especially with the used ratio between instances and independent variables, is that they are quite sensitive to the data and the fitting parameters of the models. As such, interpretation of the results is tricky and should in any case be done with much care. While it is with the combination of found predictors and coefficients indeed possible to make decent predictions for new instances, the results are deemed too uncertain for exact feature selection. Does the number of mobile phone subscriptions for example truly have a negative impact on EEA rates? And is *Strength of investor protection* indeed positively related to high-HDI TEA, but negatively related to low-HDI TEA?

However, this is not to say that this research has not produced any meaningful results. Among the deemed most important variables are several cultural variables linked to individuality and female autonomy, which is in line with extant literature. Differences in *Willingness to delegate authority* might not only be the result of cultural differences though, as it is also linked with lower perceived labor quality by managers (Lieberman & Boehe (2011)).

Further specified research into these variables seems valuable. Furthermore, there does indeed seem to be a clear and significant difference between determinants for TEA and EEA, on both levels of country development. This calls for more research into institutional determinants for intrapreneurship and the way they differ from determinants for entrepreneurship. Furthermore, this paper might help to bind the fragmented research field as described by Terjesen et al., (2016). Applying its approach in a slightly more limited setting, but a broader setting than in the more conventional way, might lead to fresh insights, and might pave the way for a new approach, as a replacement or at least complement of traditional research methods.

Considering the explorative nature of the research and the findings, no concrete policy implications arise from this research. The fact that there are significant differences between TEA and EEA determinants however implies that stimulating EEA might require different policies. Further research is needed to determine the nature and scope of these differences.

5.2 Limitations and Improvements

Although the aim of this research has been to purposely include many independent, correlation and possible non-relevant variables, the procedure might prove clearer results is only aggregate scores of the nine NES components and 12 GCI pillars were included, rather than the individual variables. This is partly due to the method's 'grouping effect'; when several highly correlated variables have a similar impact on the model's performance, only one or a few will be selected. The selected variable could therefore as well be a proxy for another conducive variable, while not necessarily be conducive itself.

Besides looking at aggregates of the variables, it might also be worthwhile to narrow the scope of the dependent and/or independent variables. For example, finding predictors for *innovative* entrepreneurial activity would possibly lead to less 'wild' results. Unfortunately, although the GEM does collect data on innovative TEA, it does not collect data on innovative EEA (at least not until 2015). The same might apply to a more limited scope of possible determinants; assessing e.g. only informal / cultural variables, or only policy variables might provide more stable and interpretable results. Possibly more than anything else, including more instances would benefit the models the most. Especially after subsetting on HDI and dividing the data in training and testing sets, some subsets became rather small, perhaps too small. Unfortunately, adding more instances was not possible given the constraints regarding data availability.

A single training set and a single test have been used. Applying bootstrapping techniques to use different samples of training and testing sets might strengthen the findings. Whether this would have a significant impact on the interpretability of the results is hard to say, but bootstrapping might indeed be worthwhile.

The lack of interpretability is not to say that the chosen methodology has proven to be not useful. By addressing the drawbacks of the procedure, and with more limited expectations, the procedure can certainly add value. A model that would provide results that show a little bit more in certain direction(s), might still contain individual predictors that are hard to interpret, but might be a starting point for further research in a more specified direction, possibly with more traditional research methods.

Machine learning algorithms might have their limits, but so do traditional research methods. Conducting an analysis on such a comprehensive dataset as the one used in this research would not be feasible at all. Given, and respecting, their constraints, methods such as the Elastic Net procedure applied in this research should therefore be seen as a potentially very powerful addition to traditional research methods.

6. CONCLUSION

Reflecting on the defined research aims and research questions, it must be concluded that the main research questions can only be partially answered. Although the desired dataset could be constructed, and the Elastic Net procedure has been executed as planned, meaningful interpretation of most variables was not considered to be appropriate and responsible. As such, no valid conclusions can, and should, be made based on individual determinants. This is not to say that these determinants are not significant. Based on a combination with other indications that these variables might be conducive, more detailed research into them could be undertaken.

However, besides the hard to interpret variables, a few clearer conclusions can be also be made. Firstly, the identified determinants for TEA and EEA, on both levels of development, are so different that it seems appropriate to conclude that the conducive determinants for both do seem to differ significantly. Secondly, cultural factors related to individualism seem to be play a significant role, which is in line with extant literature. A similar, but more targeted research might prove to be beneficial.

Furthermore, I recommend further applications of machine learning methods and use of large-scale datasets in similar research, as such methods can definitely have added value, especially as a complement to more traditional methods. This does not only apply to machine learning methods in general, but also to the model and data used in this research; several adaptations, e.g. in the form of an expanded dataset or a more limited set of independent variables might indeed produce more interpretable and verifiable results, and make it useful for further research and policy makers alike.

7. REFERENCES

- Alfons, Andreas & Croux, Christophe & Gelper, Sarah. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*. 7. 10.1214/12-AOAS575.
- Antoncic, Bostjan & Hisrich, Robert. (2004). Corporate entrepreneurship contingencies and organizational wealth creation. *Journal of Management Development*. 23. 518-550. 10.1108/02621710410541114.
- Autio, Erkkö and Pathak, Saurav and Wennberg, Karl, Consequences of Cultural Practices for Entrepreneurial Behaviors (February 6, 2013). *Journal of International Business Studies*, Forthcoming. <http://dx.doi.org/10.2139/ssrn.2229397>
- Baughn, C. C., & Neupert, K. E. 2003. Culture and national conditions facilitating entrepreneurial start-ups. *Journal of International Entrepreneurship*, 1: 313-330.
- Baumol, W. J. 1996. Entrepreneurship: Productive, unproductive and destructive. *Journal of Business Venturing*, 11: 3-20. (Reprinted from *Journal of Political Economy*, 1990)
- Bieto, E. (2008) Aproximación al Corporate Entrepreneurship en España. Tipologías y relación con los resultados empresariales, Tesis Doctoral, ESADE – Escuela Superior de Administración y Dirección de Empresas, Universitat Ramon Llull, Barcelona
- Bjørnskov, C., & Foss, N. J. (2016). Institutions, Entrepreneurship, and Economic Growth: What Do We Know and What Do We Still Need to Know? *Academy of Management Perspectives*, 30(3), 292-315. <https://doi.org/10.5465/amp.2015.0135>
- Bjuggren, Carl Magnus & Johansson, Dan & Stenkula, Mikael. (2010). Using Self-Employment as Proxy for Entrepreneurship: Some Empirical Caveats. The Ratio Institute, Ratio Working Papers. 10.2139/ssrn.1678516.
- Bosma, Niels & Stam, Erik & Wennekers, Sander. (2011). Intrapreneurship versus Independent Entrepreneurship: A cross-national analysis of individual entrepreneurial behavior.
- Bowen, H., & De Clercq, D. (2008). Institutional Context and the Allocation of Entrepreneurial Effort. *Journal of International Business Studies*, 39(4), 747-768. Retrieved February 18, 2020, from www.jstor.org/stable/25483297
- Carree, M. A., & Thurik, A. R. 2008. The lag structure of the impact of business ownership on economic performance in OECD countries. *Small Business Economics*, 30: 101-110.

- Crnogaj, K. & Hojnik, B. (2016). Institutional Determinants and Entrepreneurial Action. *Management: Journal of Contemporary Management Issues*. 21. 131-150.
- Dervis, Kemal & Klugman, Jeni. (2011). Measuring human progress: The contribution of the Human Development Index and related indices. *Revue d'économie politique*. 121. 73-92. 10.3917/redp.211.0073
- Fontana, R., Nuvolari, A., Shimizu, H. et al. Schumpeterian Patterns of Innovation and the Sources of Breakthrough Inventions: Evidence from a Data-set of R&D awards. *J Evol Econ* 22, 785-810 (2012). <https://doi.org/10.1007/s00191-012-0287-z>
- Grilo, I., & Irigoyen, J. M. 2006. Entrepreneurship in the EU: To wish and not to be. *Small Business Economics*, 26: 305-318.
- Hofstede, G. 2001. *Culture's consequences, comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
- Ireland, R. D., Tihanyi, L., & Webb, J. W. 2008. A tale of two politico-economic systems: Implications for entrepreneurship in central and eastern Europe. *Entrepreneurship Theory & Practice*, 32: 107-130.
- Justo, R., De Castro, J.O. and Maydeu-Olivares, A. (2008) 'Indicators of entrepreneurship activity: some methodological contributions', *Int. J. Entrepreneurship and Small Business*, Vol. 6, No. 4, pp.604-621.
- Kuratko, Donald F. and Montagno, Ray V. and Hornsby, Jeffrey S, *Developing an Intrapreneurial Assessment Instrument for an Effective Corporate Entrepreneurial Environment* (1990). *Strategic Management Journal*, Vol. 11, p. 49-58 1990.
- Kurnaz, F. Sevinç & Ortner, Irene & Filzmoser, Peter. (2017). Robust and sparse estimation methods for high dimensional linear and logistic regression. *Chemometrics and Intelligent Laboratory Systems*. 10.1016/j.chemolab.2017.11.017.
- Lassmann, A., Busch, C. Revisiting Native and Immigrant Entrepreneurial Activity. *Small Bus Econ* 45, 841-873 (2015). <https://doi.org/10.1007/s11187-015-9665-x>
- Lee, S.-H., Peng, M. W., & Barney, J. B. 2007. Bankruptcy law and entrepreneurship: A real options perspective. *Academy of Management Review*, 32: 257-272.
- Lieberman, Leonardo & Boehe, Dirk. (2011). Worldwide willingness to delegate and country labor quality. *The International Journal of Human Resource Management*. 22. 3477-3495. 10.1080/09585192.2011.610943.
- Lockhart, Richard & Taylor, Jonathan & Tibshirani, Ryan & Tibshirani, Robert. (2013). A significance test for the LASSO. *The Annals of Statistics*. 42. 10.1214/13-AOS1175.
- Marchiori, Danilo & Madeira, Maria & Dinis, Anabela. (2018). Antecedents of Entrepreneurial and Intrapreneurial Intentions: Evidence from Brazil. *RACE - Revista de Administração, Contabilidade e Economia*. 17. 913. 10.18593/race.v17i3.18841.

- McMullen, Jeffery & Bagby, D. & Palich, Leslie. (2008). Economic Freedom and the Motivation to Engage in Entrepreneurial Action. *Entrepreneurship: Theory and Practice*. 32. 10.1111/j.1540-6520.2008.00260.x.
- Mersha, Tigneh & Sriram, Ven & Hailu, Melkrist. (2010). Nurturing opportunity entrepreneurs in Africa: Some lessons from Ethiopia. *J. for Global Business Advancement*. 3. 10.1504/JGBA.2010.033200.
- Mitchell, R. K., Smith, J. B., Morse, E. A., Seawright, K. W., Peredo, A. M., & McKenzie, B. 2002. Are entrepreneurial cognitions universal? Assessing entrepreneurial cognitions across cultures. *Entrepreneurship Theory & Practice*, 26: 9-32.
- Morris, M., Davis, D. & Allen, J. Fostering Corporate Entrepreneurship: Cross-Cultural Comparisons of the Importance of Individualism Versus Collectivism. *J Int Bus Stud* 25, 65-89 (1994).
<https://doi.org/10.1057/palgrave.jibs.8490849>
- North, D. 1994. Economic performance through time. *American Economic Review*, 84: 359-368.
- Pinchot, G., (1987) Innovation Through Intrapreneuring, *Research Management*, 30:2, 14-19, DOI: 10.1080/00345334.1987.11757021
- Quesada, Onaindia, Ochoa Laburu. 2011. Determinant factors for intrapreneurship: a case study based analysis. *XV Congreso de Ingeniería de Organización*, pag. 56-65. Cartagena.
- Schumpeter, Joseph A., *Capitalism, Socialism, and Democracy* (1942). University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.
- Sharma, P., & Chrisman, J. J. (1999). Toward a reconciliation of the definitional issues in the field of corporate entrepreneurship. *Entrepreneurship Theory and Practice*, 23(3), 11-28.
- Shane, Scott & Venkataraman, Sankaran. (2013). The Promise of Entrepreneurship As A Field of Study. *The Academy of Management Review*. 25. 10.2307/259271.
- Stefanovici, S. (2012, September 7). Intrapreneurship – a review of recent literature. *Business Economics*. Retrieved from <http://hdl.handle.net/2105/12000>
- Terjesen, S., Hessels, J., & Li, D. (2013). Comparative International Entrepreneurship. *Journal of Management*, 42(1), 299-344.
<https://doi.org/10.1177/0149206313486259>
- Thurik, Roy & Hessels, Jolanda & Gelderen, Marco. (2008). Entrepreneurial aspirations, motivations, and their drivers. *Small Business Economics*. 31. 323-339.
- Valliere, D., & Peterson, R. 2009. Entrepreneurship and economic growth: Evidence from emerging and developed countries. *Entrepreneurship & Regional Development*, 21: 459-480.
- van Stel, A., Storey, D. J., & Thurik, A. R. 2007. The effect of business regulations on nascent and young business. *Small Business Economics*, 28: 171-186.

- Verheul, I., van Stel, A., & Thurik, R. 2006. Explaining female and male entrepreneurship at the country level. *Entrepreneurship & Regional Development*, 18: 151-183.
- Wennekers, S. & Thurik, R. Linking Entrepreneurship and Economic Growth. *Small Business Economics* 13, 27-56 (1999). <https://doi.org/10.1023/A:1008063200484>
- Woo, Hyung. (2018). Personality Traits and Intrapreneurship: The Mediating Effect of Career Adaptability. *Career Development International*. 23. 10.1108/CDI-02-2017-0046.
- Zahra, Shaker A., Predictors and Financial Outcomes of Corporate Entrepreneurship: An Exploratory Study (1991). *Journal of Business Venturing*, Vol. 6, Issue 4, p. 259-286 1991.
- Zou, Hui & Hastie, Trevor. (2005). Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B*. 2005;67(2):301-20. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 67. 301 - 320. 10.1111/j.1467-9868.2005.00503.x

8. APPENDICES

8.1 Appendix A: Overview of Independent Variables

GEM National Entrepreneurship Survey variables

1. ENTREPRENEURIAL FINANCE

- NES_A01** There is sufficient equity funding available for new and growing firms
- NES_A02** There is sufficient debt funding available for new and growing firms
- NES_A03** There are sufficient government subsidies available for new and growing firms
- NES_A04** There is sufficient funding available from informal investors who are private individuals (other than founders) for new and growing firms
- NES_A05** There is sufficient funding available from professional Business Angels for new and growing firms
- NES_A06** There is sufficient funding available from venture capitalists for new and growing firms

2. GOVERNMENT POLICY

- NES_B01** Government policies (e.g. public procurement) consistently favour new firms
- NES_B02** The support for new and growing firms is a high priority for policy at the national government level
- NES_B03** The support for new and growing firms is a high priority for policy at the local government level
- NES_B04** New firms can get most of the required permits and licenses in about a week
- NES_B05** The amount of taxes is NOT a burden for new and growing firms
- NES_B06** Taxes and other government regulations are applied to new and growing firms in a predictable and consistent way
- NES_B07** Coping with government bureaucracy, regulations, and licensing requirements it is not unduly difficult for new and growing firms

3. GOVERNMENT ENTREPRENEURSHIP PROGRAMS

- NES_C01** A wide range of government assistance for new and growing firms can be obtained through contact with a single agency
- NES_C02** Science parks and business incubators provide effective support for new and growing firms
- NES_C03** There are an adequate number of government programs for new and growing businesses
- NES_C04** The people working for government agencies are competent and effective in supporting new and growing firms
- NES_C05** Almost anyone who needs help from a government program for a new or growing business can find what they need
- NES_C06** Government programs aimed at supporting new and growing firms are effective

4. ENTREPRENEURSHIP EDUCATION

- NES_D01** Teaching in primary and secondary education encourages creativity, self-sufficiency, and personal initiative
- NES_D02** Teaching in primary and secondary education provides adequate instruction in market economic principles
- NES_D03** Teaching in primary and secondary education provides adequate attention to entrepreneurship and new firm creation
- NES_D04** Colleges and universities provide good and adequate preparation for starting up and growing new firms
- NES_D05** The level of business and management education provide good and adequate preparation for starting up and growing new firms
- NES_D06** The vocational, professional, and continuing education systems provide good and adequate preparation for starting up and growing new firms

5. R&D TRANSFER

- NES_E01** New technology, science, and other knowledge are efficiently transferred from universities and public research centers to new and growing firms
- NES_E02** New and growing firms have just as much access to new research and technology as large, established firms
- NES_E03** New and growing firms can afford the latest technology
- NES_E04** There are adequate government subsidies for new and growing firms to acquire new technology
- NES_E05** The science and technology base efficiently supports the creation of world-class new technology-based ventures in at least one area
- NES_E06** There is good support available for engineers and scientists to have their ideas commercialized through new and growing firms

6. COMMERCIAL AND LEGAL INFRASTRUCTURE

- NES_F01** There are enough subcontractors, suppliers, and consultants to support new and growing firms
- NES_F02** New and growing firms can afford the cost of using subcontractors, suppliers, and consultants
- NES_F03** It is easy for new and growing firms to get good subcontractors, suppliers, and consultants
- NES_F04** It is easy for new and growing firms to get good, professional legal and accounting services
- NES_F05** It is easy for new and growing firms to get good banking services (checking accounts, foreign exchange transactions, letters of credit, and the like)

7. ENTRY REGULATION

- NES_G01** The markets for consumer goods and services change dramatically from year to year
- NES_G02** The markets for business-to-business goods and services change dramatically from year to year
- NES_G03** New and growing firms can easily enter new markets
- NES_G04** The new and growing firms can afford the cost of market entry
- NES_G05** New and growing firms can enter markets without being unfairly blocked by established firms
- NES_G06** The anti-trust legislation is effective and well enforced

8. PHYSICAL INFRASTRUCTURE

- NES_H01** The physical infrastructure (roads, utilities, communications, waste disposal) provides good support for new and growing firms
- NES_H02** It is not too expensive for a new or growing firm to get good access to communications (phone, Internet, etc)
- NES_H03** A new or growing firm can get good access to communications (telephone, internet, etc) in about a week
- NES_H04** New and growing firms can afford the cost of basic utilities (gas, water, electricity, sewer)
- NES_H05** New or growing firms can get good access to utilities (gas, water, electricity, sewer) in about a month

9. CULTURAL AND SOCIAL NORMS

- NES_J01** The national culture is highly supportive of individual success achieved through own personal efforts
- NES_J02** The national culture emphasizes self-sufficiency, autonomy, and personal initiative
- NES_J03** The national culture encourages entrepreneurial risk-taking
- NES_J04** The national culture encourages creativity and innovativeness
- NES_J05** The national culture emphasizes the responsibility that the individual (rather than the collective) has in managing his or her own life

WEF Global Competitiveness Index variables

BASIC REQUIREMENTS
EFFICIENCY ENHANCERS
INNOVATION AND SOPHISTICATION FACTORS

1. INSTITUTIONS

- GC_1.01** Property rights
- GC_1.02** Intellectual property protection
- GC_1.03** Diversion of public funds
- GC_1.04** Public trust in politicians
- GC_1.05** Irregular payments and bribes
- GC_1.06** Judicial independence
- GC_1.07** Favouritism in decisions of government officials

6. GOODS MARKET EFFICIENCY

- GC_6.01** Intensity of local competition
- GC_6.02** Extent of market dominance
- GC_6.03** Effectiveness of anti-monopoly policy
- GC_6.05** Total tax rate
- GC_6.06** Number of procedures required to start a business
- GC_6.07** Time required to start a business
- GC_6.08** Agricultural policy costs
- GC_6.09** Prevalence of trade barriers
- GC_6.10** Trade tariffs
- GC_6.11** Prevalence of foreign ownership
- GC_6.12** Business impact of rules on FDI
- GC_6.13** Burden of customs procedures
- GC_6.14** Imports as a percentage of GDP

GO_1.08	Wastefulness of government spending
GO_1.09	Burden of government regulation
GO_1.10	Efficiency of legal framework in settling disputes
GO_1.11	Efficiency of legal framework in challenging regulations
GO_1.12	Transparency of government policymaking
GO_1.13	Business costs of terrorism
GO_1.14	Business costs of crime and violence
GO_1.15	Organized crime
GO_1.16	Reliability of police services
GO_1.17	Ethical behaviour of firms
GO_1.18	Strength of auditing and reporting standards
GO_1.19	Efficacy of corporate boards
GO_1.20	Protection of minority shareholders' interests
GO_1.21	Strength of investor protection

2. INFRASTRUCTURE

GO_2.01	Quality of overall infrastructure
GO_2.02	Quality of roads
GO_2.03	Quality of railroad infrastructure
GO_2.04	Quality of port infrastructure
GO_2.05	Quality of air transport infrastructure
GO_2.06	Available airline seat kilometres
GO_2.07	Quality of electricity supply
GO_2.08	Mobile telephone subscriptions
GO_2.09	Fixed telephone lines

3. MACROECONOMIC ENVIRONMENT

GO_3.01	Government budget balance
GO_3.02	Gross national savings
GO_3.03	Inflation
GO_3.04	Government debt

4. HEALTH AND PRIMARY EDUCATION

GO_4.03	Business impact of tuberculosis
GO_4.04	Tuberculosis incidence
GO_4.05	Business impact of HIV/AIDS
GO_4.06	HIV prevalence
GO_4.07	Infant mortality
GO_4.08	Life expectancy
GO_4.09	Quality of primary education

5. HIGHER EDUCATION AND TRAINING

GO_5.01	Secondary education enrollment rate
GO_5.02	Tertiary education enrollment rate
GO_5.03	Quality of the educational system
GO_5.04	Quality of math and science education
GO_5.05	Quality of management school
GO_5.06	Internet access in schools
GO_5.07	Local availability of specialized research and training services
GO_5.08	Extent of staff training

GO_6.15	Degree of customer orientation
GO_6.16	Buyer sophistication

7. LABOR MARKET EFFICIENCY

GO_7.01	Cooperation in labor-employer relations
GO_7.02	Flexibility of wage determination
GO_7.03	Hiring and firing practices
GO_7.04	Redundancy costs
GO_7.06	Pay and productivity
GO_7.07	Reliance on professional management
GO_7.10	Female participation in labor force

8. FINANCIAL MARKET DEVELOPMENT

GO_8.03	Financing through local equity market
GO_8.04	Ease of access to loans
GO_8.05	Venture capital availability
GO_8.06	Soundness of banks
GO_8.07	Regulation of securities exchanges
GO_8.08	Legal rights index

9. TECHNOLOGICAL READINESS

GO_9.01	Availability of latest technologies
GO_9.02	Firm-level technology absorption
GO_9.03	FDI and technology transfer
GO_9.04	Internet users
GO_9.05	Broadband Internet subscriptions
GO_9.06	Internet bandwidth

10. MARKET SIZE

GO_10.01	Domestic market size index
GO_10.02	Foreign market size index
GO_10.03	GDP (PPP\$ billions)
GO_10.04	Exports as a percentage of GDP

11. BUSINESS SOPHISTICATION

GO_11.01	Local supplier quantity
GO_11.02	Local supplier quality
GO_11.03	State of cluster development
GO_11.04	Nature of competitive advantage
GO_11.05	Value chain breadth
GO_11.06	Control of international distribution
GO_11.07	Production process sophistication
GO_11.08	Extent of marketing
GO_11.09	Willingness to delegate authority

12. R&D INNOVATION

GO_12.01	Capacity for innovation
GO_12.02	Quality of scientific research institutions
GO_12.03	Company spending on R&D
GO_12.04	University-industry collaboration in R&D
GO_12.05	Government procurement of advanced technology products
GO_12.06	Availability of scientists and engineers

8.2 Appendix B: Correlation Matrix of Independent Variables

Correlation Matrix of Independent Variables



8.3 Appendix C: Composition of Dataset

Composition of Dataset

	Country	Year	TEA	EEA	HI		Country	Year	TEA	EEA	HI		Country	Year	TEA	EEA	HI
N AMERICA	Canada	2014	13.03	4.75	0.914	EUROPE	Austria	2014	8.71	5.62	0.904	EUROPE (CONT)	Netherlands	2011	8.20	7.83	0.922
		2015	14.72	7.06	0.917		Belgium	2012	5.19	7.25	0.905			2014	9.45	6.99	0.925
	United States	2014	13.81	6.46	0.915		2014	5.40	5.41	0.910	2015			7.20	6.27	0.927	
		2015	11.88	6.96	0.917		2015	6.24	6.12	0.913	Norway		2014	5.65	7.90	0.945	
LATIN AMERICA & CARIBBEAN	Argentina	2011	20.77	3.20	0.823		Bulgaria	2015	3.45	0.35		0.807	2015	5.65	9.89	0.948	
		2014	14.40	2.42	0.825		Croatia	2012	8.27	4.43	0.820	Poland	2011	9.03	2.82	0.840	
	Brazil	2011	14.88	1.01	0.730		Denmark	2012	5.36	12.61	0.924		2013	9.27	4.27	0.851	
		2015	20.98	0.98	0.755			2014	5.47	11.43	0.928		2014	9.21	3.36	0.853	
	Chile	2011	23.68	3.45	0.812		Estonia	2012	14.25	7.29	0.859	2015	9.20	4.04	0.858		
		2012	22.58	4.79	0.818			2014	9.43	3.58	0.865	Portugal	2011	7.53	4.02	0.827	
		2013	24.33	3.57	0.830			2015	13.13	6.29	0.871		2014	9.97	3.23	0.840	
		2014	26.82	5.13	0.834		Finland	2013	5.28	5.73	0.916		2015	9.48	3.95	0.843	
		2015	25.92	5.22	0.839			2014	5.63	4.45	0.918	Romania	2012	9.21	3.49	0.796	
	Colombia	2011	21.43	1.66	0.735			2015	6.58	5.83	0.919		2013	10.12	5.05	0.800	
		2014	18.54	3.69	0.750	France	2011	5.73	4.74	0.876	2014		11.34	4.94	0.803		
		2015	22.67	2.31	0.753		2014	5.34	3.80	0.887	2015	10.82	4.60	0.806			
	Costa Rica	2012	15.04	1.91	0.774	Germany	2014	5.27	4.44	0.930	Russia	2011	4.56	0.60	0.789		
		2014	11.32	0.53	0.785		2015	4.69	4.51	0.933		2014	4.68	0.51	0.807		
	Ecuador	2012	26.61	0.89	0.740	Greece	2011	7.95	1.61	0.853	Slovakia	2012	10.21	4.59	0.841		
		2013	35.97	1.87	0.751		2012	6.50	1.63	0.855		2013	9.52	4.35	0.844		
	Jamaica	2011	13.70	0.19	0.723		2014	7.85	0.76	0.866	Slovenia	2012	5.42	5.87	0.876		
		2014	18.99	0.71	0.754	2015	6.74	1.03	0.868	2013		6.45	6.09	0.884			
	Mexico	2011	9.62	0.93	0.746	Hungary	2012	9.21	2.13	0.826		2014	6.32	4.73	0.886		
		2012	12.10	1.23	0.752		2013	9.67	3.61	0.835	2015	5.91	5.59	0.886			
		2014	18.99	0.71	0.754		2014	9.33	1.80	0.833	Spain	2011	5.81	2.71	0.870		
		2015	21.00	1.16	0.759	2015	7.92	2.12	0.835	2014		5.47	1.82	0.880			
	Panama	2011	20.78	0.15	0.764	Ireland	2011	7.24	5.91	0.894		2015	5.69	1.09	0.885		
		2014	17.05	0.19	0.779		2014	6.53	6.66	0.920	Sweden	2011	5.79	16.18	0.906		
		2015	12.80	0.50	0.782		2015	9.32	6.60	0.926		2014	6.70	5.77	0.929		
	Peru	2011	22.88	1.35	0.734	Italy	2014	4.41	0.76	0.874		Switzerland	2015	7.16	6.36	0.932	
		2012	20.20	1.71	0.734		2015	4.86	1.37	0.875	2011		6.58	4.62	0.932		
		2013	23.38	0.85	0.742	Latvia	2011	11.85	3.02	0.822	2014	7.11	6.06	0.939			
		2014	28.81	1.67	0.752		2015	14.10	3.33	0.842	2015	7.30	6.49	0.943			
		2015	22.21	0.72	0.750	Lithuania	2011	11.26	4.86	0.831	United Kingdom	2011	7.28	5.24	0.899		
Uruguay	2011	16.72	5.20	0.783	2012		6.68	5.79	0.835	2012		8.97	10.07	0.897			
	2012	14.63	5.01	0.788	2013		12.42	6.79	0.840	2013		7.13	8.51	0.914			
	2013	14.07	6.64	0.797	2014		11.32	5.08	0.852	2014		10.65	6.99	0.918			
	2014	16.07	3.81	0.800	Luxembourg		2014	7.13	7.26	0.895		2015	6.93	4.07	0.916		
	2015	14.28	4.19	0.802			2015	10.18	6.43	0.899							

ASIA	China	2012	12.83	0.58	0.719
		2013	14.01	0.60	0.727
	Georgia	2014	7.21	0.36	0.764
	India	2014	6.59	0.09	0.618
		2015	10.82	0.30	0.627
	Indonesia	2014	14.19	0.47	0.691
		2015	17.67	0.20	0.696
	Iran	2012	10.79	0.73	0.782
		2013	12.32	2.49	0.785
		2014	16.01	1.39	0.788
		2015	12.93	0.95	0.789
	Israel	2012	6.52	4.23	0.893
		2015	11.82	6.54	0.901
	Kazakhstan	2014	13.72	1.04	0.798
		2015	10.99	0.94	0.806
	Malaysia	2011	4.92	0.40	0.779
		2012	6.98	0.89	0.782
		2013	6.59	0.51	0.787
		2014	5.91	0.00	0.792
		2015	2.92	0.27	0.797
	Pakistan	2011	9.07	0.19	0.528
		2012	11.57	0.16	0.533
	Philippines	2013	18.51	0.39	0.692
		2014	18.37	0.63	0.697
		2015	17.15	2.34	0.702
	Singapore	2011	6.60	3.30	0.914
	South Korea	2012	6.63	2.06	0.890
		2013	6.85	3.97	0.893
	Thailand	2011	19.51	1.37	0.729
		2012	18.94	1.10	0.733
		2013	17.66	1.66	0.731
		2014	23.29	2.04	0.739
		2015	13.74	0.72	0.746
	Turkey	2015	12.21	0.96	0.765
	Vietnam	2013	15.35	2.70	0.673
		2014	15.30	0.30	0.675
		2015	13.65	0.55	0.680
AFRICA	Algeria	2011	9.26	0.82	0.738
		2012	8.75	0.57	0.737
		2013	4.88	0.64	0.746
	Botswana	2012	27.65	2.06	0.687
		2013	20.85	0.86	0.699
		2014	32.79	2.69	0.709
		2015	33.23	1.56	0.714
	Cameroon	2014	37.37	1.91	0.540
		2015	25.36	0.66	0.548
	Egypt	2012	7.82	1.03	0.676
		2015	7.39	1.33	0.690
	Ethiopia	2012	14.72	0.57	0.429
	Morocco	2015	4.44	0.37	0.660
	Namibia	2012	18.15	2.05	0.612
		2013	33.33	1.58	0.622
	Nigeria	2011	34.98	3.16	0.494
		2012	35.04	0.99	0.502
		2013	39.86	0.46	0.520
	Senegal	2015	38.55	2.27	0.504
	South Africa	2012	7.31	0.40	0.673
		2013	10.58	0.77	0.683
		2014	6.97	0.25	0.691
		2015	9.18	0.30	0.699
	Tunisia	2015	10.12	1.86	0.731
	Uganda	2014	35.52	2.32	0.509
OCEANIA	Australia	2011	10.50	6.15	0.928
		2014	13.14	8.37	0.929
		2015	12.78	8.51	0.933

8.4 Appendix D: Conducive Variables per Model

Conducive Variables: High HDI - TEA

Code	Variable Description	Coefficients
INTERCEPT		6.578053
NES_B05	The amount of taxes is NOT a burden for new and growing firms	- 0.4908398789
NES_D05	The level of business and management education provide good and adequate preparation for starting up and growing new firms	- 0.2228536408
NES_I03	The national culture encourages entrepreneurial risk-taking	0.6209032959
NES_I05	The national culture emphasizes the responsibility that the individual (rather than the collective) has in managing his/her own life	0.2264686318
GCI_1.13	Business costs of terrorism	0.0083357738
GCI_1.15	Organized crime	0.1292432948
GCI_1.21	Strength of investor protection	- 0.3437431756
GCI_2.05	Quality of air transport infrastructure	- 0.4655406759
GCI_3.02	Gross national savings	- 0.0131659617
GCI_4.03	Business impact of tuberculosis	0.0223444060
GCI_5.01	Secondary education enrollment rate	- 0.0052848694
GCI_5.06	Internet access in schools	0.0511977501
GCI_6.10	Trade tariffs	- 0.3181896493
GCI_7.02	Flexibility of wage determination	0.2197546626
GCI_8.04	Ease of access to loans	- 0.0836912468
GCI_9.03	FDI and technology transfer	0.8791185479
GCI_9.06	Internet bandwidth	0.0002325376
GCI_10.03	GDP (PPP\$ billions)	- 0.0001935117
N (train)		66
RMSE (train)		6.029457
N (test)		29
RMSE (test)		6.971976

Conducive Variables: High HDI - EEA

Code	Variable Description	Coefficients
INTERCEPT		- 1.750194
NES_C06	Government programs aimed at supporting new and growing firms are effective	0.032100216
NES_I05	The national culture emphasizes the responsibility that the individual (rather than the collective) has in managing his/her own life	0.105673523
GCI_2.05	Quality of air transport infrastructure	- 0.099337612
GCI_2.08	Mobile telephone subscriptions	- 0.002738772
GCI_3.01	Government budget balance	- 0.085352943
GCI_3.04	Government debt	- 0.011240684
GCI_5.06	Internet access in schools	0.114413640
GCI_6.07	Time required to start a business	- 0.025153584
GCI_6.08	Agricultural policy costs	0.322325235
GCI_6.09	Prevalence of trade barriers	- 0.033494882

GCI_7.03	Hiring and firing practices	- 0.001725863
GCI_8.08	Legal rights index	0.023303923
GCI_9.03	FDI and technology transfer	0.218281439
GCI_11.08	Extent of marketing	0.027800309
GCI_11.09	Willingness to delegate authority	0.846635329
GCI_12.01	Capacity for innovation	0.255388870
GCI_12.02	Quality of scientific research institutions	- 0.071947378
N (train)		66
RMSE (train)		1.720098
N (test)		29
RMSE (test)		1.54061

Conducive Variables: Low HDI - TEA

Code	Variable Description	Coefficients
INTERCEPT		- 1.900887
NES_C04	The people working for government agencies are competent and effective in supporting new and growing firms	0.1760214963
NES_D04	Colleges and universities provide good and adequate preparation for starting up and growing new firms	0.3279372837
NES_D05	The level of business and management education provide good and adequate preparation for starting up and growing new firms	0.1409262606
NES_I05	The national culture emphasizes the responsibility that the individual (rather than the collective) has in managing his/her own life	0.0093036624
GCI_1.06	Judicial independence	- 0.0837026621
GCI_1.11	Efficiency of legal framework in challenging regulations	- 0.3171789968
GCI_1.21	Strength of investor protection	0.2161870427
GCI_2.03	Quality of railroad infrastructure	- 0.7307355453
GCI_2.05	Quality of air transport infrastructure	- 0.0096620077
GCI_2.07	Quality of electricity supply	- 0.6795736017
GCI_2.08	Mobile telephone subscriptions	- 0.0032073222
GCI_3.01	Government budget balance	0.2470724560
GCI_4.03	Business impact of tuberculosis	- 0.0008455736
GCI_4.07	Infant mortality	- 0.0084079633
GCI_5.01	Secondary education enrollment rate	- 0.0091658913
GCI_5.07	Local availability of specialized research and training services	- 0.1328961009
GCI_5.08	Extent of staff training	- 0.0005627943
GCI_6.01	Intensity of local competition	1.7916711822
GCI_6.02	Extent of market dominance	0.1243480969
GCI_6.07	Time required to start a business	0.0434072927
GCI_6.12	Business impact of rules on FDI	0.1368237854
GCI_6.13	Burden of customs procedures	0.4969995935
GCI_6.15	Degree of customer orientation	0.8904379021
GCI_7.01	Cooperation in labor-employer relations	1.1784710324
GCI_7.02	Flexibility of wage determination	0.3179739311
GCI_7.04	Redundancy costs	0.0045919709

GCI_7.10	Female participation in labor force	0.4290972606
GCI_8.06	Soundness of banks	0.6259741803
GCI_8.08	Legal rights index	0.0745139496
GCI_9.06	Internet bandwidth	- 0.0062910886
GCI_10.02	Foreign market size index	- 0.1889320903
GCI_11.04	Nature of competitive advantage	- 0.0570336597
GCI_12.02	Quality of scientific research institutions	- 0.3564012364
GCI_12.06	Availability of scientists and engineers	- 1.8328420997
N (train)		59
RMSE (train)		8.410804
N (test)		26
RMSE (test)		2.24617

Conductive Variables: Low HDI - EEA

Code	Variable Description	Coefficients
INTERCEPT		2.444289
NES_E04	There are adequate government subsidies for new and growing firms to acquire new technology	- 0.0311333682
NES_E05	The science and technology base efficiently supports the creation of world-class new technology-based ventures in at least one area	- 0.0112153577
NES_E06	There is good support available for engineers and scientists to have their ideas commercialized through new and growing firms	- 0.0123922452
GCI_2.08	Mobile telephone subscriptions	- 0.0006568882
GCI_6.02	Extent of market dominance	- 0.1309822159
GCI_7.02	Flexibility of wage determination	0.0128614034
GCI_7.10	Female participation in labor force	0.2805800153
GCI_8.05	Venture capital availability	- 0.0599826620
GCI_8.07	Regulation of securities exchanges	- 0.0006489753
GCI_9.06	Internet bandwidth	- 0.0004120142
GCI_10.01	Domestic market size index	- 0.0146945369
GCI_10.02	Foreign market size index	- 0.0982966754
GCI_12.01	Capacity for innovation	- 0.0874052501
N (train)		59
RMSE (train)		1.357226
N (test)		26
RMSE (test)		2.095679