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

Bachelor Thesis

With the supervision of dr. W. Liebregts

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 aims to identify institutional determinants of both intrapreneurial (GEM EEA) and entrepreneurial activity (GEM TEA) and assess whether these differ, and whether they differ from determinants found in previous, more limited, research. 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, and with determinants found in previous research. Interpretation of the exact results has to be done with care, given the, considering the model, still limited dataset and capriciousness of the elastic net procedure, despite the added robustness. However, 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.

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Research Rationale and Scientific and Societal Relevance

Ever since Joseph Schumpeter developed his first, Mark I, theory on entrepreneurship 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 and Foss, 2016)). While there has been an significant amount of research on individual factors that characterize entrepreneurs, 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 and Foss, 2016). There has been research into institutional determinants of entrepreneurial activity (e.g. Bowen & de Clercq (2008), Crnogaj and Hojnic (2016), Urbano, Aparicio and Audretsch (2018)), 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 Schumpeter 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 entreneurship 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 and Dinis (2018), but research into determinant of intrapreneurial activity is still almost succinct. Bosma, Stam and 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.

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 variables 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 have to 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 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

A number of 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

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 and 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 (2008); Lassmann & Busch (2015)). The disadvantage of such stringent definitions is that they seldom cover the entire range of entrepreneurship.

However, while undoubtedly a relevant discussion, a problem arises when it comes to applying research based on such a variety of broad definitions; the gathering of valid data and the comparison, and reproducibility, of research results. The Global Entrepreneurship Monitor aims to bridge the gap between a broad definition of entrepreneurship and feasibly concerning applicability in (quantitative) research . The GEM approach will be discussed in more depth later, but its approach has become popular and has resulted in many studies using the GEM definition and data. A positive effect of this broad application of similar data, based on a single definition, is that the studies are comparable, as similar variables are used, in particular the TEA rate. A drawback of such a 'benchmark' is that the discussion on definitions for entrepreneurial more and more becomes a theoretical one, Obviously such a theoretical discussion can still be valuable, but the GEM dominance has created a sort of fait accompli. Interestingly enough, it is apparently hardly challenged as little critique on the GEM methodology can be found.

2.1.2 Intrapreneurial Activity

There are similar discussions regarding intrapreneurial activity. Bjørnskov and Foss (2016) among others point to the difficulties regarding the definition and measurement of entrepreneurship. Some, like Sharma and Chrisman (1999) apply a broader definition of what they call corporate entrepreneurship, but Bjørnskov and Foss (2016) also stricter definitions.

2.2 Institutional Conditions and Entrepreneurial Activity

In order to understand the determinants for entrepreneurial activity, quite some research has been conducted on these determinants. Bjørnskov & Foss (2016) state that national institutions do indeed affect the levels of entrepreurship. These institution can be determined are

ID → TEA: Crnogaj and Hojnic investigated institutional determinants of entrepreneurial activity

ID → TEA: Terjesen et al, and Bjornskov comparative studies

ID → TEA: Bowen & de Clercq looked into institutions that predict high growth entrepreneurship (using GEM)

Bosma et al: One of the first international comparative studies of intrapreneurship and independent entrepreneurship. However, did not look at difference in institutional determinants.

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 & De Castro (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 world [1]. 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 business [2].

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 calculations [3]. 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,

[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

the variable 'IPACLTD_ALL' has been selected as the proxy for intrapreneurial activity. This variables 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 is 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'.

[5] Retrieved from http://gem-consortium.ns-client.xyz/about/wiki. last accessed on June 09th 2020
[6] For an non-exhaustive overview of papers that have used GEM data. see PDF: https://www.gemconsortium.org/research-papers-export%2Ffull&usg=AOvVaw188wwIwYKMMDjWU4f5VwIc
[7] For 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 an 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.

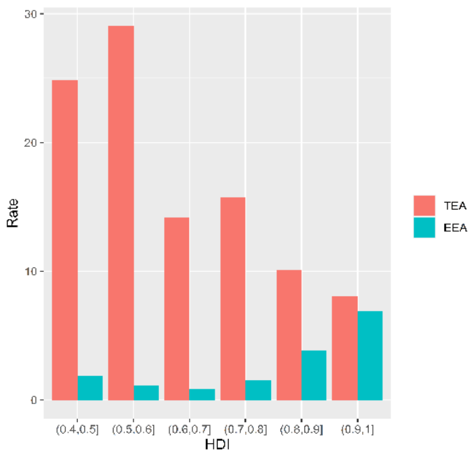
3.1.3 Country Development Indicator

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 202http://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. 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 variables 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 dataset led to a loss of a 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.

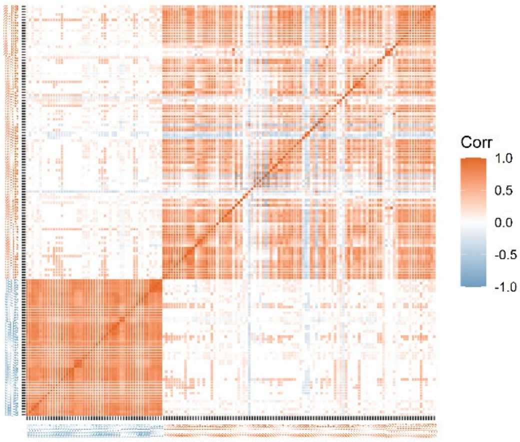
Merging GCI data was more cumbersome due to the presence of 2259 missing values. After identifying several rows and columns with a high number of missing values, all rows that contained many missing values (more than 20% of total values) were removed, leaving only 103 missing values. Several imputation 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 ≤ 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 are 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 can cope well with highly correlating variables and 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 an 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 ≪ 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:

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

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

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

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 minimize 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 looks as follows:

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

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 ≪ p datasets.

Thus, both methods have significant drawbacks; Ridge cannot shrink coefficients to zero and Lasso cannot cope well with correlating variables and n ≪ 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:

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

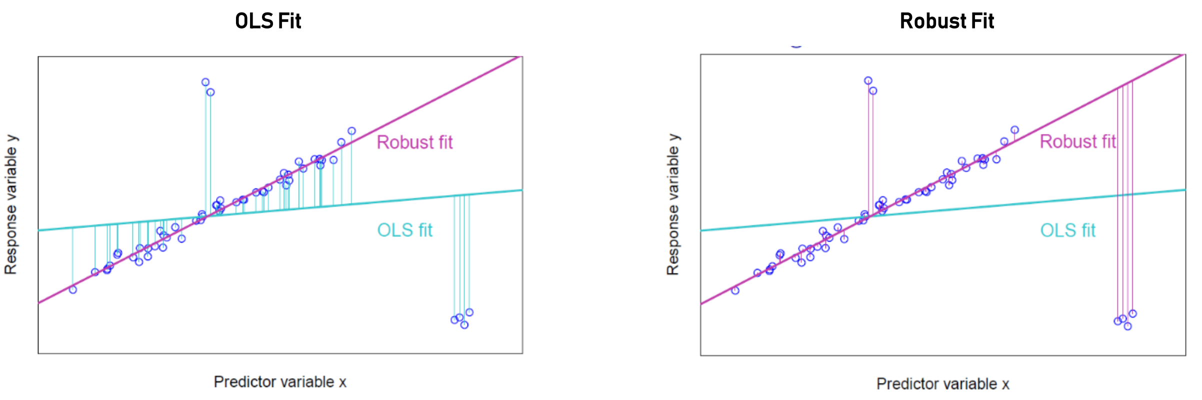
As with the regularization parameters, the optimal tuning parameter α is found by applying cross validation. Real world data and a simulation study show that the elastic net often outperforms the lasso, while enjoying a similar sparsity of representation (Zou & Hastie, 2005).

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 ≪ 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 ≪ p issue, outliers and uninformative variables can have relative 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 only. 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 approach [11]. The procedure includes cross-validation for parameter optimization and repeats this cross-validation for more stable results. The reweighing 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`

[11] For 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` [12][13].

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 reproducibly 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

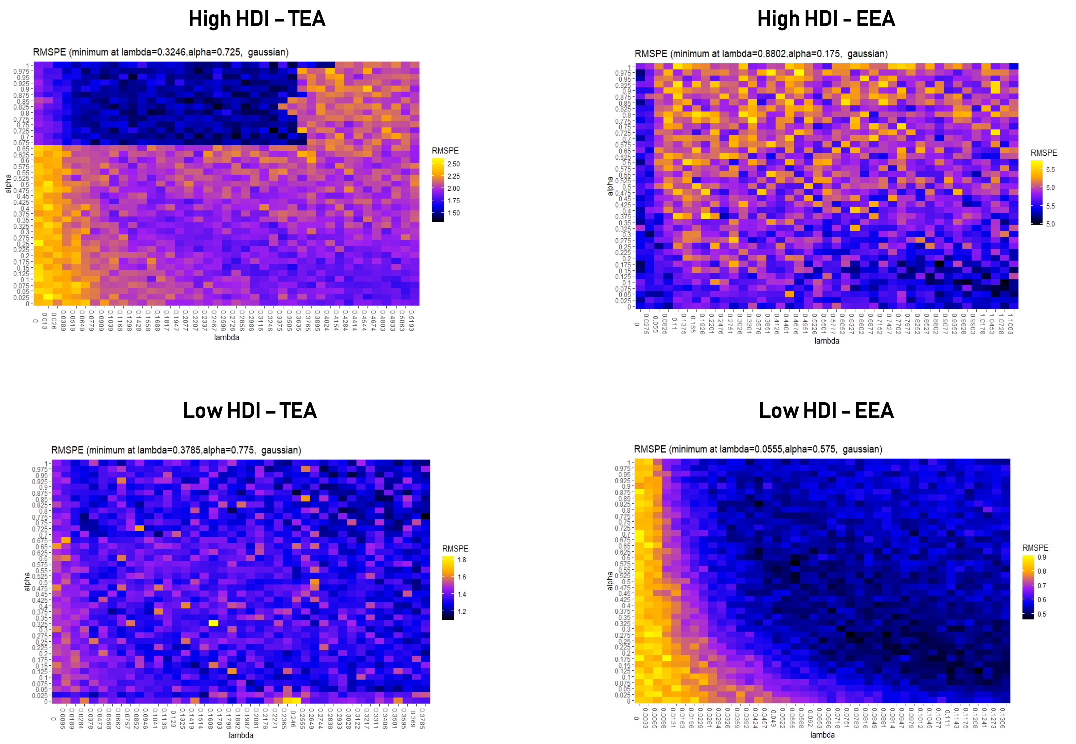
4.1 Results

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

[12] For documentation of the glmnet |R|-package, see: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf
[13] For documentation of the elasticnet |R|-package, see: https://cran.r-project.org/web/packages/elasticnet/elasticnet.pdf

combinations of lambdas and alphas for each pf the four models (metric: root mean square prediction error, or RMSPE).

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

Conducive 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

Conducive Variables for TEA and EEA for Developed Countries

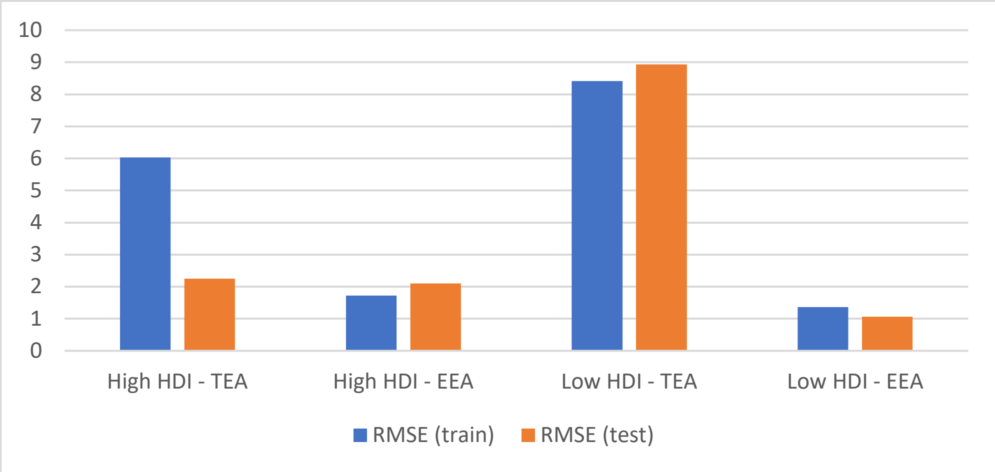
Code	Variable Description	Coef. TEA	Coef. EEA
GCI_2.05	Quality of air transport infrastructure	- 0.4655406759	- 0.099337612
GCI_5.06	Internet access in schools	0.0511977501	0.114413640
GCI_9.03	FDI and technology transfer	0.8791185479	0.218281439

Conducive 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 the 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 then the entrepreneurial framework conditions, which can be considered rather surprising.

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. 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, individual variables will, unfortunately, not be interpreted 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, the finding will also not be compared to findings of previous research. This is unfortunate, as it being one of the aims of this research, but comparing found determinants would not be a responsible approach, considering the uncertainty of the found results.

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

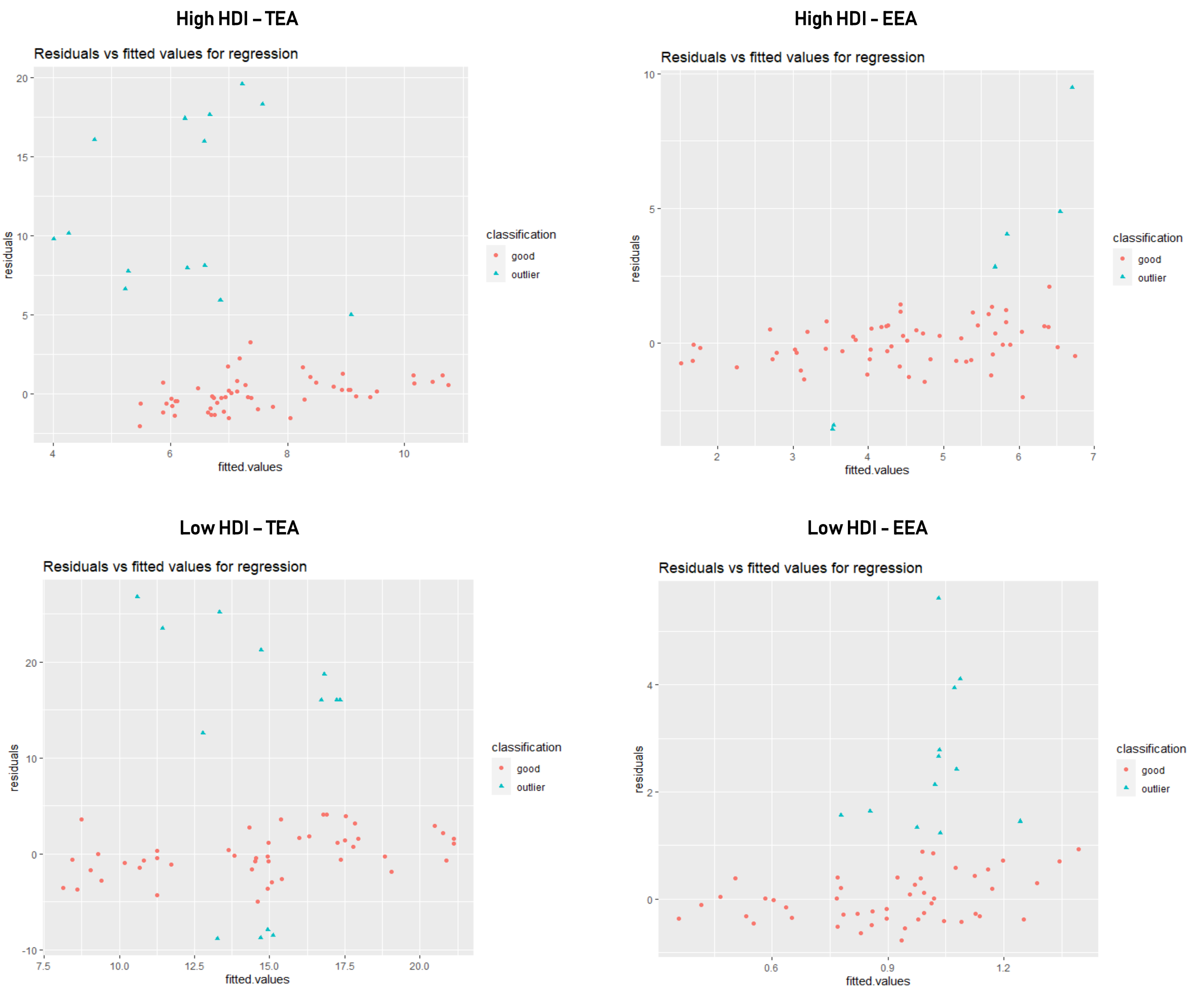
Model	Coefficients (n)	RMSE (train)	RMSE (test)	Outliers
All HDI - TEA	104	6.81	6.97	26
All HDI - EEA	33	1.41	1.54	5
High HDI - TEA	18	6.03	2.25	14
High HDI - EEA	17	1.72	2.10	6
Low HDI - TEA	34	8.41	8.93	12
Low HDI - EEA	13	1.36	1.06	9

The Elastic Net procedure only outputs a list of significant coefficients 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 the 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.

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

This study has for the first time applied Elastic Net regression for feature selection in the field of entrepreneurship research. The robustness of the method strengthens the quality of the findings, While the procedure has produced validated models with a good fit and a limited, interpretable number of variables deemed conducive, interpreting the variables, and hence assess the differences between determinants for TEA and EEA is practical not feasible. However, 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 settings of the models. As such, interpretation of the results is tricky and should 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 too uncertain for exact feature selection. Does the number of mobile phone subscriptions for example truly have a negative impact on EEA rates?

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 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 possible lead to less 'wild' results. Unfortunately the GEM 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, but 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 including 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, 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 not be answered in a satisfactory manner. Although the desired dataset could be constructed and the elastic net procedure has been executed as planned, meaningful interpretation of the variables was not considered to be appropriate and responsible. As such, no valid conclusions can, and should, be made based on the found determinants and, consequentially, a comparison with previous research is not feasible.

The found predictors for TEA and EEA, on both development levels are however so different, that it seems appropriate to conclude that the conducive determinants for both do seem to differ significantly. This is an indication that more research into EEA should be considered.

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 adaptation, 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. BIBLIOGRAPHY

Antoncic, Bostjan & Hisrich, Robert. (2003). Clarifying the Intrapreneurship Concept. Journal of Small Business and Enterprise Development. 10. 7-24. 10.1108/14626000310461187.

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.

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, N.S., Jones, K., Autio, E. & Levie, J. (2008), Global Entrepreneurship Monitor. 2007 Global Report.

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

Crnogaj, Katja & Bradac Hojnik, Barbara. (2016). Institutional determinants and entrepreneurial action. Management: Journal of Contemporary Management Issues. 21. 131-150.

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

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.

Kemal Dervis & Jeni Klugman, 2011. "Measuring human progress: the contribution of the Human Development Index and related indices," Revue d'économie politique, Dalloz, vol. 121(1), pages 73-92.

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

Levie, Jonathan and Ali, Abdul and Amoros, Ernesto and Hart, Mark and Kelley, Donna and Morris, R and Drexel, M and Eltobgy, M and Gratzke, P (2015) *Leveraging Entrepreneurial Ambition and Innovation : A Global Perspective on Entrepreneurship, Competitiveness and Development.* [Report]

Lockhart, Richard & Taylor, Jonathan & Tibshirani, Ryan & Tibshirani, Robert. (2013). A significance test for the LASSO. The Annals of Statistics. 42. 10.1214/13-AOS1175.

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.

Schumpeter, Joseph A., Capitalism, Socialism, and Democracy (1942). University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

Shane, S., & Venkataraman, S. (2000). The Promise of Entrepreneurship as a Field of Research. *The Academy of Management Review*, *25*(1), 217–226. Retrieved from www.jstor.org/stable/259271

Sharma, Pramodita & Chrisman, James. (1999). Toward A Reconciliation of the Definitional Issues in the Field of Corporate Entrepreneurship. Entrepreneurship: Concepts, Theory and Perspective. 23. 10.1007/978-3-540-48543-8_4.

Terjesen, S., Hessels, J., & Li, D. (2013). Comparative International Entrepreneurship. Journal of Management, 42(1), 299–344. https://doi.org/10.1177/0149206313486259

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.

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		
	GCI_1.01	Property rights
	GCI_1.02	Intellectual property protection
	GCI_1.03	Diversion of public funds
	GCI_1.04	Public trust in politicians
	GCI_1.05	Irregular payments and bribes
	GCI_1.06	Judicial independence
	GCI_1.07	Favouritism in decisions of government officials

6. GOODS MARKET EFFICIENCY		
	GCI_6.01	Intensity of local competition
	GCI_6.02	Extent of market dominance
	GCI_6.03	Effectiveness of anti-monopoly policy
	GCI_6.05	Total tax rate
	GCI_6.06	Number of procedures required to start a business
	GCI_6.07	Time required to start a business
	GCI_6.08	Agricultural policy costs
	GCI_6.09	Prevalence of trade barriers
	GCI_6.10	Trade tariffs
	GCI_6.11	Prevalence of foreign ownership
	GCI_6.12	Business impact of rules on FDI
	GCI_6.13	Burden of customs procedures
	GCI_6.14	Imports as a percentage of GDP

GCI_1.08	Wastefulness of government spending
GCI_1.09	Burden of government regulation
GCI_1.10	Efficiency of legal framework in settling disputes
GCI_1.11	Efficiency of legal framework in challenging regulations
GCI_1.12	Transparency of government policymaking
GCI_1.13	Business costs of terrorism
GCI_1.14	Business costs of crime and violence
GCI_1.15	Organized crime
GCI_1.16	Reliability of police services
GCI_1.17	Ethical behaviour of firms
GCI_1.18	Strength of auditing and reporting standards
GCI_1.19	Efficacy of corporate boards
GCI_1.20	Protection of minority shareholders' interests
GCI_1.21	Strength of investor protection

2. INFRASTRUCTURE

GCI_2.01	Quality of overall infrastructure
GCI_2.02	Quality of roads
GCI_2.03	Quality of railroad infrastructure
GCI_2.04	Quality of port infrastructure
GCI_2.05	Quality of air transport infrastructure
GCI_2.06	Available airline seat kilometres
GCI_2.07	Quality of electricity supply
GCI_2.08	Mobile telephone subscriptions
GCI_2.09	Fixed telephone lines

3. MACROECONOMIC ENVIRONMENT

GCI_3.01	Government budget balance
GCI_3.02	Gross national savings
GCI_3.03	Inflation
GCI_3.04	Government debt

4. HEALTH AND PRIMARY EDUCATION

GCI_4.03	Business impact of tuberculosis
GCI_4.04	Tuberculosis incidence
GCI_4.05	Business impact of HIV/AIDS
GCI_4.06	HIV prevalence
GCI_4.07	Infant mortality
GCI_4.08	Life expectancy
GCI_4.09	Quality of primary education

5. HIGHER EDUCATION AND TRAINING

GCI_5.01	Secondary education enrollment rate
GCI_5.02	Tertiary education enrollment rate
GCI_5.03	Quality of the educational system
GCI_5.04	Quality of math and science education
GCI_5.05	Quality of management school
GCI_5.06	Internet access in schools
GCI_5.07	Local availability of specialized research and training services
GCI_5.08	Extent of staff training

GCI_6.15	Degree of customer orientation
GCI_6.16	Buyer sophistication

7. LABOR MARKET EFFICIENCY

GCI_7.01	Cooperation in labor-employer relations
GCI_7.02	Flexibility of wage determination
GCI_7.03	Hiring and firing practices
GCI_7.04	Redundancy costs
GCI_7.06	Pay and productivity
GCI_7.07	Reliance on professional management
GCI_7.10	Female participation in labor force

8. FINANCIAL MARKET DEVELOPMENT

GCI_8.03	Financing through local equity market
GCI_8.04	Ease of access to loans
GCI_8.05	Venture capital availability
GCI_8.06	Soundness of banks
GCI_8.07	Regulation of securities exchanges
GCI_8.08	Legal rights index

9. TECHNOLOGICAL READINESS

GCI_9.01	Availability of latest technologies
GCI_9.02	Firm-level technology absorption
GCI_9.03	FDI and technology transfer
GCI_9.04	Internet users
GCI_9.05	Broadband Internet subscriptions
GCI_9.06	Internet bandwidth

10. MARKET SIZE

GCI_10.01	Domestic market size index
GCI_10.02	Foreign market size index
GCI_10.03	GDP (PPP$ billions)
GCI_10.04	Exports as a percentage of GDP

11. BUSINESS SOPHISTICATION

GCI_11.01	Local supplier quantity
GCI_11.02	Local supplier quality
GCI_11.03	State of cluster development
GCI_11.04	Nature of competitive advantage
GCI_11.05	Value chain breadth
GCI_11.06	Control of international distribution
GCI_11.07	Production process sophistication
GCI_11.08	Extent of marketing
GCI_11.09	Willingness to delegate authority

12. R&D INNOVATION

GCI_12.01	Capacity for innovation
GCI_12.02	Quality of scientific research institutions
GCI_12.03	Company spending on R&D
GCI_12.04	University-industry collaboration in R&D
GCI_12.05	Government procurement of advanced technology products
GCI_12.06	Availability of scientists and engineers

8.2 Appendix B: Correlation Matrix of Independent Variable

Correlation Matrix of Independent Variables

Appendix C: Composition of Dataset

Composition of Dataset

Region	Country	Year	TEA	EEA	HDI
N AMERICA	Canada	2014	13,03	4,75	0.914
		2015	14,72	7,06	0.917
	United States	2014	13,81	6,46	0.915
		2015	11,88	6,96	0.917
LATIN AMERICA & CARIBBEAN	Argentina	2011	20,77	3,20	0.823
		2014	14,40	2,42	0.825
	Brazil	2011	14,88	1,01	0.730
		2015	20,98	0,98	0.755
	Chile	2011	23,68	3,45	0.812
		2012	22,58	4,79	0.818
		2013	24,33	3,57	0.830
		2014	26,82	5,13	0.834
		2015	25,92	5,22	0.839
	Colombia	2011	21,43	1,66	0.735
		2014	18,54	3,69	0.750
		2015	22,67	2,31	0.753
	Costa Rica	2012	15,04	1,91	0.774
		2014	11,32	0,53	0.785
	Equador	2012	26,61	0,89	0.740
		2013	35,97	1,87	0.751
	Jamaica	2011	13,70	0,19	0.723
	Mexico	2011	9,62	0,93	0.746
		2012	12,10	1,23	0.752
		2014	18,99	0,71	0.754
		2015	21,00	1,16	0.759
	Panama	2011	20,78	0,15	0.764
		2014	17,05	0,19	0.779
		2015	12,80	0,50	0.782
	Peru	2011	22,88	1,35	0.734
		2012	20,20	1,71	0.734
		2013	23,38	0,85	0.742
		2014	28,81	1,67	0.752
		2015	22,21	0,72	0.750
	Uruguay	2011	16,72	5,20	0.783
		2012	14,63	5,01	0.788
		2013	14,07	6,64	0.797
		2014	16,07	3,81	0.800
		2015	14,28	4,19	0.802

Region	Country	Year	TEA	EEA	HDI
EUROPE	Austria	2014	8,71	5,62	0.904
	Belgium	2012	5,19	7,25	0.905
		2014	5,40	5,41	0.910
		2015	6,24	6,12	0.913
	Bulgaria	2015	3,45	0,35	0.807
	Croatia	2012	8,27	4,43	0.820
	Denmark	2012	5,36	12,61	0.924
		2014	5,47	11,43	0.928
	Estonia	2012	14,25	7,29	0.859
		2014	9,43	3,58	0.865
		2015	13,13	6,29	0.871
	Finland	2013	5,28	5,73	0.916
		2014	5,63	4,45	0.918
		2015	6,58	5,83	0.919
	France	2011	5,73	4,74	0.876
		2014	5,34	3,80	0.887
	Germany	2014	5,27	4,44	0.930
		2015	4,69	4,51	0.933
	Greece	2011	7,95	1,61	0.853
		2012	6,50	1,63	0.855
		2014	7,85	0,76	0.866
		2015	6,74	1,03	0.868
	Hungary	2012	9,21	2,13	0.826
		2013	9,67	3,61	0.835
		2014	9,33	1,80	0.833
		2015	7,92	2,12	0.835
	Ireland	2011	7,24	5,91	0.894
		2014	6,53	6,66	0.920
		2015	9,32	6,60	0.926
	Italy	2014	4,41	0,76	0.874
		2015	4,86	1,37	0.875
	Latvia	2011	11,85	3,02	0.822
		2015	14,10	3,33	0.842
	Lithuania	2011	11,26	4,86	0.831
		2012	6,68	5,79	0.835
		2013	12,42	6,79	0.840
		2014	11,32	5,08	0.852
	Luxembourg	2014	7,13	7,26	0.895
		2015	10,18	6,43	0.899

Region	Country	Year	TEA	EEA	HDI
EUROPE (CONT.)	Netherlands	2011	8,20	7,83	0.922
		2014	9,45	6,99	0.925
		2015	7,20	6,27	0.927
	Norway	2014	5,65	7,90	0.945
		2015	5,65	9,89	0.948
	Poland	2011	9,03	2,82	0.840
		2013	9,27	4,27	0.851
		2014	9,21	3,36	0.853
		2015	9,20	4,04	0.858
	Portugal	2011	7,53	4,02	0.827
		2014	9,97	3,23	0.840
		2015	9,48	3,95	0.843
	Romania	2012	9,21	3,49	0.796
		2013	10,12	5,05	0.800
		2014	11,34	4,94	0.803
		2015	10,82	4,60	0.806
	Russia	2011	4,56	0,60	0.789
		2014	4,68	0,51	0.807
	Slovakia	2012	10,21	4,59	0.841
		2013	9,52	4,35	0.844
	Slovenia	2012	5,42	5,87	0.876
		2013	6,45	6,09	0.884
		2014	6,32	4,73	0.886
		2015	5,91	5,59	0.886
	Spain	2011	5,81	2,71	0.870
		2014	5,47	1,82	0.880
		2015	5,69	1,09	0.885
	Sweden	2011	5,79	16,18	0.906
		2014	6,70	5,77	0.929
		2015	7,16	6,36	0.932
	Switzerland	2011	6,58	4,62	0.932
		2014	7,11	6,06	0.939
		2015	7,30	6,49	0.943
	United Kingdom	2011	7,28	5,24	0.899
		2012	8,97	10,07	0.897
		2013	7,13	8,51	0.914
		2014	10,65	6,99	0.918
		2015	6,93	4,07	0.916

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

Conducive 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